



senie 09

V Semana Nacional de Ingeniería Electrónica

V Semana Nacional de Ingeniería Electrónica

Memorias del Congreso

**Universidad de Guadalajara
Centro Universitario de la Ciénega**

**7 al 9 de Octubre de 2009
Ocotlán, Jal. México**

ISBN 978-607-477-073-5

Edición

Universidad Autónoma Metropolitana-Azcapotzalco
Av. San Pablo 180 Col. Reynosa Tamaulipas
Del. Azcapotzalco, 02200 México D.F.

Co-Edición

Universidad de Guadalajara
Centro Universitario de la Ciénega
Av. Universidad 1115, Col. Lindavista.
47820 Ocotlán, Jal. México.

Autor

Varios

Editores

Mildreth Alcaraz Mejía
Raúl Campos Rodríguez
Miguel Magos Rivera

V Semana Nacional de Ingeniería Electrónica
Memorias del Congreso
ISBN 978-607-477-073-5

México D.F. 2009

Comité Organizador

Mtra. Paloma Ibañez Villalobos

Rectora

Universidad Autónoma Metropolitana – Azcapotzalco

Dr. Raúl Medina Centeno

Rector

Centro Universitario de La Ciénega

Universidad de Guadalajara

Dr. Emilio Sordo Zabay

Director de la División de Ciencias Básicas e Ingeniería

Universidad Autónoma Metropolitana – Azcapotzalco

Mtro. Pedro Javier Guerrero Medina

Director de la División de Desarrollo Bio-Tecnológico

Centro Universitario de La Ciénega

Universidad de Guadalajara

Dr. Andrés Ferreyra Ramírez

Jefe del Departamento de Electrónica

Universidad Autónoma Metropolitana – Azcapotzalco

Mtro. Eduardo René Carrillo Íñiguez

Jefe del Departamento de Ciencias Tecnológicas

Centro Universitario de la Ciénega

Universidad de Guadalajara.

Dr. Miguel Magos Rivera

Presidente Comité Organizador

Universidad Autónoma Metropolitana - Azcapotzalco

Dr. Raúl Campos Rodríguez

Presidente Comité Local

Centro Universitario de la Ciénega

Universidad de Guadalajara.

Dra. Mildreth Alcaraz Mejía

Coordinador Local

Centro Universitario de la Ciénega

Universidad de Guadalajara.

Dr. Ramiro Velázquez Guerrero

Coordinador Logístico

Universidad Panamericana - Aguascalientes

Comité Técnico

Dra. Abascal Mena Ma. Del Rocío	UAM - Cuajimalpa
Dr. Adam Medina Manuel	CENIDET - Cuernavaca
Dra. Alcaraz Mejía Mildreth Isadora	Centro Universitario de la Ciénega - UdG
Dr. Astorga Zaragoza Carlos	CENIDET - Cuernavaca
Dr. Aviles Cruz Carlos	UAM - Azcapotzalco
Dr. Campos Rodríguez Raúl	Centro Universitario de la Ciénega – UdG
Dr. Cano González Mario Eduardo	Centro Universitario de la Ciénega – UdG
Dr. Ferreyra Ramírez Andrés	UAM - Azcapotzalco
Dr. Incera Diéguez José Alberto D.	Instituto Tecnológico Autónomo de México
Dr. Jiménez Alaniz Juan Ramón	UAM - Iztapalapa
Dr. López Ornelas Erick	UAM - Cuajimalpa
Dr. López Orozco Omar	ITESM - Morelia
M.en C. Macías Quijas Ricardo	Universidad Panamericana - Aguascalientes
Dr. Magos Rivera Miguel	UAM - Azcapotzalco
Dr. Mota Grajales Rafael	Instituto Tecnológico de Tuxtla Gutiérrez
Dr. Orozco Mora Jorge Luis	Instituto Tecnológico de Aguascalientes
Dr. Ramírez Treviño Alberto	Instituto Tecnológico Superior de Cajeme
Dr. Ramírez Treviño Antonio	Cinvestav - Guadalajara
Dr. Rodríguez Calderón Rosalino	ITESM - Morelia
Dr. Rodríguez Segura Elías José J.	Instituto Tecnológico de Celaya
Dra. Ruiz Beltrán Elvia	Instituto Tecnológico de Aguascalientes
M.en C. Varona Salazar Jorge	Snowbush - Aguascalientes
Dr. Velázquez Guerrero Ramiro	Universidad Panamericana - Aguascalientes
Dr. Zárate Águila Oscar Antonio	Centro Universitario de la Ciénega - UDG
Dr. Zavala Río Arturo	Instituto Potosino de Investigación Científica y Tecnológica

Presentación

Con la realización de la V Semana Nacional de Ingeniería Electrónica, el Departamento de Electrónica de la Universidad Autónoma Metropolitana Azcapotzalco y el Departamento de Ciencias Tecnológicas del Centro Universitario de la Ciénega de la Universidad de Guadalajara cumplen con el objetivo de proporcionar un foro para la discusión de ideas, la comparación e intercambio de resultados y la divulgación de los trabajos de investigación relacionados con la Ingeniería Electrónica y sus diversas ramas.

En esta ocasión el evento se lleva a cabo en las instalaciones del Centro Universitario de la Ciénega, institución de educación pública de reconocido prestigio ubicada en el centro del país.

Las memorias incluyen 83 artículos en extenso, los cuales fueron seleccionados por el Comité Técnico de entre más de 100 trabajos enviados por autores de todo el país. Cabe señalar que se realizó un esfuerzo importante en el arbitraje de los trabajos con el objetivo de asegurar la calidad de los mismos. Nuestro reconocimiento a los miembros del Comité Técnico por la labor realizada.

Se programaron 3 Conferencias Magistrales impartidas por especialistas, las ramas consideradas fueron: Microsistemas, Diseño de Microprocesadores y Metrología Electrónica.

Finalmente, agradecemos a los miembros del Comité Organizador Local por su esfuerzo para lograr que este evento se realice en las mejores condiciones.

Ocotlán, Jal., octubre de 2009.

Comité Organizador de la V Semana Nacional de Ingeniería Electrónica.

Miércoles 7 de Octubre

Horario	Sala 1	Sala 2	Sala 3
9:00 – 10:00		Inscripción	
10:00 – 10:30		Inauguración	
10:30 – 10:40		Receso	
10:40 – 11:40		Magistral 1 Sala 1 Microsystems Validation at Sandia National Labs <i>Dr. Anton Sumali</i> Sandia National Laboratories, Albuquerque, Nuevo México	
11:40 – 12:00		Receso	
12:00 – 13:40	<i>Análisis del comportamiento mecánico de acelerómetros piezoresistivos uniaxiales</i> Ángel Roberto Cortés Pérez Universidad de Guanajuato	<i>Empleo de series de Fourier en el modelado del caudal de un río</i> Joel Herrera Cabral Universidad Tecnológica del Suroeste de Guanajuato Universidad de Guanajuato	<i>El seguimiento de la tutoría personalizada en la Universidad de Colima mediante tecnologías de información</i> Ricardo Fuentes Covarrubias Universidad de Colima
	<i>Micropinza MEM's basada en actuación electroestática</i> Emilio Sáenz Irigoyen Universidad Panamericana-Bonaterra	<i>Revisión del estado del arte de las estufas de cocción bajo el principio de calentamiento por inducción</i> José Ángel Chávez García Instituto Tecnológico de Celaya	<i>Sistema de selección de horarios para profesores SiSHoP</i> María del Carmen Nolasco Salcedo Centro Universitario de la Ciénega Universidad de Guadalajara
	<i>Caracterización de un acelerómetro micromaquinado con tecnología POLYMUMP's</i> Rodolfo Iván Rincón Jara Universidad Autónoma de Ciudad Juárez	<i>Diseño y construcción de los sistemas de potencia y control para un robot industrial tipo cartesiano</i> Sergio Vergara Limón Benemérita Universidad Autónoma de Puebla	<i>Tecnologías de bajo costo para medir y mejorar la motricidad en los niños "Adventures in the jungle"</i> Oscar David Quezada Patiño Universidad Panamericana-Bonaterra
			<i>Uso de realidad aumentada para apoyar el aprendizaje en niños de 2 a 5 años (Juego de memoria)</i> Luis Quezada Patiño Universidad Panamericana-Bonaterra
13:40 – 15:00	COMIDA		
15:00 – 16:40	<i>Diseño de micro y macro espejos de actuación electrostática</i> Samuel Rosales Sánchez Centro Universitario de la Ciénega Universidad de Guadalajara	<i>Arquitectura genérica para el procesamiento de múltiples imágenes en tiempo real basada en FPGA</i> Juan Manuel Xicoténcatl Pérez Universidad Politécnica de Pachuca	<i>Generación de ambientes virtuales mediante modelado declarativo basado en conocimiento</i> Jaime Alberto Zaragoza Ríos Cinvestav Guadalajara
	<i>Arreglo de micro-engranes planetarios</i> Alejandra Aceves Paredes Centro Universitario de la Ciénega Universidad de Guadalajara	<i>Primera etapa de un sistema CAD para el análisis de mamografías</i> Arturo Zúñiga López Universidad Autónoma Metropolitana-Azcapotzalco	<i>Visualización Distribuida en Tiempo Real de un Ambiente Virtual</i> Alma Verónica Martínez González Cinvestav Guadalajara
	<i>Metodología para el diseño de sistemas embebidos basada en modelos de redes de Petri</i> Laura García Contreras Centro Universitario de la Ciénega Universidad de Guadalajara	<i>Implementación de algoritmos computacionales para la identificación de personas mediante la lectura del iris</i> Andrés Gerardo Fuentes Covarrubias Universidad de Colima	<i>MPISTE: Un entorno para contar historias basadas en ubicación de forma personalizada en dispositivos móviles</i> Carlos Delgado Mata Universidad Panamericana-Bonaterra
		<i>Teoría de la Resonancia Adaptiva para el análisis del aprendizaje, utilizando redes neuronales artificiales ARTI</i> Alejandro Cruz Sandoval Universidad Autónoma Metropolitana-Azcapotzalco	<i>Conducta influenciada por la personalidad en avatares ante situaciones de desastre: Una propuesta apoyada en la psicología</i> Víctor Fernández Cervantes Cinvestav Guadalajara
16:40 – 16:50	Receso		
16:50 – 18:30	<i>Corrección Cromática de Imágenes Naturales</i> Pedro Alfonso Patlán Rosales Universidad de Guanajuato	<i>Transitorios No-Lineales aplicados a un servomecanismo difuso</i> Domingo Rangel Miranda Centro de Física Aplicada y Tecnología Avanzada. UNAM	<i>Diseño de 3 diferentes tipos de aerogeneradores para uso urbano y domestico</i> José M. Carlos Pérez Aguirre Benemérita Universidad Autónoma de Puebla
	<i>Inspección de tarjetas electrónicas mediante técnica de visión por computadora</i> José Jorge Hernández Constante Centro Universitario de Ciencias Exactas e Ingenierías Universidad de Guadalajara	<i>Nonlinear observer for lateral velocity using roll angular acceleration</i> Cuauhtémoc Acosta Lúa University of L'Aquila	<i>Análisis técnico-económico de factibilidad de generación de energía eólica en Celaya, Guanajuato</i> Joel Herrera Cabral Universidad Tecnológica del Suroeste de Guanajuato Universidad de Guanajuato
	<i>Implementación óptima de un filtro espacial para mejorar el desempeño de clasificadores de señales cerebrales en interfaces cerebro-computadora</i> Diana Irazú Escalona Vargas Cinvestav Monterrey	<i>Análisis fraccional para la reducción de orden del modelo matemático de un autobús</i> Iván Vázquez Álvarez Universidad Autónoma Metropolitana-Azcapotzalco	<i>Diseño e implementación de un circuito MPPT para sistemas fotovoltaicos autónomos</i> Fernando Ponce Márquez Universidad Autónoma de Zacatecas
	<i>Aplicación de diferentes técnicas para el reconocimiento de patrones de trayectorias en un espacio euclidiano con un WiiMote</i> Blanca Miriam Lee Cosío Universidad Panamericana Campus Bonaterra		<i>Caracterización de baterías y paneles de un sistema fotovoltaico autónomo</i> Luis Manuel Segura Flores Universidad Autónoma de Zacatecas

Jueves 8 de Octubre

Horario	Sala 1	Sala 2	Sala 3
9:00 – 10:15	<i>Identificación de sistemas No Lineales usando redes dinámicas recurrentes</i> Juan Anzurez Marin Universidad Michoacana de San Nicolás de Hidalgo Facultad de Ingeniería Eléctrica	<i>Sistema de posicionamiento autocontrolable por la intensidad de un haz laser</i> Edgar Eduardo Antunez Ceron Centro de Investigación en Ingeniería y Ciencias Aplicadas. INAOE	<i>Evaluación electroacústica de iPods</i> Ilhuicamina Trinidad Servín Rivas Escuela Superior de Ingeniería Mecánica y Eléctrica. Zacatenco Instituto Politécnico Nacional
	<i>PID Wavelet Auto-Sintonizado con una red neuronal Wavenet</i> José Alberto Cruz Tolentino Universidad Autónoma del Estado de Hidalgo	<i>Desarrollo de un prototipo para corte de placa metálica</i> Francisco Javier López Jarquez Universidad Autónoma de Ciudad Juárez	<i>Desarrollo de un selector de derivaciones electrocardiográficas basado en un arreglo de CMOS comerciales</i> Jorge Alberto García Vázquez Centro Universitario de la Ciénega Universidad de Guadalajara
	<i>Implementación de un mapa Auto-Organizativo de Kohonen en FPGA</i> Sergio Garduza González Cinvestav México D.F.	<i>Aplicación de las pruebas Maximin para la calidad de estabilización robusta a un robot móvil</i> Javier Cabrera Martínez Benemérita Universidad Autónoma de Puebla	<i>Sistema digital de temporización</i> Enrique Montoya Suárez Universidad Autónoma de Nayarit
10:15 – 10:25	Receso		
10:25 – 11:25	Magistral 2 Sala 1 Retos en el Diseño de Microprocesadores <i>Expositor por confirmar</i> Intel Labs, Intel Guadalajara		
11:25 – 11:30	Receso		
11:30 – 12:00	Conferencia Técnica Sala 1 Título por confirmar <i>Expositor por confirmar</i> CIELECTRONICA S.A. de C.V.		
12:00 – 12:10	Receso		
12:10 – 13:50	<i>Desarrollo de una herramienta de apoyo para el aprendizaje del microcontrolador 16F84</i> Juan Carlos Pérez Merlos Universidad Autónoma del Estado de México	<i>El problema del diagnóstico de fallas: Diseño de observadores dedicados</i> Juan Anzurez Marin Universidad Michoacana de San Nicolás de Hidalgo Facultad de Ingeniería Eléctrica	<i>Silla de ruedas robótica para personas con discapacidad motora severa</i> Enrique Preza Hernández Universidad Panamericana-Bonaterra
	<i>Simuladores de control secuencial usados para la educación a distancia</i> Alberto Ramiro Garibay Martínez Facultad de Ingeniería UNAM	<i>Monitor cardiaco móvil en teléfonos celulares para la detección de anomalías cardiacas</i> José Alberto Juárez García UPIITA Instituto Politécnico Nacional	<i>Diseño de un control óptimo en tiempo continuo para la estabilización de un robot móvil.</i> Lucio Hernández Angulo Benemérita Universidad Autónoma de Puebla
	<i>Modelado y control de un reactor químico, como parte del aprendizaje basado en estudios de caso</i> Francisco Daniel Soria Villegas Facultad de Ingeniería UNAM	<i>GeoBus: Sistema de orientación espacial para asistir a personas invidentes utilizando teléfonos celulares</i> Andrés Jaramillo Calderón Instituto Politécnico Nacional	<i>Diseño de un control óptimo discreto para un robot móvil</i> Aline Galindo Téllez Benemérita Universidad Autónoma de Puebla
		<i>Aplicación de la teoría de grafos para encontrar las distancias más cortas entre las ciudades de la República Mexicana mediante la implementación del algoritmo de Floyd</i> Mireya Cacho Ruiz Centro Universitario de la Ciénega Universidad de Guadalajara	<i>Formación de agentes móviles basada en estrategias de Auto-Organización</i> Miguel Ángel Sánchez Acevedo Cinvestav Guadalajara
13:50 – 16:00	COMIDA		
16:00	Visita a la fábrica de muebles <i>Lilico Muebles S.A. de C.V.</i>		
20:00	Cena en el hotel: <i>La Cabaña de Yeyo</i>		

Viernes 9 de Octubre

Horario	Sala 1	Sala 2	Sala 3
9:00 – 10:15	<i>Implantación de algoritmo de filtrado adaptable QR-RLS en FPGA</i> Mauricio Alberto Martínez-García Universidad la Salle, México	<i>Novedoso sensor refractométrico de fibras ópticas con elemento de detección semicilíndrico plástico, para medir el nivel en un tanque de gasolina o diesel</i> Selene Pérez García Facultad de Ingeniería UNAM	<i>About the path loss in the RFID radio channel</i> Salvador Ricardo Meneses González Escuela Superior de Ingeniería Mecánica y Eléctrica. I.P.N.
	<i>Plataforma FPGA para el diseño y evaluación de sistemas de control</i> Felipe Santiago Espinosa Universidad Tecnológica de la Mixteca	<i>Sistema de seguridad perimetral óptico "SSPO"</i> Rubén Herrera Galicia Instituto Tecnológico de Tuxtla Gutiérrez	<i>Análisis de propagación electromagnética mediante un sistema georeferenciado</i> Fabiola Martínez Zúñiga Sección de Estudios de Posgrado e Investigación ESIME IPN
	<i>Implementación de un sistema de medición ultrasónico de distancia utilizando programación en LabVIEW sobre FPGA</i> Anna Leonor Aguilar Torres Instituto Tecnológico de la Laguna		<i>Una aproximación de radio cognitiva para aplicaciones móviles</i> Salvador Silva López Instituto Tecnológico y de Estudios Superiores de Monterrey Campus Estado de México
10:15 – 10:30	Receso		
10:30 – 11:30	Magistral 3 Sala 1 Metrología Electrónica Dr. Horacio Estrada Vázquez Centro Nacional de Metrología		
11:30 – 11:40	Receso		
11:40 – 13:20	<i>Herramienta visual para analizar secuencias genómicas</i> Gustavo Verduzco Reyes Universidad Autónoma del Carmen	<i>VCO's Diferenciales para redes de generación y distribución de señales de reloj</i> Mónico Linares Aranda Instituto Nacional de Astrofísica, Óptica y Electrónica	<i>Diseño y construcción de un sistema con base en un módulo XBee aplicado en el monitoreo inalámbrico de datos</i> Jayro Santiago Paz Universidad Tecnológica de la Mixteca / Cinvestav
	<i>Esquema de gestión basado en calidad de servicio para distribución de contenido Media Streaming sobre redes P2P</i> Raymundo Buenostro Mariscal Universidad de Colima	<i>Síntesis de equalizadores de ganancia basados en rejillas de periodo largo para EDFAs</i> Alfredo Beltrán Hernández Facultad de Ingeniería UNAM	<i>Sistema de monitoreo de la planta tratadora de aguas residuales de la UAM-Azcapotzalco</i> Luis Francisco Soldevilla Juárez / Ariel Cruz Ramírez Universidad Autónoma Metropolitana-Azcapotzalco
	<i>Planes de prevención y recuperación de desastres (PPRyD) en redes WLAN</i> Federico Felipe Durán Escuela Superior de Ingeniería Mecánica y Eléctrica. Zacatenco Instituto Politécnico Nacional	<i>Instrumento virtual para la medición y análisis de las vibraciones mecánicas en máquinas rotativas</i> Julio Laria Menchaca Facultad de Ingeniería "Arturo Narro Siller" Universidad Autónoma de Tamaulipas	<i>Actuador lineal ultrasónico piezoeléctrico para pantallas táctiles</i> Hermes Hernández Mücke Universidad Panamericana-Bonaterra
	<i>Carnet médico digital para teléfonos celulares</i> Luis Enrique Bustos Méndez UPIITA- Instituto Politécnico Nacional		<i>Diseño y primer prototipo de mano robótica a base de materiales con memoria de forma</i> Claudia García Gallegos Universidad Panamericana-Bonaterra
13:20 – 15:00	COMIDA		
15:00 – 16:40	<i>Mini-robot tetrápodo de inspiración biológica: diseño, prototipo y modelado</i> Sergio Juárez Miramontes Universidad Panamericana-Bonaterra	<i>Sistema electrónico para el monitoreo a distancia de procesos de corrosión mediante la técnica de ruido electroquímico</i> Julio Laria Menchaca Facultad de Ingeniería "Arturo Narro Siller" Universidad Autónoma de Tamaulipas	<i>Herramienta visual para el desarrollo de software mediante abstracciones gráficas</i> Alejandro Medina Santiago Tecnológico de Estudios Superiores de Ecatepec
	<i>Una biblioteca para el modelado y simulación de la locomoción humana en robots humanoides</i> Héctor Rafael Orozco Aguirre Cinvestav-Guadalajara	<i>Diseño e implementación de una interfaz de registro de datos en tarjetas de memoria Flash SD Card, para la unidad sísmica SR04</i> Lauro Santiago Cruz Instituto de Ingeniería, UNAM	<i>Generación de protocolos de interacción empleando programación genética</i> José Octavio Gutiérrez García Cinvestav-Guadalajara
	<i>Algoritmo de planeación de movimientos para un brazo en 3D</i> Cristian Eduardo Boyain y Goytia Luna Cinvestav-Guadalajara	<i>Sistema de control y adquisición de datos para un probador de válvulas cardíacas</i> Miguel Ángel Bañuelos Saucedo Centro de Ciencias Aplicadas y Desarrollo Tecnológico UNAM	<i>Modelado y simulación de sistemas multi-agentes orientados a eventos</i> Emmanuel Lopez Neri Cinvestav-Guadalajara
		<i>Interfase para la adquisición de señales puras del equipo de ultrasonido Hitachi EUB-25</i> Carlos Avilés Cruz Universidad Autónoma Metropolitana-Azcapotzalco	
16:40 – 16:50	Receso		
16:50 – 17:50	Clausura y Brindis		

INDICE

Título de la Ponencia	Página
Aplicación de las pruebas maximin para la calidad de estabilización robusta a un robot móvil	1
Transitorios no-lineales aplicados a un servomecanismo difuso	9
Identificación de sistemas no lineales usando redes dinámicas recurrentes	16
PID Wavelet Auto-Sintonizado con una Red Neuronal Wavenet	23
El problema del diagnóstico de fallas: diseño de observadores dedicados	32
Nonlinear Observer for Lateral Velocity using Roll Angular Acceleration	39
GeoBus: sistema de orientación espacial para asistir a personas invidentes utilizando teléfonos celulares	47
About the Path Loss in the RFID Radio Channel	57
Análisis de propagación electromagnética mediante un sistema Georreferenciado	63
Una aproximación de Radio Cognitiva para aplicaciones móviles	69
Planes de prevención y recuperación de desastres (PPRyD) en Redes WLAN	76
VCO's diferenciales para redes de generación y distribución de señales de reloj	84
Síntesis de ecualizadores de ganancia basados en rejillas de periodo largo para EDFAs	93
Sistema de posicionamiento autocontrolable por la intensidad de un haz láser	101
Implantación del algoritmo de filtrado adaptable QR-RLS en FPGA	108
Plataforma FPGA para el diseño y evaluación de sistemas de control	117
Desarrollo de una herramienta de apoyo para el aprendizaje del microcontrolador 16f84A	127
Simuladores de control secuencial usados para la educación a distancia	134

Título de la Ponencia	Página
Modelado y control de un reactor químico, como parte del aprendizaje basado en estudios de caso	142
Diseño e implementación de un circuito MPPT para sistemas fotovoltaicos autónomos	149
Revisión del estado del arte de las estufas de cocción bajo el principio de calentamiento por inducción	157
Caracterización de baterías y paneles de un sistema fotovoltaico autónomo	167
Diseño de 3 diferentes tipos de aerogeneradores para uso urbano y domestico	174
Análisis técnico-económico de factibilidad de generación de energía eólica en Celaya, Guanajuato	184
Empleo de series de Fourier en el modelado del caudal de un río	191
Sistema Electrónico para el Monitoreo a Distancia de Procesos de Corrosión Mediante la Técnica de Ruido Electroquímico	197
Sistema digital de temporización	203
Instrumento virtual para la medición y análisis de las vibraciones mecánicas en máquinas rotativas	211
Monitor cardiaco móvil en teléfonos celulares para la detección de anomalías cardiacas	217
Diseño e Implementación de una Interfaz de registro de datos en tarjetas de memoria Flash SD Card, para la Unidad Sísmica SR04	226
Sistema de Seguridad Perimetral Óptico “SSPO”	236
Novedoso sensor refractométrico de fibras ópticas con elemento de detección semicilíndrico plástico, para medir el nivel en un tanque de gasolina o diesel	245
Sistema de control y adquisición de datos para un probador de válvulas cardíacas	254
Diseño y construcción de un sistema con base en un módulo XBee aplicado en el monitoreo inalámbrico de datos	260
Implementación de un sistema de medición ultrasónico de distancia utilizando programación en LabView sobre FPGA.	270
Interfase para la adquisición de señales puras del equipo de Ultrasonido Hitachi EUB-25	277

Título de la Ponencia	Página
Desarrollo de un selector de derivaciones electrocardiográficas basado en un arreglo de CMOS comerciales	285
Sistema de monitoreo de la planta tratadora de aguas residuales de la UAM-Azcapotzalco	292
Actuador lineal ultrasónico piezoeléctrico para pantallas táctiles	302
Desarrollo de un prototipo para corte de placa metálica	310
Análisis fraccional para la reducción de orden del modelo matemático de un autobús	317
Diseño y primer prototipo de mano robótica a base de Materiales con memoria de forma	330
Diseño y construcción de los sistemas de potencia y control para un robot industrial tipo cartesiano	339
Silla de ruedas robótica para personas con discapacidad motora severa	348
Diseño de un Control Óptimo en Tiempo Continuo para la Estabilización de un Robot Móvil	356
Mini-robot tetrapodo de inspiración biológica: diseño, prototipo y modelado	365
Diseño de un control óptimo discreto para un Robot Móvil	371
Una Biblioteca para el Modelado y Simulación de la Locomoción Humana en Robots Humanoides	378
Simulador 3D para robots manipuladores de 3 grados de libertad	388
Análisis del comportamiento mecánico de acelerómetros piezoresistivos uniaxiales	396
Micropinza MEMs basada en actuación electrostática	406
Diseño de un sistema acondicionador de señal para sensores MEMs de reactancia variable	412
Caracterización de un acelerómetro micromaquinado con tecnología POLYMUMP's	419
Diseño de micro y macro espejos de actuación electrostática	425
Arreglo de micro-engranes planetarios	435
Corrección cromática de imágenes naturales	443
Inspección de tarjetas electrónicas mediante técnica de visión por computadora	451

Título de la Ponencia	Página
Implementación óptima de un filtro espacial para mejorar el desempeño de clasificadores de señales cerebrales en interfaces cerebro-computadora	457
Aplicación de diferentes técnicas para el reconocimiento de patrones de trayectorias en un espacio euclidiano con un <i>Wimote</i>	467
Arquitectura genérica para el procesamiento de múltiples imágenes en tiempo real basada en FPGA	474
Implementación de algoritmos computacionales para la identificación de personas mediante la lectura del iris	480
Evaluación electroacústica de iPods	487
Primera etapa de un sistema CAD para el análisis de mamografías	493
teoría de la resonancia adaptiva para el análisis del aprendizaje, utilizando redes neuronales artificiales Art1.	499
esquema de gestión basado en calidad de servicio para distribución de contenido media streaming sobre redes P2P	506
Herramienta visual para analizar secuencias genómicas	516
Generación de ambientes virtuales mediante modelado declarativo basado en conocimiento	523
El seguimiento de la tutoría personalizada en la universidad de colima mediante tecnologías de información	533
Formación de agentes móviles basada en estrategias de auto-organización	540
Implementación de un mapa auto-organizativo de Kohonen en FPGA	547
Visualización distribuida en tiempo real de un ambiente virtual	555
Aplicación de la teoría de grafos para encontrar las distancias más cortas entre las ciudades de la república mexicana mediante la implementación del algoritmo de Floyd	564
MPISTE: “Un entorno para contar historias basadas en ubicación de forma personalizada en dispositivos móviles”	570
Carnet médico digital para teléfonos celulares	578

Título de la Ponencia	Página
Algoritmo de planeación de movimientos para un brazo en 3D	589
Sistema de selección de horarios para profesores SiSHoP	596
Conducta influenciada por la personalidad en avatares ante situaciones de desastre: una propuesta apoyada en la psicología	603
Herramienta visual para el desarrollo de software mediante abstracciones gráficas	612
Generación de protocolos de interacción empleando programación genética	621
Modelado y simulación de sistemas multi-agentes orientados a eventos	629
Tecnologías de bajo costo para medir y mejorar la motricidad en los niños “Adventures in the Jungle”	639
Uso de realidad aumentada para apoyar el aprendizaje en niños de 2 a 5 años (juego memoria)	646
Metodología para el diseño de sistemas embebidos basada en modelos de Redes de Petri	654

APLICACIÓN DE LAS PRUEBAS MAXIMIN PARA LA CALIDAD DE ESTABILIZACIÓN ROBUSTA A UN ROBOT MÓVIL

Gutiérrez Arias José Eligio Moisés, Cabrera Martínez Javier, Reyes Cortés José Fernando.

Facultad de Ciencias de la Electrónica, Benemérita Universidad Autónoma de Puebla
Av. San Claudio y Blvd. 18 Sur. Col. Jardines de San Manuel, Puebla, Pue. 72570, México
Teléfono: (222) - 2198878
jmgutierrez@ece.buap.mx / javier_cm4@ece.buap.mx, / freyes@ece.buap.mx

RESUMEN

Existen situaciones en que se desea estimar la calidad de estabilización de un sistema de control, en las cuales se desconoce cómo funciona el sistema de dicho control; pero conocemos la entrada y la salida del mismo, estas situaciones permiten considerar los sistemas de pruebas de calidad de la estabilización, como lo son las pruebas maximin. Las estrategias de pruebas maximin es una nueva dirección en la teoría de control óptimo y tienen su aplicación en la práctica de pruebas. Su objetivo es evaluar el desempeño de algoritmos de estabilización desconocidos, así, mediante un sistema de pruebas determinar cuan bueno es dicho algoritmo de control. En el presente trabajo, se evalúa el desempeño que presenta un controlador desconocido, el cual es aplicado a un robot móvil, utilizando para ello las pruebas maximin.

Palabras clave: Pruebas maximin, indicador de desempeño en la estabilización, estimación inferior, ecuación de Riccati.

I. INTRODUCCIÓN

Los sistemas automáticos de control surgen inicialmente para reemplazar a los humanos en la realización de tareas de control poco deseables, el perfeccionamiento de las técnicas empleadas permite obtener sistemas de control con una elevada precisión o rapidez. Hacia 1960 y como respuesta a las nuevas necesidades (plantas con múltiples entradas y salidas, elevada complejidad), surge la teoría moderna de control, la cual está basada en el análisis y la síntesis en el dominio del

tiempo empleando variables de estado. Los desarrollos más avanzados están representados por conceptos como el control óptimo, siendo éste, un tópico especial dentro de la disciplina de las ecuaciones diferenciales. Con el desarrollo del principio del máximo por Pontryagin (1958) y el método de la programación dinámica por Bellman (1957), actualmente, la teoría de control óptimo se considera como uno de los elementos de los métodos de optimización.

La idea de control puede ser expresada como el proceso mediante el cual se ejerce una influencia sobre el comportamiento de un objeto dinámico (que varía con el tiempo) para alcanzar un propósito previamente fijado [1]; pero si ahora, se desea lograr tal propósito en un tiempo mínimo o con un mínimo uso de recursos de control, entonces, este es un problema de control óptimo, ya que se quiere minimizar un funcional que depende del estado del sistema y del control, de esta manera, podemos decir que este funcional describe la calidad de estabilización.

En el establecimiento de las pruebas de estabilización en sistemas dinámicos, la solución de un problema del tipo maximin es considerado como el principio de las pruebas. El conocimiento o descripción de un algoritmo de estabilización no es necesario para la realización de un procedimiento de pruebas. Sólo se supone el uso de una señal de entrada y salida del algoritmo (esta situación tiene lugar en la práctica). Por ejemplo, es posible examinar un sistema de estabilización cuando un humano forma parte del mismo.

II. ESTABLECIMIENTO DE LAS PRUEBAS MAXIMIN.

Un sistema dinámico controlable está formado generalmente por cuatro partes como se observa en la figura 1, éstas actúan de tal forma que se produce un mecanismo de retroalimentación [2].

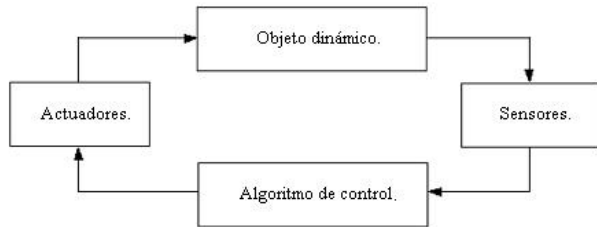


Figura 1: Diagrama a bloques general de un sistema dinámico controlable.

El problema de control estudia el proceso de llevar este sistema de un estado inicial arbitrario a un estado deseado tan rápido como sea posible aplicando para ello un algoritmo de control.

Para el establecimiento de las pruebas maximin, consideremos las ecuaciones en desviaciones siguientes del movimiento programado:

$$\begin{cases} \dot{x} = Ax + Bu + Cv, & x(t_0) \in X_0 \\ u(\cdot) \in U_0 = \{u(\cdot) \in KC : |u_i(t)| \leq v_i, i = 1, \dots, r\}, \\ v(\cdot) \in V_0 = \{v(\cdot) \in KC : |v_j(t)| \leq \mu_j, j = 1, \dots, s\} \end{cases} \quad (2.1)$$

donde $x(t)$ es el vector n -dimensional de desviaciones de una trayectoria dada, $v(\cdot)$ es el vector s -dimensional de perturbaciones permanentes, $u(\cdot)$ es el vector r -dimensional de controles estabilizadores, $x(t_0) \in X_0$ es el conjunto de desviaciones iniciales, v_i y μ_j son parámetros dados que describen los recursos de control y la perturbación permanente. U_0 y V_0 son conjuntos dados en los cuales el constructor trabaja prácticamente.

Suponemos que el sistema (2.1) es estabilizable y completamente controlable ($x(t) \rightarrow 0$ cuando $t \rightarrow \infty$) para cualquier $v \equiv const \in V_0$, si el correspondiente algoritmo de estabilización $u(\cdot) \in U_0$ es usado. Suponemos también que la

calidad de la estabilización está descrita por el criterio de desempeño cuadrático dado en (2.2)

$$L(u, v) = \int_{t_0}^{t_1} (x^T G x + u^T N u) dt + x^T(t_1) S x(t_1) \quad (2.2)$$

donde t_1 es el instante de tiempo en el cual la variedad dada en el espacio de las desviaciones se alcanza por primera vez (es posible considerar el caso de un t_1 fijo, $0 < t_1 \leq \infty$). La expresión (2.2) también es conocida como funcional objetivo y es una medida cuantitativa del comportamiento del sistema en el tiempo bajo la ley de control.

El primer sumando del funcional es una integral que depende de los valores que va tomando $x(t)$ y $u(t)$ a lo largo del horizonte temporal, por lo tanto, valora el comportamiento del sistema a través del tiempo. El segundo sumando valora el estado en que queda el sistema al final del intervalo de tiempo que constituye el horizonte temporal del problema.

Para el establecimiento de las pruebas asumimos que conocemos el sistema dinámico (2.1), el cual describe el comportamiento de las desviaciones y el criterio de desempeño (2.2) y los conjuntos U_0 y W_0 . Las matrices G , N y S son constantes y satisfacen que $G \geq 0$, $N > 0$ y $S \geq 0$.

Denotemos a $w = (x(t), v(t)) \in W_0$ donde $W_0 = X_0 \times V_0$, formulamos los siguientes problemas de síntesis y análisis de sistemas de control [4]:

a) El problema de estabilización minimax.

$$\inf_{u(\cdot) \in U_0} \sup_{w(\cdot) \in W_0} L(u, w) \quad (2.3)$$

b) El problema de las pruebas maximin.

$$\sup_{w(\cdot) \in W_0} \inf_{u(\cdot) \in U_0} L(u, w) \quad (2.4)$$

El problema (2.4) es considerado como el soporte del procedimiento de pruebas (ver el esquema funcional de la figura 2). En este trabajo, en el planteamiento general de las pruebas maximin asumimos que el funcional L alcanza su máximo y su mínimo. La cadena de desigualdades (2.5) es muy importante [5].

$$L^0 = L(u^0, w^0) = \max_{w \in W_0} \min_{u \in U_0} L(u, w) \leq \min_{u \in U_0} \max_{w \in W_0} L(u, w) \leq L(\tilde{u}, w^0) = \tilde{L} \quad (2.5)$$

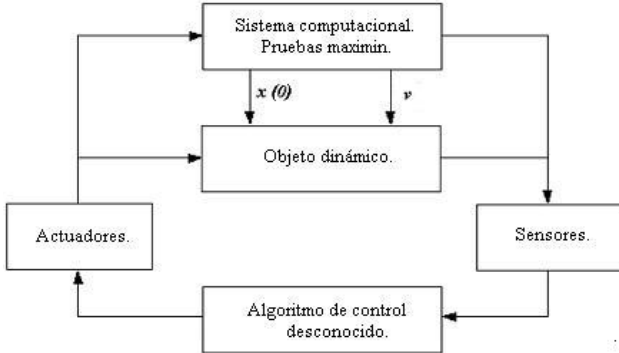


Figura 2: Esquema Funcional de las Pruebas Maximin.

El procedimiento de las pruebas maximin para la calidad de la estabilización robusta está constituido por tres etapas [6]:

En la **primera etapa**, la estrategia óptima $u^0(x, t)$, la estrategia de las pruebas $w^0(x, u, t)$ (si existe) y el valor del juego dinámico (2.4) son determinados y de la primera desigualdad en (2.5), convenimos en tomar como parámetro de evaluación o nota excelente a la solución del problema extremal (2.4). Dicho problema es considerado como el soporte del procedimiento de pruebas, ya que L^0 es la estimación cualitativa inferior de estabilización de las trayectorias programadas (ver esquema funcional de la figura 2).

La **segunda etapa** se lleva a cabo dentro del marco de la simulación dinámica, cuando generada una contraestrategia $w^0(x, u, t)$, en respuesta a la acción de salida $\tilde{u}(t)$, obtenemos la estimación de estabilización real \tilde{L} . El algoritmo de estabilización es considerado para nosotros como una caja negra, dado que la forma en que se construyó es desconocida para nosotros, únicamente tenemos conocimiento de la señal de salida de dicho algoritmo de estabilización (sin embargo, en algunas ocasiones se considera que la estructura del algoritmo a evaluar es lineal,

$\tilde{u} = K(t)x$, donde los parámetros $K(t)$, son desconocidos para nosotros).

En la **tercera etapa**, el resultado de la prueba \tilde{L} , se compara con la estimación menor L^0 :

$$k_0 = \frac{L^0}{\tilde{L}(w^0(x, u, t), \tilde{u}(x, t))} \cdot 100 \quad (2.6)$$

la relación (2.6) es una estimación para la calidad de la estabilización.

III. PRUEBAS MAXIMIN PARA UN ROBOT MÓVIL

Ahora se aplican las pruebas maximin a un ejemplo práctico, se desarrolla la estrategia de estas pruebas para evaluar un algoritmo de estabilización aplicado un robot móvil. Consideramos la clase de robots móviles autónomos que consisten de tres ruedas, dos activas y una pasiva, con restricciones no-holonómicas, que aparecen como consecuencia de la hipótesis de no deslizamiento [2]. Las velocidades del centro de las ruedas son denotadas como v_r y v_l , ver figura 3.

La posición del robot (centro del arreglo de sensores) con respecto al sistema de referencia $P_{\xi\eta}$, está dada por las siguientes relaciones:

$$\begin{cases} \dot{\alpha} = \omega \\ \dot{\xi} = v \cos \alpha - h \sin \alpha \\ \dot{\eta} = v \sin \alpha + h \sin \alpha \end{cases} \quad (3.1)$$

el robot es un sistema no-holonómico y las velocidades de las ruedas activas deben tener la dirección del eje P_x (el sistema de referencia P_{xy} , tiene situado el origen en la intersección del eje de simetría del robot y el centro del eje de las ruedas activas).

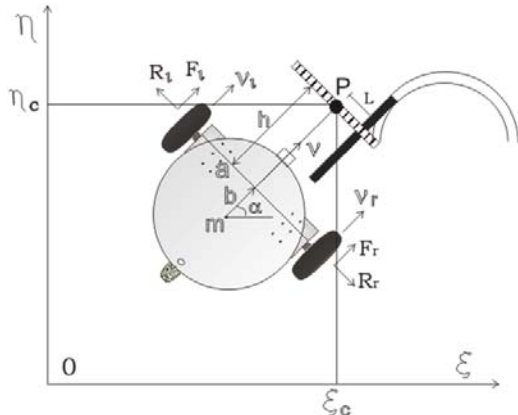


Figura 3: Sistema de coordenadas inercial.

Para obtener las ecuaciones dinámicas, se sustituyen estas restricciones por fuerzas reactivas \vec{R}_r , \vec{R}_l . Se tienen también, dos fuerzas activas \vec{F}_r , \vec{F}_l . Mediante los teoremas principales de la mecánica se describen las ecuaciones en el sistema $P^{\xi\eta}$.

Obtenemos las siguientes relaciones dinámicas

$$\begin{aligned} m(\dot{v} + b\omega^2) &= F_r + F_l \\ \tilde{J}\dot{\omega} + mb\omega v &= (F_r - F_l)a \end{aligned} \quad (3.2)$$

sustituyendo las fuerzas activas por torques de los motores y los voltajes que son aplicados a los mismos

$$M = F\rho \quad (3.3)$$

donde M es el torque del motor, y el modelo más simple del motor es:

$$M = \chi u - \sigma\dot{\phi}$$

el miembro derecho $\sigma\dot{\phi}$ es la suma de la fricción viscosa y la fuerza contraelectromotriz. Entonces para la rueda derecha

$$\begin{aligned} F_r &= \frac{\chi u_r - \sigma\dot{\phi}}{\rho} \\ \dot{\phi}_r &= \frac{v_r}{\rho} = \frac{v + \omega a}{\rho} \end{aligned} \quad (3.4)$$

Sustituyendo en la ecuación para cada rueda, y realizando operaciones en (3.2) se obtienen las ecuaciones dinámicas:

$$\begin{aligned} m\dot{v} + mb\omega^2 + \frac{2\sigma}{\rho^2}v &= \frac{\chi}{\rho}(u_r + u_l) \\ \tilde{J}\dot{\omega} + mb\omega v + \frac{2\sigma a}{\rho^2}\omega &= \frac{\chi}{\rho}(u_r - u_l) \end{aligned} \quad (3.5)$$

Finalmente se obtienen las ecuaciones del movimiento que permiten simular la dinámica de este robot:

$$\begin{cases} \dot{\xi} = v \cos \alpha - b\omega \sin \alpha \\ \dot{\eta} = v \sin \alpha - b\omega \cos \alpha \\ \dot{\alpha} = \omega \\ m\dot{v} = -mb\omega^2 - \frac{2\sigma}{\rho^2}v + \frac{\chi}{\rho}(u_r + u_l) \\ \tilde{J}\dot{\omega} = mb\omega v - \frac{2a^2\omega\sigma}{\rho^2} + \frac{\chi a}{\rho}(u_r - u_l) \end{cases} \quad (3.6)$$

Cualquier trayectoria real se puede conseguir realizando combinaciones de trayectorias rectas y semicírculos como se muestra en la figura 4.

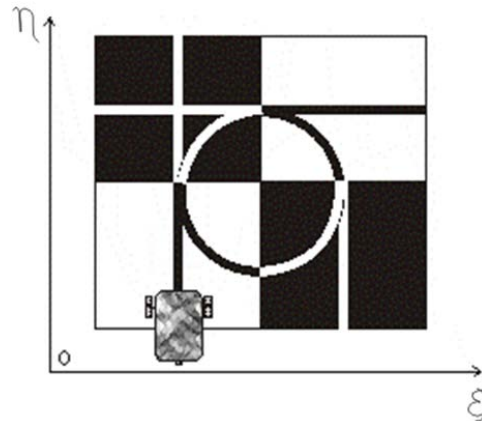


Figura 4: Conjunto de posibles trayectorias programadas.

Se observa que el robot puede realizar movimientos por líneas rectas horizontales o verticales en cualquier sentido, así como semicírculos también en ambos sentidos.

La tabla 1 muestra el conjunto de trayectorias deseadas, cuando el movimiento se realiza a lo largo de un segmento de línea paralela al eje ξ o al eje η .

Trayectoria.	ξ^d	η^d	α^d	v^d	ω^d
1. Línea horizontal sentido negativo.	$v_0 t + \xi_0$	0	0	v_0	0
2. Línea horizontal sentido positivo.	$v_0 t + \xi_0$	0	π	v_0	0
3. Línea vertical sentido negativo.	0	$v_0 t + \eta_0$	$\frac{\pi}{2}$	v_0	0
4. Línea vertical sentido positivo.	0	$v_0 t + \eta_0$	$-\frac{\pi}{2}$	v_0	0

Tabla 1: Trayectorias programadas de las líneas horizontal y vertical.

En la figura 4 se muestra un tablero de entrenamiento donde tenemos cuatro cuadrantes, cada uno con una trayectoria semicircular la cual se puede realizar en dirección a las manecillas del reloj o contraria a ella, entonces se tienen ocho posibles configuraciones como se muestra en la tabla 2.

Trayectoria.	ξ^d	η^d	$\varphi \in$	α^d	v^d	ω^d
5. Cuadrante I dirección reloj.	$R \cos \varphi$	$R \sin \varphi$	$\left[\frac{\pi}{2}, 0\right]$	$-\omega_0 t$	$\omega_0 R$	ω_0
6. Cuadrante I contra reloj.	$R \cos \varphi$	$R \sin \varphi$	$\left[0, \frac{\pi}{2}\right]$	$\omega_0 t + \frac{\pi}{2}$	$\omega_0 R$	ω_0
7. Cuadrante II dirección reloj.	$R \cos \varphi$	$R \sin \varphi$	$\left[\pi, \frac{\pi}{2}\right]$	$-\omega_0 t + \frac{\pi}{2}$	$\omega_0 R$	ω_0
8. Cuadrante II contra reloj.	$R \cos \varphi$	$R \sin \varphi$	$\left[\frac{\pi}{2}, \pi\right]$	$\omega_0 t + \pi$	$\omega_0 R$	ω_0
9. Cuadrante III dirección reloj.	$R \cos \varphi$	$R \sin \varphi$	$\left[-\frac{\pi}{2}, -\pi\right]$	$\pi - \omega_0 t$	$\omega_0 R$	ω_0
10. Cuadrante III contra reloj.	$R \cos \varphi$	$R \sin \varphi$	$\left[-\pi, -\frac{\pi}{2}\right]$	$\omega_0 t - \frac{\pi}{2}$	$\omega_0 R$	ω_0
11. Cuadrante IV dirección reloj.	$R \cos \varphi$	$R \sin \varphi$	$\left[0, -\frac{\pi}{2}\right]$	$-\omega_0 t - \frac{\pi}{2}$	$\omega_0 R$	ω_0
12. Cuadrante IV contra reloj.	$R \cos \varphi$	$R \sin \varphi$	$\left[-\frac{\pi}{2}, 0\right]$	$\omega_0 t$	$\omega_0 R$	ω_0

Tabla 2: Trayectorias programadas de los segmentos de círculo.

Si $u^d(t)$ es una entrada nominal al sistema descrito por las ecuaciones (3.6) y $y^d(t)$ es una

trayectoria nominal de dicho sistema, entonces podemos describir el sistema de ecuaciones lineales en desviaciones. Para la trayectoria 4 (Tabla 1) el sistema lineal es:

$$\begin{cases} \dot{\xi} = v_0 \alpha + h \omega \\ \dot{\eta} = -v \\ \dot{\alpha} = \omega \\ \dot{v} = -\frac{2\sigma}{m\rho^2} v + \frac{\chi}{m\rho} (u_l + u_r) \\ \dot{\omega} = \frac{mbv_0}{\tilde{J}} \omega - \frac{2a^2\sigma}{\tilde{J}\rho^2} \omega + \frac{\chi a}{\tilde{J}\rho} (u_r - u_l) \end{cases} \quad (3.7)$$

Si la trayectoria deseada es un semicírculo (trayectoria 6, tabla 2), se obtiene el siguiente sistema lineal:

$$\begin{cases} \dot{\xi} = -(\omega_0 R \sin \alpha^d + h \omega_0 \cos \alpha^d) \alpha + v \cos \alpha^d - h \omega \sin \alpha^d \\ \dot{\eta} = (\omega_0 R \cos \alpha^d + h \omega_0 \sin \alpha^d) \alpha + v \sin \alpha^d + h \omega \cos \alpha^d \\ \dot{\alpha} = \omega \\ \dot{v} = -\frac{2\sigma}{\rho^2} v - 2b\omega_0 \omega + \frac{\chi}{m\rho} (u_r + u_l) \\ \dot{\omega} = \frac{mb\omega_0}{\tilde{J}} v - \frac{mb\omega_0 R}{\tilde{J}} \omega - \frac{2a^2\sigma}{\tilde{J}\rho^2} \omega + \frac{\chi a}{\tilde{J}\rho} (u_r - u_l) \end{cases} \quad (3.8)$$

Para las demás trayectorias semicirculares los sistemas linealizados son análogos y la diferencia entre ellos dependerá del valor de α^d .

IV. ESTIMACIÓN INFERIOR DE LAS PRUEBAS MAXIMIN DE UN ROBOT MOVIL

En este epígrafe resolvemos el problema extremal (2.4) para obtener la estimación inferior (paso 1) y establecer las pruebas maximin. Para solucionar el problema interno de (2.4), hacemos uso de la programación dinámica de Bellman.

El criterio de desempeño (2.2) es minimizado, dada la dinámica del proceso controlable (2.1), ahora definimos la función de costo mínima

$$\omega(x, t) = \min_{u(t)} \left\{ \int_t^{t_f} (x^T G(t)x + u^T N(t)u) dt + x^T(t_1) Sx(t_1) \right\} \quad (4.1)$$

Asumiendo que las derivadas parciales de ω existen y son acotadas, se escribe la ecuación de Hamilton-Jacobi-Bellman, donde $\omega(x, t)$ debe satisfacer la ecuación en derivadas parciales.

$$0 = \omega_t(x(t), t) + \min_{u(t)} \left\{ x^T G(t)x + u^T N(t)u + \nabla_x \omega(x(t), t) [A(t)x + B(t)u] \right\} \quad (4.2)$$

donde

$$\omega(x(t_f), t_f) = x^T(t_1) Sx(t_1) \quad (4.3)$$

Asumimos que la solución de la ecuación (4.3) es de la forma $\omega(x, t) = x^T P(t)x$, de esta manera se deduce el siguiente control:

$$u^T = -N^{-1} B^T P(t)x \quad (4.4)$$

Sustituyendo el control (4.4) en la ecuación (4.2) se obtiene:

$$0 = x^T \left\{ \frac{\partial P(t)}{\partial t} + G + P(t)A + A^T P(t) - P(t)BN^{-1}B^T P(t) \right\} x \quad (4.5)$$

donde:

$$\frac{\partial P(t)}{\partial t} = G + P(t)BN^{-1}B^T P(t) - P(t)A - A^T P(t) \quad (4.6)$$

es la llamada ecuación diferencial matricial de Riccati [3].

La ecuación diferencial matricial de Riccati (4.6), con condiciones finales $P(t_f) = 0$ se soluciona y se obtiene la matriz $P(t)$ lo que permite obtener completamente el control (4.4), así obtenemos la solución del problema interno de (2.4).

Ahora para solucionar el problema externo de (2.4), consideramos cuatro movimientos programados; en este caso las trayectorias 2 y 4 de la tabla 1, así como las trayectorias 6 y 7 de la tabla 2 como se muestra en la figura 5.

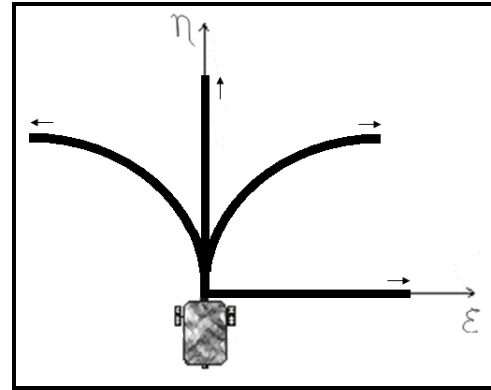


Figura 5: Trayectorias programadas.

Como consecuencia el conjunto W_0 se considera un conjunto discreto, por lo que resolvemos el problema interno 4 veces y posteriormente tomamos el máximo valor determinando con ello la estimación inferior L^0 .

En el desarrollo de esta parte consideramos los valores numéricos de los parámetros que intervienen en el robot móvil [7] (ver tabla 3).

Variable.	Valor.	Descripción.
v_0	1.5	Velocidad deseada [m/s]
a	0.4	Distancia entre las ruedas [m]
b	0.4	Distancia del centro de masa al eje de las ruedas [m]
h	0.1	Distancia del eje de las ruedas al arreglo de sensores infrarrojos [m]
m	4.5	Masa del robot [Kg]
ρ	0.08	Radio de las ruedas [m]
R	0.35	Radio de inercia del carro [m]
χ	0.01	Fricción viscosa
σ	0.009	Fuerza contraelectromotriz del motor.

Tabla 3: Parámetros del robot móvil.

Sustituimos los parámetros y encontramos el sistema de ecuaciones lineales en desviaciones, para las trayectorias mostradas en la figura 5.

La línea horizontal (sentido positivo):

$$\begin{cases} \dot{\xi} = -v \\ \dot{\eta} = -1.5\alpha - 0.1\omega \\ \dot{\alpha} = \omega \\ \dot{v} = -0.625v + 0.0277(u_l + u_r) \\ \dot{\omega} = 7.8435\omega - 0.1743(u_r - u_l) \end{cases} \quad (4.7)$$

La línea vertical (sentido positivo):

$$\begin{cases} \dot{\xi} = 1.5\alpha + 0.1\omega \\ \dot{\eta} = -v \\ \dot{\alpha} = \omega \\ \dot{v} = -0.625v + 0.0277(u_l + u_r) \\ \dot{\omega} = 7.8435\omega - 0.1743(u_r - u_l) \end{cases} \quad (4.8)$$

La línea semicircular del cuadrante I (sentido contra reloj):

$$\begin{cases} \dot{\xi} = -(0.175 \sin \alpha^d + 0.05 \cos \alpha^d)\alpha + v \cos \alpha^d - 0.1\omega \sin \alpha^d \\ \dot{\eta} = (0.175 \cos \alpha^d - 0.05 \sin \alpha^d)\alpha + v \sin \alpha^d + 0.1\omega \cos \alpha^d \\ \dot{\alpha} = \omega \\ \dot{v} = -0.625v - 0.4\omega + 0.0277(u_l + u_r) \\ \dot{\omega} = 3.1374v - 0.4706\omega - 0.1743(u_r - u_l) \end{cases} \quad (4.9)$$

Por último la línea semicircular del cuadrante II (sentido del reloj) que sería el mismo caso que (4.9) variando únicamente α^d .

Mediante la programación dinámica resolvemos los problemas internos para cada uno de los sistemas anteriores.

1. El valor del problema interno para el sistema (4.7) es : 23.5298
2. El valor del problema interno para el sistema (4.8) es : 32.0577
3. El valor del problema interno para el sistema (4.9) es : 8.0087
4. El valor del problema interno para el sistema con la línea semicircular del cuadrante II (sentido del reloj) es: 7.8776

Por lo tanto la estimación inferior para las pruebas maximin es:

$$L^0 = 32.0577 \quad (4.10)$$

V. EVALUACIÓN DE UN SISTEMA DE CONTROL DESCONOCIDO.

Ahora analizamos la calidad de estabilización de un algoritmo de control (pasos 2 y 3 del sistema de pruebas maximin) que tiene la forma lineal $\dot{\tilde{u}} = K(t)x$, donde:

$$\tilde{u} = \begin{bmatrix} u_l \\ u_r \end{bmatrix} \quad x = \begin{bmatrix} \xi \\ \eta \\ \alpha \\ v \\ \omega \end{bmatrix} \quad K = \begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} & k_{15} \\ k_{21} & k_{22} & k_{23} & k_{24} & k_{25} \end{bmatrix}$$

de esta manera el control será:

$$\begin{cases} u_l = (k_{11})\xi + (k_{12})\eta + (k_{13})\alpha + (k_{14})v + (k_{15})\omega \\ u_r = (k_{21})\xi + (k_{22})\eta + (k_{23})\alpha + (k_{24})v + (k_{25})\omega \end{cases} \quad (5.1)$$

Los parámetros de $K(t)$ están dados por:

$$\begin{aligned} k_{11} &= -(0.119411 \times 10^{-6})t^8 - (0.108389 \times 10^{-4})t^7 \\ &\quad + (0.898817 \times 10^{-3})t^6 - 0.0227861t^5 \\ &\quad + 0.271979t^4 - 1.63701t^3 + 4.69189t^2 \\ &\quad - 5.19183t + 123.582 \\ k_{14} &= -(0.354728 \times 10^{-5})t^8 + (0.25612 \times 10^{-3})t^7 \\ &\quad - (0.751453 \times 10^{-2})t^6 + 0.114973t^5 \\ &\quad - 0.978158t^4 + 4.55636t^3 - 10.6111t^2 \\ &\quad + 9.8337t + 95.42 \\ k_{22} &= (0.380492 \times 10^{-6})t^8 - (0.437618 \times 10^{-4})t^7 \\ &\quad + (0.176443 \times 10^{-2})t^6 - 0.0343873t^5 \\ &\quad + 0.35581t^4 - 1.95405t^3 + 5.24845t^2 \\ &\quad - 5.52425t + 101.122 \\ k_{23} &= -(0.559137 \times 10^{-5})t^8 + (0.404527 \times 10^{-3})t^7 \\ &\quad - 0.011901t^6 + 0.182708t^5 \\ &\quad - 1.5608t^4 + 7.30545t^3 - 17.109t^2 \\ &\quad + 15.9622t + 157.549 \\ k_{25} &= -(0.980086 \times 10^{-5})t^8 + (0.730996 \times 10^{-3})t^7 \\ &\quad - 0.0221376t^6 + 0.349339t^5 \\ &\quad - 3.06312t^4 + 14.6952t^3 - 35.228t^2 \\ &\quad + 33.6019t + 101.944 \\ &\quad k_{12} = k_{13} = k_{15} = k_{21} = k_{24} = 0 \end{aligned}$$

Utilizando el control (5.1) determinamos la nota real \tilde{L} (paso 2), para ello sustituimos el controlador en el sistema (3.6) y utilizamos el índice de desempeño dado en (2.2), con lo cual obtenemos la estimación real:

$$\tilde{L} = 10003 \quad (5.2)$$

La última etapa (paso 3) consiste en comparar la estimación real obtenida en (5.2) con la estimación inferior (4.10), para esto utilizamos la relación que nos determina la estimación para la calidad de la estabilización dada en (2.6)

$$k^0 = \frac{32.0577}{10003} * 100 = 0.3205 \quad (5.3)$$

De acuerdo a lo anterior, podemos concluir que el algoritmo de control evaluado se debe mejorar en su desempeño.

VI. CONCLUSIONES

Este trabajo se basa principalmente en el establecimiento y análisis de las pruebas maximin para la calidad de estabilización robusta y como su nombre lo dice, es una estrategia de pruebas que nos determinarán la calidad de un algoritmo de control.

El uso de la programación dinámica es de vital importancia, ya que con este método se pudo obtener la estimación inferior y de este manera tener un parámetro con quien comparar cualquier sistema de control que sea aplicado a un robot móvil como lo ejemplifica este trabajo.

VII. REFERENCIAS

- [1] Ferreyra Guillermo y Pascal Jesús, *Control Óptimo Determinista vía Programación Dinámica*, [en línea]; Divulgaciones Matemáticas Vol. 7 No. 2, 1999, pp. 167-185, [consulta: 19 de febrero de 2009] Disponible: <http://www.emis.de/journals/DM/v72/art6.pdf>
- [2] Gutiérrez Arias J. E. Moisés, *Estrategias de las Pruebas Maximin para la*

Calidad de la Estabilización, Tesis de doctorado, México 2003, pp. 1-34

- [3] Tamer Basar y Pierre Bernhard, *H_∞-Optimal Control and Related Minimax Design Problems*, Segunda edición, Editorial Series, U.S.A. 1995, pp. 14-48
- [4] V. A. Sadovnichii y otros, “*Maximin Testing of Satellite Stabilization*”, *Mathematical Modeling of Complex Information Processing Systems*, pp. 61-70
- [5] V. V. Alexandrov y Gutiérrez Arias J. E. Moisés, “*About the Maximin Testing Strategy*”, *Sociedad Matemática Mexicana, Aportaciones Matemáticas Memoria 35*, México 2004, pp. 3-11
- [6] V. V. Alexandrov y Gutiérrez Arias J. E. Moisés, “*Pruebas Maximin y Juegos Geométricos*”, *Sociedad Matemática Mexicana, Aportaciones Matemáticas Memoria 35*, México 2005, pp. 3-12
- [7] V. V. Alexandrov y otros, “*Optimization and Computer-Aided Testing of Stabilization Precision*”, *Mathematical Modeling of Complex Information Processing Systems*, pp. 49-60

VIII. AUTORES

Dr. José Eligio Moisés Gutiérrez Arias obtuvo su grado de Doctor en Ciencias Matemáticas con especialidad en Teoría de Control Optimal por la Benemérita Universidad Autónoma de Puebla (BUAP) en el año 2003.

Lic. Javier Cabrera Martínez obtuvo su título de Licenciado en Electrónica con especialidad en Robótica y Control por la Benemérita Universidad Autónoma de Puebla (BUAP).

Dr. José Fernando Reyes Cortés obtuvo su grado de Doctor en Ciencias de la Electrónica y Comunicaciones por el Centro de Investigación Científica y Estudios Superiores de Ensenada (CICESE).

TRANSITORIOS NO-LINEALES APLICADOS A UN SERVOMECANISMO DIFUSO

Rangel Miranda Domingo, Rivera López Ana Leonor y Castaño Meneses Víctor M.

Centro de Física Aplicada y Tecnología Avanzada, UNAM Campus Juriquilla Querétaro Qro., C.P. 76230
Tel. 442 2381136, correo electrónico: ranged@fata.unam.mx

RESUMEN

En este trabajo se describe un servomecanismo de posicionamiento angular controlado con lógica difusa, sujeto a la acción de cargas no-lineales generadas con un motor DC, el cual es alimentado con señal tipo PWM (pulse width modulation) cuya variación transitoria en el tiempo es de forma exponencial o logarítmica. Se hicieron pruebas a diferentes pasos angulares para verificar el rendimiento de este esquema de control actuando en tiempo real y se observó una alta precisión en el posicionamiento angular, mínimo sobrepaso y robustez al momento de perturbarlo con los transitorios mecánicos no-lineales mencionados.

Palabras clave: Posicionamiento angular, control difuso, servomecanismos, cargas no-lineales.

I. INTRODUCCIÓN

Los sistemas mecatrónicos de posicionamiento angular, formados generalmente por un motor generador de torque, un actuador y un sistema de acoplamiento hacia cargas, cuya variación en muchos casos son transitoriamente no-lineales, requieren de sistemas de control adaptables a condiciones dinámicas, que en la mayoría de las veces no pueden ser modeladas ni manejadas por métodos convencionales. Es común utilizar estrategias de control tales como PID (Proporcional-integral –derivativa) para controlar sistemas de posicionamiento en casos cuyo modelado sea representativo y pueda ajustarse a este tipo de esquemas [1]. Para describir tales sistemas mecatrónicos generalmente son utilizados modelos de cuarto orden, en los que se ha hecho uso de la modelación dinámica compleja en servoválvulas [2], control inteligente adaptable

usando autocalibración [3], retroalimentación de velocidad y servomecanismos difusos [4]. El control difuso [5] asegura seguimiento preciso del movimiento prescrito, preservando el balance dinámico [6] y puede manejar control preciso aún en presencia de fricción estática y dinámica.

La lógica difusa es una metodología que habilita al usuario a describir el comportamiento de un sistema usando variables lingüísticas que pueden tomar cualquier valor real entre 1 y 0 [7]. El control lógico difuso consiste de tres componentes básicos que son: fuzzificación, inferencia basada en reglas *si-entonces* y defuzzificación. En esta propuesta es usado un controlador tipo Takagi-Sugeno [8], el cual consiste de dos entradas: el error y la tasa de error. Es decir, el error en la posición y la velocidad angular. Por lo que en este trabajo se presenta un servomecanismo rotatorio robusto para posicionamiento angular preciso, el cual es acoplado suavemente a una carga mecánica transitoria mediante un motor de DC alimentado con una señal tipo PWM variando en forma no-lineal. Este sistema puede ser usado en mecanismos de posicionamiento angular con cargas mecánicas aplicadas, cuyos comportamientos sean descritos mediante una función exponencial o logarítmica.

II. ARREGLO EXPERIMENTAL

El sistema de posicionamiento angular es compuesto por un motor de 12 VDC de imán permanente de dos polos a 2800 rpm, el cual tiene una reducción de 55.22:1 para dar 0.845 rps. La carga no-lineal transitoria es acoplada mediante un resorte torsional de 24 vueltas de acero inoxidable. El motor de carga es de 24 VDC de imán permanente, de 4 polos a 2800 rpm, el cual tiene como estradas dos señales tipo PWM,

especialmente generadas en un microcontrolador de bajo costo, Un algoritmo de control difuso construido en ambiente de *Lab view* V 8.2, es usado en lazo cerrado para el control de la posición angular. Un esquema de sistema retroalimentado es mostrado en la figura 1.

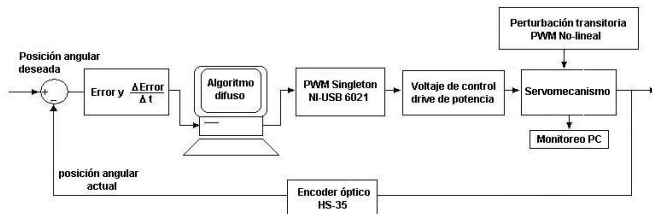


Figura 1. Bloques de sistema.

La posición angular es medida por un encoder HS-35 de cuadratura de 1024 pulsos por revolución, el cual es colocado sobre la flecha que une el motor de torque con la carga. Los pulsos son enviados a una tarjeta de adquisición de señales de *National Instruments* modelo NI-USB 6221, que cuantifica la posición en pulsos y la convierte en grados o en miliradianes. Estos datos son desplegados en la pantalla de la PC y registrados en memoria.

El voltaje de control del motor de torque es una señal de PWM a 15 Khz y el ciclo de trabajo depende del cálculo hecho en el algoritmo de control, el cual lo asigna en forma de *singletons*, es decir están dados en valores particulares de 0%, 20%, 50% y 95% del ciclo de trabajo. Este valor es dado por la regla lingüística aplicada en ese momento y el peso de ésta sobre la salida del controlador.

El posicionamiento angular va desde 0 a 1 revolución y es elegido como *set point* en el sistema. La corriente de armadura es monitoreada y registrada en archivo de Excel. También, el torque dinámico, que provoca la acción del motor de torque y el transitorio debido al motor de carga es medido con un torquímetro digital Lebow 1254-305. Estas variables son registradas a una tasa de muestro de 1ms.

El voltaje de control es aplicado al motor de torque con un *driver* de potencia y permite invertir la polaridad del motor en el proceso de posicionamiento. Los transitorios en la carga son generados en un microcontrolador PIC, se utiliza el

módulo implementado para señal tipo PWM y se escogió un tiempo desde el inicio al final de la variación no-lineal de 765ms a una frecuencia de 60 Hz, para provocar un efecto transitorio no-lineal en el motor de carga. Se escogieron dos tipos de funciones para describir la variación no-lineal en el tiempo, que fueran casos extremos en una carga. En la figura 2 se observa la variación de la señal PWM en forma exponencial, en este caso no hay potencia disponible en un buen tramo de la señal, sólo se acentúa al final.

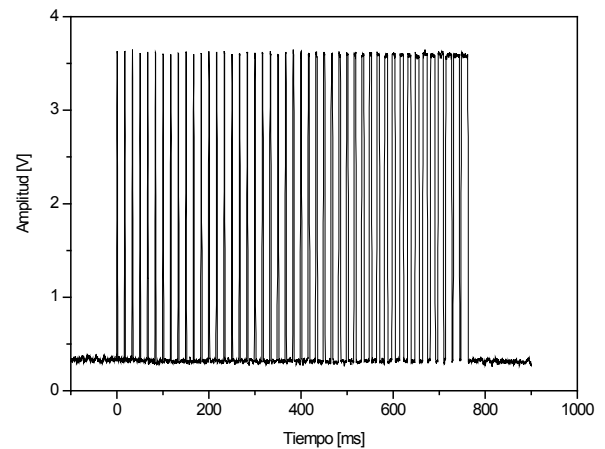


Figura 2. Señal PWM con variación exponencial.

Así mismo se generó una señal PWM con variación logarítmica en un intervalo de 765ms.

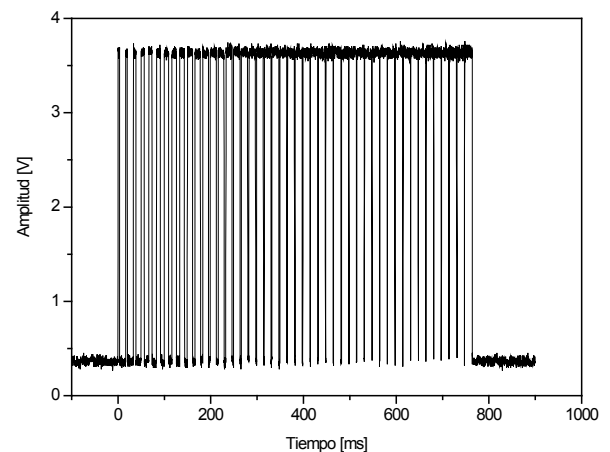


Figura 3. Señal PWM con variación logarítmica.

La figura 3 muestra la variación transitoria de potencia en forma logarítmica, con un aumento

súbito de la energía, manteniendo ésta hasta el final de la aplicación de la carga.

El control difuso fue diseñado en el *fuzzy toolkit* y cargado en el *load controller*. Cada una de las funciones de pertenencia $[\mu]$, así como los conjuntos difusos fue hecho bajo la experiencia de un operador quien maneja el posicionamiento de este sistema bajo la acción de cargas no-lineales. En la figura 4 pueden verse los conjuntos difusos para el error en la posición de 90° , 180° y 360° , dado en miliradianes, así como para la derivada del error, que es la velocidad angular dada en milirads/ms.

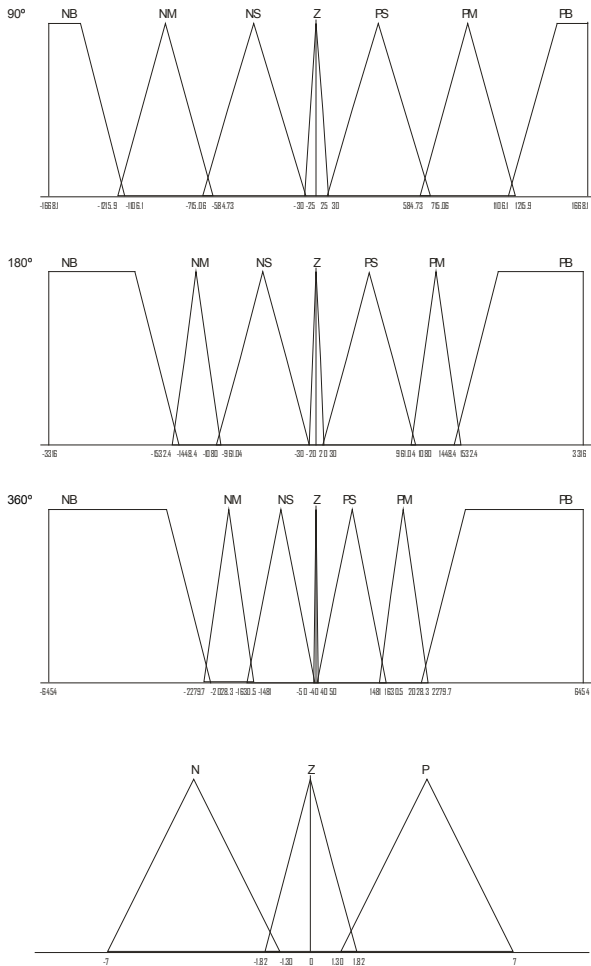


Figura 4. Conjuntos difusos para el error en la posición y la velocidad angular.

Se utilizaron por simplicidad funciones de membresía triangulares, dado que el procesamiento

del algoritmo así como la salida del controlador deben ser en tiempo real.

El universo de discurso fue dividido en siete dominios para el error que son: negativo grande (NG), negativo medio (NM), negativo pequeño (NP), cero (Z), positivo pequeño (PP), positivo medio (PM) y positivo grande (PG). Para la derivada del error, es decir la posición angular, se formaron tres conjuntos que se han llamado positivo (P), cero (Z) y negativo (N). También se formaron las matrices difusas asociadas, en las que se establecieron los porcentajes de PWM en cada caso, de donde surgen las reglas lingüísticas aplicadas.

$\Delta e/e$	NG	NM	NP	Z	PP	PM	PG
P	0.95	0.50	0.50	-0.20	-0.20	-0.50	-0.95
Z	0.95	0.50	0.20	0	-0.20	-0.50	-0.95
N	0.95	0.50	0.20	0.20	-0.20	-0.50	-0.95

Tabla 1. Matriz difusa para 90° , e y Δe .

$\Delta e/e$	NG	NM	NP	Z	PP	PM	PG
P	0.95	0.50	0.20	-0.20	-0.20	-0.50	-0.95
Z	0.95	0.50	0.20	0	-0.20	-0.50	-0.95
N	0.95	0.50	0.20	0.20	-0.20	-0.50	-0.95

Tabla 2. Matriz difusa para 180° , e y Δe .

$\Delta e/e$	NG	NM	NP	Z	PP	PM	PG
P	0.95	0.50	0.50	-0.20	-0.20	-0.50	-0.95
Z	0.95	0.50	0.20	0	-0.20	-0.50	-0.95
N	0.95	0.50	0.20	0.20	-0.20	-0.50	-0.95

Tabla 3 Matriz difusa para 360° , e y Δe .

En la tabla 1, se presentan los datos de los porcentajes de PWM aplicados en el posicionamiento de 90° , tanto para el error en la posición como en las reglas difusas. El cero por ciento significa ninguna acción de control, es decir en la intersección cero con cero, el sistema está en la posición deseada. Lo mismo sucede en las tablas 2 y 3, en las que se muestran datos para el posicionamiento de 180° y para 360° , en los que se aprecian diferentes porcentajes de PWM debido a los requerimientos diferentes en torque para cada posición, el signo positivo significa en sentido horario y negativo en sentido antihorario. En el

caso de 360°, el mecanismo lleva más inercia que en el primer caso, que es para 90°, donde la posición es alcanzada en apenas 750ms.

III. RESULTADOS

Se llevaron a cabo pruebas de posicionamiento a diferentes ángulos para verificar la robustez de este servomecanismo a transitorios no-lineales, con y sin carga. En la figura 5, podemos observar la respuesta a 90° sin carga en la que podemos notar que el sistema alcanza la posición en 750ms aproximadamente, con casi cero sobrepaso (0.37°) y solamente 0.37° de error en estado estacionario, que es la mínima resolución del encoder digital. Se observa al inicio del posicionamiento un retardo debido principalmente a umbral de torque necesario para vencer la inercia provocada por la reducción debida a la transmisión mecánica.

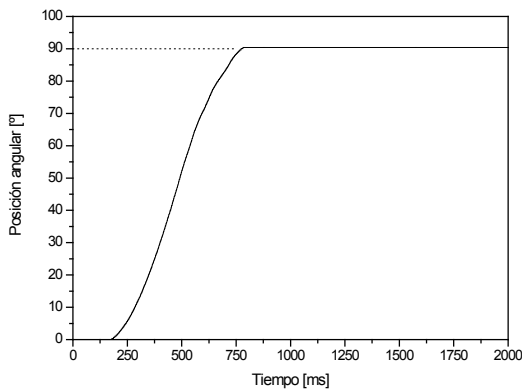


Figura 5. Respuesta a 90° sin carga.

Así mismo se llevó al sistema a una posición de 180° sin aplicar ningún transitorio como se muestra en la figura 6.

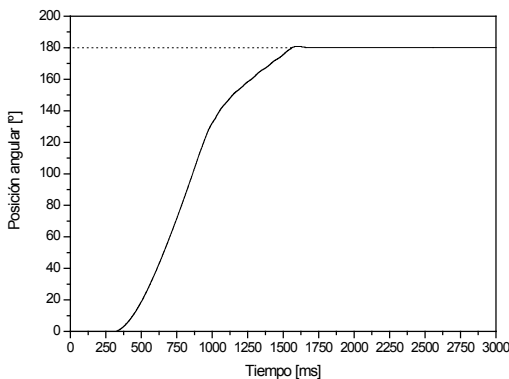


Figura 6. Respuesta a 180° sin carga.

El servomecanismo inicia el posicionamiento con las mismas características que en el caso anterior, sin embargo cerca de los 1000ms sufre una desaceleración para poder alcanzar la posición con poco sobrepaso 0.75°, y error en estado estable de 0.05° en estado estacionario. La posición es alcanzada con el doble del tiempo que la anterior. El sistema también fue posicionado en 360°, es decir una revolución y la posición fue alcanzada con 2382ms, como se muestra en la figura 7, aquí se presentan oscilaciones debida al efecto inercial, pero inmediatamente son reducidas hasta tener un error en estado estable de 0.10° y el sobrepaso máximo fue de 7.13°.

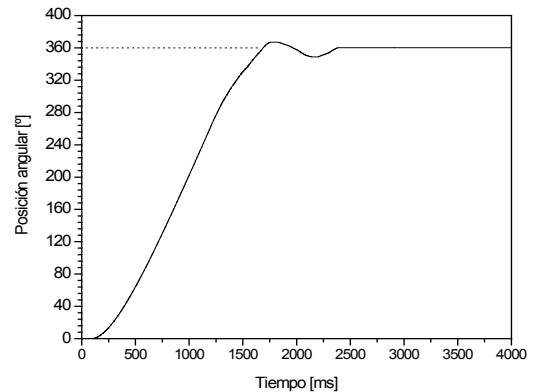


Figura 7. Respuesta a 360° sin carga.

Por otro lado se aplicaron los transitorios de tipo exponencial y logarítmico al comienzo del posicionamiento es decir, en las condiciones más críticas de torque para el servomecanismo. En la figura 8 se aprecian las respuestas del sistema difuso con transitorio exponencial y logarítmico cuando el sistema es programado a 90°.

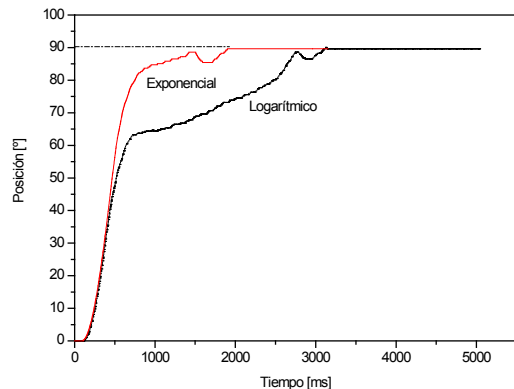


Figura 8. Respuestas transitorias a 90°.

Se observa inmediatamente el cambio en la respuesta debido a la acción del transitorio exponencial que llega con oscilación y provoca que el servomecanismo baje su velocidad, retrasando el posicionamiento. El transitorio logarítmico afectó de manera importante el alcance de la posición y redujo aún más la velocidad, aumentando así el tiempo de establecimiento. Sin embargo, en ambos casos el sistema llega a la posición deseada. La figura 9, muestra la respuesta obtenida para un posicionamiento de 180°, en el cual se consigue llegar mejor, a pesar de la perturbación exponencial.

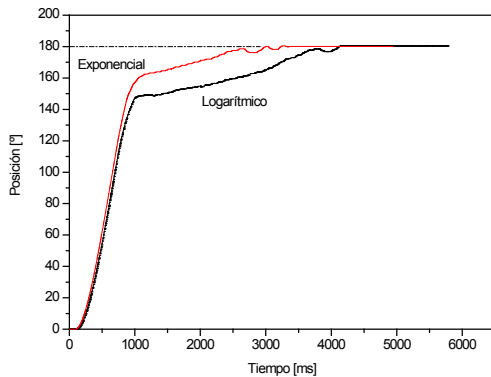


Figura 9. Comparación de los transitorios a 180°.

La acción mecánica de la función logarítmica, retrasa hasta 4 segundos la llegada a la posición, el contenido energético de este transitorio es más fuerte y provoca una notoria desaceleración cuando éste es aplicado. En la figura 10 se muestra la acción de la carga exponencial y logarítmica a un posicionamiento de 360°.

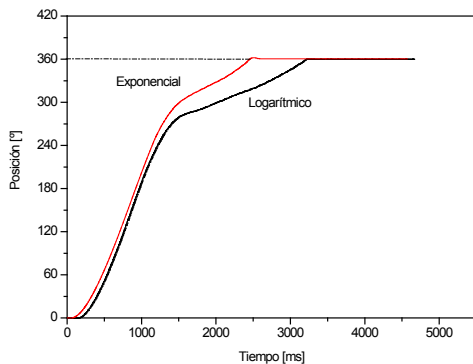


Figura 10. Transitorios aplicados para 360°.

En ella se observa que en ambos casos la trayectoria de subida es similar, sin embargo en

cuanto se aplica la acción de carga transitoria se nota un cambio de velocidad con distintas desaceleraciones. El control difuso trata de recomponer la trayectoria, pero ésta es modificada en relación a la trayectoria obtenida sin carga. Aún así, el servomecanismo fue capaz de alcanzar la posición deseada.

En la tabla 4 se muestra el rendimiento de la respuesta al posicionamiento en ambos casos, sin carga y con carga transitoria. La evolución en el tiempo de estas cargas transitorias permite que sea caracterizado su contenido energético y que sean evaluados sus efectos en servomecanismos de precisión. Se ha escogido el tiempo de establecimiento, el sobrepaso y en error en estado estable para evaluar la estrategia de control difuso.

Carga transitoria/Posición	Parámetro de Control	90°	180°	360°
Sin carga transitoria	Tiempo de establecimiento [ms]	807	1667	2382
	Sobrepaso [°]	0.37	0.75	7.13
	Error estado estable [°]	0.37	0.05	0.10
Exponencial	Tiempo de establecimiento [ms]	1950	3339	2587
	Sobrepaso [°]	0.0	0.40	1.15
	Error estado estable [°]	-0.33	0.05	0.45
Logarítmica	Tiempo de establecimiento [ms]	3147	4133	3302
	Sobrepaso [°]	0.02	0.40	0.10
	Error estado estable [°]	-0.33	0.40	-0.25

Tabla 4. Parámetros bajo transitorios no-lineales.

El tiempo en el que ocurre el posicionamiento a 90°, 180° y 360°, así como para la carga transitoria exponencial y logarítmica, muestran tendencias muy lineales para el caso sin carga y tendencias no lineales cuando es aplicado el transitorio exponencial y logarítmico en cualquiera de los posicionamientos angulares. En todos los casos el error en estado estable fue menor de 1°, el sobrepaso también fue menor de 1°, excepto para 360°. El signo negativo de los valores en la tabla significa que la posición fue alcanzada antes de la posición deseada. Los tiempos de establecimiento de la posición, así como los demás parámetros fueron alcanzados y optimizados por la experiencia de un operador al momento de hacer el posicionamiento. Cabe destacar que el controlador difuso por su característica no lineal, permite un

mayor intervalo de operación, es decir que las reglas y conjuntos difusos para la posición de 360° pueden ser utilizadas en cualquier posicionamiento que quede dentro de universo de discurso, que en este caso corresponde a una revolución.

El transitorio logarítmico, que es el más energético para fines de perturbación al posicionamiento, se aplicó también al servomecanismo en estado estable. En la figura 11 se muestran las respuestas a posicionamientos de 90° , 180° y 360° ; en la se observa el efecto de la carga no-lineal transitoria actuando sobre el sistema una vez posicionado.

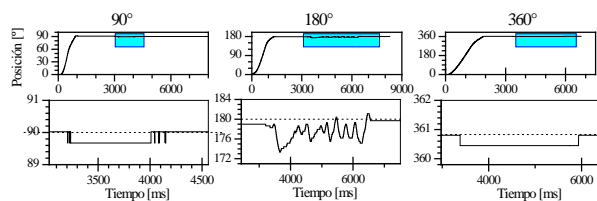


Figura 11. Los transitorios no-lineales como perturbaciones.

La perturbación aplicada no afectó de manera significativa el posicionamiento, ya que en 90° y 360° , lo alejó del punto de control tan solo 0.5° , mientras que para 180° lo alejó 5° del punto de control, pero lo recuperó nuevamente y además logró mejorar la precisión.

Con ello se logró probar también la robustez del servomecanismo ante transitorios no-lineales aplicados como perturbaciones.

IV. CONCLUSIONES

Se ha diseñado y construido un sistema mecatrónico robusto de posicionamiento angular, al cual se le han aplicado transitorios mecánicos producidos mediante un motor de carga que es alimentado con dos tipos de señales moduladas por ancho de pulso de manera exponencial y logarítmica. Para el control de la posición se implementó un controlador difuso basado en reglas lingüísticas, en el cual tanto la posición como la velocidad angular son monitoreadas y procesadas por un algoritmo difuso, quien inteligentemente aplica cada una de ellas, dependiendo del

requerimiento de torque durante el seguimiento de la posición deseada. Se hicieron pruebas a diferentes magnitudes angulares tales como 90° , 180° y 360° para verificar el rendimiento de este esquema de control durante el tiempo de levantamiento así como en estado estable. La posición fue alcanzada con gran precisión, siendo el error menor a 1° en todos los casos, prácticamente cero sobrepaso y tiempos de establecimiento dependiendo del efecto producido por cada uno de los transitorios no-lineales sobre el servomecanismo. El transitorio exponencial no tiene grades efectos sobre la posición si consideramos que la energía disponible en este caso es prácticamente nula en gran parte del intervalo de acción, pero en el caso de la variación transitoria logarítmica el contenido armónico así como la energía disponible para torque hace que sea mas grande el esfuerzo de control para esta condición. El alcance de la posición sin carga fue realizado para comprobar el rendimiento del servomecanismo en condiciones ideales. Se comprobó la robustez de este sistema de posicionamiento angular bajo transitorios no-lineales pudiendo ser aplicada esta estrategia de control en aquellos servomecanismos sujetos a la acción de este tipo de transitorios, o similares a los aquí implementados.

V. REFERENCIAS

- [1] Johnson, M. A. and Moradi, M. H., *PID Control*, Springer Berlín, 2005.
- [2] Li, S. and Song, Y., *Mechatronics*, 17, 8, 442-447, 2007.
- [3] Cecco, M.; Debei, S. and Zaccariotto, M., *Mech. Sys. & Signal Proc.*, 20, 2305-2320, 2006.
- [4] Huang, S.J. and Wang, S.S., *Mechatronics*, 19, 1, 14-28, 2009.
- [5] Tanaka, K. and Wang, H.O., *Fuzzy control systems design and analysis: a linear matrix inequality approach*, Wiley Inter-Science, New York, 2001.

- [6] Dusko, M.K.; Aleksandar, D.R. and Miomir, K.V., *J. Intell. Robot Systems*, **51**, 1, 3-30, 2008.
- [7] Zadeh, L.A., Fuzzy sets, *Information and Control*, 8, 338-353, 1965.
- [8] Takagi & Sugeno, *IEEE Trans. on Syst., Man and Cyber.*, 15, 1, 116-132, 1985.

VI. AUTORES

Domingo Rangel Miranda, obtuvo el grado de Maestro en Ingeniería con especialidad en Sistemas Eléctricos y Electrónicos, por la Facultad de Ingeniería de la Universidad Nacional Autónoma de México, su área de interés son los sistemas mecatrónicos inteligentes.

Ana Leonor Rivera López, es Doctora en Física por la Facultad de Ciencias de la Universidad Nacional Autónoma de México, su área de interés es el análisis de señales de sistemas complejos.

Víctor M. Castaño Meneses, es Doctor en Física por la Facultad de Ciencias de la Universidad Nacional Autónoma de México, realizó postdoctorado en el IBM Thomas J. Watson Research Center, Yorktown Heights, New York, USA. Su área de interés son los sistemas físicos con dinámica compleja.

IDENTIFICACIÓN DE SISTEMAS NO LINEALES USANDO REDES DINÁMICAS RECURRENTE

Omar Cuevas Silva, Juan Anzures Marín, Isidro I. Lázaro Castillo.

Facultad de Ingeniería Eléctrica, División de Estudios de Posgrado
Universidad Michoacana de San Nicolás de Hidalgo
Av. Fco. J. Mújica S/N, Ciudad Universitaria, Morelia, Mich. México
mayrux@hotmail.com, j.anzures@ieee.org, ilazaro@umich.mx

RESUMEN

En este artículo se revisan las propiedades de las redes neuronales para la identificación de sistemas no lineales. La red empleada es de tipo dinámico recurrente de sólo una capa oculta con función tangente sigmoideal. La función de activación en la capa de salida consiste de una función lineal. Estas características facilitan la prueba de estabilidad del modelo neuronal en sistemas donde el origen es un punto de equilibrio. El análisis realizado está enfocado hacia sistemas de segundo orden representados mediante variables de estado debido al fácil trazo en dos y tres dimensiones. La identificación es ilustrada con dos ejemplos de modelos no lineales autónomos.

Palabras clave: Identificación no lineal, redes neuronales, redes dinámicas recurrentes, estabilidad de Lyapunov.

I. IDENTIFICACIÓN DE SISTEMAS DINÁMICOS

A menudo la principal, y a veces la única, información de que dispone un proceso es una serie numérica temporal. Dicha serie corresponde con los valores de una función, de una o más variables, que contiene la dinámica interna del sistema. El problema radica entonces en la identificación de la función generadora de la serie de valores, esta tarea se ha realizado ampliamente con las redes neuronales [1].

La tarea de la identificación de sistemas es esencialmente encontrar mapeos convenientes que puedan aproximar los mapeos implicados en un

sistema dinámico. Un sistema dinámico puede ser descrito por dos tipos de modelos: modelo de entrada salida y modelo estado-espacio [2].

Modelo de Entrada-Salida

Un modelo de entrada-salida describe un sistema dinámico basado en datos de entrada salida. En el dominio discreto, un modelo de entrada-salida puede ser del tipo ARX [3] (AutoRegressive, eXternal input) como se muestra en la Fig.1.

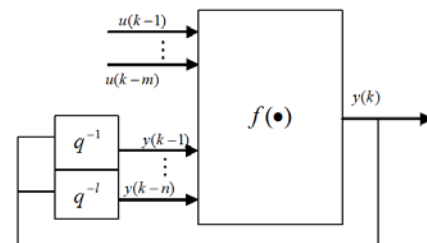


Figura 1. Estructura ARX (AutoRegressive, eXternal input).

Un modelo de entrada-salida supone que la salida nueva del sistema se puede predecir por las entradas y salidas pasadas del sistema. Si un sistema es además supuesto determinístico, invariante en el tiempo, una entrada una salida (SISO), el modelo se convierte en

$$y(k) = f[y(k-1), y(k-2), \dots, y(k-n), u(k-1), u(k-2), \dots, u(k-m)] \quad (1)$$

donde m y n son la cantidad de retrasos en las señales de entrada y salida respectivamente. La función $f(\cdot)$ puede ser una función estática que

mapea las entradas y salidas pasadas a una nueva salida. Si el sistema es lineal, $f(\cdot)$ es una función lineal y la ecuación 1 puede ser rescrita como

$$y(k) = a_1 y(k-1), a_2 y(k-2), \dots, a_n y(k-n), \quad (2)$$

$$b_1 u(k-1), b_2 u(k-2), \dots, b_m u(k-m)$$

Así, desde este enfoque el problema de identificación consiste en una única tarea: encontrar $f(\cdot)$.

Modelo de Estado-Espacio

Un sistema dinámico puede ser descrito también por un modelo de estado-espacio. El modelo de estado-espacio de un sistema dinámico de orden n , no-lineal, invariante en el tiempo, múltiple-entrada-múltiple-salida (MIMO) es como sigue

$$x(k+1) = \Phi[x(k), u(k)] \quad (3)$$

$$y(k) = \psi[x(k)]$$

donde $x(k) = [x_1(k), x_2(k), \dots, x_n(k)]^T$ es el vector de estados de n componentes del sistema, $u(k) = [u_1(k), u_2(k), \dots, u_r(k)]^T$ es el vector de entrada del sistema, y $y(k) = [y_1(k), y_2(k), \dots, y_m(k)]^T$ es el vector de salida del sistema. Φ y ψ son mapeos no-lineales. La función $\Phi(\cdot)$ es un mapeo del estado $x(k)$ y la entrada $u(k)$ dentro del nuevo estado $x(k+1)$. El otro es un mapeo $\psi(\cdot)$ que transforma el estado $x(k)$ dentro de la salida $y(k)$. La Fig 2 muestra el diagrama de bloques del modelo espacio-estado.

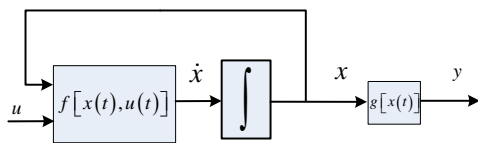


Figura 2. Modelo de estado-espacio

En la identificación de un sistema pueden conocerse previamente el orden del sistema y desconocer de forma precisa sus parámetros o desconocer totalmente el orden del sistema y sus parámetros. La identificación de un sistema mediante redes neuronales no pretende encontrar las funciones Φ y Ψ , sino un mapa de la relación entre un espacio de entrada y un espacio de salida con base a los patrones con que fue entrenada. Si el sistema es lineal, la ecuación 3 se convierte en

$$x(k+1) = Ax(k) + Bu(k) \quad (4)$$

$$y(k) = Cx(k)$$

donde A, B y C son matrices de orden $(n \times n), (n \times r)$ y $(m \times n)$ respectivamente y de coeficientes constantes.

II. ESTABILIDAD DE LYAPUNOV

Un punto de equilibrio se dice estable si todas las soluciones que se inicien en las cercanías del punto de equilibrio permanecen en las cercanías del mismo, de otro modo el punto de equilibrio es inestable. Un punto de equilibrio se dice asintóticamente estable si todas las soluciones que inicien en las cercanías del punto de equilibrio no sólo permanecen en las cercanías del punto de equilibrio, sino que además tienden hacia el equilibrio a medida que el tiempo se aproxima a infinito [4].

Teorema 1 (Método Directo): Sea el origen $x=0$ un punto de equilibrio del sistema no lineal $\dot{x} = f(x)$ donde $f : D \subset \mathbb{R}^n$ es una función continuamente diferenciable y $D \subset \mathbb{R}^n$ es un entorno del origen. Sea

$$A = \left. \frac{\partial f(x)}{\partial x} \right|_{x=0}$$

Entonces, el origen es **ASINTÓTICAMENTE ESTABLE** si todos los valores propios de A tienen parte real negativa. El origen es **INESTABLE** si uno o más valores propios de A tienen parte real positiva.

III. REDES DINÁMICAS RECURRENTE

Las redes multicapa son de naturaleza estática, es decir, su salida no evoluciona con el tiempo (para un patrón de entrada existe una salida asociada), pero pueden adquirir un comportamiento dinámico (para un patrón entrada la salida posee un estado transitorio y converge a un valor en el estado estacionario) realimentando sus entradas con estados anteriores de sus salidas. Como se puede ver en la Fig. 3, la red esta compuesta de una capa estática la cual generalmente posee un número de neuronas superior al número de variables de estado del sistema a identificar, la salida de la capa estática va a un sumador donde se le resta el valor anterior de la variable de estado z_i identificada por el sistema, de esta operación se genera la derivada de cada una de las i variables de estado identificadas por el sistema.

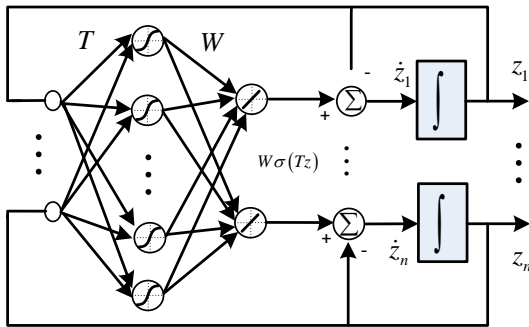


Figura 3. Red Dinámica multicapa.

La red recurrente dinámica multicapa cuyo comportamiento lo describe la ecuación 5 puede identificar el comportamiento de un sistema autónomo ($u = 0$) descrito por la ecuación 6.

$$\frac{\partial z}{\partial t} = f(z) = Az + ws(Tz) \quad (5)$$

$$\frac{\partial x}{\partial t} = f(x) = Ax + f_0(x) \quad (6)$$

donde $x, z \in \mathfrak{R}^n$, $A \in \mathfrak{R}^{n \times n}$, $f(x): \mathfrak{R}^n \rightarrow \mathfrak{R}^n$, $\bar{f}(x): \mathfrak{R}^n \rightarrow \mathfrak{R}^n$, $W \in \mathfrak{R}^{n \times N}$, $T \in \mathfrak{R}^{n \times N}$, $\sigma(z) = [\sigma(z_1), \sigma(z_2), \dots, \sigma(z_n)]$ y función de transferencia $\sigma(\theta) = \text{tansig}(\theta)$, n es el número

de variables de estado del sistema y N el número de neuronas en la capa oculta.

De acuerdo a [5], sin pérdida de generalidad, si el origen se asume como punto de equilibrio, el sistema 6 será identificado con la red 5 alrededor de su región de atracción y se garantiza que el error en la aproximación $e(t)$ es limitado.

La etapa estática que hace parte de la red multicapa dinámica recurrente generalmente es entrenada con algún procedimiento de propagación inversa (Backpropagation [6]), ya que son más eficientes y rápidos. Los patrones de entrenamiento de la capa estática de la Fig. 3 son diferentes combinaciones de valores de las variables de estado y los patrones objetivos están dados por la suma de cada variable de estado con su correspondiente derivada como se muestra en la Fig. 4.

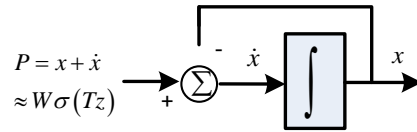


Figura 4. Patrones de entrenamiento de la red

Así, el error generado para el entrenamiento será:

$$e(t) = P - W\sigma(Tz) \quad (7)$$

La red después de entrenada tiene la estructura de la ecuación 8.

$$\frac{\partial}{\partial t} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} -z_1 \\ -z_2 \\ \vdots \\ -z_n \end{bmatrix} + \begin{bmatrix} W_{11} & W_{12} & \dots & W_{1N} \\ W_{21} & W_{22} & \dots & W_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n2} & \dots & W_{nN} \end{bmatrix} \begin{bmatrix} \sigma(T_{11}z_1 + T_{21}z_2 + \dots + T_{n1}z_n) \\ \sigma(T_{12}z_1 + T_{22}z_2 + \dots + T_{n2}z_n) \\ \vdots \\ \sigma(T_{1N}z_1 + T_{2N}z_2 + \dots + T_{nN}z_n) \end{bmatrix} \quad (8)$$

Para garantizar que la red ha identificado la dinámica del sistema, el Jacobiano de la red en el origen (ecuación 9) debe tener valores propios muy cercanos a los del sistema que ha sido aproximado.

$$J_m = -I_n + TW \quad (9)$$

IV. CASOS DE ESTUDIO

Caso 1: Sistema de ecuaciones no lineales

Para el siguiente sistema, una red neuronal identificará su comportamiento autónomo y se verificará la similitud entre ambos vía eigenvalores en el origen como punto de equilibrio.

$$\begin{aligned} \dot{x} &= (x_1 - y)(y^2 + x^2 - 1) \\ \dot{y} &= (x + y)(y^2 + x^2 - 1) \end{aligned} \quad (10)$$

Haciendo $\dot{x} = \dot{y} = 0$

$$\begin{aligned} (x_1 - x_2)(x_2^2 + x_1^2 - 1) &= 0 \\ x^3 + x + y - x - x^2y - x^3 + y &= 0 \end{aligned}$$

y con $y = 0$ y $x = 0$, el PE es (0,0).

Para los eigenvalores es necesario calcular el jacobiano,

$$A = \begin{bmatrix} x^2 + y^2 - 1 + 2(x - y)x & -x^2 - y^2 + 1 + 2(x - y)y \\ x^2 + y^2 - 1 + 2(x + y)x & x^2 + y^2 - 1 + 2(x + y)y \end{bmatrix}$$

Y evaluar en el PE, entonces se obtiene

$$A = \begin{bmatrix} x^2 + y^2 - 1 + 2(x - y)x & -x^2 - y^2 + 1 + 2(x - y)y \\ x^2 + y^2 - 1 + 2(x + y)x & x^2 + y^2 - 1 + 2(x + y)y \end{bmatrix}_{\substack{x_1=0 \\ x_2=0}} = \begin{bmatrix} -1 & \\ & -1 \end{bmatrix}$$

Finalmente el $\det(sI - A)$

$$\det(sI - A) = \det \begin{bmatrix} s+1 & -1 \\ 1 & s+1 \end{bmatrix} = s^2 + 2s + 2 = (s+1-i)(s+1+i)$$

Puesto que los eigenvalores tienen parte real negativa, entonces el punto de equilibrio es un foco estable.

Para la identificación con una red neuronal recurrente se propone una red de 2 nodos de entrada para los estados, 4 nodos en la capa oculta y sus 2 respectivas salidas. La Fig 5(a) muestra la evolución del sistema de ecuaciones (línea continua) y la estimada de la red (línea punteada)

para las condiciones iniciales C.I. $(x,y)=(-0.5,0.5)$. Dicha condición inicial, corresponde a una de las trayectorias del diagrama de fase de la Fig. 5. Con un coeficiente de aprendizaje de 0.001 y después de una iteración la red tiene el comportamiento mostrado en la Fig. 6(b). Para 5 y 14 iteraciones se muestran en la Fig. 6(c) y Fig. 6(d) respectivamente.

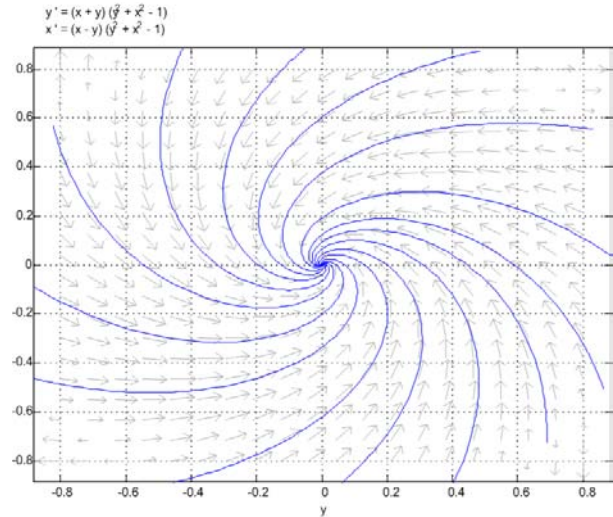


Figura 5. Retrato de fase del sistema no lineal.

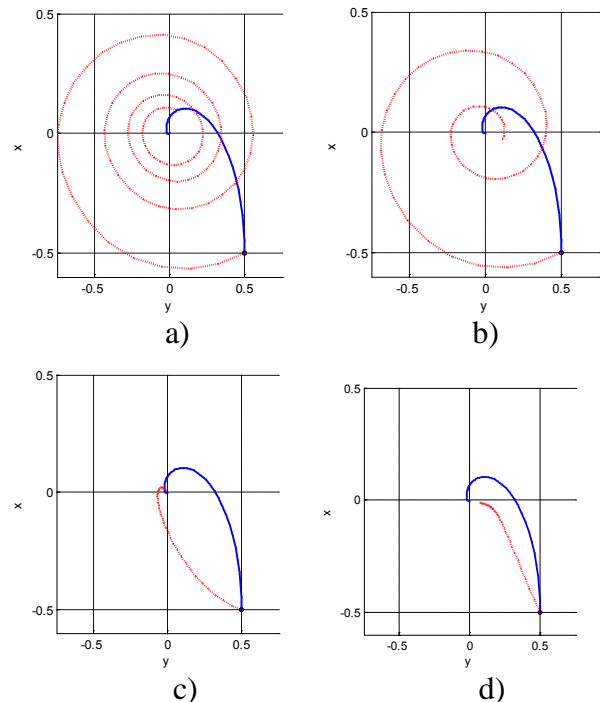


Figura 6. Respuestas para diferentes instantes en el entrenamiento de la red.

Después de 600 iteraciones, el error medio cuadrático desciende por debajo de 1×10^{-5} , como se muestra en la Fig. 7.

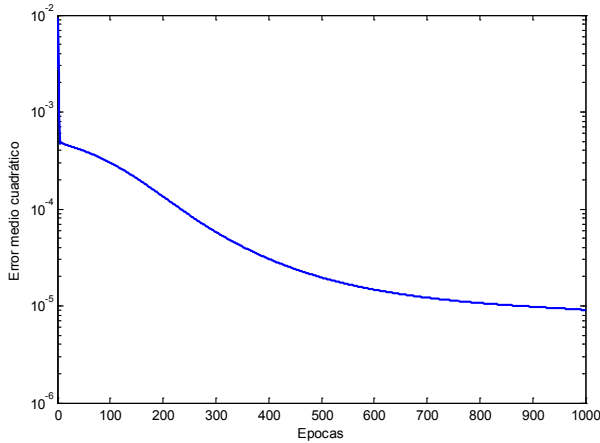


Figura 7. Error medio cuadrático en el entrenamiento.

Finalmente, el conjunto de pesos de la capa oculta y la capa de salida quedan como:

$$T = \begin{bmatrix} 1.193544379257324 & 0.195174618327510 & -0.498339933183938 & -0.637843923175222 \\ 0.403326254967031 & 0.430748853425264 & 0.875252929513446 & 0.692477337561748 \end{bmatrix}$$

$$W = \begin{bmatrix} 0.775984647898979 & -0.250943357452627 \\ -0.562698857678695 & -1.100868262176126 \\ 0.230076245441069 & 0.650376171892572 \\ 0.685686078017679 & 0.206345167179368 \end{bmatrix}$$

Así el jacobiano de dicha configuración es obtenido como:

$$J_{Red} = \begin{bmatrix} -0.735669298740707 & -0.970098005705944 \\ 0.746790071722791 & -0.863276784553033 \end{bmatrix}$$

y sus eigenvalores

$$\lambda_1 = -0.799473041646870 + 0.848757115817266i$$

$$\lambda_2 = -0.799473041646870 - 0.848757115817266i$$

son similares a los obtenidos para el sistema original.

En la Fig. 8 se puede ver la respuesta para cuatro condiciones iniciales y se verifica su dinámica alrededor del origen.

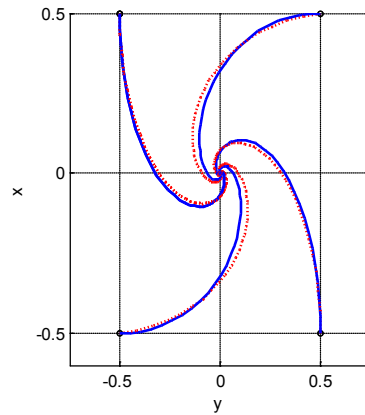


Figura 8. Respuesta para diferentes condiciones iniciales.

Caso 2: Péndulo invertido

El péndulo invertido es un sistema muy comúnmente usado para ejemplificar sus características no lineales. Las ecuaciones no lineales que describen el comportamiento de este sistema son:

$$L_f \frac{\partial i_a}{\partial t} + R_{fa} i_{fa} + V_b = u$$

$$ml^2 \ddot{\theta} + mgl \sin \theta = \tau_m \quad (11)$$

$$\tau_m = N_t K_m i_a$$

$$V_b = N_t K_b \theta_m$$

K_m : Constante de torque = 0.1Nm/A

K_b : Constante de fuerza contraelectromotriz = 0.1v/rad

N_t : Relación de transformación entre engranes = 10

$L_a = 1mH, R_a = 1\Omega, m = 1kg, l = 1m, g = 9.8m/s$

Con $x_1 = \theta, x_2 = \dot{\theta}, x_3 = i_a$ como variables de estado las ecuaciones quedan como:

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = -9.8 \sin(x_1) + x_3 \quad (12)$$

$$\dot{x}_3 = -10x_2 - 10x_3$$

y el jacobiano

$$\begin{bmatrix} 0 & 1 & 0 \\ -9.8 \cos(x_1) & 0 & 1 \\ 0 & -10 & -10 \end{bmatrix}$$

evaluado en el origen

$$\begin{bmatrix} 0 & 1 & 0 \\ -9.8 & 0 & 1 \\ 0 & -10 & -10 \end{bmatrix}$$

sus valores propios son:

$$\lambda_1 = -0.495178632826933 + 3.260679833820912i$$

$$\lambda_2 = -0.495178632826933 - 3.260679833820912i$$

$$\lambda_3 = -9.009642734346132$$

En la Fig. 9(a) se muestra la evolución del sistema de ecuaciones (línea continua) y la estimada de la red, aun sin entrenar, (línea punteada) para las condiciones iniciales C.I. $(x_1, x_2, x_3) = (-0.5, 0.5)$. Para las iteraciones 5, 50 y 100 corresponden a la Fig. 9(b), Fig. 9(c), Fig. 9(d) respectivamente.

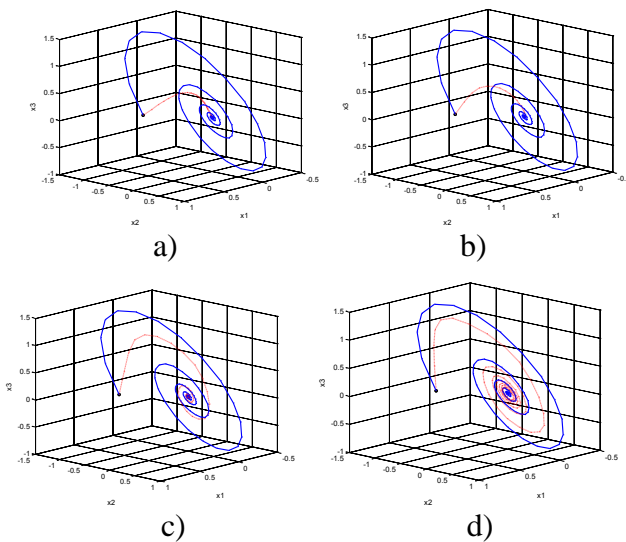


Figura 9. Proceso de identificación del péndulo.

Los valores finales en los pesos de la red son:

$$T = \begin{bmatrix} 1.9336 & 1.1072 & 0.1334 & 0.7694 & -0.4750 \\ -2.0317 & -0.7135 & -0.3941 & -0.4198 & -0.3941 \\ -1.4494 & -1.1213 & -0.9286 & 0.9602 & -1.2162 \end{bmatrix}$$

$$W = \begin{bmatrix} -0.0052 & -2.2753 & 1.7799 \\ -0.1292 & -1.5027 & 1.3498 \\ 0.8391 & 0.3704 & 1.1357 \\ -0.3298 & 0.4050 & 1.3161 \\ 0.2739 & -0.3676 & 0.3636 \end{bmatrix}$$

y el jacobiano

$$J_{Red} = \begin{bmatrix} -0.4251 & -5.5275 & 5.9276 \\ -0.1973 & 5.5237 & -5.7227 \\ -1.2764 & 5.4747 & -4.3265 \end{bmatrix}$$

con los valores propios:

$$\lambda_1 = -0.4885 - 3.6597i$$

$$\lambda_2 = -0.4885 + 3.6597i$$

$$\lambda_3 = -1.2510$$

similares a los del sistema a identificar. Con un tiempo de muestreo de 0.01 después de 500 iteraciones (Épocas) el error cae por debajo de 1×10^{-6} . Como se puede apreciar en la Fig. 10.

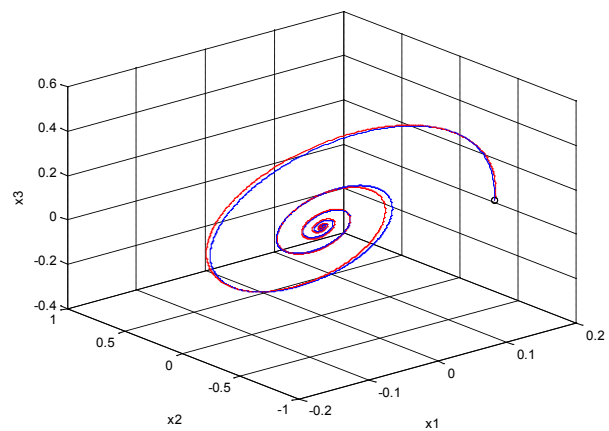


Figura 10.- Respuesta del proceso de identificación para el sistema del péndulo invertido.

V. CONCLUSIONES

Las redes neuronales son muy efectivas para modelar no linealidades en sistemas. En especial las redes recurrentes ya que pueden adquirir un comportamiento dinámico aun siendo de naturaleza estática. Otra característica importante en éste tipo de redes, es la capacidad de poder demostrar la estabilidad mediante la comparación de los eigenvalores de la red y del sistema a identificar. Si bien en muchos sistemas es difícil y en algunos casi imposible obtener su modelo matemático, las redes dinámicas recurrentes son

una excelente alternativa para abordar este problema de identificación de sistemas.

Sistemas Digitales, Programación Visual, Redes Neuronales, Robótica y Diagnóstico de fallas.

VI. REFERENCIAS

- [1] J. Chen and R.J. Patton, *Robust Model-Based fault Diagnosis for Dynamic Systems*, Boston: Kluwer Academic Publishers, 1999.
- [2] Andres Ferreyra Ramirez; tesis *Redes Neuronales Difusas para modelado via agrupamiento en-línea: aplicación a un condensador de aspiración*, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Departamento de Control Automático, México, D.F., 2005.
- [3] W.C.Chan, C.W.Chan, K.C.Cheung and Y.Wang, “*Modeling of nonlinear stochastic dynamical systems using neurofuzzy networks*”, 38th IEEE Conference on Decision and Control, 2643-2648, Phoenix, USA, 1999.
- [4] Hassan. K. Khalil, *Nolinear systems*, New Jersey, Prentice Hall, 2002.
- [5] A. Delgado, *Inteligencia Artificial y Minirobots*, Ecoe Ediciones. Santafé de Bogotá: 1998.
- [6] Abhijit S. Pandya, Robert B. Macy, *Pattern Recognition with Neural network in C++*, CRC press. Florida Atlantic University.

Juan Anzurez Marín, Nació en el estado de Puebla en 1968, recibió el título de: Ingeniero Electricista por la Facultad de Ingeniería Eléctrica de la Universidad Michoacana de San Nicolás de Hidalgo en 1991; Maestro en Ciencias en Ingeniería Electrónica opción instrumentación por la División de Estudios Superiores del Instituto Tecnológico de Chihuahua en 1997 y Dr. en Ciencias en Ingeniería Eléctrica opción Control Automático por el Centro de Investigación y de Estudios Avanzados del IPN, CINVESTAV, Unidad Guadalajara en 2007. Es profesor de la Facultad de Ingeniería Eléctrica de la Universidad Michoacana desde 1987 actualmente colabora en los programas Licenciatura y Posgrado de la misma Facultad. Sus áreas de interés son Instrumentación y Control de Sistemas así como el desarrollo de algoritmos para el diagnóstico de fallas en sistemas no lineales.

M.I. Isidro Ignacio Lázaro Castillo. Recibió el título de Ingeniero Electricista por la Facultad de Ingeniería Eléctrica de la Universidad Michoacana de San Nicolás de Hidalgo, el grado de maestro en ingeniería por la misma institución en 1992 y 1999 respectivamente. Actualmente es profesor de tiempo completo en la misma facultad. Ha publicado diversos artículos científicos nacionales e internacionales y de revistas. Autor del libro “*Ingeniería de Sistemas de Control Continuo*”. Sus áreas de interés son Calidad de la Energía, Electrónica de Potencia, Control e Instrumentación.

VII. AUTORES

Omar Cuevas Silva. Nació en Cd. Lázaro Cárdenas Michoacán en 1981. Realizó sus estudios en el de Licenciatura en el Tecnológico de Lázaro Cárdenas, obtuvo el Título de Ingeniero Electrónico en el 2005, actualmente se encuentra realizando estudios de maestría en la división de Estudios de Posgrado de la Facultad de Ingeniería Eléctrica de la Universidad Michoacana de San Nicolás de Hidalgo. Sus áreas de interés son

PID WAVELET AUTO-SINTONIZADO CON UNA RED NEURONAL WAVENET

Cruz Tolentino J. Alberto, Ramos Velasco L. Enrique, Espejel Rivera M. Angélica.

Centro de Investigación en Tecnologías de Información y Sistemas, División de ICBI
Universidad Autónoma del Estado de Hidalgo
Carr. Pachuca-Tulancingo, Km. 4.5, Pachuca, Hidalgo. 42000, México
Tel. +52 771 712000 ext. 6734, correo electrónico: jactmac@live.com.mx

RESUMEN

En este artículo, se presenta un control PID wavelet el cual emplea una red neuronal wavenet para sintonizar sus ganancias en línea. Debido a que no existe un método y a la dificultad para sintonizar las ganancias del PID wavelet, al hacerse de manera práctica, se propone hacer uso de las redes neuronales wavenet, para sintonizar las ganancias del control PID wavelet, además sin necesidad de conocer el modelo matemático del sistema dinámico. Los parámetros de la red neuronal wavenet son actualizados por la minimización de una función de energía E , mediante técnicas de adaptación basadas en algoritmos de mínimos cuadrados medios (LMS).

Palabras clave: PID, wavelet, control, wavenet, redes neuronales.

I. INTRODUCCIÓN

El PID es el control más empleado en la industria debido a que es fácil de sintonizar y de implementar mediante circuitos analógicos [1]. Con la llegada de la era digital, fue necesario emplear un PID que pudiese programarse en una PC, PIC's, DSP, FPGA's, etc. El algoritmo del PID discreto [7] generado a partir de la discretización del PID clásico [1, 3, 8], permitió programar un PID en estas plataformas. Existen diferentes variantes del PID que también son programados mediante algoritmos, como: PID no lineal, PID robusto, PID genético, PID difuso, PID neuro-difuso, PID wavelet y PID wavenet. El PID wavelet de interés en este artículo, presentado por [10], cuya ecuación que describe al PID wavelet está dada por (26), teniendo la ventaja de proveer una señal de control suave, mejora el rechazo al

ruido y a las perturbaciones, en comparación a un PID clásico [10]. Pero a diferencia del PID clásico en el PID wavelet no existe un método para sintonizar sus ganancias y debido a que el PID wavelet puede tener más de tres ganancias dependiendo de su nivel (i.e. $N+1$ ganancias, siendo N el nivel de descomposición de la señal del error y a su vez describe el nivel del PID wavelet), por lo que, sintonizarlo de manera práctica puede ser difícil y en algunos casos imposible dependiendo del tipo de sistema dinámico tal es el caso de un sistema inestable, un sistema no lineal, donde no se conozca el modelo matemático del sistema, etc.. Pero si los únicos datos accesibles son la entrada y la salida del sistema, una solución a este problema es emplear el esquema de un PID wavenet [11], para sintonizar sus ganancias.

El presente artículo está organizado de la siguiente manera: en la Sección II, se presentan la teoría empleada para el desarrollo del PID wavelet, el algoritmo de la red neuronal wavenet, el algoritmo para la sintonización de las ganancias del control y el algoritmo para actualizar los parámetros de la red neuronal wavenet. En la Sección III, se presentan los resultados de las simulaciones, las conclusiones son dadas en la Sección IV y por último en la Sección V se dan las referencias empleadas para el desarrollo de este artículo.

II. TEORÍA

A. DESCRIPCIÓN DEL SISTEMA

Considere un sistema dinámico no lineal SISO (una entrada – una salida) descrito por la ecuación de estado discreto [4, 5]

$$x(k+1) = f[x(k), u(k), k], \quad (1)$$

$$y(k) = g[x(k), k], \quad (2)$$

donde $x(k) \in \mathbb{R}^n$ y $u(k), y(k) \in \mathbb{R}$. Además con las funciones desconocidas $f, g \in \mathbb{C}^1$. Donde los únicos datos accesibles son la entrada y la salida del sistema, entonces, si el sistema linealizado alrededor del punto de equilibrio es observable [4, 5], existe una representación entrada-salida la cual es descrita por

$$y(k+1) = \Psi[\mathbf{Y}(k), \mathbf{U}(k)], \quad (3)$$

donde

$$\mathbf{Y}(k) = [y(k) \quad y(k-1) \quad \dots \quad y(k-n+1)], \quad (4)$$

$$\mathbf{U}(k) = [u(k) \quad u(k-1) \quad \dots \quad u(k-n+1)], \quad (5)$$

i.e. existe una función Ψ que mapea a $y(k), u(k)$ y sus $n-1$ valores pasados, dentro de $y(k+1)$.

Entonces un modelo de una red neuronal wavenet $\hat{\Psi}$, puede ser entrenado para aproximar a Ψ sobre el dominio de interés. Prácticamente, incluso si el modelo de la planta está disponible, los modelos aproximados se adaptan para actualizar los parámetros del control en línea. Un modelo alternativo de una planta desconocida que puede simplificar el algoritmo de la entrada de control es descrito por

$$y(k+1) = \Phi[\mathbf{Y}(k), \mathbf{U}(k)] + \Gamma[\mathbf{Y}(k), \mathbf{U}(k)] \cdot u(k), \quad (6)$$

donde $y(k)$ y $u(k)$ denotan la salida y la entrada del sistema en el instante de tiempo k , respectivamente. Si los términos $\Phi(\bullet)$ y $\Gamma(\bullet)$ son exactamente conocidos, el control requerido $u(k)$ para lograr una salida deseada $y_{ref}(k+1)$, puede ser calculada para cada instante de tiempo k usando la siguiente fórmula

$$u(k) = \frac{y_{ref}(k+1) - \Phi[\mathbf{Y}(k), \mathbf{U}(k)]}{\Gamma[\mathbf{Y}(k), \mathbf{U}(k)]}. \quad (7)$$

Sin embargo si los términos $\Phi(\bullet)$ y $\Gamma(\bullet)$ son desconocidos, entonces aquí es donde se emplea un modelo de redes neuronales wavenet para

aproximar el sistema dinámico, como se verá más adelante en los incisos **B** y **C**.

En la Figura 1, se muestra el esquema empleado para el control de un sistema SISO con un control PID wavelet auto-sintonizado con una red neuronal wavenet.

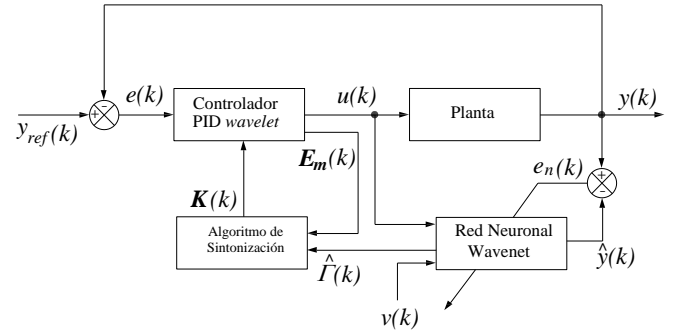


Figura 1: Esquema a bloques del PID wavelet auto-sintonizado con una red neuronal wavenet para un sistema dinámico SISO.

B. LAS WAVELETS, EL ANÁLISIS MULTIRRESOLUCIÓN Y LAS WAVENETS

Una wavelet se define como una onda de duración muy corta, algunas veces llamada “ondita” u “ondoleta”. La función wavelet $\psi(t)$ es llamada wavelet madre porque de las diferentes wavelets generadas a partir de ella, por su dilatación o contracción y translación, se les llama wavelets hijas $\psi_{a,b}(\tau)$ y la representación matemática de la wavelet hija es

$$\psi_{a,b}(\tau) = \frac{1}{\sqrt{a}} \psi\left(\frac{\tau-b}{a}\right) \quad a > 0; \quad a, b \in \mathbb{R}; \quad \tau = \frac{t-b}{a}, \quad (8)$$

donde a es la variable de escala, que permite hacer dilataciones y contracciones; y b es la variable de translación, que permite el desplazamiento en el tiempo. Existen diferentes wavelet madres como: Haar, Mexican hat, Morlet, Meyer, Daubechies, RASP1, entre otras [2]. En este artículo se hace uso de la wavelet madre Morlet, para generar a partir de ella las wavelet hijas empleadas en la red neuronal wavenet. La representación matemática de la wavelet Morlet y su derivada parcial con respecto a b son

$$\psi(t) = e^{-\frac{t^2}{2}} \cos(5t), \quad (9)$$

$$\frac{\partial \psi_{a,b}(\tau)}{\partial b} = \frac{1}{a} [\omega_0 \sin(\omega_0 \tau) e^{-0.5\tau^2} + \tau h(\tau)]. \quad (10)$$

En el análisis multirresolución introducido por [6], para funciones cuadráticas integrables $f[n] \in L^2(\mathbb{R})$, consiste de una secuencia de subespacios cerrados sucesivos, V_j que satisfacen [2, 6]

$$V_j \subset V_{j-1} \quad \forall j \in \mathbb{Z}, \quad (11)$$

$$\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R}), \quad (12)$$

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\}, \quad (13)$$

donde V_j es el subespacio generado por funciones base wavelet, descritas como

$$\phi_{j,k}[n] = 2^{-\frac{j}{2}} \phi(2^{-j}n - k) \quad \forall j, k \in \mathbb{Z}. \quad (14)$$

Las funciones base wavelet son generadas por la dilatación y translación de una función wavelet $\phi[n] \in V_0$ llamada función de escalonamiento o función wavelet padre. Si para cada V_j existe un espacio complementario W_j que cumple con las siguientes condiciones

$$V_{j-1} = V_j \oplus W_j, \quad (15)$$

$$V_j \perp W_j = 0 \quad \forall j \in \mathbb{Z}, \quad (16)$$

donde W_j es el subespacio generado por funciones base wavelet, descritas como

$$\psi_{j,k}[n] = 2^{-\frac{j}{2}} \psi(2^{-j}n - k) \quad \forall j, k \in \mathbb{Z}. \quad (17)$$

Las funciones base wavelet son generadas por la dilatación y translación de una función wavelet $\psi[n]$ llamada función wavelet madre.

Así las funciones base wavelets inducen la descomposición ortogonal de una función $f[n] \in L^2(\mathbb{R})$, como

$$f[n] = \sum_k c_{N,k} \phi_{N,k}[n] + \sum_{m=1}^N \sum_k d_{m,k} \psi_{m,k}[n], \quad (18)$$

$$c_{m,k} = \sum_k f[n] \overline{\phi_{m,k}[n]}, \quad (19)$$

$$d_{m,k} = \sum_k f[n] \overline{\psi_{m,k}[n]}, \quad (20)$$

donde N es el nivel al cual se descompone la función $f[n]$. Las funciones $\overline{\phi[n]}$ y $\overline{\psi[n]}$ son las funciones conjugadas de $\phi[n]$ y $\psi[n]$, respectivamente. Una forma eficiente de realizar el análisis multirresolución, es usar el esquema de codificación de subbanda [12], el cual usa únicamente los filtros $h[k]$ y $g[k]$, los cuales son encontrados por [2], y se relacionan de la siguiente manera

$$h[k] = \sqrt{2} \sum_n \phi[n] \overline{\phi[2n - k]}, \quad (21)$$

$$g[k] = \sqrt{2} \sum_n \psi[n] \overline{\psi[2n - k]}, \quad (22)$$

$$g[k] = (-1)^k \overline{h[-k + 1]}. \quad (23)$$

En la Figura 2, muestra el esquema empleado para la descomposición de la señal del error e para un nivel 3 de descomposición, usada para el control PID wavelet.

De la ecuación (18), si N es lo suficientemente grande, entonces $f[n] \in L^2(\mathbb{R})$ puede ser aproximada arbitrariamente cercana en W_j . Esto es que para cualquier $\varepsilon > 0$, tal que

$$\left\| f[n] - \sum_k c_{N,k} \phi_{N,k}[n] \right\| < \varepsilon, \quad (24)$$

la aproximación por la descomposición wavelet truncada puede ser expresada como [14, 15]

$$f[n] \approx \sum_k c_{N,k} \phi_{N,k}[n], \quad (25)$$

la ecuación (25) dice, que algunos componentes de baja escala (alta frecuencia) que pertenecen al espacio wavelet W_j para la función son despreciados y los componentes de alta escala (baja frecuencia) que pertenecen al espacio escalamiento V_j son preservados para aproximar la función original bajo la escala N . Esta expresión tiene una similitud con una red neuronal de tres capas, como se muestra en la Figura 3. En la Figura 3 muestra una estructura que es similar a una red neuronal de base radial, donde la función de escalamiento wavelet reemplaza el rol de una función de activación en la capa oculta, teniendo la siguiente relación $w_l = c_{N,l}$, $h_l(\tau) = \phi_{N,l}[n]$, donde $\tau = \frac{k-b_l(k)}{a_l(k)} \quad \forall l, l=1,2,\dots,L-1,L$ siendo L igual al número de neuronas en la red neuronal wavelet.

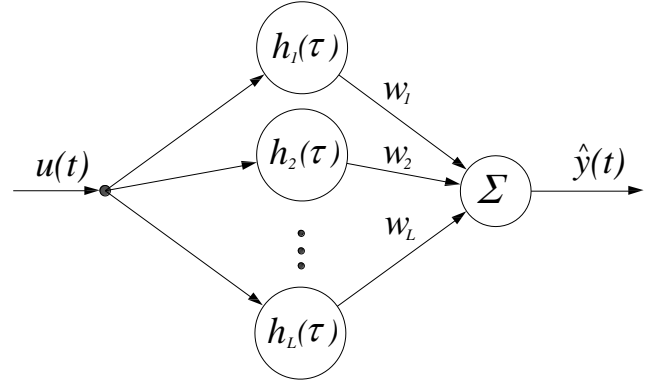


Figura 3: Estructura de la red neuronal wavelet de 3 capas.

C. CONTROL PID WAVELET

El control PID wavelet, descompone la señal del error e haciendo uso del análisis multirresolución, escalando cada componente de la señal del error con sus respectivas ganancias y adicionados todos juntos para generar la señal de control u , como

$$u = K_H e_H + K_{m_1} e_{m_1} + \dots + K_{M_{N-1}} e_{M_{N-1}} + K_L e_L, \quad (26)$$

por simplicidad, para emplearlo en el esquema mostrado en la Figura 1, se usará la siguiente forma matricial para un instante de tiempo k

$$\mathbf{K}(k) = [K_H(k) \quad K_{M_1}(k) \quad \dots \quad K_i(k) \quad \dots \quad K_{M_{N-1}}(k) \quad K_L(k)]^T, \quad (27)$$

$$\mathbf{E}_m(k) = [e_H(k) \quad e_{M_1}(k) \quad \dots \quad e_i(k) \quad \dots \quad e_{M_{N-1}}(k) \quad e_L(k)], \quad (28)$$

$$u(k) = \mathbf{K}^T(k) \mathbf{E}_m(k). \quad (29)$$

Puesto que hay un número diferentes de wavelet empleada para la descomposición, la selección de la wavelet afecta en el funcionamiento del controlador. Para mayores detalles sobre la selección de la wavelet, ver [9]. En la construcción del PID wavelet en este artículo se emplea la wavelet madre Daubechies de orden 2 [2], basado en los resultados [10] al encontrarse muy conveniente para control.

Todos los sistemas físicos están sujetos algún tipo de señales externas o ruidos durante la operación. En la práctica, las perturbaciones y comandos son algunas veces señales de baja frecuencia y el ruido es de alta frecuencia. Por lo que emplear un control

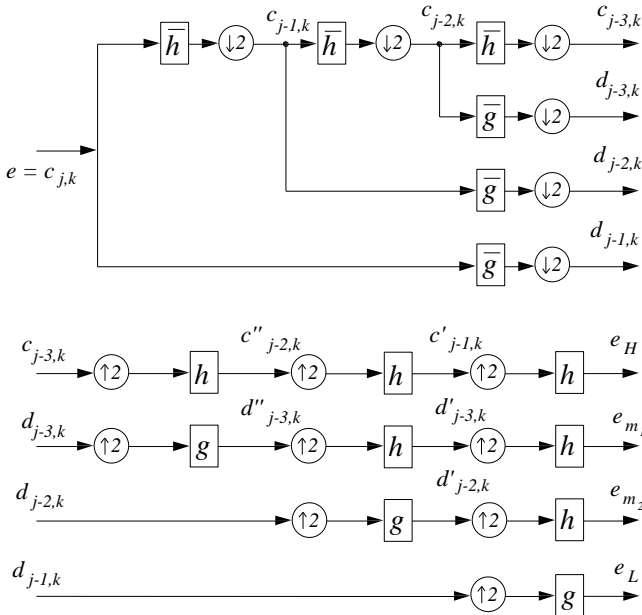


Figura 2: Esquema de codificación subbanda para la descomposición de la señal del error a un nivel de N=3.

PID wavelet que genera una señal de control muy suave y que reduce drásticamente los efectos del ruido en la salida del sistema dinámico, es muy conveniente, sobre todo porque, la señal suave de control produce un esfuerzo mínimo, mejorando la vida del actuador (comúnmente motores de corriente directa) y todo el comportamiento del sistema dinámico [3, 8].

Para poder sintonizar las ganancias del control PID wavelet $\mathbf{K}(k)$, primero se hace uso de la red neuronal wavenet junto con un filtro IIR (Respuesta al Impulso Infinito) en cascada como se muestra en la Figura 4, para aproximar las funciones $\Phi(\bullet)$ y $\Gamma(\bullet)$ de la ecuación (6), como [11]

$$\hat{\Phi}[\mathbf{Y}(k), \mathbf{U}(k)] = \mathbf{D}^T(k) \hat{\mathbf{Y}}(k) \cdot v(k), \quad (30)$$

$$\hat{\Gamma}[\mathbf{Y}(k), \mathbf{U}(k)] = \mathbf{C}^T(k) \mathbf{Z}(k), \quad (31)$$

$$z(k) = \mathbf{H}^T(k) \mathbf{W}(k), \quad (32)$$

por lo tanto

$$\hat{y}(k+1) = \hat{\Phi}[\mathbf{Y}(k), \mathbf{U}(k)] + \hat{\Gamma}[\mathbf{Y}(k), \mathbf{U}(k)] \cdot u(k), \quad (33)$$

donde

$$\mathbf{W}(k) = [w_1(k) \ w_2(k) \ \dots \ w_{L-1}(k) \ w_L(k)]^T, \quad (34)$$

$$\mathbf{A}(k) = [a_1(k) \ a_2(k) \ \dots \ a_{L-1}(k) \ a_L(k)]^T, \quad (35)$$

$$\mathbf{B}(k) = [b_1(k) \ b_2(k) \ \dots \ b_{L-1}(k) \ b_L(k)]^T, \quad (36)$$

$$\mathbf{C}(k) = [c_0(k) \ c_1(k) \ \dots \ c_m(k) \ \dots \ c_{M-1}(k) \ c_M(k)]^T, \quad (37)$$

$$\mathbf{D}(k) = [d_1(k) \ d_2(k) \ \dots \ d_j(k) \ \dots \ d_{J-1}(k) \ d_J(k)]^T, \quad (38)$$

$$\mathbf{H}(k) = [h_1(\tau) \ h_2(\tau) \ \dots \ h_l(\tau) \ \dots \ h_{L-1}(\tau) \ h_L(\tau)]^T, \quad (39)$$

$$\mathbf{Z}(k) = [z(k) \ z(k-1) \ \dots \ z(k-m) \ \dots \ z(k-M+1) \ z(k-M)]^T, \quad (40)$$

$$\hat{\mathbf{Y}}(k) = [\hat{y}(k-1) \ \hat{y}(k-2) \ \dots \ \hat{y}(k-j) \ \dots \ \hat{y}(k-J+1) \ \hat{y}(k-J)]^T. \quad (41)$$

Los parámetros de la red neuronal wavenet y las ganancias del control wavelet son optimizados con el algoritmo de mínimos cuadrados medios (LMS) por medio de la minimización de una función de energía \mathbf{E} , a lo largo de todo el tiempo $k \ \forall \ k=1,2,\dots,T$. Si denotamos al error entre la

salida de la planta y , y la salida de la red neuronal wavenet \hat{y} como $e_n(k) = y(k) - \hat{y}(k)$ para un instante de tiempo k , entonces la función de energía se define como (42)

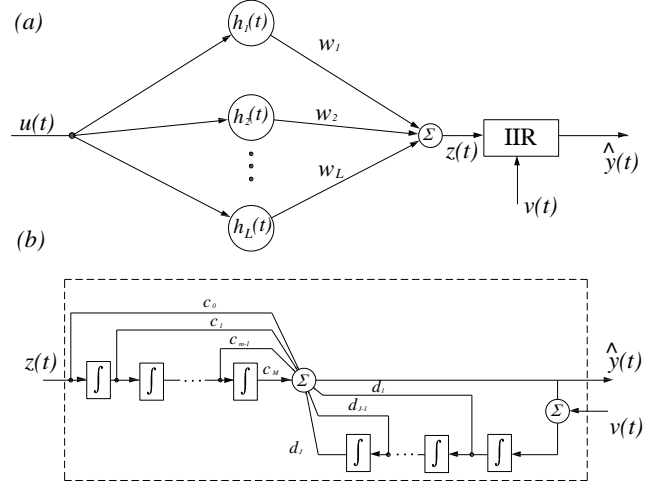


Figura 4: a) Esquema de una red neuronal wavenet con IIR, y b) la estructura del filtro IIR.

$$\mathbf{E} = \frac{1}{2} \mathbf{E}_n^T \mathbf{E}_n, \quad (42)$$

con

$$\mathbf{E}_n = [e_n(1) \ e_n(2) \ \dots \ e_n(k) \ \dots \ e_n(T-1) \ e_n(T)]. \quad (43)$$

Entonces, para minimizar \mathbf{E} se usa el método del gradiente de pasos descendentes, el cual requiere los gradientes $\frac{\partial \mathbf{E}}{\partial \mathbf{W}(k)}$, $\frac{\partial \mathbf{E}}{\partial \mathbf{A}(k)}$, $\frac{\partial \mathbf{E}}{\partial \mathbf{B}(k)}$, $\frac{\partial \mathbf{E}}{\partial \mathbf{C}(k)}$, $\frac{\partial \mathbf{E}}{\partial \mathbf{D}(k)}$ y $\frac{\partial \mathbf{E}}{\partial \mathbf{K}(k)}$. Para actualizar los parámetros con la siguiente regla [11]

$$\Delta \boldsymbol{\theta}(k) = - \frac{\partial \mathbf{E}}{\partial \boldsymbol{\theta}(k)}, \quad (44)$$

$$\boldsymbol{\theta}(k+1) = \boldsymbol{\theta}(k) + \mu_0 \Delta \boldsymbol{\theta}(k), \quad (45)$$

donde $\boldsymbol{\theta}$ puede ser \mathbf{W} , \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{D} , o \mathbf{K} , y $\mu_0 \in \mathbb{R}$ es el coeficiente de velocidad de aprendizaje para cada uno de los parámetros.

El algoritmo de sintonización para las ganancias del control PID wavelet auto-sintonizado, es descrito como

$$\frac{\partial \mathbf{E}}{\partial \mathbf{K}(k)} = \left[\frac{\partial \mathbf{E}}{\partial K_H(k)} \quad \frac{\partial \mathbf{E}}{\partial K_{m_1}(k)} \quad \cdots \quad \frac{\partial \mathbf{E}}{\partial K_i(k)} \quad \cdots \quad \frac{\partial \mathbf{E}}{\partial K_{m_{N-1}}(k)} \quad \frac{\partial \mathbf{E}}{\partial K_L(k)} \right], \quad (46)$$

$$\frac{\partial \mathbf{E}}{\partial K_i(k)} = -e_n(k) \hat{\Gamma}[\mathbf{Y}(k), \mathbf{U}(k)] \cdot e_i(k). \quad (47)$$

III. RESULTADOS DE SIMULACIÓN

El control PID wavelet auto-sintonizado con una red neuronal wavenet es simulado en el control de movimiento para un motor de CD con la siguiente función de transferencia, tomado de [12]

$$G(s) = \frac{22}{s(s+4)} \quad (48)$$

Los resultados son comparados con el control PID wavelet, el cual se sintonizo con las ganancias que se muestran en la Tabla 1, obtenidas de manera manual.

PID wavelet	K_H	K_{m_1}	K_{m_2}	K_L
	0.15	2.5	2.5	0

Tabla 1: Ganancias del control PID wavelet.

El control PID wavelet auto-sintonizado con una red neuronal wavenet, tiene los siguientes parámetros para la red neuronal wavenet mostrados en la Tabla 2 y con valores iniciales mostrados en la Tabla 3.

Neuronas	5
Wavelet madre	Wavelet Morlet
Coefficientes del filtro IIR, c	2
Coefficientes del filtro IIR, d	2

Tabla 2: Parámetros de la red neuronal wavenet.

El sistema es simulado con una red neuronal wavenet sin entrenamiento previo, pero si se tiene un periodo de aprendizaje $0 \leq t \leq 120$ segundos, para identificar al sistema.

En la Figura 5, se observa como después del periodo de aprendizaje se obtiene una respuesta del PID wavelet auto-sintonizado aceptable con un sobre impulso no mayor del 10 % y también se muestra la respuesta del sistema con un PID wavelet sintonizado de manera manual.

W	$[-0.5 \quad -0.5 \quad -0.5 \quad -0.5 \quad -0.5]$	μ_W	0.2
A	$[10 \quad 10 \quad 10 \quad 10 \quad 10]$	μ_A	0.2
B	$[0 \quad 30 \quad 60 \quad 90 \quad 120]$	μ_B	0.2
C	$[0.1 \quad 0.1]$	μ_C	0.2
D	$[0.1 \quad 0.1]$	μ_D	0.2
K	$[0.1 \quad 0.5 \quad 0.5 \quad 0.1]$	μ_K	0.5

Tabla 3: Valores iniciales de los parámetros de la red neuronal wavenet y las ganancias del control PID wavelet a sintonizar.

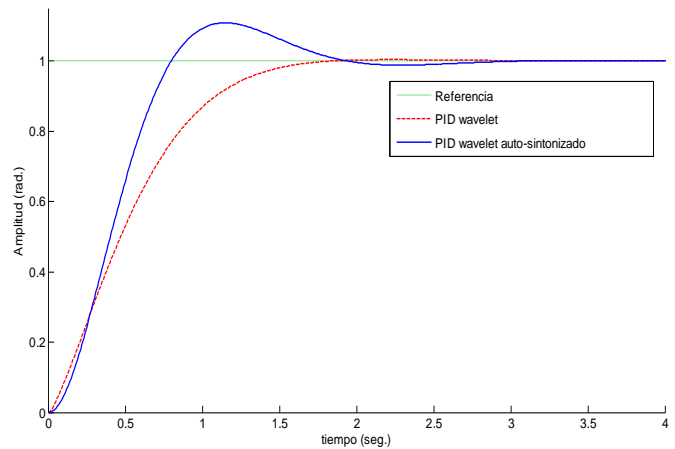


Figura 5: Señal de salida del sistema con un control PID wavelet auto-sintonizado y un control PID wavelet.

En la Figura 6, se observa como la señal de control es mucho menor que la que se genera con el control wavelet, y aun así se tiene una rápida respuesta como se ve en la Figura 5, y esto se debe a la sintonización de las ganancias realizado por el algoritmo de sintonización.

En la descomposición de la señal del error mostrada en la Figura 7, podemos observar los diferentes componentes que componen a la señal del error e mediante el análisis multirresolución.

En las Figuras 8, 9 y 10, se muestran los comportamientos de los parámetros de la red neuronal wavenet y del filtro IIR, así como el comportamiento de las ganancias del PID wavelet auto-sintonizado durante el periodo de identificación.

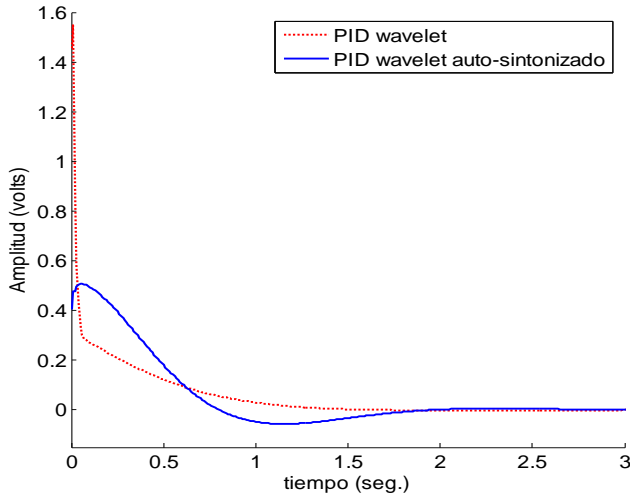


Figura 6: Señal de control generada por el control PID wavelet auto-sintonizado y el control PID wavelet.

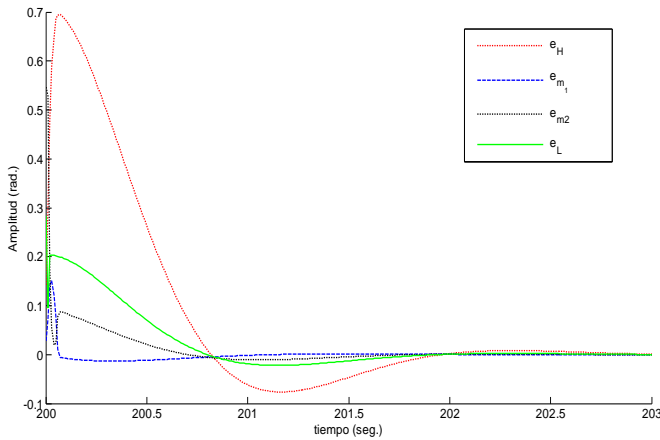


Figura 7: Descomposición de la señal del error del sistema con un control PID wavelet auto-sintonizado.

Observándose que con poco periodo de identificación y sin previo entrenamiento de la red neuronal wavenet, se logra una buena salida del sistema como se muestra en la Figura 5. Esto se debe a que el sistema dinámico es lineal, para sistemas con no linealidades es posible que se

requiera un entrenamiento previo de la red neuronal wavenet.

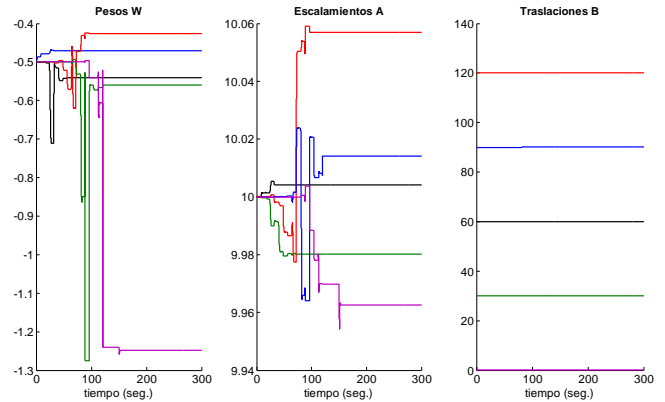


Figura 8: Comportamiento de los parámetros de la red neuronal wavenet durante la simulación.

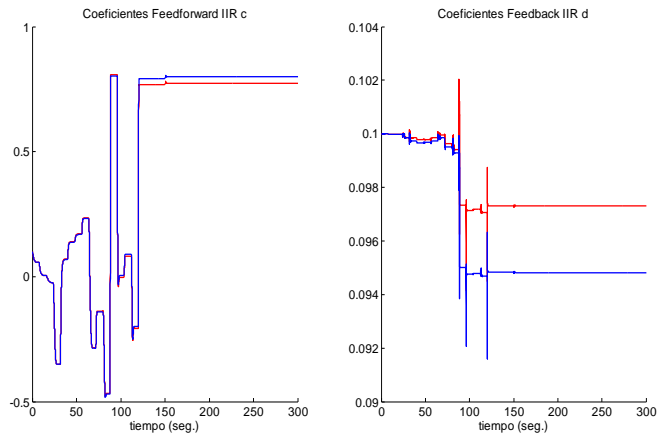


Figura 9: Comportamiento de los coeficientes del filtro IIR durante la simulación.

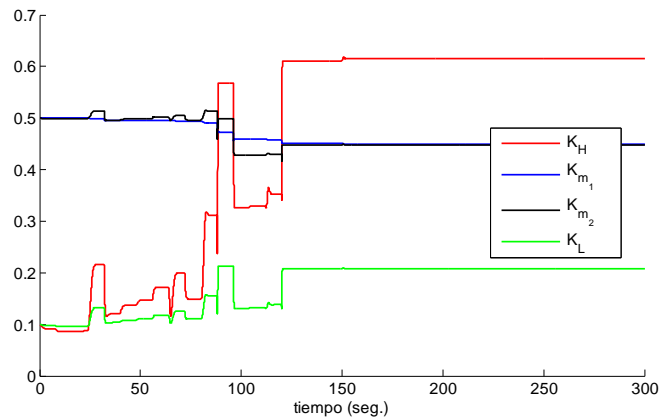


Figura 10: Comportamiento de las ganancias del control PID wavelet auto-sintonizado durante la simulación.

Los valores finales son mostrados en la siguiente tabla

W	[-0.426 -0.47 -0.54 -0.56 -1.25]
A	[10.057 10.014 10.004 9.98 9.962]
B	[0.003 30.019 60.065 90.15 120.018]
C	[0.775 0.8]
D	[0.0973 0.0948]
K	[0.615 0.45 0.448 0.2]

Tabla 4: Valores finales de los parámetros de la red neuronal wavenet y las ganancias del control PID wavelet auto-sintonizado.

IV. CONCLUSIONES

Se ha presentado un control PID wavelet auto-sintonizado con una red neuronal wavenet, de los resultados obtenidos en la simulación se puede concluir:

- El emplear una descomposición basada en wavelet nos permite ser selectivos en los componentes de la señal del error y con ello poder obtener una señal de control suave, incluso ante la presencia de ruido.
- El emplear redes neuronales wavenet permite la sintonización de un PID wavelet en línea, sin necesidad de conocer el modelo matemático del sistema dinámico.
- De lo anterior se puede decir que es posible emplear un control PID wavelet auto-sintonizado con una red neuronal wavenet en sistemas no lineales

Trabajo Futuro.

Actualmente, se está implementando este algoritmo para un sistema no lineal.

V. REFERENCIAS

[1] K. Aström and T. Häggglund. *Advanced PID Control*. ISA, 2006.

[2] I. Daubechies. *Ten lectures on Wavelets*. SIAM, 1992.

[3] B. Kuo. *Sistemas de Control Digital*. C.E.C.S.A., 1997.

[4] A. Levin and K. Narendra. *Control of Nonlinear Dynamical Systems Using Neural Networks: Controllability and Stabilization*. IEEE Transactions on Neural Networks, Vol. 4, pp. 192-206, 1993.

[5] A. Levin and K. Narendra. *Control of Nonlinear Dynamical Systems Using Neural Networks - Part II: Observability, Identification, and Control*. IEEE Transactions on Neural Networks, Vol. 7, pp. 30-42, 1996.

[6] S. Mallat. *A Theory Multiresolution Signal Decomposition: The Wavelet Representation*. IEEE Transactions Pattern Analysis and Machine Intelligent, Vol 7, pp. 674-693, Jul. 1989.

[7] K. Ogata. *Sistemas de Control en Tiempo Discreto*. Prentice-Hall, 1996.

[8] K. Ogata. *Ingeniería de Control Moderno*. Prentice-Hall, 1998.

[9] S. Parvez. *Advance Control Techniques for Motion Control Problem*. Ph. D. Dissertation, Cleveland State University, Cleveland, Ohio, USA, Dec. 2003.

[10] S. Parvez and Z. Gao. *A Wavelet-Based Multiresolution PID Controller*. IEEE Transaction on Industry Applications, Vol. 41, pp. 537-543, Mar.-Abril. 2005.

[11] M. Sedighzadeh and A. Rezazadeh. *Adaptive PID Control of Wind Energy Conversion Systems Using RASPI Mother Wavelet Basis Function Network*. Proceeding of World Academy of Science, Engineering and Technology, Vol. 27, pp. 269-273, Feb. 2008.

[12] J. Tang. *PID Controller Using The TMS320C31 DSK with on-line Parameter Adjustment for Real-Time DC Motor Speed and Position Control*. ISIE, pp. 786-791, Pusan, Korea, 2001.

- [13] M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding*. Prentice-Hall PTR, 2007.
- [14] Q. Zhang and A. Benveniste. *Wavelet Networks*. IEEE Transactions on Neural Networks, Vol. 3, pp. 889-898, Nov. 1992.
- [15] J. Zhang, G. Walter, Y. Miao and W. Lee. *Wavelet Neural Networks for Function Learning*. IEEE Transactions on Signal Processing, Vol. 43, pp. 1485-1497, Jun. 1995.

VI. AUTORES

Ing. José Alberto Cruz Tolentino es Ingeniero Eléctrico por el Instituto Tecnológico de Pachuca, Actual alumno de la Maestría en Automatización y Control, impartida en el Centro de Investigación en Tecnologías de Información y Sistemas CITIS en la Universidad Autónoma del Estado de Hidalgo, en Pachuca, Hidalgo, México.

Dr. Luis Enrique Ramos Velasco por la Universidad Checa en Praga, República Checa. Actualmente Profesor investigador en la Universidad Autónoma del Estado de Hidalgo, en Pachuca, Hidalgo, México.

M. en C. María Angélica Espejel Rivera por el Instituto Tecnológico de Torreón, en Torreón Coahuila. Actualmente Catedrático en el Instituto Tecnológico de Pachuca, en Pachuca, Hidalgo, México.

EL PROBLEMA DEL DIAGNÓSTICO DE FALLAS: DISEÑO DE OBSERVADORES DEDICADOS

Celso Cortes Cervantes, Juan Anzures Marín, Isidro I. Lázaro Castillo

Facultad de Ingeniería Eléctrica, División de Estudios de Posgrado
Universidad Michoacana de San Nicolás de Hidalgo
Av. Fco. J. Mújica S/N, Ciudad Universitaria, Morelia, Mich. México
Tel. 443 322 3500 ext4354, j.anzures@ieee.org, celso_2c@hotmail.com, ilazaro@ieee-sco.org

RESUMEN

El presente trabajo plantea un esquema para el diagnóstico de fallas basada en el diseño de observadores dedicados, cuyo objetivo es evitar la redundancia en hardware. La idea general del esquema de observadores dedicados es construir un observador Luenberger por cada salida del sistema, de forma tal que el problema del diagnóstico de fallas se resuelve mediante la comparación de la salida del sistema y la salida estimada generando señales conocidas como residuos. Si los residuos son cero el sistema está libre de fallas en caso contrario hay presencia de fallas en el sistema. En el presente artículo se muestra una aplicación al problema de diagnóstico para sistemas no lineales para lo cual se emplea como ejemplo el modelo del péndulo invertido.

Palabras clave: Diagnóstico de fallas, esquema de observadores dedicados, instrumentación y control.

I. INTRODUCCIÓN

El área de control siempre ha sido tema interesante para avances en ingeniería y la ciencia, además de su gran importancia en procesos industriales y de manufactura para mejorar la productividad evitar el exceso de operaciones manuales o rutinarias etc.

Conforme las plantas modernas con muchas entradas y salidas se vuelven más y más complejas, la descripción de un sistema de control moderno requiere de una gran cantidad de ecuaciones. Debido a que la disponibilidad de las computadoras digitales (1960) hizo posible el análisis en el dominio del tiempo de sistemas

complejos, la teoría de control moderna basada en el análisis en el dominio del tiempo y la síntesis a partir de variables de estados, se ha desarrollado para enfrentar la creciente complejidad de las plantas modernas [7].

Aunque desde los primeros años de la década de los 70 se vienen produciendo avances significativos en el campo del diagnóstico de fallas tanto en la teoría como en la práctica, ha sido en los últimos años cuando las investigaciones en este campo han aumentado considerablemente [1], [6], [8], [9].

Tal es el caso de los observadores o estimadores que permiten estimar las variables o estados de un sistema basados en mediciones de las señales de salida y señales de control [2], [9], [10].

El presente trabajo muestra el estudio y desarrollo del Esquema de Observadores Dedicados (DOS, por sus siglas en inglés “Dedicated Observer Schemes”). El esquema básico se muestra en la Fig. 1, el cual fue propuesto por Clark y fue probado por primera vez en 1974 [3], [5], [9]. Este esquema propone utilizar un observador por cada salida del sistema o planta, de ahí es que toma su nombre. Cada uno de estos observadores se basa en el diseño de Luenberger para cada salida [1], [9].

Esto permite tener un esquema de aislamiento de fallas en la salida de manera natural, debido a la propiedad de estimación de cada salida del sistema o planta bajo análisis. El concepto de aislamiento

se refiere a la identificación plena del elemento que dentro del sistema presenta una falla. [1-5], [9].

II. DESCRIPCIÓN DE DOS

En el esquema de observadores dedicados presentado por Clark cada observador es manejado por el vector de entrada u y la salida de un sensor. De esta manera para m sensores, tendrían que existir m observadores. Estos observadores generan un vector de estados estimados, el cual se puede emplear para detectar fallas en el sistema mediante la comparación con las salidas (señales reales de los sensores), de forma tal que: "Una falla en el k -ésimo sensor provocará que el estado estimado del k -ésimo observador difiera del estado estimado de los otros $m-1$ observadores, lo cual permite la detección y aislamiento de la falla de manera directa" [1], [2], [9], [10]. Lo anterior significa que el residuo será igual a cero si el sistema está libre de fallas, en caso contrario el sistema presenta una falla.

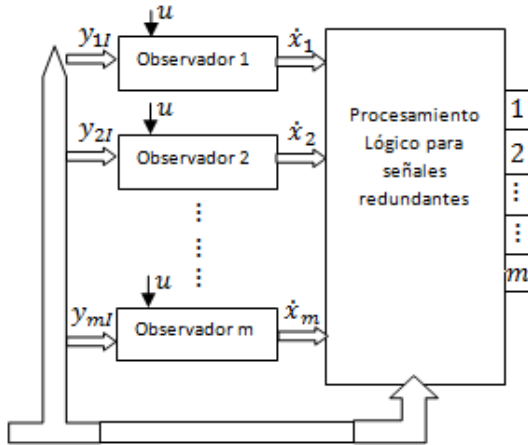


Fig. 1 Esquema de Observadores Dedicados, DOS.

Por lo que las fallas están desacopladas, es decir que cada falla de encuentra asociada con una salida, tal como se ilustra en la Fig. 2.

El modelo para este esquema expresado en variables de estado es representado de la siguiente manera:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y_1 &= C_1 x(t) + f_1(t) \\ &\vdots \\ y_k &= C_k x(t) + f_k(t) \end{aligned} \quad (1)$$

$$y_m = C_m x(t) + f_m(t)$$

Donde $x(t) \in \mathbb{R}^n$ es el vector de estado, $y(t) = [y_1(t), \dots, y_m(t)]^T$ es el vector de salida medible además A, B, y C son matrices constantes de dimensiones apropiadas ($C = [C_1 \dots C_k \dots C_m]^T$) y $f_m(t)$ representa la falla en el elemento m de salida [4].

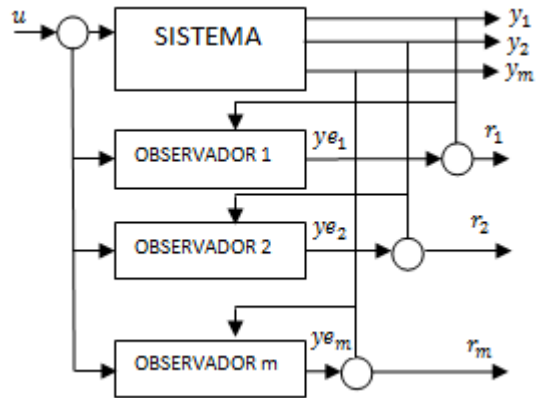


Fig. 2 Esquema de Clark para la generación de Residuos basado en DOS.

III. DISEÑO DE OBSERVADORES

Consideremos un sistema modelado en variables de estado representado por las ecuaciones:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t), x(0) = x_0 \neq 0 \end{aligned} \quad (2)$$

donde $x(t) \in \mathbb{R}^n$, el par $(u(t), y(t))$ es escalar y las matrices (A, B, C) son perfectamente conocidas.

El estimador de estados está compuesto por una reproducción del sistema más un término adicional de corrección, como se indica en la ecuación (3).

$$\dot{\xi}(t) = A\xi(t) + Bu(t) + L[y(t) - C\xi(t)] \quad (3)$$

Nótese que el observador reproduce la salida del sistema y además corrige la ecuación dinámica con un término que es proporcional al error entre la salida del sistema real ($y(t)$) y la salida estimada ($C\xi(t)$).

Definimos el error entre los estados reales del sistema y los estimados como:

$$e(t) = x(t) - \xi(t) \quad (4)$$

Entonces, la dinámica del error ($\dot{e}(t)$) a partir de (1) y (3) será:

$$\begin{aligned} \dot{e}(t) &= \dot{x}(t) - \dot{\xi}(t) \\ \dot{e}(t) &= A[x(t) - \xi(t)] - LC[x(t) - \xi(t)] \\ \dot{e}(t) &= (A - LC)e(t) \end{aligned} \quad (5)$$

De esta manera, $\xi(t) \rightarrow x(t)$ si los eigenvalores de la matriz $A-LC$ son todos estables; es decir, están en el semiplano izquierdo del plano complejo. Debido a que la solución de la ecuación:

$$\dot{x}(t) = (A - LC)e(t) \text{ es:}$$

$$\begin{aligned} e(t) &= e^{(A-LC)t} e_0 \\ e(t) &= e^{(A-LC)t} (x(0) - \xi(0)) \\ e(t) &= e^{(A-LC)t} x_0 \end{aligned} \quad (6)$$

De modo que, si los eigenvalores de $A - LC$ tienen todas partes reales negativas, el error ($e(t)$) tiende a cero sin importar su condición inicial. Lo que quiere decir que, el observador de Luenberger, bajo la suposición de conocimiento perfecto del sistema, a medida que pasa el tiempo, mejora asintóticamente la estimación de los estados.

De esta forma, el problema de diseño de un estimador de estados se reduce a la determinación de una ganancia L del observador tal que la matriz $A - LC$ sea Hurwitz, en ese sentido, el problema de diseño de un observador es equivalente al cálculo del control por realimentación de estados. En este artículo para el diseño del observador se propone utilizar las herramientas de *MATLAB*[®], por ejemplo, los comandos: `place` o `acker`, para calcular L . Para lo cual, $A^T=A$ y $C^T=B$. La K obtenida de esta manera es la transpuesta de L , ($L = K^T$).

Es importante indicar que esta técnica está basada en la propuesta de modificar la dinámica del sistema mediante la reubicación de los polos de la planta en lazo abierto. Para la ubicación de los polos del observador se debe considerar entonces que la dinámica de éste debe ser más rápida que la del sistema a observar se desea que el estado estimado alcance al estado de la planta lo más rápido posible.

Es necesario mencionar que el observador Luenberger es aplicable únicamente a sistemas deterministas lineales e invariantes en el tiempo. El observador Luenberger extendido se puede aplicar a sistemas deterministas no lineales y variantes en el tiempo.

Si el observador se emplea para estimar todas las variables de estado del sistema, independientemente de si se pueden medir o no, se le denomina observador de orden completo. En otras ocasiones esto no será necesario, pues habrá variables de estado medibles de forma directa. Las variables de estado y las de salida están relacionadas, debido a que estas últimas son medibles, sólo será necesario observar ($n-m$) variables de estado, donde n es la dimensión del vector de estado y m es la dimensión del vector de salida. Cuando el observador únicamente estima el número mínimo de variables de estado, se le denomina observador de orden mínimo. En determinadas ocasiones puede resultar más sencilla la aplicación de un observador de orden mínimo, cuando el número de variables del sistema es muy elevado.

IV. CASO DE ESTUDIO

En el presente artículo para mostrar la técnica de Diagnóstico de Fallas basada en el Esquema de Observadores Dedicados, se toma como ejemplo de aplicación el conocido modelo del péndulo invertido el cual es un sistema no lineal cuyo modelo se puede representar mediante variables de estado linealizando alrededor de un punto de operación. El diagrama del ejemplo de aplicación se muestra en la Fig 3 y para el cual se emplean los parámetros indicados en la tabla 1.

Siendo los criterios de diseño son los siguientes:

- ✓ Tiempo de establecimiento menor que 5 segundos.
- ✓ Ángulo del Péndulo nunca mayor que 0.05 radianes de la vertical.

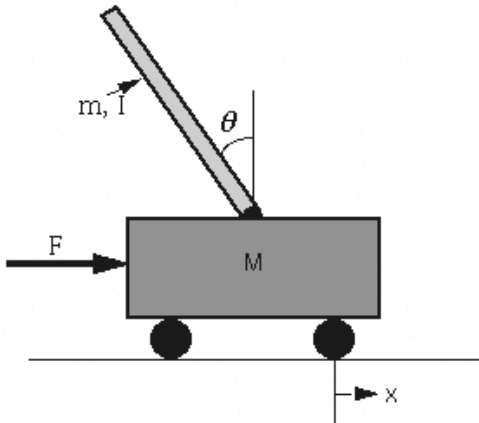


Fig. 3 Diagrama del caso de estudio (péndulo invertido)

Tabla 1.- Parámetros usados en el caso de estudio

Variable	Definición	Valor
M	masa del carro	0.5 kg
m	masa del péndulo	0.5 kg
b	fricción del carro	0.1 N/m/seg
l	longitud al centro de masa del péndulo	0.3 m
i	inercia del péndulo	0.006 kg*m ²
F	fuerza aplicada al carro	
x	coordenadas de posición del carro	
Φ	ángulo del péndulo respecto de la vertical	

Se ha considerado detectar fallas en los sensores de posición del carro y posición del péndulo respecto a la vertical. Con los parámetros anteriores el modelo en espacio de estado se puede representar como se muestra a continuación.

$$\begin{bmatrix} \dot{x} \\ \dot{\dot{x}} \\ \dot{\Phi} \\ \dot{\dot{\Phi}} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -0.1818 & 2.6727 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -0.4545 & 31.1818 & 0 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \\ \Phi \\ \dot{\Phi} \end{bmatrix} + \begin{bmatrix} 0 \\ 1.8182 \\ 0 \\ 4.5455 \end{bmatrix} u$$

$$y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{\dot{x}} \\ \dot{\Phi} \\ \dot{\dot{\Phi}} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} u$$

Para el diseño del observador se considera el de orden completo, bajo el esquema mostrado en la Fig 4. En el que además se propone sea de la forma de compensador es decir el conjunto Controlador-Observador.

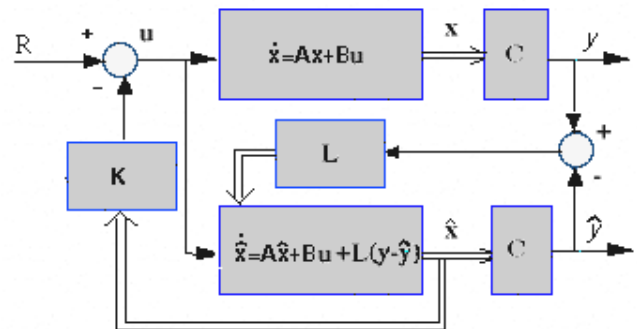


Fig. 4 Esquema de un Observador Luenberger de orden completo.

Usando métodos de espacio de estado es relativamente simple trabajar con un sistema de salida múltiple, así en este ejemplo de aplicación se diseña el controlador considerando tanto el ángulo del péndulo como la posición del carrito. El esquema del controlador se muestra en la Fig. 5.

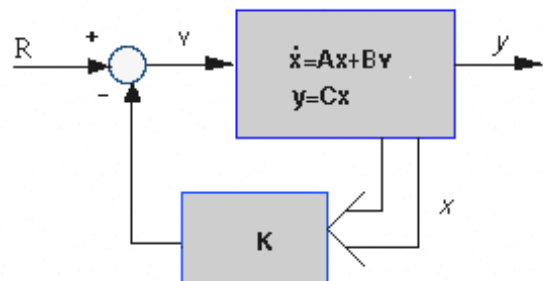


Fig. 5 Esquema de un controlador mediante retroalimentación de estados

En la Fig. 5, **R** representa una entrada al carrito del tipo escalón. Los 4 estados son: La posición y velocidad del carro; el ángulo y velocidad angular

del péndulo. La salida y contiene tanto la posición del carro como el ángulo del péndulo. El objetivo es diseñar un controlador de modo tal que para una entrada escalón, el péndulo pueda ser desplazado, y eventualmente retornar a cero (a la vertical) y el carrito debiera moverse a su posición nueva.

Para ello es necesario determinar los polos en lazo abierto de sistema:

$$p = \begin{matrix} -8.4910 + 7.9283i \\ -8.4910 - 7.9283i \\ -4.7592 + 0.8309i \\ -4.7592 - 0.8309i \end{matrix}$$

Se busca que los polos del controlador diseñado sean cerca de 4-10 veces más rápidos que el más chico de los polos, por ejemplo en -40. Se emplea el comando *acker* para hallar el vector **L**.

Las señales de salida consideradas son el ángulo de del péndulo y la posición del carro. Para el diseño del observador puesto que el sistema no es observable si usamos como salida sólo el ángulo del péndulo. Esto tiene sentido: Ya que si lo único que puede medir es el ángulo del péndulo, no puede determinar cuál será la posición del carro.

De esta manera las ganancias Luenberger **L** de cada estado son:

$$L = 1.0e+03 \begin{matrix} * \\ 0.0826 & -0.0010 \\ 1.6992 & -0.0402 \\ -0.0014 & 0.0832 \\ -0.0762 & 1.7604 \end{matrix}$$

El sistema se probó bajo dos escenarios: Primero el sistema libre de fallas en los sensores y luego se introducen fallas en los sensores tanto en el de desplazamiento y posición angular. En la Fig. 6 observamos las señales de los Residuos Generados, para el sistema libre de falla, en la que se puede apreciar una señal cerca de cero, lo que indica un sistema sano.

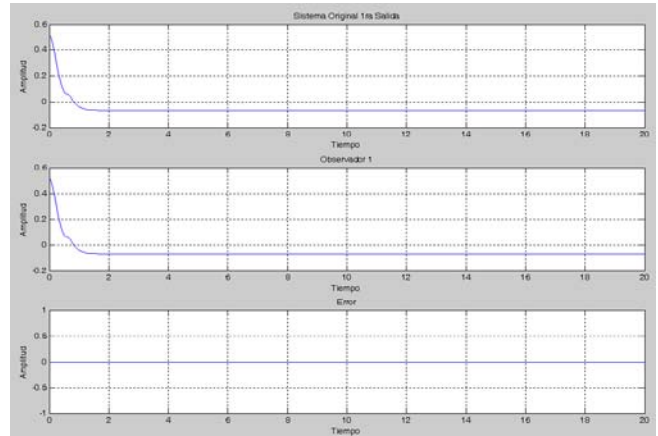


Fig. 6 Salida del sensor 1 sin fallas

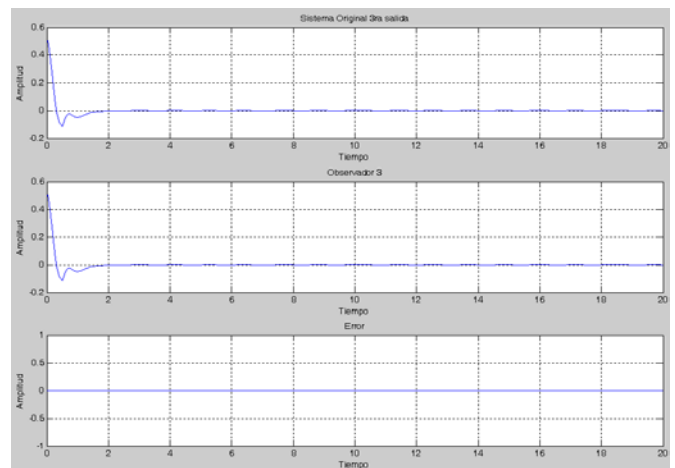


Fig. 7 Salida del sensor 2 sin fallas

Se provoca una Falla en el sensor 1 mediante una señal escalón en un tiempo de 1 seg. y amplitud final de 0.5, los residuos generados se muestran en la Fig 8. Aquí es fácil ver un residuo diferente de cero lo cual es un indicativo de que una falla ocurrió, en este caso en el sensor 1.

De igual manera la Fig 9 muestra los residuos generados por los observadores para una Falla en el sensor 2 provocada con una señal escalón en un tiempo de 2 seg. y amplitud final de 0.8

V. CONCLUSIONES

El trabajo presentado muestra que este esquema es práctico y sencillo de implementar. En las simulaciones mostradas se observa que la detección y el aislamiento se hacen de manera

directa, lo que facilita su manejo e implementación.

Como se mencionó la parte fundamental es el diseño del observador Luenberger, para el cálculo de tales ganancias las cuales deben de ser de cuatro y diez veces mayores que el valor de los polos del sistema en lazo abierto.

Se ha observado que en la actualidad el tema de robustez es considerablemente estudiado al igual que los diagnósticos de fallas dado a la importancia que representan estos temas en aplicaciones industriales, científicas y humanas, lo que permite omitir redundancia de hardware he introducir actualmente redundancia analítica. Lo cual ha permitido reducir costos y obtener mejores beneficios.

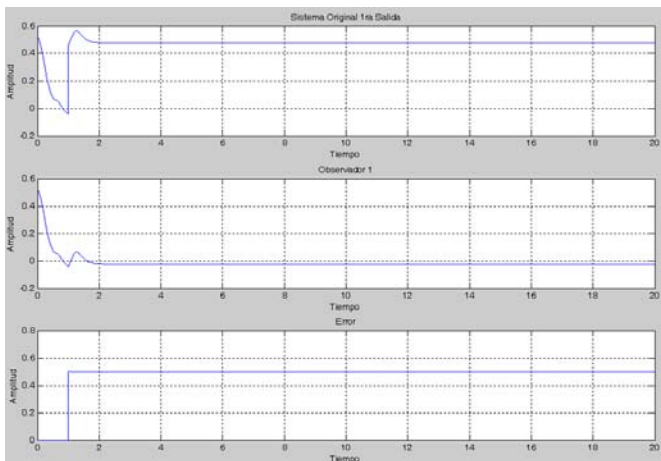


Fig. 8 Salida del sensor 1 con fallas

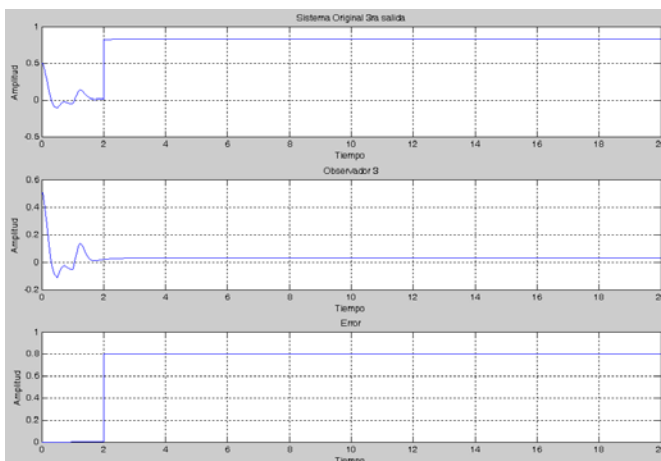


Fig. 9 Salida del sensor 2 con fallas

VI. REFERENCIAS

- [1] J. Anzurez, “*Diagnóstico de Fallas en Sistemas no Lineales usando lógica difusa y observadores con modos deslizantes*”, Tesis Doctoral, CINVESTAV, Unidad Guadalajara, 2007.
- [2] J. Chen, R.J. Patton, “*Robust Model-Based fault diagnosis for Dynamic Systems*”. Boston: Kluwer Academic Publisher, 1999.
- [3] R.N. Clark, “*The Dedicated Observer Approach to Instrument Failure Detection*”, Proceeding of the 15th IEEE Conference on Decision & Control, FL, USA, 1979, pp. 237-241.
- [4] R. Isermann, “*Supervision, fault-detection and fault diagnosis methods and introduction*”, Control Eng. Practice, Vol. 5, No 5, 1997, pp. 639-652.
- [5] R.J. Patton, P.M. Frank, R.N. Clark, “*Issues of Fault Diagnosis for Dynamic Systems*”, New York, Springer-Verlag, 2000.
- [6] B. C. Toledo and J. A. Marin, “*Model-Based Fault Diagnosis Using Sliding Mode Observers to Takagi-Sugeno Fuzzy Model*”, Proceedings of the 2005 IEEE International Symposium on Intelligent Control, Cyprus, 2005, pp. 652-657.
- [7] H. Hammouri, M. Kinnaert and E. H. El Yaagoubi , “*Application of nonlinear observers to fault detection and isolation*”, Lecture Notes in Control and Information Sciences, Nijmeier and Fossen Eds., 1999, pp. 423-443.
- [8] Alcorta Garcia, E. Frank, P.M., “*Analysis of a class of dedicated observer schemes to sensor fault isolation*”, Control '96, UKACC International Conference on (Conf. Publ. No. 427), Volume: 1, pp. 60-65, 1996.
- [9] Steven X.Ding, “*Model-Based Fault Diagnosis Techniques Design Schemes, Algorithms and Tools*”, ISBN 978-3-540-76303-1, Springer-Verlag, Berlin 2008.

- [10] S.Nagarajan, J.Shanmugam and T.R. Rangaswamy, “*MANFIS Observer-Based Sensor Fault Detection and Identification in interacting Level Process with NN Based Threshold Generator*”, International Journal of Soft Computing 3, ISSN 1816-9503, 2008.
- “*Ingeniería de Sistemas de Control Continuo*”. Sus áreas de interés son Calidad de la Energía, Electrónica de Potencia, Control e Instrumentación.

VII. AUTORES

Dr. Juan Anzurez Marín, Recibió el título de: Ingeniero Electricista por la Facultad de Ingeniería Eléctrica de la Universidad Michoacana de San Nicolás de Hidalgo en 1991; Maestro en Ciencias en Ingeniería Electrónica opción instrumentación por la División de Estudios Superiores del Instituto Tecnológico de Chihuahua en 1997 y Dr. en Ciencias en Ingeniería Eléctrica opción Control Automático por el Centro de Investigación y de Estudios Avanzados del IPN, CINVESTAV, Unidad Guadalajara en 2007. Es profesor de la Facultad de Ingeniería Eléctrica de la Universidad Michoacana desde 1987 actualmente colabora en los programas Licenciatura y Posgrado de la misma Facultad. Sus áreas de interés son Instrumentación y Control de Sistemas así como el desarrollo de algoritmos para el diagnóstico de fallas en sistemas no lineales y Energy Harvesting.

P. Ing. Celso Cortés Cervantes. Se encuentra terminando el último semestre de la Carrera de Ingeniero en Electrónica en la Facultad de Ingeniería Eléctrica de la Universidad Michoacana de San Nicolás de Hidalgo. Está desarrollando la tesis Diseño de Observadores Dedicado para el Diagnóstico de Fallas en Sistemas lineales y no lineales. Sus áreas de interés el Control e Instrumentación, El Diagnóstico de Fallas y Automatización de Procesos.

M.I. Isidro Ignacio Lázaro Castillo. Recibió el título de Ingeniero Electricista por la Facultad de Ingeniería Eléctrica de la Universidad Michoacana de San Nicolás de Hidalgo, el grado de maestro en ingeniería por la misma institución en 1992 y 1999 respectivamente. Actualmente es profesor de tiempo completo en la misma facultad. Ha publicado diversos artículos científicos nacionales e internacionales y de revistas. Autor del libro

NONLINEAR OBSERVER FOR LATERAL VELOCITY USING ROLL ANGULAR ACCELERATION

Cuauhtemoc Acosta Lua*, Bernardino Castillo Toledo** and Stefano Di Gennaro*.

*Department of Electrical and Information Engineering and Center of Excellence DEWS
University of L'Aquila, Poggio di Roio, 67040 L'Aquila, Italy
E.mail: cacosta@gdl.cinvestav.mx, digennar@ing.univaq.it

**Centro de Investigación y de Estudios Avanzados { CINVESTAV del IPN, Unidad Guadalajara
Av. Científica, Col. El Bajío, Zapopan, 45010, Jalisco, México
E.mail: toledo@gdl.cinvestav.mx

ABSTRACT

In this paper we present a nonlinear observer for the estimation of the lateral velocity and the roll position and velocity, based on measurements of the longitudinal and lateral acceleration, longitudinal velocity, yaw rate and the steer angle. The nonlinear observer use a tyre-road friction model by the Pacjeka magic formula and produce the observer produce reliable estimates throughout the linear and nonlinear handling region. Exponentially stability of the observer is obtained. The nonlinear observer has been validated by numerical simulations to demonstrate the efficacy of our nonlinear observer.

Keywords: Ground vehicles, Nonlinear Observer, Vehicle Stability.

I. INTRODUCTION

The feedback control system for active safety in automotive applications have entered production cars. Many of these systems, which Electronic Stability Program (ESP) [19], Vehicle Stability Assist (VSA), Dynamic Stability Control (DSC), and nonlinear control [1] and [15] have in common that the control action depends on information about vehicle velocity. However the lateral velocity is rarely measured directly due to the cost and space in the vehicle and must therefore inferred from other measurement, such as longitudinal and lateral acceleration, longitudinal velocity, yaw rate and steer angle.

Earlier works on observer for estimation of lateral velocity without roll angular velocity are mainly based on linear techniques [5], [4] or quasilinear techniques for example [20], [17]. A nonlinear observer linearizing the observer error dynamics is proposed in [9] and [11]. The same type of observer, in addition to an observer on forcing the dynamics of the nonlinear estimation error to the dynamics of a linear reference system is investigated in [7]. Other types of observers linear and nonlinear using the sliding mode techniques are [3] and [14].

Other observer proposed found in literature are based on the Estimation Kalman Filter (EKF). In this case an EKF is used for estimating vehicle velocity and tire forces in [13], thus without the explicit use of friction models. An EKF based on a tire-road friction model which also includes estimation of the adhesion coefficient and road inclination angle is suggested in [16]. In [2], the use of an EKF based on a nonlinear tire-road friction model is considered, which also includes estimation of cornering stiffness. The strategy proposed in [10] combine dynamic and kinematics models of the vehicle with numerical band-limited integration of the equations to provide a side-slip estimate. In [6] the side-slip angle is estimated along with yaw rate in an approach which is similar to the one considered herein, but without yaw rate measurements.

The goal of this work is to develop and efficiently nonlinear observer for the vehicle velocity and roll position using only the information available, i.e. information from the sensor like longitudinal and lateral acceleration, longitudinal velocity and yaw rate in addition to the steer angle and principally this observer guarantees the asymptotically stability. A significative advantage of the proposed approach over the EKF is that real-time solution of the Riccati differential equations is avoided then our observer is implemented more efficiently because the gains obtained in this observer are easy for the implementation on the electronic control unit.

We proposed a kinematical model relating the longitudinal velocity, lateral velocity, longitudinal acceleration, lateral acceleration, yaw rate and position and velocity roll dynamics. This model contain an advantages that contain vehicle parameters, and hence the kinematical model based observer is unaffected by changes in the vehicle parameters. The observer produce reliable longitudinal, lateral velocity and position and velocity roll angle throughout the linear and nonlinear handling regions.

II. MATHEMATICAL PROBLEM

The dynamic models presented in this section are used in the numerical simulations to estimate the vehicle behavior. In the Figure 1, we show the vehicle model used for the numerical simulation within the study has four degrees of freedom: longitudinal velocity v_x , lateral velocity v_y , yaw rate ω_z and body roll α_x . The vehicle longitudinal, lateral and yaw dynamics can be expressed using the so called bicycle model [8], [18].

$$\begin{aligned} m(\dot{v}_x - \omega_z v_y) &= a_x - m_s h \omega_z \dot{\alpha}_x \\ m(\dot{v}_y + \omega_z v_x) &= a_y - m_s h \omega_z \ddot{\alpha}_x \\ J_z \dot{\omega}_z &= l_f F_{y,f} - l_r F_{y,r} \end{aligned} \quad (1)$$

with m, J_z the vehicle mass and inertia with respect to the perpendicular axis, l_f, l_r the distances from the vehicle center of gravity (CG) to the tires,

a_x, a_y are longitudinal and lateral accelerations of the vehicle, v_x, v_y are the longitudinal and lateral velocity of CG of vehicle and ψ, ω_z are the yaw angle and yaw rate, m_s the so-called sprung mass, h the center of gravity height. In (1), the vehicle roll acceleration can be expressed as follows

$$J_r \ddot{\alpha}_r = -b_x \dot{\alpha}_x - (k_x - m_s g h) \alpha_x - m_s h (\dot{v}_y + \omega_z v_x) \quad (2)$$

with J_r is the vehicle moment of inertia momentum about the longitudinal axis, b_x, k_x are the suspension roll damping and the suspension roll stiffness respectively.

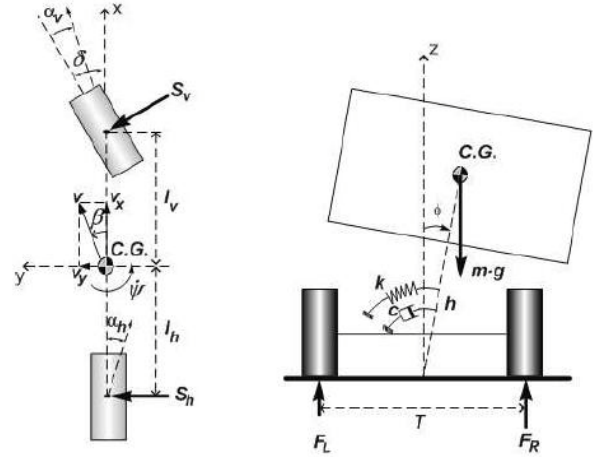


Figure. 1 Linear bicycle model with roll degree of freedom.

Hence, the system equations are

$$\begin{aligned} \dot{v}_x &= \omega_z v_y + \frac{1}{m} (a_x - m_s h \omega_z \omega_x) \\ \dot{v}_y &= -\omega_z v_x + k_j \left[a_y + \frac{m_s h}{J_r} (b_x \omega_x + (k_x - m_s g h) \alpha_x) \right] \\ \dot{\omega}_z &= \frac{1}{J_z} (l_f F_{y,f} - l_r F_{y,r}) \\ \dot{\alpha}_x &= \omega_x \\ \dot{\omega}_x &= -k_m \left[b_x \omega_x + (k_x - m_s g h) \alpha_x - \frac{m_s h}{m} a_y \right] \end{aligned} \quad (3)$$

with $k_m = \frac{m}{m J_r - m_s^2 h^2}$ and $k_j = \frac{J_r}{m J_r - m_s^2 h^2}$.

This model can be used to describe the vehicle under the following assumptions

- Pitch dynamic are neglected.
- The system is rigid.
- We considered the passive suspension.

The longitudinal and lateral accelerations have the following expressions

$$a_x = \mu (F_{x,f} + F_{x,r})$$

$$a_y = \mu (F_{y,f} + F_{y,r})$$

where μ is the tire-road friction coefficient and $F_{x,f}, F_{x,r}$ the front/rear longitudinal tyre forces, $F_{y,f}, F_{y,r}$ the front/rear lateral tyre forces, defined as follows [12]

$$F_{i,j} = D_{i,j} \sin \left[C_{i,j} \arctan \left(B_{i,j} \alpha_j \right) \right] \quad i = x, y; j = f, r$$

The constant $D_{i,j}, C_{i,j}, B_{i,j}$ can be determined experimentally and the slip angles tyre are

$$\alpha_f = \delta_d - \frac{v_y + l_f \omega_z}{v_x} \quad \alpha_r = -\frac{v_y - l_r \omega_z}{v_x}$$

where δ_d is the driver angle.

In the following, a_x, a_y, ω_z and v_x are supposed measurable, and the aim is to determine an observer to estimate the lateral velocity v_y and the position and velocity roll α_x, ω_x .

For, let we consider an observer of the following form

$$\dot{\hat{v}}_x = \omega_z \hat{v}_y + \frac{1}{m} (a_x - m_s h \omega_z \omega_x) + k_1 (v_x - \hat{v}_x)$$

$$\dot{\hat{v}}_y = -\omega_z v_x + k_j \left[a_y + \frac{m_s h}{J_r} (b_x \omega_x + (k_x - m_s g h) \alpha_x) \right] + k_2 (v_x - \hat{v}_x)$$

$$\dot{\hat{\alpha}}_x = \omega_x + k_3 (v_x - \hat{v}_x)$$

$$\dot{\hat{\omega}}_x = -k_m \left[b_x \omega_x + (k_x - m_s g h) \alpha_x - \frac{m_s h}{m} a_y \right] + k_4 (v_x - \hat{v}_x)$$

The dynamic of the estimation errors

$$e_x = v_x - \hat{v}_x$$

$$e_y = v_y - \hat{v}_y$$

$$e_\alpha = \alpha_x - \hat{\alpha}_x$$

$$e_\omega = \omega_x - \hat{\omega}_x$$

and its derivative

$$\begin{aligned} \dot{e}_x &= -k_1 e_x + \omega_z e_y - \frac{1}{m} m_s h \omega_z e_\omega \\ \dot{e}_y &= -k_1 e_x + k_s k_\alpha e_\alpha + k_s b_x e_\omega \end{aligned} \quad (5)$$

$$\dot{e}_\alpha = -k_3 e_x + e_\omega$$

$$\dot{e}_\omega = -k_4 e_x - k_m k_\alpha e_\alpha - k_m b_x e_\omega$$

where $\kappa_\alpha = k_x - m_s g h$ and the gains $k_i, i = 1, \dots, 4$ can be selected to render them asymptotically stable, under the following hypothesis.

Hypothesis 1: The yaw angular velocity ω_z is different from zero for all time instants $t \geq t_0$ with t_0 the initial time.

For the previous hypothesis $|\omega_z| \geq \omega_{\min} > 0$ for $t \geq t_0$. In fact, let us consider the following Lyapunov function candidate

$$\begin{aligned} V(t, e) &= \frac{1}{2} (\gamma e_x^2 + e_y^2) - \delta_1 \text{sign}(\omega_z) e_x e_y + \frac{1}{2} e_\alpha^2 \\ &+ \frac{1}{2} \gamma_1 e_\omega^2 - \delta_2 \omega_{z, \min} e_y e_\alpha \end{aligned} \quad (6)$$

with $\delta_1, \delta_2, \gamma, \gamma_1 > 0$ constants and the $\text{sign}(\omega_z)$ represent the signum function of the yaw rate.

Hence,

$$\begin{aligned} \dot{V}(t, e) &= \gamma e_x \left(-k_1 e_x + \omega_z e_y - \frac{1}{m} m_s h \omega_z e_\omega \right) \\ &+ e_y \left(-k_2 e_x + k_s k_\alpha e_\alpha + k_s b_x e_\omega \right) \\ &- \delta_1 \text{sign}(\omega_z) e_x \left(-k_2 e_x + k_s k_\alpha e_\alpha + k_s b_x e_\omega \right) \\ &- \delta_1 \text{sign}(\omega_z) e_y \left(-k_1 e_x + \omega_z e_y - \frac{1}{m} m_s h \omega_z e_\omega \right) \\ &- \delta_1 \delta_D(\omega_z) \dot{\omega}_z e_x e_y + \gamma_1 e_\omega \left(-k_4 e_x - k_m k_\alpha e_\alpha - k_m b_x e_\omega \right) \\ &- \delta_2 \omega_{z, \min} e_y \left(-k_3 + e_\omega \right) + e_\alpha \left(-k_3 + e_\omega \right) \\ &- \delta_1 \omega_{z, \min} e_\alpha \left(-k_2 e_x + k_s k_\alpha e_\alpha + k_s b_x e_\omega \right) \end{aligned}$$

with δ_D the Dirac distribution.

Under Hypothesis (1), $\delta_D(\omega_z) = 0$, so that

$$\begin{aligned} \dot{V}(t, e) = & -(\gamma k_1 - k_2 \delta_1 \text{sign}(\omega_z)) e_x^2 - \delta_1 |\omega_z| e_y^2 \\ & - k_s \kappa_\alpha \delta_2 \omega_{z, \min} e_\alpha^2 - \gamma_1 k_m b_x e_\omega^2 \\ & + (\gamma \omega_z - k_2 + k_1 \delta_1 \text{sign}(\omega_z) + k_3 \delta_2 \omega_{z, \min}) e_x e_y \\ & + (k_2 \delta_2 \omega_{z, \min} - k_s \kappa_\alpha \delta_1 \text{sign}(\omega_z) - k_3) e_x e_\alpha \\ & + \left(-\frac{\gamma}{m} m_s h \omega_z \right) - k_s b_x \delta_1 \text{sign}(\omega_z) - \gamma_1 k_4 e_x e_\omega \\ & + k_s \kappa_\alpha e_y e_\alpha + (1 - \gamma_1 k_m \kappa_\alpha - k_s b_x \delta_2 \omega_{z, \min}) e_\alpha e_\omega \\ & + \left(k_s b_x + \frac{1}{m} \delta_1 m_s h |\omega_z| - \delta_2 \omega_{z, \min} \right) e_y e_\omega \end{aligned} \quad (7)$$

with $\omega_{z, \min} = \min_{t \geq t_0} \omega_z$, and

$$\begin{aligned} \gamma k_1 - \delta_1 \text{sign}(\omega_z) k_2 &= \delta_1 |\omega_z| \\ \gamma \omega_z - k_2 + k_1 \delta_1 \text{sign}(\omega_z) + k_3 \delta_2 \omega_{z, \min} &= 0 \\ k_2 \delta_2 \omega_{z, \min} - k_s \kappa_\alpha \delta_1 \text{sign}(\omega_z) - k_3 &= 0 \\ -\frac{\gamma}{m} m_s h \omega_z - k_s b_x \delta_1 \text{sign}(\omega_z) - \gamma_1 k_4 &= 0 \end{aligned} \quad (8)$$

Considering the following inequalities

$$\begin{aligned} \alpha_1 e_y e_\alpha &\leq \alpha_1 \frac{\sqrt{|\omega_z|}}{\omega_{z, \min}} e_y^2 + \alpha_1 \frac{\omega_{z, \min}}{\sqrt{|\omega_z|}} e_\alpha^2 \\ \alpha_2 e_y e_\omega &\leq \alpha_2 \frac{1}{\lambda_2} e_y^2 + \alpha_2 \lambda_2 e_\omega^2 \\ \alpha_3 e_\omega e_\alpha &\leq \alpha_3 \frac{1}{\lambda_3} e_\alpha^2 + \alpha_3 \lambda_3 e_\omega^2 \end{aligned}$$

where $|\omega_z| \geq \omega_{z, \min} > 0$ and from the Hypothesis (1) and $\lambda_2, \lambda_3 > 0$, one gets

$$\begin{aligned} \dot{V}(t, e) \leq & -\delta_1 \omega_z e_x^2 - \left(\delta_1 |\omega_z| - \alpha_1 \frac{\sqrt{|\omega_z|}}{\omega_{z, \min}} - \frac{1}{\lambda_2} \alpha_2 \right) e_y^2 \\ & - \left(k_s \kappa_\alpha \delta_2 \omega_{z, \min} - \alpha_1 \frac{\omega_{z, \min}}{\sqrt{|\omega_z|}} - \frac{1}{\lambda_3} \alpha_3 \right) e_\alpha^2 \\ & - (\gamma_1 k_m b_x - \lambda_2 \alpha_2 - \lambda_3 \alpha_3) e_\omega^2 \end{aligned}$$

Choosing

$$\begin{aligned} \delta_1 &> \frac{m \lambda_2 - m_s h}{m \lambda_2} \left[\frac{1}{\omega_{z, \min}} \left(\alpha_1 + \frac{\delta_2 \omega_{z, \min}}{\alpha_2} - \frac{1}{\lambda_2} k_s b_x \right) \right] > 0 \\ \delta_2 &> \frac{\lambda_2}{\lambda_3 k_s \kappa_\alpha + k_s b_x} \left[\frac{1}{\omega_{z, \min}} \left(\alpha_1 + \frac{1}{\lambda_3} (1 - \lambda_1 k_m \kappa_\alpha) \right) \right] > 0 \end{aligned}$$

from (8)

$$\begin{aligned} k_1 &= \frac{1}{\gamma - \gamma \delta_2^2 \omega_{z, \min} - \delta_1} \left[\delta_1 |\omega_z| + \gamma \delta_1 |\omega_z| \right. \\ &\quad \left. - \delta_1 \delta_2^2 \omega_{z, \min}^2 |\omega_z| - k_s \kappa_\alpha \delta_1^2 \delta_2 \omega_{z, \min} \right] \\ k_2 &= \frac{1}{\gamma - \gamma \delta_2^2 \omega_{z, \min} - \delta_1} \left[\omega_z (\gamma^2 + \delta_1^2) \right. \\ &\quad \left. - \gamma k_s \kappa_\alpha \delta_1 \delta_2 \omega_{z, \min} \text{sign}(\omega_z) \right] \\ k_3 &= \frac{1}{\gamma - \gamma \delta_2^2 \omega_{z, \min} - \delta_1} \left[\gamma^2 \delta_2 \omega_{z, \min} \omega_z \right. \\ &\quad \left. + \delta_1^2 \delta_2 \omega_{z, \min} \omega_z - k_s \kappa_\alpha \delta_1 \text{sign}(\omega_z) (\gamma^2 - \delta_1^2) \right] \\ k_4 &= \frac{1}{\gamma_1} \left[-\frac{\gamma}{m} m_s h \omega_z - k_s b_x \delta_1 \text{sign}(\omega_z) \right] \end{aligned}$$

where $\gamma > \frac{\delta_1^2}{1 - \delta_2^2 \omega_{z, \min}^2}$.

Therefore, one finally obtains

$$\begin{aligned} \dot{V}(t, e) \leq & -\delta_1 |\omega_z| e_x^2 - \lambda_{\delta, 1} |\omega_z| e_y^2 \\ & - \lambda_{\delta, 2} \omega_{z, \min} e_\alpha^2 - \lambda_{\delta, 3} e_\omega^2 \end{aligned} \quad (9)$$

with $\lambda_{\delta, 1} > 0, \lambda_{\delta, 2} > 0$ and $\lambda_{\delta, 3} = (1 - \gamma_1 k_m \kappa_\alpha) > 0$ then the equation (9) is negative definite, hence the error system (5) is exponentially stable.

III. SIMULATION RESULTS

This section presents the behavior of the nonlinear observer for the lateral velocity and roll position and velocity. The simulation consists of a single lane maneuver with the longitudinal velocity at 100 km/h Figure (2). The results obtained by the kinematics model are shown in the Figure (3,4,5). In the Figure (3a) we shown the effectiveness of the nonlinear observer for the lateral velocity

because dynamic observer tracks the state v_y ; in the Figure (3b) we present the yaw rate ω_z used in this simulation. In the Figure (4) is shown the roll position α_x and velocity ω_x and its observer $\hat{\alpha}_x, \hat{\omega}_x$. This observer is obtained only with information based on measurements of the longitudinal and lateral acceleration, longitudinal velocity, yaw rate and the steer angle. Finally, in the Figure (5) we present the errors where we can see the error is identically to zero.

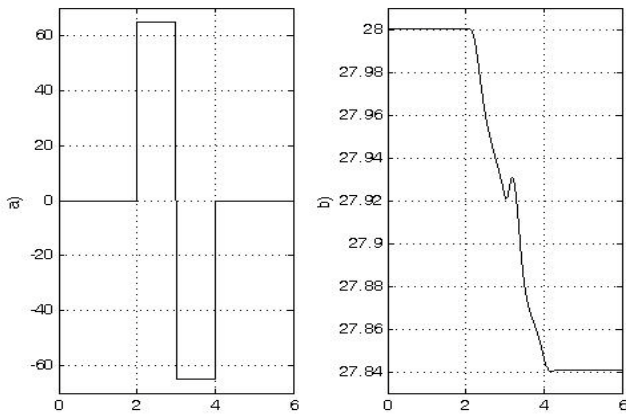


Figure 2. a) Single lane change maneuver δ_d [deg/s]; b) Longitudinal velocity v_x (solid) and its observer \hat{v}_x (dotted) [m/s].

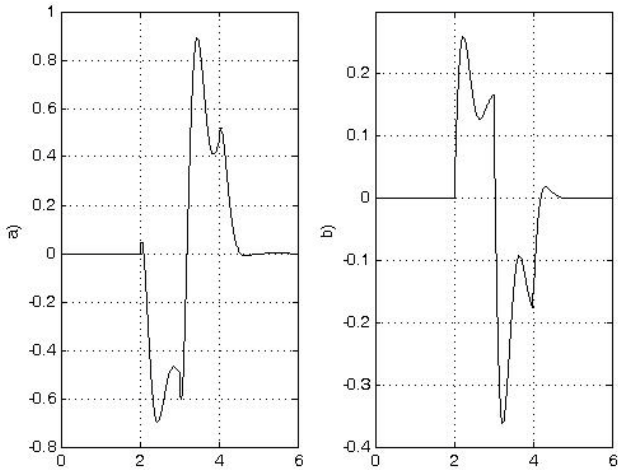


Figure 3. a) Lateral velocity v_y (solid) and its observer \hat{v}_y (dotted)[m/s]; b) Yaw rate ω_z [rad/s].

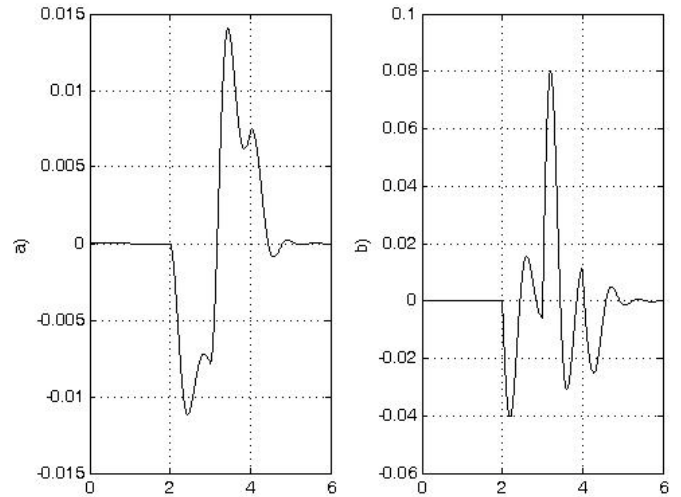


Figure 4. a) Position Roll α_x (solid) and its observer $\hat{\alpha}_x$ (dotted) [m]; b) Velocity roll ω_x (solid) and its observer $\hat{\omega}_x$ (dotted) [m].

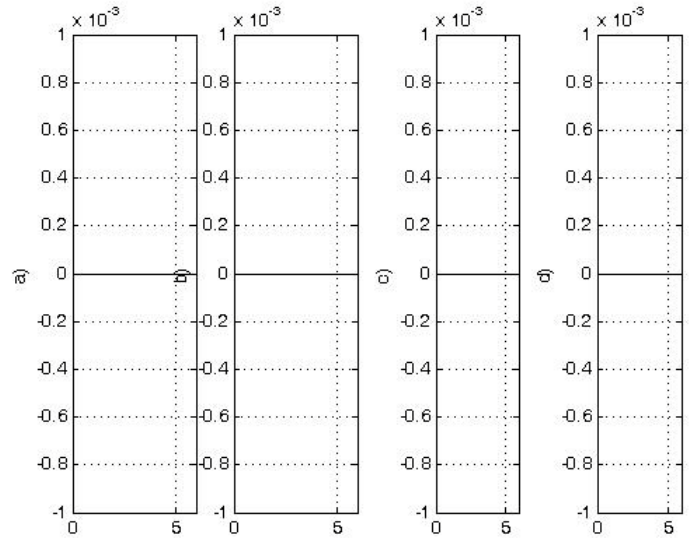


Figure 5. Errors a) $e_x = v_x - \hat{v}_x$; b) $e_y = v_y - \hat{v}_y$; c) $e_\alpha = \alpha_x - \hat{\alpha}_x$ and d) $e_\omega = \omega_x - \hat{\omega}_x$.

IV. CONCLUSIONS

This paper proposes the nonlinear observer for the lateral velocity using the dynamics of roll position and velocity. This observer is based in the available information for the common sensor in the vehicle, i.e. the longitudinal and lateral acceleration a_x, a_y , longitudinal velocity v_x , yaw rate ω_z and the steer angle δ_d .

Therefore, the observer is not sensitive to vehicle parameter changes. The mathematical model used in this article is the bicycle model together the dynamics of passive suspension of utility vehicle. Finally, we demonstrated the theory of a nonlinear observer and its effectiveness with numerical simulations where the error between the system and the observer states tends exponentially to zero.

V. REFERENCIAS

- [1] C. Acosta-Lua, B. Castillo Toledo and S. Di Gennaro “*Output Feedback Regulation of Electromagnetic Valves for Camless Engines*” Proceedings of the 2007 American Control Conference, New York City, USA, July 11-13, 2007
- [2] M. C. Best, T. J. Gordon and P.J. Dixon, “*An extended adaptive Kalman filter for real-time state estimation of vehicle handling dynamics*”, Vehicle System Dynamics, Vol 34, pp. 5775. 2000.
- [3] G. Baffet, J. Sthphant and A. Charara, “*Lateral vehicle-dynamic observers: simulations and experiments*”, International Journal of Vehicle Autonomous Systems, vol. 5, num. 3-4, pp. 184-203, December, 2007.
- [4] J. Farrelly and P. Wellstead, “*Estimation of vehicle lateral velocity*”, Proceedings of the 1996 IEEE international conference on control applications pp. 552557, 1996.
- [5] Y. Fukada, “*Slip-angle estimation for stability control*”, Vehicle Systems Dynamics, vol 32, pp. 375388. 1999.
- [6] A. Hac and M. D. Simpson, “*Estimation of vehicle side slip angle and yaw rate*”, In SAE 2000 World Congress, Detroit, MI, USA., 2000.
- [7] M. Hiemer, A. VonVietinghoff, U. Kiencke and T. Matsunaga, “*Determination of vehicle body slip angle with non-linear observer strategies*”, In Proceedings of the SAEWorld Congress, Paper no. 2005-01-0400, 2005.
- [8] G. J. Heydinger, W. R. Garrott, J. P. Chrstos, and D. A. Guenther, “*A methodology for validating vehicle dynamics simulations*”, SAE, Paper 900128, 1990.
- [9] U. Kiencke and A. Daiss, “*Observation of lateral vehicle dynamics*”, Control Engineering Practice, Vol 5(8), pp. 11451150, 1997.
- [10] J. Lu and T.A. Brown, “*Vehicle side slip angle estimation using dynamic blending and considering vehicle attitude information*”, Patent US 6671595, 2003.
- [11] U. Kiencke and L. Nielsen, “*Automotive control systems*”, Berlin: Springer, 2000.
- [12] H. B. Pacejka, “*Tyre and vehicle dynamics*”, London: Butterworth-Heinemann
- [13] L. R. Ray, “*Nonlinear state and tire force estimation for advanced vehicle control*”, IEEE Transactions on Control Systems Technology, Vol. 3(1), pp. 117124, 1995
- [14] J. Stéphane, A. Charara and D. Meizel, “*Evaluation of sliding mode observer for vehicle sideslip angle*”, Control Engineering Practice, vol. 15, pp. 803-812, 2007.
- [15] P. Setlur, J. R. Wagner, D. M. Dawson and D. Braganza, “*A Trajectory Tracking SteerbyWire Control System for Ground Vehicles*”, IEEE Transactions on Vehicular Technology, Vol. 55, No. 1, pp. 7685, 2006.
- [16] A. Suissa, Z. Zomotor and F. Bttiger, “*Method for determining variables characterizing vehicle handling*”. Patent US 5557520, 1996.
- [17] A.Y. Ungoren, H. Peng and H. Tseng, “*A study on lateral speed estimation methods*”, International Journal on Vehicle Autonomous Systems, Vol 2, pp. 126144. 2004.
- [18] J. Wong, “*Theory of Ground Vehicles.*” New York: Wiley, 1978.

- [19] A. T. Van Zanten, “*Bosch ESP system: 5 years of experience*”, In Proceedings of the automotive dynamics and stability conference, pp. 354, Paper no. 2000-01-1633, 2000.
- [20] P. J. Venhovens, and K. Naab, “*Vehicle dynamics estimation using Kalman filters*”, Vehicle System Dynamics, Vol. 32, pp. 171-184. 1999.

VI. APPENDIX

We considered simulations based on data from a prototype vehicle obtained of the Ford Motor Company. The nominal parameters are

$$m = 1880 \text{ kg.}$$

$$J_z = 830 \text{ kg m}^2$$

$$lr = 1.53 \text{ m}$$

$$lf = 1.38 \text{ m}$$

$$l = f + lr \text{ m}$$

$$\mu =$$

$$J_r = 71 \text{ kg m}^2$$

$$m_s = 9 \text{ kg}$$

$$g = .81 \text{ m/s}^2$$

$$h = 0.3 \text{ m}$$

$$b_x = 1000 \text{ N m rad/s}$$

$$k_x = 16000 \text{ N m/rad}$$

$$B_{f,y} = 7.2$$

$$C_{f,y} = 1.81$$

$$D_{f,y} = 8854$$

$$B_{r,y} = 11$$

$$C_{r,y} = 1.68$$

$$D_{r,y} = 8394$$

VII. AUTHORS

Cuahtemoc Acosta-Lua was born in Lázaro Cárdenas, Michoacán, México, in 1977. He received the B.Sc. degree in electronic engineering from the Morelia Technologic Institute, Morelia, Mexico in 1996 and the M.Sc. and Ph.D. degrees in electrical engineering from the Advanced

Studies and Research Center, National Polytechnic Institute, Guadalajara, México, in 2003 and 2007, respectively. Since April 2008, he has been with the Department of Electrical and Information Engineering, University of L'Aquila, L'Aquila, Italy, where he was External Consulter in a project between the Center of Excellence DEWS and Ford Forschungszentrum Aachen. His research interest focus on regulator theory, hybrid systems, regulation theory, applications of nonlinear control, automotive control systems.

Bernardino Castillo-Toledo received the B.Sc. degree in electrical engineering from the National Polytechnic Institute (IPN), Mexico City, Mexico, the M.Sc. degree from the Research and Advanced Studies Center of the National Polytechnic Institute (CINVESTAV-IPN), Guadalajara, Mexico, and the Ph.D. degree from the University of Rome La Sapienza, Rome, Italy, in 1981, 1985, and 1992, respectively. He was a Lecturer with the School of Electrical and Mechanics Engineering, IPN, from 1985 to 1989. From 1985 to 1995, he was with the Automatic Control Section, Department of Electrical Engineering, CINVESTAV-IPN, and, since 1995, with CINVESTAV-IPN, Guadalajara Campus. He held several research stages with the University of Rome La Sapienza and University of L'Aquila, L'Aquila, Italy, and was a visiting Professor at the Laboratoire d'Automatique et d'Analyse des Systemes, French Council for Scienti_c Research, during 1999 to 2000, and at the University of Compiègne, Compiègne, France, during the first semester of 2002. His main research interests include nonlinear control design, the robust regulation problem, and the application of artificial neural networks and fuzzy logic techniques to control and fault diagnosis of dynamical systems.

Stefano Di Gennaro received the M.S. degree (summa cum laude) in nuclear engineering and the Ph.D. degree in system engineering from the University of Rome La Sapienza, Rome, Italy, in 1987 and 1992, respectively. Since October 1990, he has been with the Department of Electrical and Information Engineering, University of L'Aquila, L'Aquila, Italy, where he was an Assistant

Professor of Automatic Control and, since 2001, has been an Associate Professor of Automatic Control. He holds courses on automatic and nonlinear controls. He is also with the Center of Excellence DEWS Design methodologies of Embedded controllers, Wireless interconnect and Systems on- chip, University of L'Aquila. In 1986, he was a Visiting Scientist at the Nuclear Research Center ENEA, Casaccia. He has been a Visiting Professor at the Laboratoire des Signaux et Systmes, Centre National de la Recherche Scientifique (CNRS), Paris, at the Department of Electrical Engineering, Princeton University, Princeton, NJ, at the Department of Electrical Engineering and Computer Science, University of California, Berkeley, and at the Advanced Studies and Research Center, National Polytechnic Institute, Unidad Ciudad de México and Unidad Guadalajara, México. He is working in the area of hybrid systems, regulation theory, and applications of nonlinear control.

GEOBUS: SISTEMA DE ORIENTACIÓN ESPACIAL PARA ASISTIR A PERSONAS INVIDENTES UTILIZANDO TELÉFONOS CELULARES

Jaramillo Calderón Andrés, Mata Rivera Félix Miguel

Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas, IPN. Avenida Instituto Politécnico Nacional No. 2580, Colonia Barrio la Laguna Ticomán, Delegación Gustavo A. Madero, C.P. 07340 México D.F. Correo Electrónico: andresjcal@gmail.com, migfel@sagitario.cic.ipn.mx

RESUMEN

GeoBus es una aplicación móvil para teléfonos celulares para ayudar a orientar espacialmente a los pasajeros y usuarios del sistema de transporte de la Ciudad de México, "Metrobus". El sistema está enfocado para ayudar a los pasajeros invidentes o con visión limitada. La aplicación móvil (instalada en un teléfono celular), funciona utilizando menús audibles. GeoBus indica el nombre de una estación del Metrobus y la ubicación de las puertas de embarque (a la derecha, izquierda, atrás, en frente). El sistema interactúa con un dispositivo GPS y una brújula digital. El GPS proporciona la localización geográfica del usuario. La brújula digital indica la dirección para encontrar un lugar determinado (por ejemplo, la salida de la estación de Metrobus). La comunicación entre estos dispositivos se logra con la tecnología Bluetooth. El software de GeoBus se ejecuta utilizando el lenguaje Java (plataforma J2ME), y Java API para ubicación. Las pruebas se realizaron en un teléfono celular Nokia N73 por usuarios invidentes.

I. INTRODUCCIÓN

El MetroBus (Indios Verdes – Dr. Gálvez) del Distrito Federal es usado por 260 mil personas en promedio en día hábil. Si bien está diseñado incluso para personas con capacidades especiales, no está totalmente condicionado para que una persona invidente se pueda trasladar con toda seguridad y confianza. En pocas estaciones del MetroBus existen placas con información en

Braile y guías para bastón. Información tipo guía espacial es vital para una persona invidente, la cual, solo la pueden obtener por otra persona. Esta dependencia hacia otras personas hace que la persona invidente prefiera no salir a lugares que no conozca quedándose limitada a desenvolverse en ciertas zonas. Impulsar su independencia en su vida cotidiana ayuda a este sector de la población a ampliar sus horizontes y desenvolverse dentro de la sociedad en que vive. Atendiendo a la necesidad de estas personas de tener acceso a la tecnología con la que contamos, me veo motivado a desarrollar un sistema para teléfono celular que ayude a una persona invidente a orientarse para que pueda hacer uso del sistema de transporte MetroBus con más tranquilidad y seguridad.

El uso de teléfonos celulares en nuestros días es ya prácticamente una necesidad. Para la población invidente el uso de estos también se ha difundido en consideración. La implementación en el teléfono celular de un software conversor de texto a voz que genera idéntico sonido al producido por una persona al leer un texto en voz alta hace posible que una persona invidente pueda manejar un celular. Los servicios que un celular con este tipo de software ofrece son decir quién está llamando, las llamadas perdidas y realizadas, o enviar y recibir mensajes de texto, en sí acceso a recursos del celular.

Por otro lado, el auge que se ha visto en los últimos años de los dispositivos GPS para el uso diario ha impulsado al desarrollo de aplicaciones para celular nunca antes vistas. Uno de ellos sería el uso de la información satelital para que una persona invidente pueda, mediante

audio proporcionado por su teléfono celular, ubicarse dentro del sistema de transporte colectivo MetroBus (Indios Verdes - Dr. Gálvez) del Distrito Federal a través de la implementación de una brújula digital y un GPS en el teléfono celular. Así, una vez que se sabe en qué estación se encuentra la persona, se pueda saber hacia dónde está la salida o hacia dónde debe caminar para abordar el MetroBus que va por ejemplo en dirección Doctor Gálvez.

Las posibilidades de una persona con discapacidad visual para poder hacer uso de los sistemas de comunicaciones existentes aumentarían gradualmente, ya que este sistema se podría extender a otros ambientes como son MetroBus Eje 4 Sur, el sistema de transporte colectivo Metro (aplicando triangulación con celdas celulares), autobuses, central de camiones, parques, etc. Por otro lado, el mercado de telecomunicaciones crecería abarcando un sector de la población un tanto marginado por las empresas que brindan tales servicios. Además las posibilidades de una persona con discapacidad visual se ensancharían ayudándoles a tener una vida cotidiana más independiente y segura.

II. ANTECEDENTES

Hasta este momento, los avances en los teléfonos celulares se han enfocado en aplicaciones para uso de la mayoría de la población, dejando a un lado las funcionalidades para usuarios con alguna capacidad diferente haciendo que las aplicaciones para celular existentes en nuestro país para personas con discapacidades visuales sean de precios elevados y difíciles de conseguir.

Por otra parte se han realizado estudios que tienen como finalidad brindarle a una persona invidente la oportunidad de interactuar con las tecnologías que hoy en día se encuentran en casa. Analizando situaciones donde las áreas de trabajo son vistas como un campo de trabajo en el diseño de artefactos para invidentes. GeoBus es también

una aplicación que promueve al usuario invidente a interactuar con tecnología presente hoy en día como son GPS, teléfonos celulares y la electrónica.

Otro tipo de sistemas existentes son para asistir tecnológicamente a una persona invidente a orientarse dentro de lugares cerrados. Esto con un sensor usado como lámpar para buscar objetivos en tercera dimensión. Apretando teclas se consultan características del objeto como son posición, orientación y navegación enviada a una computadora portátil o a un servidor. Finalmente, estas consultas son contestadas mediante un conversor de Texto a Voz. El sistema GeoBus sigue esta línea al servir como guía dentro del sistema de transporte colectivo MetroBus.

III. METODOLOGÍA

El objetivo principal de este trabajo consiste en orientar a una persona invidente para que pueda caminar con cierto grado de tranquilidad dentro del sistema de transporte MetroBus necesitando de información satelital proporcionada por un GPS. Aprovechando el hecho de que el sistema de transporte colectivo MetroBus puede recibir la señal GPS sin problemas, se diseñó un sistema de guía basado en GPS que ubica en qué estación del MetroBus se encuentra la persona. Registrando el grupo de coordenadas GPS que identifica a una estación en particular, como se puede ver en la Figura 1.



Figura 1. Coordenadas GPS de una estación del MetroBus.

Se generó una serie de archivos planos en el teléfono celular que relacionen el nombre de una estación con un grupo de coordenadas, por ejemplo: “Estación Nuevo León” -> {19°24’06.02’’N 99°10’11.57’’O}. Ahora bien, tomando como referencia el centro de la estación y mirando hacia las puertas en dirección Indios Verdes, se registró el valor de la brújula digital para así poder orientar a la estación respecto a los puntos cardinales, como se ve en la Figura 2:

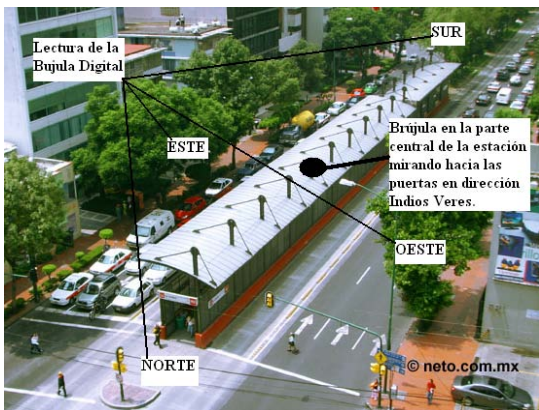


Figura 2. Orientación cardinal de una estación del MetroBus.

Generando un registro de archivos planos con la siguiente estructura: “Estación Nuevo León” -> {Puertas en dirección Indios Verdes: Este, Puertas en dirección Dr. Gálvez: Oeste, Salida a la calle de Nogales: Sur, Salida a la calle de Hermosillo: Norte}

Con la estación registrada geoespacialmente en archivos planos del teléfono celular, se procedió a diseñar una interfaz mediante la cual el teléfono celular pueda obtener información del GPS de forma inalámbrica y así ubicar la estación en la que está el usuario. Esta interfaz está diseñada con menús audibles para el uso total de la persona invidente. El usuario navegará por la aplicación escuchando audio y presionando botones claves.

Hasta este punto, el usuario tendrá conocimiento de la estación en la que se encuentra. Lo que procede es responder a cuestiones como:

- ¿Hacia dónde está la salida?
- ¿Hacia dónde están las puertas en dirección Indios Verdes?
- ¿Hacia dónde están las puertas en dirección Doctor Gálvez?

La brújula digital, mediante comunicación Bluetooth (incorporada en una PC), brinda la información necesaria para responder estas cuestiones. Una vez que el celular ha identificado la estación en la que se encuentra, el usuario debe elegir una de las siguientes opciones del menú audible:

- a) Saber en qué dirección están las salidas
- b) Saber en qué dirección están las puertas en dirección Indios Verdes
- c) Saber en qué dirección están las puertas en dirección Dr. Gálvez

Seleccionada la petición, el celular solicitará la información a la brújula. Por ejemplo si el usuario tecleó la opción “a” y la brújula nos devuelve Oeste y retomando la Figura 2 donde la salida esta al Norte, el sistema le informa auditivamente al usuario que la salida está a su derecha. A continuación se muestra el marco de trabajo del sistema en la Figura 3.



Figura 3. Marco de trabajo del sistema GEOBUS.

Entonces, de acuerdo a este esquema de trabajo, se definen todas las funcionalidades que GeoBus necesitará, esto, consultando a un grupo de personas invidentes que asisten a la Biblioteca Central Ricardo Garibay del estado de Hidalgo, con el fin de conjuntar consejos de personas invidentes y así atender todas las necesidades.

El siguiente paso consiste en analizar los dispositivos GPS y brújula digital de acuerdo a los siguientes requerimientos: tamaño reducido, ligero, comunicación Bluetooth y precio reducido. El caso del teléfono celular requiere que tenga comunicación Bluetooth. Posteriormente, se estudia la configuración de las estaciones del MetroBus y se diseña un esquema de orientación hacia un punto cardinal de cada una de ellas. Considerando los siguientes parámetros:

- Nombres y posición GPS de cada estación del MetroBus
- Orientación hacia un punto cardinal de las salidas y puertas para abordar de cada estación

La posición GPS de la salida y puertas estará referenciada hacia un punto cardinal para así poder compararla con la posición del usuario cuando se realicen peticiones. Esta información, se almacena en el teléfono celular.

Se desarrolló un algoritmo capaz de identificar una estación respecto a un rango de coordenadas GPS para identificar en cual estación se encuentra el usuario; así como también para orientar a la persona hacia la salida o puertas para abordar, con mensajes de “a tu derecha”, “a tu izquierda”, etc. Con base en la información adquirida de la brújula digital.

IV. RESULTADOS Y PRUEBAS

Registro de archivos planos de identificación de estaciones para el teléfono celular

GeoBus necesita de información geoespacial previamente almacenada en el teléfono celular para poder identificar las estaciones del MetroBus en tiempo real. Se hizo la lectura de la latitud y la longitud de cada estación del MetroBus. Esta lectura se hizo en la parte central de cada estación con el fin de tomar éste como punto de referencia para ubicar la estación, con un rango de aproximadamente 35 metros de radio. Los valores de la latitud y la longitud se suman para obtener un solo valor y así identificar a cada estación por este valor, esto es, para la estación del MetroBus Indios Verdes la latitud es 19,49666 y la longitud es 99,11972, realizando la suma obtenemos: 1186164. En la tabla 1 se pueden ver las sumas de ocho estaciones.

	Estación del MetroBus	Latitud	Longitud	Suma
11	San Simón	19,45968	99,14638	1186061
12	Manuel González	19,45676	99,14944	1186062
13	Buenavista	19,44698	99,1531	1186001
14	El Chopo	19,44334	99,15437	1185977
15	Revolución	19,44038	99,1554	1185958
16	Tabacalera	19,43609	99,15737	1185935
17	Reforma	19,43299	99,15867	1185917
18	Hamburgo	19,42780	99,1611	1185889

Tabla 1. Tabla con las sumas de la latitud y longitud para 8 estaciones del MetroBus

Dado que se implementará una búsqueda binaria cada vez que desee encontrar una estación en el archivo plano, a las sumas se les dieron un formato especial. Este formato especial coloca cada suma en una línea de caracteres dentro del archivo plano, anteponiendo un número entero a cada suma para identificar cada una de ellas con la búsqueda binaria. Por ejemplo, para la estación Tabacalera: **161185935**.

Registro de archivos planos de orientación espacial para el teléfono celular

Se necesita de información geoespacial previamente almacenada en el teléfono celular para poder orientar las estaciones del MetroBus respecto a los puntos cardinales en tiempo real. Esta información juega un papel importante en el funcionamiento del sistema ya que para poder orientar a una persona invidente de a qué dirección se encuentra. Por ejemplo, la salida de la estación, se necesita saber primeramente hacia qué punto cardinal está orientada la salida de dicha estación, hacia qué punto cardinal está mirando el usuario y a qué altura de la estación se encuentra el usuario (para ubicar cada una de las puertas).

En la parte central de cada estación se realizó la lectura cardinal con una brújula magnética (la cual está calibrada con la brújula digital a usar) de la salida o salidas, las puertas para abordar en dirección Indios Verdes y las puertas para abordar en dirección Dr. Gálvez. Esta orientación cardinal será la medición base. Respecto a la distancia a la que se encuentre el usuario de la parte central de la estación (medición base) la orientación de las puertas cambiará.

En la tabla 2 se muestran los resultados obtenidos de las lecturas base para las primeras trece estaciones.

Numero	Estación	Puerta I.V.	Puerta Dr. Gálvez	Salida
1	Indios Verdes	-----	E	S
2	18 de Marzo (Dr. Gálvez)	-----	NW,W	N
3	18 de Marzo (Indios Verdes)	SE,E	-----	N
4	Euskaro (Dr. Gálvez)	-----	NW,W	S
5	Euskaro (Indios Verdes)	SE,E	-----	S
6	Potrero (Dr. Gálvez)	-----	NW,W	S
7	Potrero (Indios Verdes)	SE,E	-----	S
8	La Raza	-----	-----	-----
9	Circuito(Dr. Gálvez)	-----	NW,W	N
10	Circuito(Indios Verdes)	SE,E	-----	NE
11	San Simón	SE,E	NW,W	S
12	Manuel González	SE,E	NW,W	NE
13	Buenavista	SE,E	NW,W	S

Tabla 2. Mediciones de orientación cardinal para 15 estaciones del MetroBus

Donde: N= Norte, NE=Noreste; E=Este; SE=Sureste; S=Sur; SW=Suroeste; W=Oeste; NW=Noroeste;--- =No existe; ND=Estación no Disponible.

GeoBus procesará los datos de forma numérica es por ello que se realiza la una transformación que se muestra en la tabla 3.

Numero	Estación	Puerta I.V.	Puerta Dr. Gálvez	Salida
1	Indios Verdes	-----	4	8
2	18 de Marzo (Dr. Gálvez)	-----	14,12	16
3	18 de Marzo (Indios Verdes)	4,6	-----	16
4	Euskaro (Dr. Gálvez)	-----	14,12	8
5	Euskaro (Indios Verdes)	4,6	-----	8
6	Potrero (Dr. Gálvez)	-----	14,12	8
7	Potrero (Indios Verdes)	4,6	-----	8
8	La Raza	ND	ND	ND
9	Circuito(Dr. Gálvez)	-----	14,12	16
10	Circuito(Indios Verdes)	4,6	-----	2
11	San Simón	4,6	14,12	8
12	Manuel González	4,6	14,12	2
13	Buenavista	4,6	14,12	8

Tabla 3. Tabla de las medidas de orientación cardinal.

A los valores de orientación cardinal se le dio un formato parecido al de las estaciones. A cada estación le corresponde un valor constituido por el número de estación, por ejemplo San Simón “11”, seguido por los puntos cardinales correspondientes a las puertas para abordar en dirección Indios Verdes (pueden ser hasta 3, en caso de que existan menos su orientación corresponde a un 0) “040600”, seguido por los puntos cardinales correspondientes a las puertas para abordar en dirección Dr. Gálvez “141200” y al final los puntos correspondientes a las salidas (en caso de que no existan más de una se llena con 33) “0833”, estos valores se concatenan:110406001412000833. El final se rellena con dos ceros para que nos dé un valor de longitud de 20 dígitos: 11040600141200083300.

Creación del menú audible de GeoBus

Dado que el sistema está enfocado a usuarios invidentes, la interfaz debe ser de tipo auditiva. Se programó el menú de GEOBUS en código J2ME a bajo nivel. Cada menú de

GEOBUS está referido por un número, esto es: menú 1, menú 2, menú 3, menú 4, menú 5. En cada selección de menú se escuchará un audio para identificar el menú y en cada salto de opción (TECLA #2 ó TECLA #8) se escuchará un audio para identificar la opción a la que se pasó. Con estos audios el usuario podrá manejar las funcionalidades de GEOBUS.

El resultado obtenido es que GEOBUS cuenta con 5 menús: Menú Principal, Menú Estación, Menú Nombre, Menú Dirección y SubMenú Dirección. Cada uno identificado con un número: Menú Principal = 0, Menú Estación = 1, Menú Nombre = 2, Menú Dirección = 3 y SubMenú Dirección = 4. Las opciones de cada menú son las siguientes:

- **Menú Principal:** Estación y Salir del Sistema
- **Menú Estación:** Nombre de la Estación, Ubicar Dirección, Ubicar Salida y Regresar al menú anterior
- **Menú Nombre:** Repetir Audio y Regresar al menú anterior
- **Menú Dirección:** Indios Verdes, Dr. Gálvez y Regresar al menú anterior.
- **SubMenú Dirección:** Repetir Audio, Actualiza Audio y Regresar al menú anterior

Comunicación del teléfono celular con el GPS

El teléfono en el que se instaló la aplicación fue el Nokia N73 debido a que éste teléfono cuenta con Symbian S60, CLCD 1.0, MIDP 2.0 y Bluetooth 1.2 así como su precio en el mercado. Se tiene contemplado instalar GEOBUS en celulares que cuenten con GPS integrado, la principal limitante en este momento son los precios de estos celulares.

Se realizó la programación en lenguaje J2ME de la comunicación entre el teléfono celular Nokia N73 y el GPS G – Trender. La especificación *JSR 179 Location API* para J2ME pone a disposición una serie de clases que

permiten acceder desde java a información relacionada con la posición y el movimiento del GPS. La *Location API* está preparada para que el dispositivo tenga acceso a todo tipo de *LocationProviders*, que es como se denominan las fuentes de datos, en este caso el GPS G – Trender.

Comunicación del teléfono celular con la Brújula Digital

En el diseño y análisis se decidió usar la brújula digital incluida en el GPS G – Trender. Realizando las pruebas con el GPS se descubrió que las tramas NMEA que contienen a los datos cardinales del GPS no pueden ser recuperados vía Bluetooth o USB. Se comunicó con la empresa Wintec y se corroboró que dichos datos no pueden ser extraídos del dispositivo. Dado que este problema es crítico se procedió a buscar una solución buscando otro tipo de brújulas digitales que puedan satisfacer nuestras necesidades. La brújula digital que se decidió usar fue la CMPS03 de la marca Sparkfun, con las siguientes características: 5 Volts de alimentación, 0.1 grados de resolución y datos de salida I2C.. Los datos I2C de la brújula digital son transformados a USB mediante el convertidor I2C – USB: Obteniendo la siguiente conexión que se muestra en la Figura 4.

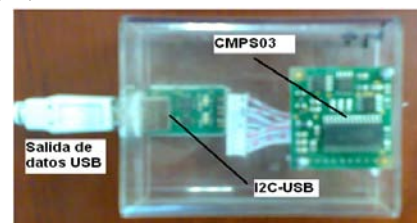


Figura 4. Conexión real del circuito CMPS03 con el conversor I2C - USB

Teniendo los datos con salida USB, es necesario hacérselos llegar al teléfono celular. Una alternativa es usar un dispositivo bluetooth que realice esta tarea pero debido a los costos que esto implicaba se tuvo que usar un USB – Host que se encargue de manejar los datos de la brújula digital. Hasta este momento los dispositivos que funcionan como USB – Host (On The Go) son

dispositivos que funcionan para transferir información entre memorias flash, discos duros o reproductores de MP3, por lo que el dispositivo USB-Host a que se utilizó fue una PC.

Hasta este punto se tuvo una brújula digital conectada a una computadora por medio de USB. El siguiente paso fue hacerle llegar estos datos al celular del usuario. Para esto, se programó una aplicación en JAVA que sea capaz de recuperar los datos de la brújula digital y enviarlos por Bluetooth al celular. La siguiente figura muestra el esquema de funcionamiento de la conexión de la brújula digital con el teléfono celular.

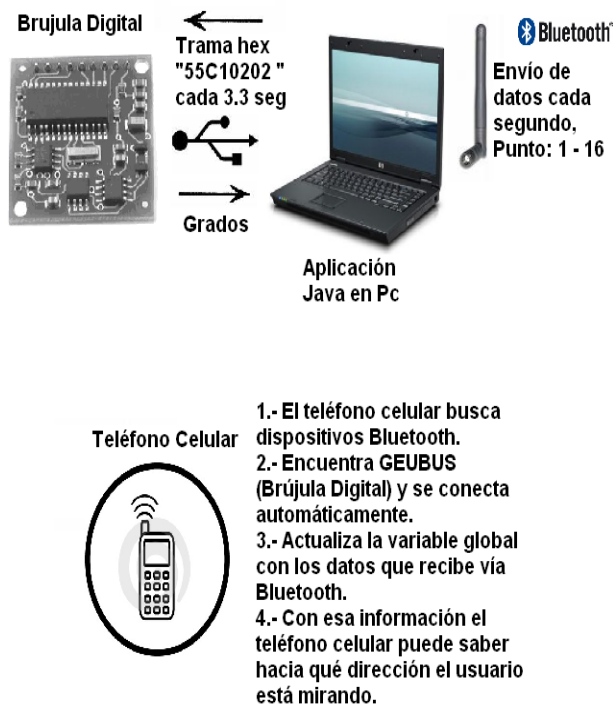


Figura 5. Comunicación brújula digital – teléfono celular

Algoritmo de Decisión

Se desarrolló un algoritmo para decidir qué instrucción se le dará al usuario (modo horaria, a la una, a las dos, etc) dependiendo de hacia dónde está viendo (lectura de la brújula digital) y hacia qué punto cardinal está el objetivo (puertas Indios Verdes, Dr. Gálvez o Salida, *BusquedaBinariaOrient*). El algoritmo le asigna a

cada punto cardinal un número (0=N, 1 = NNE,... 15 =NN0) y haciendo la diferencia del punto cardinal al que el invidente esta viendo contra el del objetivo módulo 15 se obtiene una cifra que hace la relación de qué tan desviado está el usuario respecto al objetivo y así poder darle una respuesta en audio.

Movimiento del usuario respecto a la Medición Base

Para poder ubicar cada una de las puertas de la estación, GeoBus necesita identificar los movimientos del usuario dentro de una estación del MetroBus para así poder darle una respuesta dependiendo de la posición en la que está dentro de la estación. El sistema tomará como punto central las coordenadas GPS del centro de la estación (Punto 1) y las lecturas en tiempo real del GPS será el punto en tiempo real (Punto 2). Para obtener la distancia entre Punto 1 y Punto 2, GEOBUS se basa del hecho de que la latitud y la longitud son distancias en grados al meridiano 0 y al ecuador.

La latitud la suponemos para ambos puntos en latitud Norte (para una latitud sur nos bastaría con cambiar el signo del resultado). Si 90° es 10000 km (mas o menos un cuadrante terrestre) se emplea una regla de tres. Para la latitud de la estación San Simón 19,45968 el cálculo es: 19,45968*100000/90 = 2162,186667. Para la longitud, se sigue un proceso similar, cada grado de longitud nos da (40000/360) unos 111,111 km. Siguiendo con la estación San Simón, 99,14638*111,111 = 11016,26334. Obteniendo el Punto 1 (2162,186667, 11016,26334), suponiendo una lectura del GPS en tiempo real: 19,45973 99,14640, nuestro Punto 2 seria (2162,192222, 11016,26567). Ahora sólo resta aplicar la siguiente fórmula para obtener la distancia entre dos puntos:

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$\sqrt{0.005555^2 + 0.002225^2}$$

Distancia = 0.00598km

El cálculo dio un movimiento de la posición base a la posición en tiempo real de 5.9 metros. Obteniendo este movimiento podemos saber qué tan alejado está el usuario respecto a las puertas de abordaje.

Un movimiento de 4 metros respecto a la medición base conlleva a un cambio de la orientación cardinal de cada puerta de dos valores. Para San Simón, el movimiento fue de 5 metros, por lo que los valores que arroja la *BusquedaBinariaOrient* sufren un cambio de $5/4 = 1$. Si para San Simón teníamos las puertas de abordaje en dirección Indios Verdes: 04 06, por el movimiento presentado, los valores cambiarían a 05 y 07. Una vez hecho este cambio se procede a aplicar *AlgoritmoD* para poder brindarle una respuesta auditiva.

Pruebas. La prueba inicial consistió en solicitarle a dos personas invidentes que asisten a la Biblioteca Central del Estado de Hidalgo Ricardo Garibay, José Luis Piña Ugalde y Gabriela Lara Alcalá, que probaran GEOBUS adaptado a una calle de la colonia Plutarco Elías Calles en Pachuca Hidalgo. La calle simuló una estación del metro bus con las puertas de abordaje y salidas marcadas con listones. Para esto, se incorporaron los datos geográficos y de orientación correspondientes a la calle en cuestión. La siguiente figura muestra el escenario de prueba:

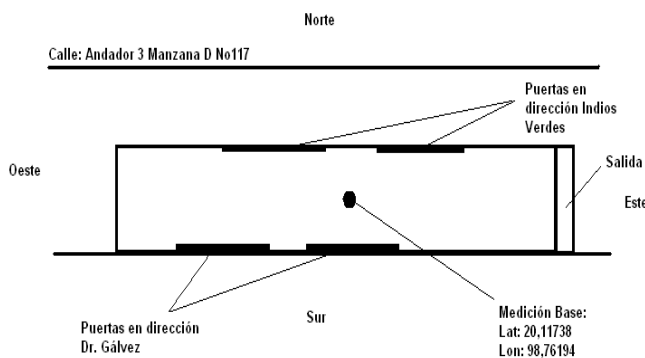


Figura 6. Escenario de prueba Pachuca, Hidalgo

El tamaño que se manejó para la estación fue de 35 metros y los usuarios probaron el sistema en este escenario donde se obtuvieron los siguientes resultados:

1. El sistema identifica la estación dentro de ella y fuera de ella a unos 20 metros a la redonda.
2. El sistema identifica el lado (izquierda, derecha, enfrente o atrás) de las puertas de abordaje de la estación. En esta prueba GEOBUS no identificó con precisión cada una de las puertas, en promedio, se hubo un error de 5 metros.
3. En los extremos de la estación el sistema no pudo identificar las puertas. Esto se debió a que en los extremos, el movimiento respecto a la medición base era el máximo y el movimiento que el sistema realizaba de las puertas de abordaje lo puso al máximo. En nuestro escenario de prueba, viendo al Sur y pidiendo la orientación de las Puertas en dirección Dr. Gálvez, el audio esperado sería el correspondiente a las 10 en punto pero dado al movimiento máximo, el sistema nos decía que la puerta estaba a las 9, se pasó por un valor.
4. La comunicación Bluetooth presentó problemas, a lo largo de la prueba se perdía la conexión del teléfono celular con la brújula digital (PC), esto debido a que la PC y el dispositivo Bluetooth se calentaron mucho por el sol.
5. Por otro lado, los usuarios comentaron que el menú audible era comprensible y cumplía con el objetivo de que puedan navegar a lo largo de éste sin ningún problema. Las teclas de navegación (tecla #5 para aceptar, tecla #2 para subir de opción y tecla #8 para bajar de opción) son fáciles de ubicar gracias a las marcas que los teléfonos celulares tienen en la tecla número 5.
6. El GPS brindó la latitud y longitud en todo momento que GeoBus las requería.

7. La variación que se registró de los valores de la latitud y la longitud a lo largo de la prueba fue de un error de aproximadamente de 3 metros. Este error es un factor delimitante ya que depende de estos datos para su correcto funcionamiento.
8. GeoBus ubicó la salida de manera satisfactoria.

Con los resultados obtenidos en esta prueba se procedió a ajustar el código para resolver el problema presentado en los extremos de la estación. Se tomaron en cuenta movimientos máximos para los valores de orientación dentro de la estación. En lo que se refiere a los problemas con la comunicación Bluetooth, para que no se sigan presentando, se buscó un dispositivo USB-Bluetooth para la PC de mejor calidad que contrarrestará el efecto causado por la radiación del sol. Este escenario de prueba estuvo expuesto al sol en su totalidad, dentro del sistema de transporte MetroBus, la exposición al sol no será tan fuerte pero probarlo en condiciones extremas sirve para asegurar el funcionamiento dentro de las condiciones esperadas.

El sistema fue probado dentro de la Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas (UPIITA). La prueba consistió en simular que una parte de la plaza principal fuese una estación del MetroBus con sus delimitaciones, puertas de abordaje y salida, básicamente la misma metodología que para la Prueba Inicial. Realizadas las pruebas en el escenario UPIITA se concluye que GEOBUS cumple con los siguientes objetivos:

- Contar con un menú audible, el cual, usable para una persona invidente.
- Identificar la estación y reproducir el nombre de ésta, detectando cuando el usuario ha salido de la estación.
- Ubicar el lado de las puertas en cierta dirección.
- Ubicar la salida de la estación

- Ubicar el movimiento del usuario dentro de la estación
- Ubicar cada una de las puertas de abordaje, teniendo en cuenta un error de aproximadamente 4 metros.
- Reproducir las respuestas horarias (“A la una en punto”, “A las seis en punto”, “A las diez en punto”, etc) con audio.

Es necesario realizar pruebas de campo, esto es, en el sistema de transporte colectivo MetroBus para así poder resolver problemáticas sólo ahí se presenten. Para esto se solicitará un permiso al sistema de transporte que nos permita permanecer en las estaciones el tiempo necesario para hacer las pruebas. Este es un trámite ajeno a la escuela y dependemos totalmente de ellos.

V. CONCLUSIONES

Del presente análisis, diseño y desarrollo del sistema se concluye que GeoBus logra asistir tecnológicamente (teléfono celular, GPS y brújula digital) a una persona invidente a saber el nombre de la estación. Así como también la ubicación de las puertas de abordaje en cierta dirección. Sin embargo, es necesaria la adaptación de un dispositivo que funcione de USB-Host y se pueda eliminar la dependencia que tiene la brújula digital de la PC. Así como también cambiar el dispositivo Bluetooth de la PC por uno menos sensible al ruido electromagnético. En cuanto a la precisión del GPS se presenta la siguiente solución, disponer a cada estación con un dispositivo inalámbrico capaz de brindarle información geográfica fija que GEOBUS pueda comparar y así eliminar errores, un GPS-Diferencial. Se propone esta solución para poder asegurar que el error en los datos de la latitud y longitud sea mínima y así asegurarle al usuario la correcta orientación de las puertas de abordaje.

Del trabajo realizado en este proyecto se pueden encontrar el siguiente trabajo a futuro:

- La creación de la funcionalidad del trazado de rutas, tiempos estimado de viajes y estaciones próximas, tomando en cuenta la nueva línea del MetroBus Eje 4 Sur.
 - La creación de la funcionalidad que haga a GEOBUS capaz de actualizarse con información diaria referente a avisos a los usuarios del MetroBus, ya que, la mayoría de estos anuncios son desplegados en mensajes escritos que no se hacen llegar a los usuarios invidentes.
- [6] Enrique Palacios, Fernando Remiro y Lucas J. López, *Microcontrolador PIC16F84*, 2da Edición, Alfaomega
- [7] Ceballos Francisco Javier, *JAVA 2 – Fundamentos de Programación*, 3ra Edición, , Alfaomega
- [8] Prieto Martin, Manuel Jesus, *Desarrollo de Juegos con J2ME*, Editorial Alfaomega
- [9] Letham, Lawrence, *GPS Fácil uso del Sistema de Posicionamiento Global*, Editorial Paidotribo

VII. AUTORES

Del trabajo realizado en este proyecto se desprenden las siguientes aportaciones:

- Un proyecto enfocado para la población mexicana, fomentando la investigación nacional.
- Un proyecto para fomentar el estudio de las aplicaciones para personas con capacidades diferentes.
- El crecimiento del mercado de las telecomunicaciones al abarcar una población que no ha sido considerada.
- Se comenzará con el trámite de la patente de esta idea.

Andrés Jaramillo Calderón es estudiante de la carrera de Telemática de la Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas UPIITA-IPN. Este artículo es parte de su tesis para obtener el título de ingeniero en Telemática.

Dr. en C Miguel Félix Mata Rivera, es doctor en ciencias de la computación y profesor investigador en la academia de telemática en la UPIITA-IPN. Actualmente, dirige diversos proyectos relacionados con el cómputo móvil, las telecomunicaciones y sistemas web orientados al conocimiento y los SIG.

VI. REFERENCIAS

- [1] *Evaluation of Text and Speech Systems*, Laila Dybkjaer, Holmer Hemsén, Wolfgang Minker
- [2] Joseph Rose, "Closet buddy: dressing the visually impaired", Proceedings of the 44th annual southeast regional conference, March 10-12, 2006, Melbourne, Florida
- [3] David Arnow, *Introducción a la Programación con JAVA: un enfoque orientado a objetos*, Ed Prentice Hall.
- [4] *NMEA Reference Manual*, SiRF Technology, Inc.
- [5] Francisco M. Márquez, *UNIX Programación Avanzada*, 3ra Edición, Alfaomega

ABOUT THE PATH LOSS IN THE RFID RADIO CHANNEL

S. Ricardo Meneses G., Laura Montes and Roberto Linares
Instituto Politécnico Nacional

Escuela Superior de Ingeniería Mecánica y Eléctrica, Campus Zacatenco
U.P.A.L.M. Edif. 5, Jefatura de ICE, Col. Lindavista, C. P. 07738, México, D. F.
rmenesesg@ipn.mx; rlinaresy@ipn.mx

ABSTRACT

Most terrestrial wireless communication systems operate in environments that present a multitude of RF propagation problems. The Radio Frequency Identification System, a wireless communication system, applied to vehicular control, is a typical point to point communication system, which one, the radio link formed between the reader, working at the same time as base station and as an end user site, and the tag, the intermediate site, is exposed to electromagnetic disturbances, due reflections off objects in the environment. Meanwhile any misalignment between the antennas at the reader and the tag will result in a polarization mismatch, from which, added to the multipath propagation, the system will see a slight fall of some level of reduction efficiency.

The focus of this article will be a discussion of the path loss in the RFID communication system applied to vehicular control, owed to antennas polarization mismatch and the ground reflections. Experimental and measurements were taken in the field and in the electromagnetic compatibility laboratory

I. INTRODUCTION

The RFID system uses the near and far field electromagnetic coupling concept, for detecting objects. A wireless communication link, is provided between a remote transponder (antenna and integrated circuit), known as the tag, and an interrogator or reader, as showed in the figure 1.

The main functions performed by a reader are demodulating the data retrieved from the tag, decoding the received data, energizing in the case of passive and semi-passive tags, and as an antenna radiating above of a lossy surface [1].

The RFID Technology works in the low, high, ultra high and the super high band frequency. The 18006-C standard establishes the requirements of the physical and logical levels, where the reader starts the communication in the RFID system, operating in the 860-960 MHz frequency range [2], [3].

Recently, this technology is applied to the vehicular control in order to have all the motor vehicle information, for example, owner name, license, fines, etc. It is a radio mobile typical application, showed in the figure 2, where the reader is installed in a fixed position, strapped on a fixed post, as the transmitter, and the tag is strapped on the motor vehicle, as a mobile receiver and starting the data back transmission to the reader, living side by side with other communication systems, in a complex and violent electromagnetic environment. In this application the RFID reader is based in the monostatic antenna configuration, with a single transmit and receive antenna [4], where a directional coupler is used as an isolator, which one, introduces additional path attenuation loss. Making up the propagation media, we have the free space, the reflections from ground [5], [6], this one, made of homogeneous isotropic material with linear

properties, the electric permittivity ϵ_r , and the magnetic permeability μ_r , and the reflections from the various obstacles, some of them statics, and/or some of them, in continuous movement,

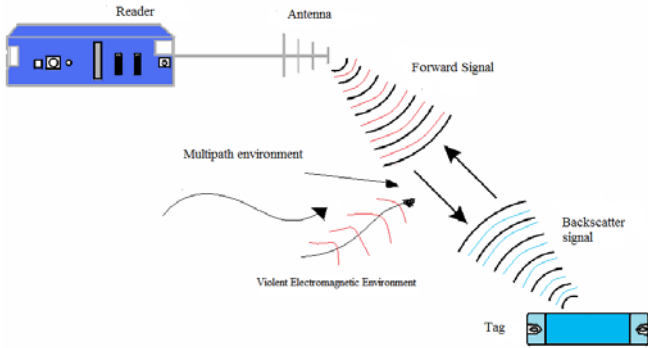


Fig.1. The Radio Frequency Identification Channel

The polarization is an important consideration for RFID reader antennas. For maximizing the tag range, antenna polarization of the tag must be matched to that of the reader antenna. In the RFID applied to the vehicular control, the alignment between the reader antenna and the tag antenna is a difficult situation, due to on motion of the mobile, resulting in a mismatch polarization adding path attenuation loss.

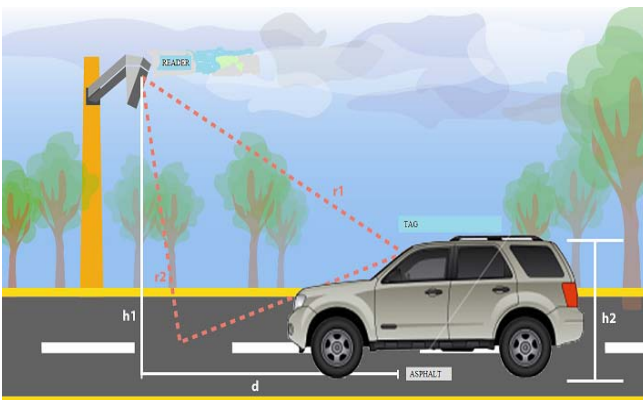


Fig.2. The RFID technology applied to the vehicular control.

The figure shows the components of the radio channel, the direct ray and the ray reflected from the ground.

II. STATEMENT OF THE PROBLEM

A. Polarization Mismatch

During the tests, a Yagi antenna is used as reader antenna, and it was installed forming an angle with the post. In the other hand, a horizontal dipole is used as tag antenna, which one is installed on the windshield of the motor car.

In order to get a maximum electric field in the antenna of the tag, this one must be matched with the reader antenna, but as long as the mobile is closing to the reader, a mismatch polarization between the antennas is present. Polarization mismatch adds an extra loss. The mutual polarization efficiency is expressed as [7]:

$$P = \frac{1 + |\hat{\rho}_1|^2 |\hat{\rho}_2|^2 + 2|\hat{\rho}_1| |\hat{\rho}_2| \cos(\theta_1 - \theta_2)}{(1 + |\hat{\rho}_1|^2)(1 + |\hat{\rho}_2|^2)} \quad \dots (1)$$

where:

- θ_1 , reader antenna polarization ratio phase
- θ_2 , tag antenna polarization ratio phase
- $\hat{\rho}$, complex polarization ratio

The equation 6 represents the correlation between the antennas polarization, and it is a loss due to polarization mismatch. If θ_1 and θ_2 are the phases of the polarization ratios of the antenna, the tag antenna and the incident wave from the Yagi antenna, the maximum and minimum polarization efficiencies occur when $\theta_1 - \theta_2$ equals 0° and 180° respectively. The polarization efficiency becomes:

$$P_{\max/\min} = \frac{(1 \pm |\hat{\rho}_1| |\hat{\rho}_2|)^2}{(1 + |\hat{\rho}_1|^2)(1 + |\hat{\rho}_2|^2)} \quad \dots (2)$$

In the other hand the polarization loss factor, PLF, can be expressed as:

$$\vec{E}_i = \hat{a}_w E_i, \text{ the incoming wave electric field}$$

$$\vec{E}_a = \hat{a}_a E_a, \text{ receiving antenna electric field}$$

Then:

$$PLF = |\hat{a}_w \cdot \hat{a}_a^*|^2 = |\cos \theta_p|^2 \quad \dots (3)$$

where:

θ_p , angle between the two unit vectors

The fig. 3 shows the geometry formed by the antennas position during the tests.

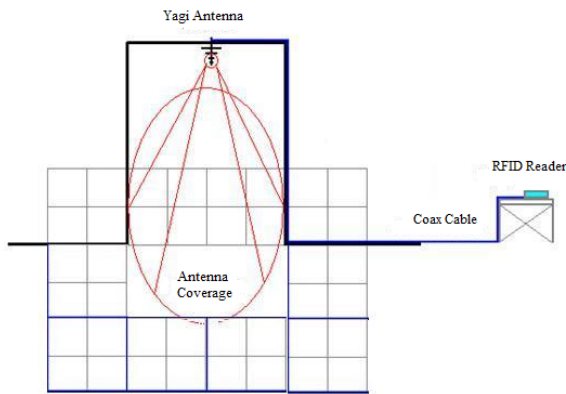


Fig.3a. Polarization Antennas Geometry Frontal View.

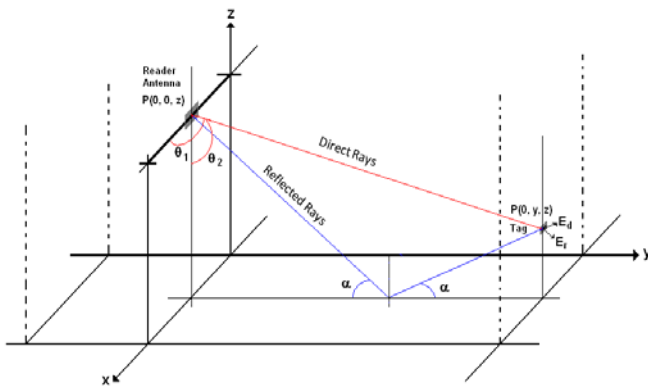


Fig.3b. Polarization Antennas Geometry Lateral View.

A 45° angle is made between the Yagi Antenna and the vertical. Meanwhile the mobile is closing to the fixed station, the distance between them is decreasing and the polarization angle is changing too. In this situation, the polarization loss factor is expressed as 3, where:

$$\Psi = \tan^{-1} \left[\frac{h_1 - h_2}{d} \right] \quad \dots (4)$$

B. Reflections on the ground

In other hand, the two ray model [8] considers the direct wave and the reflected wave, and the reflection coefficient is based in the electric properties of the reflecting surface and the angle of incidence. So, due to contribution of the polarization loss factor, now the received power can be expressed in this way:

$$P_r = P_t G_t G_r (PLF) \left(\frac{\lambda}{4\pi} \right)^2 \left| \frac{1}{r_1} \exp(-jkr_1) + \sum_{i=1}^N \Gamma(\alpha_i) \frac{1}{r_i} \exp(-jkr_i) \right|^2 \quad \dots (5)$$

where:

P_t , transmission power

G_t , Antenna gain of transmitter

G_r , Antenna gain of receiver

d , distance between the transmitter and receiver

r_1 , distance of the direct ray

r_i , distance of the reflected ray

$\Gamma(\alpha)$, reflection coefficient

being:

$$r_1 = \sqrt{d^2 + (h_1 - h_2)^2} = \sqrt{d^2 + h_1^2 + h_2^2 - 2h_1h_2}$$

$$r_2 = \sqrt{d^2 + (h_1 + h_2)^2} = \sqrt{d^2 + h_1^2 + h_2^2 + 2h_1h_2}$$

... (6)

and:

$$\sin \alpha = \frac{\sqrt{d^2 + (h_1^2 + h_2^2)}}{(h_1 + h_2)}$$

where:

α , incidence angle

The first part of expression (5) represents the direct ray, now affected by the polarization loss factor, and the second part represents the reflected ray.

The reflection coefficient depends of the angle of incidence, signal polarization and the relative permittivity of the reflecting surface. So, due to contribution of the polarization loss factor, the reflection coefficient is given by [13]:

$$\Gamma(\alpha) = \frac{\varepsilon_r' \cos(90 - \alpha) - \sqrt{\varepsilon_r' - \sin^2(90 - \alpha)}}{\varepsilon_r' \cos(90 - \alpha) + \sqrt{\varepsilon_r' - \sin^2(90 - \alpha)}}$$

Vertical Polarization

$$\Gamma(\alpha) = \frac{\cos(90 - \alpha) - \sqrt{\varepsilon_r' - \sin^2(90 - \alpha)}}{\cos(90 - \alpha) + \sqrt{\varepsilon_r' - \sin^2(90 - \alpha)}}$$

Horizontal Polarization

... (7)

where:

σ , conductivity of the reflecting surface (asphalt $\sigma = 0.005$)

And the electric permittivity is given by [9]:

$$\varepsilon_r = 15 - j60\sigma\lambda$$

The graphic in the figure 4 shows the received power performance.

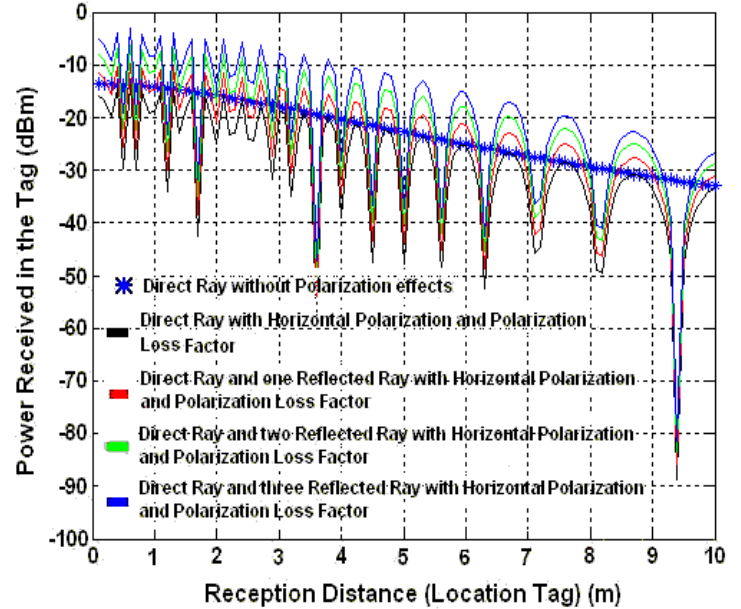


Fig.4. Tag Received Power

The constructive contribution of the reflected waves in the received power (the green curve and the sky blue curve), is showed in the graphic 4.

The facts speak for themselves about the destructive and constructive contributions of the reflections. The figure 5 shows one of the several experiments and measurements made in the field.



Fig.5a. Measurements in the road.



Fig.5b. Measurements in the road.

III. CONCLUSION

Multipath propagation, polarization mismatch antennas, speed of the mobile, speed of surrounding objects, ground reflections and the transmission bandwidth of the signal are the most important physical factors in the radio propagation channel influence. For mobile radio applications, the RFID channel is time-variant because motion between the transmitter and receiver results in propagation path changes.

A propagation model allows us to predict the power, the electric field spatial distribution, and the polarization state of the wave, in order to optimize the receiver operation.

IV. REFERENCES

- [1] Finkenzeller, K., *RFID Handbook: Radio-Frequency Identification Fundamentals and Applications*, 2nd edition, Wiley, New York, 2003.
- [2] *International Standard ISO/IEC 18043-6, Information Technology-Radio Frequency Identification Device Performance Tests Methods*, First edition 2007-09-15.
- [3] *860MHz – 930MHz Class I Radio Frequency Identification Tag Radio Frequency & Logical Communication Interface Specification Candidate Recommendation, version 1.0.1*. Auto-Id Center
- [4] M. Reynolds and C. Weigand, “*Design considerations for embedded software-defined RFID readers*”, Emerging Technology, August 2005.
- [5] S. Sa-e, M. Chamchoy and S. Promwong, “*Study on Propagation Path Loss and VBR Performance for Fixed Broadband WiMax*”, Proceeding of Asia-Pacific Conference on Communications 2007.
- [6] Dolukhanov, M. *Propagation of Radio Waves*, MIR Publishers, Moscow, USSR, 1971.
- [7] Thomas A. Milligan, *Modern Antenna Design*, IEEE Press
- [8] H.S. Kim, R.M. Narayanan, “*A New Measurement Technique for Obtaining the Complex Relative Permittivity of Terrain Surfaces*”, IEEE Transactions on Geoscience and Remote Sensing, 2002.
- [9] Anderson, H. A. “*A Ray Tracing Propagation Model for Digital Broadcast Systems in Urban Areas*”, IEEE Transaction on Broadcasting Vol. 39, Number 3, September 1993.
- [10] Parsons, J. D. *The Mobile Radio propagation Channel*, Second Edition John Wiley & Sons, England 2000.
- [11] P. V. Nikitin, Rao, K.V.S., “*Antennas and Propagation in UHF RFID Systems*”, IEEE RFID 2008 Conference Proceedings (Las Vegas, NV, April 16-17, 2008).
- [12] P. V. Nikitin, K. V. S. Rao, and S. Lazar, “*An overview of near field UHF RFID*”, IEEE RFID Conference, pp.167-174, 2007

- [13] Xia H.H., Bertoni H.L., Maciel L.R., Lindsay-Stewart A. and Rowe R., “*Radio Propagation Characteristics for Line-of-Sight Microcellular and Personal Communications*”, IEEE Trans. Antennas Propagat., vol 41, pp. 1439-1446, 1993.
- [14] T. Rappaport, *Wireless Communications Principles and Practice*, Prentice Hall, New Jersey 1996.
- [15] K. I. Blackard, M. J. Feuerstein, T. S. Rappaport, S. Y. Seidel. And H. H. Xia, “*Path loss and delay spread models as function of antenna height for microcellular system design*”, Proceedings of the 1992 IEEE Vehicular Technology Conference, Denver, CO, pp. 333-337, May 1992.

Laura Leticia Montes Peralta Communications and Electronic Engineer from ESIME IPN. She is actually professor and departmental assistant manager of Communications and Electronic Engineering of the ESIME Zacatenco, IPN.



Dr. Roberto Linares y Miranda, Ph. D. degree from Centro de Investigación y de Estudios Avanzados del IPN (CINVESTAV- IPN). He is actually professor in the Post-graduate Section of the ESIME Zacatenco.



V. BIOGRAPHIES

Salvador Ricardo Meneses González. Communications and Electronic Engineer from ESIME IPN (1981). M. S. E. degree in electrical engineering from Centro de Investigación y de Estudios Avanzados del IPN (CINVESTAV IPN). He is actually professor and departmental head of Communications and Electronic Engineering of the ESIME Zacatenco, IPN.



ANÁLISIS DE PROPAGACIÓN ELECTROMAGNÉTICA MEDIANTE UN SISTEMA GEORREFERENCIADO

Jorge R. Sosa-Pedroza, Fabiola Martínez-Zúñiga,
Carlos Quiroz Pichardo Francisco Daniel Sánchez Dámaso

Sección de Estudios de Posgrado e Investigación
Escuela Superior de Ingeniería Mecánica y Eléctrica
Instituto Politécnico Nacional
jsosa@ipn.mx; fmartinezzu@ipn.mx

RESUMEN

Los sistemas de información geográfica (SIG) son herramientas de aplicaciones muy amplias, las telecomunicaciones no son la excepción; mediante la superposición de capas de Bases de Datos de Terreno (BDT) es posible usarlos en el cálculo de enlaces o análisis de sistemas ya establecidos, se propone en este trabajo el uso de los SIG para comparar los efectos de la altura de edificios sobre la propagación electromagnética para aplicaciones de comunicaciones móviles en la Ciudad de México. La superposición georeferenciada de bases de datos de uso de suelo y de mediciones en la zona de estudio, permite analizar cómo los edificios o las zonas arboladas influyen en la distribución de campo sobre el terreno.

Palabras clave: Sistemas de Información Geográfica, sistemas georeferenciados, Bases de Datos Digitales, Propagación Electromagnética, comunicaciones móviles.

I. INTRODUCCIÓN

Las características de propagación en un sistema de comunicaciones móviles es un problema ampliamente estudiado en muchas partes del mundo; mediante la medición específica en las ciudades de interés, se han propuesto modelos de atenuación que después se han generalizado para otros sitios, sin embargo, para cada lugar es importante corroborar si son aplicables o si las mediciones específicas tienen diferencias que deben llevar a

un ajuste de los parámetros del modelo, sin embargo este proceso suele ser costoso y pocas ciudades tienen una caracterización ad-hoc por lo que se prefiere confiar en el desarrollo empírico original aplicándolo prácticamente como un hecho. El modelo más ampliamente usado para comunicaciones móviles en el mundo, para frecuencias entre 150 MHz y 1500 MHz, es el desarrollado por Hata [1], que define una formulación empírica a partir de mediciones de propagación hechas en Japón por Okumura [2], posteriormente el modelo fue modificado para incluir frecuencias en las bandas de telefonía celular en el rango de 2 GHz y se propuso para su aplicación por parte de la UIT en la recomendación COST 231 [3]. Los autores de este trabajo han hecho mediciones en la ciudad de México y procesado la información obtenida, para comparar sus resultados con el modelo de Hata en 400 MHz, encontrando diferencias importantes de hasta 6 dB en algunas zonas específicas. Se considera que el procedimiento se puede invertir, de modo que para una ciudad tipo, moderna, de edificios promedio en altura y materiales de construcción, usando sistemas de información geográfica y bases digitales de terreno se puede predecir el comportamiento de la propagación. En este artículo se usan los sistemas de información geográfica, aplicados a los sistemas de comunicaciones, definiendo los datos necesarios requeridos en el diseño: alturas de terreno, la presencia de obstáculos, de áreas de reflexión y refracción electromagnética; en las ciudades la

altura de edificios, así como su posición geográfica, la caracterización media de materiales de construcción, el régimen de uso de suelo, etc., la digitalización de la información representa la Base Digital de Terreno (BDT), que permite sistematizar el análisis y el diseño de enlaces de comunicaciones. Presentamos los primeros análisis en esta propuesta de metodología, usando un sistema computarizado de información geográfica, con capas representando las calles de la ciudad de México, las bases de datos de usos de suelo y altura de edificios y la superposición de un mapa de mediciones de propagación en UHF, la comparación de las capas superpuestas permite determinar las razones de atenuación debido a la presencia de obstáculos: edificios, zonas arboladas o zonas abiertas.

II. MODELO OKUMURA-HATA

El modelo Hata es el más usado en la planeación y predicción de propagación de los sistemas de comunicaciones móviles, en el rango de 150 MHz a 1500 MHz; partiendo de los estudios de Okumura, modificó el modelo proponiendo el uso como una nueva variable el factor de corrección de altura de la antena del móvil, definida a partir del “tamaño” de la ciudad, en relación con la altura de los edificios, así una ciudad mediana es aquella que tiene edificios de alturas promedio inferiores a 10 m y una ciudad grande la que tiene edificios de altura superior; por otro lado una zona suburbana es la que tiene edificios bajos y espacios amplios y una zona rural la conformada por espacios abiertos con alguna vegetación, por supuesto en una misma ciudad puede haber condiciones diferentes y debe aplicarse aquella que corresponda a la zona en que se desea determinar la atenuación. El modelo de Hata permite determinar las pérdidas de propagación a partir de la ecuación empírica dada por:

$$L_{50}(\text{dB}) = 69.55 + 26.16 \log f_c - 13.82 \log h_{te} - a(h_{re}) + (44.9 - 6.55 \log h_{te}) \log d \quad (1)$$

En donde:

f_c = Frecuencia portadora [150 - 1500 MHz]

h_{te} = Altura de antena transmisora en metros para 30-200 m

h_{re} = Altura de antena receptora en metros para 1-10 m.

$a(h_{re})$ = Factor de corrección para la altura efectiva de la antena móvil,

d = Distancia entre transmisor y receptor en km

La corrección para la altura de la antena móvil se establece en la siguiente forma:

Para ciudades pequeñas y medianas:

$$a(h_{re}) = (1.1 \log f_c - 0.7) h_{re} - (1.56 \log f_c - 0.8) \text{ dB} \quad (2)$$

Para ciudades grandes:

$$a(h_{re}) = 8.29 (\log 1.54 h_{re})^2 - 1.1 \text{ dB} \quad (3)$$

donde h_{re} es una función de la frecuencia, dada por:

$$a(h_{re}) = 8.29 (\log 1.54 h_{re})^2 - 1.1 \text{ dB} \quad (4)$$

para $f_c < 300$ MHz

$$a(h_{re}) = 3.2 (\log 11.75 h_{re})^2 - 4.97 \text{ dB} \quad (5)$$

para $f_c > 300$ MHz

Para ambientes suburbanos o rurales, se aplican las siguientes correcciones:

Áreas suburbanas:

$$L(\text{dB}) = L_{50}(\text{urbano}) - 2[\log(f_c/28)]^2 - 5.4 \quad (6)$$

Áreas rurales:

$$L(\text{dB}) = L_{50}(\text{urbano}) - 4.78(\log f_c)^2 + 18.33 \log f_c - 40.94 \quad (7)$$

El procedimiento de Hata ha sido usado por los autores para compararlo con un plan de mediciones sobre la Ciudad de México [4], sin embargo, mediante el conocimiento de elevaciones de terreno, complementada con altura de edificios definida a través de las características de uso de suelo, es posible definir características de cobertura usando un análisis gráfico superponiendo los resultados de esas mediciones con bases de datos, georeferenciados usando el sistema ARCGIS. La comparación entre el modelo Hata, las mediciones y el análisis gráfico puede establecerse una nueva metodología, que

permita predecir las características de propagación a través de ese análisis gráfico.

III. SISTEMAS DE INFORMACIÓN GEOGRÁFICA

Los sistemas de información geográfica [5], [6], tienen una amplia variedad de aplicaciones: la simple descripción del terreno en función de sus coordenadas espaciales, la caracterización de asentamientos humanos, la definición de uso de suelo, el diseño de mapas de riesgo específico, etc. En las telecomunicaciones los sistemas de información geográfica juegan un papel fundamental en el diseño de enlaces radioeléctricos, a partir de ellos, frente a una computadora, se puede definir la línea de vista del enlace o la presencia de obstáculos en el terreno que permiten o no tal enlace y por tanto la búsqueda de alternativas que lo hagan viable. En una ciudad, ya sea un enlace punto a punto o de radiodifusión, los obstáculos que representan los edificios deben ser caracterizados en el cálculo de atenuación por trayectoria. Una base de datos, que parte del sistema de información geográfica, es determinante en el diseño del enlace punto a punto o el área de cobertura en radiodifusión. Este trabajo emplea los SIG para analizar los efectos de obstáculos sobre la propagación de un sistema de comunicaciones móviles en una zona específica de la Ciudad de México, para ello se usan bases digitales (BD) tanto de alturas de terreno (BDT) como de uso de suelo (BUS) de la ciudad para sistematizar el análisis. A partir de las BD se puede reconocer la presencia de obstáculos y su composición con información georeferenciada con longitud, latitud y elevación; para un enlace rural los obstáculos se refieren a montañas, colinas, bosques, reservas de agua etc, mientras que para enlaces ciudadanos la altura y posición geográfica de edificios, así como la caracterización media de materiales de construcción, el régimen de uso de suelo y otros datos. Partiendo de la información georeferenciada es posible conocer la atenuación por trayectoria y a partir de ello la viabilidad del sistema de comunicaciones. El

propósito de este trabajo es superponer 3 capas de información en el ARCGIS: La base de datos de las calles del área de análisis, la base de datos de altura de edificios, que incluye la de altura de terreno, a partir de información sobre el uso de suelo de la zona y el mapa de cobertura obtenido de mediciones hechas en la misma zona.

IV. MEDICIÓN DE COBERTURA

La medición de potencia recibida se desarrolló montando la antena receptora sobre el techo de un automóvil, conectada a un receptor ANRITSU que guarda en memoria 3 datos por segundo, generando un archivo cada minuto, lo que representa aproximadamente uno cada 400 m dependiendo de la cantidad de tráfico y de las vías recorridas, que determinaron la velocidad del móvil; aunque el experimento se realizó en toda la ciudad de México, los resultados que se presentan incluyen únicamente el Centro Histórico, que es cubierto por seis estaciones base, de las cuales presentamos resultados de dos de ellas, la figura 1 muestra un mapa del área de estudio. Aunque cada estación base tiene varias frecuencias asignadas, por razones de confidencialidad, definimos genéricamente la banda de trabajo como UHF y no se establece su posición exacta. Todas las mediciones están georeferenciadas de tal modo que la posición relativa de cada estación base y los puntos de cada archivo, pueden ser localizados en un mapa igualmente georeferenciado. El recorrido se diseñó procurando cumplir con las siguientes directivas:

- Definir el área de análisis
- Incluir todas las frecuencias de las estaciones base que dan servicio en esa área
- La ruta se estableció viajando primero sobre el perímetro seleccionado y después cubriendo la mayor cantidad de calles de norte a sur y oriente a poniente.

- Se cubrieron alrededor de 80 Km.
- Se generaron 204 archivos.
- Cada archivo se procesó tomando la potencia promedio por estación.



Figura 1. Mapa de la zona de estudio

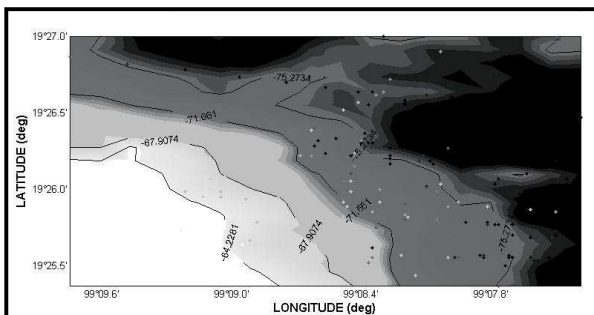


Figura 2. Mapa de cobertura

V. PROCESAMIENTO DE LA INFORMACIÓN

La información se procesó usando el sistema ARCGIS [5] y el software Easy Krieg, construyendo los mapas de cobertura a partir de la extrapolación de los datos de valores puntuales. El proceso de Easy Krieg [6] permite construir mapas continuos a partir de valores puntuales, como el que se muestra en la figura 2.

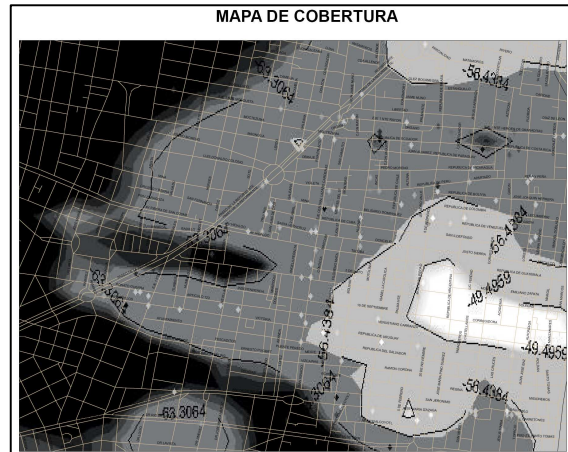


Figura 3. Superposición de capa de calles y de cobertura

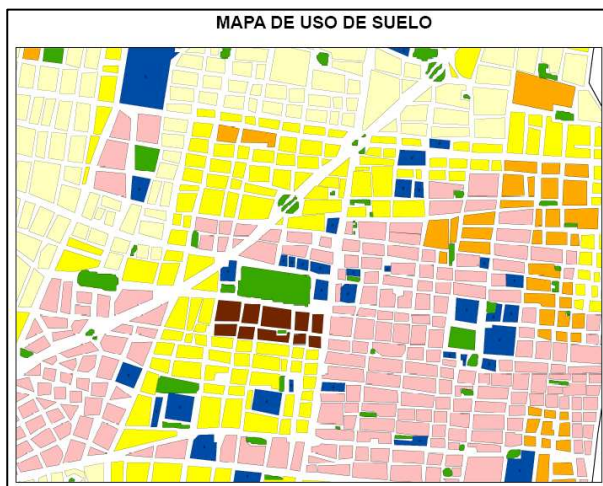
Los mapas continuos fueron definidos como una capa en el sistema ArcGIS, superpuestos en otra capa georeferenciada de las calles del área de análisis, como muestra la figura 3. Se observa en la figura una gama de tonos de grises que ilustran cómo se comporta la potencia en el área de medición, el blanco indica mayor nivel, que disminuye en la escala de grises, hasta llegar al negro que indica el valor menor de la escala, cuantitativamente la potencia está definida por las isolíneas que se observan en la figura, que muestran valores entre -49.4959 y -63.3064 dBm.



Figura 4. Mapa de uso de suelo y cobertura estación base 1.

VI. MAPA DE USO DE SUELO

Para analizar el efecto de los edificios sobre la propagación, se superpuso en el ArcGIS otra capa que representa el mapa de uso de suelo de la zona de análisis, que representa el promedio de altura de edificios, marcado por colores, como muestra la figura 4.



	PISOS	METROS
	2	7.2
	3	10.8
	4	14.4
	5	18.0
	>6	>21.6
	Zona verde	Promedio 20

Figura 5. Mapa de uso de suelo

VII. SUPERPOSICIÓN DE LAS TRES CAPAS EN ARCGIS

La superposición de las tres capas: calles, uso de suelo y propagación, permite determinar la correlación entre ellas, como muestran las figuras 5 y 6, para las dos estaciones base consideradas. Se puede ver que en la estación base 1, existe una pérdida significativa en la zona de la Alameda, el Palacio de Bellas Artes y la Torre Latinoamericana, marcada con rojo, se observa una caída drástica de 10 dBm, que se

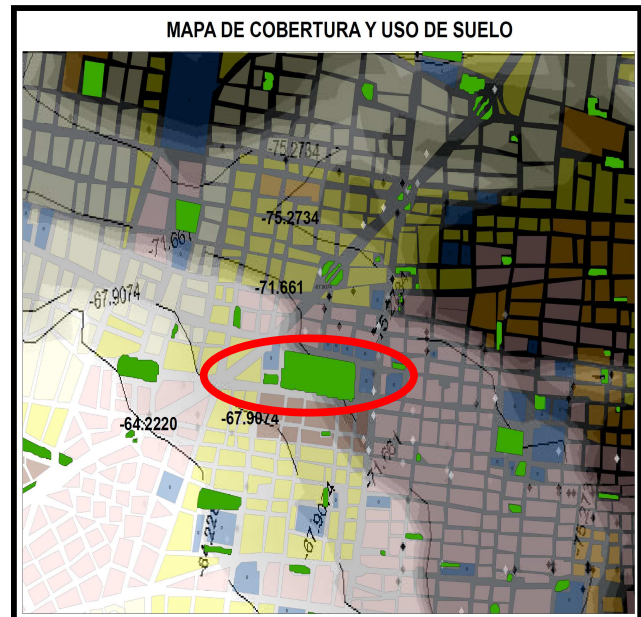


Figura 6. Mapa de uso de suelo y cobertura estación Base 2.

puede interpretar por la sombra de esos edificios, pero principalmente por la atenuación que generan los árboles de la Alameda, un efecto similar se observa en la estación base 2, aunque menos marcada porque no “ve” los edificios altos de la primera y el efecto se debe exclusivamente a los árboles. Por otro lado se observa que mientras la altura promedio de los edificios se mantenga constante (definida por el color de uso de suelo), la potencia no sufre cambios drásticos, como se ve cerca de la plaza de la Constitución y en otras áreas, fácilmente identificada.

VIII. CONCLUSIONES

Hemos presentado en este trabajo la comparación georeferenciada de la propagación electromagnética en UHF y la altura de edificios, definida por el mapa de uso de suelo, sobre el centro de la Ciudad de México. Se muestra la congruencia de la distribución de campo en relación con los obstáculos que encuentra, tanto de edificios como de áreas verdes. Este primer ensayo es el principio para definir una metodología de ingeniería inversa, que partiría de la altura de edificios y su

distribución en la zona de interés, tomando en cuenta la altura de la antena transmisora.

IX. REFERENCIAS

- [1] Masaharu Hata, "Empirical Formula for Propagation Loss in Land Mobile Radio Services", IEEE Transactions in Vehicular Technology", Vol. VT-29 No.3, August 1980
- [2] Y. Okumura et al. "Field Strength and its variability in VHF and UHF Land Mobile Radio Service" Rev. Elec. Commun. Lab., Vol. 16, 1968.
- [3] European Cooperative for Scientific and Technical Research, Rec. COST 231.
- [4] Jorge Sosa-Pedroza, Fabiola Martínez Zúñiga, Miguel Sánchez-Meraz, "Comparing Hata Model and Mexico's City Measurements in 400 MHz". IEEE Latincom, Medellín Colombia, 2009.
- [5] Fabiola Martínez Zúñiga, "Los sistemas de información Geográfica Aplicados en Telecomunicaciones", Tesis de Maestría, Febrero 2008, Director de Tesis: Dr. Jorge Sosa Pedroza.
- [6] Turner A.K. "The role of three-dimensional Geographic Information Systems in subsurface characterization for telecommunications applications" at J. "Raper: Three-dimensional applications in Geographic Information Systems", London, Taylor & Francis, pp115 – 128, 1989.
- [7] Doyle, A.G Clausen, F. L. (1978:1481). "Practical universal kriging and automatic contouring", Geo-Procesing, núm 1, pp. 377-394.

X. AUTORES



Dr. Jorge R. Sosa Pedroza Profesor-Investigador Titular de tiempo completo de la ESIME-IPN de 1973 a la fecha. Ingeniero en Comunicaciones y Electrónica de la ESIME IPN, con estudios de Maestría en Ciencias con especialidad en

Telecomunicaciones en la Sección de Posgrado de la ESIME y doctorado en Telecomunicaciones por la misma institución.



M en. C. Fabiola Martínez Zúñiga Profesora-Investigadora de la ESIME-IPN de 2003 a la fecha. Ingeniera en Comunicaciones y Electrónica, con una Maestría en Ciencias con especialidad en Telecomunicaciones en la Sección de Posgrado de la ESIME. Profesora Titular de las academias de Electromagnetismo y Comunicaciones.



Carlos Quiroz Pichardo Francisco Pasante de la carrera de Ing. en Comunicaciones y Electrónica con especialidad en Comunicaciones de ESIME-IPN



Francisco Daniel Sánchez Damaso Pasante de la carrera de Ing. en Comunicaciones y Electrónica con especialidad en Comunicaciones de ESIME-IPN

UNA APROXIMACIÓN DE RADIO COGNITIVA PARA APLICACIONES MÓVILES

Salvador Silva López¹, José Luis Cuevas Ruiz².

Departamento de Ingeniería Eléctrica y Electrónica
 Escuela de Graduados en Ingeniería y Ciencias
 Tecnológico de Monterrey Campus Estado de México
 Carretera Lago de Guadalupe Km 3.5, Atizapán de Zaragoza, Estado de México, C.P. 52926

correo electrónico: A01165974@itesm.mx¹, jose.cuevas@itesm.mx²

RESUMEN

El incremento en los requerimientos de capacidad de los sistemas de comunicaciones móviles traen consigo una creciente demanda de espectro radioeléctrico, lo que ha generado problemas de congestión en dichas bandas. Radio cognitivo ofrece una solución viable a éste problema al proponer el uso oportuno de rangos de frecuencia dentro de la banda de operación que no estén ocupadas por usuarios con licencia. En este artículo, se analiza el impacto que la radio cognitiva puede tener en el aprovechamiento eficiente del espectro, además de una metodología para la asignación del espectro radioeléctrico de forma dinámica y sin interferencias a los usuarios en la banda de GSM. Se hace uso de esquemas de simulación emulando un escenario real.

Palabras clave: Radio cognitiva, espectro radioeléctrico, agujero espectral, comunicaciones móviles.

I. INTRODUCCIÓN

La tendencia actual de los sistemas inalámbricos modernos es ofrecer una gran variedad de aplicaciones a altas tasas de transmisión a una gran cantidad de usuarios al mismo tiempo. Para llevar a cabo éste tarea se tiene que lidiar con las limitantes prácticas, impuestas por los recursos que se necesitan, tales como son potencia y espectro electromagnético. Ante el incremento del número de usuarios de estos sistemas, la escasez de los recursos mencionados, principalmente del espectro

frecuencial, se convierte en un problema vital día a día.

Las corporaciones legislativas de varios países (como la Comisión Federal de Telecomunicaciones en Estados Unidos ó FCC y la Office of Communications OFCOM en Inglaterra) han observado que la mayor parte del espectro de radio frecuencia está siendo ineficientemente utilizado. Por ejemplo, las bandas de la red celular están sobrecargadas en la mayor parte del mundo, pero la banda de radioaficionados ó las frecuencias de localización están siendo bastante desaprovechadas, tal y como se muestra en la gráfica a continuación:

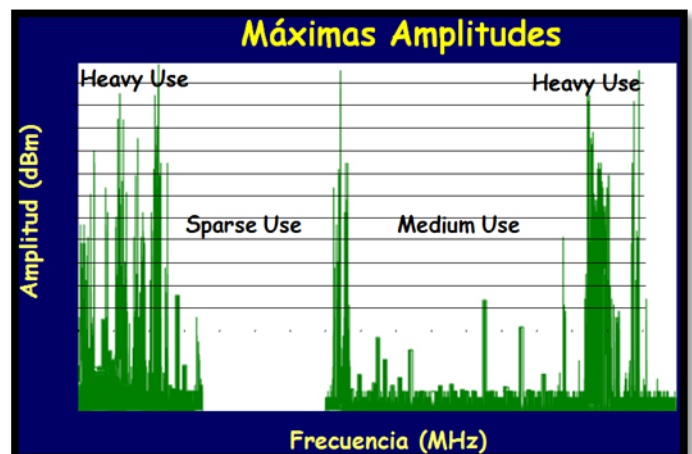


Fig. 1 Utilización del espectro fijo frecuencial
 [Fuente: FCC]

La radio cognitiva [1] es un concepto que apunta a solucionar el problema de congestión al

proponer un uso eficiente del espectro [2], en las cuales, las bandas de frecuencia que no estén ocupadas por sus usuarios licenciados sean utilizadas mediante radio cognitiva; éste acercamiento conlleva a un uso altamente económico y eficiente del espectro frecuencial.

Las propiedades primarias de la radio cognitiva son: su capacidad de sensado, adaptación, capacidad de aprender, así como su memoria. El diseño presentado en éste trabajo se enfoca principalmente en sensado y aspectos de adaptación. Es posible sensar numerosos factores, tales como espectro utilizado, ambientes RF, nodos disponibles en la red y potencia disponible. La capacidad de adaptación, incluye adaptar los parámetros del sistema, tales como transmisión y recepción sin necesidad de intervención del usuario.

Los sistemas de radio cognitiva continuamente escanean el espectro y detectan espacios en el espectro, que son las bandas frecuenciales pertenecientes al usuario licenciado, pero no están siendo usadas en ese tiempo en particular en una localización específica [3][4]. Se explotan los huecos espectrales haciendo uso de la habilidad espectral para adaptar sus parámetros a condiciones de cambio muy rápidas. Cada vez que el espectro es sensado y se detectan los huecos espectrales, se le da forma a su pulso a transmitir, de tal forma que el espectro ocupado por el pulso, encaje en los huecos espectrales detectados. La figura 2 muestra el concepto de agujero espectral:

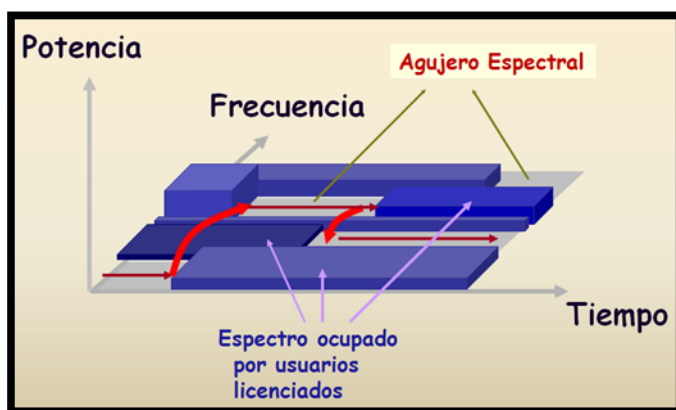


Fig. 2 Concepto de agujero espectral

La radio cognitiva habilita el uso de espectro no usado, temporalmente (agujero espectral ó espacio en blanco), si ésta banda ya está siendo usada por un usuario con licencia (usuario licenciado) RC se mueve a otro agujero espectral ó se queda en la misma banda, altera su nivel de potencia de transmisión ó esquema de modulación para evitar interferencia.

En las comunicaciones de radio cognitiva uno de los mayores problemas es como compartir la información de sensado del espectro con otros usuarios cognitivos. En la literatura, se considera tener un canal de asignación de control para transmitir ésta información [5]. En algunos trabajos, se propone tener un controlador centralizado que reúna ésta información, decida la disponibilidad espectral y asigne distintas bandas a diferentes usuarios cognitivos [6],[7].

En años recientes, muchos algoritmos diferentes para el control de radio cognitiva han sido propuestos, incluidos los basados en teoría de juegos [8], la aproximación por diseño de experimentos [9], ó modelos LP [10]. Éstos algoritmos son la componente clave de un sistema de radio cognitiva, debido a que administran como una radio cognitiva aprenderá y reaccionará al ambiente exterior.

Uno de las componentes más importantes del concepto de radio cognitiva es la habilidad para medir, sensar, aprender e identificar los parámetros relacionados a las características del canal de radio, disponibilidad del espectro y potencia, ambiente operacional de radio, requerimientos de usuario y aplicaciones, redes disponibles y nodos, políticas locales y otras restricciones de operación. En la terminología de radio cognitiva los *usuarios primarios* se definen como los usuarios que tienen mayor prioridad en el uso de una parte específica del espectro. Por el otro lado, *usuarios secundarios*, los cuales tienen baja prioridad, explotan el espectro de tal forma que no causen interferencia a usuarios primarios. Por lo tanto, usuarios secundarios necesitan tener capacidades, tales como sensar la confiabilidad del espectro para revisar, ya sea que esté siendo usado por un

usuario primario y cambiar los parámetros de radio para explotar la parte no usada del espectro.

La reconfigurabilidad es la capacidad de ajustar parámetros de operación para la transmisión “al vuelo” sin ninguna modificación a componentes de hardware. Ésta capacidad habilita la RC para adaptarse fácilmente al ambiente de radio dinámico.

Los parámetros reconfigurables son:

- i) Frecuencia de operación
- ii) Modulación
- iii) Potencia de transmisión
- iv) Tecnología de comunicación

Los algoritmos de sensado de espectro convencionales no consideran tratar con señales que usan espectro disperso, códigos de salto de frecuencia ó de tiempo. Como resultado, éstos tipos de señales constituyen un reto mayor en el sensado del espectro.

Otro uso muy importante que se prevé en varios estudios, es la adaptación de la radio cognitiva para introducir servicios y aplicaciones 4G. La novedad en 4G radica en el receptor, se debe adaptar para encontrar el rango de frecuencias y proveer el mismo servicio, por poner un ejemplo, se plantea el escenario donde un usuario viaja a una zona remota donde solo se cuenta con el servicio de una red inalámbrica (802.11 ó WiFi), el dispositivo celular debe ser capaz de usar esa banda para adaptarse y tener disponible el servicio celular.

II. MÉTODO DE SIMULACIÓN

La metodología de simulación de un esquema de radio cognitiva en un ambiente de sistemas de comunicaciones móviles se basa en las siguientes etapas:

- 1) Determinar cuáles porciones (canales) del espectro estén disponibles y detectar la presencia de usuarios licenciados cuando un usuario opere en una banda licenciada (*Sensado del espectro*).

Una RC monitorea las bandas de espectro disponibles, captura la información y detecta los agujeros espectrales.

- 2) Seleccionar el mejor canal disponible (*Decisión espectral*).

Basado en la disponibilidad espectral, los usuarios de RC pueden determinar un canal. Ésta operación no solo depende en la disponibilidad espectral, sino que también está determinada basada en políticas internas (y posiblemente externas).

- 3) Coordinar acceso a éste canal con otros usuarios (*Compartición del espectro*).

Múltiples usuarios de RC intentan acceder el espectro.

El acceso a la red de RC debe ser coordinado para prevenir que colisionen múltiples usuarios en porciones traslapantes del espectro.

- 4) Hacer vacante el canal disponible cuando un usuario con licencia sea detectado (*Movilidad Espectral*).

A los usuarios de RC se les considera como visitantes del espectro. Si los usuarios primarios necesitan una porción específica del espectro, entonces los usuarios de RC deben irse hacia otra porción vacante del espectro

A continuación se muestra el proceso general que se necesita implementar para llevar a cabo el proceso de radio cognitiva

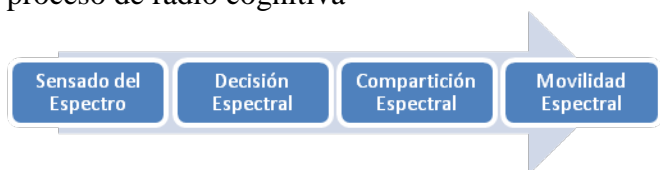


Fig. 3 Flujograma para implementar algoritmo de radio cognitiva.

Una RC determina los parámetros de comunicación adecuados y se adapta al ambiente dinámico de radio. Las tareas requeridas para la operación adaptativa en un ambiente abierto se les conoce como *ciclo cognitivo*.



Fig 4. Ciclo Cognitivo

A GSM ha sido asignada una frecuencia operacional de 890 MHz hasta 960 MHz. Para reducir posibles interferencias, el MS (Mobile Switching) y la BS (Base Station) usan diferentes rangos de frecuencias (MS emplea 890 MHz a 915 MHz y la BS opera en 935 MHz a 960 MHz). GSM se basa en FDMA y permite mas de 124 MSs en servicio al mismo tiempo (la banda de frecuencia de 25 MHz se divide en 124 canales de multiplexaje por división de frecuencia (FDM), cada uno de 200 kHz como se muestra en la figura 5. Se usa una trama de guarda de 8.25 bits entre dos tramas transmitidas, ya sea por la BS ó el MS.

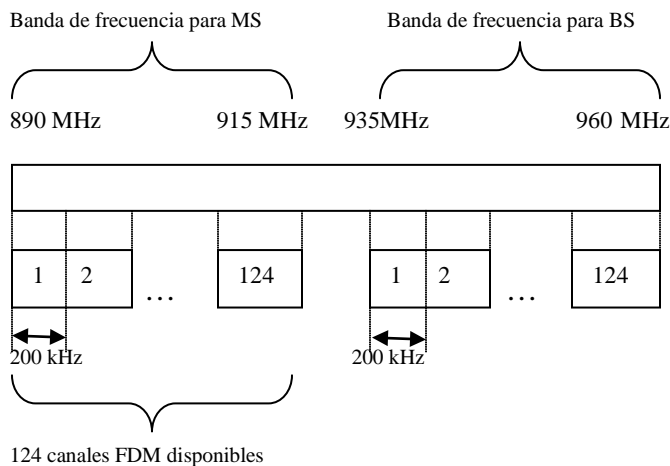


Fig 5. Bandas de frecuencia usadas por GSM.

Para la implementación en código Matlab se propone un escenario con 5 usuarios primarios

GSM. Para fines de ejemplificarlo proponemos limitarlo a 5 canales, uno por cada usuario. También existe un usuario secundario, el cual continuamente sensa el espectro para encontrar el agujero espectral, éste se dará, ya sea apagando el dispositivo ó cuándo un usuario salga de la red.

Durante la simulación se pregunta al usuario cuáles canales están ocupados, se introducen solo 3 canales dejando al canal 2 y 4 desocupados, por poner un ejemplo. También se puede simular otro escenario donde todos los usuarios primarios hagan uso de la red y en ese caso el usuario secundario tiene que esperar por la oportunidad ó lo que se conoce como un agujero espectral.



Fig 6. Escenario de comunicaciones propuesto para la simulación.

La Densidad de Potencia Espectral (Power Spectral Density) de una señal es una función matemática que nos informa de cómo está distribuida la potencia de dicha señal sobre las distintas frecuencias de las que está formada, es decir, su espectro.

Para obtener la potencia de densidad espectral se hace uso de la definición de densidad de potencia espectral como se describe a continuación.

Para el caso de señales definidas en potencia:

Una señal $x(t)$ es definida en potencia si su potencia media es finita, i.e, $0 < P_x < \infty$ y por tanto, su energía media es infinita, $E_x = \infty$.

La DPE (Densidad de Potencia Espectral) se calcula usando el teorema de Wiener-Khinchin el cual relaciona la DPE con la transformada de Fourier de la función de autocorrelación.

$$S_{xx}(f) = TF\{R_{xx}(\tau)\} = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-2\pi i f \tau} d\tau$$

expresado en [W/Hz]

donde TF significa Transformada de Fourier y R_{xx} es la función de autocorrelación de $x(t)$.

El valor $S_{xx}(0)$ es la potencia de la componente continua (DC) de la señal. La integral de esta función en todo el eje f es el valor de la potencia total de la señal $x(t)$

$$P = \int_{-\infty}^{\infty} S_{xx}(f) df$$

Usando el concepto de correlación cruzada es posible definir también la densidad espectral cruzada.

$$S_{1xy}(f) = TF\{R_{1xy}(\tau)\}$$

Como se puede ver en la figura 3, es posible detectar los huecos ó espacios espectrales a partir de la definición de Densidad de Potencia Espectral; es en éstos sitios donde los nuevos usuarios pueden introducirse.

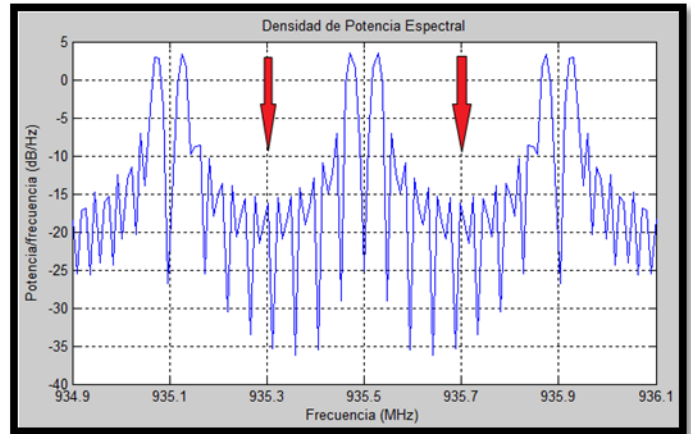


Fig. 7 Densidad de potencia espectral de usuarios primarios.

A continuación se introduce algún usuario secundario y si hay espacio en el medio se introduce en el canal desocupado, de otra forma se pide desocupar algún canal ó esperar a que esté desocupado. En la siguiente figura se muestra el primer canal que se asigna al nuevo usuario, aquí es la parte del algoritmo que interviene para sensar el espectro, se toma la decisión de cuál está disponible, en éste caso es el canal dos y se comparte ese canal del espectro. Se sigue sensando por si regresa un usuario licenciado, abandonar ese canal que se tomó “prestado”.

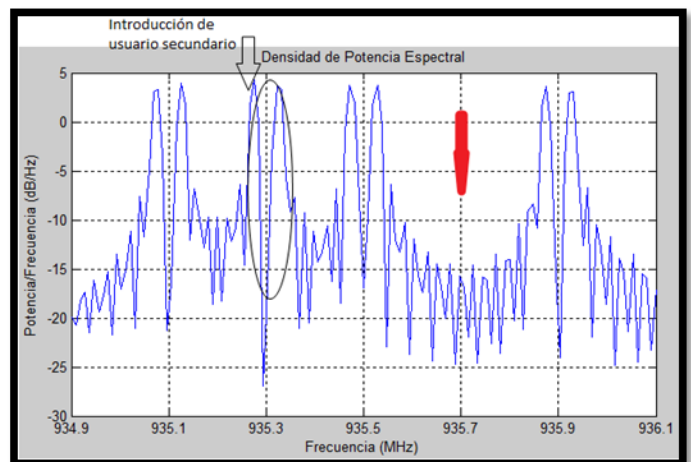


Fig. 8 Introducción de usuario secundario a espectro.

La definición de oportunidad determina las formas de medir y explotar el espacio espectral. La definición convencional de oportunidad espectral, la cual es comúnmente definida como “una banda

de frecuencias que no están siendo usadas por el usuario primario de esa banda en un tiempo particular, en un área geográfica” [11], solamente explota tres dimensiones del espacio espectral: frecuencia, tiempo y espacio. Así mismo, hay otras dimensiones que necesitan explotarse para oportunidad espectral como lo son la dimensión del código del espacio espectral.

Los principios de la radio cognitiva permiten al usuario hacer vacante algún canal para que en el momento deseado se introduzca al nuevo usuario secundario.

III. CONCLUSIONES

En éste artículo se presentan un diseño de sistema de radio cognitiva aplicado a un escenario GSM. Se implementa y se obtienen gráficas que muestran un escenario real donde usuarios esperan ser introducidos en la banda licenciada una vez ésta se desocupe. Se discuten los detalles de sensado del espectro. Se obtienen gráficas como resultado de sensado e introducción al espectro de un nuevo usuario.

El espectro es un recurso muy valioso en la actualidad, uno de los elementos más importantes de la radio cognitiva son las oportunidades disponibles del sensado del espectro.

Se trabajó con un rango de frecuencias que van desde [935-936] MHz con 5 canales de 200 kHz cada uno de acuerdo al estándar GSM, también se pudo haber ido a otros estándares y trabajar en esos rangos de frecuencias.

La radio cognitiva es un área de investigación ampliamente extensiva y abierta. Posibles tópicos de investigación futuros incluyen determinar el rango espectral más conveniente para la radio cognitiva, investigar la complejidad y confiabilidad de diferentes técnicas de sensado del espectro, determinar que algoritmos son los más eficientes y determinar el óptimo tamaño para redes cognitivas.

IV. REFERENCIAS

- [1] J. Mitola, "Cognitive radio for flexible mobile multimedia communications", in Proc. Mobile Multimedia Common. (MoMuC '99), Nov. 1999, pp. 3-10.
- [2] S. Mangold Z. Zhong, K. Challapali, and C.-T. Chou' "Spectrum agile radio: radio resource measurements for opportunistic spectrum usage", in Proc. IEEE Globazl Telecomm. Cou.f (GLOBECOM), vol. 6, Dallas, TX, Dec. 2004, pp. 3467- 3471.
- [3] S. Haykinl, "Cognitive radio Brain-empowered wireless communications", IEEE J. Select. Ameats Common., vol. 23, nlo. 2, pp. 201-220, Feb. 2005.
- [4] D. Cabric, S. M. Mishra, and R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios", in Pilroc. IEEE Asiloinar Conf on Signals, Syst., Computers, vol. 1, Pacific Grove, CA, Nov. 2004, pp. 772-776.
- [5] X. Jing and D. Raychaudhuri, "Spectrum co-existence of IEEE 802.11b band 802.16a networks using the CSCC etiquette protocol", in Proc. IEEE Int. Synmp. Dynamic Spectrum lAccess Networks (DySPAN) 2005, Baltimore, MD, Nov. 2005, pp. 243 -250 .
- [6] J. Hillenbrand, T. Weiss, and F. Jondral, "Calculation of detection and false alarm probabilities in spectrum pooling systems", IEEE Comm. Letters, vol. 9, no. 4, pp. 349-351, April 2005.
- [7] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio networks", in Proc. IEEE Int. Symp. Dynamic Spectrum Access Networks (DySPAN) 2005, Baltimore, MD, Nov 2005, pp. 137 143 .
- [8] J. O. Neel and J. H. Reed, "Performance of distributed dynamic frequency selection schemes for Interference reducing networks", in Military Communications Conference MILCOM, 2006.

- [9] T. Weingart, G. V. Yee, D. Sicker, and D. Grunwald, "A dynamic cognitive radio configuration algorithm", IEEE Communications, 2006.
- [10] C. Raman, R. Yates, and N. Mandayam, "Scheduling variable rate links via a spectrum server", in First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005.
- [11] P. Kolodzy et al., "Next generation communications: Kickoff meeting", in Proc. DARPA, Oct. 2001.
- [12] Weiss, T.A.; Jondral, F.K., "Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency", Communications Magazine, IEEE Volume 42, Issue 3, Mar 2004 Page(s): S8 – 14
- [13] Niels Hoven, Rahul Tandra, and Prof. Anant Sahai, Some Fundamental Limits on Cognitive Radio, Wireless Foundations, EECS University of California at Berkeley February 11, 2005
- [14] Cabric, D.; Mishra, S.M.; Brodersen, R.W, "Implementation issues in spectrum sensing for cognitive radios", Signals, Systems and Computers, 2004. Conference Record of the Thirty-Eighth Asilomar Conference on Volume 1, Issue , 7-10 Nov. 2004 Page(s): 772 - 776 Vol.1

V. AUTORES

Ing. Salvador Silva López obtuvo su título de Licenciatura en Ingeniería electrónica especialidad en sistemas digitales por la Universidad Autónoma Metropolitana Unidad Azcapotzalco y actualmente cursa la Maestría en Ciencias de la Ingeniería con especialidad en Comunicaciones Móviles en el ITESM Campus Estado de México.

PLANES DE PREVENCIÓN Y RECUPERACIÓN DE DESASTRES (PPRYD) EN REDES WLAN

F. Felipe, M. A. Acevedo, I Martínez S.

ESIME Zacatenco Academia de Computación Departamento Académico de ICE
Instituto Politécnico Nacional
Av. IPN S/N Unidad Profesional Adolfo López Mateos Col Lindavista, México, D.F. 07738, México
Tel. 57-29-60-00, ext 54755, correo electrónico: ffelipe100@hotmail.com

RESUMEN

Los desastres informáticos y de telecomunicaciones tienen costos muy altos para las organizaciones, si bien muchos de ellos son inevitables ante fenómenos naturales como temblores, inundaciones y otro tipo de eventos. El presente trabajo se centra en establecer Planes de Recuperación y Prevención de Desastres en redes WLAN

Palabras clave: Metodología, Prevención Recuperación, WLAN, Desastres, PPRyD, contingencia, Seguridad Informática, AP, antenas 802.11, Sistemas de Respaldo.

I. INTRODUCCIÓN

Las redes de área local inalámbricas (WLAN) [3]; y en general las tecnologías inalámbricas para las telecomunicaciones, tienen una expansión muy alta en la actualidad. Las posibilidades de movilidad casi ilimitada, el bajo costo de instalación y las cada vez más altas tasas de transferencia de datos las hacen muy atractivas para todo tipo de organizaciones. Si bien se presentan problemas, principalmente de seguridad y puntos ciegos a la señal, esto no ha limitado su continuo crecimiento.

La cultura de prevención y recuperación de desastres en México está muy poco extendida. La mayoría de las organizaciones ni siquiera realiza tareas mínimas de respaldo de su información que les permitan salvaguardar la misma en caso de cualquier tipo de incidente sea pequeño o grande. Los documentos básicos para establecer un PPRyD como tener un inventario completo y actualizado de los equipos de cómputo y telecomunicaciones, planos de distribución física de tales recursos y catálogos de perfiles de puestos del personal

encargado de tales no existen en las organizaciones y tampoco se tiene contemplado tenerlos en orden. Las tareas siguientes como establecer las fuentes de riesgo, diagramas de Fortalezas/Debilidades (SWOT en inglés) y concluir en sí un PPRyD completo no son considerados en ningún momento.



Fig. 1 Inundación en la ciudad de Nueva Orleans por el Huracán Katrina

Las pérdidas que las empresas pueden tener, por no tener contempladas estas situaciones, son cuantiosas y pueden llegar a millones de dólares y a paralizar empresas o grandes áreas de compañías. Para México es difícil obtener datos precisos debido a la misma falta de cultura anteriormente mencionada.

Algunos ejemplos son los siguientes: El quince de octubre del 2006 ocurrió un temblor de magnitud 6.7 grados en escala de Richter en Hawaii. El tráfico aéreo fue suspendido, también en la energía eléctrica. Por precaución se solicitó que el agente permaneciera en sus casas. Debido a la sobre carga de llamadas en telefonía celular por habitantes que

deseaban enterarse del estado de sus familiares y amigos, a la interrupción de la energía eléctrica que alimentaba Modems, ruteadores y switches el sistema quedo inhabilitado. EL impacto no fue tan grande debido a que el temblor sucedió en domingo, si hubiera ocurrido en un día hábil los daños económicos hubieran sido mucho mayores. El huracán Katrina provocó grandes daño en Nueva Orleans que afectaron todos los sistemas de la ciudad incluyendo los de comunicaciones e informáticos, Fig 1.

II. MÉTODO

La Para la elaboración e implementación de un PPRyD, de cualquier tipo se acostumbra seguir una metodología consistente en los siguientes pasos [7].

1. Levantamiento de datos: Se deben incluir Equipos de Cómputo, de Telecomunicaciones, Infraestructura Física, Personal, Seguridad Física e Informática, Sistemas de Respaldo Eléctrico e Informático.
2. Diagnóstico de riesgos y seguridad. Identificación de todas las posibles fuentes de Desastres: Incendio, Terremotos, Inundaciones, Deslizamientos de tierra, Derrumbes, Intrusiones Físicas e Informáticas, Fallas Eléctricas, Determinación de Vida Útil de Equipos incluyendo Software.
3. Análisis de Impacto: Cuantificar el costo monetario que tendrá la organización en caso de presentarse un Desastre con sus diversos escenarios, pérdida total, parcial. También se debe considerar las interrupciones que se tengan en la operación de la misma.
4. Guía Prototipo: Los responsables del Proyecto deben realizar un Prototipo del PPRyD que sirva de base para elaborar el documento final. Debe presentarse a todos los posibles actores de las contingencias, como documento de trabajo, para que participen activamente en su elaboración ya que ellos serán los

que tendrían que ponerlo en práctica cuando fuera necesario.

5. Desarrollo del PPR y D: Con la ayuda de la Guía Prototipo se elabora una versión dura del Plan. Como ya se comentó la participación de los actores, auxiliados por especialistas, es importantísima, ellos conocen mejor que nadie su entorno de trabajo, generalmente han detectado e incluso sufrido por pequeñas contingencias y finalmente serán quienes operen el PRR y D en su momento.
6. Definir la declaración de desastre: El momento en que se deben iniciar las actividades del PRR y D es crítico.
7. Establecer Responsables: Esta tarea es de vital importancia. Si bien todos los actores que participan en una organización deben saber que acciones tomar durante una contingencia, debe existir uno o varios responsables encargados de la coordinación de las acciones y que sobre tomen en su momento las decisiones críticas necesarias.
8. Pruebas y mejoras: Una vez que se tiene establecido el Plan, se debe probar y hacer los simulacros necesarios para verificar su efectividad y sobre todo ir familiarizando a los usuarios finales del Plan sobre el mismo.
9. Iniciar Plan de Prevención. El ideal de un PPyRD es nunca llegar a la contingencia, esto es que la parte dedicada a la prevención evitará que el desastre suceda.

El ideal de un PPyRD sería que la prevención evitará que se presente un desastre de cualquier tipo. Sin embargo esto es imposible. Existen desastres como los terremotos, incendios, derrumbes y huracanes, incluso ataques terroristas que si afectan seriamente la infraestructura física es casi seguro que afecten los sistemas informáticos y de telecomunicaciones y posiblemente los destruyan en buena parte. En el caso de los sistemas informáticos y las redes WLAN existen desastres, que aunque no lleguen a afectar físicamente a sus elementos pueden destruir casi en

su totalidad la información del sistema. Es en estas indeseables situaciones en donde cobra importancia la parte del plan que involucra la recuperación de las funciones de los sistemas.

Es muy importante remarcar que si bien las actividades de prevención deben ser permanentes la declaración de contingencia es muy importante. La declaración oportuna permite que los daños sean menores y un retraso en el inicio de las actividades marcadas en el PPyRD puede llevar incluso a la destrucción total de los sistemas.

III. LAS REDES WLAN

Una WLAN [6] está compuesta por un servidor, uno o mas elementos de transmisión/Recepción como Puntos de Acceso o Antenas para 802.11 y un conjunto de usuarios de la red que cuentan con una computadora con tarjeta de acceso inalámbrico o bien un dispositivo móvil como una PDA, un iPhone que cuente con la capacidad de conectarse bajo el estándar 802.11 [3]. Por lo general y como sucede en todos los sistemas inalámbricos el servidor es la parte fija de la WLAN y la que tiene acceso vía cable a sistemas de telecomunicación más robustos y que brindan enlace hacia el exterior.

Servidor. Es muy frecuente tener un solo servidor en una WLAN encargado de los servicios de Internet, de correo, de impresión o de otro tipo, pero nada impide que se tengan más con funciones distribuidas entre todos ellos. La naturaleza y necesidades de la WLAN dictarán la cantidad y función de cada servidor de la misma.

Puntos de Acceso. Los Puntos de Acceso fueron los primeros dispositivos de Transmisión/Recepción que permitieron crear redes inalámbricas. El estándar marca que en espacio cerrados la señal alcanza alrededor de los cien metros, pero en espacios cerrados, se ha determinado que su alcance real y con buena efectividad es de alrededor de los 25 metros. Uno de los problemas típicos de los Puntos de Acceso son los puntos ciegos debido a los materiales de

construcción donde se instala la WLAN. En esos casos la solución va desde instalar más AP's, buscar lugares con líneas de vista para entre todos los usuarios de la red y el AP hasta instalar antenas específicas para la WLAN.



Fig. 2 Imagen de un Access Point (AP)

Antenas 802.11. Las antenas [9] nacieron de la necesidad de superar varios de los problemas de los AP's, especialmente su alcance en lugares amplios de una organización. Por ese motivo las Universidades, Empresas Oficinas de Gobierno tienden a usar antenas en lugar de varios AP's tanto por el costo como por la facilidad de instalación. Esto es que para dar mayor alcance se deben instalar varios AP's incrementado costo y tareas de instalación, mientras que con una sola antena se tendrían los mismos resultados. Incluso existen antenas que compiten en precio con los AP's. El alcance de las antenas está ligado a la potencia que es construida, en la práctica los autores han probado antenas que cubre los cien metros con buena eficiencia.



Fig. 3 Imagen de una antena para 802.11 en 5.4 Ghz

Computadoras con tarjeta inalámbrica. En un principio las computadoras debían instalar una tarjeta con capacidad de hacer el enlace al AP, en un tiempo breve la mayoría de las laptops incluyeron en su hardware la tarjeta de forma que se volvió una característica estándar de estos equipos. En la actualidad las computadoras de escritorio deben instalar una tarjeta para el acceso inalámbrico y las laptops ya la tienen incluida.

Dispositivos móviles. En los últimos cinco años, dispositivos netamente móviles como PDA's y iPhone, entre otros, tienen la capacidad de conectarse a las WLAN. En forma empírica se ha constatado que su alcance es similar a las tarjetas de las laptops.



Fig. 4 Modelo de Blackberry

IV. PPRD PARA REDES WLAN

Los servicios que proporciona una WLAN pueden incluir varios aspectos: Proporcionar servicio a Sistemas Informáticos, Servicios de INTERNET en varios aspectos como correo electrónico, consultas y operaciones en línea, servicios educativos en aulas, videoconferencias, Soporte para INTRANET's entre otros.

Si bien todos los componentes de la WLAN son importantes para su operación puede considerarse que los componentes indispensables para su funcionamiento son el medio de comunicación ya sean AP's y/o Antenas 802.11 y el o los servidores. En otra dimensión, que tiene que ver con la

energía, los sistemas de alimentación eléctrica son prioritarios para que la WLAN pueda operar. En el primer caso y pensando en una operación mínima se requiere al menos un servidor y al menos un AP o una antena para que puedan seguirse realizando las operaciones. Para el segundo caso se deben considerar los sistemas auxiliares de energía, desde simple No-breaks que su soporte es de minutos hasta un par de horas, Unidades de Energía de Emergencia conocidos como UPS que proporcionan energía por horas hasta las plantas de emergencia que pueden soportar de forma permanente los fallos de energía eléctrica.

Especial interés para la operación de cualquier red de computadoras es la seguridad informática en todas sus vertientes: virus informáticos, robo de identidades, denegación de servicios, correo indeseado o SPAM entre otras amenazas. Las herramientas para prevenir y remediar los problemas cuando se presentan estas situaciones son amplias y variadas. Últimamente se han desarrollado medios que pueden proteger el espacio que rodea una WLAN de forma que intrusos o espías no puedan afectar la red como las proporcionadas por la empresa Air Defense.

La seguridad física es un tema que por lo general se olvida, tal vez sea por haberse comprobado que la mayoría de los ataques y robos a redes de computadoras provienen del interior de la organización o por gente que recientemente dejó de pertenecer a ella. El control de acceso físico por medio de vigilancia, los sistemas de entrada automáticos por medio de tarjetas magnéticas o los sistemas de acceso biométrico son medios que permiten un mejor control de las personas que entran a la empresa y deben de considerarse en un Sistema de Prevención de Desastres.

Un tema olvidado en México es el catalogo de perfiles de puestos. Saber que habilidades y que preparación tiene cada uno de los actores en una contingencia o desastre es primordial para poder asignar las responsabilidades y tareas a realizar

Los sistemas y políticas de respaldo no son de uso frecuente en México. En el mundo es variable la

importancia que se les da. Los centros de cómputo, sobre todo los que manejan bases de datos deben contar con una política de respaldos establecida clara y que sea seguida por el personal que labora en ellos. También los usuarios de cualquier computadora, sea de uso personal o que esté conectada a una Red debe establecer su propia política.

Los elementos mínimos para establecerla son un medio de almacenamiento, que puede ser desde una memoria USB, CD's o DVD's, o discos duros externos o sistemas de respaldo especializados, ya sean en cinta o en disco duro. Una recomendación importante es que los respaldos deberían estar en una locación física diferente a donde se tiene la computadora o servidor, porque en caso de un desastre mayor que involucre destrucción de las instalaciones físicas los respaldos también serían dañados.

	Desastres previstos				Herramientas para enfrentarlo	Nivel de preparación
	Sistema Evaluado	Tipo o Prioridad	Causa	Efectos o pérdidas		
1						
2						
3						
4						
5						

Tabla 1. Previsión de Desastres

Una opción reciente en México son los DATA CENTERS [10], que son centros especializados, con tecnología de punta, de manejo de información y entre los servicios que brindan está el de respaldo de información, con una opción interesante puede hacerse de vía remota sin necesidad de llevar la información respaldada de manera personal.

El censo o levantamiento de riesgos es crucial. Algunos no son factibles de presentarse en alguna situaciones, por ejemplo temblores en zonas no sísmicas, deslaves un zonas planas entre otros. En cambio las descargas eléctricas, los incendios, la

seguridad informática no puede faltar en el censo de riesgos.

Principales Desastres	Pérdidas Posibles	Sistemas de Prevención	Disponibilidad	Recursos Necesarios Para Enfrentarlos
Fuego en edificios				
Afectaciones por agua				
Fallas de energía eléctrica				
Daños por Descargas eléctricas				
Fallas en servidores				
Fallas en AP o Antenas				
Seguridad Informática				
Personal con perfil inadecuado				
Mal ambiente laboral				

Tabla 2. Clasificación de Riesgos.

Una WLAN presenta aspectos interesantes para la recuperación de un desastre en telecomunicaciones, si se usan dos o más antenas o AP's.

Estos dispositivos permiten varias configuraciones y pueden montarse redes redundantes que permiten más de un circuito de transmisión colocando los dispositivos en diferentes ubicaciones físicas y por lo tanto pueden servir como red alternativa en caso de desastre o destrucción total de algunos de los dispositivos de transmisión.

Las Tablas 1 y 2 presentan formatos que permiten registrar información necesaria para diagnosticar y preparar la Prevención y la Recuperación del Desastre [4] [2]. El primer formato evalúa las capacidades de una organización para enfrentar una contingencia o desastre en los sistemas que son de su interés:

Actividad	E n	F e	M a	A b	M a	Ju	Ju	Ag
Inicio								
Levantamiento de Datos								
Análisis de riesgos y seguridad								
Prototipo o guía inicial								
Desarrollo de módulo de prevención								
Desarrollo del módulo de Recuperación								
Definición de Situaciones de contingencia y/o desastre								
Definición de Autoridad y Responsabilidad								
Pruebas y Mejoras								
Inicio del Plan								

Fig. 5 Calendario de Actividades

La organización puede establecer las calificaciones para los niveles de pérdidas, autocalificar las herramientas con que cuenta y el nivel de preparación para enfrentarlos. Los especialistas sugieren calificar de 1 a 5, por ejemplo si se considera que existen excelentes herramientas con 5 y 1 para la ausencia o mala calidad de las mismas.

El siguiente formato, en la Tabla 2, es para evaluar a la empresa en su preparación y recursos disponibles para enfrentar la contingencia. Ya se encuentra particularizada para las WLAN, pero nada impide que el usuario o diseñador del Plan incorpore los apartados que considere necesarios:

Cada organización debe particularizar el tipo de riesgo que puede enfrentar y de esa forma mejorar el formato de las dos tablas anteriores

La figura 3 presenta una recomendación de calendario de actividades a seguir, nuevamente, cada organización puede y debe ajustar de acuerdo a su tamaño, necesidades, misión y objetivos de tal calendario.

Otro formato que se debe desarrollar es el diagrama SWOT, por sus siglas en inglés, que consiste en enumerar las fortalezas, debilidades riesgos y oportunidades del medio en que se desenvuelve la organización. Existen varias versiones, todas muy similares, y bastante extendidas por lo que no se colocarán aquí

V. CONCLUSIONES

El presente trabajo presentó un Plan de Prevención y Recuperación de Desastres tanto informáticos como de telecomunicaciones para redes WLAN. Se partió de metodologías generales para cualquier tipo de organización y cualquier tipo de sistema para particularizar en los elementos de la WLAN y considerar sus características y problemáticas específicas, de forma que los responsables de las mismas cuenten con herramientas propias de los problemas a los que se pueden enfrentar. En la ESIME Zacatenco del IPN se han iniciado trabajos básicos para elaborar un Plan de este tipo, pero la inercia y falta de cultura informática han impedido que se tengan resultados tangibles.

Una aportación que se considera fundamental es incorporara diferentes puntos de vista para tener una visión más amplia de la situación. También se considera muy valiosa la recomendación de involucrar en la elaboración y operación del plan a los usuarios de la WLAN y en general a todas las personas de la organización ya que todas serán actores, de alguna forma, tanto en la prevención como en durante la contingencia del plan. Aprovechar su conocimiento del medio y hacerlos corresponsables a mostrado un gran éxito en otros proyectos como los Círculos de Mejora Continua y ese éxito se recomienda aprovecharlo en estos Planes.

Un punto muy descuidado y que también se sugiere reforzar es la elaboración de sistemas de respaldo informático. La información es en muchas organizaciones un valor muy alto y no considerado, que debe salvaguardarse [5]. Existen muchos ejemplos de pérdidas económicas debido a

no conservara adecuadamente los archivos de las organizaciones...

VI. RECOMENDACIONES

Existe en México un retraso en varias actividades que si se llevaran a cabo sistemáticamente simplificarían las tareas tener “vivo” un Plan de Prevención y Recuperación de Desastres Informáticos y de Telecomunicaciones. Tareas como tener un inventario actualizado de los recursos de cómputo y telecomunicaciones, tener perfiles de puestos técnicos acordes con las funciones que realizan los recursos humanos, elaborar planes de mediano y largo plazo tanto para mantenimiento y actualización de los mismos recursos no se acostumbra. Por esta última razón los administradores de redes de computadoras, incluidos los de WLAN, investigan poco sobre tecnologías recientes de seguridad informática y de monitoreo. Cubrir de forma continua estas tareas simplificará la posterior elaboración los respectivos planes de prevención y recuperación. Para finalizar también se recomienda que las organizaciones analicen el clima organizacional existente

VII. REFERENCIAS

- [1] Currión P., Silva Ch., Walle B. “*Open Source software for disaster management*”, Communications of the ACM, March 2007, Volume 50, number 3.
- [2] P Hayes, A. Hammons, “*Picking up the pieces*”, JIEEE Industry Applications Magazine, Dec 2002.
- [3] R. Olexa “*Implementing 802.11, 802.16 and 802.20 Wireless networks*” Elsevier 2005, United States of America.
- [4] R. Oelxa “*Implementing 802.11, 802.16 and 802.20 Wireless networks*” Elsevier 2005, United States of America.
- [5] G. E. Mahdy “*Disaster management in telecommunications, Broadcasting and*

Computer Systems”, Wiley 2001, United States of America.

- [6] R. Rohde, R. Haskett “*Disaster Recovery Planning for academic computing Systems*” Communications of the ACM, June 2009, Volume 33, number 6.
- [7] www.ieee.org/standards
- [8] www.recovery-disaster.net
- [9] www.squidoo.com/telecomdisasterprep
- [10] www.trendnet.com
- [11] www.triara.com

VIII. AUTORES

Federico Felipe Durán. Recibió su título de Licenciatura en Ingeniería en Comunicaciones y Electrónica en la Escuela Superior de Ingeniería Mecánica y Eléctrica del IPN. Actualmente es profesor titular en la ESIME Zacatenco del IPN. Realizó estudios de Maestría en Ciencias con especialidad en Ingeniería Eléctrica en el CINVESTAV del IPN México. Sus intereses en la investigación incluyen las redes de computadoras, la ingeniería de sistemas aplicada a las telecomunicaciones y la Inteligencia artificial.

Marco Antonio Acevedo Mosqueda. Realizó sus estudios de licenciatura en Ingeniería en Comunicaciones y Electrónica en la Escuela Superior de Ingeniería Mecánica y Eléctrica del IPN. Actualmente es profesor titular en la ESIME Zacatenco. Obtuvo su grado de M. en C. con especialidad en Electrónica en la ESIME Zacatenco del IPN. Sus principales áreas de investigación son el Procesamiento Digital de Señales, el diseño y construcción de antenas y los Sistemas de Telecomunicaciones

Ignacio Martínez Sánchez. Realizó sus estudios de licenciatura en Ingeniería Mecánica en la Escuela Superior de Ingeniería Mecánica y Eléctrica del IPN. Actualmente es profesor titular en la ESIME Zacatenco. Obtuvo su grado de M. en

C. con especialidad en Ingeniería de Sistemas en la ESIME Zacatenco del IPN. Sus principales áreas de investigación son la ingeniería de Sistemas de Calidad y la elaboración de proyectos.

VCO'S DIFERENCIALES PARA REDES DE GENERACIÓN Y DISTRIBUCIÓN DE SEÑALES DE RELOJ

Carlos Ramón Báez Álvarez, Oscar González Díaz, Mónico Linares Aranda

Instituto Nacional de Astrofísica, Óptica y Electrónica, Departamento de Electrónica
Luís Enrique Erro No. 1, Sta. María Tonantzintla, P.O. Box 51, 72000, Puebla, México
mlinars@inaoep.mx, ogonzalez@inaoep.mx, ramon060287@prodigy.net.mx

RESUMEN

En este trabajo se presenta el diseño y caracterización de osciladores controlados por voltaje (VCO's) del tipo diferencial de una, dos y tres etapas de retardo, para la generación y distribución simultánea de señales de reloj en Sistemas Integrados en un solo Chip (SoC). Los diferentes osciladores son diseñados utilizando parámetros de la tecnología CMOS de 0.35 μm de la fundidora AustriaMicrosystems, simulados mediante Hspice en el ambiente de Mentor Graphics, y caracterizados considerando como figuras de mérito: frecuencia de oscilación, rango de entonado, consumo de potencia y excursión de voltaje de salida.

Palabras clave: VCO's, Osciladores, Circuitos Integrados.

I. INTRODUCCIÓN

Actualmente, los circuitos integrados (CI's) o chips de diferentes sistemas electrónicos contienen millones de diminutos dispositivos semiconductores (transistores) que ocupan un área de varios milímetros cuadrados y operan a frecuencias del orden de GigaHertz (10^9 Hz). El manejo masivo de información y su procesamiento en tiempo real, especialmente en aplicaciones de video y comunicaciones, exigen que estos sistemas sean cada vez más rápidos y consuman menor potencia.

Para obtener mayores frecuencias de operación y lograr una mayor densidad de integración en estos CI's (mayor número de dispositivos por

unidad de área), generalmente se utilizan las tecnologías de fabricación de circuitos integrados más avanzadas disponibles. Además, debido a la necesidad de realizar constantemente un mayor número de funciones mediante un mismo chip, es necesario aumentar el tamaño del chip y en consecuencia que crezcan las interconexiones globales, tales como: las redes que le proporcionan energía (V_{dd} y G_{nd}), las redes de comunicación (buses) y las redes de temporización y sincronización (distribución de reloj). En el presente trabajo estamos interesados en la resolución de los problemas que se presentan en las redes de generación y distribución de reloj.

La mayoría de los microprocesadores procesan la información de forma digital y realizan sus operaciones de forma digital síncrona, esto es, requieren de al menos una señal de reloj que active (temporice) y transfiera (sincronice) la información entre elementos apropiadamente para su correcto funcionamiento. Típicamente esta señal de reloj es generada y modulada fuera del chip, e introducida al mismo a través de una de sus terminales de entrada/salida. A continuación la señal de reloj es repartida por una red en forma global a todos los elementos de carga (generalmente registros o flip-flops) que la necesiten y que por lo general se encuentran en todo el chip. La frecuencia máxima de operación del procesador dependerá del sistema de reloj, por lo que la red de distribución de reloj junto con el generador,

constituyen el corazón de todos los sistemas síncronos.

Las redes de distribución de señales de reloj (RDSR) operan a la frecuencia más alta, manejan la mayor carga y ocupan el área más grande de un circuito integrado.

Por otra parte, el oscilador controlado por voltaje constituye el bloque básico de construcción de sintetizadores de frecuencia para transceptores de comunicación inalámbrica y recientemente en sistemas de sincronización para microprocesadores. Particularmente, en sincronización de sistemas de muy alta velocidad y gran tamaño, los osciladores controlados por voltaje/corriente (VCO's/ICO's) en anillo no resonantes, se están convirtiendo en el elemento esencial para generar y distribuir simultáneamente las señales de reloj necesarias para sincronizar y temporizar sistemas digitales.

Con el fin de aprovechar los diseños de VCO's/ICO's, en el presente trabajo se diseñan y caracterizan diferentes VCO's de una, dos y 3 etapas de retardo, con el fin de aplicarlos a redes de generación y distribución de reloj para sistemas en un solo chip (SoC).

El presente artículo es organizado de la siguiente manera: En la sección II se define someramente lo que es un oscilador, particularizando en el oscilador en anillo no resonante y se presentan los diferentes VCO's considerados para su caracterización. En la sección III se presentan los resultados de simulación y la comparación de los mismos. Finalmente, en la sección IV se derivan las respectivas conclusiones.

II. OSCILADORES CONTROLADOS POR VOLTAJE EN ANILLO.

Un oscilador es un circuito capaz de producir una señal periódica (senoidal, cuadrada, triangular etc.) por sí solo, esto es sin la necesidad de una señal de entrada. Un oscilador en anillo es un oscilador constituido por celdas de retardo (etapas) conectadas en cascada y en

lazo cerrado, y cuya frecuencia de oscilación es dependiente del número de celdas o etapas de retardo N y del retardo t_{del} de cada una de las etapas de retardo que lo conforman. La ecuación que rige su comportamiento es [3]:

$$f_{osc} = \frac{1}{2Nt_{del}} \quad (1)$$

De acuerdo con esta ecuación, el incremento de frecuencia de oscilación puede lograrse mediante: a) la reducción del retardo de cada t_{del} , b) la reducción del número de etapas del anillo N y c) el uso de excursión pequeña (barrido) de la señal oscilante. No obstante, la disminución en el número de etapas resulta más atractiva no solo por el incremento en la velocidad de operación sino por la reducción del consumo de potencia y área requerida. Si el número de etapas N es reducido por un factor de dos, entonces el circuito puede oscilar al doble de la frecuencia debido a que el retardo total del lazo es reducido a la mitad, y por consiguiente, el oscilador disipará a lo mucho la mitad de la potencia. De esta forma para una frecuencia de oscilación dada, la disipación de potencia disminuye con el cuadrado del factor de reducción en el número de etapas. Esta tendencia hace atractivo considerar osciladores de una sola etapa; sin embargo, a medida que N disminuye, es más difícil satisfacer simultáneamente los criterios de oscilación de Barkhausen dados en la ecuación 2 [3]:

$$\angle H(jw_0) = 180^0 \quad \text{Condición de fase} \quad (2a)$$

$$|H(jw_0)| \geq 1 \quad \text{Condición de ganancia} \quad (2b)$$

En tecnología CMOS existen osciladores de diferente número de etapas de retardo, los cuales a su vez pueden ser de tipo diferencial y de terminación simple (Fig. 1). En el presente trabajo nos enfocamos en la configuración diferencial debido a las múltiples ventajas que presenta [3].

En base a artículos reportados en la literatura [1-16], previamente se analizaron

cuantitativamente 16 VCO's, de los cuales 8 fueron seleccionados, por presentar mejores características, para ser caracterizados

cuantitativamente en el presente trabajo. Los osciladores considerados se muestran en las figuras 2 a 9.

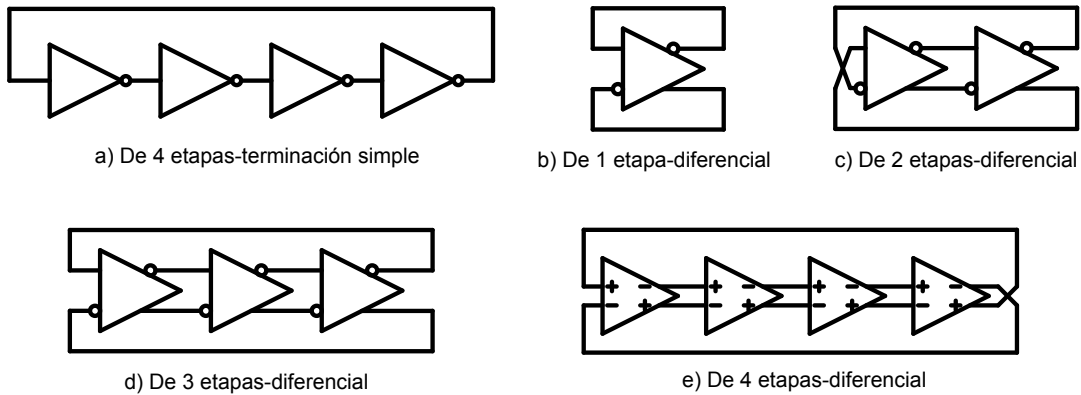


Fig. 1 Osciladores controlados por voltaje en anillo

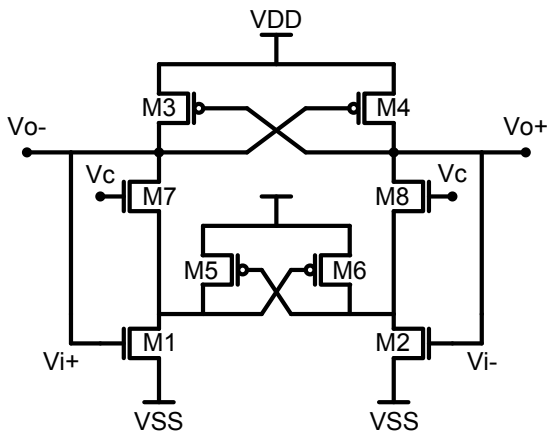


Fig. 2 VCO de una sola etapa de Ahmed [1]

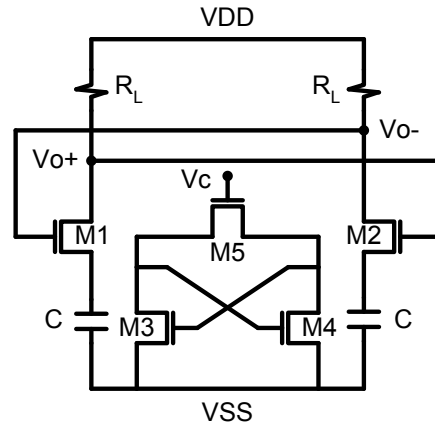
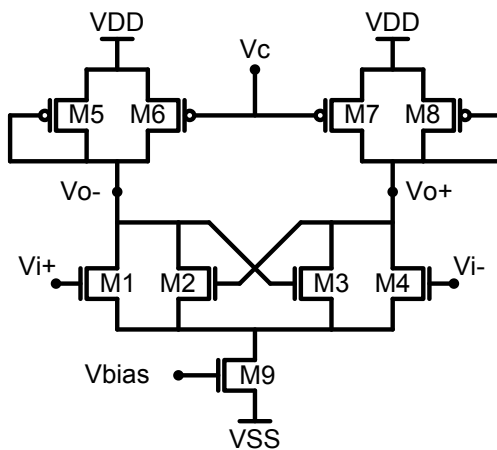
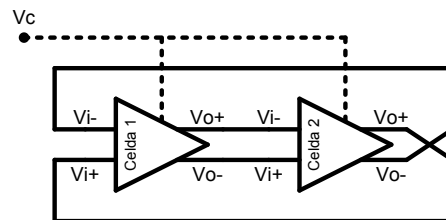


Fig. 3 VCO de una sola etapa de Mostafa [2]



(a) Celda de retardo



(b) Oscilador

Fig. 4 VCO de dos etapas de Daniel [3]

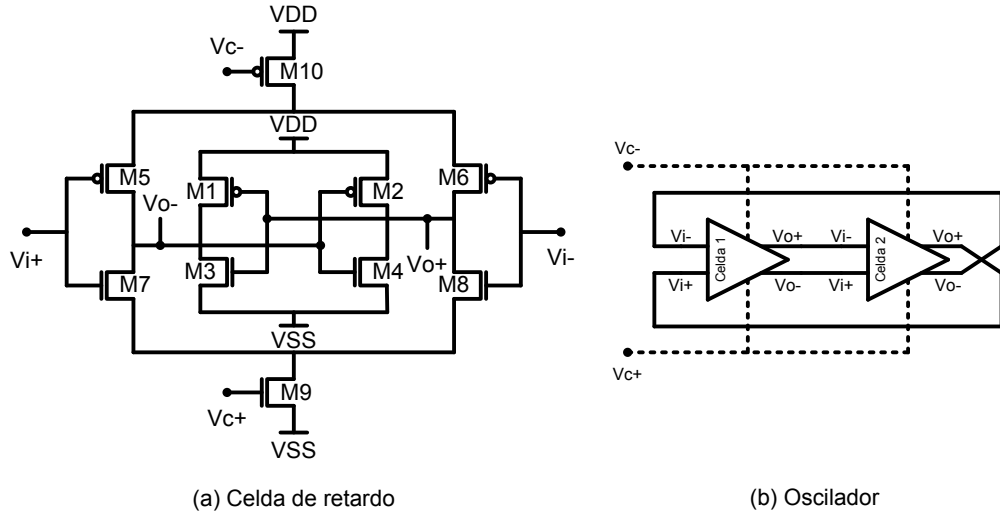


Fig. 5 VCO de dos etapas de Liang [4]

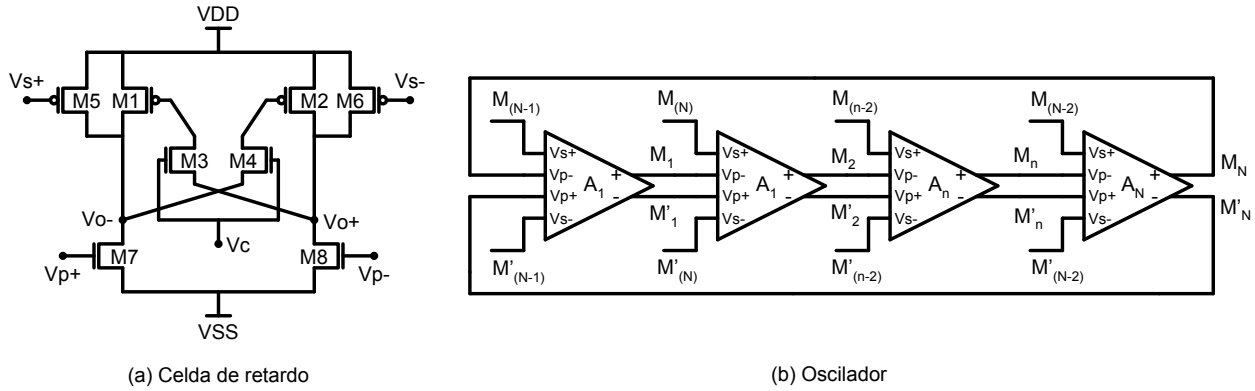


Fig. 6 VCO de cuatro etapas de Yalcin [5]

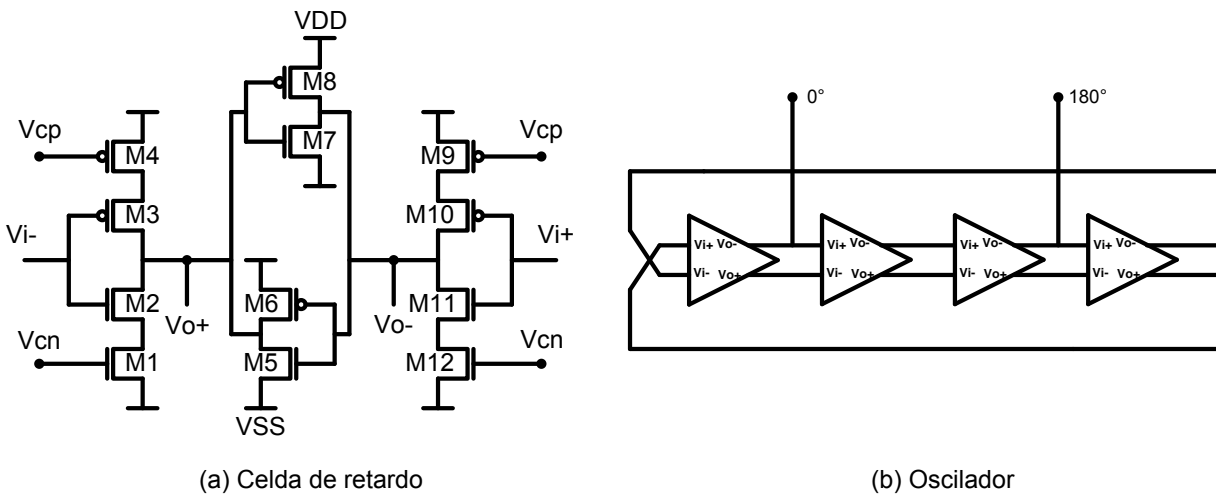


Fig. 7 VCO de cuatro etapas de Joeres [6]

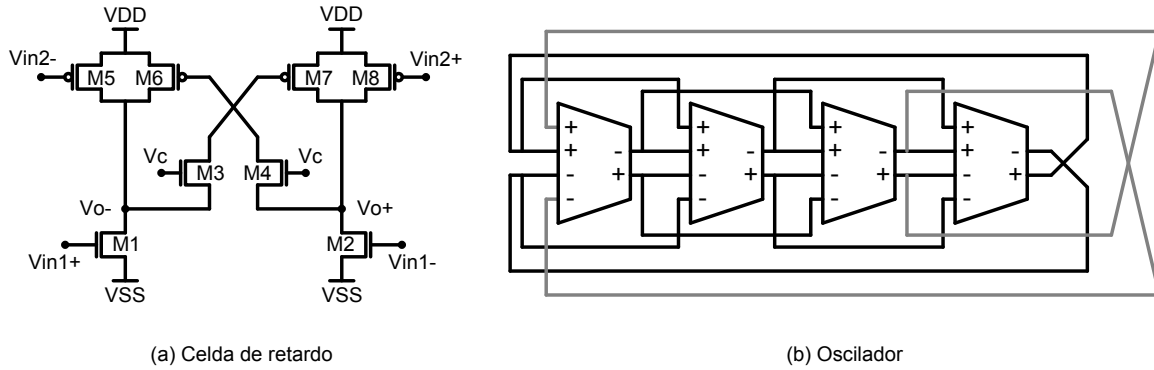


Fig. 8 VCO de cuatro etapas con doble lazo de retardo de Park [7]

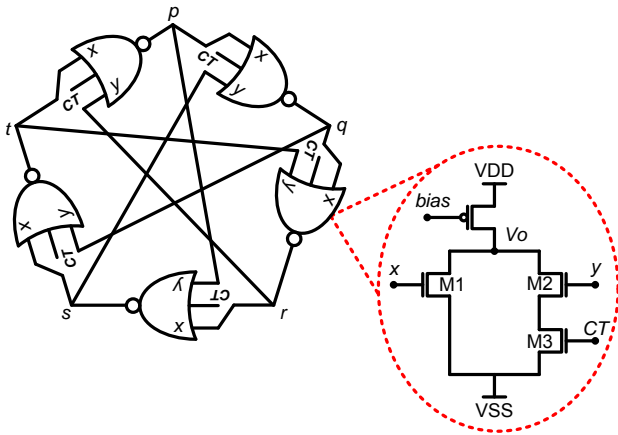


Fig. 9 VCO de 5 etapas con técnica de interpolación de Fadi [8]

III. RESULTADOS Y COMPARACIÓN

Cada oscilador se diseñó para un óptimo desempeño utilizando una tecnología CMOS de 0.35 μ m. Los osciladores fueron probados para dos tipos de carga: un flip-flop diferencial y un flip-flop dinámico (Fig. 10). Esto con el fin de verificar su funcionamiento para aplicaciones tanto analógicas como digitales.

Los resultados de simulación Hspice se muestran en la Tabla 1. Las funciones de transferencia de los osciladores se presentan en la Figura 11, mientras que el consumo de los diferentes osciladores es dado en la figura 12. En la figura 13 se muestran los patrones geométricos (layout) de tres de los 8 osciladores diseñados.

A partir de la tabla I y las gráficas de las figuras 11 y 12 se puede observar que las más altas frecuencias de oscilación las presentan los VCO's de una etapa, sin embargo su amplitud de voltaje a altas frecuencias es reducida comparada con la de los demás. El oscilador de Yalcin aún cuando consta de tres etapas, constituye el mejor VCO desde el punto de vista de consumo de potencia, ya que solo consume un 25% comparado con Ahmed y 37.3% respecto de Mostafa. El VCO con mayor rango de entonado lo constituye el de Daniel (1.81 GHz).

Considerando las formas de onda, de la figura 14 se observa que la señal de salida del VCO de Ahmed es similar a una señal triangular. Los osciladores de Mostafa y de Yalcin generan una señal más senoidal y de mayor frecuencia de oscilación. El oscilador de Yalcin produce una onda de mayor cuadratura debido al mayor barrido de voltaje (hasta 3 volts de acuerdo con la tabla I).

Concluimos que a pesar de las diferencias, los VCO's de Yalcin y de Mostafa son los que presentan el mejor desempeño.

IV. CONCLUSIONES

En este trabajo se diseñaron, simularon y analizaron cuantitativamente ocho VCO's de anillo de diferente número de etapas. De acuerdo con resultados de simulación, los diferentes

VCO's mostraron capacidad para operar a alta frecuencia de hasta cerca de 3.7 GHz.

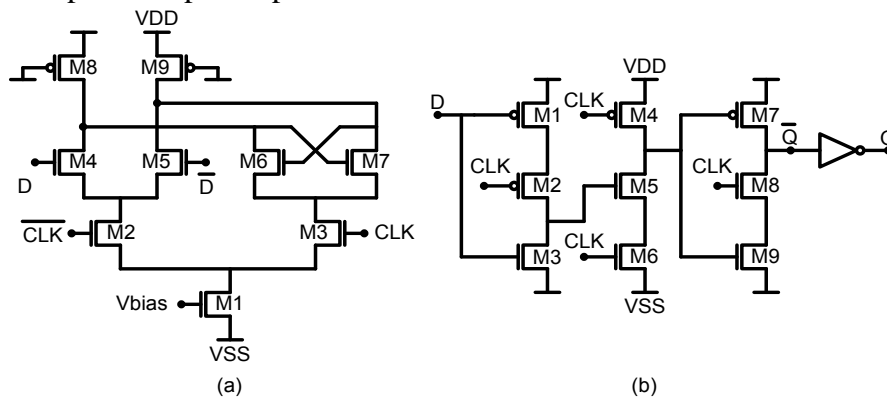


Fig. 10 Flip-flops utilizados como carga: a) Estático diferencial, b) Dinámico de terminación sencilla

TABLA 1. RESULTADOS DE SIMULACIÓN HSPICE

Parámetro	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Frecuencia de operación (GHz)	1.8 – 3.2	2.2 – 3.7	0.19 – 2	1.3–1.9	2.6-1.4	0 – 0.96	0.94 – 0.8	1.2 – 2.4
Rango de entonado (GHz)	1.4	1.5	1.81	0.6	1.2	0.96	0.14	1.2
Consumo Potencia (mW)	7.5 – 8.5	13.4–12	1 – 14	9.2 - 12.6	2-3	0 - 5	12 - 13	6.8 – 8
Barrido de voltaje (V)	1.1 – 0.3	1.5–0.75	0.7–0.75	2.3 – 1.9	2.3	2.1 – 2.6	3.17 - 3	2.3 - 1.4
No. de celdas	1	1	2	2	3	4	4	5

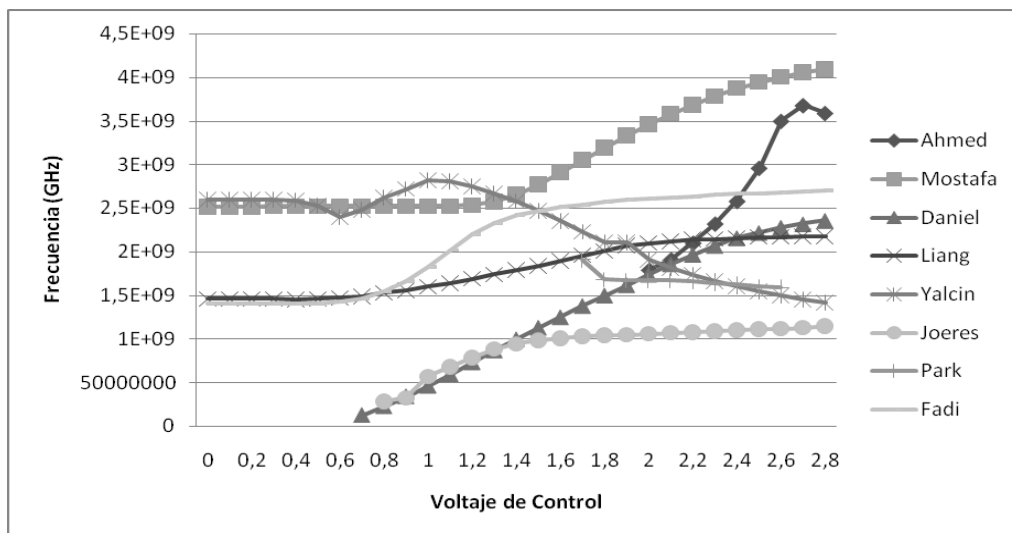


Fig. 11 Funciones de transferencia de osciladores, simuladas utilizando tecnología CMOS 0.35µm.

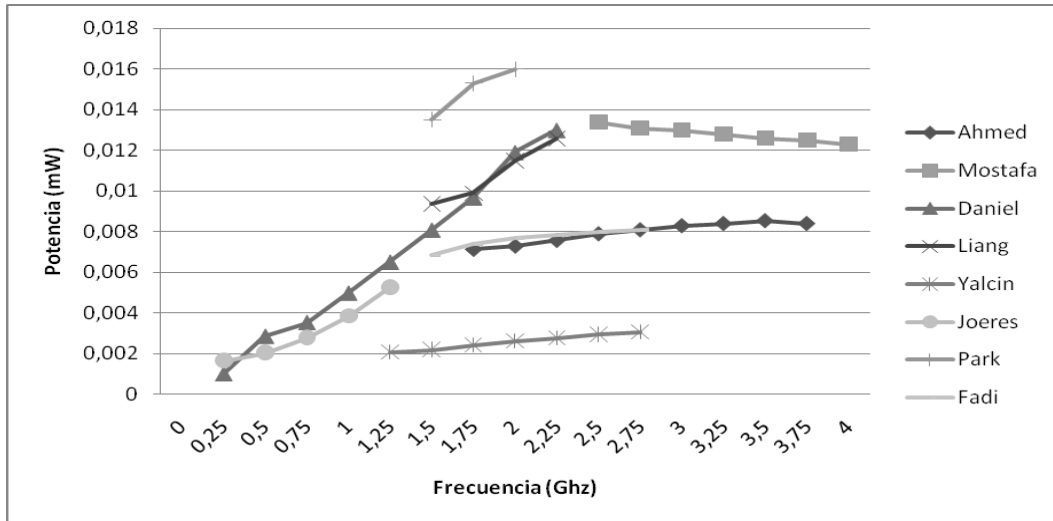
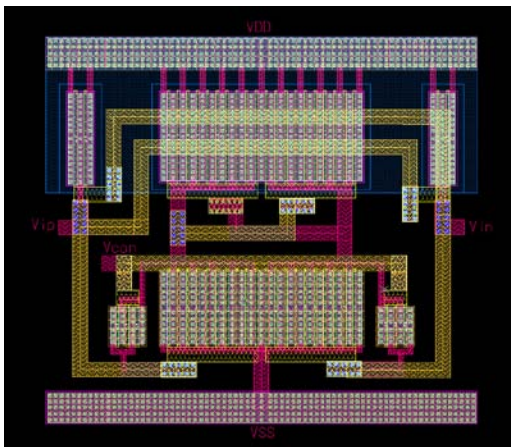
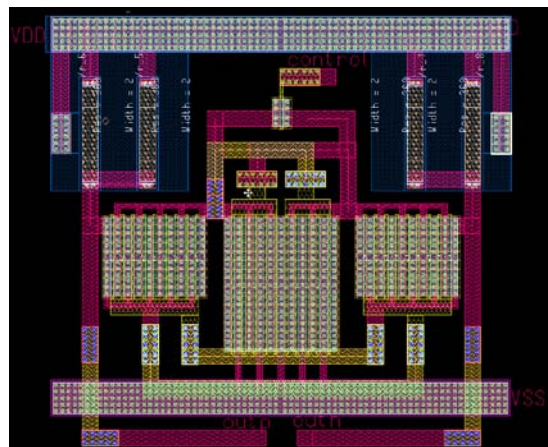


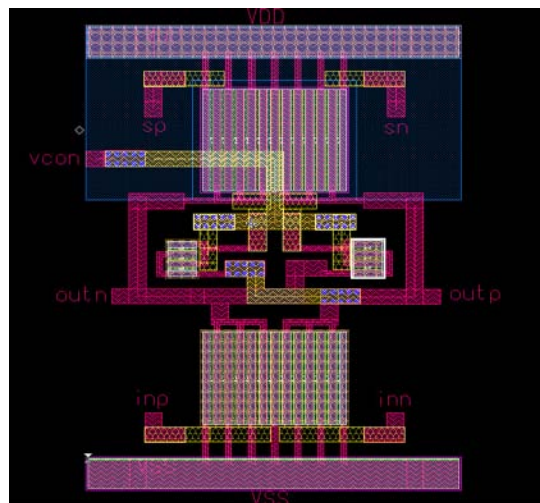
Fig. 12 Consumo de potencia de los VCO's simulados en tecnología CMOS 0.35um.



a) Oscilador de Ahmed [1]

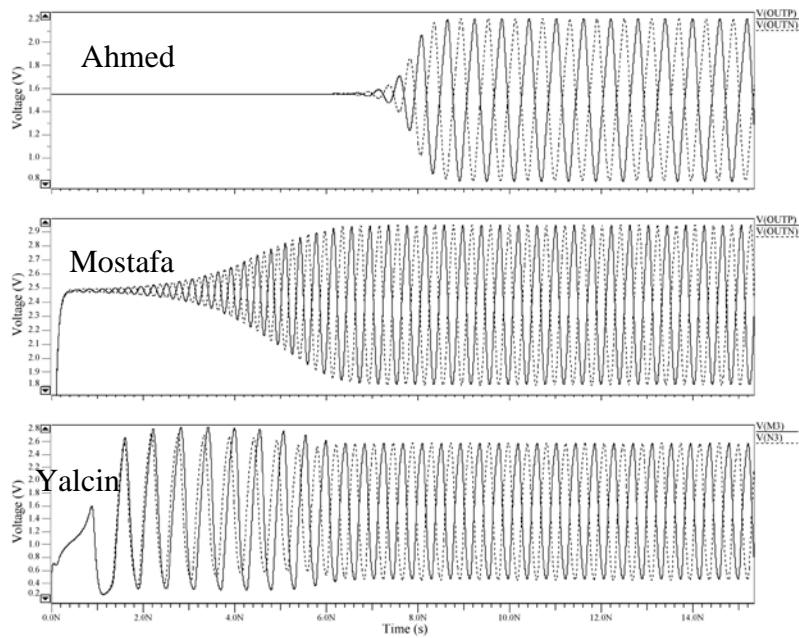


b) Oscilador de Mostafa [2]

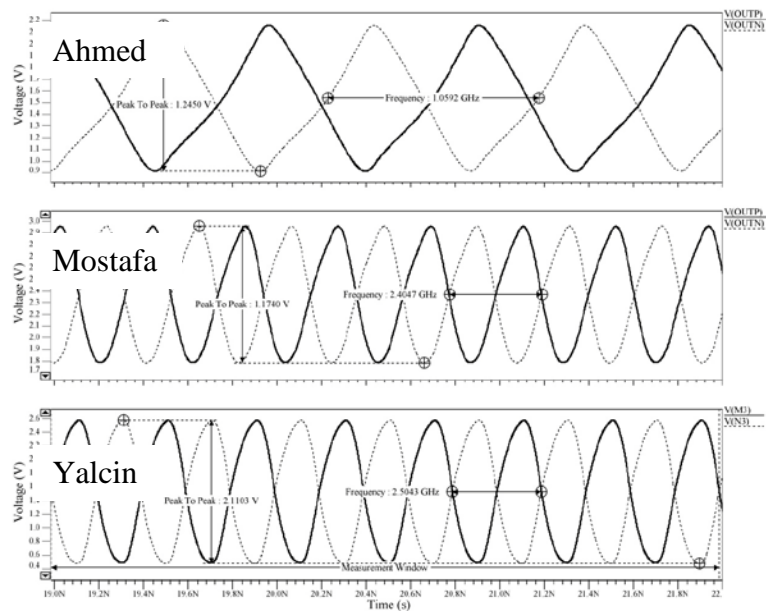


c) Oscilador de Yalcin [5]

Fig. 13 Layouts de osciladores controlados por voltaje en anillo



a)



b)

Fig. 14 Formas de onda de salida de tres VCO's: a) Mostrando inicio de oscilación, b) acercamiento

Considerando la frecuencia de operación, rango de entonado y consumo de potencia, el VCO de una sola etapa de Mostafa y el VCO de tres etapas de Yalcin resultaron ser los más atractivos para su utilización en sistemas de alta velocidad y bajo consumo de potencia, lo cual indica que no

necesariamente los VCO's con un mínimo número de etapas son los mejores.

Agradecimientos:

Al Conacyt-México por el apoyo parcial a través del proyecto 51511-Y y la beca de doctorado 166074.

V. REFERENCIAS

- [1] A. Ahmed, K. Sharaf, H. Haddara, H.F. Ragai, "CMOS VCO-prescaler cell-based design for PLL frequency synthesizer", IEEE International Symposium on Circuits and Systems, Vol. II, pp. 737-739. May 2000.
- [2] Mostafa Savadi Oskooei, Ali Afzali-Kusha, S. M. Atarodi, "A High-Speed and Low-Power Voltage Controlled Oscillator in 0.18- μm CMOS Process". IEEE International Symposium on Circuits and Systems, pp. 933-936, May 2007.
- [3] Daniel Pacheco Bautista, Mónico Linares Aranda, "A Low Power and High Speed CMOS Voltage-Controlled Ring Oscillator", IEEE International Symposium on Circuits and Systems, pp.752-755, 2004.
- [4] Liang Dai, Ramesh Harjani, *Design of high performance cmos voltage controlled oscillators*. Kluwer Academic Pub. 2003.
- [5] Yalcin Alper Eken, John P. Uyemura, "A 5.9-GHz Voltage-Controlled Ring Oscillator in 0.18- μm CMOS", IEEE JSSC, Vol. 39, No. 1, pp. 230-233, January 2004.
- [6] S. Joeres, A. Kruth, O. Meike, G. Ordu, S. Sappok, R. Wunderlich and S. Heinen, *Design of a Ring-Oscillator with a Wide Tuning Range in 0.13 μm CMOS for the use in Global Navigation Satellite Systems*.
- [7] Chan-Hong Park, Beomsup Kim, "A Low-Noise, 900-MHz VCO in 0.6- μm CMOS". IEEE JSSC, Vol. 34, No. 5, pp. 586-591, May. 1999.
- [8] F. H. Gebara, J. D. Schaub, A. J. Drake, K. J. Nowka, R. B. Brown, "4.0GHz 0.18 μm CMOS PLL Based on an Interpolative Oscillator", Symposium on VLSI Circuits Volume, June 2005, pp. 100 – 103.
- [9] Wei-Hsuan Tu, Jyh-Yih Yeh, Hung-Chieh Tsai, Chornng-Kuang Wang, "A 1.8V 2.5-5.2 GHz CMOS Dual-input Two-stage Ring VCO", EEE Asia-Pacific Conference on Advanced System Integrated Circuits. Aug. 4-5, 2004, pp. 134-137.
- [10] Luciano Severino de Paula, Eric Fabris, Sergio Bampi, Altamiro Amadeu Susin, "A High Swing Low Power CMOS Differential Voltage-Controlled Ring Oscillator". 14th IEEE International Conference on Electronics, Circuits and Systems, 2007. Pp. 498-501.
- [11] Xin Wu, Dunshan Yu, Shimin Sheng, "A Full Swing and Low Power Voltage-Controlled Ring Oscillator", IEEE Conference on Electron Devices and Solid-State Circuits, pp. 141- 143, Dec. 2005.
- [12] Vincent Cheynet de Beaupré, Wenceslas Rahajandraibe, Lakhdar Zaid, "2.45-GHz-CMOS temperature compensated multi-controlled oscillator for IEEE 802.15 wireless PAN", Analog Integr Circ Sig Process (2006) 49:257–266.
- [13] H. Q. Liu, W. L. Goh, L. Siek, "Design and frequency/phase-noise analysis of a 10-GHz CMOS ring oscillator with coarse and fine frequency tuning", Analog Integrated Circuit and Signal Processing (2006). 48:85–94.
- [14] Abel Vámos, "Quadrature-Phase Topology of a High Frequency Ring Oscillator", Design and Diagnostics of Electronic Circuits and Systems, Volume 11, Issue 13 April 2007 pp. 1- 4.
- [15] Chan-Hong Park, Beomsup Kim, "A Low-Noise, 900-MHz VCO in 0.6- μm CMOS", IEEE JSSC, Vol. 34, No. 5, May. 1999. pp. 586-591.
- [16] Ali Hajimiri and Thomas H. Lee. *The design of low noise oscillators*. Kluwer Academic publishers. Boston/ Dordrecht/ London, 2000.

SÍNTESIS DE ECUALIZADORES DE GANANCIA BASADOS EN REJILLAS DE PERIODO LARGO PARA EDFAS

Flores Llamas Ignacio, Beltrán Hernández Alfredo, Khotiaintsev Serguei

Departamento de Ingeniería en Telecomunicaciones, División de Ingeniería Eléctrica
Facultad de Ingeniería, Universidad Nacional Autónoma de México
C.U., Circuito Interior, Edificio Luis G. Valdés Vallejo 3er piso, México D.F., C.P. 04510
Tel. 5622 3074, 3062, 3055

RESUMEN

En este artículo se presenta la adaptación de un algoritmo genético (*GA*) para la tarea de la síntesis de una rejilla de periodo largo (*LPFG*). Esta rejilla está destinada a ecualizar la ganancia de un amplificador de fibra dopada con erbio (*EDFA*). La adaptación del *GA* descrita en este trabajo proporciona los parámetros de la rejilla dentro de rangos prácticos para su fabricación. Se presentan los resultados obtenidos y se discuten las ventajas ofrecidas por este método de síntesis.

Palabras clave: Amplificadores ópticos, amplificador de fibra dopada con erbio, algoritmo genético, ecualizador de ganancia, rejillas de periodo largo.

I. INTRODUCCIÓN

Hoy en día, debido a la necesidad de transmitir información a gran escala, el uso de amplificadores en los sistemas de comunicaciones ópticas juega un papel muy importante; se ha evolucionado al uso de amplificadores completamente ópticos para evitar la conversión necesaria en amplificadores electrónicos (*OEO, Óptica-Eléctrica-Óptica*), que en el caso de sistemas de Multiplexación por División de Longitud de onda (*WDM, Wavelength Division Multiplexing*) representa un costo elevado ya que es necesario colocar un amplificador electrónico para cada canal usado en el envío de datos.

Al respecto existen varias tecnologías capaces de realizar una amplificación a nivel óptico; entre ellas podemos encontrar los amplificadores ópticos

de semiconductores y los amplificadores de fibra óptica. En esta última categoría se encuentran los más utilizados en sistemas *WDM*, que son los basados en fibra dopada (*DFA, Doped Fiber Amplifier*). Estos amplificadores están compuestos por iones de elementos pertenecientes a las tierras raras, como el neodimio, el iterbio (*YDFA, Ytterbium Doped Fiber Amplifier*), el praseodimio (*PDFA, Praseodymium Doped Fiber Amplifier*), el tulio y el erbio (*EDFA, Erbium Doped Fiber Amplifier*). Este último posiblemente es el más importante [1], ya que trabaja en la tercera ventana de transmisión, los 1550 nm, y cuyo ancho espectral está entre los 30 nm y los 40 nm [2].

En la Fig. 1 se muestra la curva del espectro de ganancia de un *EDFA* comercial con línea punteada. Como se observa, esta curva no tiene un comportamiento plano en toda la banda de operación, esto representa una desventaja importante de dichos dispositivos pues puede generar problemas en el receptor, provocando una saturación. Este inconveniente se puede disminuir por medio de diversas técnicas capaces de generar un espectro de ganancia plano, o casi plano, en la banda espectral deseada.

Dentro de las técnicas más comunes para el aplanado o ecualización de la curva de ganancia se encuentran las siguientes:

1. *Composición de la fibra:* se cambia la estructura interna de la fibra con diversos materiales, por lo general aluminio o materiales no dopados [2].

2. *Híbridos*: esta técnica utiliza la mezcla de varios elementos, como silicio y fluoruro de silicio, así como la superposición de varios tipos de amplificadores, por ejemplo un EDFA con un amplificador Raman en cascada [2]; su principal ventaja es su espectro ancho.
3. *Procesos de filtrado*: en esta técnica se modela la respuesta como una señal cualquiera que es necesario filtrar; existen diversos tipos de filtros como los Mach-Zehnder, filtros en cascada o multietapas, filtros acústico-ópticos ajustables, filtros de Fourier, y las rejillas de periodo largo [2].

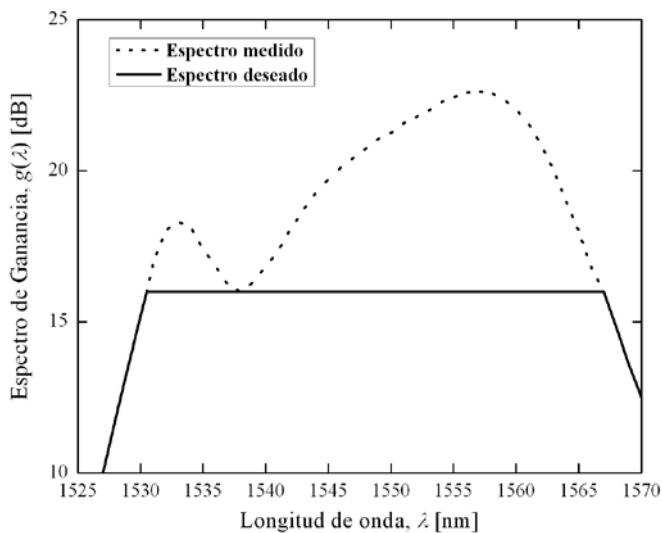


Figura 1. Ganancia de un EDFA comercial; espectro medido (línea punteada) y espectro deseado (línea continua) [3].

En nuestro caso se ha optado por un método de filtrado basado en rejillas de periodo largo (*LPFGs, Long-Period Fiber Gratings*) no uniformes (*con parámetros variables*). El uso de este tipo de rejillas es apropiado en este caso ya que operan como filtros de rechazo de banda [4].

Para el diseño de la rejilla es necesario conocer su espectro de transmisión deseado; con base en el espectro de ganancia medido del EDFA (*Fig. 1, línea punteada*) se fija un espectro de ganancia deseado (*Fig. 1, línea continua*). Con estas dos características se obtiene el espectro de transmisión de la rejilla que será capaz de ecualizar la ganancia

del amplificador, este espectro de la rejilla debe ser inverso al espectro medido en la banda espectral deseada.

El procedimiento de diseño de la rejilla se puede llevar a cabo por medio de la síntesis de la misma, partiendo de su espectro de transmisión deseado. Existen varios métodos para realizar dicha síntesis, entre los que destacan el de pelado de capas (*Layer Peeling*) [5-6], los algoritmos genéticos (*Genetic Algorithms, GAs*) [7-8] y la programación evolutiva (*Evolutionary Programming, EP*) [9].

Existen trabajos donde proponen la ecualización de la ganancia por medio de métodos de síntesis de rejillas, en particular con rejillas de periodo largo [3, 9, 10]. Algunos de los inconvenientes que presentan estos trabajos son: 1. Rango muy grande de la variación del coeficiente de acoplamiento de las rejillas sintetizadas [10], lo que complica la fabricación de las mismas. 2. Variación muy amplia del periodo de las rejillas sintetizadas [3], lo que resulta en el acoplamiento del modo fundamental del núcleo y varios modos de la cubierta dentro del ancho de banda de operación del amplificador; tal comportamiento debe ser considerado en la etapa de análisis de la rejilla, además puede reducir la estabilidad operacional del EDFA. 3. El espectro de transmisión de las rejillas sintetizadas, presentado en dichos trabajos, muestra importantes discrepancias con respecto al espectro de transmisión obtenido con métodos de análisis probados.

En este trabajo presentamos un método de diseño de ecualizadores de ganancia basados en una sola LPFG no uniforme. Este método está fundamentado en la aplicación de un GA especializado que realiza la síntesis de los perfiles de magnitud y fase del coeficiente de acoplamiento de la LPFG, los cuales aseguran una eficiente ecualización de la curva de ganancia del amplificador. Con la utilización de nuestro método, al asegurar que la magnitud y la fase del coeficiente de acoplamiento son los parámetros variables en la rejilla, solo existe acoplamiento entre el modo fundamental del núcleo y un único modo de la cubierta en la banda de operación del

amplificador. De igual forma, al limitar el rango del coeficiente de acoplamiento se asegura que la fabricación del dispositivo es viable con la tecnología disponible en la actualidad. Finalmente, se presentan dos casos particulares de aplicación del método de diseño y, además, se analizan y comparan la efectividad y las ventajas del algoritmo propuesto.

II. MÉTODO DE SÍNTESIS DE LPFGs

Un algoritmo genético es capaz de simular la evolución natural, donde se busca obtener la mejor solución a un problema dado; la obtención de ésta se logra mediante la cuantificación del error que existe entre la solución obtenida y la solución deseada.

Para comenzar el algoritmo es necesario crear una población con un número N de individuos, también conocidos como soluciones potenciales del problema; cada individuo se compone de los parámetros que forman la solución. Como se trata de un algoritmo iterativo, se crea una nueva generación basada en las mejores soluciones, las que entregan una respuesta más cercana a la deseada; por medio de la combinación de los parámetros de dos o más individuos se crea un número de “hijos”; también pueden existir mutaciones (*pequeños cambios en los parámetros de los individuos*). Se pretende que los individuos de la nueva generación sean más cercanos a la solución deseada, este grado de aproximación es cuantificado con una función de error propuesta.

La creación de una nueva generación es realizada de forma iterativa y termina cuando la función de error tiene un valor tan pequeño que puede ser despreciable o se ha alcanzado un número máximo de generaciones propuesto.

En el caso particular de la ecualización de ganancia de un EDFA es necesario tener una LPFG cuyo espectro de transmisión sea asimétrico, de tal manera que sea el inverso de la curva de ganancia del amplificador. Por lo tanto, la rejilla debe ser no uniforme, es decir, uno o varios de sus parámetros físicos, tales como el periodo, el coeficiente de acoplamiento y la fase, deben variar longitudinalmente. En nuestro caso se hace una

variación del coeficiente de acoplamiento y de la fase, dejando constante el periodo.

A continuación se presenta la adaptación particular del GA para el problema de síntesis de ecualizadores de ganancia basados en LPFGs.

2.1 Respuesta deseada

Para poder definir la respuesta deseada es necesario tomar como punto de referencia el espectro de ganancia del EDFA $g_m(\lambda)$, al igual que el valor de la ganancia deseado donde se quiere tener la región plana $g_d(\lambda)$. La respuesta espectral deseada de la rejilla, $t_d(\lambda)$, expresada en decibeles (dB), está dada por

$$t_d(\lambda) = g_d(\lambda) - g_m(\lambda) \quad (1)$$

2.2 Población inicial

En nuestro caso las soluciones potenciales son rejillas de periodo largo no uniformes divididas en M secciones, cada sección con sus propios parámetros: periodo Λ_m , longitud L_m , fase φ_m y coeficiente de acoplamiento κ_m . El algoritmo mantiene invariables el periodo y la longitud de cada sección; de esta manera, tanto el coeficiente de acoplamiento como la fase se crean de forma aleatoria dentro de rangos predefinidos. Con lo anterior la población inicial queda conformada por N soluciones potenciales.

2.3 Error

Para conocer el grado de aproximación de las soluciones se implementa una función de error, también llamada función objetivo. En este caso se hizo mediante el error cuadrático medio entre el espectro deseado, $t_d(\lambda)$, y el espectro obtenido de cada solución, $t_o(\lambda)$

$$E = \frac{1}{K} \sum_{k=1}^K [t_d(\lambda) - t_o(\lambda)]^2 \quad (2)$$

donde K es el número de puntos, o valores de longitud de onda discretos, considerados en el espectro de la rejilla.

2.4 Selección y recombinación

Para la selección de las mejores soluciones y su recombinación se eligen dos “padres”, uno denotado por P_a , elegido de forma aleatoria dentro de un porcentaje p_{ra} de las mejores soluciones, y el otro denotado por P_b , tomado en orden ascendente de un porcentaje p_{rb} de las mejores soluciones. Los “padres” seleccionados son modificados, con una cierta probabilidad p_r , de la siguiente forma

$$\begin{aligned} P'_a &= 0.75P_a + 0.25P_b \\ P'_b &= 0.25P_a + 0.75P_b \end{aligned} \quad (3)$$

Ambas operaciones son realizadas sobre los parámetros de interés, κ_m y φ_m . El primer “hijo” es creado como $H_1 = P'_a$; el proceso de recombinación se repite hasta que se haya creado el número total de hijos por pareja previamente definido. La selección y recombinación se repiten hasta que el número total de soluciones tiene un tamaño N .

2.5 Mutación

En este punto del algoritmo se debe definir una probabilidad de mutación, p_c ; este valor nos indica el porcentaje de la población que sufrirá un cambio en los parámetros, de forma independiente de su función de error. Para nuestro diseño se hizo una modificación, en todas las secciones, tanto del coeficiente de acoplamiento como de la fase; esta modificación es de la siguiente forma

$$\begin{aligned} \kappa'_{i,m} &= \kappa_{i,m} + \Delta\kappa_1 - \Delta\kappa_2 \\ \varphi'_{i,m} &= \varphi_{i,m} + \Delta\varphi_1 \cdot \pi - \Delta\varphi_2 \cdot \pi \end{aligned} \quad (4)$$

donde $\kappa'_{i,m}$ y $\varphi'_{i,m}$ son el nuevo coeficiente de acoplamiento y la nueva fase, respectivamente, de la sección m de la solución i ; los valores $\Delta\kappa_1$, $\Delta\kappa_2$, $\Delta\varphi_1$ y $\Delta\varphi_2$, son valores aleatorios dentro de un rango que va decreciendo cuando el error de la mejor solución de la generación actual es mayor que el error de la mejor solución de la generación anterior.

Al terminar este proceso se verifican las condiciones de error o número máximo de

generaciones para decidir si se ha encontrado la solución buscada o se continúa con la siguiente generación.

III. RESULTADOS

Para la síntesis de las rejillas se utilizó el GA descrito en la sección anterior, mediante el cual se diseñaron dos equalizadores de ganancia para dos EDFAs con distinto comportamiento.

En un primer caso se realizó la síntesis de una rejilla de periodo largo para la equalización del espectro de ganancia del EDFA mostrado en la Fig. 1 con línea punteada. Basados en el GA se tomó un espectro de ganancia deseado con una ganancia constante de 16 dB, porque en la banda espectral a equalizar es el punto con menor ganancia.

El espectro de transmisión de la rejilla se limitó de $\lambda_{min} = 1525$ nm a $\lambda_{max} = 1570$ nm; en la Fig. 2 se muestra el espectro de transmisión deseado con línea punteada, se puede observar que existe un punto de mínima transmisión (*máxima ganancia en el espectro medido*), este punto es conocido como longitud de onda de resonancia, $\lambda_r = 1557$ nm. Suponiendo que los índices de refracción efectivos del modo del núcleo y del modo de la cubierta son $n_{co} = 1.446$ y $n_{cl} = 1.443$, respectivamente, el periodo de la rejilla está dado por

$$\Lambda = \frac{\lambda_r}{(n_{co} - n_{cl})} \quad (5)$$

resultando en un valor de $\Lambda = 469.204$ μm .

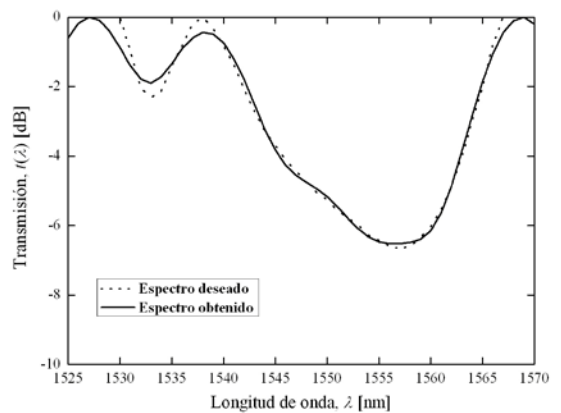


Figura 2. Transmisión de la rejilla de periodo largo; espectro deseado (línea punteada) y espectro obtenido (línea continua).

La longitud total de la rejilla se estableció en $L = 7.0$ cm y se dividió en $M = 25$ secciones del mismo tamaño cada una. El rango del coeficiente de acoplamiento se limitó de $\kappa_{min} = -2.5$ cm^{-1} a $\kappa_{max} = 2.5$ cm^{-1} , estos límites fueron propuestos para evitar que dicho coeficiente tomara valores muy grandes que dificultaran o imposibilitaran la fabricación de la rejilla sintetizada. Para el caso de la fase, se propuso un rango de $\varphi_{min} = -\pi/2$ rad a $\varphi_{max} = \pi/2$ rad, para no tener ambigüedades en la fase de las secciones (debido a que un cambio de signo en el coeficiente de acoplamiento implica un cambio de fase de π rad).

Después de varias ejecuciones del GA, los resultados óptimos de síntesis se obtuvieron con los valores de los parámetros del algoritmo mostrados en la Tabla 1.

En la Fig. 3 se muestra el resultado final de la ecualización, con línea punteada se tiene el espectro medido del EDFA y con línea continua el espectro de ganancia ecualizado. En este último se observa una oscilación al inicio de la banda deseada; en la región de 1540 nm a 1565 nm existe una zona más plana con variaciones entre -0.16 dB y 0.27 dB. Para una banda mayor, de 1531 nm a 1570 nm, la variación de ganancia obtenida en el espectro ecualizado del amplificador se encuentra entre -0.46 dB y 0.42 dB. Los resultados anteriores muestran que este método de síntesis, basado en el GA, presenta un desempeño satisfactorio, pues dichos resultados son comparables con los reportados en trabajos previos. Es importante mencionar que el error casi no disminuyó después de la generación número 300.

Parámetro	Valor
Número total de generaciones	600
Número total de soluciones, N	500
Probabilidad de mutación, p_c	1.00
Porcentaje de las mejores soluciones para P_a, p_{ra}	4.2 %
Porcentaje de las mejores soluciones para P_b, p_{rb}	20 %
Probabilidad de modificación de los	0.86

padres, p_r
 Número de hijos por pareja 5

Tabla 1. Parámetros óptimos del GA.

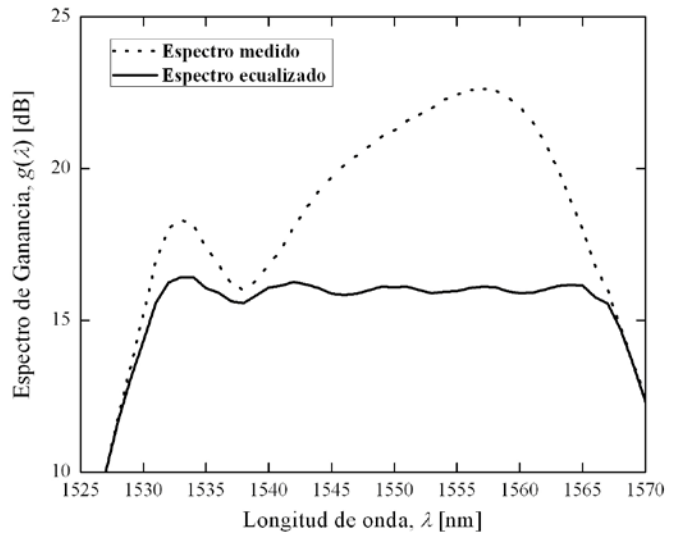


Figura 3. Espectro de ganancia del EDFA; sin ecualizar (línea punteada) y ecualizado (línea continua).

En la Fig. 4 se muestran los perfiles del coeficiente de acoplamiento (línea continua) y de la fase (línea punteada) de la rejilla sintetizada. Para el coeficiente de acoplamiento se tienen valores pequeños, entre -1.71 cm^{-1} y 1.50 cm^{-1} , comparados con lo presentado en [11], esta es una de las ventajas de la utilización de este algoritmo, pues la fabricación de la rejilla sintetizada es más factible con el uso de las tecnologías actuales. Para el caso de la fase las variaciones se encuentran entre -0.72 rad y 0.98 rad.

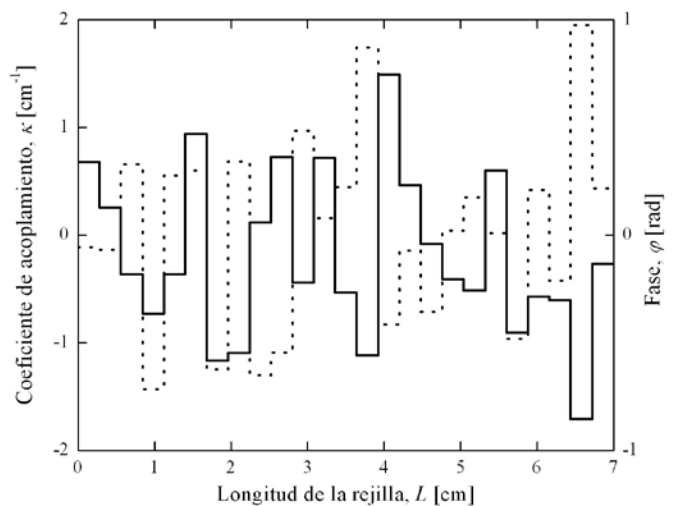


Figura 4. Perfil del coeficiente de acoplamiento (línea continua) y de la fase (línea punteada).

En el segundo caso se consideró un EDFA cuyo espectro de ganancia se muestra en la Fig. 5, línea punteada. El espectro de ganancia deseado se definió desde $\lambda_{min} = 1520$ nm hasta $\lambda_{max} = 1580$ nm, con una ganancia constante de 30 dB. Este nivel de ganancia se eligió porque al tener una ganancia considerable en el amplificador, es necesario contar con un límite para evitar problemas de saturación en algún punto.

En la Fig. 6 se muestra el espectro de transmisión deseado de la rejilla con línea punteada, donde se observa que la longitud de onda de resonancia se encuentra en $\lambda_r = 1556$ nm. En este caso los valores del índice de refracción efectivo del modo del núcleo y del modo de la cubierta se supusieron de $n_{co} = 1.456$ y $n_{cl} = 1.446$, respectivamente. Así que el periodo de la rejilla, obtenido mediante la Ec. (5), es de $\Lambda = 155.6$ μm .

En este caso la longitud total de la rejilla se propuso de $L = 5.0$ cm, dividida en $M = 25$ secciones del mismo tamaño. El coeficiente de acoplamiento nuevamente se limitó de $\kappa_{min} = -2.5$ cm^{-1} a $\kappa_{max} = 2.5$ cm^{-1} . En el caso de la fase también se propuso en un rango de $\varphi_{min} = -\pi/2$ rad a $\varphi_{max} = \pi/2$ rad.

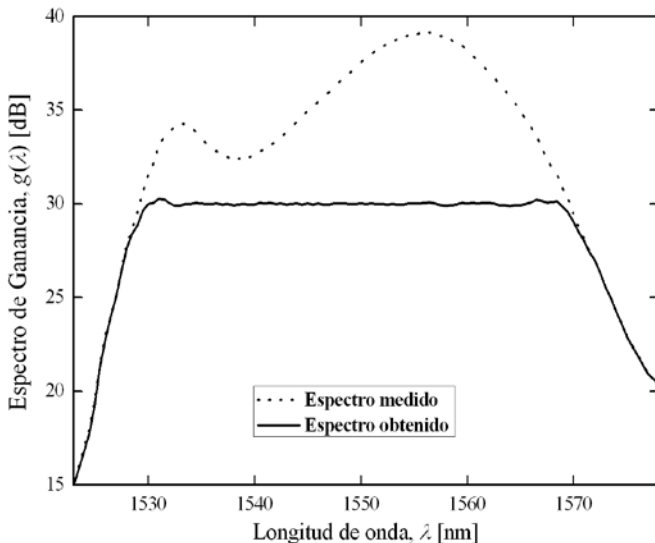


Figura 5. Espectro de ganancia del EDFA; sin ecualizar (línea punteada), y después de la ecualización (línea continua) [9].

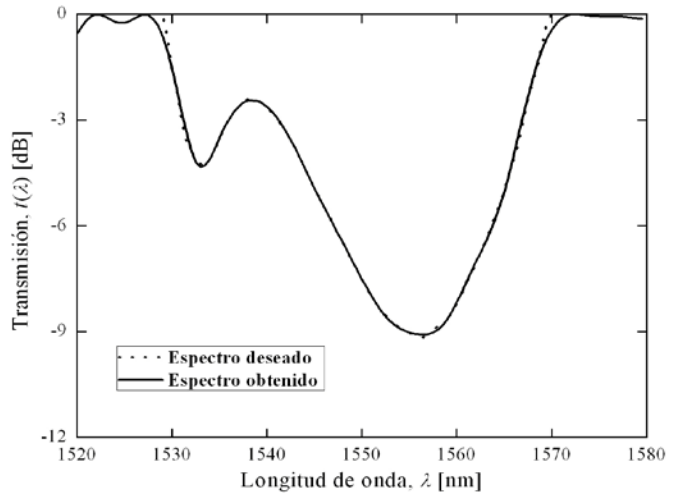


Figura 6. Transmisión de la rejilla de periodo largo; espectro deseado (línea punteada) y espectro obtenido (línea continua).

En la Fig. 5 se ilustra el espectro de ganancia ecualizado del amplificador con línea continua; como se observa se tiene una ecualización más eficiente que en el primer caso, ya que las variaciones en la banda deseada son menores en casi todo el ancho de banda de trabajo del amplificador; tales variaciones se encuentran entre -0.30 dB y 0.50 dB. Si únicamente tomamos una banda espectral de 30 nm, desde $\lambda_1 = 1535.0$ nm hasta $\lambda_2 = 1565.0$ nm, las variaciones en esta región son mínimas, entre -0.08 dB y 0.12 dB. En este caso se modificaron algunos de los parámetros del algoritmo genético para lograr una síntesis óptima de la rejilla, los valores de estos parámetros se presentan en la Tabla 2. Cerca de la generación 600 el error dejó de decrecer.

Parámetros	Valor
Número total de generaciones	1000
Número total de soluciones, N	600
Probabilidad de mutación, p_c	0.90
Porcentaje de las mejores soluciones para P_a, p_{ra}	4.2 %
Porcentaje de las mejores soluciones para P_b, p_{rb}	20 %
Probabilidad de modificación de	0.86

los padres, p_r
 Número de hijos por pareja 5

Tabla 2. Parámetros óptimos del GA.

Finalmente, en la Fig. 7 se muestra el perfil del coeficiente de acoplamiento (*línea continua*), el cual varía entre -1.29 cm^{-1} y 2.15 cm^{-1} , y el perfil de la fase (*línea punteada*), el cual está entre -0.77 rad y 0.97 rad . Estos rangos se encuentran dentro de un margen razonable, lo que nos indica que la fabricación de la rejilla es factible.

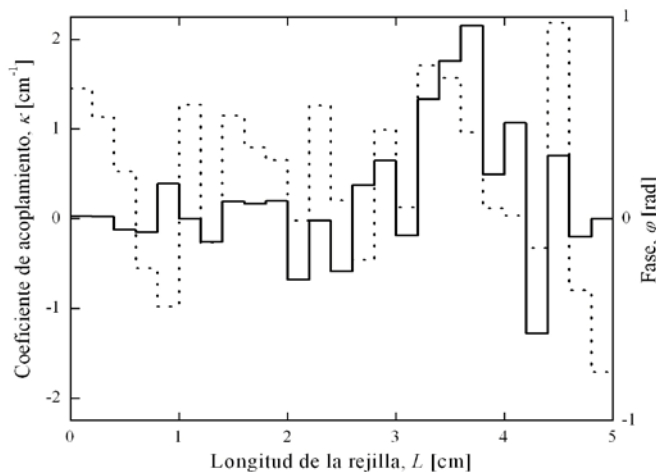


Figura 7. Perfil del coeficiente de acoplamiento (*línea continua*) y de la fase (*línea punteada*).

IV. CONCLUSIONES

El método de síntesis de rejillas en fibra óptica, basado en un algoritmo genético, se ha adaptado con éxito para el diseño de ecualizadores de ganancia para EDFAs basados en rejillas de periodo largo. Se demostró que con este método se pueden lograr curvas de ganancia ecualizadas con variaciones menores a $\pm 0.50 \text{ dB}$, llegando a obtener bandas de operación de hasta 40 nm . A diferencia de trabajos presentados previamente, donde se hace uso de dos o más rejillas de periodo largo, el diseño de estos ecualizadores está basado en una sola LPFG no uniforme, con variaciones del coeficiente de acoplamiento entre -1.71 cm^{-1} y 2.15 cm^{-1} , y de la fase de -0.77 rad a 0.98 rad . Estos rangos de los parámetros mencionados están dentro de los límites tecnológicamente alcanzables, lo que hace factible la fabricación de las rejillas para ser implementadas como ecualizadores.

Es importante mencionar que el uso de los GAs puede ser extrapolado a muchas otras disciplinas (*tales como la economía, la ecología, la genética de población, el diseño de antenas, etc.*) en las que se requiera de métodos de optimización para la solución de problemas específicos.

V. REFERENCIAS

- [1] T. Li, "The impact of optical amplifiers on long-distance lightwave telecommunications", Proceedings of the IEEE, Vol. 81, no. 11, pp. 1568-1579, Nov. 1993.
- [2] M. J. F. Digonnet, *Rare-earth-doped fiber lasers and amplifiers*, CRC Press, 2001.
- [3] J. Bae, J. Chun, S. B. Lee, "Synthesis of long-period fiber gratings with the inverted erbium gain spectrum using the multiport lattice filter model", J. of Lightwave Technol., Vol. 22, no. 8, pp. 1976-1986, Aug. 2004.
- [4] A. M. Vengsarkar, P. J. Lemaire, J. B. Judkins, V. Bhatia, T. Erdogan, J. E. Sipe, "Long-period fiber gratings as band-rejection filters", J. of Lightwave Technol., Vol. 14, no. 1, pp. 58-68, Jan. 1996.
- [5] L. Wang, T. Erdogan, "Layer peeling algorithm for reconstruction of long-period fibre gratings", Electron. Lett., Vol. 37, no. 3, pp. 154-156, 2001.
- [6] J. Skaar, L. Wang, T. Erdogan, "On the synthesis of fiber Bragg gratings by layer peeling", J. of Quantum Electron., Vol. 37, pp. 165-173, 2001.
- [7] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, 1992.
- [8] J. Skaar, K. M. Risvik, "A genetic algorithm for the inverse problem in synthesis of fiber gratings", J. of Lightwave Technol., Vol. 16, no. 10, pp. 1928-1932, 1998.

- [9] C. L. Lee, Y. Lai, “*Long-Period Fiber Grating Filter Synthesis Using Evolutionary Programming*”, *Fiber and Integrated Optics*, Vol. 23, no. 4, pp. 249-261, 2004.
- [10] C. L. Lee, Y. Lai, “*Evolutionary programming synthesis of optimal long-period fiber grating filters for EDFA gain flattening*”, *IEEE, Photon. Technol. Lett.*, Vol. 14, no. 11, pp. 1557-1559, 2002.

VI. AUTORES

Dr. Ignacio Flores Llamas obtuvo el grado de Maestro en Ciencias en la Universidad de Sheffield, Reino Unido, y el grado de Doctor en la Universidad Nacional Autónoma de México. Actualmente se desempeña como investigador y profesor en la Facultad de Ingeniería de la UNAM.

Alfredo Beltrán estudió la Ingeniería en Telecomunicaciones en la Universidad Nacional Autónoma de México y actualmente continúa sus estudios en la Maestría en Ingeniería Eléctrica en la misma Universidad.

Dr. Serguei Khotiaintsev recibió el grado de Maestro en Ciencias y el grado de Doctor de la Facultad de Ingeniería en Radio del Instituto Politécnico de Kiev (hoy en día Universidad Nacional Tecnológica de Ucrania). Actualmente es Profesor de Tiempo Completo de la Facultad de Ingeniería de la Universidad Nacional Autónoma de México. Es autor o co-autor de más de 120 artículos, algunos libros y tiene 19 patentes. El Dr. Khotiaintsev es *Senior Member* del IEEE.

SISTEMA DE POSICIONAMIENTO AUTOCONTROLABLE POR LA INTENSIDAD DE UN HAZ LÁSER

Antúñez Cerón Edgar Eduardo, Zamudio Lara Álvaro, Márquez Aguilar Pedro Antonio.

Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE) -
 Centro de Investigación en Ingeniería y Ciencias Aplicadas (CIICAp).
 Av. Universidad 1001, Col. Chamilpa C.P. 62209, Cuernavaca, Morelos.
 Correo electrónico: lalo.ance@gmail.com

RESUMEN

En el presente artículo, se propone el diseño e implementación de un sistema de posicionamiento de control automático por la intensidad de la luz de un láser con el propósito de encontrar el punto de máxima intensidad en la distribución de potencia de un haz con perfil gaussiano. Utilizando para su construcción dispositivos programables (microcontroladores), dispositivos ópticos (medidor de potencia y sensor de detección), dispositivos electrónicos de bajo costo y material reciclado.

Palabras clave: control automático, haz gaussiano, microcontrolador, sensor de detección.

I. INTRODUCCIÓN

En nuestros días la tecnología ha evolucionado a tal grado que es necesario automatizar gran parte de las tareas donde se requiere la obtención de resultados precisos y confiables de algún proceso bajo estudio.

La mayoría de los haces de láser tienen distribuciones no uniformes de potencia o energía (gaussianas). Sin embargo, en muchas técnicas ópticas es necesario encontrar el punto de máxima intensidad como en la técnica de barrido en Z (Z Scan) que es un método práctico para la medición de no linealidades ópticas. Esta técnica está basada en el principio de distorsión espacial de un haz, y consiste en desplazar la muestra en dirección del eje óptico, a lo largo de la distancia de Rayleigh, y detectar sobre este mismo eje los datos de transmitancia (punto máximo de la intensidad). La relación de los valores de transmitancia en función de la posición de la muestra nos muestra el signo

de no linealidad, y nos permite el cálculo del índice no lineal de refracción de los materiales. Así, la detección de un valor mínimo de transmitancia (valle) seguido de uno máximo (pico) en el barrido, nos indica un comportamiento no lineal positivo, y viceversa [1]. El arreglo para esta técnica se muestra en la Figura 1.

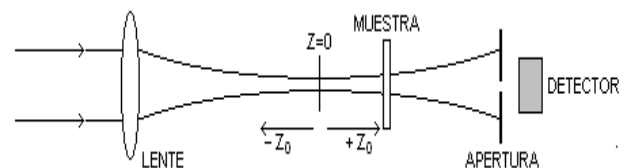


Figura 1. Técnica de barrido en Z.

En la medición de la distribución transversal de irradiancia de un haz láser el parámetro de mayor interés es el diámetro, dado por el punto $1/e^2$ en irradiancia para haces gaussianos. Por otro lado, es importante conocer la forma de la distribución transversal para valorar si el haz puede ser aproximado por un haz gaussiano. [2]

Un haz gaussiano es una forma de propagación de la radiación en un medio no guiado caracterizada por las siguientes propiedades: simetría cilíndrica alrededor del eje de propagación, perfil transversal de la intensidad en forma de campana de Gauss, con valores despreciables para puntos del espacio suficientemente alejados del eje de propagación y las normales a los frentes de onda son rayos paraxiales, es decir, el haz posee una baja divergencia angular. [3]

En la Figura 2 puede apreciarse la distribución de la intensidad de un haz con perfil gaussiano.

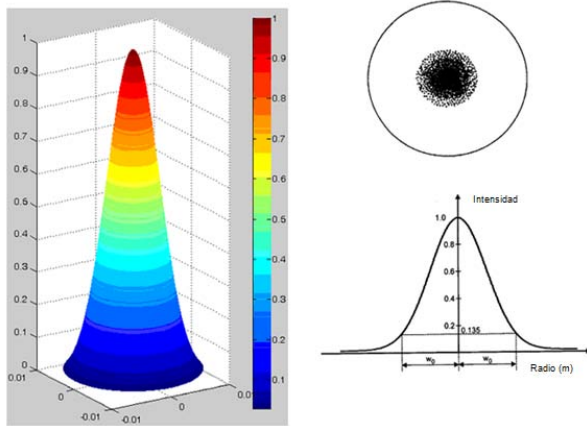


Figura 2. Distribución de la intensidad de un haz con perfil gaussiano. [4]

Existen diversas técnicas para resolver esta búsqueda del punto máximo en la distribución de potencia de un haz láser y una de ellas es la técnica manual. En esta técnica el usuario hace uso de dispositivos ópticos como el sensor de detección y el medidor de potencia óptica que interconectados le asisten en la búsqueda de dicho punto. Para ello, el usuario deberá sostener el sensor de detección con la mano y realizar el barrido por el cuerpo del haz bajo estudio de tal forma que se acerque a la mayor lectura que pueda encontrar y continuar realizando la búsqueda del punto alrededor de esta determinada área.

Durante este proceso el usuario cuenta únicamente con un medio de aproximación y orientación: el medidor de potencia óptica. El medidor es el instrumento con el que el usuario se guía para encontrar la máxima lectura. Conforme se realiza la toma de lecturas, el usuario debe de visualizarlas en la pantalla del medidor para orientarse y realizar el rastreo en determinadas regiones del haz y asegurarse de que ha encontrado la máxima.

A pesar de lo complicada y lo poco confiable que puede resultar esta técnica, existen otros factores que advierten aún más complicaciones al proceso. Entre éstos podemos mencionar la poca estabilidad con la que el usuario sostiene el sensor de detección y la escasa posibilidad de mantener o repetir la

lectura del máximo de intensidad más de una vez durante el proceso.

Una de las técnicas más versátiles es la que se propone a continuación, se ha implementado un mecanismo de posicionamiento bidireccional en dos dimensiones (X, Y) para realizar un barrido sobre los dos diámetros principales de la circunferencia del haz que incide sobre la superficie del sensor. El sensor de detección utiliza un obturador que se coloca sobre él, este obturador tiene una pequeña apertura (de aproximadamente 0.15 mm) en el centro, de tal forma, que ésta se encuentre sobre el centro de la matriz de detección del sensor para asegurar una captura de datos confiable. Es a través de esta apertura que el haz penetra para ser detectado por el sensor y recibir una lectura de intensidad luminosa puntual. Este método nos asegura realizar una toma de lecturas puntuales dentro del perfil del haz incidente y nos evita la sobre saturación de información del sensor.

El mecanismo de posicionamiento que se propone se caracterizó como un sistema de autocontrol o sistema de lazo cerrado para asegurar una evaluación confiable de la variable bajo estudio (intensidad luminosa) y determinar el mejor desempeño del sistema en la búsqueda del punto de máxima intensidad durante el proceso.

Existen dispositivos de posicionamiento comerciales del orden de micras y nanómetros con la desventaja de que para un movimiento en más de una dimensión es necesario utilizar dos o más dispositivos de posicionamiento adicionales sin mencionar que para su funcionamiento se debe de adquirir una unidad de control específica.

II. MÉTODO

En todo sistema de control podemos considerar una señal de entrada que actúa sobre el mismo y una señal de salida proporcionada por el sistema. En la Figura 3 se puede observar la señal de entrada del sistema que corresponde a la intensidad de la luz del haz láser y la señal de salida del mismo que es la posición de los ejes coordenados en el punto de máxima intensidad.

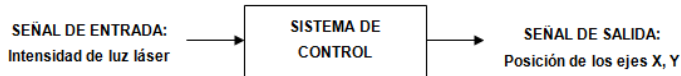


Figura 3. Señales del sistema.

Se puede decir que un sistema de control es un arreglo de dispositivos que permiten desarrollar alguna función de control o de posición. También se puede definir varias categorías generales de los sistemas de control como los sistemas de control de lazo cerrado y los sistemas de control de lazo abierto.

Un sistema de control de lazo cerrado tiene como característica principal el contar con una etapa de comparación entre la señal de entrada y la señal de salida del mismo y de acuerdo con esta evaluación accionar los mecanismos correspondientes para la respuesta más fiable del sistema.

El modelado a bloques del sistema de posicionamiento de lazo cerrado se puede apreciar en la Figura 4.

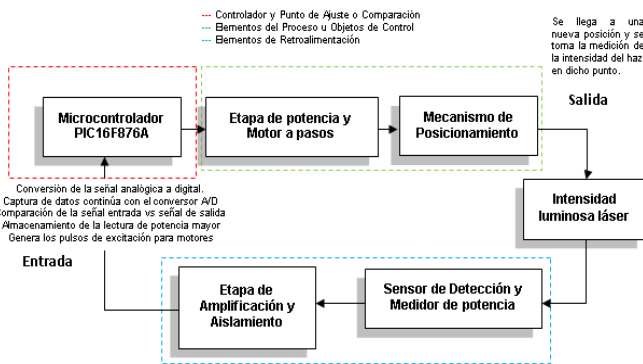


Figura 4. Modelado a bloques del sistema de posicionamiento autocontrolable por la intensidad de un haz láser de tipo lazo cerrado.

El modelado anterior hace referencia a las etapas de las funciones que se llevan a cabo en cualquiera de los dos ejes de desplazamiento. Es decir, el bloque anterior es idéntico en operación y caracterización tanto para el mecanismo de desplazamiento en X como para el mecanismo en Y.

Como se puede ver en la Figura 4 la etapa de adquisición de datos utiliza el sensor de detección y el medidor de potencia óptica. Estos dos dispositivos interconectados nos permiten realizar la medida y lectura de la intensidad en el haz del láser bajo estudio. El medidor proporciona una salida analógica que es la que se dirige hacia la terminal correspondiente del microcontrolador y así darle el tratamiento requerido para el funcionamiento del sistema.

Antes de que la señal de entrada de nuestro sistema sea dirigida a la terminal correspondiente del microcontrolador, ésta deberá pasar por la etapa de amplificación y aislamiento de impedancia, ya que generalmente muchas de las señales que provienen de los transductores que son los encargados de capturar la variable o fenómeno físico real nos proporcionan una señal de salida en voltaje o corriente variable en el tiempo muy atenuada, es decir, su amplitud es muy pequeña, generalmente del orden de los mV (milivolts) o mA (miliamperes).

Esta etapa logra una amplificación de la señal analógica por un determinado factor y el aislamiento de las impedancias de los diferentes dispositivos interconectados para evitar de este modo la atenuación o “caída” de la señal analógica cuando se dirija hacia la terminal del microcontrolador correspondiente.

La base de todo el sistema motriz recae sobre los ejes de desplazamiento en dirección X, Y. Éstas bases móviles están conformadas por dos unidades de reproductores DVD, una montada perpendicularmente sobre la otra y con el soporte para el transductor de luz (sensor de detección) sobre el eje de desplazamiento Y. El eje de desplazamiento en X tiene una guía con movimiento bidireccional y sobre ésta se encuentra sujeta ortogonalmente el eje de desplazamiento Y, el cual al igual que el eje X cuenta con una guía de movimiento bidireccional. El sistema cartesiano de ejes de desplazamiento para este sistema se puede observar en la Figura 5.

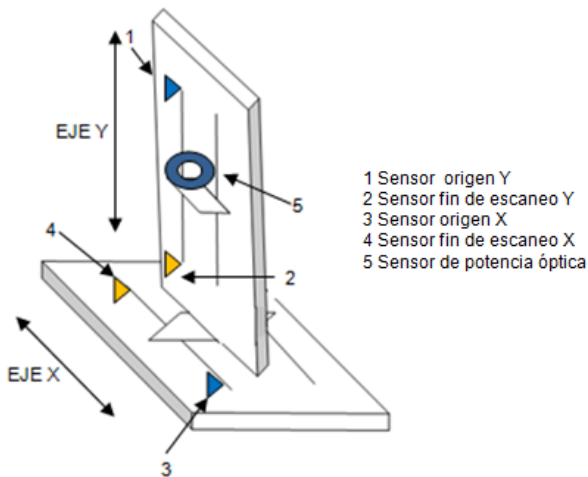


Figura 5. Ejes coordenados de desplazamiento del sistema

En los extremos de cada uno de los ejes de desplazamiento se encuentra colocado un sensor de interrupción, éste tendrá la tarea de avisar al microcontrolador mediante su interrupción que la guía del eje ha llegado a su extremo final.

Los motores son el elemento mecánico por excelencia en todo sistema de posicionamiento, y en este caso no es la excepción, para ello se hace uso de dos motores a pasos para accionar cada uno de los ejes de desplazamiento. Este tipo de motores tiene como principal característica el realizar un avance mecanizado en pasos incrementales, todos en una misma medida, lo cual da ventaja ante cualquier otro motor ya que podemos posicionarlo a voluntad propia con la secuencia de pulsos adecuada.

La secuencia de pulsos utilizada para la configuración de los motores fue la “*secuencia de medio paso*”, ésta es generada por el microcontrolador y haciendo uso de específicos trenes de engranajes implementados se obtiene una resolución de desplazamiento del orden de décimas de micras.

Una de las partes importantes en la etapa de diseño es la parte concerniente a los engranes. Para nuestro sistema en particular se implementaron dos trenes de engranajes; uno para cada base móvil del sistema y se adaptaron entre el eje de desplazamiento (guía móvil) y el motor a pasos

correspondiente de cada eje. Con los trenes de engranajes implementados se logra obtener una resolución de desplazamiento en cada uno de los ejes lo más parecida posible. Obteniendo una resolución de avance para el eje X de 0.003010 mm y para el eje Y una de 0.003141 mm.

Por otro lado, la etapa de potencia de este sistema tiene básicamente la tarea de hacerla de interfaz entre la etapa de control digital y los motores del sistema, y nos permite conmutar niveles lógicos (0V y 5V) con voltajes y corrientes grandes. Para ello hacemos uso de dos circuitos integrados dedicados especialmente para realizar esta tarea.

Finalmente, el microcontrolador es el encargado mediante su programación de ser la unidad de control automática principal.

De esta forma, el microcontrolador es el encargado de realizar la adquisición de la señal física de intensidad luminosa del haz láser con ayuda del medidor y el sensor de detección, dicha señal es previamente tratada en la etapa de amplificación y aislamiento para posteriormente llevar a cabo una conversión analógica-digital de la señal mediante el convertidor A/D y generar los resultados correspondientes en formato digital, esta información es almacenada en registros de la memoria interna del microcontrolador para llevar a cabo una comparación entre los datos obtenidos, el resultado de dicha comparación es evaluado para determinar la generación de las secuencias de excitación y así energizar las bobinas que activarán a cada uno de los motores que se encuentran instalados en cada base móvil de desplazamiento. Mediante la constante toma de lecturas y su correspondiente comparación es como el microcontrolador puede realizar el proceso de automatización y el rastreo del punto de mayor intensidad sobre los diámetros principales de la circunferencia del haz que incide sobre la superficie del obturador que recubre al sensor.

El funcionamiento del sistema es el siguiente: Se realiza un “escaneo” sobre el diámetro principal del eje; mientras el sistema avanza de extremo a extremo en una determinada dirección toma las

lecturas de la intensidad por cada paso que avanza, al finalizar el “escaneo” ha almacenado la mayor lectura registrada durante el proceso. Finalmente, el sistema regresa a la posición en dónde registró la mayor lectura de la intensidad láser. Este proceso lo realiza para ambos ejes de desplazamiento.

Como resultado, tenemos a la mano un sistema automático en el cual la intervención humana queda en la mayor medida posible fuera del proceso.

La Figura 6 muestra el montaje del sistema de posicionamiento y los diferentes dispositivos que lo conforman.

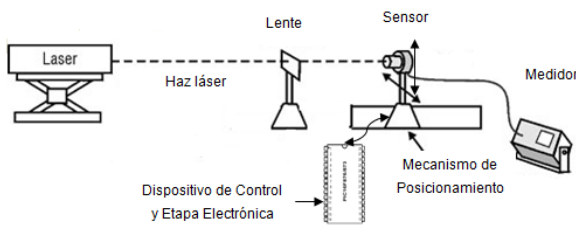


Figura 6. Montaje del sistema de posicionamiento.

Las pruebas realizadas con el sistema de posicionamiento se llevaron a cabo bajo la siguiente condición: la lente convergente estará colocada a 10 cm del dispositivo emisor láser y ésta deberá alinearse propiamente a la altura y ángulo requerido para un buen desempeño del haz. A continuación, se coloca a una distancia “*d*” el sistema de posicionamiento implementado. La distancia de prueba “*d*” no es más que la distancia entre la lente y el mecanismo de posicionamiento. A esta distancia el haz tiene un diámetro “*D*”, éste aumentará su magnitud cuanto mayor sea la distancia “*d*”. Lo anterior se debe a la propiedad de la divergencia de un haz con perfil gaussiano. Esta característica se puede apreciar en la Figura 7.

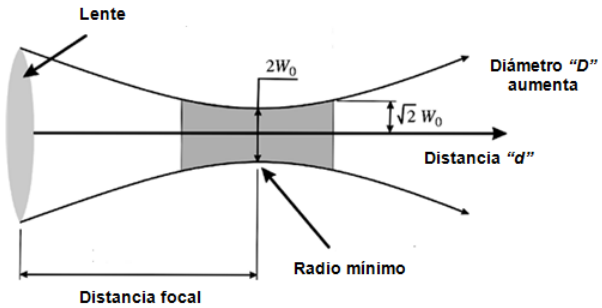


Figura 7. Divergencia de un haz gaussiano.

Como se observa en la Figura 7 el diámetro del haz aumenta conforme se aleja del radio mínimo o foco. A mayor distancia “*d*” se obtendrá un mayor diámetro “*D*”.

El criterio bajo el cual se realizó la evaluación de las pruebas y la obtención de los resultados correspondientes se ilustra en el arreglo experimental de la Figura 8.

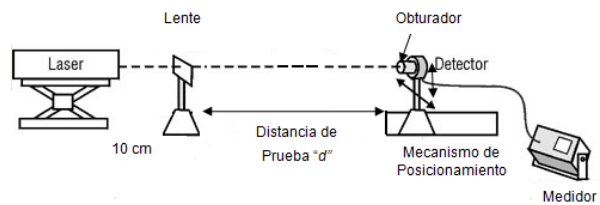


Figura 8. Arreglo experimental para la evaluación de las pruebas.

En la figura anterior se puede ver que la distancia entre el láser y el pie que soporta la lente óptica es de 10 centímetros y esta distancia entre ambos dispositivos ópticos se conserva a lo largo de la realización de las pruebas. El parámetro a modificar es la distancia de prueba “*d*”(y como consecuencia “*D*”, cambiará su valor).

La Tabla 1 muestra los resultados de las diferentes pruebas y puestas en marcha del sistema para diversos valores de “*d*” y “*D*”, respectivamente. Las pruebas se evaluaron cada 10 centímetros de distancia a partir de la posición del soporte de la lente y se incrementa gradualmente hasta llegar a los 150 centímetros.

Distancia de prueba "d"	Diámetro aproximado del haz láser "D"	Intensidad láser obtenida
10 cm	Aprox. 2 mm	3.250 mW
20 cm	3 mm	1.091 mW
30 cm	4 mm	0.692 mW
40 cm	6 mm	0.312 mW
50 cm	9 mm	0.238 mW
60 cm	11 mm	0.165 mW
70 cm	13 mm	0.131 mW
80 cm	15 mm	0.085 mW
90 cm	17 mm	0.047 mW
100 cm	19 mm	0.043 mW
110 cm	20 mm	0.037 mW
120 cm	23 mm	0.035 mW
130 cm	25 mm	0.028 mW
140 cm	27 mm	0.025 mW
150 cm	30 mm	0.023 mW

Tabla 1. Resultados de las pruebas realizadas con el sistema de posicionamiento para diferentes valores de "d".

Una fotografía del sistema real puede apreciarse en la Figura 9. En ésta se muestra el dispositivo láser bajo estudio (HeNe, $\lambda=633\text{nm}$, Potencia máxima = 41 mW) disparando el haz, el cuál pasa a través de una lente plano convexa con distancia focal de 2.0 cm. y finalmente, el cuerpo del haz incide sobre la superficie del sensor de detección (obturador), el cual a su vez se encuentra colocado sobre el eje de desplazamiento en dirección Y, y éste se encuentra montado ortogonalmente sobre el eje de desplazamiento en dirección X. Se aprecian también, el medidor de potencia y la fuente de alimentación DC a un costado del sistema de posicionamiento.

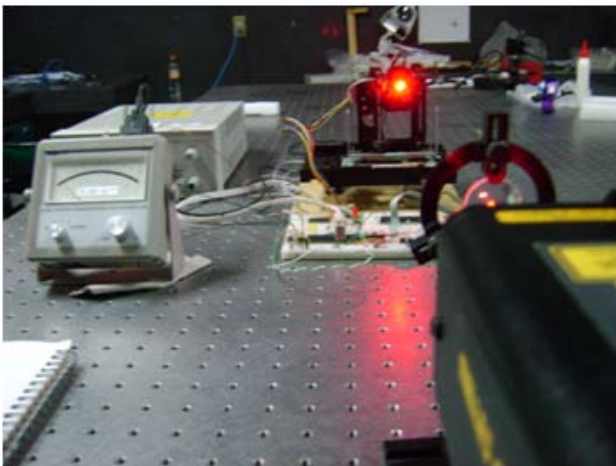


Figura 9. Sistema real de posicionamiento.

III. CONCLUSIONES

Se ha implementado un sistema de posicionamiento bidimensional cartesiano para la búsqueda del punto de máxima intensidad en la distribución de perfil gaussiano de un haz láser.

El sistema opera de manera automática minimizando la intervención de la mano humana durante el proceso de adquisición, medición y evaluación de los datos.

Se diseñó e implementó un arreglo de engranajes para maximizar la resolución de avance en ambos ejes de desplazamiento. Esta resolución resultó tener una magnitud en el orden de las décimas de micras.

La repetitividad de los resultados es muy aproximada, por lo que se puede decir que el sistema es preciso dentro de un límite considerado mínimo de error.

IV. REFERENCIAS

- [1] P.B. Chapple, J. Staromlynska, J. A. Hermann, T. J. McKay, R.G. Mcduff. "Single-beam Z-scan: measurement techniques and analysis", Journal of Nonlinear Optical Physics and Materials. Vol. 6, No. 3 251-293, (1997).
- [2] Boreman, Glenn D., *Fundamentos de electro-óptica para ingenieros*, Editorial SPIE Press 1ra Edición 1999.
- [3] España, María Carmen, *Comunicaciones Ópticas: Conceptos esenciales y resolución de ejercicios*, Ediciones Díaz de Santos, 2005.
- [4] Fuentes de la Figura 1:
<http://zone.ni.com/devzone/cda/ph/p/id/239>
<http://www.hope-academic.org.uk/Academic/researchdevelopment/Themes/Neurosciences/Pain/LaserDevelopment.htm>

V. AUTORES

Edgar Eduardo Antúnez Cerón es estudiante de la Ingeniería Tecnológica en Electrónica por el Instituto de Astrofísica, Óptica y Electrónica (INAOE).

Dr. Álvaro Zamudio Lara es Profesor Investigador de Tiempo Completo del Centro de Investigación en Ingeniería y Ciencias Aplicadas (CIICAp), en la Universidad Autónoma del Estado de Morelos. Realizó sus estudios de maestría y doctorado en el Instituto Nacional de Astrofísica, Óptica y Electrónica.

Dr. Pedro A. Márquez Águilar es Profesor Investigador de Tiempo Completo del Centro de Investigación en Ingeniería y Ciencias Aplicadas (CIICAp), en la Universidad Autónoma del Estado de Morelos. Realizó sus estudios de doctorado en el Instituto Nacional de Astrofísica, Óptica y Electrónica.

IMPLANTACIÓN DEL ALGORITMO DE FILTRADO ADAPTABLE QR-RLS EN FPGA

Méndez-Cuanalo Ana Antígona., Martínez-García Mauricio Alberto.

Universidad la Salle, México

Benjamín Franklin 47, Campus II, Col. Condesa, México06140, D.F.

Tel. (55) 5278-9500 ext 2278, correo electrónico: fierro.mendez@hotmail.com, mmartine@lci.ulsula.mx

RESUMEN

En el presente trabajo se muestra que es posible construir de forma sencilla un filtro adaptable RLS mediante arreglo sistólico utilizando un FPGA. Iniciando con la descripción de las características del algoritmo de filtrado QR-RLS en forma de arreglo sistólico, se identifican sus requerimientos de operaciones aritméticas en paralelo. Seguidamente, se propone la programación esquemática para FPGA, lo cual resulta una alternativa interesante debido a su capacidad de realizar operaciones aritméticas y lógicas en paralelo. Se finaliza con mediciones en laboratorio que confirman el correcto funcionamiento del sistema.

Palabras clave: Filtro adaptable RLS, arreglo sistólico, FPGA

I. INTRODUCCIÓN

Los filtros digitales encuentran innumerables aplicaciones en áreas tales como control, comunicaciones y cibernética. Estos sistemas se han implantado tradicionalmente mediante procesadores de señales o *DSPs*. La creciente necesidad de sistemas de filtrado cada vez más eficientes ha llevado a los diseñadores a probar con diversas arquitecturas de hardware. En particular, los *FPGA* (*Field-Programmable Gate Array*: Arreglo de compuertas programables en campo) resultan una alternativa interesante, debido a su capacidad de realizar operaciones aritméticas y lógicas en paralelo.

En este trabajo se muestra que es posible construir de forma relativamente sencilla un arreglo sistólico, utilizando un *FPGA*. Proponemos implantar el algoritmo de filtrado adaptable *QR-RLS* [1], así como la conveniencia de usar un *FPGA* en lugar de un microprocesador.

En la literatura se encuentran varias investigaciones sobre filtrado adaptable aplicado a diferentes casos específicos. Algunos trabajos proponen utilizar los filtros *LMS* y *RLS* [2], [3],[4], Otros autores nombran el uso de *FPGA*'s [5],[6] o [7] quienes reportan un mejor rendimiento de procesamiento con *FPGA*'s en comparación con el uso de diferentes microprocesadores, los cuales dependen del área de aplicación, como por ejemplo control, comunicaciones o visión. Un artículo más reciente [8] describe un sistema de filtrado de señal mediante arreglos sistólicos.

Este trabajo se organiza de la siguiente manera. En la segunda sección se describen características del filtro *RLS* y del algoritmo QR-RLS: arreglo sistólico, la implantación de arreglos sistólicos en *FPGA*, así como la identificación de sistemas mediante arreglo sistólico *RLS*. En la tercera sección se describe la simulación del algoritmo programado para *FPGA*, mostrando el funcionamiento del arreglo sistólico, con el fin de que los resultados obtenidos sirvan para evaluar el correcto funcionamiento del sistema que se implantará mediante un *FPGA*. De esta manera, será posible medir el desempeño total del sistema en términos del tiempo de procesamiento y así concluir en la cuarta sección, mostrando las ventajas del uso de los *FPGA*'s sobre los sistemas tradicionales, al ejecutar un arreglo sistólico.

II. ANTECEDENTES

II.1. Algoritmo RLS

El algoritmo *RLS* parte de la solución global de problemas de estimación de parámetros por mínimos cuadrados (*LS - Least Squares*) [1]. Este algoritmo estima los coeficientes $C[n]$ de un filtro transversal, por cada muestra de entrada $x[n]$ se hace una actualización de los coeficientes en el tiempo discreto n tomando en cuenta las N entradas anteriores [2].

Los coeficientes $C_N[n]$ se pueden actualizar mediante la recursión:

$$C_N[n] = C_N[n-1] + K_N[n] e[n]$$

donde:

$$K_N[n] = R_N^{-1}[n] x_N[n]$$

$$x_n[n] = [x[n] \ x[n-1] \ \dots \ x[n-NH]]^T$$

es el vector de datos de entrada, y

$$e[n] = d[n] - C_N[n] x_N[n]$$

es el error, siendo $d[n]$ la señal deseada, el vector de coeficientes del filtro adaptable es

$$C_N(n) = [C[n] \ C[n-1] \ \dots \ C[n-N+1]],$$

y $R_N[n]$ es la matriz de autocorrelación de los datos de entrada.

El vector de ganancias $K_N[n]$ se conoce como ganancia de Kalman [9] y el algoritmo *QR-RLS* lo puede calcular recursivamente sin necesidad de invertir la matriz $R_N^{-1}[n]$ en cada iteración, la cual se puede calcular recursivamente como

$$R_N[n] = \lambda R_N[n-1] + x_N[n] x_N[n]^T$$

Las operaciones efectuadas por el algoritmo *RLS*, por dato de entrada, son $4N^2 + 4N$ multiplicaciones y $3N^2 + N-1$ sumas. Algunos autores reportan que la convergencia del algoritmo *RLS* es del orden $2N$ iteraciones cuando la media del error de convergencia es pequeña comparada con la señal deseada $d[n]$.

Un criterio de desempeño frecuentemente utilizado, es la minimización del error cuadrático $e[n]$ multiplicado por un factor de peso

exponencial λ , conocido como factor de olvido, el cual tiene un valor muy cercano a uno. Al inverso de $1 - \lambda$ se le conoce como la memoria del algoritmo.

II.2. Algoritmo QR-RLS: arreglo sistólico.

En la Figura 1 se muestra una estructura básica de un filtro adaptable [1], donde $x[n]$ es la entrada al filtro en el tiempo n , $d[n]$ es la señal deseada, $y[n] = C_N[n] x[n]$ es la señal de salida del filtro y $e[n]$ es la señal de error definida por $d[n] - y[n]$. El filtro adaptable es representado por N pesos representados en el vector $C_N[n]$.

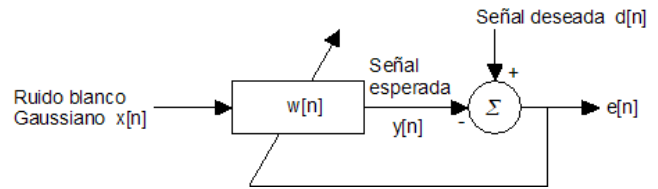


Figura 1. Estructura básica de filtro adaptable

El objetivo del algoritmo *QR-RLS* es encontrar el conjunto de pesos $w = C_N[n]$, los cuales minimizan

$$\xi = \|e\|^2 = \|d - A w^* [n]\|^2$$

Dónde $\|e\|^2$ es definido como $\|e\|^2 = e^H e$, con $(\cdot)^H$ representando al vector transpuesto conjugado. Los otros elementos de la ecuación inmediata anterior están definidos por:

$$e = \begin{bmatrix} e[0] \\ e[1] \\ \dots \\ e[n] \end{bmatrix}, d = \begin{bmatrix} d[0] \\ d[1] \\ \dots \\ d[n] \end{bmatrix}, A = \begin{bmatrix} x_0^T \\ x_1^T \\ \dots \\ x_n^T \end{bmatrix}, x_k = \begin{bmatrix} x[k] \\ x[k-1] \\ \dots \\ x[k-N+1] \end{bmatrix}$$

$$w^* [n] = \begin{bmatrix} w_0[n] \\ w_1[n] \\ \dots \\ w_{N-1}[n] \end{bmatrix}$$

El algoritmo de *QR-RLS* puede ser ejecutado siguiendo recursivamente los pasos mostrados abajo, donde el subíndice n denota una dependencia de tiempo en el instante n :

$$\text{Paso 1: } \begin{bmatrix} R_n \\ 0 \end{bmatrix} = Q \begin{bmatrix} \lambda R_{n-1} \\ x_n^T \end{bmatrix}$$

$$\text{Paso 2: } \begin{bmatrix} p_n \\ \gamma \end{bmatrix} = Q \begin{bmatrix} \lambda p_{n-1} \\ d[n] \end{bmatrix}$$

$$\text{Paso 3: } p_n = R_n C_n^*$$

R_N es la matriz de autocorrelación de la entrada $x[n]$, donde λ es el “factor de olvido” del algoritmo. Notando que en el paso 1 el objetivo es encontrar una matriz Q que pone a cero todos los elementos de la fila inferior y hace R_n una matriz triangular superior.

Esto puede ser logrado por una serie de *rotaciones de Givens* para poner en cero los elementos requeridos en la matriz:

$$\begin{bmatrix} \lambda R_{n-1} \\ x_n^T \end{bmatrix}$$

Los coeficientes de rotación de Givens $\cos \theta$ y $\sin \theta$ han sido representado en algunas ecuaciones a través de éste trabajo como c y s , respectivamente.

II.3. Implantación de arreglos sistólicos en FPGA

Un arreglo sistólico es un conjunto de elementos de proceso (*EP*'s) simples e idénticos, los cuales están interconectados únicamente entre procesadores adyacentes (comunicación local) y los cuales interactúan entre sí (Figura 2).

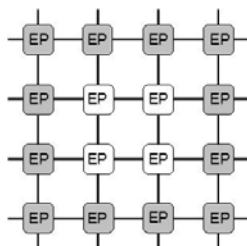


Figura 2. Estructura de un arreglo sistólico

Los datos se mueven a una velocidad constante a través del arreglo pasando de un elemento de proceso al siguiente. Cada uno de los elementos realiza una función específica de manera que el conjunto contribuye a la culminación del proceso. Los datos que se reciben en el sistema desde una memoria externa, son pulsados rítmicamente a través de los procesadores.

El término *arreglo sistólico* fue introducido en las ciencias computacionales por H.T. Kung en 1982 [10]. Un arreglo sistólico se caracteriza por las siguientes propiedades:

- Modularidad. El arreglo está integrado por elementos de proceso cuya función está definida en una forma clara y simple.
- Regularidad. El arreglo está compuesto por módulos o submódulos similares.
- Interconexión local. Cada elemento de proceso tiene conexión solo con sus vecinos.
- Alto grado de procesamiento en línea. Para arreglos que operan en forma serial, la característica de “pipeline” opera de una manera natural debido a la propia arquitectura.
- Multiprocesamiento sincronizado. Ya que todo el arreglo está bien sincronizado por una señal de reloj, cada elemento de proceso ejecuta su correspondiente proceso en forma rítmica

Es conocido que para el diseño de sistemas de Procesamiento Digital de Señales (*DSP*) se requiere de circuitos multiplicadores de alto desempeño, el diseño de estos multiplicadores se ha abordado en sistemas *VLSI* con diversas arquitecturas, técnicas de diseño y celdas básicas particulares [11][12].

Las ventajas de los sistemas diseñados con arreglos sistólicos son: el alto rendimiento operacional utilizando poca memoria, el control de información es sencillo y regular con número reducido de celdas simples y uniformes, la comunicación local frecuentemente a alta velocidad, así como facilidad de construcción, de expansión y de reconfiguración [13].

Debido al paralelismo existente en el algoritmo *QR-RLS* (en su forma sistólica) es posible utilizar *FPGA's* para mejorar su tiempo de convergencia. Cabe aclarar que el costo computacional, que interesa para el presente trabajo no disminuye, pero sí el tiempo de procesamiento por muestra, el cual utilizamos como un parámetro de desempeño.

Este circuito integrado contiene celdas lógicas idénticas que se interconectan por medio de una matriz de cables y switches programables (Figura 3). Las *FPGA's* son matrices de compuertas lógicas eléctricamente programables. Se caracterizan por altas densidades de puerta, alto rendimiento, un número grande de entradas y salidas definibles por el usuario, un esquema de interconexión flexible, y un entorno de diseño similar al de matriz de puertas. Además contienen una matriz interna configurable de relojes lógicos (*CLB* – Clock Logical Blocks) y un anillo de bloques de Entrada/Salida (*IOB*- I/O Blocks) [14][15].

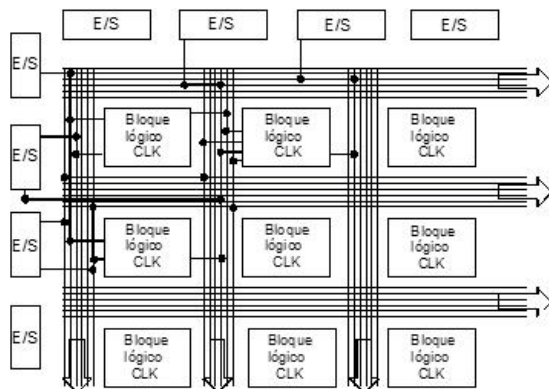


Figura 3. Estructura interna de un FPGA. Línea clara: trayectorias de conexión posibles, línea oscura: interconexión programable o segmento de conexión, y cada bloque lógico tiene su entrada de reloj.

Las *FPGA's* se utilizan en controladores, codificadores, decodificadores y en el prototipo de circuitos *VLSI* y microprocesadores a medida.

II.4. Identificación de sistemas mediante arreglo sistólico RLS

La programación del algoritmo *RLS* conlleva la programación de dos bloques funcionales (Figura 4). Los cuales se denominarán bloque o celda de frontera (redondo) y bloque o celda interna (cuadrado), los cuales interactúan de forma que algunas de las operaciones aritméticas y lógicas pueden realizarse de forma síncrona durante la ejecución del algoritmo.

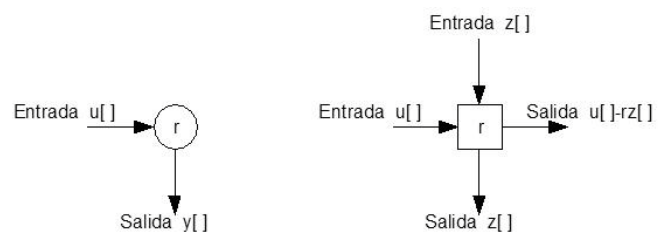


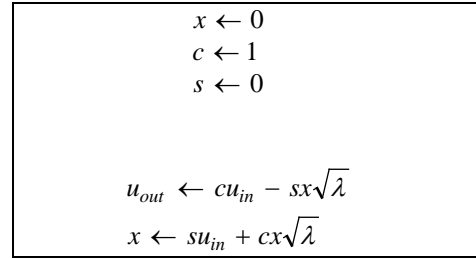
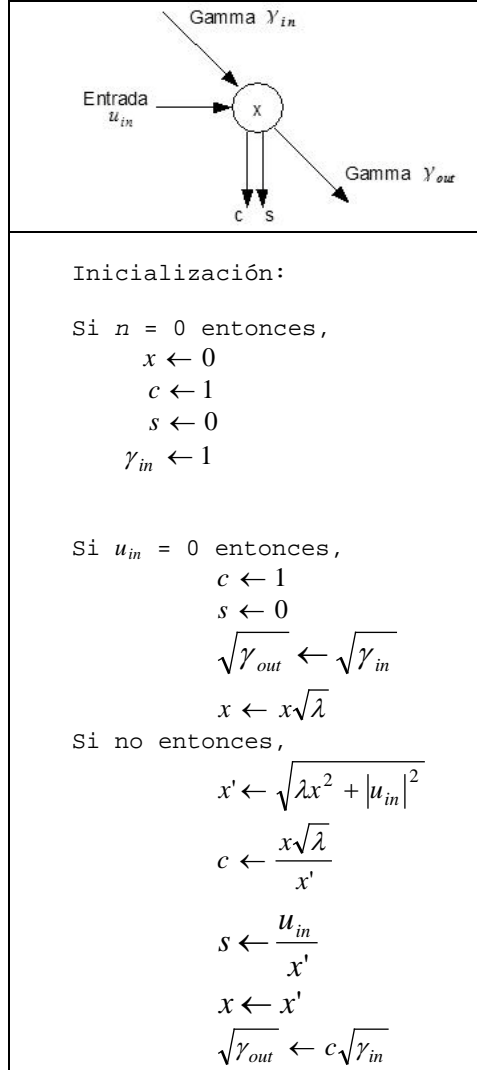
Figura 4. Celdas básicas para el arreglo sistólico. Celda frontera (izquierda) y celda interna (derecha)

Estas celdas son básicas para construir el arreglo sistólico, se ordenan de acuerdo al orden del algoritmo, por ejemplo un sistema de segundo orden utiliza dos bloques redondos (o celda frontera) y tres cuadrados (o celda interna), mientras que uno de tercer orden usa tres redondo y seis cuadrados. Entre mayor sea el orden del algoritmo más bloques serán necesarios.

Independientemente del número de orden, la última fila debe tener por entrada la señal deseada $d[]$, ya que el bloque final realiza la operación de un multiplicador entre la señal estimada $u[]$ y la deseada $d[]$, generándose la señal de salida que, para este caso es el error. El arreglo sistólico se forma por una sección triangular y una lineal; la primera tiene el inicio de las operaciones y la mayoría de las operaciones, mientras que la segunda son las operaciones finales para calcular la señal de salida.

Siguiendo con la estructura del arreglo sistólico, las operaciones a usar, de acuerdo a los diagramas anteriores son:

1. Bloque básico redondo o “celda frontera”:



Para la identificación de sistemas utilizando un arreglo sistólico *RLS*, que se explica en la sección 2.2, donde se estima la señal de salida $y[n]$, tomando en cuenta que los coeficientes para este algoritmo se pueden actualizar recursivamente, sin necesidad de estar invirtiendo la matriz, correspondiente a la Figura 5:

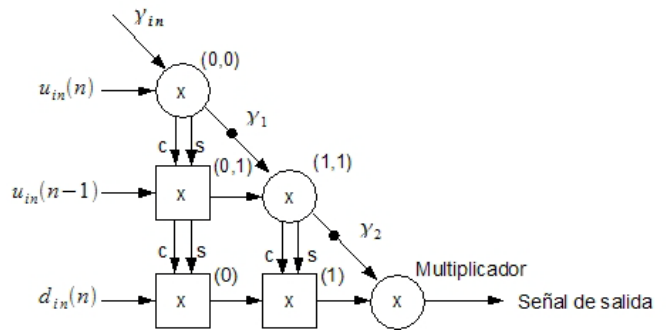
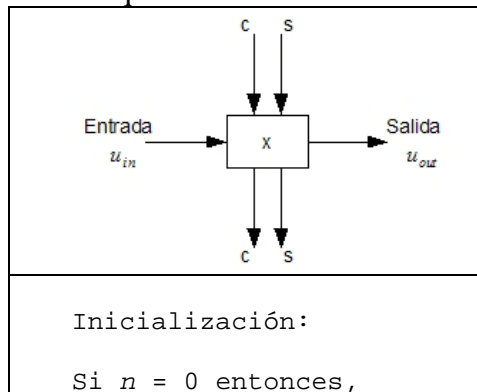


Figura 5. Arreglo sistólico del algoritmo *QR-RLS* de segundo orden. Los puntos en las líneas de unión entre los bloques redondos significan que existe un retraso en la señal.

2. Bloque básico cuadrado o “celda interna”:



Para la programación de este algoritmo se definen constantes como la “memoria” lambda (λ), orden del algoritmo, número de realizaciones m y número de muestras n . Además se requieren definir la señal de entrada $u[]$, y la señal deseada $d[]$. En cada iteración el algoritmo calcula una señal de salida, llamada también estimada.

Para comenzar con el ciclo para el cálculo de la señal estimada, comienza a partir del número de orden hasta el número de muestras a calcular, para inmediatamente comenzar con el siguiente ciclo que, utiliza una señal retrasada de la deseada recorriendo cada uno de los valores de esta señal.

Para cada uno de ellos se actualiza en cada iteración el valor de gamma (γ).

Dicho valor, junto con la señal deseada, forman la entrada para el procedimiento de *bloque redondo()*, el cual actualiza los valores de c y s que son valores de entrada para el *bloque cuadrado()*.

El *bloque cuadrado()* tiene por entrada los valores de c y s actualizados, la señal deseada en retraso, y la señal de salida de este bloque es la actualización de la señal estimada.

Es importante señalar que, las operaciones entre bloques se realizan de manera síncrona, es decir que al mismo tiempo que se realizan operaciones en algún *bloque redondo()*, se pueden estar realizando operaciones en algún *bloque cuadrado()*. Es decir el algoritmo va teniendo un “ritmo sistólico”, donde cada bloque realiza su función de acuerdo con el ciclo de reloj establecido.

III. EXPERIMENTOS REALIZADOS (PROGRAMACIÓN DE UN ARREGLO SISTÓLICO)

La programación de un arreglo sistólico puede realizarse en *Quartus II* (la herramienta provista por Altera, fabricante del *FPGA* utilizado). Para este trabajo se describe la programación de un arreglo de primer orden (Figura 6). La idea principal de programarlo en *FPGA*, es que las operaciones pueden llevarse a cabo en forma paralela, reduciendo en forma notable el tiempo requerido por dato de entrada para el cálculo de la salida.

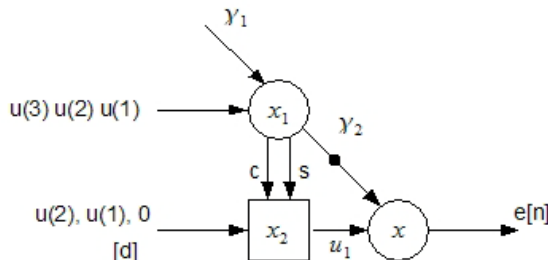


Figura 6. Gráfica de arreglo sistólico de primer orden.

Para iniciar con la programación gráfica del algoritmo se definen los datos de entrada, como lambda, señal de entrada, señal de reloj y la señal deseada. En nuestra simulación, las señales entrada y salida serán comparadas con las obtenidas utilizando *Matlab* para verificar el correcto funcionamiento de nuestro sistema.

IV. RESUMEN DE RESULTADOS (SIMULACIÓN Y COMPARACIÓN DE DESEMPEÑO)

Las pruebas de simulación comienzan con la ejecución del programa en *Matlab*, especificando las señales de entrada y el orden deseado del algoritmo. Dicho programa genera las señales de entrada en forma aleatoria. Una vez ejecutado, se almacenan los resultados de salida. Todos estos datos se convierten a su representación de punto flotante en formato hexadecimal *IEEE 754*, para así definir los datos de entrada para el programa en *Quartus II*, y de esta manera poder comparar los resultados obtenidos por ambas implementaciones.

Haciendo referencia al código de *Matlab* se generaron las señales de entrada $u[]$, y la deseada $d[]$ como valores aleatorios. El arreglo sistólico calcula la señal de error $e[n]$, que es $d[n] - y[n]$, la cual sirve de referencia para verificar el correcto funcionamiento del algoritmo en la tarjeta *FPGA*.

Debido a los requerimientos limitados de espacio (número de unidades lógicas) del *FPGA* disponible (*DE2 de Altera*), se decide realizar pruebas con un arreglo sistólico de primer orden.

Los valores de las señales de entrada $u[]$ y $d[]$ fueron generadas en forma aleatoria del programa en *Matlab*. Tomando en cuenta los valores iniciales: $\gamma = 1$, $c = 1$ y $s = 0$, se obtuvieron los siguientes resultados: (Tabla 1)

	n=0	n=1	n=2	n=3	n=4	n=5
Señal						
$u[]$	-0.1166	-0.1343	-0.0102	-0.0449	-0.1071	-0.1429
$d[]$	0	-0.1166	-0.1343	-0.0102	-0.0949	-0.1071

Tabla 1. Resultados cuando $\gamma = 1$, $c = 1$ y $s = 0$

Después de las primeras cinco iteraciones (o “latidos”) del arreglo sistólico, los valores obtenidos con *Matlab* para la señal de error (*e*) son: (Tabla 2)

n	
0	0
1	-0.0764
2	-0.0846
3	0.0135
4	0.0089
5	-0.0263

Tabla 2. Resultados después de 5 iteraciones.

Posteriormente, para la prueba de simulación del arreglo sistólico programado en *Quartus II*, los valores de las señales de entrada *u[]* y *d[]* junto con el del error *e[]*, se representan en punto flotante con forma hexadecimal *IEEE 754*, como se muestra en la tabla siguiente: (Tabla 3)

n	<i>u[]</i>	<i>d[]</i>	<i>e[]</i>
0	BDEECBFB	0	0
	-0.1166	0	0
1	BE0985F0	BDEECBFB	BD9C8C9C
	-0.1343	-0.1166	-0.076440066
2	BC271DE7	BE0985F0	BDAD3FC0
	-0.0102	-0.1343	-0.08459425
3	B137E910	BC271DE7	3C5C5F9B
	-2.6762486E-9	-0.0102	0.0134505285
4	BDDDB573F	BD37E910	3C115D71
	-0.1071	-0.0449	0.008872376
5	BE125461	BDDDB573F	BCD73CAB
	-0.1429	-0.1071	-0.026274046

Tabla 3. Señales de entrada *u[]*, *d[]* y *e[]*, en punto flotante con forma hexadecimal *IEEE 754*

Se observa que los datos de salida del programa en *Quartus II*, tienen una gran similitud con respecto a los obtenidos en *Matlab*, lo cual confirma el correcto funcionamiento del sistema. Posteriormente, se realizaron pruebas con respecto al tiempo de procesamiento del algoritmo.

Teóricamente el programa en *FPGA* debe ser más rápido puesto que tiene la capacidad de ser programado para realizar operaciones en paralelo, adecuándose perfectamente a la forma del arreglo sistólico, mejorando el tiempo de respuesta.

Una vez programado el *FPGA*, se determinó mediante simulaciones que el sistema requiere 54 ciclos de reloj en obtener un resultado, inmediatamente de que tiene los datos de entrada disponibles.

La frecuencia máxima de reloj de la tarjeta *FPGA* de Altera tipo *Cyclone II* es de 50MHz, lo cual corresponde a un periodo de reloj mínimo de 200ns. Entonces:

$$200ns \times 54 \text{ ciclos de reloj} = 11\mu s$$

Esto significa que cada resultado tarda 11µs en ser determinado. Lo cual corresponde aproximadamente a 92,000 datos en 1 segundo.

En el programa de *Matlab* se insertaron comandos para medir tiempo de procesamiento, del cual se obtuvo un tiempo de 0.017578 segundos con 250 ciclos de reloj. En una computadora con microprocesador de 1.73 GHz se obtuvo lo siguiente:

$$250 \text{ ciclos de reloj} \div 0.017578 \text{ segundos} = 14,222 \text{ datos en 1 segundo}$$

Comparando el número de datos que se obtienen en un segundo entre la tarjeta *FPGA* y *Matlab*, existe una diferencia de:

$$92,000 - 14,222 = 77,778 \text{ datos en 1 segundo}$$

Esto indica que el número de resultados obtenidos de la tarjeta *FPGA* por unidad de tiempo es mucho mayor al que se obtiene por *Matlab*, estableciendo un mejor desempeño del sistema en *FPGA*.

De acuerdo a los resultados en nuestras pruebas para el tiempo de procesamiento, se puede afirmar que un algoritmo de primer orden en ambos programas es notablemente diferente en cuanto a desempeño se refiere. El programa en *FPGA*, tiene un mejor desempeño sobre el de *Matlab* debido a que las operaciones del arreglo sistólico se hacen

paralelamente mientras que en un microprocesador se realizan en cascada.

Se puede afirmar que una de las ventajas del algoritmo en la tarjeta *FPGA* es el tiempo de ejecución, además el uso de recursos de la tarjeta permiten que la desventaja del algoritmo *QR-RLS* de ser computacionalmente complejo, se ve superada.

V. CONCLUSIONES

Con la programación de la tarjeta *FPGA* de un arreglo sistólico, se comprueba la posibilidad de combinar el alto desempeño del algoritmo *RLS* con la rapidez de convergencia del algoritmo *LMS*, logrando una implantación eficiente de un sistema de filtrado adaptable.

El algoritmo programado en *Quartus II*, puede ser medido en los términos arriba mencionados, esto debido a que los datos de entrada y salida con los que fue probado son los obtenidos del programa desarrollado en *Matlab*.

Recordando lo que se planteo desde un inicio, la eficiencia de un sistema de filtrado adaptable, se mide para este trabajo, en términos del desempeño que incluye el error cuadrático y el tiempo de convergencia, donde la convergencia es el tiempo que tarda un filtro adaptable en llegar a una solución óptima. Lo cual fue comprobado al comparar los resultados obtenidos por la programación en *Matlab* y en *Quartus II*.

Si el costo computacional se define como el número de operaciones por unidad de tiempo, el arreglo sistólico programado al ejecutar operaciones en paralelo cubre en menos tiempo las operaciones necesarias, superando los dispositivos tradicionales para ejecutar filtrado adaptable.

Al obtener resultados muy similares (iguales) entre ambas herramientas de programación, es posible afirmar que esta señal de error (de salida), es el factor que indica el buen funcionamiento de la implantación del algoritmo en la tarjeta *FPGA* y la

ventaja de este desarrollo sobre los sistemas tradicionales.

Los resultados aquí obtenidos, llevan a plantear trabajo futuro como por ejemplo el desarrollo en lenguaje *VHDL* o *Verilog*, para optimizar el tiempo de compilación del programa y eliminar la redundancia de código que se genera por la programación esquemática o programación en bloques. Aclarando que el hecho de optimizar la programación no significa que cambien los resultados.

De las pruebas que se hicieron se deriva la posibilidad de que si el sistema se programa con aritmética entera, las operaciones se simplificarían notablemente y el tiempo de ejecución podría reducirse.

VI. REFERENCIAS

- [1] Haykin, Simon, *Adaptive Filter Theory*; Tercera edición, Prentice Hall, 1996.
- [2] Escobar S., Larry, *Estudio comparativo y aplicaciones de algoritmos de filtrado adaptable*, Ingeniería Investigación y Tecnología UNAM abril-junio, año/vol. V, número 002, pp. 89-108, 2004.
- [3] Mario A. García Martínez, *Construcción de operadores básicos sobre campos finitos GF2m*. PhD thesis, Centro de investigación y de estudios del Instituto Politécnico Nacional, México, D.F., diciembre 2004.
- [4] Carlos Espinosa, Perla; Villarroel, *Proposición y simulación de un algoritmo adaptativo para sistemas de antenas inteligentes*, Ingeniare, Revista chilena de ingeniería, 15(3):344–350, 2007.
- [5] Martinez, Mauricio, *Efficient Filter Bank Adaptive Filtering*, Ph.D. Thesis, Northeastern University, Boston Massachusetts 2002.

- [6] D. García-Alís-I.G. Stirling-G.W. Rice-R.W, *Stewart. Bit true adaptive equalization design and simulation library*, Department of Electronic and Electrical Eng UK, 2006.
- [7] Ruiz-Merino, Garrigos-Guerrero, Martínez Alvarez, Toledo-Moreo, *Especificaciones y arquitectura de un prototipo de estetoscopio inteligente*, Technical Report PB-63-007, Universidad Politécnica de Cartagena (Departamento de Electrónica, Tecnología de Computadoras y Proyectos), España 2006.
- [8] Abbes Amira Meher Kumar, Shrutisagar Chandrasekaran, “*FPGA realization of FIR filters by efficient and flexible systolization using distributed arithmetic*”, IEEE Transaction on signal processing, 56(7), July 2008.
- [9] Lev-Ari, Hanoach, *Notas de clase. Statistical Signal Processing*, ECE 3497, Spring 1998, Northeasten University, Boston, E.U.A.
- [10] H.T. Kung, “*Why systolic architectures?*”, IEEE Trans. on Comp., pp. 37-46, 1982.
- [11] Meyer-Baese, Uwe, *Digital Signal Processing with field programmable gate arrays*, Springer, USA 2001.
- [12] Manolakis, Dimitris, Ingle, Vinay, and Kogon, Stephen, *Statistical and Adaptative Signal Processing*, McGraw Hill Series in Electrical and computer engineering, 2000.
- [13] Stephen Brown and Jonathan Rose, “*Architecture of FPGAs and CPLDs: A Tutorial*”, IEEE Design and Test of Computers, Vol. 13, No. 2, pp. 42-57, 1996.
- [14] Albó, Jordi Canals, *FPGAs – 1.0*. La Salle:Universitat Ramon Llull, Seminario especial ULSA, Mexico, 2007.
- [15] ALTERA, “*Implementing Multipliers in FPGAs Devices*”, Application Note 306, July 2004.

http://www.salle.url.edu/~gpazienza/seminarios/presentaciones/P_FPGAs_1_0.pdf

VII. AUTORES

Ing. Ana Antígona Méndez-Cuanalo. Egresada de la Carrera de Ingeniería en Cibernética y Ciencias Computacionales en la Universidad La Salle (1986-1991), México. Estudiante de la Maestría en Ciencias – Área Cibernética de esta misma Universidad.

Dr. Mauricio Alberto Martínez-García Ingeniero en Energía Eléctrica y en Sistemas Electrónicos en 1994 por la Universidad La Salle (ULSA), el grado de Maestro en Ingeniería (Eléctrica) 1997, por Facultad de Ingeniería de la UNAM, y el grado de Doctor en Filosofía (Ph.D.) en Ingeniería Eléctrica, en 2002 por Northeastern University, Boston, ambos con apoyo de CONACyT. Profesor de la Escuela de Ingeniería de la ULSA, y de la Facultad de Ingeniería de la UNAM desde 2002. Integrante del Sistema Nacional de Investigadores (SNI) con el nombramiento de Candidato.

PLATAFORMA FPGA PARA EL DISEÑO Y EVALUACIÓN DE SISTEMAS DE CONTROL

Santiago Espinosa Felipe, Barahona Avalos Jorge Luis

Instituto de Electrónica y Computación
Universidad Tecnológica de la Mixteca
Carretera a Acatlima Km 2.5, Huajuapán de León, Oaxaca, 69000, México
Tel. 9535320214, correo electrónico: fsantiago@mixteco.utm.mx, jbarahona@mixteco.utm.mx

RESUMEN

Este trabajo está relacionado con el desarrollo de módulos, para la evaluación de algoritmos de control, en una tarjeta *Spartan 3E Starter Board*. Dicha tarjeta contiene un FPGA Spartan 3E de Xilinx y, además, incluye un DAC, un ADC, LCD y diversos periféricos de entrada/salida. Sin embargo, el DAC y el ADC utilizan una interfaz SPI (*Serial Port Interfaz*), lo que implica que para obtener o generar una señal analógica se requiere de un módulo secuencial que maneje esta interfaz. Se desarrollaron, en VHDL, los módulos necesarios para el manejo del LCD y de los distintos periféricos de entrada/salida empleados, así como de los algoritmos de control evaluados. Se presentan resultados experimentales de las pruebas realizadas con la plataforma desarrollada en conjunto con un sistema de control de temperatura.

Palabras clave: FPGA, SPI, Picoblaze, VHDL, Sistemas de Control.

I. INTRODUCCIÓN

El uso de FPGA's para la implementación de sistemas electrónicos cada vez es más demandante, por su velocidad de procesamiento y flexibilidad. En este sentido, en el mercado se ofertan diferentes tarjetas de evaluación y desarrollo, las cuales incluyen una diversidad de periféricos para aumentar su versatilidad. Un ejemplo es la tarjeta *Spartan 3E Starter Board*, manufacturada y distribuida por Digilent Inc [2], en la figura 1 se muestra la vista superior de esta tarjeta. Esta tarjeta es una alternativa de bajo costo (\$149.00 USD)

para el diseño y evaluación de sistemas de control, principalmente por la inclusión del DAC y ADC.

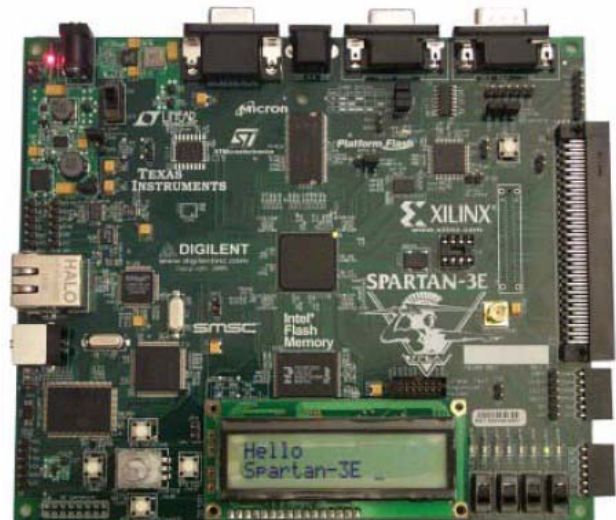


Fig. 1 Vista superior de la tarjeta *Spartan 3E Starter Board*.

Las características principales de la tarjeta son [3]:

- FPGA XC3S500E, familia Spartan-3E de Xilinx.
- Memoria Flash de 4 Mbit para configuración.
- CPLD XC2C64A, familia CoolRunner.
- DDR SDRAM de 64 MByte (512 Mbit), interfaz x16, 100+ MHz.
- Memoria Flash 16 MByte (128 Mbit) para aplicaciones.
- Memoria Flash 16 Mbits acceso serial, via SPI.
- Pantalla LCD de 16 caracteres por 2-líneas.
- Puerto PS/2 y Puerto VGA.
- Capa física Ethernet 10/100.
- Dos puertos RS-232 de 9 terminales.
- Interfaz USB para descarga y depuración
- Oscilador de 50 MHz.

- Convertidor Digital a Analógico SPI de cuatro salidas (DAC), con resolución de 12 bits.
- Convertidor Analógico a Digital SPI de dos entradas (ADC) con resolución de 14 bits y pre-amplificador con ganancia programable.
- Botón rotatorio.

Motivados por el interés práctico de desarrollar sistemas empotrados que permitan la implementación de algoritmos de control para sistemas físicos reales a bajo costo, en éste trabajo se presenta el desarrollo de los módulos necesarios en la tarjeta anteriormente descrita. Este documento está organizado como sigue: en la Sección II se muestra la organización de la plataforma desarrollada; en la Sección III se presenta el desarrollo de la interfaz SPI que permite el manejo de los convertidores ADC y DAC; en la Sección IV se presenta la interfaz de usuario; la Sección V está dedicada al desarrollo del módulo para el algoritmo de control a evaluar; la Sección VI muestra la plataforma experimental que se integró para la realización de pruebas; la sección VII presenta los resultados experimentales obtenidos y finalmente en la sección VIII se dan algunas conclusiones.

II. ORGANIZACIÓN DE LA PLATAFORMA

La plataforma se diseñó con base en la metodología Top-Down [4], en el nivel de mayor jerarquía se tienen los módulos mostrados en la figura 2.

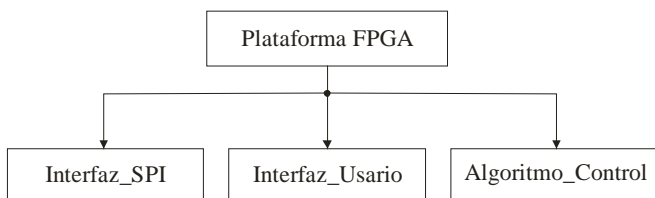


Fig. 2 Módulos principales en la plataforma

El módulo *Interfaz_SPI* establece una interfaz para manipular a 3 dispositivos de la tarjeta, el DAC, el ADC y el pre-amplificador con ganancia programable. El protocolo SPI establece un esquema Maestro-Eslavos manejado por 3 señales: MOSI (*Master Output, Slave Input*),

MISO (*Master Input, Slave Output*) y SCK (Reloj del Sistema). En este caso el maestro es el FPGA y los esclavos son: El DAC LTC2624, el ADC LTC1407A-1 y el pre-amplificador LTC6912-1, todos manufacturados por Linear Technology. Puesto que los esclavos comparten el bus SPI, además de estas tres señales cuentan con señales propias para su habilitación, de manera que no existan conflictos para un dispositivo mientras otro se está utilizando. En la figura 3 se ilustran las conexiones que se establecen entre el módulo y los esclavos. Sólo se maneja una salida analógica, su valor digital se coloca en D_OUT, y una entrada analógica, su valor digital es obtenido en D_IN.

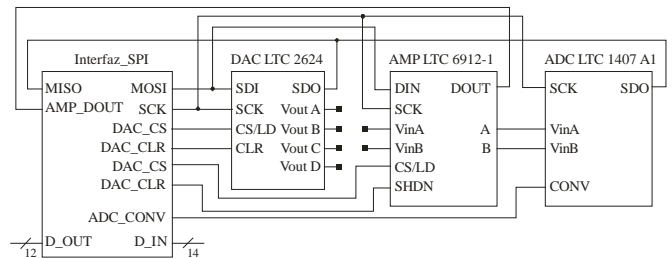


Fig. 3 Dispositivos manejados por el módulo Interfaz_SPI

El módulo *Interfaz_Usuario* tiene por objetivo manipular la pantalla LCD para mostrar 3 datos importantes en cualquier sistema de control: La referencia, el valor actual de la planta y el error. Con este módulo también se maneja al botón rotatorio e interruptores, necesarios para definir la referencia y los LED's, únicamente para mostrar el signo del error. El LCD se emplea utilizando una interfaz de 4 bits. En la figura 4 se muestran sus entradas y salidas, relacionadas con los periféricos correspondientes.

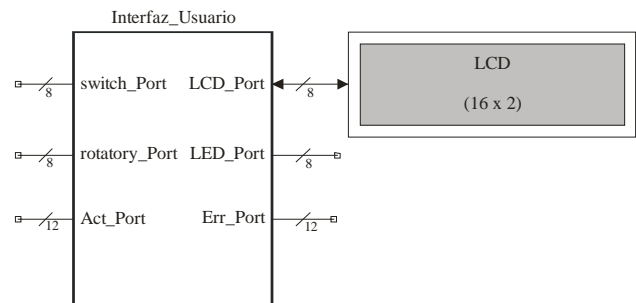


Fig. 4 Periféricos manejados por el módulo Interfaz_Usuario

El módulo Algoritmo_Control recibe como entradas la magnitud (12 bits) y signo del error para que, con base en el algoritmo programado genere la señal que actuará sobre la planta (12 bits). El error proviene del módulo Interfaz_Usuario y la salida que actúa sobre la planta lo hace a través del módulo Interfaz_SPI.

En la figura 5 se muestra la relación existente entre los tres módulos que conforman la plataforma desarrollada en el FPGA.

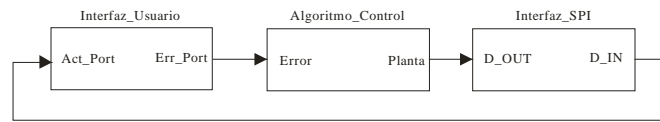


Fig. 5 Relación entre módulos dentro del FPGA

III. MÓDULO INTERFAZ_SPI

La interfaz SPI basa su operación en registros de desplazamiento. El maestro determina la velocidad y marca el inicio de la transferencia. Los datos recibidos por el esclavo son retornados para verificación. Esto se ilustra en la figura 6, donde *ss* es para seleccionar un esclavo en particular, dado que pueden ser varios los que comparten la interfaz.

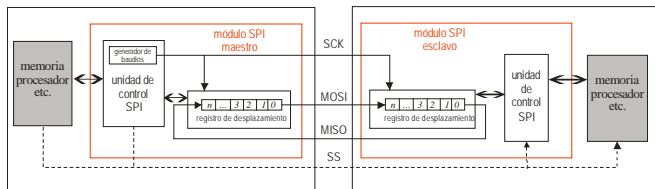


Fig. 6 Comunicación vía SPI

Este módulo se planteó con base en registros de desplazamiento y un control basado en una máquina de estados, en [5] se encuentran notas de aplicación en donde se manipula a los dispositivos SPI empleando al Picoblaze. Se prefirió desarrollar un módulo propio para aprovechar la frecuencia de 50 MHz generada por el oscilador incluido en la tarjeta, frecuencia soportada por el ADC y el DAC.

Puesto que son 3 los dispositivos a manejar y el número de bits difiere entre ellos, para cada uno se requirió de un registro de desplazamiento con sus bits de carga y desplaza, luego, un módulo de

control se encarga de habilitarlos e ir encausando la información a la terminal de salida (MOSI). En la figura 7 se observa la organización del módulo Interfaz_SPI. Todos los componentes se desarrollaron en VHDL.

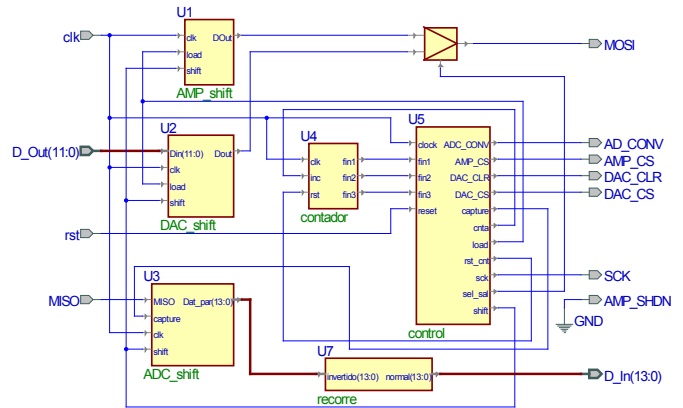


Fig. 7 Organización del módulo Interfaz_SPI

El contador indica al control cuando se han realizado los desplazamientos requeridos por cada uno de los dispositivos.

El control es quien irá generando las tramas SPI e irá activando a los dispositivos para que solo trabaje uno a la vez. En la figura 8 se muestra la trama empleada para el DAC. Para conocer el estado de las señales y el número de bits de cada componente, es necesario revisar sus hojas de especificaciones [6, 7, 8].

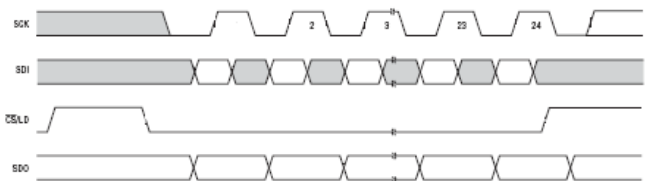


Fig. 8 Trama utilizada para el manejo del DAC

El control se basa en una máquina de estados, básicamente se tienen 4 estados jerárquicos, éstos se muestran en la figura 9. En el estado *inicio* el control coloca señales inofensivas en espera que los dispositivos se estabilicen tras un reinicio en la alimentación. Luego se tiene al estado *AMP*, el cual incluye la secuencia para configurar al pre-amplificador con una ganancia unitaria. Posteriormente se tiene una secuencia infinita entre el

estado *ADC* y el estado *DAC*, con lo que se consigue que ambos convertidores trabajen bajo un esquema de carrera libre.

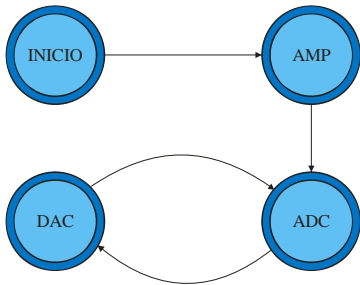


Fig. 9 Comportamiento del control

El estado jerárquico *ADC* requiere de 18 estados simples, de los cuales 5 se ejecutan 14 veces, por lo tanto, requiere de 83 ciclos de reloj. El dato digital se captura cuando han concluido las conversiones. El estado *DAC* únicamente emplea 7 estados, pero 2 de ellos se ejecutan 23 veces. Por lo tanto, son necesarios 51 ciclos de reloj, el dato a convertir queda atrapado durante el proceso de conversión.

Esto significa que cada iteración en el lazo infinito va a emplear 134 ciclos de reloj. Puesto que el FPGA trabaja a una frecuencia de 50 MHz, las conversiones se están realizando a una frecuencia de 373.134 KHz.

IV. MÓDULO INTERFAZ_USUARIO

Este módulo basa su funcionamiento en el procesador Picoblaze. El cual es un procesador de 8 bits que puede implementarse en dispositivos Spartan-3, Virtex-II o dispositivos más recientes. También es conocido como KCPSM3, es de distribución libre y fue desarrollado por Ken Chapman; ingeniero de desarrollo de Xilinx, Inc. Aunque puede ser usado para procesamiento de datos, generalmente es empleado para aplicaciones que requieren una compleja máquina de estados. Un factor importante es su tamaño, ocupa solo el 3% de los recursos en un dispositivo XC3S500 [1].

Requiere de un único bloque de RAM para formar una ROM en donde es posible almacenar un programa de hasta 1024 instrucciones. Los programas con requerimientos mayores pueden ser

redireccionados de manera que utilicen múltiples procesadores, cada uno con su respectivo bloque de memoria, para distribuir tareas en diversos sistemas. Su desempeño es de aproximadamente 43 a 66 millones de instrucciones por segundo (*MIPS, Millions Instruction Per Second*) y el núcleo puede acondicionarse para su síntesis en VHDL o Verilog.

Su repertorio de instrucciones contiene los comandos suficientes para procesamiento de datos de 8 bits y está formado por 42 instrucciones con que puede ejecutar instrucciones lógicas, aritméticas y de transferencia. Incluye memoria de datos, 16 registros de propósito general, banderas, unidad aritmética lógica (*ALU, Arithmetic Logic Unity*), una interrupción externa, la posibilidad de manejar hasta 256 puertos de entrada y 256 puertos de salida [1].

En la figura 10 se muestra al procesador Picoblaze con su correspondiente memoria de código. Pueden observarse las señales disponibles para acondicionar puertos. Los puertos son necesarios para el manejo del LCD, botón rotatorio, interruptores y botones, conocer el estado actual de la planta (desde el ADC, 14 bits) y proporcionar el valor del error (12 bits).

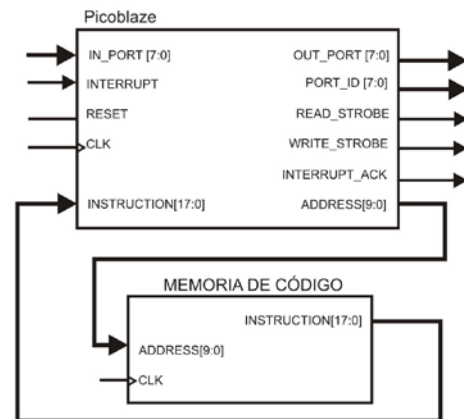


Fig. 10 El núcleo Picoblaze y su memoria de código

Los puertos se agregan codificándolos en VHDL y en total se requiere de 9, 5 puertos de entrada y 4 de salida. El estado actual de la planta y el error se maneja en 2 puertos, dado que los puertos son de 8 bits. En la tabla 1 se listan todos los puertos.

Las direcciones de los puertos de entrada no tienen bits en común con las direcciones de los puertos de salida, por lo tanto, el acondicionamiento de los 5 puertos de entrada básicamente se consigue con un multiplexor, como se muestra en la figura 11. La selección de la entrada se determina con los 3 bits menos significativos de Port_ID.

Puerto	Dir.	E/S	Uso
switch_port	00	E	4 botones y 4 interruptores
rotary_port	01	E	Codificador Rotatorio
LCD_input_port	02	E	Datos del módulo LCD
Act_port_LOW	03	E	Estado actual (byte bajo)
Act_port_HIGH	04	E	Estado actual (byte alto)
LCD_output_port	10	S	Datos y Control del LCD
Err_port_LOW	20	S	Error (byte bajo)
Err_port_HIGH	40	S	Error (byte alto)
LED_port	80	S	Manejo de los LEDs

Tabla 1. Puertos requeridos

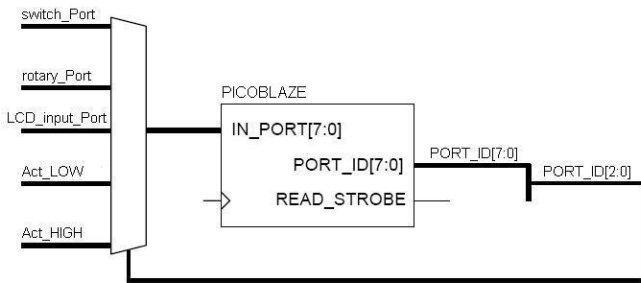


Figura 11. Puertos de Entrada

Para los puertos de salida, el nibble más significativo cambia 1 bit para diferenciar a cada uno de los puertos, por lo tanto, la escritura en un registro queda determinada por el habilitador de escritura y el bit que distingue al puerto. Esto se muestra en la figura 12.

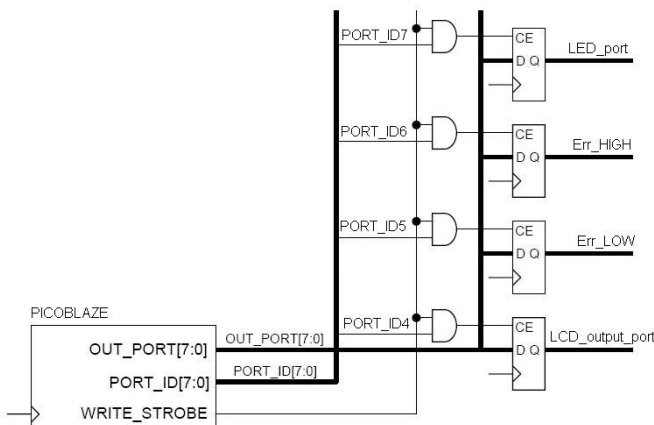


Fig. 12 Puertos de salida

En relación con el LCD, el programa que comande al Picoblaze podrá disponer de dos direcciones, una para un puerto de salida y la otra para un puerto de entrada. Sin embargo, realmente el LCD [9] es conectado con el FPGA a través de un bus bi-direccional de datos, además de las señales de control, por lo tanto, debe hacerse el acondicionamiento mostrado en la figura 13.

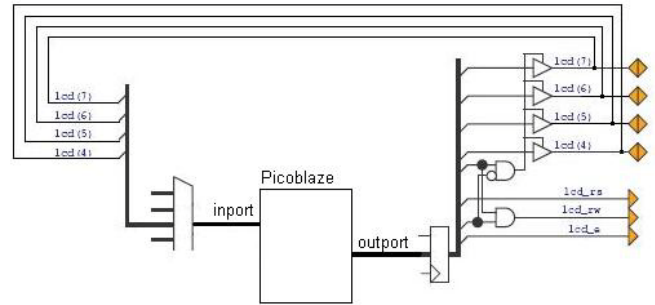


Figura 13. Puertos bidireccional para el LCD

Con la señal *lcd_rw* se determina si se escribirá o leerá en el LCD, con 0 se escribe y 1 se lee. Se agrega la señal interna *lcd_drive*, para que el LCD solo trabaje cuando es habilitado. El código VHDL correspondiente es:

```

lcd_rw <= lcd_rw_control and lcd_drive;
lcd_d <= lcd_output_data when
    (lcd_rw_control='0' and lcd_drive='1')
    else "ZZZZ";
    
```

Para el botón rotatorio también debe hacerse un acondicionamiento de hardware [10], esto para que los giros se detecten inmediatamente por interrupción, y por el estado de los botones se conozca la dirección del giro. El eje no está centrado, con ello hace que dos botones sean presionados en diferente orden, esto se muestra en la figura 14.

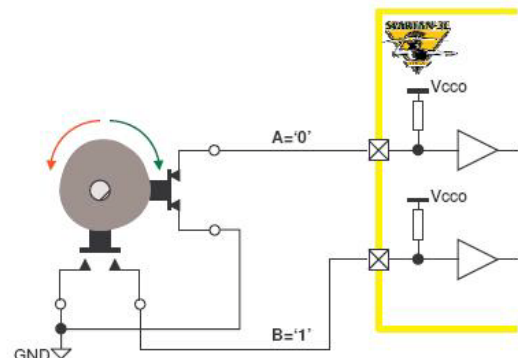


Fig. 14 Apariencia del botón rotatorio

En la figura 15 se observa como por la forma de las señales puede determinarse la dirección del movimiento.



Fig. 15 Resultado de los giros del botón rotatorio

La forma inmediata para detectar la dirección de rotación es: Si 'B' va a un nivel Alto mientras 'A' está en Bajo, la rotación es a la Izquierda, y si 'A' va a un nivel Alto mientras 'B' está en Bajo, la rotación es a la Derecha. Sin embargo, la misma irregularidad del eje hace que se inserten cambios inesperados, por lo que debe establecerse un filtro, de manera que una vez que se detecte un cambio en una señal, la actividad subsecuente en ella sea ignorada hasta que ocurra un cambio en la otra señal.

El filtro modifica dos señales internas (q1 y q2), que posteriormente son útiles para indicar si hubo algún movimiento y su dirección. En la figura 16 se observa como en q1 se aplica una XNOR, mientras que en q2 una XOR. Anulando los rebotes de los botones.



Fig. 16 Un filtro elimina rebotes y modifica señales internas

En la figura 17 se muestra la aplicación del filtro en diferentes movimientos, puede notarse que un flanco de subida en q1 indica que se realizó un

movimiento y el nivel de q2 indica si el movimiento fue a la izquierda.

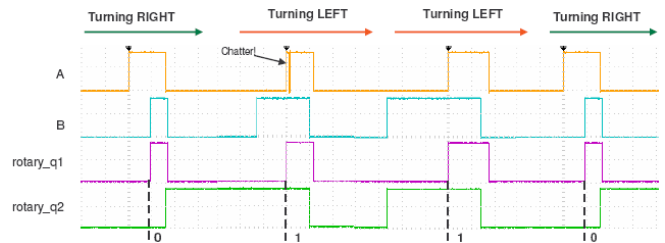


Fig. 17 Señales intermedias

Estas señales son las que se utilizan para detectar los movimientos y su dirección, acondicionando el hardware mostrado en la figura 18.

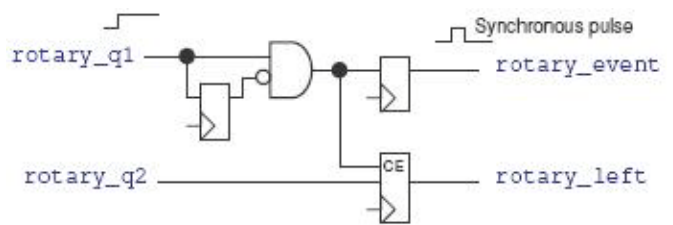


Fig. 18 Las señales intermedias interrumpen al Picoblaze

Dentro de la plataforma, se utiliza al botón rotatorio para aumentar o reducir la referencia, en un factor determinado por los interruptores. Esto se hace en la rutina que atiende a la interrupción, en la figura 19 se muestra esta evaluación.

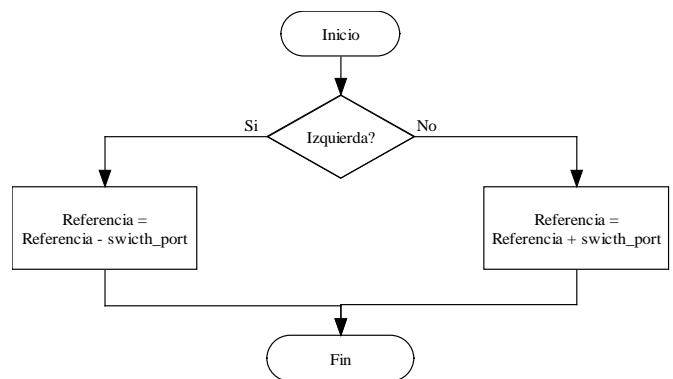


Fig. 19 Rutina que da servicio a la interrupción

El programa que ejecuta el Picoblaze en el módulo Interfaz_Usuario sigue el diagrama de flujo ilustrado en la figura 20. Los datos dentro del programa se manejan como números binarios de

12 bits, sin embargo, al momento de hacer la impresión en el LCD, se normalizan a decimal.

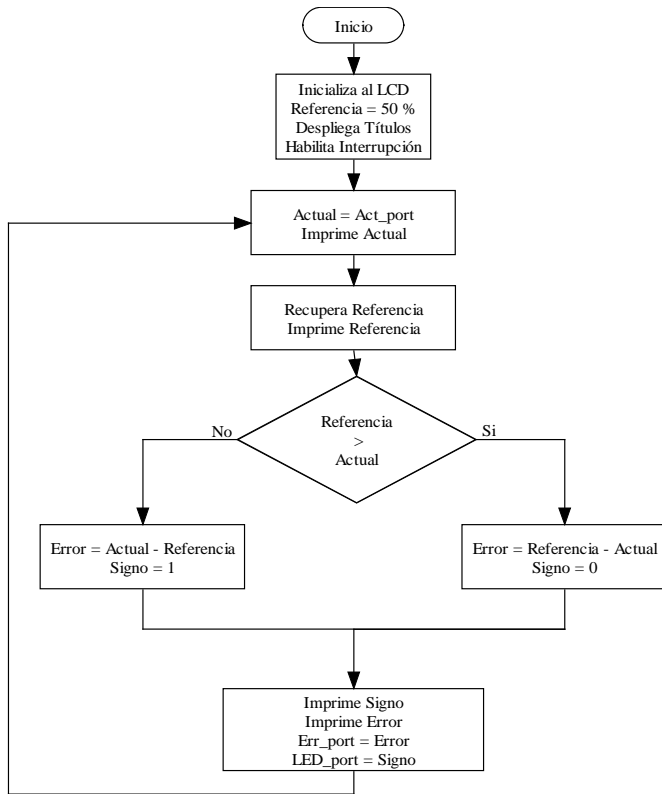


Fig. 20 Programa principal en el Picoblaze

V. MÓDULO ALGORITMO_CONTROL

Básicamente se implementó un algoritmo de control proporcional, para después probarse en un sistema de control de temperatura. La planta incluye un calefactor y un sensor. Sin embargo, al actuador no se le pueden suministrar voltajes negativos, por lo que este módulo básicamente realiza el algoritmo mostrado en la figura 21.

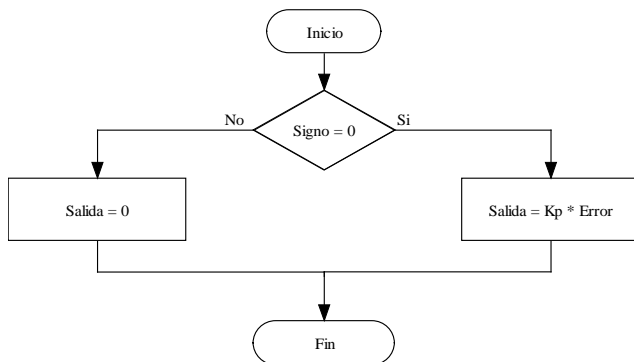


Fig. 21 Flujo del módulo Algoritmo_Control

VI. PRUEBA DE LA PLATAFORMA

La plataforma con el algoritmo de control proporcional se probó sobre el sistema LTR701, el cual es un sistema de control de temperatura desarrollado por Amira [11]. En la figura 22 se observa la planta a controlar.



Fig. 22 Planta empleada para evaluar la plataforma

El actuador de la planta trabaja en un rango entre 0 y 10 V, y la salida del DAC integrado en la tarjeta está en un rango entre 0 y 3.3 V, por lo tanto, fue necesario realizar un amplificador con ganancia de 3, cuyo diagrama electrónico se muestra en la figura 23.

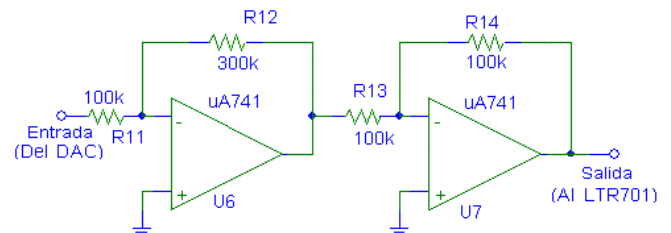


Fig. 23 Circuito de acondicionamiento para el actuador

El sensor de la planta entrega 0.1 volts por grado centígrado. Puesto que el valor máximo para la temperatura es de 100° C, el voltaje proporcionado por el sensor está entre 0 y 10 V, en contraste con el voltaje que puede recibir el ADC, el cual está en un rango entre 0.4 y 2.9 V, por lo tanto también se requirió de un circuito de acondicionamiento, consistente en un reductor de voltaje por un factor de 4 seguido por un sumador para que el valor mínimo sea de 0.4, el diagrama de este circuito se muestra en la figura 24.

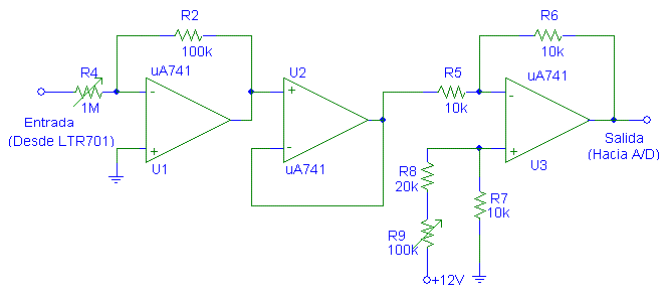


Fig. 24 Circuito de acondicionamiento para el sensor

En la figura 25 se muestra el montaje de la plataforma desarrollada con sus circuitos de acondicionamiento y su conexión con la planta.

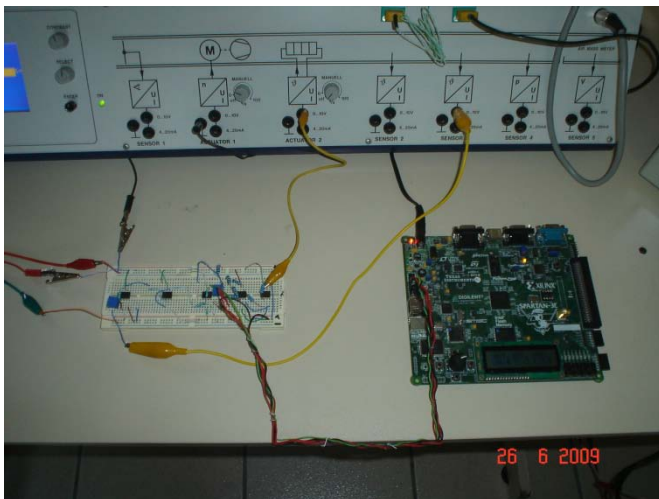


Fig. 25 Conexión de la plataforma con la planta, utilizando los circuitos de acondicionamiento

Los circuitos de acondicionamiento se basan en Amplificadores Operacionales, por lo tanto, se requirió de una fuente de alimentación simétrica (+/- 12 V), así como la fuente de 5 V para la tarjeta de desarrollo. El programa del FPGA se grabó en su memoria Flash, con ello, no se pierde su configuración aún en ausencia de alimentación.

VII. RESULTADOS

Se probó la plataforma utilizando una constante y $K_p = 1$ en el módulo Algoritmo_Control, la plataforma funcionó correctamente, ante una referencia de 50°C partiendo de la temperatura ambiente, tiende a minimizar el error, aunque no lo nulifica, dado que solamente se utilizó la contribución proporcional. En la figura 26 se

muestra la gráfica con el comportamiento de la temperatura que puede visualizarse en la misma planta.

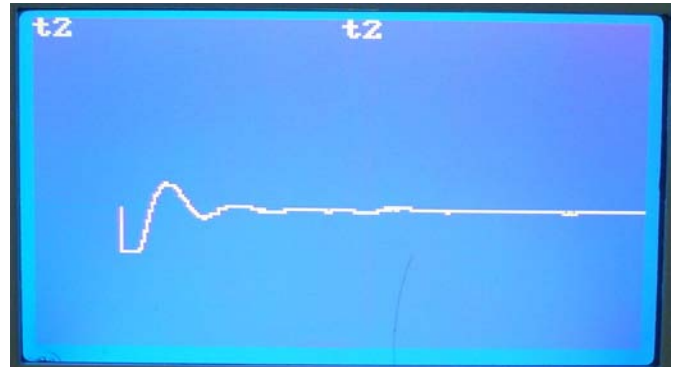


Fig. 26 Resultados observados en la planta

En la figura 27 se muestra la información que se presenta en la plataforma de desarrollo, la cual coincide con la información que entrega la planta.



Fig. 27 Resultados en la pantalla LDC de la plataforma

Una vez que se verificó el funcionamiento de la plataforma, se conectó una tarjeta de adquisición NI-6024E de National Instruments, manipulada con el RTWT de Matlab/Simulink y se hicieron pruebas con diferentes valores para K_p , en las figuras 28 a la 30 se muestran los resultados experimentales obtenidos para los algoritmos de control proporcional evaluados. Se muestra el estado actual de la planta, el proporcionado por el sensor y la señal de control generada por la plataforma.

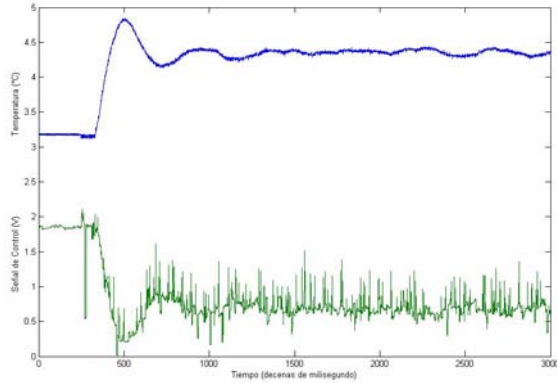


Fig. 28 Comportamiento del sistema con $K_p = 1$

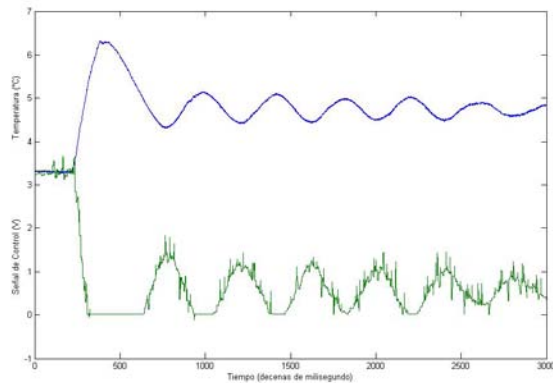


Fig. 29 Comportamiento del sistema con $K_p = 2$

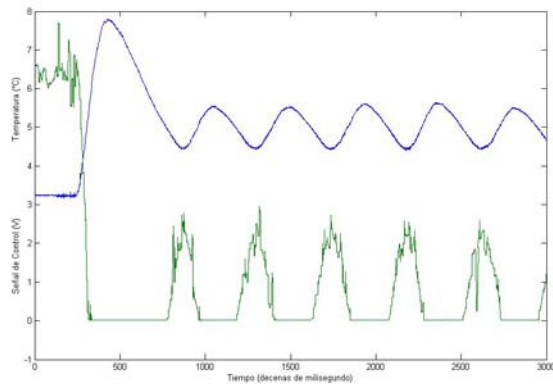


Fig. 30 Comportamiento del sistema con $K_p = 8$

En la figura 28 puede verse que aunque el sistema se estabiliza, presenta un error en estado estable. En la figura 30 se observan demasiado sobretiro y oscilaciones alrededor de la temperatura deseada, esto concuerda con la teoría [12], dado que un control proporcional no es suficiente para alcanzar un error cero sin evitar que el sistema oscile.

VIII. CONCLUSIONES

La aplicación de FPGA's en el área de control automático proporciona una herramienta flexible, rápida, de bajo costo y con alta capacidad de procesamiento, sin embargo, es necesario el desarrollo de una suite de módulos que faciliten el diseño de sistemas.

Los módulos desarrollados son transportables a FPGA's Xilinx de diferentes densidades y con excepción del Picoblaze, la transportabilidad se puede dar con FPGA's de diferentes fabricantes, esto por el uso de VHDL, el cual es un lenguaje estandarizado. Aunque el Picoblaze está elaborado en VHDL, su arquitectura utiliza una definición de bajo nivel, que explota las características únicamente en FPGA's de Xilinx.

La densidad en los FPGA's actuales proporciona un medio adecuado para desarrollar sistemas empotrados, en la tabla 2 se muestran los principales recursos empleados por la plataforma, se observa que aún hay muchos elementos disponibles para implementar algoritmos de control con mayor complejidad.

Recurso	Usado	Disponible	Uso
Slices	196	4,656	4 %
Ent/Sal	39	232	16 %
Bloques RAM	1	20	5 %
GCLKs	1	24	4 %

Tabla 2. Recursos empleados por la plataforma

El algoritmo de control empleado como prueba es demasiado simple, pero este no es el aspecto relevante en el presente trabajo, lo relevante es que la plataforma queda disponible y con ella en un futuro inmediato se plantea comparar el comportamiento de un controlador PID digital con un controlador difuso, o bien aplicar la plataforma en otras plantas que se tienen disponibles en el laboratorio de control de la Universidad Tecnológica de la Mixteca.

Para el diseño de un sistema secuencial dentro de un FPGA se tienen dos alternativas: organización modular bajo una metodología Top-Down o bien, el uso de un procesador empotrado, como el

Picoblaze. La elección de una u otra está en función de los requerimientos. Para el módulo Interfaz_SPI se prefirió el desarrollo de módulos propios para que los convertidores trabajen a su frecuencia más alta. Para la Interfaz_Usuario se optó por el uso del Picoblaze, porque en caso contrario se requerirían máquinas de estado muy complejas para permitir todos los comandos del LCD. Además, la información se ha normalizado para presentarla de manera decimal (y no en binario, como internamente se maneja). Esto requirió un algoritmo para la división aritmética dentro de la función que imprime en el LCD.

IX. REFERENCIAS

- [1] Chapman, Ken: “PicoBlaze KCPSM3 8 bit Microcontroller”; Xilinx, Inc; June, 2005.
- [2] Digilent Inc., Digital Design Engineer’s Source < <http://www.digilentinc.com/> >
- [3] Spartan-3E Starter Kit Board User Guide, Xilinx, Inc., version 1.0a, March 9, 2006
- [4] ROMERO, René: “Sistemas Digitales con VHDL”; Legaria Ediciones
- [5] Xilinx: Prototyping and Development Boards; Spartan-3e Starter Kit Design Examples, http://www.xilinx.com/products/boards/s3e_starter/reference_designs.htm
- [6] LTC2624 Quad DAC Data Sheet <http://cds.linear.com/docs/Datasheet/2604fd.pdf>
- [7] LTC6912 Dual Programmable Gain Amplifiers with Serial Digital Interface <http://cds.linear.com/docs/Datasheet/6912fa.pdf>
- [8] LTC1407A-1 Serial 14-bit Simultaneous Sampling ADCs with Shutdown <http://cds.linear.com/docs/Datasheet/14071fa.pdf>
- [9] Chapman, Ken: “Initial Design for Spartan-3E Starter Kit (LCD Display Control)”, Xilinx Inc., February 2006
- [10] Chapman, Ken: “Rotary Encoder Interface for Spartan-3E Starter Kit”, Xilinx Inc., February 2006
- [11] Manual de usuario del sistema LTR701.
- [12] Katsuhiko, Ogata: “Ingeniería de control utilizando Matlab”, Prentice Hall, España, 1998.

X. AUTORES

M. C. Felipe Santiago Espinosa es Maestro en Ciencias con especialidad en Electrónica por parte del INAOE, incorporado al departamento de Electrónica de la Universidad Tecnológica de la Mixteca, desde 1998. Sus intereses de investigación están relacionados con Sistemas Digitales y Arquitectura de Computadoras.

M. C. Jorge Luis Barahona Ávalos es Maestro en Ciencias de la Electrónica Opción en Automatización por parte de la BUAP, incorporado al departamento de Mecatrónica de la Universidad Tecnológica de la Mixteca, desde 2008. Sus intereses de investigación están relacionados con control y automatización.

DESARROLLO DE UNA HERRAMIENTA DE APOYO PARA EL APRENDIZAJE DEL MICROCONTROLADOR 16F84A.

Pérez Merlos J.C., Portillo Medina J.C., Salgado Gallegos M., Albarrán Trujillo S.

Universidad Autónoma del Estado de México.

Facultad de Ingeniería.

División de Electrónica.

Dirección: cerro de Coatepec. s/n Ciudad Universitaria, Toluca Edo. De México.

Teléfono: (01) (722) (2-14-08-55) ext. 111, (2-15-14-53) Fax (2-15-45-12)

Correo electrónico: jcjc63@yahoo.com, jcpm_74@yahoo.com.mx, seat@uaemex.mx

RESUMEN

En la Facultad de Ingeniería de la Universidad Autónoma del Estado de México existen las carreras de ingeniería electrónica y computación, en ambas se imparten asignaturas que contienen temas de sistemas digitales en las cuales hay tópicos relacionados a microprocesadores y microcontroladores. Los microcontroladores son muy populares por su facilidad y versatilidad de aplicación en sistemas electrónicos, ya que permiten automatizar trabajos rutinarios de manera sencilla, asimismo el espacio que ocupan en dichos sistemas es pequeño, debido a que internamente contienen muchos periféricos que son programables, lo cual hace que tengan un gran potencial en sistemas de desarrollo, por lo que es muy importante que los estudiantes los conozcan y manejen con habilidad. En este contexto el siguiente trabajo describe el “desarrollo de una herramienta de apoyo en el aprendizaje del microcontrolador 16F84A”, el cual consta de un programa multimedia y un panel de prácticas, en éste se pueden realizar varios ejercicios relacionados a la interconexión del microcontrolador con diferentes interfaces. Además, se pretende proporcionar al alumno de material de buena calidad, que le permita consolidar los conocimientos adquiridos en clase o incluso generar un auto aprendizaje.

Palabras Claves: Multimedia, microcontrolador, aprendizaje del microcontrolador, tutorial multimedia.

I. INTRODUCCIÓN

Es un hecho, que actualmente la electrónica digital ha tenido un gran desarrollo especialmente en el área relacionada con los microcontroladores. Éstos son dispositivos que internamente tienen periféricos como memoria, convertidor analógico digital y digital analógico, comparadores, comunicación serie, etc., lo cual permite una versatilidad y facilidad de aplicaciones con ellos, aunado a lo anterior, ayudan al ahorro de espacio en el desarrollo electrónico debido a la integración de componentes que tienen y la programación de éstos. La aplicación del microcontrolador está por doquier en la industria automotriz, en el hogar, en equipo de medición, etc. Es por ello que estudiantes de áreas afines a la electrónica deben conocer y manejar con habilidad a los microcontroladores.

Existen herramientas de desarrollo y manuales que cada empresa de microelectrónica genera cuando un nuevo dispositivo sale al mercado. Los manuales describen toda la información pertinente pero en muchos casos no es accesible o son poco entendibles. El uso de simuladores y emuladores también son una herramienta de apoyo para el aprendizaje pero se tiene que contar con una persona como asesor, debido a que estos sistemas son para simular los programas y emular los resultados reales por lo que la persona debe ya tener conocimientos previos sobre ellos, éstos no son muy recomendables para estudiantes que inician en el área y que desean aprender de manera autodidacta, además el emulador comúnmente no

tiene un precio económico. Lo anterior, no está dirigido para estudiantes principiantes de electrónica.

Por otro lado, los nuevos escenarios de aprendizaje propiciados por la tecnología de información y comunicación implican cambios organizacionales en cuanto a combinación de estos escenarios y los servicios integrados de aprendizaje [1]. Las tendencias actuales en la formación superior están promoviendo cierta evolución hacia modalidades de aprendizaje abierto, con una oferta educativa flexible, que sirva tanto para aquellos alumnos que siguen la enseñanza presencial, como aquellos que la siguen a distancia [2]. Ello requiere de modelos pedagógicos nuevos y un fuerte apoyo de tecnologías multimedia interactiva [3][4].

Wolfgang [5], define multimedia como la combinación de múltiples aparatos técnicos que permiten presentar la información en múltiples formatos como textos, imágenes realistas o gráficos, a través de múltiples modalidades sensoriales. Los materiales didácticos universitarios tiene que formar verdaderos paquetes didácticos integrados en los formatos antes mencionados para un doble uso, dirigidos a estudiantes tanto presenciales como para los que no pueden estar físicamente presentes [6].

Es indudable que la aparición de los medios de masas (como la radio y la t.v.) han afectado la forma en que los ciudadanos aprenden. Sin embargo, el desarrollo de estos medios no ha afectado de manera profunda a la institución educativa [2]. Asimismo, en disciplinas técnicas como la electrónica, la robótica, etc., se hace necesario el acercamiento a la realidad y a las sesiones demostrativas, en este sentido la herramienta multimedia es un esfuerzo de apoyo para estudiantes favoreciendo el autoaprendizaje.

Varias empresas e instituciones de educación hoy en día trabajan con sistemas multimedia para la enseñanza de diferentes áreas, desde cursos de inducción hasta Matemáticas y Medicina, pero comúnmente el material es para nivel medio

superior. Cuando se encuentran aplicaciones multimedia para nivel superior son sólo para temas básicos. En ingeniería, el material multimedia o también llamado e-learning es limitado, debido muchas veces a la complejidad de los temas avanzados que se manejan en esta área. Actualmente el uso de las tecnologías de información está cambiando el paradigma de la educación y esto se debe al avance de los dispositivos móviles, que ya empiezan a ser usados con recursos multimedia llamado m-learning [7].

Cabe señalar, que no todos los estudiantes aprecian este tipo de herramienta educativa lo cual parece estar relacionado con lo que los expertos llaman estilos cognitivos o de aprendizaje (teóricos, activos, reflexivos, pragmáticos) [8][9]. Woodhead [10] menciona que una clasificación de los programas multimedia son los tutoriales y que son los más utilizados a nivel superior. En este sentido, Ponsa [11] es su trabajo describe un multimedia para robótica, pero como una traducción de apuntes de clase a la modalidad multimedia.

También, Vallejo [12] realizó un video para microncontroladores, pero éste no proporciona información suficiente para el nivel de ingeniería y sólo propone una práctica, además el formato usado para el video es VHS. Con base a lo anterior, se puede decir, que son pocos los materiales que se realizan para asignaturas avanzadas en ingeniería.

Por otro lado, en la Facultad de Ingeniería de la UAEM, en las carreras de Ingeniería Electrónica y de Computación hay asignaturas donde se imparten sistemas digitales y se abordan los temas relacionados a microprocesadores y microcontroladores, éstos se enseñan de manera tradicional. Por lo que, tratando de aportar un complemento al estudiante y un apoyo al profesor, este trabajo describe una herramienta multimedia para el aprendizaje del microcontrolador 16F84A.

II. METODOLOGÍA

- Investigar y analizar la información sobre el PIC 16F84A.
- Seleccionar la información apropiada correspondiente al microcontrolador.
- Seleccionar el ambiente de programación para el desarrollo del tutorial usando multimedia.
- Diseñar la página para el vaciado de la información, seleccionar el tipo de navegación, así como la presentación de la información, colores, videos, sonidos etc.
- Realizar pruebas tanto de la funcionalidad del programa, como de la apreciación de los estudiantes, como una herramienta de apoyo para aprendizaje del microcontrolador.
- Seleccionar las prácticas para el panel.
- Diseñar y construir el panel de prácticas.
- Realizar las pruebas de funcionamiento de todas las prácticas seleccionadas.
- Redactar conclusiones.

- Que incluya diferentes tipos de interfaces.
- Que integre todo lo necesario para trabajar en situ.
- Que tenga una presentación aceptable y que sea funcional.

Programa multimedia.

Partiendo de lo anterior, el programa multimedia fue desarrollado pensando en alumnos principiantes, por ello sólo se enfocó al microcontrolador PIC 16F84 de Microchip, con el fin de no confundir a un estudiante que apenas comienza en este tema. El material fue hecho en Auto Play Media Studio Professional versión 6, debido a su facilidad de integrar diferentes tipos de archivos, ya que varias partes que conforman al tutorial fueron realizadas con otras herramientas como Corel Draw (para líneas, texturas), Macromedia Flash MX (para gráficos así como animaciones), Camtasia Studio (para videos demás de animaciones) y Adobe Audition (para creación y edición de sonidos).

III. DESARROLLO

El proyecto se desarrolló en dos vertientes, la primera consistió en realizar un software multimedia y la segunda, en construir un panel de prácticas. Inicialmente, se establecieron algunos criterios que debería de cubrir la herramienta a desarrollar, éstos se describen a continuación.

Sobre el programa multimedia:

- Que se instale fácilmente.
- Que se pueda navegar de manera práctica.
- Que pueda abordar el tema deseado en cualquier momento.
- Que tenga colores agradables.
- Que presente información detallada así como tips.
- Que la información más específica y delicada sea guiada.

Sobre el panel de prácticas:

- Fácil de conectar.
- Que se puedan hacer varias prácticas.
- Que sea ligero.

El desarrollo de esta herramienta está basado en el proceso de ingeniería de software aplicando el modelo de cascada, integrando los componentes del elemento humano, específicamente del alumno de ingeniería (memoria, representación del conocimiento, el pensamiento y razonamiento, percepción visual y construcción del diálogo) [13].

El material se organizó en varios temas: contenido, introducción, generalidades, programación, simulación, panel de prácticas y anexos como se muestra en la fig. 1. La página fue diseñada de tal manera, que el usuario siempre pueda acceder a cualquier tema independientemente donde esté ubicado.

Asimismo, su navegación dentro del tutorial es fácil, ya que los botones en la parte baja de la página son sólo cuatro, el de avanzar, retroceder, principal y el de salir como se puede ver en la fig. 2. Es decir, se trató de no saturar la página con tantas indicaciones. El método de navegación es del tipo no lineal. Los colores utilizados son

suaves y combinados entre azul, blanco y gris. El azul para el fondo y cuadros de relleno, el blanco para textos (títulos, temas y algunos cuadros de relleno), el gris para botones, viñetas, así como las barras de desplazamiento.

Dentro del contenido del tutorial se presentan figuras pequeñas pero éstas pueden ampliarse dando un botazono (click) en cada una ellas. Algo muy especial que distingue a este trabajo es que en la parte de programación y simulación, se presentan detalles de cómo iniciar con la plataforma de Microchip que es el MPLAB, esto se describe a través de pequeñas guías automáticas (simulaciones) que le indican al estudiante cómo debe realizar las actividades desde abrir un archivo hasta cómo editarlo y ensamblarlo.

De la misma manera, en la parte de cómo hacer un programa, se indica a través del mismo método los detalles finos de lo que debe contener un programa en ensamblador de este microcontrolador (encabezado, directivas, comentarios, etiquetas, mnemónicos, etc.). También se describe el desarrollo de 10 prácticas propuestas, las cuales fueron seleccionadas de tal manera que incluyeran conexión del microcontrolador con interfaces de diferentes niveles y tipos de voltaje (corriente directa y corriente alterna) para el manejo de los dispositivos más comunes en el área de electrónica, asimismo, se incluye desde el material a utilizar hasta la programación y el diseño de la circuitería electrónica necesaria, así como un manual de usuario del panel de prácticas.

Por otro lado, su instalación no requiere altos conocimientos de computación, simplemente se inserta el disco en la unidad de C.D. y automáticamente realiza la instalación, por lo que sólo hay que presionar el botón continuar o siguiente. Los requerimientos mínimos del tutorial son: Windows 98, Me, XP o Vista, 128 de memoria RAM, microprocesador a 350 Mhz. o superior, 200 Megabytes en disco duro, lector de C.D., tarjeta de sonido y altavoces.

Panel de prácticas.

El panel de control mostrado en la fig. 3, es una parte complementaria y de función independiente de este tutorial, se realizó con base en las declaraciones ya establecidas anteriormente, pero también con la idea de que se cubrieran las prácticas de conexión del microcontrolador con diferentes interfaces, así que, el panel cuenta internamente con dos fuentes de alimentación (una de 5 volts y otra de 24 volts), tiene en el centro una base donde se inserta el microcontrolador ya programado (esta base ya cuenta con lo necesario para el funcionamiento del microcontrolador como por ejemplo su cristal de cuarzo también llamado oscilador de cuatro mega hertz, su reinicio, su alimentación de 5 volts).

Asimismo, alrededor de éste se encuentran distribuidos los bornes de conexión para las diferentes prácticas o aplicaciones como encender con el microcontrolador un diodo emisor de luz (led), sacar información hacia visualizadores 7 segmentos, insertar datos con un teclado, realizar acciones según la posición de interruptores, control del giro y paro de motores de corriente directa, control de posición con motores paso a paso, encendido y apagado de lámparas a 110 volts corriente alterna, así también, el control de bobinas para accionar electroválvulas y desplazar y contraer el vástago de un pistón, disparar tiristores para encender y apagar gradualmente una lámpara.

Adicionalmente, se pueden mezclar los dispositivos dispuestos en el panel, que cubren las 10 prácticas ya mencionadas. Lo anterior, está integrado en una caja mitad madera (12 mm de espesor) y mitad acrílico (3 mm de espesor) de 62 cm de largo por 31.5 de ancho y 6.3 cm de alto, el acrílico fue cubierto con una calcomanía de vinil color negro, los pines de conexión de los elementos se identificaron en el panel en color blanco y de rojo los títulos de las prácticas, las conexiones son realizadas con jacks y conectores tipo bananas como lo muestra la fig. 3, la caja se conecta directamente a 110 volts de corriente alterna y todo internamente está disponible al momento.

IV. PRUEBAS Y RESULTADOS

Las pruebas realizadas al tutorial, fueron de instalación correcta (que se crearan los accesos directos y las carpetas necesarias), funcionamiento de los botones de navegación, de los videos, del sonido, de las animaciones, barras de desplazamiento, desinstalación, etc. Las pruebas al panel de prácticas desarrollado, fueron las siguientes: Encendido del panel, función de reinicio, encendido de visualizadores (displays), funcionamiento elementos luminosos como diodos emisores de luz y lámparas y de dispositivos actuadores como botones y teclado, salida de voltaje del panel, operatividad correcta de las dos fuentes de alimentación para 5 y 24 volts, protecciones y ventilación interna en el panel.

Asimismo, se verificaron cada una de las prácticas en cuanto a su funcionamiento, incluyendo al microcontrolador programado para ver que las conexiones de éste funcionaran adecuadamente.

Se realizó una comparación en relación a la información, cantidad y tipo de prácticas con el video VHS de Vallejo [12]. También, se hizo una valoración por parte de los alumnos de electrónica, la cual fue a nivel de apreciación hacia el tutorial.

V. CONCLUSIONES

El producto obtenido, es una herramienta en el área de sistemas digitales del cual el profesor o el alumno tienen ya a su disposición.

El desarrollo de materiales multimedia requiere de un grupo multidisciplinario, lo cual sólo fue posible integrarlo con profesores del área de computación y electrónica.

Se realizó el tutorial y el panel de control únicamente con el PIC 16F84A, debido a que es un microcontrolador pequeño con pocos periféricos internos, que permite a un principiante en el área comprender fácilmente cómo funciona éste. Los microcontroladores más avanzados de la misma marca trabajan igual, por lo que, escalar a ellos será más fácil.

El material aquí desarrollado va dirigido a estudiantes principiantes en el manejo de microcontroladores.

El panel de prácticas permite al profesor o estudiante realizar in situ los ejercicios propuestos con sólo conectar.

El trabajo desarrollado cumple con lo establecido al inicio, pero puede seguir desarrollándose.

La herramienta se ha utilizado por más de dos años en la Facultad de Ingeniería de la Universidad Autónoma del Estado de México, ha sido de gran utilidad y aceptada muy bien por lo alumnos.

La valoración que se hizo por parte de los estudiantes fue meramente cualitativa y no cuantitativa, pero demuestra aceptación por parte de los mismos.

VI. REFERENCIAS

- [1] Hidalgo, N.; Arrabal, M.; Salinas, J. *Centros de Recursos multimedia para la educación flexible y a distancia*, Comunicación presentada a Edutec '99, Sevilla. ISBN 84-89673-79-9, 1999.
- [2] Salinas, J., *Nuevos ambientes de aprendizaje para una sociedad de la información*, Revista Pensamiento Educativo, Pontificia Universidad Católica de Chile, 1997.
- [3] Hiltz, S.: *The Virtual Classroom, Software for Collaborative Learning*. En Barret, E. (Ed): Sociomedia. The MIT Press, Cambridge (Massachusetts), 1992.
- [4] Salinas, J., Sureda, J., *Aprendizaje abierto y educación a distancia*, Grupo Tecnología Educativa, Universidad de las Islas Baleares, Palma Mallorca, 1992.

- [5] Wolfgang Schnotz, traducción de Raúl Alelú Paz, *Aprendizaje multimedia desde una perspectiva Cognitiva*, Revista de docencia Universitaria, Vol. 2, No 2, 1999.
- [6] Lewis, R: *Open Learning- the future*. En Paine, N. (Ed) *Open Learning in transition*, London, Kogan Page, 89-104, 1988.
- [7] Álvarez Sánchez, D.; Schachter Edward M., *El teléfono móvil: una herramienta eficaz para el aprendizaje activo*, Current Developments in Technology-assisted Education, Formatex, 2006.
- [8] Honey, P., Mumford, A., *The manual of learning Styles*, ISBN 0-9508444-7-0, 1992.
- [9] Kogan, N., *Educational implications of Cognitive Styles*, G. S. Lesser, Scott Foresman, 1992.
- [10] Woodhead, N., *Hipertext and Hypermedia*, Wokingham, Addison Wesley Publishing Compa, 1990.
- [11] Ponsa, Pere; Aranda Joan, *Creación de un aplicativo multimedia en robótica*, Universidad Politécnica de Catalunya, 2002.
- [12] Vallejo, Horacio D. *Microcontroladores PIC*. Editorial QUA RK S.R.L, 2002.
- [13] S. Pressman Roger, *Software engineering A Practitioners Approach*, sixth edition, Mc. Graw Hill, 2005.
- Angulo, J. M. Martín, E. *Microcontroladores PIC. La solución en un chip. Diseño Práctico de Aplicaciones*, 1^{ra} Parte: PIC16F84. Madrid, España. Editorial ITP Paraninfo, 2^{da} edición. 1999.
- Angulo, J. M., Romero, S. y Angulo, I. *Microbótica. 1^{ra} Parte*. Argentina. Editorial Paraninfo, 1^{ra} edición. 1999.
- Bartolomé Pina R. Antonio, *Multimedia interactivo y sus posibilidades en educación superior*, Revista Pixel Bit, Número 1, Universidad de Barcelona, 1994.
- Duque C., Edison. *Curso Avanzado de Microcontroladores PIC*. Editorial Cedit. 1998.
- Iovine, John. *PIC Robotics: A Beginner's Guide to Robotics Projects Using the PIC Micro*. Editorial Mc. Graw-Hill, 2004.
- Pallás, Areny Ramón. *Adquisición y Distribución de Señales*. Barcelona, España. Editorial Marcombo, 1^{ra} edición. 1993.

VII. AUTORES

M. en I. Juan Carlos Pérez Merlos es Ingeniero en Electrónica y obtuvo la Maestría en Ingeniería en Informática en la Facultad de Ingeniería de la UAEM.

Ing. Juan Carlos Portillo Medina es Ingeniero en Electrónica y es pasante de la M en A. con especialidad en Contabilidad Gerencial por la Facultad de Contaduría y Administración de la UAEM.

M. en I. Mireya Salgado Gallegos es Ingeniero en Computación con Maestría en Informática por la Facultad de Ingeniería de la UAEM.

M. en A. Silvia Edith Albarrán Trujillo es Ingeniero en Computación y tiene la Maestría en Administración por la Facultad de Contaduría y Administración de la UAEM.

Otras consultas.

Angulo, Martínez Ignacio y Angulo Usategui José Ma. *Microcontroladores PIC. Diseño Práctico de aplicaciones*, 1ra parte: PIC 16F84, España. Editorial Mc. Graw Hill, 3ra edición, 2003.

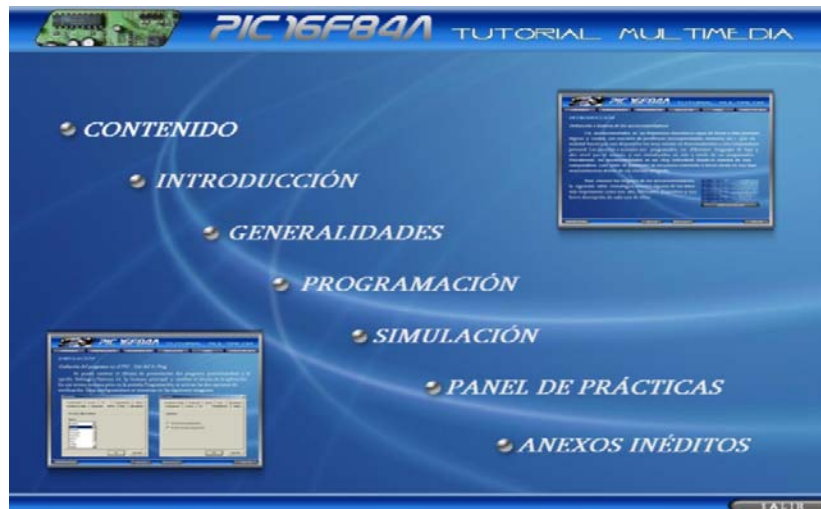


Fig. 1.- Pantalla principal del tutorial multimedia, para el microcontrolador 16F84A.

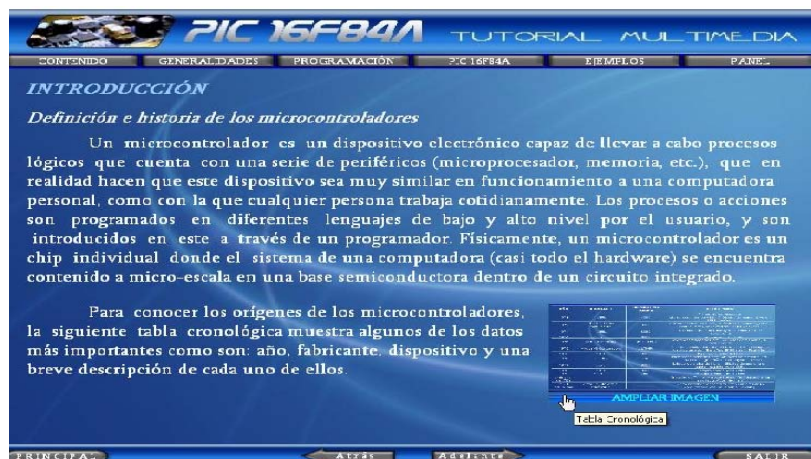


Fig.2.- Página del tutorial, donde se muestran los temas que puede acceder el usuario, los botones de navegación y una imagen la cual puede ser ampliada.



Fig. 3.- Se muestra el panel de prácticas, con las diferentes conexiones para los 10 ejercicios propuestos en el tutorial.

SIMULADORES DE CONTROL SECUENCIAL USADOS PARA LA EDUCACIÓN A DISTANCIA

Garibay Martínez Alberto Ramiro, Garibay Jiménez Ricardo.

Departamento de Ingeniería de Control, Departamento de Ingeniería Eléctrica, Facultad de Ingeniería
Universidad Nacional Autónoma de México
Tel. 56223130, e-mail: agaribay@ctrl.fi-b.unam.mx, rgaribay@servidor.unam.mx

RESUMEN

En la Facultad de Ingeniería de la UNAM, los programas institucionales de apoyo al desarrollo de metodologías educativas, como el PAPIME¹, han permitido realizar cursos a distancia relativos a circuitos eléctricos y control automático. En particular, en este trabajo se exponen los resultados concretos y las experiencias obtenidas durante la realización de objetos y modelos de aprendizaje de la teoría de control secuencial, interactivos, flexibles y de fácil acceso, en los cuales, el alumno tiene la opción de manipular las simulaciones en línea, y como consecuencia, obtener un mejor aprovechamiento de sus actividades de aprendizaje. Se hace énfasis en el uso de las tecnologías actuales de información y comunicaciones (TIC), las cuales han permitido un gran avance de los modelos de educación a distancia, como herramienta indispensable para complementar satisfactoriamente el proceso de enseñanza aprendizaje.

Palabras clave: Ciencias de la ingeniería, educación a distancia, objeto de aprendizaje.

I. INTRODUCCIÓN

Los nuevos modelos de aprendizaje basados en el uso de las tecnologías de la información y las comunicaciones (TIC), como los esquemas modernos de educación a distancia deben fortalecer los procesos educativos para actualizar los

¹ Este trabajo ha sido financiado con fondos del Proyecto perteneciente al Programa de Apoyos para la Innovación y el Mejoramiento de la Enseñanza (PAPIME) con clave PE102206, otorgado por la Dirección General de Asuntos del Personal Académico de la UNAM.

paradigmas de la educación presencial tradicional, de tal forma que el alumno complemente su proceso educativo ingresando a los espacios virtuales, de manera síncrona o asíncrona, para reforzar los conocimientos adquiridos en clase. En este proceso el profesor tendrá la facilidad de usar simuladores, presentaciones, guías, artículos y demás material que se elabore para complementar su curso sin necesidad de llevarlo al aula tradicional.

En el proyecto que se presenta se ha abordado una de las disciplinas más representativas de las llamadas ciencias de la ingeniería, las cuales se desarrollan conforme a un campo disciplinario específico, para crear una plataforma teórica, organizada y justificada científicamente, en la cual se basan ciertos campos de la ingeniería aplicada; es decir, los conceptos, formulaciones y métodos de las ciencias de la ingeniería son tomados como base para desarrollar los conocimientos particulares de las múltiples formas de la ingeniería aplicada.

La teoría de control, como parte de las ciencias de la ingeniería, se compone de conceptos y métodos analíticos con un alto grado de abstracción, lo que hace necesario el uso de recursos didácticos que faciliten el proceso cognoscitivo, como el uso de plataformas computacionales para el manejo de modelos matemáticos, la solución de ecuaciones diferenciales y la simulación numérica para la validación de algoritmos de diseño [1]. La teoría de control es un ejemplo paradigmático de las ciencias de ingeniería porque requiere de una amplia gama de antecedentes físico-matemáticos y porque constituye una plataforma para diversas

ingenierías de aplicación, como control de procesos industriales, control automático, sistemas dinámicos, análisis de transitorios en redes eléctricas, automatización, control neuro-difuso, robótica y mecatrónica, entre otras, como se muestra en la Fig. 1.

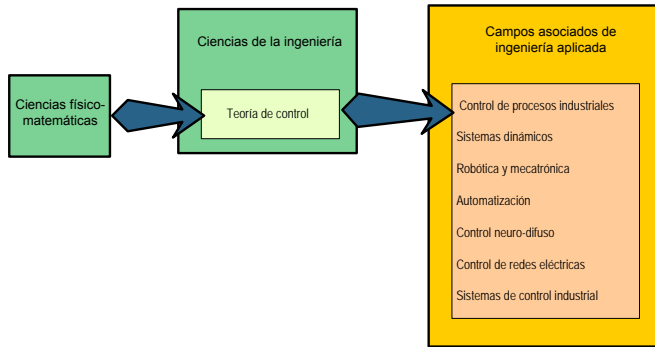


Figura 1. Teoría de control e ingenierías aplicadas.

II. SIMULADORES EN EL ESTUDIO DE LA TEORÍA DE CONTROL

El proceso de aprendizaje de la teoría de control ha propiciado, desde hace mucho tiempo, la construcción y uso de objetos, físicos y virtuales, enfocados a la experimentación, la cual constituye una parte fundamental de este proceso [4]. La experimentación, que hace posible verificar las formulaciones analíticas, comprobar las propiedades de los sistemas y validar los elementos de diseño, se ha realizado tradicionalmente en los laboratorios, empleando equipo y dispositivos de propósito específico; sin embargo, dada la dificultad de contar con equipo experimental en número suficiente y con las características idóneas para reproducir físicamente los elementos y propiedades que se plantean en el desarrollo teórico, se ha tenido que recurrir a la simulación en computadoras digitales y, recientemente, al trabajo en laboratorios remotos [3]. Es importante reconocer que la simulación es una herramienta fundamental en el diseño de sistemas en ingeniería, ya que permite probar una teoría o una solución propuesta, antes de pasar a la realización, que por lo general consume significativos recursos económicos y de tiempo. Incluso, se tiene

metodologías en las que, la simulación es parte importante del desarrollo de los proyectos y se tienen diversos productos computacionales para ese propósito.

En la teoría de control secuencial, la simulación se ha empleado para verificar los algoritmos diseñados (en este caso Diagramas de Estado, Diagramas de escalera, Redes de Petri, etc.) y las propiedades asociadas a la solución, además se ha ampliado su uso para verificar los conceptos teóricos de mayor complejidad o de difícil exposición.

III. CURSOS A DISTANCIA Y OBJETOS DE APRENDIZAJE INTERACTIVOS

El desarrollo de cursos a distancia se basa en la integración de múltiples objetos que contienen la expresión completa de un tópico o tema de conocimiento, incluyendo definiciones, conceptos, formulaciones, gráficas, animaciones, applets, actividades y ejercicios, que hacen posible la comprensión y la asimilación detallada de los conocimientos propuestos. La creación de los objetos de aprendizaje tiene como base la participación de los académicos expertos en la enseñanza de los temas que se abordan, cuya amplia experiencia y práctica educativas son los elementos esenciales para el diseño fundamental de los procesos y componentes de los cursos de educación a distancia. La capacidad, iniciativa y creatividad del experto, deben soportar el diseño de alta calidad de los objetos, en conjunción con una metodología precisa que haga posible la transformación del material fuente a documentos y elementos electrónicos que constituyen el curso.

La metodología que se ha empleado para elaborar los cursos a distancia es la propuesta por la Coordinación de Universidad Abierta y Educación a Distancia (CUAED) de la UNAM, cuyas principales etapas se muestran en la Fig. 2.

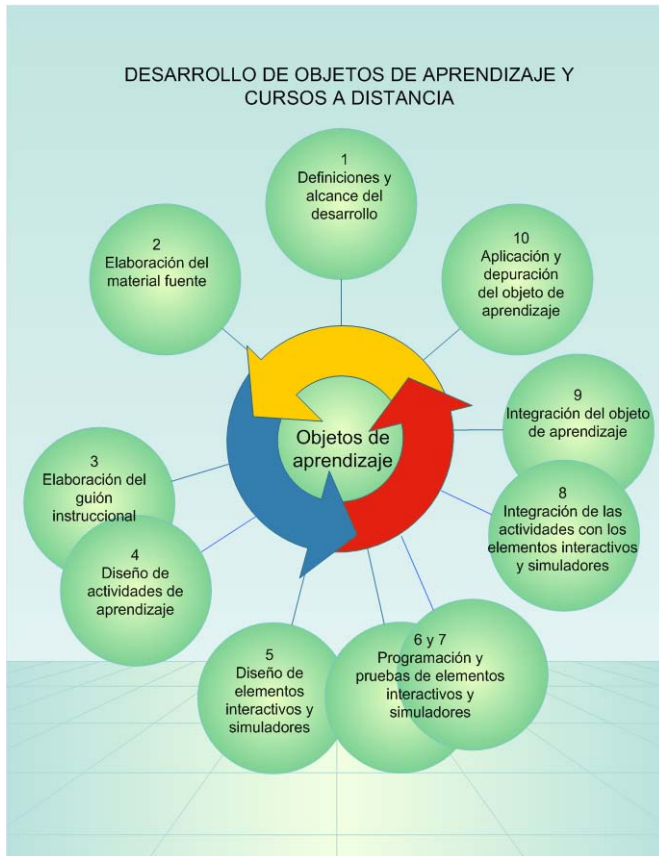


Figura 2. Proceso de desarrollo de objetos de aprendizaje

En este trabajo se detallarán algunas de las características de esta metodología. Consecuentemente, es importante destacar de la figura 2, que el guión instruccional requiere el diseño de actividades de aprendizaje para realimentar al alumno sobre su elaboración cognoscitiva; dichas actividades pueden ser un cuestionario de auto-evaluación, ejercicios, un estudio de caso o la experimentación a distancia por medio de simuladores en línea.

La creación de los elementos virtuales interactivos y simuladores implica su diseño, programación y validación, para que funcionen de acuerdo con los propósitos del curso y sus objetos de aprendizaje.

IV. SIMULADOR DE CONTROL SECUENCIAL

El control secuencial está constituido por secuencias, explica como usarlas para controlar mecanismos, y como construirlas. El componente principal de una secuencia es un estado o un objeto de movimiento, los cuales constituyen cada tarea o paso que la secuencia realiza y son enlazados entre sí para construir secuencias. El control secuencial también implica saber de que forma iniciarlas y detenerlas, enlazar las salidas físicas, y el enlazarlas a otras secuencias.

Para ayudar a comprender el funcionamiento y la aplicación del simulador cabe recordar que una secuencia es una lista ordenada de pasos que son requeridos para ser realizados por un grupo de mecanismos. Las secuencias contienen el tiempo de los mecanismos y las condiciones que se requieren previamente a la ejecución de cada paso, siguen los diagramas de estados o los métodos usados en el diseño o la documentación textual e incluyen hasta el mínimo detalle necesario para realizar la función deseada.

Para construir secuencias, lo primero que se tiene que hacer es agrupar los mecanismos del equipo en grupos de funciones o estaciones y después determinar los pasos para cada uno de estos grupos [5]. Una secuencia se construye para cada grupo programando un objeto de movimiento para cada paso. Cuando programamos secuencias no necesariamente nos preocupamos por las salidas o como las fallas accionan mensajes para el operador. Esas son tareas distintas, las cuales se realizan después de que las secuencias son programadas. Lo importante aquí es concentrarse únicamente en los pasos de la secuencia. Esto nos permite realizar un mejor y más rápido trabajo para programarlas.

El diagrama mostrado en la Fig. 3 ilustra los conceptos básicos de las secuencias aplicados al simulador. El segmento muestra los pasos que realiza un mecanismo que recoge/acomoda (mano mecánica) para recoger y mover la parte de un componente.

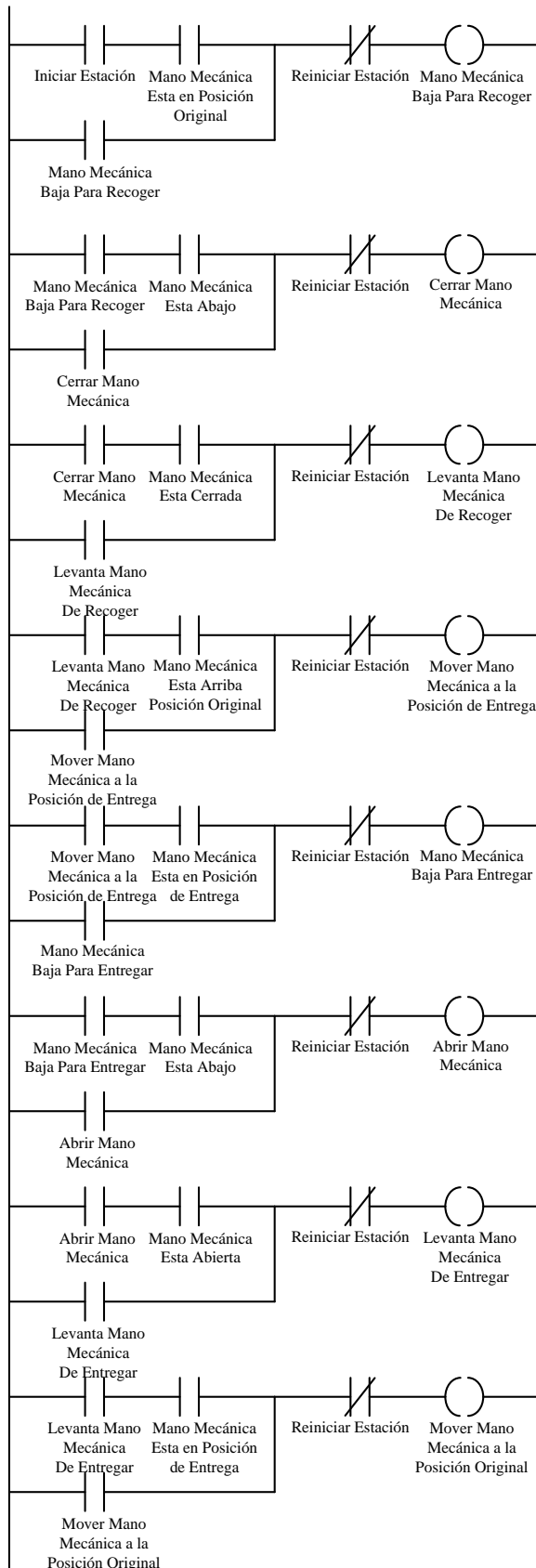


Figura 3. Diagrama de escalera del simulador.

Los pasos en la Fig. 3 usan memorias (latches) de secuencia para controlar cada movimiento del mecanismo. En el paso n la lógica realiza una tarea específica no sin antes estar segura que el último paso se realizó para poder continuar. Por ejemplo los contactos “Cerrar Mano Mecánica” y “Mano Mecánica Esta Cerrada” son condiciones de parámetros para el latch “Levanta Mano Mecánica de Recoger”. Ambas condiciones tienen que satisfacerse antes de que la mano mecánica se levante. El contacto “Cerrar Mano Mecánica” viene del paso anterior y es usado para mantener que cada paso se ejecute en cierto orden. El contacto “Mano Mecánica Esta Cerrada” es la retroalimentación que le dice a la lógica que el paso anterior ha finalizado.

El hecho de estar seguros de que cada movimiento regresa a donde empezó hace que la secuencia se cíclica. El contacto “Reiniciar Estación” se usa para reiniciar de forma incondicional todos los latches de secuencia después de que esta termina, para que así pueda empezar de nuevo.

Como se puede ver en la Fig. 3, el segmento de código es muy fácil de leer. Mientras el código para el mecanismo de la mano mecánica es ejecutado, los latches de secuencia se activan uno después de otro hasta que toda la secuencia termina. Si se desarrolla algún problema, se puede trazar a través de los latches de secuencia y de esa forma encontrar o identificar que falló en activarse. Viendo las condiciones de activación para el latch, se puede diagnosticar la causa del problema, como por ejemplo cual sensor falló en energizarse.

Un laboratorio virtual es un entorno distribuido de herramientas de simulación y animación cuyo objetivo es realizar la simulación interactiva de un proceso físico, proporcionando un método flexible y amigable para definir los experimentos que se llevan a cabo [2].

Los simuladores están programados en el lenguaje de programación LabVIEW y las prácticas se realizan mediante una interfaz gráfica que contiene instrumentos virtuales. El proceso se comporta de acuerdo con el modelo programado y puede incluir

componentes no lineales, cuyo uso queda a criterio del diseñador de las prácticas y en última instancia en la decisión del alumno.

Estos simuladores se encuentran en la plataforma Moodle del Departamento de Ingeniería de Control y Robótica el cual se muestra en la Fig. 4, y pueden ser usados por los profesores y alumnos en cualquier momento, desde cualquier PC con acceso a Internet.

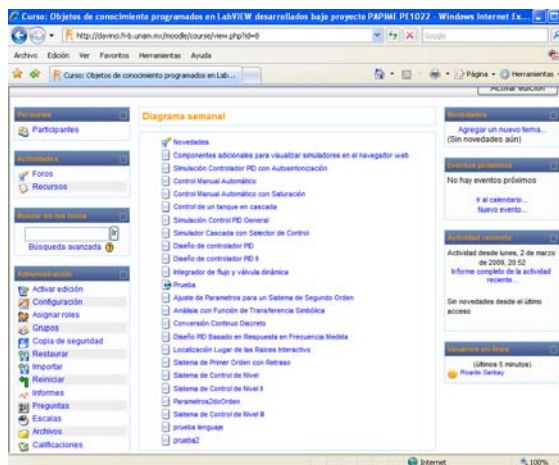


Figura 4. Portal de acceso a los simuladores de sistemas de control.
(<http://davinci.fi-b.unam.mx/moodle>)

El simulador mostrado en la Fig. 5 incluye diferentes tipos de controles e indicadores por medio de los cuales el alumno realiza las actividades en línea e interacciona de forma sencilla e intuitiva, con lo cual, realiza las actividades que se le indican en el contenido del curso. En este caso dichos controles e indicadores son personalizados en base a la teoría de control secuencial industrial, por ejemplo se puede identificar un control de tipo booleano el cual “Habilita cierta Estación”, esto se aprovecha para enfatizar la organización por estaciones lo cual se menciona en unidades posteriores que se enfocan en control jerárquico y control concurrente. De forma similar tenemos un control booleano para reiniciar la estación lo cual lleva a un estado lógico que indica la posición original del brazo mecánico.

Además, se tiene un control de paro de estación el cual detiene por completo el funcionamiento del simulador, también se cuenta con un diagrama de estados y un panel de accionamientos. En esta parte el alumno relaciona los diagramas de escalera vistos en la primera parte del curso con los diagramas y tablas de estados, ya que en base al estado en que se encuentre el brazo mecánico, el color del indicador de estado cambia, cabe mencionar que la animación del brazo mecánico se programó con controles booleanos personalizados y nodos de propiedad los cuales son funciones propias de LabVIEW y nos permiten mostrar u ocultar dichos controles en base al estado en que se encuentra el simulador.

El panel de accionamientos le sirve al alumno para relacionar la función y activación de los transductores simulados los cuales indican, por ejemplo, si la mano mecánica está arriba, abierta, próxima a cerrar, etc. Esta misma información se usa para que ocurra la transición de estados a esto también se le conoce por el nombre de “acciones y condiciones”, estas se usan en un anexo del curso donde el alumno, ya que conoce los diagramas y las tablas de estado, relaciona dichas representaciones con las Redes de Petri.

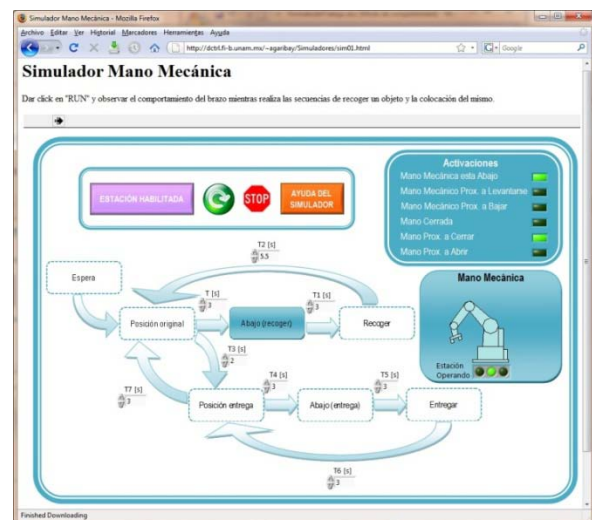


Figura 5. Elemento interactivo para ejercitar el diseño de sistemas secuenciales.
(<http://ctrl.fi-b.unam.mx/~agaribay/Simuladores/sim01.html>)

Finalmente el simulador cuenta con un botón de ayuda con el cual invoca un subprograma que incluye la descripción de todos los controles e indicadores y de la forma en que el alumno puede interactuar con el mismo.

Designación	Función	Comentarios
	Inicia el simulador	
HABILITAR ESTACIÓN	Coloca la máquina en las condiciones de arranque.	Es la condición necesaria para que el ciclo se inicie.
	Reinicia el proceso desde el estado "Posición original"	Vuelve a la posición inicial una vez que se han efectuado las activaciones del estado actual
	Detiene la ejecución del simulador	Detiene la simulación cuando terminan las activaciones del estado actual
Activaciones	Indica la activación correspondiente al estado actual del brazo mecánico.	
Estación operando	Muestra el código binario asociado al estado actual.	<ul style="list-style-type: none"> Posición original Abajo (recoger) Recoger Posición entrega Abajo (entrega) Entregar
Mano mecánica	Muestra una representación gráfica del estado actual de la máquina.	
Tn [s]	Ajusta el valor del temporizador n.	Valores en segundos.

Figura 6. Ayuda del simulador.

En la Fig. 7 se muestra el código del programa principal donde se pueden ver algunos de los nodos de propiedad usados para programar la animación del diagrama de estados y el movimiento de la mano mecánica.

La subrutina mostrada en la parte inferior invoca el subprograma que contiene la información de ayuda del simulador. También se puede notar la estructura de programación basada en máquinas de estados usando, por mencionar las más importantes, registros de corrimiento; constantes de tipo numéricas "Enum"; estructuras de secuencia (Flat Sequence Structure); estructura de control por casos (Case Structure); comparadores;

ciclos While (While Loop); y funciones de temporización.

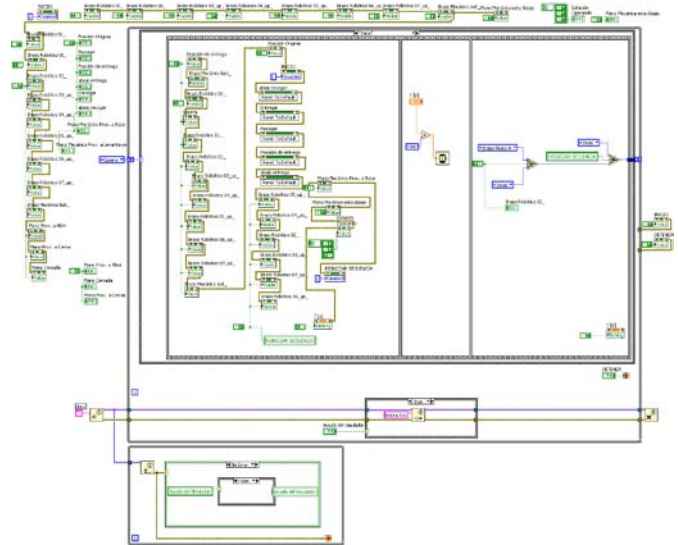


Figura 7. Diagrama de bloques del simulador.

Finalmente para poder ejecutar el simulador en su computadora no es necesario tener LabVIEW instalado, ni tener una licencia válida, el alumno simplemente tiene que descargar e instalar el LabVIEW Run-Time Engine 8.2 o superior el cual permite que el explorador de internet despliegue de forma embebida las aplicaciones programadas en LabVIEW, este se descarga de forma gratuita de la red. Además de esto si no se tiene LabVIEW instalado hay que guardar el archivo LVBrowserPlugin.ini el cual se puede descargar del siguiente enlace <ftp://ftp.ni.com/pub/devzone/LVBrowserPlugin.ini> en la carpeta "Mis documentos\ LabVIEW Data" (Si la carpeta "LabVIEW Data" no existiera hay que crearla).

Debido a las complicaciones que han tenido algunos de los alumnos para realizar la configuración antes mencionada, se ha diseñado una aplicación que configura el sistema automáticamente, con lo cual se evitó que el alumno tuviese que configurar de forma manual la computadora. Cabe mencionar que esta configuración solo es válida para sistemas con Windows XP/Vista de 32 y 64 bits, la descarga del archivo de instalación antes mencionado se puede

realizar por medio del siguiente enlace:
<http://dctrl.fi-b.unam.mx/~agaribay/Insta-Config.zip>

Cabe destacar que este es solo uno de varios simuladores que se han desarrollado para los diferentes cursos de control e instrumentación virtual que se imparten en la facultad de ingeniería de la UNAM, el resultado ah sido satisfactorio, esto se ha verificado mediante la retroalimentación proporcionada por los alumnos en la misma plataforma Moodle, ya que la mayoría lo consideran fácil de operar, y con el programa instalador antes mencionado, no hay necesidad de configurar de forma manual absolutamente nada.

Además, los resultados obtenidos han producido un incremento en el aprovechamiento de los cursos debido el cual se ha visto reflejado en las calificaciones y en la retroalimentación como antes se menciona la cual se encuentra en los foros de discusión con los que se cuenta en la plataforma Moodle.

V. CONCLUSIONES

El alumno necesita elementos didácticos para su estudio, gráficas que expresen las propiedades de los sistemas que ha analizado o diseñado, como en el caso de las simulaciones cuyas respuestas resultan útiles para entender el comportamiento y funcionamiento de un sistema secuencial en este caso una mano mecánica que recoge y posiciona un objeto.

En Internet se encuentran diversos tutoriales, desde aquellos que incluyen solo texto, hasta los que por medio de interfaces gráficas proporcionan ejemplos, en donde los usuarios pueden cambiar parámetros. En este sentido, las soluciones que se han obtenido otorgaron un alto beneficio a los estudiantes que hicieron uso de ellos.

La programación de brazo mecánico y de la mayoría de los simuladores, requieren de cierto tiempo de programación, como se menciona anteriormente, dicho trabajo es realizado en LabVIEW, con lo cual se reduce

significativamente el tiempo de programación comparado con el tiempo que llevaría en algún otro lenguaje de programación tradicional orientado a objetos (C++, Java, Visual Basic, etc.).

Impacto en la actividad docente. El docente cuenta con herramientas para crear y/o modificar contenidos en línea, sin necesidad de una gran inversión de tiempo o de tener un conocimiento profundo de la plataforma educativa, pues los objeto desarrollados se han integrado en cursos a distancia, y han fomentado una nueva cultura para la creación e intercambio de material didáctico por parte de los docentes, además de esto, el hecho de incorporar la plataforma LabVIEW de forma embebida en un explorador de internet ha facilitado el intercambio de información entre el profesor y el alumno, de esta forma en los cursos impartidos en las aulas o laboratorios, el profesor ha reducido notoriamente la cantidad de tiempo que utilizaba para describir y dibujar el planteamiento del problema o en explicar el funcionamiento del sistema, ya que el alumno tiene acceso al simulador y al material del curso previamente a la clase, lo cual representa una ganancia en tiempo debido a que el docente aprovecha y se enfoca directamente en explicar y desarrollar el contenido del curso.

Impacto en las actividades de los alumnos. Los alumnos han fortalecido su proceso de enseñanza aprendizaje a través de un espacio virtual, han reforzado sus conocimientos, también han obtenido información adicional de la asignatura y efectuado prácticas a través de simuladores, desde la ubicación y hora en han podido y han deseado trabajar.

También se han llegado a suscitar problemas debido a que la computadora donde la persona desea acceder al curso, está mal configurada o tiene algún tipo de limitante (virus, spyware, etc.) lo que causa que la persona no pueda ejecutar el simulador, también se ha dado el caso que utilizan algún otro sistema operativo distinto al que se le indica, o algún explorador que no está soportado por esta plataforma.

Posiblemente se pueda ver como una desventaja el hecho de que la interacción con el alumno se reduce al usar este método de enseñanza a distancia en los cursos tradicionales, aunque el profesor no esté presente cuando los alumnos realizan las actividades y ocurran las dudas, con un solo docente se puede atender a múltiples alumnos, ya que el docente no tiene que atender de forma simultánea a todo el grupo, debido a que las asesorías son vía remota (correo electrónico, foros de discusión, y en algunos casos si el alumno lo requiere vía telefónica). En contraste, algunas de las ventajas serían que el material elaborado puede ser por múltiples usuarios de forma simultánea, además, puede ser utilizado en otros cursos y queda como acervo o material de consulta para la universidad.

VI. REFERENCIAS

- [1] Garibay J. R., Espinosa P. G. “*La investigación en Control y Robótica en la Facultad de Ingeniería, Trayectoria y Tendencias*”, Primer Congreso sobre Investigación en Facultades, 2005.
- [2] Mata H. G, Garibay J. R. “*Instrumentación virtual: una alternativa para el equipamiento de laboratorios*”, Congreso Nacional COPEI, 2002.
- [3] Flores G. R., Aguado C. E., Vázquez F. R. “*La web en la enseñanza del control*”, Memorias del Tercer Congreso Nacional CIMEEM, 2008.
- [4] Granados P., Valerio M., Díaz G. “*Enseñanza del control automático mediante tutoriales y laboratorios virtuales por Internet*”, Monografía del curso IE431, Escuela de Ingeniería Eléctrica, Universidad de Costa Rica, 2003.
- [5] Kirckof, Gary, P.E. *Cascading logic. A machine control methodology for programmable logic controllers*. The Instrumentation, Systems, and Automation Society, 2003 ISA

VII. AUTORES

Ing. Alberto R. Garibay Martínez es Ingeniero en Electrónica por parte de la Universidad Autónoma Metropolitana – Azcapotzalco. Actualmente profesor de asignatura y experto en el desarrollo de simuladores. Facultad de Ingeniería de la UNAM.

M. en I. Ricardo Garibay Jiménez es Jefe del Departamento de Ingeniería de Control y Robótica de la Facultad de Ingeniería de la UNAM.

MODELADO Y CONTROL DE UN REACTOR QUÍMICO, COMO PARTE DEL APRENDIZAJE BASADO EN ESTUDIOS DE CASO

Soria Villegas Francisco Daniel.

Departamento de Ingeniería Eléctrica, DIE
UNAM, Facultad de Ingeniería
Av. Universidad 3000, Ciudad Universitaria, Coyoacán D.F. 04600, México
Teléfono 56223297, correo electrónico ing_fdsoriav@yahoo.com.mx

RESUMEN

El trabajo tiene la intención de compartir un ejemplo del control de una mini planta para la producción de etilenglicol y ácido acético empleado como un estudio de caso, que puede ser útil para asignaturas afines a instrumentación, modelado y control de procesos para diferentes campos de la ingeniería. Los alumnos deben estar cursando asignaturas de ciencias de la ingeniería, tener una preparación adecuada, con bases que les permitan realizar el modelado de sistemas físicos, conocer la aplicación de la instrumentación y determinar la mejor forma de control.

Para modelar el comportamiento hidráulico y térmico de los componentes de la planta se empleó la teoría de sistemas lineales (con ecuaciones diferenciales de parámetros concentrados e invariantes con el tiempo). Por otro lado se proporcionan a los alumnos datos prácticos obtenidos experimentalmente del mini reactor químico por lotes (**Fig. 1**) empleado en la producción de etilenglicol, que es un componente de algunos anticongelantes.

Palabras clave: reactor químico, estudio de caso, modelado de sistemas hidráulicos, balance de energía y control.

I. INTRODUCCIÓN

Una forma de mejorar la enseñanza y el aprendizaje en asignaturas asociadas al campo del modelado, de la instrumentación y del control es empleando ejemplos que sean asociados a procesos o sistemas de producción industrial, en nuestro caso una mini planta para la producción de etilenglicol, que fue construida y posteriormente desmantelada en 2009. La reacción no es peligrosa, ni tóxica, no genera contaminación y es de tipo endotérmica. La mini planta era operada automáticamente a través de una interfase constituida por una tarjeta electrónica a la que se introducía la información mediante un teclado; la misma tarjeta permitía el monitoreo de las variables de la miniplanta mediante el despliegado de éstas en un display, además de contener el circuito de potencia.

Conforme al modo de operación los reactores químicos se clasifican en continuos o por lotes. La selección o diseño del reactor depende de las características de los reactivos y productos, del volumen, de las temperaturas, de las velocidades de cambio y de la reacción que se quiere realizar. En el presente trabajo se empleó un mini reactor químico por lotes.

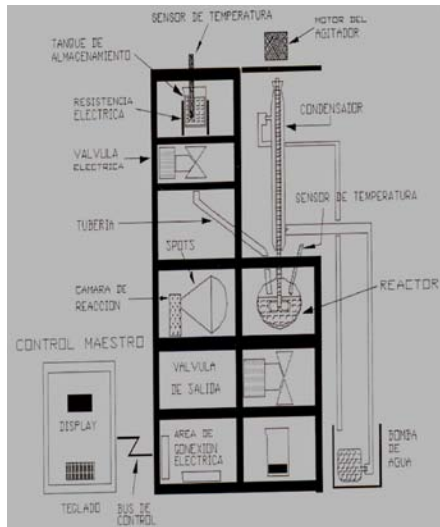


Fig. 1 Mini planta piloto con un reactor químico

La mini planta (**Fig. 1**) constaba de dos tanques, uno de ellos contenía agua, (en lo sucesivo reactivo A) y el otro monoacetato de etilenglicol (reactivo B).

Se tenía una fuente de calor en cada tanque. Una vez alcanzada la temperatura a la que debían alimentarse cada uno de los reactivos, se abrían las válvulas de salida de los tanques y se conducían las sustancias a otro recipiente (reactor) para ser combinados y obtener por reacción una mezcla en la que se encuentra presente el etilenglicol.

La reacción es denominada “hidrólisis del diacetato de etilenglicol” (agua + monoacetato de etilenglicol \leftrightarrow etilenglicol + ácido acético), es de tipo endotérmico, a volumen constante y bajo determinadas condiciones puede ser reversible.

En la mini planta el reactor es un recipiente formado por un matraz de tres bocas esmeriladas: dos para la alimentación de los reactantes y una para permitir la salida del vapor generado por la reacción. En el fondo del matraz se adaptó una oliva con una tercera válvula para la salida del producto de la reacción.

La homogeneidad en la mezcla reactiva se lograba por medio de un mezclador. Se producía vapor que

se condensaba y recuperaba mediante un condensador químico (**Fig. 1**). La reacción se iniciaba con el suministro externo de energía y se mantenía durante media hora.

Al término de la reacción se obtenía una mezcla que al pasar por un proceso de destilación permite obtener ácido acético y etilenglicol.

Cada tanque contaba con un sensor de temperatura, la señal de éste se comparaba con la información programada en un microprocesador, del cual salían las señales para un circuito de potencia que activaba las válvulas para dejar salir los reactivos. El control de temperatura se realizaba por medio de la comparación de los datos de un sensor con un setpoint, generando una señal de error que aumentaba o disminuía la potencia de la fuente de calor. El control se realizaba por medio de un control tradicional P-I con el cual se mantenía la temperatura dentro del reactor en un valor específico durante la reacción.

II. MÉTODO

Se inicia el estudio de caso dando a los alumnos una introducción que los sensibilice respecto a la importancia de la ingeniería en la producción de sustancias o artículos empleados en la vida diaria.

Por ejemplo, para el enfriamiento de los motores se emplea actualmente la recirculación de anticongelante que por sus propiedades es más eficiente que la recirculación de agua pura.

Imagine el lector que al haber escasez de anticongelantes durante varias semanas, se empleara agua como refrigerante de los motores, pero, con el tráfico de las grandes ciudades, a vuelta de rueda, casi parados los vehículos, seguramente se sobrecalentarían los motores, provocando caos vial, enojos, posiblemente accidentes y tensión nerviosa en los que conducen las unidades.

Por los motivos anteriores se debe asegurar la producción de los anticongelantes. La mini planta

descrita permite obtener uno de los constituyentes principales de los refrigerantes, el etilenglicol.

El situar el estudio de caso en un contexto práctico y relacionado con la vida cotidiana, permite motivar al alumno, comprometiéndolo a participar activamente en su solución. En el proceso de aprendizaje esta participación activa es fundamental; se debe fomentar que los alumnos asuman un rol protagónico en el análisis del caso y en el planteamiento colectivo de alternativas para el modelado y control del proceso.

La organización de trabajo

Para abordar el tema de ciencias de la ingeniería que nos ocupa, los alumnos deberán haber aprobado asignaturas que les provean conocimientos en temas como: las ecuaciones diferenciales, álgebra, termodinámica, mecánica de fluidos, sensores y actuadores.

Los equipos de trabajo, de máximo 5 alumnos, tendrán un moderador y un secretario. El respeto a las opiniones y la incorporación de ideas de todos es fundamental para desarrollar un buen trabajo en equipo. La participación individual y colaborativa en ambientes agradables produce excelentes resultados.

Es recomendable dividir el caso planteado en partes (clases), proporcionando el tiempo suficiente para que cada alumno entienda y planteé soluciones al problema que se ha establecido. Después se lleva a efecto el trabajo de los equipos y se finaliza cada parte con las conclusiones del grupo.

Los tiempos para solucionar los problemas que componen el caso, dependerán del número de horas requeridas para abordarlos con la profundidad que se quiera lograr (basada en el objetivo inicial). Al principio del planteamiento se dan a conocer los tiempos y las reglas que deben ser claras para todos.

El profesor debe conocer el tema a tratar, adicionalmente debe tener experiencia en el manejo de equipos de trabajo como guía y, capacidad de generar confianza en los alumnos, quienes deben resolver las preguntas en tiempo y forma.

III. EL EJEMPLO

Los tanques de almacenamiento y distribución.

Se tienen dos tanques a cielo abierto (**Fig. 1**) con los reactivos, cada uno tiene un volumen inicial a la temperatura ambiente. El diámetro y altura es de 7.5 cm, el nivel de líquido cubre 2/3 partes del volumen.

Los reactivos (A y B) se precalentaban a 75 °C valor cercano a la temperatura de reacción (80 °C). Había una disminución de la temperatura durante el transporte. La capacidad de cada tanque era de 300 ml.

Actividades a realizar.

1. El alumno consulta información de diferentes tipos de actuadores, en particular clases de válvulas y por escrito realizará un resumen.
2. El equipo comenta y observa los comportamientos de diferentes válvulas y sus ecuaciones, según presiones y características de flujos.
3. El equipo selecciona las mejores válvulas para el proceso analizado.

Las válvulas

Para regular la cantidad de flujo intermitente que se suministra al reactor químico, se operaban con señales eléctricas válvulas solenoides de dos vías (**Fig. 1**) normalmente cerradas, a 24 V CD.

Las presiones antes y después de la válvula no influían en la descarga, siendo el tiempo de apertura de 4 segundos aproximadamente.

El reactor químico

Una vez alimentados los componentes, la reacción se iniciaba por medio de suministro de calor durante 30 minutos, a través de una pared semejante a un encapsulado, tenía poco desprendimiento de vapor controlado por un condensador químico 30 mm (Kimex # 18150); terminado el tiempo de reacción se abría una válvula normalmente cerrada y se volvía a repetir el ciclo.

El modelado de cada tanque

Las consideraciones establecen dentro del recipiente el trabajo igual a cero, la energía cinética despreciable, la energía potencial muy pequeña, la entalpía (h) es función esencialmente de la temperatura (T) y de la capacidad calorífica promedio (C_p), de forma que $h = C_p T$.

Actividades a realizar.

1. Cada alumno determina la ecuación de balance de energía, balance de volumen y efecto inercial del flujo por la tubería.

2. El equipo de trabajo determina la ecuación diferencial asociada al comportamiento de la temperatura, del nivel y del efecto inercial en el transporte de la componente.

3. Determinar en el grupo por consenso, las respuestas del punto anterior.

Nota. Se puede opinar sobre otros materiales y diferentes características de recipientes y fuentes de calor.

Se tenía como fuente de calor una resistencia eléctrica de 300 W, de abrazadera, cubriendo el

área superficial (lateral) del recipiente de almacenamiento para cada componente.

La constante de tiempo para alcanzar 75 °C es de 5 segundos para el reactivo A y 6 segundos para el reactivo B, de tal manera que las válvulas se abrían después de 7 segundos.

El modelo considera parámetros concentrados. Si bien la temperatura se comporta de forma distribuida, al ser el calentamiento rápido y contar con mezclado puede considerarse una temperatura promedio uniforme para la transferencia de calor (**Fig. 2**).

$$C_v dT_v/dt + h_v A_v (T_v - T_i) = u$$

Donde la resistencia térmica (R_T) está dada por $R_T = 1/(h_v A_v)$ siendo A_v (m^2) el área en contacto con la fuente de calor y h_v ($cal / s \text{ } ^\circ C \text{ } m^2$) el coeficiente de transferencia de calor. C_v es la capacidad de almacenamiento de calor en el recipiente de cristal.

$$C_v = \rho C_e V_v \text{ en } J/^\circ C \text{ o } cal/^\circ C$$

La transferencia de calor de la pared del recipiente a su contenido (componente) para incrementar su temperatura se realiza por convección:

$$Q_v = h_i A_i (T_v - T_i)$$

El incremento de temperatura del reactivo, T_i , tiene un efecto transitorio sumado al estado permanente:

$$T_i = a \{ 1 - e^{-t/\tau} \}$$

Se considera la temperatura ambiente como la referencia o estado inicial, en el estado permanente la temperatura debe ser 75 °C. Con estos datos se puede evaluar la constante a .

Para la respuesta de la temperatura en el reactivo A y el B, se tiene la constante de tiempo τ . El valor

obtenido experimentalmente es de $\tau_A=5$ segundos para A y $\tau_B = 6$ segundos.

Para la salida de los reactivos

La variación de la altura del líquido en los tanques debido a la salida de flujo se puede expresar por medio de las ecuaciones asociadas a la velocidad de salida del flujo y a la variación de altura del tanque durante el tiempo de suministro de los reactivos al reactor.

Se considera el flujo igual a la velocidad de salida por el área de la tubería para determinar dv/dt y a partir de ésta se determina la variación de altura en el tanque (dh/dt), considerando que el cambio de volumen es equivalente al cambio de altura por el área del tanque (A_T).

$$A_T \{dh/dt\} = F_e - F_s = F_e - h/ R_h$$

R_h se considera como el efecto de regulación de flujo de la válvula.

Para la variación de flujo por la tubería

$$dF_s/dt = \dot{A} (\Delta P)/(\rho L)$$

Siendo F_e el flujo de entrada al tanque, F_s el flujo de salida del tanque, \dot{A} es el área de la tubería, ΔP diferencia de presiones, ρ densidad y L longitud de la tubería.

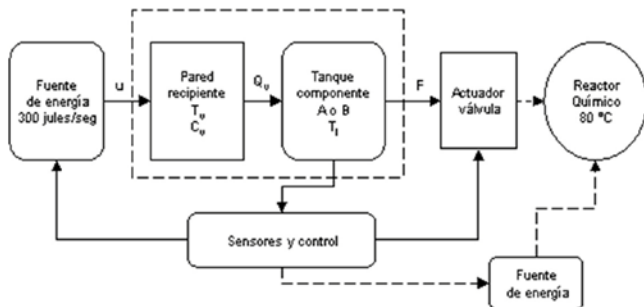


Fig. 2 Diagrama de bloques del tanque conectado al reactor químico.

Los reactivos son alimentados al reactor (**Fig. 2**) con el volumen especificado de 200 ml por cada uno de ellos, son uniformemente combinados para

tener una mezcla homogénea, a la cual se suministra calor por medio de una fuente de 500 W en forma de malla eléctrica sobre la superficie del reactor.

Actividades a realizar.

1. El alumno determina las ecuaciones para después determinar como varían el nivel y la temperatura de los reactivos con el tiempo durante la reacción.
2. El equipo de trabajo realiza el punto anterior, hasta lograr sus conclusiones.

Nota. Se entrega al grupo una gráfica con los resultados experimentales y se puede pedir la comparación de resultados teóricos con respecto a los prácticos.

La variación de los reactivos

La disminución de la concentración molar del componente A (diacetato de etilenglicol) está dada por:

$$dA/dt = - C_1 A$$

Conforme disminuye la concentración del diacetato de etilenglicol, aumenta la concentración del monoacetato de etilenglicol.

Balance de energía en el reactor

Suministrado el calor Q_R a la mezcla de reactivos, el balance térmico es el siguiente:

$$(V \rho C_p) dT_i/dt = - \{ \lambda C_1 A \} + Q_R$$

V representa el volumen de líquido en el reactor.

La λ es un coeficiente positivo en unidades de cal/g-mol para reacciones endotérmicas (negativo para reacciones exotérmicas).

C_1 es un coeficiente asociado a la velocidad de la reacción y depende de la temperatura a la que se lleva a cabo ésta.

$$C_1 = b e^{(-E/RT)}$$

ecuación de Arrhenius.

La b es una constante, E energía de activación, T es temperatura absoluta, R es la constante de los gases perfectos igual a 1.99 cal/g mol °K

Las actividades para la siguiente sección.

1. El alumno realiza el modelado del sistema térmico, por medio del balance de energía en el reactor químico.
2. El equipo de trabajo determina el modelado del sistema térmico.

El sistema de control de la temperatura en el reactor

El control se realiza por lazo cerrado (Fig. 3), con una señal de error efecto de la suma algebraica de la señal de la temperatura dentro de la mezcla del reactor y el SP.

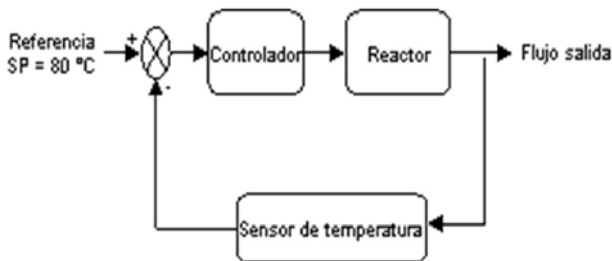


Fig. 3 Control de temperatura para el reactor químico.

En ocasiones se necesita un filtro entre el sensor y el comparador; así la señal de error sin perturbaciones es la entrada al controlador.

Para mantener la temperatura en 80 °C se tiene un controlador que puede ser proporcional (P),

proporcional integral (P-I) o proporcional derivativo e integral (P-I-D). Debe ser sintonizado de conformidad con las características de respuesta de la temperatura debido a la reacción.

El sensor de temperatura seleccionado fue el LM35 cuyo comportamiento es lineal, para 20 grado tiene un potencial de salida de 2 volts y, para 95 °C 11 volts.

La salida del comparador (error) se utilizaba, para actuar sobre un circuito de potencia que proporciona la energía calorífica para la reacción de las componentes A y B en el reactor químico.

El incremento de temperatura dentro del reactor (T_i) se comportaba como un sistema de primer orden, con variaciones que eran corregidas por el control proporcional P-I.

$$P-I = K_p \{ 1 + 1/s\sigma_i \}$$

Los valores resultantes experimentalmente fueron $K_p = 50$, que producía un error de estado estable que era corregido por el efecto integral con valor de tiempo $\sigma_i = 0.1$ segundos.

La constante de tiempo producida por incremento de temperatura dentro de la mezcla de reactivos, es obtenida en forma experimental, $\tau_R = 100$ segundos.

Después de terminada la reacción, se abría la válvula de salida durante 10 segundos. El volumen se almacenaba en un recipiente que pasaba a otra etapa.

1. Al grupo se le entregarán los resultados de la experimentación con diferentes controles para que comparen los resultados de las ecuaciones que han obtenido con esos datos asociados al reactor químico.
2. Se solicita a los alumnos los comentarios finales.

IV. CONCLUSIONES

Se ha presentado un ejemplo que corresponde a un estudio de caso para el modelado y control de una miniplanta para la producción de etilenglicol.

Las consideraciones de valores obtenidos prácticamente y las ecuaciones diferenciales básicas, son medios que han permitido realizar la aplicación de la teoría de modelado de sistemas físicos, para la representación adecuada de la respuesta de un proceso con fines académicos.

La meta del ejercicio es lograr con la guía del profesor el mejor modelo y acciones de control para la planta analizada.

El trabajo de análisis y síntesis individual, de equipo y del grupo lleva al consenso de la determinación del modelo teórico lineal con el menor error al compararlo con los datos prácticos.

En un reactor para procesos industriales químicos o petroquímicos cambiarían las especificaciones. En esencia el ejemplo realizado es cercano a la realidad.

Quedan pendientes para otro capítulo el modelo de parámetros distribuidos, en tres dimensiones, con señales de sensores en tiempo discreto (muestreadas) y control digital.

V. REFERENCIAS

- [1] Brogan W. L., Modern Control Theory, Quantum, USA, 1974.
- [2] Dorf Richard C. et al, Sistemas de Control Moderno, Editorial Parson, 2005.
- [3] Franklin Gene F. et al, Feedback Control of Dynamics Systems, Addison Wesley, USA, 1994.

- [4] L. Luyben W., Process Modeling, Simulation and Control for Chemical Engineers, McGraw Hill. USA, 1990.
- [5] Ogata Katsuhico, Modern Control Engineering, 4th, Prentice Hall, 2002.
- [6] Soria Francisco et al, Diseño, Construcción y Control de un Reactor Químico, Revista de Ingeniería LXIV 4/4, Facultad de Ingeniería, UNAM, 1994.
- [7] Umez-Eronini Eronini, Dinámica de Sistemas y Control, Thomson, México, 2001.

VI. AUTOR

M. en I. Francisco D. Soria Villegas obtuvo su título de Ingeniero Mecánico Electricista (1976) y de Maestría en Ingeniería (1996) con especialidad en planeación en la Facultad de Ingeniería de la UNAM, es profesor de tiempo completo en el Departamento de Ingeniería Eléctrica de Potencia.

DISEÑO E IMPLEMENTACIÓN DE UN CIRCUITO MPPT PARA SISTEMAS FOTOVOLTAICOS AUTÓNOMOS

Ponce Márquez Fernando, Zapata Domínguez Julio Adrián, De la Torre y Ramos Jorge, Reta Hernández Manuel.

Unidad Académica de Ingeniería Eléctrica - Universidad Autónoma de Zacatecas.
Av. López Velarde No. 810, Col. Centro Zacatecas, Zacatecas 98000, México
Tel. 49239407 Ext. 1505, correo electrónico: nan2_pom@hotmail.com

RESUMEN

En el presente trabajo se muestran los resultados sobre el diseño y la implementación de un circuito rastreador del punto de máxima potencia (MPPT) aplicado a un sistema fotovoltaico (SFV) autónomo utilizando un microcontrolador PIC. El circuito aquí presentado es de bajo costo comparado con otro tipo de convertidores, debido a lo simple del algoritmo de control. Este prototipo permitiría incrementar la eficiencia total de sistemas ya existentes o de sistemas en proyecto de instalación en comunidades alejadas de la red eléctrica nacional, mejorando así la relación costo/beneficio para el usuario final.

Palabras clave: MPPT, sistemas fotovoltaicos, rastreador del punto de máxima potencia, Microcontrolador PIC.

I. INTRODUCCIÓN

Un alto porcentaje de la electricidad producida en las grandes plantas generadoras en el mundo requiere del consumo de combustibles fósiles. En México, por ejemplo, de los casi 50,000 MW de capacidad eléctrica instalada, el 67% es producido en plantas termoeléctricas a base de combustóleo y/o gas natural y el 5.2% en plantas a base de carbón [1]. Desafortunadamente, el consumo de los energéticos consumidos en la producción de la electricidad agrava el problema de contaminación ambiental y reduce las reservas de hidrocarburos.

Es por ello que la búsqueda de fuentes alternas de energía eléctrica, que sean renovables y que emitan residuos mínimos de contaminación, se ha

intensificado en las últimas décadas. Un sistema de generación que cumple con dichas características lo constituye el sistema fotovoltaico.

La radiación solar incidente en la Tierra puede aprovecharse, por su capacidad para calentar, o directamente, a través del aprovechamiento de la radiación en dispositivos ópticos o de otro tipo. Es un tipo de energía renovable y limpia, conocidas como energía verde. La potencia de la radiación varía según el momento del día, las condiciones atmosféricas que la amortiguan y la latitud. Se puede asumir que, en buenas condiciones de irradiación, el valor es de aproximadamente $1000\text{W}/\text{m}^2$ en la superficie terrestre. A esta potencia se la conoce como irradiancia. La radiación es aprovechable en sus componentes directa y difusa, o en la suma de ambas. La radiación directa es la que llega directamente del foco solar, sin reflexiones o refracciones intermedias. La difusa es la emitida por la bóveda celeste diurna gracias a los múltiples fenómenos de reflexión y refracción solar en la atmósfera, en las nubes y el resto de elementos atmosféricos y terrestres. La radiación directa puede reflejarse y concentrarse para su utilización, mientras que no es posible concentrar la luz difusa que proviene de todas las direcciones. La irradiancia directa normal (o perpendicular a los rayos solares) fuera de la atmósfera, recibe el nombre de constante solar y tiene un valor medio de $1354\text{W}/\text{m}^2$, correspondiendo a un valor máximo en el perihelio de $1395\text{W}/\text{m}^2$ y un valor mínimo en el afelio de $1308\text{W}/\text{m}^2$.

Una de las desventajas de utilizar sistemas fotovoltaicos es el alto costo y la baja eficiencia que dichos sistemas presentan en la actualidad. Se han propuesto en la literatura diversas alternativas para mejorar la eficiencia de los SFV sin impactar en el costo, tales como incrementar la eficiencia de las celdas solares [2-4] y el uso de convertidores que extraigan en todo momento la máxima potencia del panel [5].

El objetivo del presente trabajo se enfoca en el diseño e implementación de un prototipo de circuito MPPT, de bajo costo, que haga eficiente el uso de los sistemas fotovoltaicos autónomos.

II. MÉTODO

El punto máximo de potencia generado por un panel fotovoltaico es aquel en el cual el voltaje y la corriente extraída del mismo son óptimos para generar la mayor cantidad de potencia posible. En la Fig. 1 se muestran gráficamente las curvas teórica y real de un panel fotovoltaico.

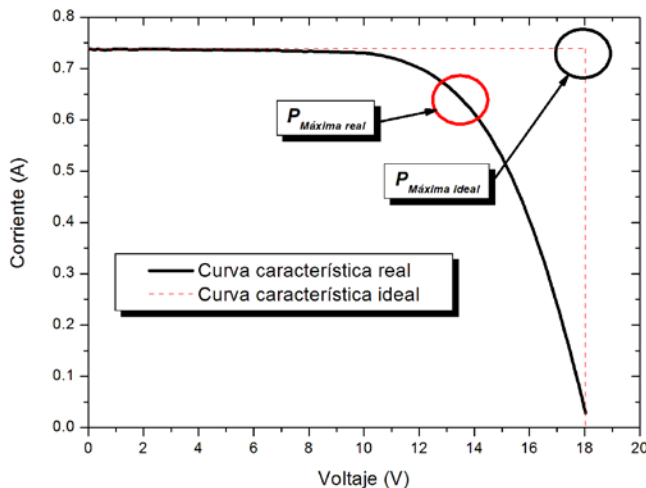


Figura 1- Ilustración del punto de máxima potencia en un panel fotovoltaico.

Dado que el punto de máxima potencia real está en función de las condiciones climáticas y la temperatura del panel, se requiere de un circuito rastreador que represente una carga electrónica variable al panel para de este modo cambiar la impedancia del circuito y extraer la potencia máxima.

II.1 Diseño del MPPT

A. Selección de la configuración.

En la literatura se pueden encontrar diferentes topologías para la fabricación de circuitos MPPT. La primera de ellas y la más utilizada es la configuración en serie [6-7], la cual debe su nombre a la forma en que está conectado el convertidor con el panel. En la Fig. 2 se muestra la topología de conexión en serie.

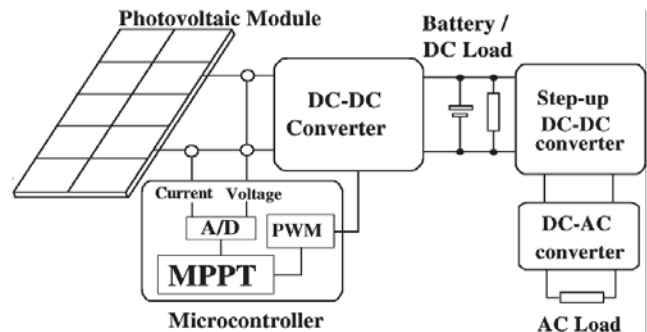


Figura 2- Ejemplo de un topología de conexión en serie [4].

La segunda configuración es la topología en paralelo [8-9], la cual, de acuerdo a pruebas reportadas por los proponentes de dicha configuración, tiene la ventaja de reducir las pérdidas de potencia en todo el sistema. En la Fig. 3 se muestra la topología de conexión en paralelo.

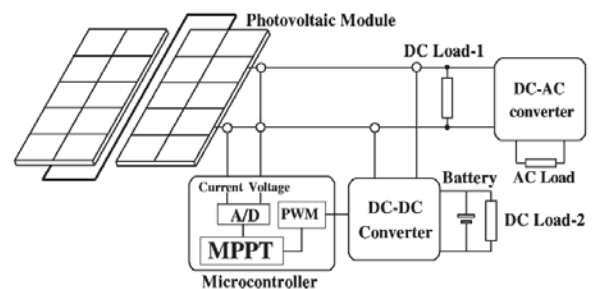


Figura 3- Ejemplo de un topología de conexión en paralelo[4].

En este trabajo se decidió utilizar una configuración en serie debido a que dicha topología permite utilizar al circuito como MPPT y como cargador de baterías simultáneamente.

B. Diseño de la etapa de potencia.

El diseño de la etapa de potencia se efectuó considerando una carga puramente resistiva, un valor del ciclo de trabajo mínimo de 10% y máximo de 90%. La potencia máxima del convertidor se calculó en función de la potencia del panel fotovoltaico a utilizar ($V_{\max} = 20V$, $I_{\max} = 5.5A$, $P_{\max}=110W$).

El primer elemento a seleccionar fue el transistor de potencia (M1), el cual deberá de ser capaz de bloquear el voltaje máximo de entrada (20V) y soportar la corriente máxima de la fuente (5.5A) considerando también un rango de seguridad del 50% para el voltaje y de 25% para la corriente. En base a estas consideraciones se tiene:

$$V_{DSS\min} = (1 + S_v)V_i = (1 + 0.5)(20) = 30V$$

$$I_{S\min} = (1 + S_i)I_i = (1 + 0.25)(5.5) = 6.9A$$

donde:

$V_{DSS\min}$ → Voltaje mínimo que M1 debe bloquear entre drenador y fuente.

$I_{S\max}$ → Corriente mínima fuente-drenador

S_v → Parámetro de seguridad de voltaje.

V_i → Es el voltaje de entrada.

S_i → Parámetro de seguridad de corriente.

I_i → Es la corriente de entrada.

De acuerdo a los valores obtenidos, se seleccionó un transistor MOSFET (IRFBC40A) cuya corriente pulsada máxima es 25A y un voltaje de bloqueo de 600V.

La selección del inductor (L1) afectará el valor del capacitor (C1) y el rizo de voltaje a la salida. Típicamente, se buscará un valor de rizo menor al 5% del valor máximo de la corriente del inductor. Para este caso se determinó $I_{\text{rizo}}=0.275A$ tomando en cuenta la corriente máxima del panel fotovoltaico. El valor de la inductancia se determina, entonces, utilizando la Ec. (1), correspondiente a un convertidor DC/DC tipo boost en un régimen de conducción continua [10].

$$L = \frac{DV_i}{\Delta I_i f_{sw}} \quad (1)$$

donde:

L → El valor del inductor.

D → Es el ciclo de trabajo.

ΔI_i → Es el rizo de la corriente.

f_{sw} → Es la frecuencia de conmutación.

Sin embargo, es necesario determinar el voltaje de salida máximo y el ciclo de trabajo necesario para obtenerlo. Para un valor fijo de $V_{\text{omax}}=70V$, el valor del ciclo de trabajo es de $D=0.7$, obtenido de la Ec. (2) de voltaje de salida de chopper tipo boost:

$$V_o = \frac{V_i}{1 - D}; \quad D = 1 - \frac{V_i}{V_o} \quad (2)$$

Sustituyendo estos valores en la Ec. (1) se obtiene un valor de inductancia de:

$$L = \frac{0.7(20)}{(0.275)(10,000)} = 5mH$$

El transistor de potencia (M1) se protegió utilizando a un diodo de rueda libre (D2) el MUR8100, el cual es un diodo de recuperación ultra-rápida (60ns) con un voltaje de bloqueo de 1000V y una corriente máxima pulsada de 150A. La frecuencia de conmutación se fijó a 10kHz ya que a dicha frecuencia se podrá disminuir el tamaño del capacitor e inductor en el circuito sin afectar significativamente las pérdidas por conmutación en el transistor. Por otra parte, para la determinación del rizado en el voltaje de salida se utilizó la Ec. (3)[10].

$$\frac{\Delta V_o}{V_o} = \frac{D}{RCf_{sw}} \quad (3)$$

donde:

ΔV_o → Es el rizo de voltaje de salida.

R → Es la resistencia de la carga.

C → Valor del capacitor para ajustar el valor de rizo.

Esta ecuación permite determinar el valor del capacitor (C1) necesario para ajustar el factor de rizo del MPPT. Por lo tanto, especificando un factor de rizo de 5% para un ciclo de trabajo de 90% (condición de mayor rizo a la salida), se tiene un valor del capacitor de:

$$C = \frac{V_o D}{\Delta V_o R f_{sw}} = \frac{0.9}{18(0.05)(10,000)} = 100\mu F$$

Finalmente, el diodo D3, cuya función es impedir la descarga del capacitor cuando el transistor está conduciendo, se seleccionó tomando en cuenta la corriente máxima de salida del circuito (5.5A) y un margen de seguridad de 25%.

$$I_{diodo} = (1 + S_i)I_o = (1 + 0.25)(5.5)$$

$$I_{diodo} = 6.87 A$$

Adicionalmente, el diodo debe bloquear un voltaje mayor a 50V, por lo cual se seleccionó el diodo 1N4007, el cual tiene un bloqueo de voltaje de 1000V y 30A de corriente pulsada máxima.

En la Fig. 4 se muestra la topología final de la etapa de potencia del MPPT.

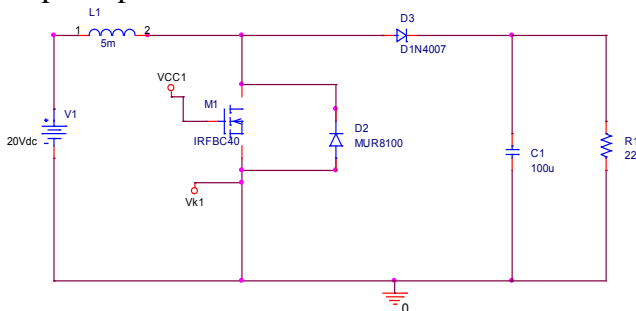


Figura 4- Etapa de potencia del MPPT propuesto.

C. Diseño de la etapa de control.

En la literatura se han propuesto una gran cantidad de opciones para el control de convertidores MPPT tales como Hill Climbing [11-13], P&O [14-15], conductancia incremental (IncCond) [16], sólo por mencionar algunos. De entre estos tipos de esquemas de control, el más

utilizado es el Hill Climbing (escalamiento de pendiente) debido a su relativa facilidad de implementación.

En el caso del circuito propuesto en este trabajo se eligió utilizar dicha técnica, la cual consiste en provocar una perturbación en el ciclo de trabajo del convertidor con la consecuente perturbación en la corriente y el voltaje que entrega el panel modificando así la potencia de salida. En la Fig. 5 se muestra de forma esquemática el principio de esta técnica de control.

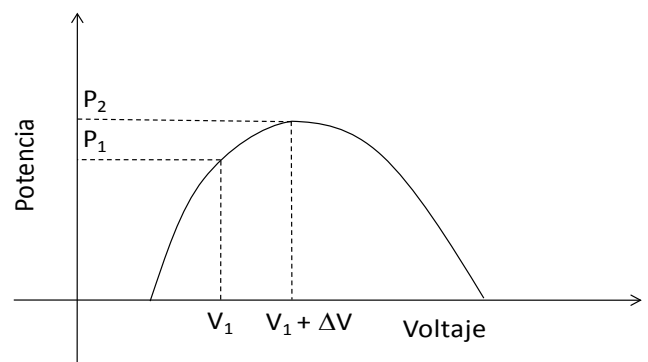


Figura 5- Principio de la técnica de control “Hill Climbing” para el rastreo del punto de máxima potencia.

La etapa de control fue implementada utilizando un microcontrolador PIC16F877A. Dicho microcontrolador recibe los valores de corriente y tensión en la carga a través de los puertos A/D y envía la señal de conmutación al driver IR2110 para conmutar al transistor MOSFET. El microcontrolador ejecuta el algoritmo de control que se muestra en la Fig. 6.

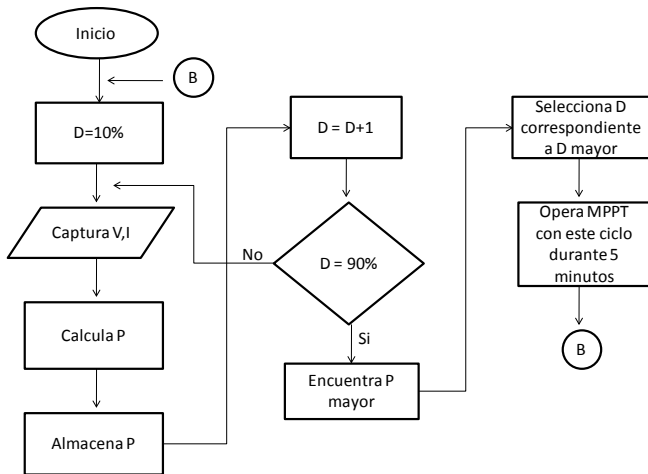


Figura 6- Algoritmo de control utilizado para el rastreo del punto de máxima potencia.

En este algoritmo, el microcontrolador inicia la operación del MPPT con un ciclo de trabajo de 10% y captura los valores de I_o y V_o para posteriormente calcular la potencia correspondiente. Este proceso lo ejecuta hasta terminar con el ciclo de trabajo de 90% y selecciona el ciclo de trabajo con el que se obtuvo la mayor potencia y con el cual se mantendrá operando durante 5 minutos antes de recomenzar el rastreo.

D. Simulación del circuito.

Antes de la implementación del circuito se efectuó la simulación del diseño en Orcad PSpice con el fin de verificar los valores de voltaje y corriente de salida, así como los valores de rizo, tal como se observa en la Fig.7. La simulación demuestra la elevación del voltaje en la carga así como un factor de rizo a la salida menor a 5%, lo cual es consistente con los cálculos efectuados durante el diseño.

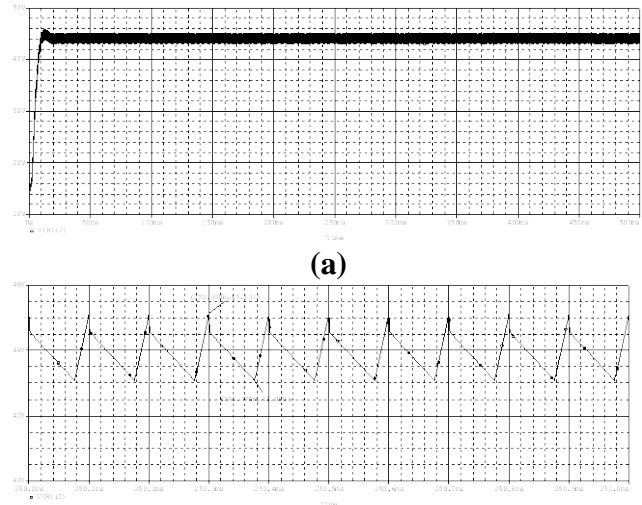


Figura 7- Curvas de voltaje de salida (a) y voltaje de rizo (b) obtenidas de la simulación en Orcad PSpice del circuito propuesto.

De acuerdo a la simulación, el voltaje promedio de salida es de 43V para un ciclo de trabajo de 0.76, mientras que el voltaje de rizo es igual a 2.04V, por lo que el factor de rizo del voltaje de salida es de 4.7%.

II.2 Ensamble y pruebas

Para la fabricación del inductor de 5mH, se utilizó un núcleo de hierro (permeabilidad relativa de 12) con 345 vueltas de alambre magneto de calibre No. 18 utilizando la Ec. (4) [17].

$$L = \frac{N^2 A \mu_0 \mu_r}{l} \quad (4)$$

donde:

N → Número de vueltas.

A → Área transversal del núcleo.

l → Longitud del núcleo.

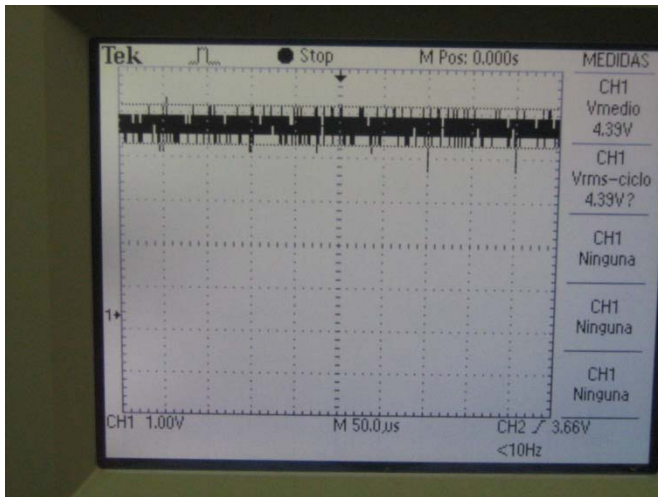
μ_0 → Permeabilidad del espacio libre.

μ_r → Permeabilidad relativa del núcleo.

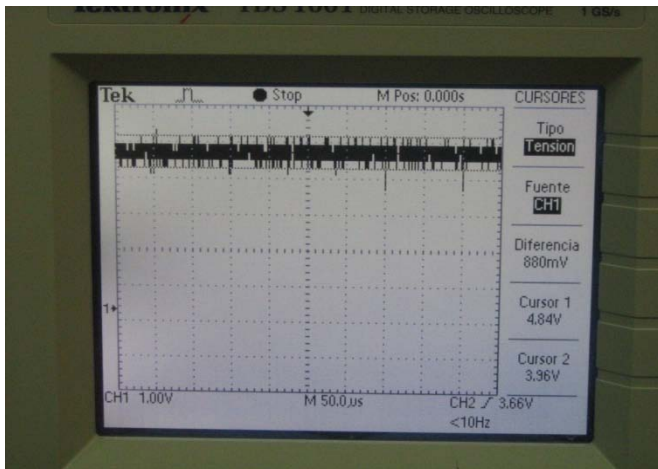
La entrada del circuito fue conectada a un panel fotovoltaico marca Kyocera de 85Wp (110W máximo), con una carga puramente resistiva (lámpara incandescente) con $R=18\Omega$. La señales de retroalimentación para la etapa de control (voltaje y corriente en la carga) fueron adquiridas

utilizando un sensor de efecto Hall (corriente) y un sensor de voltaje.

En la Fig. 8 se muestra el circuito MPPT operando en el punto de máxima potencia (ciclo de trabajo de 0.76), entregando un voltaje a la salida de 43.9V, una corriente de 2.44A y un voltaje de rizo de 0.88V.



(a)



(b)

Figura 8- Curvas de voltaje de salida (a) y voltaje de rizo (b) obtenidas en la resistencia de carga. La escala de la figura (a) está atenuada x10.

A partir de los valores obtenidos se puede calcular el factor de rizo, el cual es igual a 2%, consistente con el valor de diseño (menor a 5%), aunque considerablemente inferior al valor simulado.

Finalmente, en la Fig. 9 se muestra una imagen del panel fotovoltaico utilizado para las pruebas así como del circuito MPPT ensamblado en una tableta de pruebas.



(a)



(b)

Figura 9- (a) Panel solar de 85W utilizado para las pruebas del MPPT. (b) Montaje provisional del circuito.

III. CONCLUSIONES

Los resultados preliminares de este trabajo demuestran que el circuito propuesto permite extraer la máxima potencia de cualquier panel fotovoltaico menor o igual a 110W utilizando un algoritmo de control tipo Hill Climbing implementado con la ayuda de un microcontrolador PIC16F877A.

El circuito aquí propuesto representa una opción de bajo costo y fácilmente adaptable a sistemas fotovoltaicos nuevos o existentes permitiendo así un incremento de la relación beneficio/costo para el usuario final.

IV. REFERENCIAS

- [1] *Prospectiva del Sector Eléctrico 2008-2017, Secretaria de Energía* [En Línea], <http://www.energia.gob.mx> (consultado en febrero del 2009).
- [2] A. Luque, A. Marti et al., “Increasing the efficiency of ideal solar cells by photon induced transition at intermediate levels”, *Phys. Rev. Lett.*, Vol. 78, No. 26, p. 5014, 1997.
- [3] J. De la Torre, G. Bremond et al., “Silicon nanostructured layers for improvement of silicon solar cells efficiency”, *Mat. Sci. and Eng. C*, Vol. 26, No. 2-3, p. 427, 2006.
- [4] V. Svrcek, A. Slaoui et al., “Silicon nanocrystals as light converter for solar cells”, *Thin Solid Films*, Vol. 451-452, p. 384, 2004.
- [5] Wernher Swiegers and Johan H.R., *An Integrated Maximum Power Point Tracker for Photovoltaic Panels*, *Thin Solid Films*, Vol. 451-452, p. 384, 2004.
- [6] G. R. Walker, P. C. Sernia, “Cascaded DC-DC Converter Connection of Photovoltaic Modules”, *IEEE Trans. On Power Electron.*, Vol. 19, No. 4, p. 1130, 2004.
- [7] J. H. R. Enslin, M. S. Wolf et al. “Integrated photovoltaic maximum power point tracking converter”, *IEEE Trans. Ind. Electron.*, Vol. 44, p. 769-773, 1997.
- [8] R. Gules, J. De Pellegrin Pacheco et al., “A Maximum Power Point Tracking System With Parallel Connection for PV Stand-Alone Applications”, *IEEE trans. On Ind. Electron.*, Vol. 55, p.2674, 2008.
- [9] J. H. R. Enslin, D. B. Snyman, “Combined low-cost, high-efficient inverter, peak tracker and regulator for PV applications”, *IEEE Trans. Power Electron.*, Vol. 6, No. 1, p. 73, 1991.
- [10] A.M. Triznadlowsy, *Introduction to Modern Power Electronics*, 2nd Edition, Ed. Wiley Interscience, New York, 1998. ISBN 0471153036.
- [11] L. L. Buciarelli, B. L. Grossman, E. F. Lyon, and N. E. Rasmussen, “The energy balance associated with the use of aMPPT in a 100 kW peak power system,” in *IEEE Photovoltaic Spec. Conf.*, 1980, pp. 523–527.
- [12] Y. Kim, H. Jo, and D. Kim, “A new peak power tracker for cost-effective photovoltaic power system”, in *Proc. 31st Intersociety Energy Convers. Eng. Conf.*, 1996, pp. 1673–1678.
- [13] E. Koutroulis, K. Kalaitzakis, and N. C. Voulgaris, “Development of a microcontroller-based, photovoltaic maximum power point tracking control system”, *IEEE Trans. Power Electron.*, vol. 16, no. 21, pp. 46–54, Jan. 2001.
- [14] C. Hua and J. R. Lin, “DSP-based controller application in battery storage of photovoltaic system”, in *Proc. IEEE IECON 22nd Int. Conf. Ind. Electron., Contr. Instrum.*, 1996, pp. 1705–1710.
- [15] C.-C. Hua and J.-R. Lin, “Fully digital control of distributed photovoltaic power systems”, in *Proc. IEEE Int. Symp. Ind. Electron.*, 2001, pp. 1–6.
- [16] O. Wasynczuk, “Dynamic behavior of a class of photovoltaic power systems”, *IEEE Trans. Power App. Syst.*, vol. 102, no. 9, pp. 3031–3037, Sep. 1983.
- [17] R.L. Boylestad, *Análisis introductorio de circuitos*, 3a edición, Ed. Trillas, México, 2005, p.275.

Autores

Fernando Ponce Márquez. Estudiante del 9° semestre de la carrera de Ingeniería Eléctrica en la Universidad Autónoma de Zacatecas. Actualmente se encuentra realizando su proyecto de tesis.

Julio Adrian Zapata Domínguez. Estudiante del 9° semestre de la carrera de Ingeniería Eléctrica en la Universidad Autónoma de Zacatecas. Actualmente se encuentra realizando su proyecto de tesis.

Dr. Jorge de la Torre y Ramos. Profesor investigador de tiempo completo en la Unidad Académica de Ingeniería Eléctrica de la Universidad Autónoma de Zacatecas.

Dr. Manuel Reta Hernández. Profesor investigador de tiempo completo en la Unidad Académica de Ingeniería Eléctrica de la Universidad Autónoma de Zacatecas.

REVISIÓN DEL ESTADO DEL ARTE DE LAS ESTUFAS DE COCCIÓN BAJO EL PRINCIPIO DE CALENTAMIENTO POR INDUCCIÓN

Chávez García J. Ángel, Rodríguez Segura Elías J. J., Oroasco Guerrero Rodolfo,
Vázquez Nava Nimrod

Área de Posgrado, Departamento de Electrónica
Instituto Tecnológico de Celaya
Av. Tecnológico y A. García Cubas S / N, C.P. 38010, Celaya Guanajuato
Tel. 01(461) 611 75 75, correo electrónico: jacg04@hotmail.com

RESUMEN

En el presente artículo, se expone el estado del arte referente a una de las aplicaciones tecnológicas del calentamiento por inducción electromagnética cada vez más empleada gracias al desarrollo de los dispositivos semiconductores de potencia, esta tecnología son las estufas de calentamiento por inducción electromagnética. Este tipo de estufas ya compiten tecnológicamente con las estufas de quemadores convencionales. Asimismo, se expondrán las diferentes topologías que son empleadas en la construcción de estos dispositivos, así como las tecnologías que se emplean en su implementación práctica, y los factores que se deben de considerar cuando se desee construir un prototipo.

Palabras clave: Calentamiento por inducción (IH), inversor resonante, corrección del factor de potencia (CFP), contenido armónico, bobina de excitación, modulación por anchura de pulsos (PWM).

I. INTRODUCCIÓN

Un sistema de calentamiento por inducción puede calentar recipientes para cocción directamente debido a las pérdidas por corrientes de Eddy [1]. La corriente de alta frecuencia en una bobina de excitación produce una variación en el campo magnético el cual induce corrientes de Eddy que fluyen en una vía circular (fig. 1).

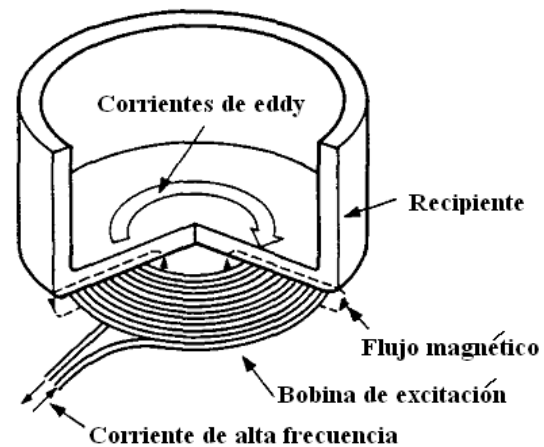


Figura 1. Diagrama esquemático de un sistema de calentamiento por inducción.

Las estufas de cocción son aplicaciones bien conocidas de la inducción electromagnética basándose en el manejo de la energía para calentar. En años recientes, los aparatos de calentamiento por inducción como son las estufas de IH, son cada vez más y más populares debido a su limpieza, alta eficiencia, seguridad y controlabilidad [2]. Las variantes de estas estufas tienen su diversidad, la cual depende de la aplicación final a la cual están destinadas y debido a ello se exponen diversas topologías que buscan cumplir con las necesidades planteadas. Equipos de IH empleando topologías de inversores de alta frecuencia tiene ventajas prácticas de seguridad, rentabilidad, ahorro de energía, ambientes limpios, eficiencia muy alta en la conversión de energía, proceso de calentamiento

enfocado y rápido, alta densidad de potencia, alta rentabilidad, ambientes libres de ruido electromagnético. Las estufas de inducción están diseñadas para competir con las estufas de quemadores convencionales. Las ventajas y desventajas principales son:

- ✓ La bobina de la estufa permanece fría y por lo tanto es más segura
- ✓ Alta eficiencia, gastos de electricidad bajos
- ✓ Potencia de salida constante
- ↓ Costo inicial alto y,
- ↓ Solo recipientes con alta resistividad y permeabilidad relativa pueden ser utilizados en principio.

Con el gran avance en los dispositivos semiconductores de potencia MOS, IGBTs, etc. y sus módulos IC de disparo, el desarrollo del estado del arte en aplicaciones de calentamiento de fluidos o metales basado en inducción electromagnética incorporando inversores resonantes alimentados en voltaje o en corriente, tiene un interés agregado en el campo de la electrónica de potencia moderna.

Debido a la naturaleza inductiva de la bobina, los convertidores empleados son usualmente del tipo resonante, por lo que esta bobina se hace acompañar de un capacitor [3]. La topología de circuitos inversores resonantes está dividida generalmente en tres tipos [4]:

- 1) Tipo resonante serie con capacitor de compensación.
- 2) Tipo resonante paralelo con capacitor de compensación.
- 3) Tipo resonante serie- paralelo con capacitor de compensación.

Estas topologías están acompañadas de circuitos adicionales cuando se emplean en prototipos de calentamiento por inducción debido principalmente a que presentan una desventaja muy significativa. El rango de operación para la conmutación suave

bajo un esquema de modulación PWM es estrecho. Esto es un factor indeseable, porque se requieren amplios rangos de operación de conmutación suave para operar en alta frecuencia.

II. PRINCIPIO DE OPERACIÓN

La fig. 2 muestra la configuración esquemática de una estufa de IH para el hogar. Un inversor de alta frecuencia suministra corriente de alta frecuencia a una bobina de excitación plana, entonces las corrientes de Eddy inducidas fluyen dentro del recipiente o cacerola en base a la ley de inducción electromagnética de Faraday. De acuerdo con dicha ley, una gran conversión térmica es producida con una alta eficiencia [5].

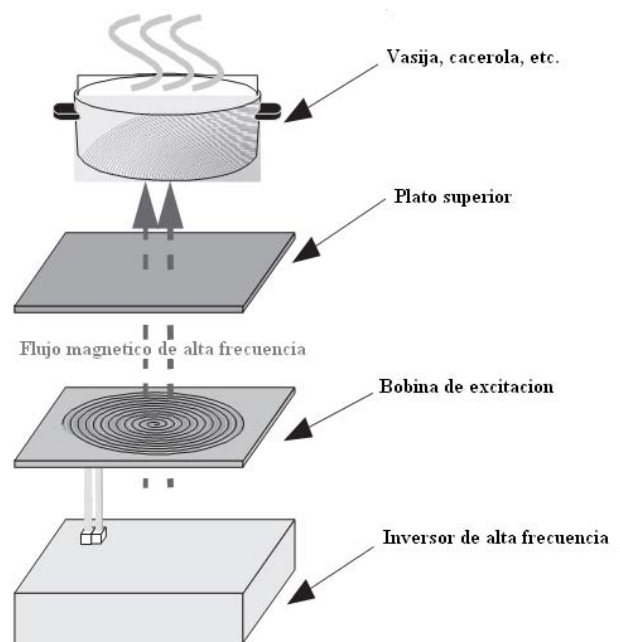


Figura 2. Esquema de un aparato de cocción por inducción.

En el caso de equipos multiquemador IH, la potencia de salida de cada quemador debe ser controlada bajo un esquema de frecuencia constante. De otra manera tendría lugar un choque acústico provocado por la diferencia de frecuencias de operación de los inversores.

Por otro lado, para disminuir la disipación de potencia en la bobina de excitación, es necesario bloquear componentes de corriente de baja frecuencia, las cuales no son útiles para el calentamiento por inducción debido a que provoca interferencia de baja frecuencia en la bobina de excitación, la cual es amplificada por los recipientes de carga[6]. Este medio de regulación de potencia a frecuencia constante es indispensable para aplicaciones de calentamiento por inducción. Más aún, para reducir los armónicos de la corriente de línea de CA, una etapa de CFP debe ser agregada, a fin de cumplir con las regulaciones [2].

III. TOPOLOGÍAS

Las topologías que se van a mencionar comparten elementos comunes entre ellas, como son el uso de inversores resonantes, bobinas de excitación y ponen especial énfasis en las técnicas de conmutación de los interruptores de los que consta el prototipo propuesto en cada caso. Estas van desde topologías básicas hasta puente completo, las cuales determinan la potencia de salida que puede entregar cada estufa. En todos los casos, los autores implementaron sus diseños validando y comparando mediciones de eficiencia, así como de corriente y voltaje donde se observe la corrección del factor de potencia.

Inversor elevador resonante PWM a conmutación suave medio puente

La fig. 3(a) muestra el desarrollo previo de un convertidor PWM con conmutación suave empleando un circuito de enclavamiento activo, el cual está compuesto de un interruptor activo Q_s y un capacitor C_s . La función del circuito de enclavamiento es reducir los esfuerzos en tensión – voltaje del interruptor principal de potencia Q_1 y habilitar la operación VPCF (frecuencia constante potencia variable). Este convertidor tiene algunas limitaciones como son que no puede elevar el voltaje de salida. Adicionalmente, la corriente en la bobina de excitación tiene una componente de la línea de CA, la cual no es útil para el IH e

incrementa las pérdidas en la bobina de excitación y en los interruptores SW_1 y SW_s [5].

Se incorpora una configuración de una sola etapa al convertidor con conmutación suave a alta frecuencia, lo cual se puede conseguir combinando la función de convertidor elevador y la de inversor medio puente en una conversión de una etapa. La fig. 3b muestra la configuración de un nuevo circuito del convertidor propuesto PWM a conmutación suave de una etapa, incorporando solo dos interruptores para este convertidor. El circuito inversor incorpora una topología de dos dispositivos de conmutación $Q_1(S_1/D_1)$, $Q_s(S_s/D_s)$, dos capacitores en serie divididos en C_s y C_{2b} y un capacitor compensador de carga inductiva C_1 , en paralelo con la bobina de excitación L_1 . También incluye el elevador de voltaje (*charge – up*), el cual está compuesto de un inductor elevador L_{1b} y un interruptor de potencia $Q_1(S_1/D_1)$ [5].

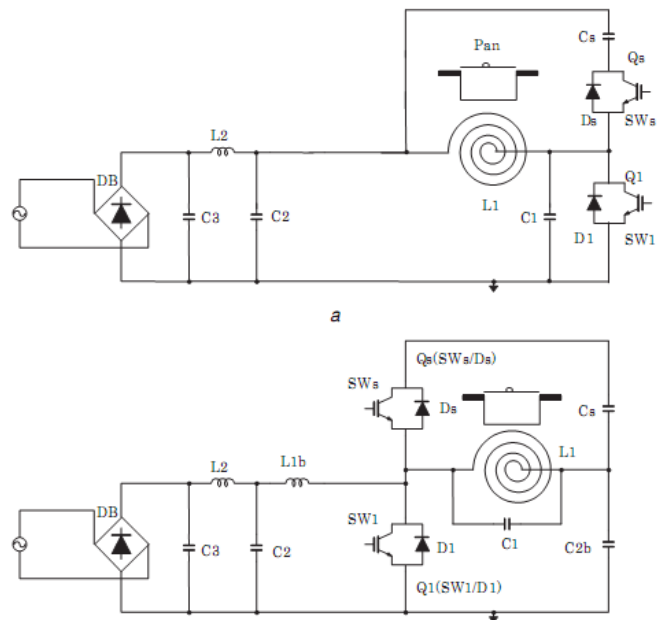


Figura 3. a Desarrollo previo de convertidor de alta frecuencia con conmutación suave. b Convertidor de potencia propuesto.

Como se puede ver en la configuración del circuito de la topología propuesta, el dispositivo de conmutación Q_1 (SW_1/D_1) comparte y desarrolla la operación tanto de convertidor elevador

monofásico, como de inversor PWM ZVS de alta frecuencia. Adicionalmente, los capacitores C_s y C_{2b} se emplean para filtrar la componente de baja frecuencia que fluye a través de la bobina de excitación del IH, la cual se construye de hilo de litz.

Convertidor resonante serie medio puente

La topología resonante serie, medio puente (fig. 4) se elige de las topologías de conmutación simple debido a las siguientes razones [7]:

- El voltaje en los dispositivos de conmutación es fijado por medio de los dos capacitores C_r . Aunque son necesarios dos interruptores, al menos la capacidad de bloquear la mitad del voltaje es requerida.
- Debido a la topología empleada, se usa un ciclo de trabajo del 50 %, lo cual simplifica el control requerido.

Interruptores con diodo anti paralelo deben ser usados como se muestran en la fig. 5. IGBTs son elegidos por ser adecuados para una aplicación de frecuencia media y alta corriente.

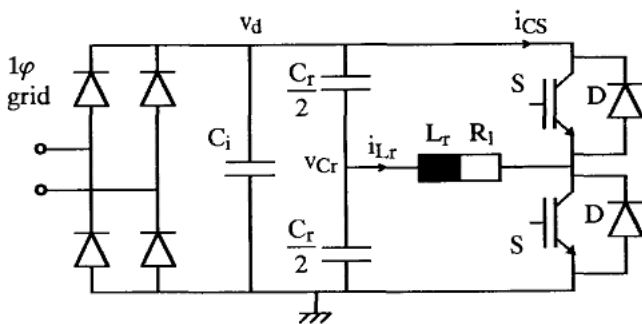


Figura 4. Circuito resonante serie medio puente.

Existen dos posibles estrategias para este convertidor con el fin de obtener el control de potencia deseada sin variar el voltaje de entrada, éstas son: Por conmutación de carga y conmutación forzada.

El convertidor a conmutación forzada es elegido debido a la eliminación de la corriente de

recuperación inversa de los diodos y a que la máxima potencia es obtenida en el punto bajo del rango de la frecuencia de conmutación, usualmente cercana a 20 KHz.

El factor de calidad no es constante como una función de la frecuencia de conmutación f_s , pero disminuye cuando f_s se incrementa arriba de f_o debido a que:

- R_l se incrementa por el efecto piel.
- El incremento en R_l debido al contenido armónico de orden superior cuando la forma de onda de la corriente en la bobina comienza a ser de forma sinusoidal y
- El decremento en L_r cuando mejora el acoplamiento magnético.

Inversor puente completo con control CFP en modo de corriente discontinua

Un inversor a conmutación suave de cambio de fase con una función de elevación y corrección de factor de potencia es propuesto para equipos de potencia en [2]. El sistema propuesto consiste de dos etapas de potencia, una primera etapa que consta de un convertidor elevador de voltaje con corrección del factor de potencia y conmutación suave y una segunda etapa conformada por un inversor puente completo de cambio de fase, con ramas que desarrollan conmutación a voltaje y corriente cero (ZVS y ZCS). Las características de operación del sistema de potencia propuesto realizables emplean interruptores IGBTs y diodos de carburo de silicio (SiC), lo cual se evalúa e ilustra con base a resultados experimentales. Este presenta un método de control simple para la etapa de corrección, el cual puede reducir el contenido armónico de corriente.

La primera etapa es un convertidor elevador CFP y tiene dos funciones, una tiene que ver con la corrección del factor de potencia del sistema total con modo de operación de corriente discontinua; la otra función es la operación de regulación de la potencia de salida para diversos recipientes, por medio de la elevación del voltaje V_c . El

dispositivo de conmutación Q1 opera bajo la técnica de conmutación ZCS en el encendido por medio del inductor *snubber* L2 y conmutación ZVS en el apagado por medio del capacitor *snubber* C1, asimismo, la corriente del inductor elevador i_{L1} es controlada bajo el modo de corriente discontinua (DCM) con un ciclo de trabajo constante.

En la etapa del circuito inversor, dispositivos de conmutación Q2 y Q3 en la rama de corriente operan bajo conmutación ZVS y ZCS en el encendido y conmutación ZVS en el apagado por medio del capacitor *snubber* Csn. Por otro lado, los dispositivos de conmutación Q4 y Q5 en la rama de corriente principal operan bajo conmutación ZCS al encendido asistida por el inductor *snubber* Lsn y conmutación ZVS & ZCS al apagado. La potencia de salida de alta frecuencia es regulada por el cambio del ángulo de fase ϕ , lo cual se consigue al variar el voltaje VCO en la entrada de la etapa inversora. La fig. 5 muestra la configuración del circuito. Esta topología de circuito consiste de una etapa de elevación CFP que consta de un filtro paso bajas formado por Lf y Cf, un inductor elevador L1, dispositivos de conmutación Q1 (S1 / D1) con su inductor *snubber* de reducción de pérdidas L2 y un capacitor *snubber* C1, C2, un capacitor de alisamiento de cd intermedio C0.

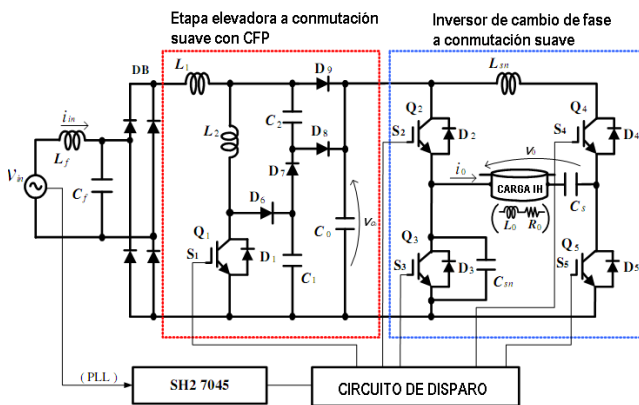


Figura 5. Inversor propuesto a conmutación suave y cambio de fase con convertidor CFP elevador.

Adicionalmente, la etapa de conversión de alta frecuencia comprende la función ZVS, integrada por los dispositivos de conmutación Q2 (S2/ D2) y Q3 (S3/ D3), un capacitor *snubber* Csn. Asimismo, la función ZCS, la integran los dispositivos de conmutación Q4 (S4/ D4) y Q5 (S5/ D5) y un inductor *snubber* Lsn. La carga a calentar junto con la bobina de excitación es descrita como L0 & R0 con un capacitor en serie Cs. El capacitor de filtrado de cd C0 debe ser diseñado para eliminar la componente de corriente de baja frecuencia de la línea de CA a través de la bobina de excitación. La frecuencia de conmutación del circuito propuesto es de 60 KHz.

Esta topología presenta la ventaja de poder regular el ciclo de trabajo en la etapa del convertidor, así como el cambio del ángulo de fase ϕ , para conseguir regular cargas distintas en términos de permeabilidad magnética y resistividad, como lo es el cobre y el acero inoxidable. Para el caso de recipientes de acero inoxidable esta regulación está en el rango de 0.5 a 2.5 KW. En el caso de recipientes de cobre la regulación estará en el orden de 0.5 a 1.7 KW, como se puede observar en la figura 6.

La eficiencia en la conversión de energía en el circuito propuesto versus potencia de entrada se muestra en la figura 7. En esta figura se observa, por ejemplo que la eficiencia de conversión de energía a una $P_{in} = 2.5$ KW con una cacerola de acero inoxidable es del 84 %.

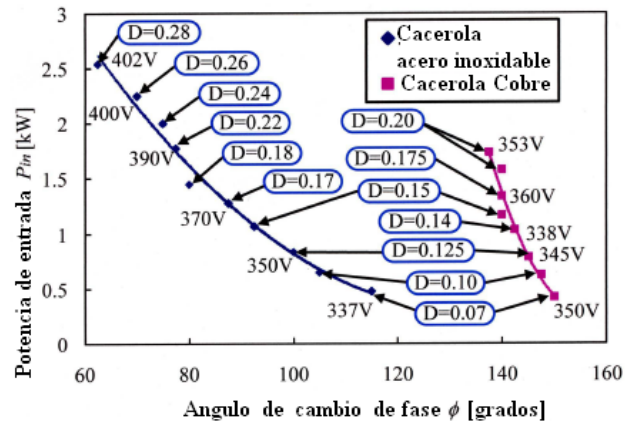


Figura 6. Potencia de salida en función del ángulo de cambio de fase.

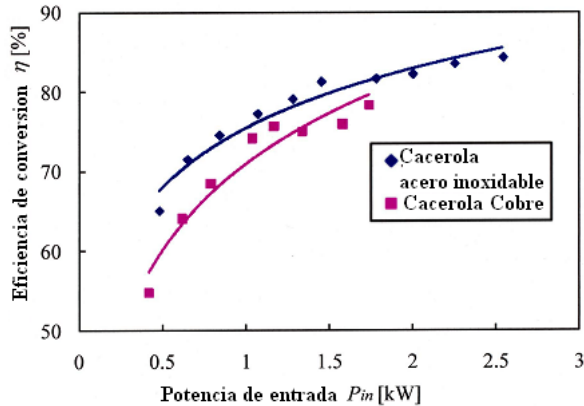


Figura 7. Eficiencia en la conversión de energía.

La figura 8 muestra las formas de onda de voltaje y corriente de línea de CA con un ciclo de trabajo $D = 0.28$ y el ángulo de cambio de fase $\phi = 62.5$ grados, empleando una cacerola de acero inoxidable y potencia de entrada de 2.5 KW. Como se muestra en 8(a), i_{L1} es controlada bajo un modo de corriente discontinua con un ciclo de trabajo constante. La figura 8(b) muestra el voltaje y la corriente de línea, se puede apreciar que la corriente tiende a ser sinusoidal y tiende a presentar un factor de potencia de 0.97.

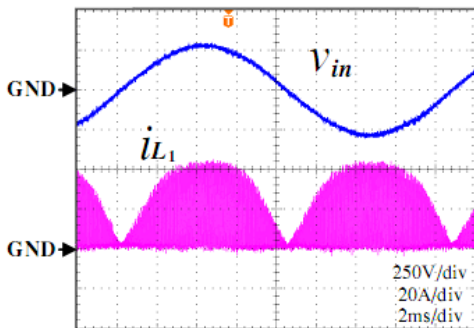


Figura 8(a). Corriente en el inductor elevador y voltaje de entrada de línea.

Inversor ZVS – PWM cuasi – resonante con corrección del factor de potencia

Se presenta un prototipo mejorado de un inversor resonante alimentado en voltaje que emplea IGBTs como interruptores bajo un esquema de control ZVS – PWM, el cual hace uso de los modos de operación resonante y cuasi resonante.

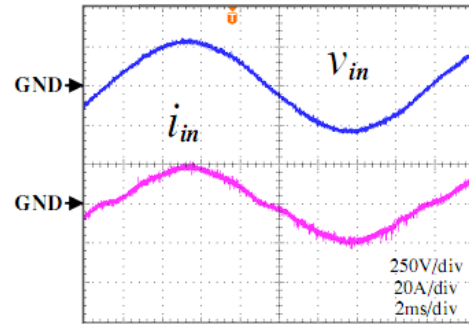


Figura 8(b). Corriente y voltaje de entrada de la línea de CA.

Este inversor es aplicado a estufas de IH de alta potencia y también puede ser empleado en calentadores de agua de IH, así como a vaporizadores de alta temperatura. El inversor ZVS – PWM de carga resonante con función de operación de filtro activo, es puesto en práctica empleando una estufa de IH que va de 2 – 10 KW de potencia de salida.

Para aplicaciones domesticas, un inversor cuasi – resonante ZVS – PWM con un solo interruptor está siendo usado actualmente. De igual forma, de acuerdo al uso de aplicación específica en negocios, seleccionar topologías de medio puente y puente completo puede ser adecuado. Los inversores cuasi - resonantes de un solo interruptor operado bajo conmutación suave a voltaje o corriente cero, son usados en aparatos caseros que están en el rango de 1 – 2 KW y un voltaje de línea de 110 V. También, los inversores cuasi – resonantes tipo *push – pull* terminación simple también pueden ser empleados para aplicaciones de negocio en el rango de potencia de 3 – 50 KW o más, para un voltaje de línea de 220 V.

La fig. 9, muestra las formas de onda de voltaje y corriente que corresponden al interruptor del inversor por modulación de frecuencia de pulsos ZVS (ZVS– PFM) empleando un solo IGBT (tipo 1) y de un inversor medio puente PWM a conmutación dura sin *snubber* capacitivo o inductivo, o bien con *snubber* resonante activo (tipo 2) comparándolo con el inversor ZVS – PWM asimétrico cuasi resonante (tipo 3).

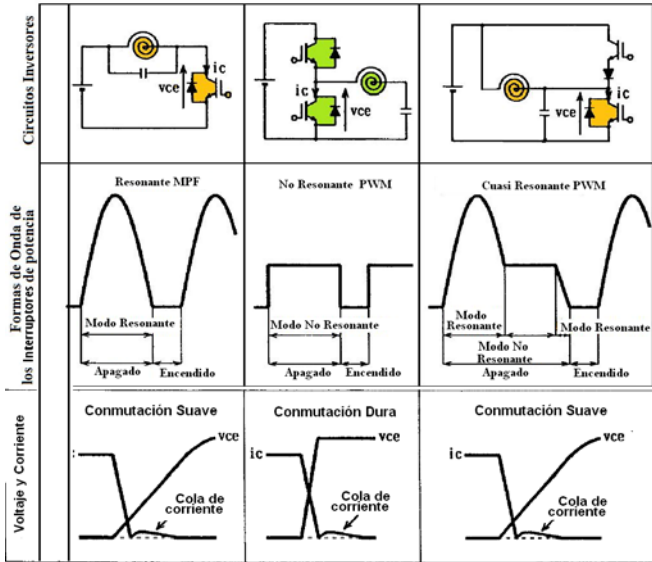


Figura 9. Formas de onda de voltaje y corriente de tres topologías de inversores resonantes.

En primer lugar, el inversor cuasi resonante ZVS – PFM (Modulación por pulsos de frecuencia) del tipo 1 puede ser operado bajo una condición de control a conmutación suave y voltaje cero para la regulación de potencia de salida debido al esquema simple PFM. La topología del tipo 1 tiene la desventaja significativa de un alto voltaje pico en el interruptor. Esta topología no está disponible para conectarse a voltajes de línea de 220 V para aplicaciones de alta potencia. En segundo término, el inversor PWM medio puente a conmutación dura sin *snubber* resonante, operado a frecuencia fija, tiene la gran ventaja de no tener ruido acústico que puede ser generado por una estufa de inducción de dos o más quemadores, sin embargo, el inversor tipo 2 tiene el problema inherente de generar grandes pérdidas por conmutación a frecuencias mayores de 20 KHz, adicionalmente se genera una gran cantidad de ruido EMI /RFI debido a la operación de conmutación dura, con una nueva topología de inversor cuasi – resonante ZVS – PWM *push – pull* a conmutación suave y con una estrategia de control de potencia variable a frecuencia constante (CFVP).

Inversor ZVS – PWM resonante

La fig. 10 muestra un circuito esquemático de la configuración del sistema para estufas de cocción.

El cual está compuesto de un inversor resonante ZVS – PWM tipo *push–pull* (tipo 3) empleando una nueva generación de IGBTs como interruptores. Esta estufa de IH se conecta a la línea de CA de 220 V, destinado especialmente para una estructura de quemadores múltiples para aplicaciones domesticas y de negocio. Este inversor resonante de alta frecuencia con función CFVP es conectado a un puente rectificador de diodos con un filtro de alisamiento, el cual puede realizar sin problema la corrección del factor de potencia unitario. Con esta topología y esquema de control se pueden combinar diversas cargas de recipientes o cacerolas acopladas a la bobina de calentamiento.

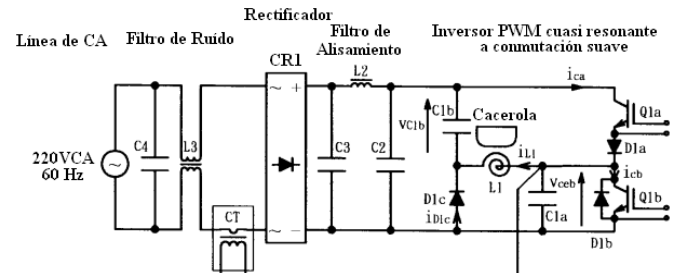


Figura 10. Configuración de sistema cuasi – resonante ZVS – PWM.

Su circuito de control está basado en la regulación de potencia y protección del sistema mediante una PC. En este circuito inversor, el interruptor de SW1a funciona como un interruptor auxiliar y el interruptor SW1b funciona como interruptor principal, en ambos casos se incorpora un IGBT de canal estrecho y bajo voltaje de saturación.

Con esta topología de inversor medio puente ZVS – PWM cuasi – resonante, la operación de conmutación suave a voltaje cero bajo la estrategia de control a potencia variable y frecuencia constante (CFVP) puede ser desarrollada fácilmente para varios tipos de cargas posibles del IH, estas cargas son de hierro, de acero inoxidable, cobre y aluminio.

La fig. 11 muestra las formas de onda medidas de la corriente y el voltaje de entrada (I_{in} y V_{in}) así como su espectro en frecuencia. Mediciones típicas de voltaje y corriente del inversor empleando los nuevos IGBTs se muestran en la figura 12. Como

se puede ver, está claramente probado que Q1b puede operar bajo condición de conmutación suave a voltaje cero. Adicionalmente, la operación de conmutación suave de Q1b se puede obtener en un rango seguro para el proceso de regulación PWM.

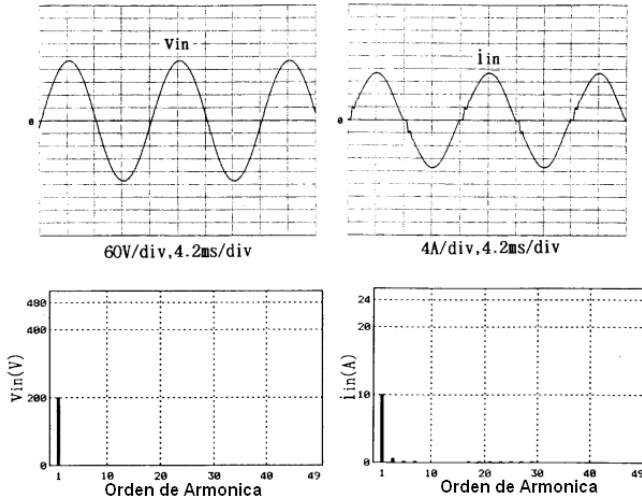


Figura 11. Formas de onda del voltaje y corriente de entrada y sus espectros armónicos.

Es importante mencionar que las operaciones de conmutación suave de Q1a y Q1b pueden ser implementadas en base a una estrategia de regulación PWM a frecuencia constante para regular la potencia de salida de este último.

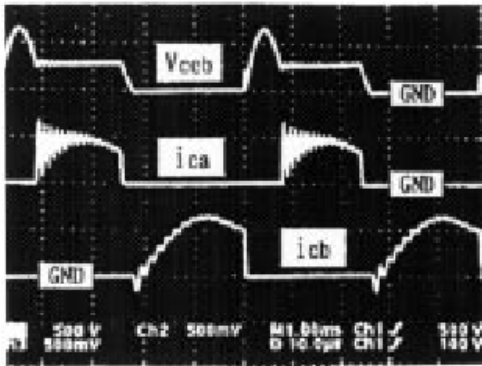


Figura 12a. Forma de onda de corriente y voltaje bajo la condición de operación de conmutación suave Vceb: 500 V/div; ica: 50 A/div; icb: 50 A/div.

La fig. 13 muestra la regulación de la potencia de entrada del inversor controlando el ciclo de trabajo

de Q1a (IGBT) bajo una condición de operación a frecuencia constante, la cual está fija en 20.5 KHz.

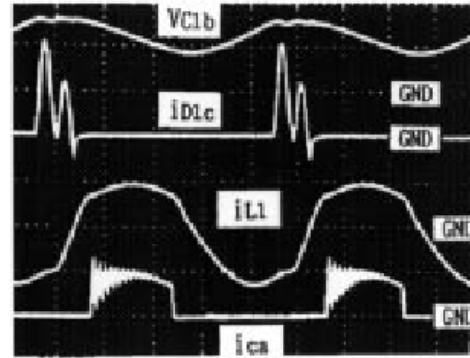


Figura 12b. Forma de onda de corriente y voltaje bajo la condición de operación de conmutación suave Vc1b:500 V/div; iD1c: 50 A/div; iL1: 50 A/div; 10 A/div.

El ciclo de trabajo definido como D_a puede ser controlado con el propósito de regular la potencia de salida bajo un esquema de frecuencia constante, en este caso, la potencia de entrada máxima del inversor puede ser obtenida para un ciclo de trabajo $D_a \approx 0.5$ en Q1a.

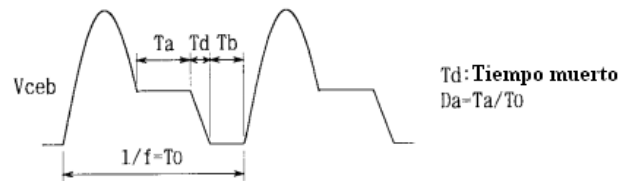
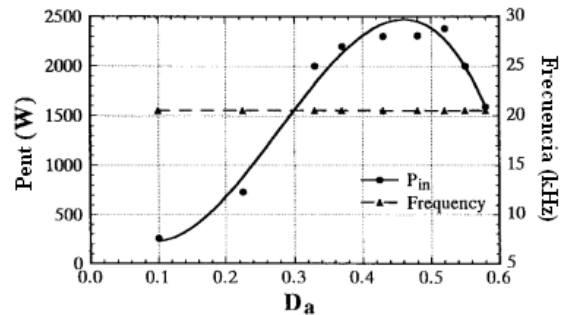


Figura 13. Característica de la regulación de potencia del inversor PWM cuasi resonante con CFPV.

IV. FACTORES A CONSIDERAR

Desde el punto de vista práctico, uno de los factores a considerar involucra a la bobina de

excitación, dicha bobina debe cumplir con tres funciones [8]:

- Transferir la energía eléctrica de la red hacia el recipiente por medio de un campo magnético.
- Distribuir dicha energía de manera uniforme en el fondo del recipiente.
- Formar parte del tanque resonante del inversor.

La bobina de excitación tiene la forma de una bobina plana, buscando reducir las pérdidas de potencia en el inductor por concepto de la frecuencia de operación del inversor, la bobina se construye con alambre de litz. Aunado a esto, es deseable que las líneas de flujo magnético se concentren lo más cerca posible de la bobina, para que sea posible transferir la máxima potencia al recipiente, para ello se emplean concentradores magnéticos, contruidos regularmente de ferrita.

Cuando ocurre un corto circuito, se incrementa abruptamente la frecuencia de resonancia, ya que el cortocircuito hace que el valor de la inductancia de la bobina de excitación baje notablemente, llevándolo a una condición de conmutación inductiva o bien capacitiva. Por tal motivo deben apagarse los interruptores para el caso de un inversor serie y encenderse si se trata de un inversor paralelo. Todo lo anterior con un retardo entre el momento que ocurre el cortocircuito y el momento que se conmutan los interruptores [9].

La frecuencia del inversor debe mantenerse en la frecuencia de resonancia, esto se puede lograr por medio de un circuito de sincronización de fase (PLL –por sus siglas en inglés – *Phase Locked Loop*), el cual consiste de un oscilador controlado por voltaje (VCO, por sus siglas en inglés – *Voltage Controlled Oscillator*-) y un detector de fase, donde el voltaje de entrada del VCO es proporcional a la frecuencia de resonancia [1].

V. CONCLUSIONES

Se presenta una revisión al estado del arte de las estufas de cocción de calentamiento por inducción, como resultado de ello, se ha pretendido con el presente artículo resaltar las características de operación, así como los principios de funcionamiento que presenta cada topología expuesta. En los artículos revisados se contemplan las características de los inversores que constituyen la estufa de inducción, asimismo la técnica de control que emplea cada una de las topologías, en la cual existe Modulación por anchura de pulsos (PWM) acompañado de conmutación suave, la cual es una técnica a frecuencia fija y representa la técnica mas empleada en el control de inversores resonantes en sus variantes PWM asimétrico, ZVS – PWM. Modulación por Pulsos de Frecuencia (PFM), la cual es una técnica a frecuencia variable. El control de cambio de fase, es otra técnica empleada, con la que es posible calentar recipientes tanto con alta como baja resistividad y permeabilidad magnética. Una técnica más compleja es la llamada cancelación de voltaje asimétrico [10]. Para el caso de las técnicas de control a frecuencia variable, la cual es una forma tradicional de control de potencia o voltaje de inversores resonantes, presenta diversas desventajas: dificultad de reducir la interferencia electromagnética (EMI) debido a un amplio espectro de ruido, filtrado más complejo del rizo del voltaje de salida, así como un pobre aprovechamiento de los componentes magnéticos. Estos problemas se solucionan empleando técnicas de control a frecuencia fija o bien con un rango estrecho de frecuencia. Sin embargo estas técnicas no garantizan conmutación suave ante variaciones en la potencia de salida, o bien ante grandes cambios de carga (es decir, empleando recipientes de diversos materiales), lo cual condiciona a incrementar la frecuencia de conmutación para lograr la conmutación suave.

Otro factor de suma importancia es el cumplimiento de las normas de inyección de contenido armónico a la red de eléctrica. La frecuencia de operación del inversor para

aplicaciones de estufas de inducción va desde los 20 KHz hasta los 50 KHz, sabiendo ahora que su elección tiene que ver con el tipo de material de los recipientes de cocción, que cuando son de hierro o de acero inoxidable, materiales que tienen una resistividad y una permeabilidad magnética alta, están en el rango de 20 KHz y cuando los recipientes son de cobre o de aluminio, materiales que tienen una resistividad y permeabilidad baja, su frecuencia de operación está en 50 KHz [1].

El desarrollo de un equipo de inducción, particularmente de una estufa de inducción enfoca su diseño a considerar aspectos de rentabilidad y confiabilidad. La rentabilidad ha de estar determinada por el costo final del sistema diseñado, factor relacionado con la simplicidad del diseño y el beneficio que se obtendrá de este mismo.

VI. REFERENCIAS

- [1] Teruya Tanaka, *A new induction cooking range for heating any kind of metal vessels*, Consumer Electronics, IEEE Transactions on Volume 35, Issue 3, Aug 1989 Page(s):635 – 641
- [2] Kawaguchi, Y.; Hiraki, E.; Tanaka, T.; Nakaoka, M.; Fujita, A.; Omori, H., *Feasible evaluation of a full-bridge inverter for induction heating cooking appliances with discontinuous current mode PFC control*, Power Electronics Specialists Conference, 2008. PESC 2008. IEEE Volume , Issue , 15-19 June 2008 Page(s):2948 – 2953
- [3] Forest François, Faucher Sébastien, Gaspard Jean-Yves, Montloup Didier, Huselstein Jean-Jacques, Joubert Charles *Frequency-Synchronized Resonant Converters for the Supply of Multiwinding Coils in Induction Cooking Appliances*, IEEE Transactions on Industrial Electronics, Vol. 54, No. 1, February 2007.
- [4] Shengpei Wang, Kiyoshi Izaki, Izuo Hirota, Hidekazu Yamashita, Hideki Omori, Mutsuo Nakaoka, *Induction-Heated Cooking Appliance Using New Quasi-Resonant ZVS-PWM Inverter With Power Factor Correction*, IEEE Transactions on Industry Applications, Vol. 34, no. 4, July/August 1998.
- [5] N.A. Ahmed, M. Nakaoka, *Boost-half-bridge edge resonant soft switching PWM high-frequency inverter for consumer induction heating appliances*, IEE Proc.-Electr. Power Appl., Vol. 153, No. 6, November 2006.
- [6] M. Burdío José, Monverde Fernando, R. García José, Barragán Luis A., Martínez Abelardo, *A Two-Output Series-Resonant Inverter for Induction-Heating Cooking Appliances*, IEEE Transactions on Power Electronics, Vol. 20, No. 4, July 2005.
- [7] Henry W. Koertzen, Jacobus D. Van Wyk, Jan A. Ferreira, *Design of the half-bridge, series resonant converter for induction cooking*, Power Electronics Specialists Conference, 1995. PESC apos; 95 Record., 26th Annual IEEE Volume 2, Issue , 18-22 Jun 1995 Page(s):729 - 735 vol.2
- [8] Acero Jesús, J. Hernández Pablo, Burdío José M., Alonso Rafael, Barragán Luis A., *Simple Resistance Calculation in Litz-Wire Planar Windings for Induction Cooking Appliances* IEEE Transactions on Magnetics, Vol. 41, no. 4, April 2005.
- [9] Burdío José M., Barragán Luis A, Monverde Fernando, Navarro Denis, Acero Jesús, *Behaviour of Series and Parallel Resonant Inverters for Induction Heating in Short-circuit Conditions*, IEEE Power Electronics and Motion Control Conference, PIEMC 2000.
- [10] Burdío José M., Barragán Luis A, Monverde Fernando, Navarro Denis, Acero Jesús, *Asymmetrical Voltage-Cancellation Control for Full-Bridge Series Resonant Inverters*, IEEE Transactions on Power Electronics, Vol. 19, No. 2, March 2004.

CARACTERIZACIÓN DE BATERÍAS Y PÁNELES DE UN SISTEMA FOTOVOLTAICO AUTÓNOMO

Segura Flores Luis Manuel, De La Torre y Ramos Jorge, Reta Hernández Manuel

Unidad Académica de Ingeniería Eléctrica
 Universidad Autónoma de Zacatecas
 Av. López Velarde No. 801, Col. Centro, Zacatecas, Zacatecas 98000, México
 Tel. 49239407 ext. 1505, correo electrónico: luiveris@hotmail.com

RESUMEN

En el presente trabajo se muestran los resultados de la caracterización de baterías y paneles solares utilizados en un sistema fotovoltaico autónomo. La caracterización de las baterías demuestra que es posible utilizar un modelo propuesto en la literatura para predecir el estado de carga (SOC) y el estado de salud (SOH) de las mismas, permitiéndole al usuario conocer la vida útil restante de las mismas. Por otra parte, la caracterización de los paneles solares permite determinar, o bien confirmar, la potencia máxima real generada indicada por el fabricante así como la vida útil del panel. Igualmente, se puede determinar la eficiencia y potencia generada por módulos de los cuales no se tenga información.

Palabras clave: SFV, Baterías, Panel fotovoltaico.

I. INTRODUCCIÓN

Cada día, los sistemas fotovoltaicos (SFV) tienen mayor presencia en México y, aunque es una tecnología con varios años de desarrollo a nivel mundial, en nuestro país su implementación y difusión aún es incipiente. Un aspecto trascendental para fomentar el uso generalizado de estas tecnologías, es la existencia de laboratorios y de personal calificado en el análisis y mantenimiento de los sistemas fotovoltaicos. El conocer las características de los elementos de un SFV es una prioridad si se quiere tener un criterio eficaz para la toma de decisiones.

Entre los diversos elementos de los SFV, las baterías y los paneles solares son fundamentales, ya que tienen un gran impacto en la eficiencia total

del sistema, por lo que se requiere un amplio estudio de su comportamiento.

II. MÉTODOS EXPERIMENTALES

II.1 CARACTERIZACIÓN DE BATERÍAS

La caracterización de las baterías se efectúa bajo los regímenes de operación de carga y descarga. En la literatura se han propuesto distintos métodos para la determinación del estado de carga (SOC) de la batería, entre los que se puede mencionar el conteo de cargas [1], la medida de la concentración de electrolito [2], y la medición de la impedancia electroquímica [3].

A. Conteo de cargas (coulomb/ampere-hora)

El método de conteo de cargas es común para la estimación del estado de carga de la batería (SOC), y requiere de conocer el estado de carga inicial (SOC₀). La Ec. (1) muestra la expresión para obtener el SOC por este método.

$$SOC = SOC_0 + \frac{1}{C_0} \int_{t_0}^t (I_{bat} - I_{pérdidas}) dt \quad (1)$$

donde SOC₀ es la capacidad inicial de la batería, I_{bat} es la corriente de descarga de la batería e I_{pérdidas} es la corriente de pérdidas debido a las reacciones electroquímicas.

B. Caracterización de descarga de baterías

Las caracterizaciones de descarga presentadas en este trabajo se efectuaron utilizando un equipo de medición marca Keithley 2420 y una interfaz de adquisición de datos desarrollada en LabView.

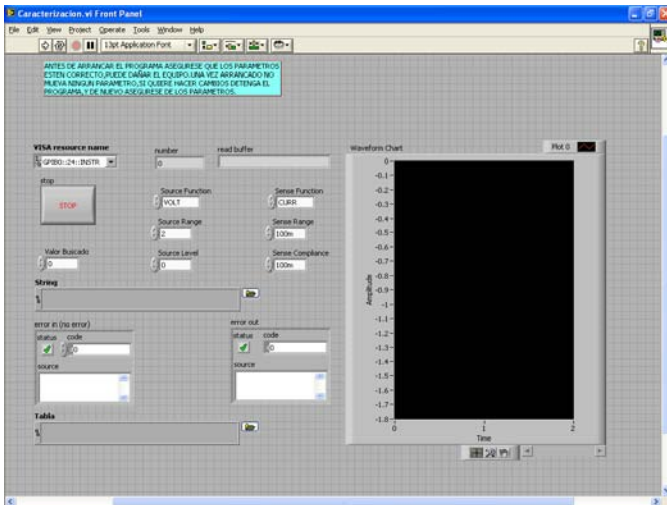


Fig. 1. Panel frontal de la interfaz en LabView desarrollada para la caracterización de baterías.

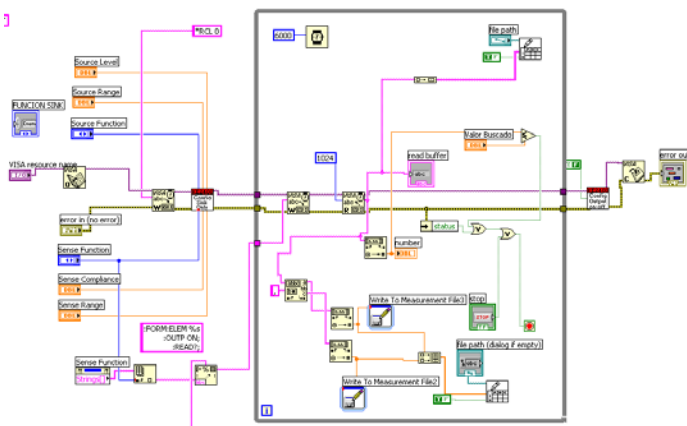


Fig. 2. Diagrama de bloques en LabView de la interfaz desarrollada para la caracterización de baterías.

Las Fig. 1 y 2 muestran, respectivamente, el panel de control y el diagrama de bloques del programa desarrollado.

En las pruebas realizadas a una batería nueva tipo plomo-ácido de 20 Ah, se observó una caída inicial del voltaje, seguida de una ligera recuperación (Fig. 3). A este fenómeno se le conoce como *coup-de-fouet* (latiguelo) [5]. En la figura se pueden distinguir dos regiones de la curva obtenida: el camellón de voltaje (trough voltage) y la meseta de voltaje (plateau voltage).

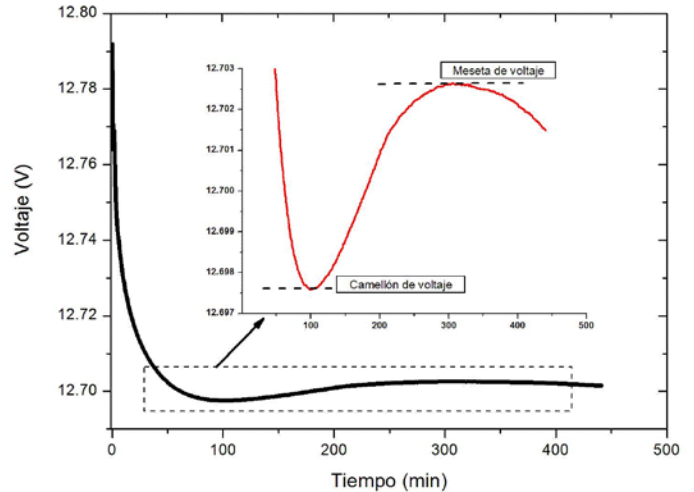


Fig. 3. Observación del fenómeno de latiguelo (coup de fouet) al inicio de la descarga de una batería con plena carga.

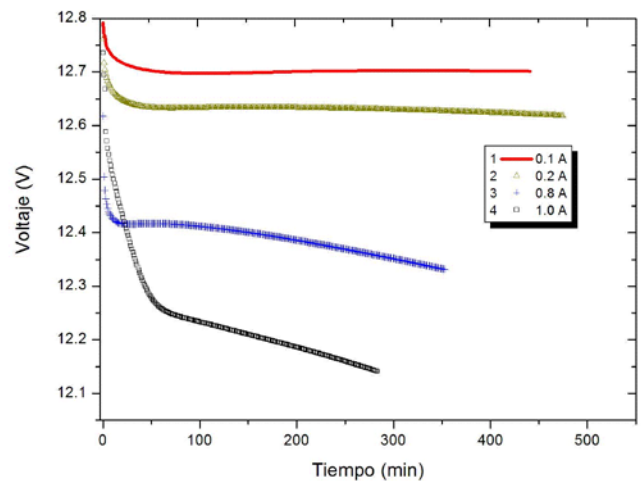


Fig. 4. Comparación de las curvas V-t a diferente corriente de descarga.

En la literatura, el fenómeno de latiguelo se interpreta como resultado del proceso químico que se genera al momento de descarga; el proceso de descarga convierte dióxido de plomo a sulfato de plomo, y esta reacción química se vuelve más eficiente cuando la molécula de sulfato de plomo es generada [5].

La profundidad del latiguelo es un claro indicador del estado de carga de la batería. En la Fig. 4 se comparan las curvas de caracterización V-t a diferentes corrientes de descarga de la batería, observándose que el fenómeno de latiguelo se va reduciendo conforme la batería disminuye su SOC.

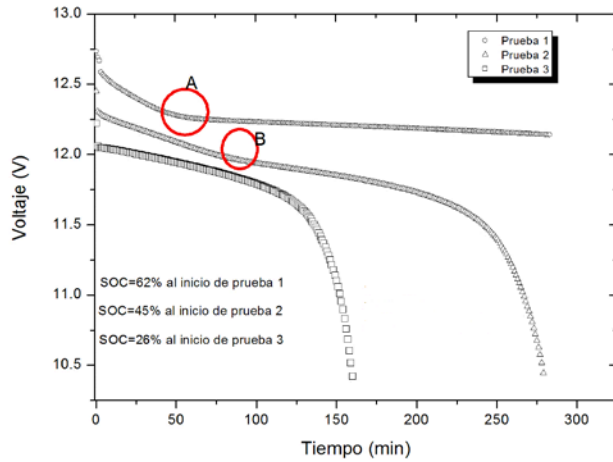


Fig. 5. Comportamiento de la zona de transición en función del SOC.

La Fig. 5 muestra las caracterizaciones del estado de carga realizadas a la batería, las cuales revelan que siempre está presente una zona de transición en todas las curvas obtenidas. Sin embargo, cuando el nivel de carga alcanza aproximadamente el 26 % del SOC inicial, esa zona de transición desaparece.

Este comportamiento indica que la batería se llevó a una zona de sobredescarga, lo cual siempre debe de evitarse para no reducir el tiempo de vida útil de las baterías.

C. Caracterización de carga de baterías

La caracterización de la carga de la batería es tan importante como la descarga, pues es volver a restituírle la energía extraída sin dañar a esta. En la Fig. 6 se muestra una comparación de las curvas de carga a corriente constante (1A) obtenidas en la batería a diferentes niveles de SOC.

En la Fig. 6 se puede observar que en la prueba No. 1, la batería alcanza un voltaje más alto en menor tiempo que el obtenido en la prueba No. 2. Con ello se puede deducir que la caracterización a la carga también es un indicador del SOC de una batería.

D. Ajuste de datos experimentales

El modelo de Copetti [6] es uno de los modelos más utilizados para la descripción de la evolución

del voltaje de la batería con respecto al tiempo de carga o descarga. En el presente trabajo se efectuó

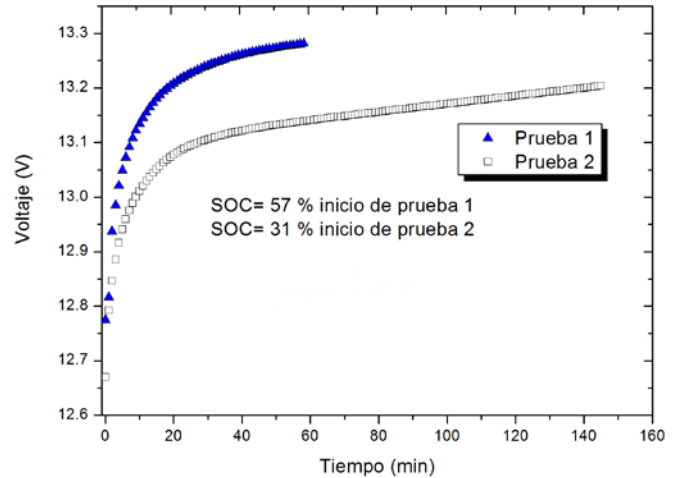


Fig. 6. Evolución del SOC para diferentes ciclos de descarga.

un ajuste de los datos obtenidos experimentalmente con respecto a este modelo. De acuerdo a Copetti, las ecuaciones para el voltaje de descarga $V_d(t)$ y el voltaje de carga $V_c(t)$ son calculadas en las Ec. (2) y (3), respectivamente.

$$V_d(t) = n_b [V_{bod} - K_{bod}(1 - SOC(t))] - n_b \frac{|I_d(t)|}{Cc} \left[\frac{k_{1d}}{1 + |I_d(t)|^{k_{2d}}} + \frac{k_{3d}}{SOC(t)^{k_{4d}}} + k_{5d} \right] (1 - \alpha_d \Delta T) \quad (2)$$

$$V_c(t) = n_b [V_{boc} + K_{boc}SOC(t)] + n_b \frac{|I_c(t)|}{Cc} \left[\frac{k_{1c}}{1 + |I_c(t)|^{k_{2d}}} + \frac{k_{3c}}{(1 - SOC(t))^{k_{4c}}} + k_{5c} \right] (1 - \alpha_c \Delta T) \quad (3)$$

donde n_b es el número de celdas en serie de la batería, V_{bod} y V_{boc} son los voltaje remanentes cuando la batería está descargada, en la descarga y en la carga, respectivamente, K_{bod} y K_{boc} son las relaciones del voltaje a circuito abierto y el estado de carga, en la descarga y en la carga, respectivamente, $I_d(t)$ es corriente de descarga, $I_c(t)$ es corriente de carga, K_{1d} , K_{2d} ,... y K_{1c} , K_{2c} ,... son los coeficientes de pérdidas asociados a la resistencia interna en la descarga y en la carga, respectivamente, α_d y α_c son los coeficientes relacionados con la temperatura en la descarga y en

la carga, respectivamente y ΔT es la diferencia de temperatura a 25°C.

Es necesario mencionar que el modelo no permite aplicarlo a las zonas de transición. Esto significa que el modelo no es aplicable ni en la zona de latiguo ni en las zonas de transición de la carga.

E. Ajuste de las curvas de descarga

En la Fig. 7 se observa la curva del modelo y de los datos experimentales, así como los parámetros extraídos del modelo de Copetti.

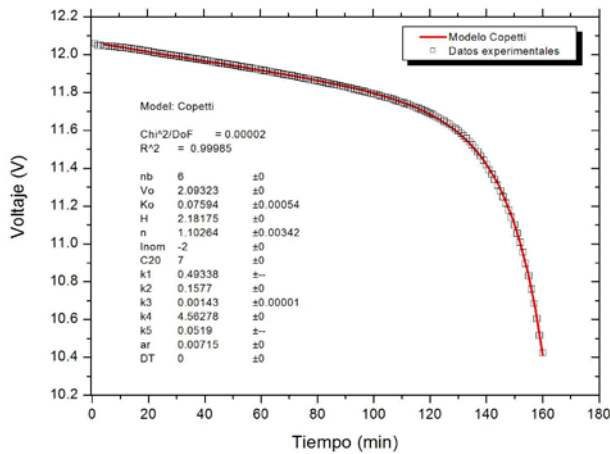


Fig. 7. Ajuste de datos experimentales en la descarga.

Como se puede observar en este caso, el modelo se ajusta a los datos experimentales con un muy buen índice de correlación (0.99). Los parámetros principales que se extrajeron de esta caracterización son el coeficiente n de Peukert y la capacidad de la batería (C_c). El coeficiente de Peukert obtenido del ajuste ($n=1.1$) cae dentro de los rangos típicos del tipo de una batería usada mientras que la capacidad de la batería (7 Ah) indica una degradación importante de la misma. Con estos datos se determina un valor para el SOH de 0.35 utilizando la ecuación 4 mientras que el SOC calculado es de 0.26 utilizando la ecuación 5. Estos datos representan las condiciones de la batería al inicio de la prueba.

$$SOH = \frac{\text{Capacidad actual}}{\text{Capacidad nueva}} \tag{4}$$

$$SOC = 1 - \frac{I H}{\left[\frac{H}{t} \right]^n} \tag{5}$$

F. Ajuste de las curvas de carga

En la Fig. 8 se muestra la comparación del modelo y de los datos experimentales para la carga.

A partir de estos ajustes se puede extraer un parámetro de suma importancia que es el coeficiente de Peukert (n), el cual está relacionado a la calidad de la batería [7].

II.2 Caracterización de paneles solares

A. Curvas I-V (paneles 85Wp y 10Wp)

La caracterización de los paneles fotovoltaicos es necesaria para determinar rangos de trabajo y eficiencias. Las caracterizaciones realizadas en el presente trabajo se realizaron utilizando el equipo de medición Keithley 2420, un piránometro marca Kipp and Zonen modelo CMP-3 y una interfaz en LabView desarrollada por el fabricante del equipo de medición Keithley (Fig. 9).

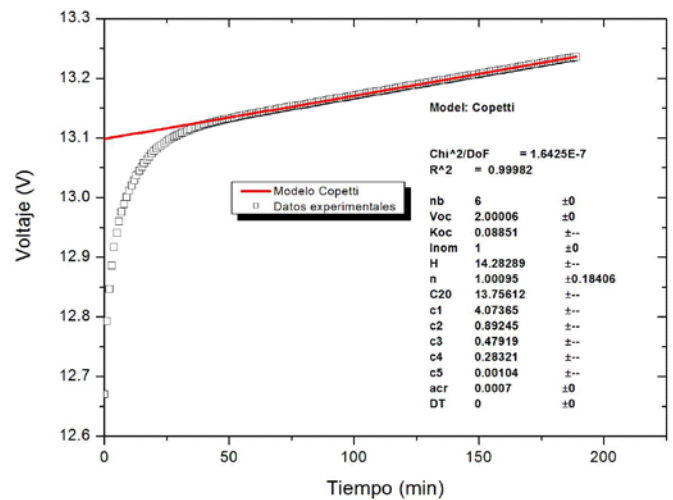


Fig. 8. Ajuste de datos experimentales en la carga.

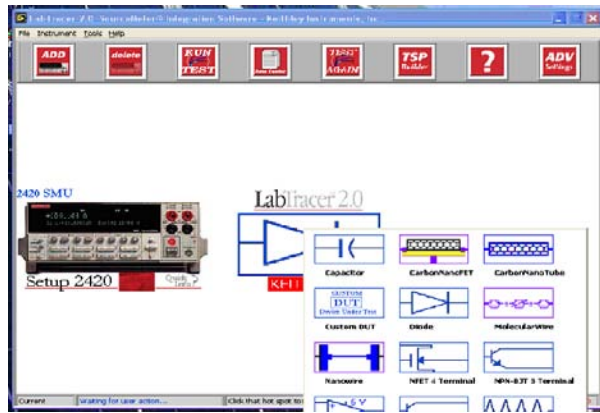


Fig. 9. Diagrama de bloques de la interfaz en LabView de Keithley.

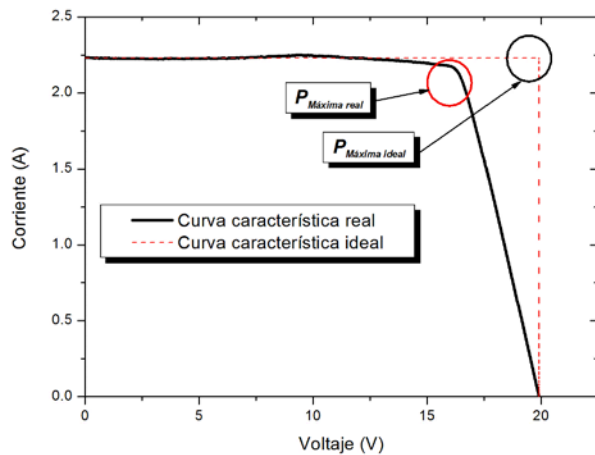


Fig. 10. Curva característica I-V del panel 85 Wp

Los paneles que se caracterizaron fueron: a) un panel tipo monocristalino de 85 Wp y un panel tipo policristalino de 15 W. En las Fig. 10 y 11 se muestran las curvas características de corriente y voltaje obtenidas de los paneles de 85 Wp y 10 Wp, respectivamente.

En la comparación de las curvas características I-V de las figuras 10 y 11, se puede observar la influencia de la resistencia serie en esta última, debido al pronunciamiento del codo. Por otra parte, no se observa un efecto importante de la resistencia paralelo. Adicionalmente, de las curvas se puede observar los puntos de máxima potencia ideal y real de cada uno de los paneles caracterizados.

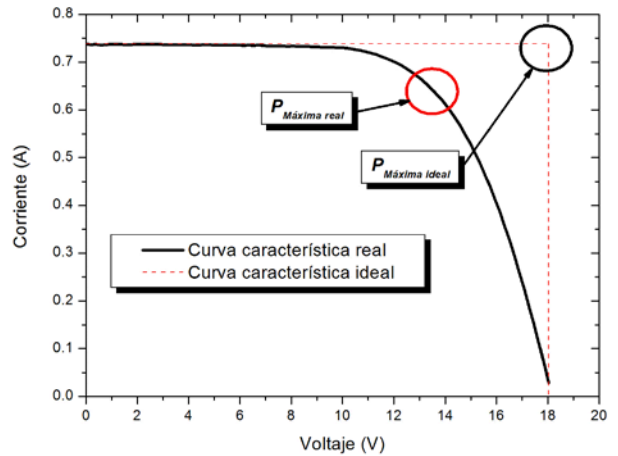


Fig. 11. Curva característica I-V del panel 10 Wp.

B. Cálculo de factor de forma

Para la determinación del factor de forma de una celda o panel fotovoltaico se utiliza la Ec. (6).

$$FF = \frac{I_{mp} V_{mp}}{I_{sc} V_{oc}} \quad (6)$$

donde FF es el factor de forma (o de llenado) [8], el cual es un parámetro con el que se puede extraer información sobre la degradación del panel y la disminución de la eficiencia debido al deterioro del material base.

En la Fig. 11 se puede observar que la curva característica I-V tiene una tendencia de forma hiperbólica, lo que indica que el punto de máxima potencia real fue desplazado. También es un indicador del incremento de la resistencia serie y paralelo intrínsecas del panel.

La tabla I muestra el factor de forma calculado a partir de los datos experimentales extraídos de la caracterización de los paneles.

Tabla I. Factor de forma de los paneles de 85 Wp y 10 Wp.

Panel monocristalino 85 Wp	FF=0.786
Panel policristalino 10 Wp	FF=0.65

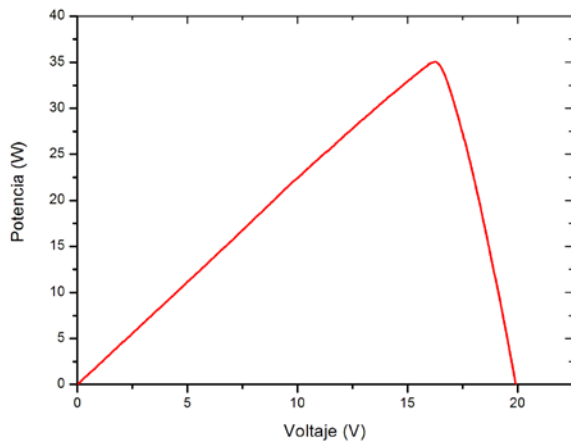


Fig. 12. Curva característica P-V del panel 85 Wp.

C. Cálculo de eficiencias

Para la determinación de la eficiencia de una celda solar o del panel fotovoltaico se utiliza la Ec. (7) o la Ec. (8). La tabla II muestra las eficiencias calculadas a partir de los datos experimentales extraídos de la caracterización de los paneles.

$$\eta = \frac{I_{mp} V_{mp}}{P_{in}} \tag{7}$$

$$\eta = \frac{I_{sc} V_{oc} FF}{A E} \tag{8}$$

Tabla II. Eficiencias calculadas de los paneles de 85 Wp y 10 Wp.

Panel monocristalino 85 Wp	$\eta = 12.5 \%$
Panel policristalino 10 Wp	$\eta = 8.4 \%$

D. Curvas P-V (panel 85Wp y 10Wp)

En las Figs. 12 y 13 se muestran las curvas características de potencia de los paneles de 85 Wp y 10 Wp.

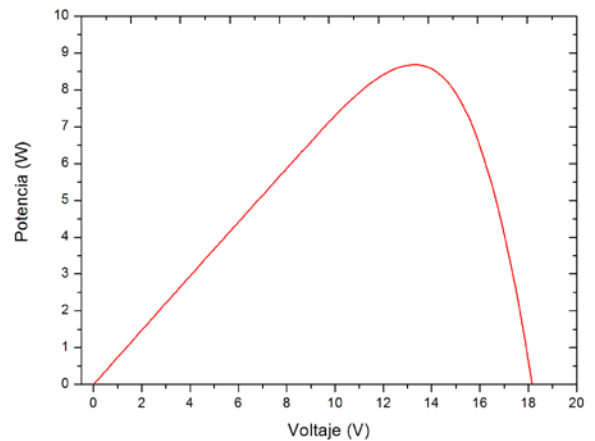


Fig. 13 Curva característica P-V del panel 10 Wp.

En las figuras 12 y 13 de las curvas características de la potencia se puede observar que el codo pronunciado que se observa en la figura 12 para el caso del panel 10 Wp (figura 13) este ha desaparecido mostrando así la degradación del panel.

III. CONCLUSIONES

El presente trabajo demuestra que es posible utilizar el modelo de Copetti para describir el comportamiento de carga y descarga de las baterías en un SFV y así extraer el coeficiente de Peukert.

La caracterización de paneles fotovoltaicos permitió extraer el factor de forma y la eficiencia para así confirmar los datos reportados por los fabricantes. Adicionalmente, las curvas de caracterización permiten observar claramente la degradación de un panel FV.

IV. REFERENCIAS

[1] P. Singh, “*Development of Fuzzy Logic-Based Lead Acid Battery Management Techniques with Applications to 42V Systems*”, ECE Department, Villanova University Villanova, PA 19085.

[2] J. Heitner, *Mecánica Automotriz, principios y prácticas*, Edit. Diana, 1986.

- [3] J. H. Jang, and J. Y. Yoo, “*Impedance-Based and Circuit-Parameter-Based Battery Models For Hev Power Systems*”, International Journal of Automotive Technology, Vol. 9, No. 5, pp. 615–623 (2008)
- [4] P. H. Pascoe *et al*, “*Coup de fouet based VRLA battery capacity estimation*”, Proceed. IEEE Int. Workshop on Electron. Design, Test and App. (DELTA '02), p. XXX, Vol XX, 2002
- [5] Electropaedia, “*Electropaedia Battery Technology Glossary*”, [en línea], <http://www.mpoweruk.com/glossary.htm>
- [6] A. O. Mohamed Yahya, A. Ould Mahmoud, and I. Youm, “*Modélisation d’un système de stockage intégré dans un système hybride (PV / Eolien / Diesel)*”, Revue des Energies Renouvelables Vol. 10 N_2 (2007), pp. 205-214.
- [7] Ch. Gibson, Peukert in Brief, [en línea] <http://www.smartgauge.co.uk/peukert3.html>
- [8] M. J. Morgan, G. Jakovidis, and I. McLeod, “*An experiment to measure the I-V characteristics of a silicon solar cell*”, Department of Physics, Monash University, Clayton, Victoria 3168, Australia.

V. AUTORES

Ing. Luis Manuel Segura Flores. Estudiante de la Maestría en Ingeniería de la Unidad Académica de Ingeniería Eléctrica de la UAZ.

Dr. Jorge de la Torre y Ramos. Profesor investigador de la Unidad Académica de Ingeniería Eléctrica de la UAZ.

Dr. Manuel Reta Hernández. Profesor investigador de la Unidad Académica de Ingeniería Eléctrica de la UAZ.

DISEÑO DE 3 DIFERENTES TIPOS DE AEROGENERADORES PARA USO URBANO Y DOMESTICO

Área de Energías Alternativas
 Facultad de Ciencias de la Electrónica de la
 Benemérita Universidad Autónoma de Puebla
 18 sur y Av. Sn. Claudio Col. Sn Manuel Puebla, Pue
 Tel: 2295500 ext. 7400

RESUMEN

En este artículo, se propone una metodología que busca introducir al estudiante interesado en el área de Mecatrónica a diseñar los elementos mecánicos y electrónicos de tres diferentes tipos de aerogeneradores tanto horizontales como verticales para la conversión de energía del viento en energía eléctrica.

El objetivo del proyecto es:

- a).- Diseñar un aerogenerador horizontal de 600 watts de potencia para uso domestico.
- b).- Diseñar un aerogenerador vertical tipo Savonius de 500 watts de potencia para uso domestico y urbano.
- c).- Diseñar un aerogenerador vertical tipo Darrieus de 500 watts de potencia para uso domestico y urbano.

Estos aerogeneradores cubrirán las necesidades energéticas de una vivienda de las zonas urbanas, así como a pequeños consumos en zonas aisladas de la red eléctrica en zonas rurales de la población, y en las carreteras con el propósito de, entre otros, iluminar sectores peligrosos, energizar sistemas de comunicación de emergencia y monitoreo, proporcionar energía a señalética y publicidad vial.

Objetivos de diseño del aerogenerador

- Maximizar la eficiencia en la captura y transformación de la energía cinética del aire.
- Minimizar el tiempo y el costo de las operaciones de mantenimiento.
- Desarrollar, diseñar y fabricar siguiendo una normativa, lo que garantiza un producto que cumple los estándares de calidad requeridos por la industria.

I. INTRODUCCIÓN

Los aerogeneradores son las maquinas que se encargan de capturar la energía del viento y transformarla en energía eléctrica, funcionan de manera opuesta a un ventilador.

Un aerogenerador debe transformar la energía (cinética) que lleva el viento en movimiento de un eje (energía mecánica). La energía cinética cumple la siguiente relación:

$$E = \frac{1}{2} \rho * v^2 \quad (1)$$

Donde E = energía cinética

ρ = densidad del aire

v = velocidad del viento

Un aerogenerador ralentiza el viento al pasar por el rotor hasta 2/3 de su velocidad inicial, lo que significa que no aprovecha toda la energía cinética que el viento aporta al rotor, por lo que la Ley de Betz dice:

“Una turbina eólica puede convertir en energía mecánica como máximo un 59,26 % de la energía cinética del viento que incide sobre ella.”

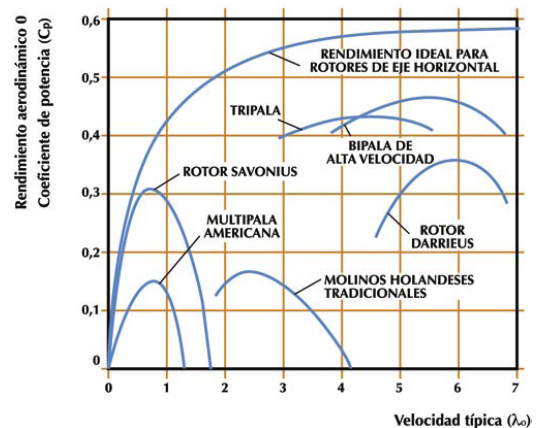


Fig.1 Eficiencia de los diferentes aerogeneradores

A modo de referencia se utiliza la siguiente nomenclatura en función del tamaño del aerogenerador:

Según la **potencia eléctrica que pueden generar**:

- Pequeños aerogeneradores cuya potencia está por debajo de los 30 kW
- Aerogeneradores de tamaño medio. Es el sector más desarrollado y extendido, en donde se engloban la mayor parte de las instalaciones conectadas a la red. Su potencia está entre 30 y 600 kW.
- Máquinas multimegawatio. Son máquinas con potencia por encima del megawatio, representan lo último en tecnología y los principales fabricantes de aerogeneradores disponen en el mercado de máquinas de este tipo.

Por el **tipo de eje**.

Se definen en general, los aerogeneradores según la posición de su eje de rotación, con relación a la dirección del viento.

Aerogeneradores de eje horizontal

Los aerogeneradores de eje horizontal con el eje paralelo a la dirección del viento, en la actualidad son las máquinas más difundidas y con mayores rendimientos que las otras existentes, algo muy importante en el momento de comenzar un diseño. En este grupo se incluyen aquellas que tienen 1,2,3 o 4 palas, además de las típicas multipalas para el bombeo de agua.

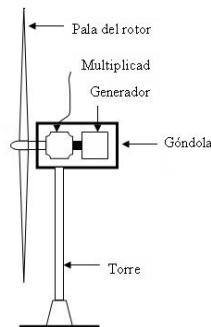


Fig.2 Aerogeneradores de eje horizontal.

Este tipo de aerogeneradores tienen unas ventajas y unos inconvenientes:

Ventajas: Este sistema es el más estudiado y el que mejores rendimientos proporciona; se ha llegado a un coeficiente de aprovechamiento (C_p) de 0,45 a 0,50 de la potencia mecánica, aunque al transformarla en eléctrica, el rendimiento

desciende. Con una elección adecuada del emplazamiento dan buen rendimiento medio, independientemente de la estación del año y de la hora del día, y emplean menos superficie de terreno que otros sistemas para la misma potencia producida.

Inconvenientes: La energía producida en cada momento tiene variaciones instantáneas muy elevadas y difíciles de calcular. Para su emplazamiento deben elegirse zonas sin obstáculos y elevadas. La necesidad de un *sistema de orientación* para adaptarse a la dirección del viento complica el sistema y lo hace más propenso a fallos. Dicho sistema suele constar de una veleta que detecta la dirección del viento.

Aerogeneradores de eje vertical.

En este caso el eje de giro está situado verticalmente y las palas giran alrededor de él en un plano horizontal. El viento, por tanto, hace fuerza por un lado sobre la parte delantera de las palas, dando velocidad, pero a la vez presiona sobre la parte trasera de otra pala, frenándola. El diseño crítico consiste en diseñar la estructura de la pala de modo que por delante sea capaz de captar el mayor viento posible y que simultáneamente ofrezca la menor resistencia al viento por su parte posterior. No necesitan ningún sistema de orientación, lo que constituye una ventaja constructiva. En funcionamiento, las palas, los rodamientos y los ejes no están sometidos a esfuerzos importantes por cambios de orientación, además son de fácil construcción. El rendimiento es mediocre (el rotor Savonius un 20% del límite de Betz). No se experimentó un gran desarrollo en estos aparatos.

Existen diferentes tipos de aerogeneradores de eje vertical los dos principales son;

Savonius: Dos o más filas de semicilindros colocados opuestamente.

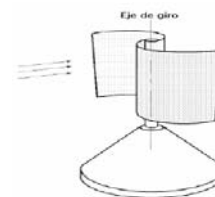


Fig. 3 Aerogenerador Savonius

Esencialmente utilizan el arrastre diferencial creado por las palas que pueden ser de diversas formas. El par de arrastre es elevado, pero la velocidad máxima es claramente inferior a la de los rotores de eje horizontal.

Darrieus: Consisten en dos o tres arcos que giran alrededor del eje

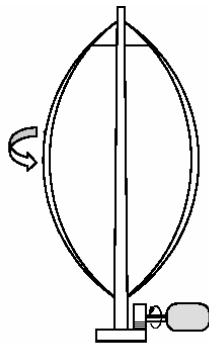


Fig.4 Aerogenerador Darrieus.

Emplea la sustentación de las palas y están caracterizados por débil par de arranque y velocidad de rotación elevada que permite la recuperación de una gran potencia.

Este tipo de aerogeneradores tiene las siguientes ventajas e inconvenientes:

Ventajas La disposición del rotor le permite captar el viento proveniente de cualquier dirección sin necesidad de *dispositivo de orientación*. Al estar dispuesto verticalmente el eje, la maquinaria de aprovechamiento de la energía puede situarse directamente sobre el terreno y no es necesario el empleo de una torre como en los sistemas de eje horizontal. Por tanto se produce un ahorro en costes de construcción y se facilita el aprovechamiento de la energía.

Inconvenientes: Estos sistemas producen menos rendimiento que un rotor de eje horizontal de dimensiones equivalentes. La *baja velocidad de giro* dificulta su empleo en instalaciones de obtención de energía eléctrica, en las que se necesitan velocidades muy elevadas.

Un punto que vale hacer notar de los dos tipos de aerogeneradores, es la autonomía frente a la fuente más cercana accesible, que en este caso es la Comisión Federal de Electricidad (CFE). Esta última no siempre se presenta en los pueblos

alejados, por el costo que supone instalar una red hacia aquellos.

II. MÉTODO

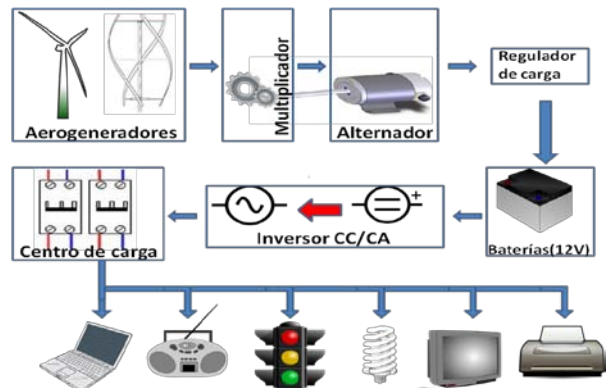


Fig.5 Partes de una instalación Eólica

La fig.5 muestra un pequeño diagrama a bloques de las partes y componentes en las cuales se divide y forman una instalación eólica de pequeña potencia.

El primer paso para construir un aerogenerador fue un estudio **anemométrico** usualmente se realiza en las primeras fases de diseño para evaluar factores:

De factibilidad: determinar si es posible técnica y económicamente la instalación de un aerogenerador en una zona o región geográfica.

De localización: elegir el lugar óptimo de emplazamiento

Técnicos: obtener parámetros necesarios para el diseño tanto del aerogenerador como de la estructura de soporte.

Debido a que el lugar de instalación fue elegido en base a consideraciones especiales, el estudio se realizó en este caso para:

- Aprender a realizar un estudio anemométrico
- Conocer el potencial eólico del lugar de emplazamiento teniendo como resultados en un periodo de medio año velocidades de 2.5 m/s hasta la mayor medición es de 12 m/s idóneas para los prototipos

Para describir las distribuciones de la velocidad del viento que se obtienen a través de la estación meteorológica, se utilizan distintas funciones de densidad de probabilidad; entre ellas la Weibull, Rayleigh (caso particular de la anterior), la Beta y la Log Normal. La mayoría de los autores se

inclina por la de Weibull de dos parámetros y es la que se utilizó.

Parámetros Hallados de Distribución de Weibull

Es representada por la siguiente expresión:

$$f(v) = (k/C) \cdot (v/C)^{(k-1)} \cdot \exp(-(v/C)^k) \quad (2)$$

Donde:

$f(v)$ = función de Weibull

k = Parámetro de Forma

C = Parámetro de Escala

$K_{weibull}$ = 2.92 constante

$C_{weibull}$ = 7.3 m/s

Velocidad Promedio = 4.0m/s

La secuencia de pasos necesaria para diseñar y dimensionar un aerogenerador es:

- Definir la necesidad energética a satisfacer.
- Magnitud de la instalación (micro, pequeña, mediana o grande):
- Potencia máxima a instalar
- Energía anual media a entregar
- Realizar un estudio anemométrico.
- Elegir un generador eléctrico acorde y realizar los ensayos correspondientes.
- Diseñar la hélice para entregar al generador la potencia mecánica y cupla que éste requiere.
- Calcular las solicitaciones mecánicas que debe soportar la estructura y dimensionarla en consecuencia.
- Elegir y diseñar un sistema de control apropiado

a).-El prototipo del **aerogenerador horizontal**

Para el cálculo y diseño de las dimensiones del rotor se han empleado las ecuaciones aerodinámicas:

Cálculo del diámetro del rotor

Se obtiene a partir de la ecuación de potencia de la turbina eólica:

$$d = \sqrt{\left(\frac{8 \cdot P}{\pi \cdot \rho \cdot V^3 \cdot C_p \cdot \eta_e \cdot \eta_t} \right)} \quad (3)$$

Donde:

d : Diámetro del rotor de la turbina eólica (m)

P : Potencia de diseño del aerogenerador (W)

ρ : Densidad del aire, variable con la a.s.n.m (1.23 kg/m³ a nivel del mar)

V : Velocidad del viento (m/s)

A : Área barrida por la turbina (m²)

C_p : Coeficiente de potencia (adimensional)

η_e : Eficiencia del generador

η_t : Coeficiente de transmisión

La **velocidad de giro** de la turbina eólica se obtiene de la siguiente expresión:

$$N = \left(\frac{60 \cdot \lambda \cdot V_D}{\pi \cdot d} \right) \quad (4)$$

Donde:

λ : Celeridad

N : Velocidad de giro de la turbina eólica (RPM)

V_D : Velocidad de diseño (m/s)

Cálculo de la sección del álabe y selección del perfil

Para el cálculo de la sección del álabe se utilizaron las siguientes ecuaciones:

$$\left. \begin{aligned} \lambda_r &= \frac{\lambda}{R} \\ \phi &= \frac{2}{3} \cdot \arctan \left(\frac{1}{\lambda_r} \right) \\ C &= \frac{8 \cdot \pi \cdot r (1 - \cos \phi)}{z \cdot C_1} \\ \beta &= \phi - \alpha \\ N &= \frac{w \cdot c}{\mu} \quad (N \text{ de Reynolds}) \end{aligned} \right\} \quad (5)$$

Donde:

λ_r : Celeridad local para el radio r

λ : Celeridad de diseño

r : Distancia del centro del rotor a la sección evaluada (m)

R : Radio de la turbina (m)

β : Angulo formado por la velocidad relativa con el plano de giro del rotor

C : Cuerda de la sección del álabe

z : Número de álabes o palas

C_1 : Coeficiente de sustentación del álabe

ϕ : Angulo formado por el álabe con el plano de giro

α : Angulo de ataque, tomado del perfil seleccionado

w : Velocidad relativa al perfil m/s

μ : Viscosidad cinemática del aire

En resumen, los parámetros de diseño considerados son los siguientes:

Potencia eléctrica:	600 W
Velocidad de funcionamiento	3 m/s
Velocidad de nominal de diseño	8 m/s
Angulo de incidencia del alabe	15 grados
Numero de palas:	3
Diámetro del rotor	3 metros
Rendimiento de la hélice: $C_p = 0,38 \pm 0,10$ a 8 m/s	
Rendimiento del generador eléctrico: $0,50 \pm 0,02$ a 800 rpm	



Fig.6 Vista frontal de la construcción de las palas



Fig.7 Aerogenerador horizontal

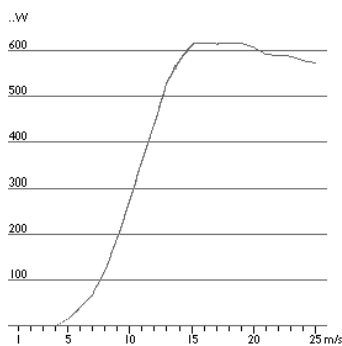


Fig. 8 La curva de potencia de un aerogenerador es un gráfico que nos indica la potencia eléctrica que es capaz de generar para cada velocidad del viento.

b).-Para desarrollar el **aerogenerador vertical tipo Savonius**, se tom en cuenta un diferente diseño de aspa en base a la involuta del círculo.

La **involuta** (o envolvente) es una curva que se puede generar al desenrollar un cordel tirante desde un cilindro de enrollado, como se ve en la figura 9. Las periferias de los cilindros desde los cuales se desenrollan los cordeles, se denominan **círculo base**.

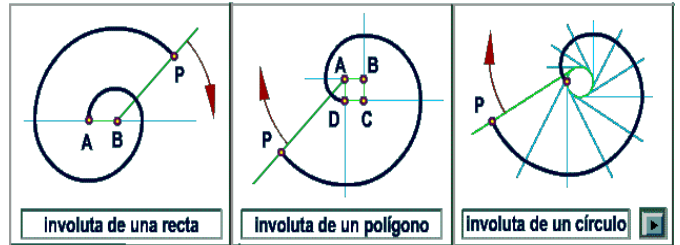


Fig.9 Involuta o Envolvente

El prototipo consiste en trazar una involuta a partir de círculo base de 5 cm de diámetro, se traza un número de tangentes a la curva C dada. Se construye en cada punto P de la curva C un círculo que sea tangente a la curva C en P, luego se obtiene un punto de la curva Ce considerando el centro de este círculo. Se trazaron 3 involutas distribuidas proporcionalmente, cuya longitud fue de 54 cm , teniendo con esto la figura base.

Una vez teniendo la figura base, se hicieron 2 estructuras de la misma, en solera de 1/2 pulgada de ancho y 1/16 de pulgada de grosor.

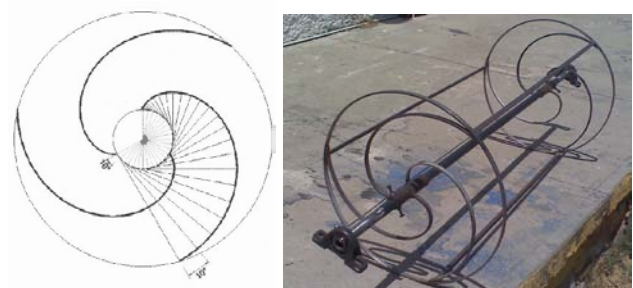


Fig. 10 Figura y estructura base

Teniendo las estructuras se unieron en forma cilíndrica a una distancia de 1 metro de altura, usando como eje central un tubo de pulgada y media , y en cada extremo una chumacera de piso de 1 1/2 de diámetro.

En la estructura se cortaron 3 rectangulos de lamina de 1/16 de pulgada de grosor, de 1metro de alto por 50 cm de ancho que tendran la función de velas antes mencionadas y se acoplan por medio de remaches a nuestra exstructura base.

Como exstructura soporte es un rectangulo de 2 metros de alto por 1.50 de ancho, con patas rigidas para asegurar su funcionamiento. El modo de funcionamiento se asemeja al de un remolino, el aire entra en el extremo de nuestro aerogenerador produciendo una diferencia de presiones en las velas del aerogenerador, y en el centro como es hueco se aremolina el viento dando a este una velocidad mayor de la que proporciona el aire en el extremo. Los equipos de conversión y control están en la base del aerogenerador. En contrapartida, como la velocidad de viento es menor al nivel del suelo, el rendimiento es bajo.

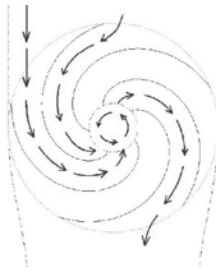


Fig.11 Captura del viento

El nuevo diseño de turbina de viento puede impulsar el uso de esta energía en el área urbana, en las azoteas de casas y edificios o en los camellones de las avenidas o autopistas, aprovechando la energía de las turbulencias que producen los automóviles a su paso a grandes velocidades.



Fig. 12 Prototipo del aerogenerador

En resumen, los parámetros de diseño considerados son los siguientes:

Potencia eléctrica:	500 W
Velocidad de funcionamiento	5.5 m/s
Velocidad de nominal de diseño:	25 m/s
Numero de alabes:	3
Diámetro del rotor	.5 metros
Rendimiento de la hélice: $C_p = 0,19 \pm 0,10$ a 10 m/s	
Rendimiento del generador eléctrico: $0,25 \pm 0,02$ a 800 rpm	

c).-El prototipo del aerogenerador vertical tipo Darrius consiste en tres arcos de pvc de 1.20 metros que giran alrededor de un eje (fig.13), se caracteriza por las palas o álabes están arqueadas con una forma parecida a la de una batidora de huevos.



Fig. 13 Aerogenerador vertical

Los álabes son biconvexos con un ángulo de separación de 60 grados entre la parte superior e inferior de donde se encuentran sujetas a estrellas de 3 picos separadas 120 grados (fig.13), y la superficie descrita tiene forma esférica.

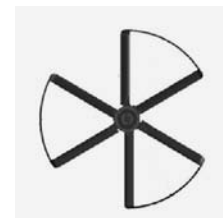


Fig.13 Posicion de las aspas

El giro del rotor está provocado por la acción aerodinámica del viento sobre los álabes, que origina fuerzas aerodinámicas que dan lugar al par de rotación. El par de arranque de un rotor Darrius es muy pequeño.



Fig.15. Prototipo del aerogenerador

En resumen, los parámetros de diseño considerados son los siguientes:

Potencia eléctrica:	500 W
Velocidad de funcionamiento	5 m/s
Velocidad de nominal de diseño:	25 m/s
Angulo de incidencia del alabe	15 grados
Numero de palas:	3
Diámetro del rotor	3 metros
Rendimiento de la hélice: $C_p = 0,18 \pm 0,10$ a 10 m/s	
Rendimiento del generador eléctrico: $0,3 \pm 0,02$ a 800 rpm	

III. CONCLUSIONES

a).-Aerogenerador Horizontal

El proyecto está en la fase intermedia de su desarrollo, donde durante los meses de Enero a Mayo del 2009 se cumplieron con el 50% de las etapas planteadas. El aerogenerador se encuentra actualmente instalado y funcionando en CU junto a LUPOTEC.

El proceso de diseño ha sido dividido en etapas. Se puede resaltar que el aerogenerador hace uso de un sistema multiplicador entre la hélice y el generador eléctrico. Este acople directo representa una ventaja en términos de eficiencia.

Se están utilizando alternadores de carro modificando su devanado con imanes de neodimio lo cual permiten una mayor eficiencia en la producción, lo que en las siguientes etapas se propone construir el propio generador que sería lo ideal para cada tipo de aerogenerador.

El desempeño del aerogenerador es muy bueno en términos generales. Su funcionamiento es suave, es decir, los giros vertical y del sistema de control no producen vibraciones o movimientos bruscos. Es

silencioso y la torre presenta sólo oscilaciones de pequeña amplitud con velocidades de viento de 10m/seg.

El funcionamiento mecánico aparenta ser adecuado, aunque son necesarios más ensayos para determinar la velocidad de viento a la que comienza a actuar y cómo es la generación de potencia para velocidades mayores.

Este proyecto, debido a su naturaleza multidisciplinaria, permite: la puesta en práctica de los conocimientos de diferentes áreas del conocimiento como son Física, Matemáticas, Mecánica, Mecatrónica, Computación y Electrónica.

Proyección del trabajo

Algunos aspectos que podrían desarrollarse son:

- Estudiar la factibilidad de realizar un *scale-up* del proyecto.
- Diseño de aspas de materiales más apropiados para aplicaciones de mayor potencia.
- Fabricar un generador eléctrico más eficiente, atendiendo específicamente a las necesidades del proyecto.
- Optimización del sistema de adquisición de datos, para permitir la lectura y procesamiento de las variables en tiempo real, abriendo la posibilidad a la implementación de un sistema de control mecánico / electrónico.
- Estudios más detallados de la performance del aerogenerador y el sistema de control.

b).-Aerogeneradores Verticales

Se ha presentado el trabajo sobre diseño de los 2 tipos principales de aerogeneradores verticales obteniendo estimulantes resultados en ambos como: un buen coeficiente de potencia, una curva de potencia ideal, idóneo para altas velocidades, arranque con baja velocidad, se propone continuar en su investigación y desarrollo pues se cree que revolucionará el campo energético y en concreto el de la energía eólica, de tal forma que cualquier inversión en este proyecto presentará una opción segura y de alta rentabilidad a corto o medio plazo.

El análisis costo/beneficio de aplicación de estos sistemas energéticos en zonas rurales y mejoramiento de calidad de vida de los habitantes, es en el fondo la meta principal por la cual se

desarrollan este tipo de sistemas. Por el tipo de diseño hace posible pensar y analizar en la factibilidad de aplicación de sistemas híbridos energéticos a procesos productivos artesanales, principalmente de un sistema solar-eólico, que es una buena respuesta a la demanda considerando que para ambos medios, nuestra región presenta facilidades naturales y geográficas para su instalación.

Innovación tecnológica.

En la actualidad el proyecto se encuentra en fase de desarrollo y tras las primeras pruebas de laboratorio y de campo se ha observado que este modelo presenta una serie de novedades tecnológicas respecto a otros aerogeneradores:

- Aprovecha la fuerza directa del viento para provocar el giro del rotor por lo que se obtiene un gran par motor.
- Coeficiente de potencia muy superior a los aerogeneradores existentes.
- Curva de potencia con crecimiento exponencial.
- Funciona a velocidades del viento superiores a los 55 m/s.
- Puede funcionar a velocidades del viento por debajo de los 5 m/s.

También presenta otras ventajas técnicas:

- Posee un diseño simplificado. Bajo impacto medioambiental: fácilmente camufladle en el entorno. Fácil montaje y mantenimiento.
- Mayor seguridad tanto para trabajadores (al no trabajar con grandes componentes ni a gran altura), como para la fauna (al poder instalar una malla de protección en el perímetro del aerogenerador).
- Las principales dificultades se presentan en la eficiencia energética de los distintos elementos constituyentes implementados, tales como la caja multiplicadora, el alternador y la solución estructural exigida por un desbalance dinámico.
- En respuesta al problema, se sugiere el diseño de un generador de imanes permanentes de múltiples polos que tenga una velocidad de operación nominal acorde con la velocidad rotacional del aerogenerador. Esto permitirá

redimensionar la caja amplificadora e inclusive, pensar en eliminarla del diseño.

Sin embargo, a pequeña y mediana escala, los aerogeneradores de eje vertical vuelven a ser competitivos respecto a los de eje horizontal si se consideran los siguientes factores:

- estructura comparativamente simple,
- bajo costo de implementación,
- bajo costo de instalación,
- fácil mantenimiento.

IV. REFERENCIAS

- [1] J. M. Escudero López *Manual de energía eólica*, Ediciones Mundi-Prensa, 2004
- [2] M. Villarubia, *Energía eólica*, CEAC, 2004
- [3] J.L. Rodríguez Amenedo, J.C. Burgos Díaz, S. Arnalte Gómez, *Sistemas eólicos de producción de energía eléctrica*, RUEDA S.L., 2003
- [4] Guillen S. O. *Energías Renovables. Una perspectiva ingenieril*, 2004.
- [5] Gipe P. *Energía Eólica Práctica*, 2000
- [6] G. M; Colmena S. A; Sánchez N. C. *Monografías técnicas de energías renovables*. Energía Eólica, 2001.
- [7] Instituto de investigaciones eléctricas. "Energías renovables"
<http://genc.ije.org.mx/genc/index2.html>
- [8] Mott L. Robert *Diseño de elementos de máquinas*, 2000.

V. AUTORES

M. en C. José M. Carlos Pérez Aguirre

Profesor Investigador de la Facultad Ciencias de la Electrónica

Área o centro: Sistemas Digitales

Dirección: 14 sur y Sn. Claudio Ciudad Universitaria

Tel: 2295500 ext. 7400 FAX:

Correo electrónico: capeagui@yahoo.com.mx

Formación académica:

Licenciatura: BUAP - Facultad de Físico Matemáticas Lic. En Electrónica 1989

Especialidad: Sistemas Digitales y control digital

Maestría: Universidad de la Américas Puebla
Maestría en Ciencias con especialidad en
Electrónica 2001

M. en C. Aurelio Jacinto Nolasco

Profesor Inv. Asociado B TC
Unidad Académica de adscripción: Facultad de
Ciencias de la Electrónica
Área o centro: Sistemas Digitales
Dirección: 14 Sur y Sn. Claudio Ciudad
Universitaria
Tel: 2295500 ext. 7400 FAX:
Correo electrónico: ajnolasco@yahoo.com.mx
Formación académica:
Licenciatura: BUAP, Facultad de Ciencias Físico
Matemáticas 1988.
Especialidad: Inteligencia artificial y
reconocimiento de patrones
Maestría: BUAP Maestría en Ciencias con
especialidad en Computación 1997.

M. en C. A. Eloy Cruz Machicado

Profesor inv. TC. Titular A
Unidad Académica de adscripción: Facultad de
Ciencias de la Electrónica
Área o centro: Sistemas Digitales
Dirección: 14 Sur y Sn. Claudio Ciudad
Universitaria

Tel: 2295500 ext. 7400 FAX:
Correo electrónico: acruzma@hotmail.com
Formación académica:
Licenciatura: BUAP, Facultad de Ciencias Físico
Matemáticas 1987.
Especialidad: Sistemas de control y digitales
Maestría: INAOE Maestría en Ciencias con
especialidad en Electrónica 1997.

Colaboradores:
Estudiante tesista de Ingeniería Mecánica y
Eléctrica

Vega Fuentes Jesús

Estudiantes de 6to y 7mo semestre de la Lic. en
Mecatrónica:

Castellanos García Arturo
López Flores Edwin
Rojas Anzaldo Jonathan
Sánchez Carmona Rodolfo

IMÁGENES



Fig. 16 Estación Meteorológica, monitor receptor de datos y programa de computadora.



Fig. 17 Vistas del Aerogenerador Horizontal y pequeño parque eólico.



Fig.18 Aerogenerador Vertical tipo Savonius



Fig. 19 Aerogenerador Vertical tipo Darrius.

ANÁLISIS TÉCNICO-ECONÓMICO DE FACTIBILIDAD DE GENERACIÓN DE ENERGÍA EÓLICA EN CELAYA GUANAJUATO

J.A. Gordillo Sosa¹, R. Guzmán-Cabrera² A. González Parada², M.A Gómez Martínez², Joel Herrera Cabral¹

¹UTSOE, Universidad Tecnológica del Suroeste de Guanajuato
Departamento de Tecnologías de la Información y la Comunicación
antgor@antoniogordillo.com

²FIMEE, Universidad de Guanajuato, México
Departamento de Ingeniería Eléctrica
guzmanc@salamanca.ugto.mx

RESUMEN

La energía eólica está estableciéndose en un lugar preponderante a nivel mundial, como una de las futuras fuentes principales en el suministro de energía eléctrica. La capacidad para no depender de combustibles fósiles en el proceso de generación de energía eléctrica al utilizar el viento como combustible, hace que la generación eólica contribuya a dejar de generar toneladas de gases contaminantes al medio ambiente. El presente trabajo plantea el estudio técnico-económico del uso del recurso eólico con el que se cuenta en el estado de Guanajuato para generar energía eléctrica que solventa parte de la demanda que tiene el sistema eléctrico. Para llevar a cabo el análisis técnico económico, primeramente se obtiene un modelo matemático utilizando series de Fourier. Este modelo nos permitirá realizar estimaciones futuras. Los resultados obtenidos permiten ver la viabilidad del método propuesto.

Palabras Clave: Fuentes alternas, Eólico, Generación, Factibilidad.

I. INTRODUCCIÓN

El aumento que se tienen en la población a nivel mundial implica todo un reto para los gobiernos en sus diferentes niveles al tener que ofrecer los servicios que se requieren los habitantes como son la energía eléctrica, alumbrado público, agua etc. Este hecho ocasiona que se presente un aumento en la demanda en la demanda de energía eléctrica.

El incremento en la demanda de energía también se debe a la actividad industrial. En México se tiene una fuerte dependencia al uso de combustibles fósiles para la generación de energía eléctrica. El uso de combustibles fósiles impacta de manera negativa al medio ambiente provocando en algunos casos efectos irreversibles como es el calentamiento global, el cual puede llegar a exterminar por completo a algunas especies de animales.

Ante este escenario surge entonces la necesidad de buscar nuevas fuentes de generación de energía eléctrica que no impacten de manera negativa al medio ambiente y rindan igual o más que las energías convencionales.

Las energías renovables han tenido en los últimos años la atención ante la necesidad de disminuir las emisiones contaminantes al medio ambiente. Algunos países le han apostado a esta fuente alternativa de generación de energía eléctrica, entre estos destaca el caso de España, el cual es el segundo productor de energía eléctrica utilizando energía eólica a nivel mundial, el primer lugar lo siguen teniendo los alemanes.

Además, la generación de energía eléctrica utilizando energía eólica permite la generación de nuevas fuentes de trabajo, como lo muestra el informe publicado de Comisiones Obreras, dado a conocer por la agencia EFE, en el que se menciona que el sector de las energías renovables generó en el año 2008 15.342 empleos sólo en Cataluña, siendo el 80% de estos con contratos indefinidos.

La proyección nacional sobre las cifras de empleo generados en el sector se estiman en 89.000 empleos directos y 99.000 indirectos, con una tendencia que indica que en el año 2020 esta cantidad podría superar los 270.000 puestos de trabajo, como se puede apreciar es toda una área de oportunidad para la generación de nuevas fuentes de empleo.

El potencial estimado de generación de energía eléctrica para los próximos años se encuentra entre los 10 y los 14 GW. El mercado de las energías renovables se ha colocado entre los más importantes de Europa, causando disminuciones significativas en el costo de la tecnología y de la energía generada por este medio. Actualmente los costos promedio de la energía eléctrica generada a partir del viento se ubican entre 4 y 8 centavos de dólar por KWH generado, los cuales ya están muy cerca de los 2.5 centavos de dólar por KWH que actualmente cuesta la generación de energía eléctrica utilizando combustibles fósiles y tecnología de punta [1].

El presente trabajo está organizado de la siguiente manera: en la sección dos se presentan los principales modelos eólicos existentes en el mercado y se presenta una breve descripción de cada uno de ellos. En la sección tres se presenta el desarrollo metodológico utilizado en el presente trabajo de investigación. En la sección cuatro se presentan los resultados obtenidos y por último en la sección cinco se presentan las conclusiones y las posibles líneas que puede tomar el trabajo futuro relacionados con la investigación desarrollada.

II. MODELOS EÓLICOS

El viento contiene energía cinética (de las masas de aire en movimiento) que puede convertirse en energía mecánica o eléctrica por medio de aeroturbinas, las cuales se componen de un arreglo de aspas, generador y torre, principalmente. Las aeroturbinas pueden ser clasificadas, por la posición de su eje, en horizontales y verticales. En la figura 1 se muestran la imagen de una aero-turbina vertical y una horizontal en sus diseños más típicos.

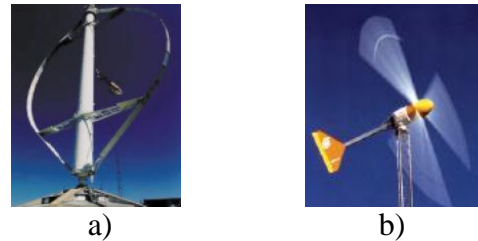


Figura 1. Aeroturbinas
a) vertical, b) horizontal

Los aerogeneradores de eje vertical tienen la ventaja de adaptarse a cualquier dirección de viento y por ello se les llama penémons (todos los vientos). No precisan dispositivos de orientación; trabajan por la diferencia de coeficiente de arrastre entre las dos mitades de la sección expuesta al viento. Esta diferencia de resistencia al viento hace que el rotor sea propenso a girar sobre su eje en una dirección específica. Además los aerogeneradores de eje vertical se emplean para generar potencias que van de los 200 W a los 4 MW. En estricto rigor no necesitan de una torre (pueden operar a nivel de piso, colocarlos en una torre les permitiría tener acceso a vientos más fuertes), también se caracterizan por tener altos torques de partida.

Los aerogeneradores horizontales por su forma pueden generar más potencia debido a que se les puede modificar tanto la altura (a mayor altura la velocidad del viento es mayor) como el tamaño de las palas para poder generar una mayor cantidad de potencia con poco viento. El inconveniente es que no pueden aprovechar el flujo de viento en ciertas direcciones como lo hace el aerogenerador de eje vertical.

III. DESARROLLO

Existen estaciones de monitoreo eólico, las cuales son las encargadas de recolectar información acerca de la velocidad y dirección del viento dominante, en el caso del estado de Guanajuato esta información es utilizada por el Departamento de Ecología del Gobierno del Estado para tener información sobre el comportamiento del mismo.

Los datos utilizados en este trabajo se obtuvieron del Instituto de Ecología del Estado de Guanajuato, las ciudades que cuentan con estaciones de mediciones sobre el viento son Salamanca, León, Irapuato, Celaya y Silao. Las estaciones de monitoreo tienen la finalidad de registrar la velocidad del viento durante el día en las ciudades mencionadas con la finalidad de prevenir a la población en caso de contingencias en las ciudades con alto grado de contaminación, como puede ser la ciudad de Salamanca. Nosotros dependemos de esta información ya que hasta ahora no tenemos la infraestructura para realizar las mediciones por cuenta propia. Por esta razón los datos corresponden al lugar donde se encuentran ubicadas las estaciones de monitoreo, sin embargo, dado que nosotros lo que buscamos es determinar la viabilidad de un método de análisis técnico económico, podemos probar nuestro método con estos datos.

Se obtuvieron los promedios mensuales para cada una de estas ciudades durante el año 2008, resultando la ciudad de Celaya con los promedios más altos. Debido a esto, el estudio se llevó a cabo utilizando los datos de la ciudad mencionada. La finalidad es crear un modelo matemático que nos permita hacer estimaciones sobre el comportamiento futuro del viento, en base a las mediciones con que se cuenta. Con la finalidad de entender el comportamiento de los datos en la figura 2 se muestra el diagrama de dispersión correspondiente. Como se puede apreciar no siguen patrón alguno, lo cual nos habla de la complejidad del problema.

Con la finalidad de encontrar una representación matemática que nos sirva para modelar estos puntos, se realizaron varias aproximaciones preliminares utilizando diversos métodos matemáticos como fueron las redes neuronales y el método de los mínimos cuadrados; sin embargo, el método que nos permitió tener la mejor aproximación fue el método de las series de Fourier.

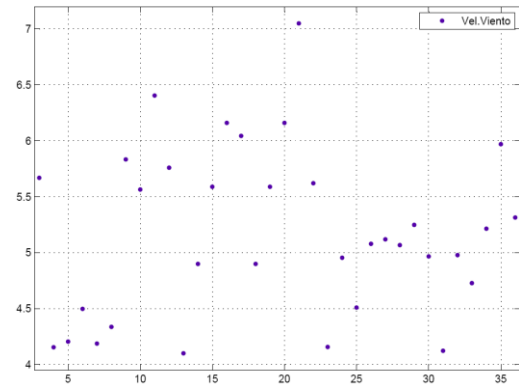


Figura 2 Diagrama de dispersión de los promedios mensuales.

En el uso de las series de Fourier se hicieron aproximaciones utilizando un número diferente de constantes, con la finalidad de encontrar la representación más adecuada, al comparar la gráfica del modelo obtenido con el diagrama de dispersión original. Como ejemplo, en la figura 3 se muestra una aproximación usando hasta cuatro constantes de Fourier; como se puede apreciar varios puntos del diagrama de dispersión no son siquiera tocados por la curva, lo cual representa un error, al alejar el valor real del valor estimado significativamente en algunos casos. Como se puede observar, lo que se busca es un modelo matemático cuya representación gráfica “toque” la mayor cantidad de los puntos del diagrama de dispersión mostrado en la figura 2.

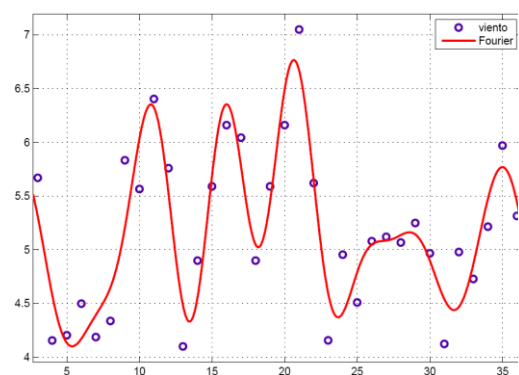


Figura 3 Aproximación matemática a cuatro constantes de Fourier.

En la figura 4 se muestra la mejor representación obtenida para el conjunto de datos proporcionado, esta se logra usando ocho constantes de Fourier.

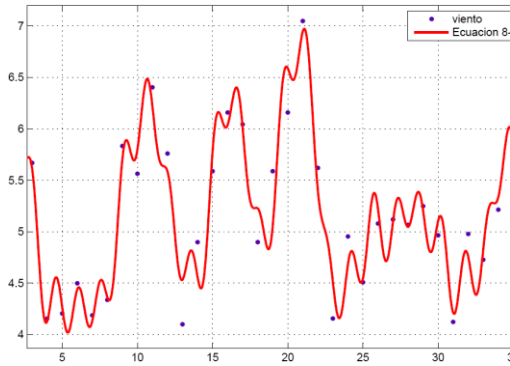


Figura 4 Aproximación matemática a ocho constantes de Fourier

Como se puede apreciar ahora la gran mayoría de los puntos son tocados por la curva modelada, lo cual nos habla de la buena aproximación del método.

El modelo matemático resultante de la aproximación mostrada en la figura 4 se muestra a continuación:

$$f(x) = a_0 + a_1 \cos(2\pi * x/c_1) + b_1 \sin(2\pi * x/c_1) + a_2 \cos(2\pi * x/c_2) + b_2 \sin(2\pi * x/c_2) + a_3 \cos(2\pi * x/c_3) + b_3 \sin(2\pi * x/c_3) + a_4 \cos(2\pi * x/c_4) + b_4 \sin(2\pi * x/c_4) + a_5 \cos(2\pi * x/c_5) + b_5 \sin(2\pi * x/c_5) + a_6 \cos(2\pi * x/c_6) + b_6 \sin(2\pi * x/c_6) + a_7 \cos(2\pi * x/c_7) + b_7 \sin(2\pi * x/c_7) + a_8 \cos(2\pi * x/c_8) + b_8 \sin(2\pi * x/c_8)$$

Los valores de las constantes utilizadas se muestran en la tabla 1.

IV. RESULTADOS OBTENIDOS

Como se pudo observar en las curvas mostradas, la velocidad del viento no es constante durante los 365 días del año.

Valores a	Valores b	Valores c
a0 = 5.283	b1 = 0.2212	c1 = 10.06
a1 = 0.1504	b2 = - 0.3385	c2 = -4.706
a2 = -0.3787	b3 = - 0.3506	c3 = 5.936
a3 = -0.3234	b4 = 0.1961	c4 = 9.068
a4 = 0.4157	b5 = - 0.0310	c5 = -1.512
a5 = 0.2563	b6 = - 0.4894	c6 = 23.35
a6 = 0.00438	b7 = 0.09759	c7 = 2.094
a7 = -0.0461	b8 = 0.08006	c8 = 3.79
a8 = -0.1456		

Tabla. 1. Valores de las constantes

Debido a esta condición, los aerogeneradores producirán solo una pequeña parte de la energía eléctrica para la cual fueron diseñados. Con los datos recabados de la velocidad del viento real por la estación ubicada en el Instituto Tecnológico en la ciudad de Celaya, se analizó el número de horas al día en las cuales se tiene una velocidad igual o mayor a la velocidad del viento promedio, en este caso es de 5.2 m/s., se obtuvo un factor de planta de 46.3%, el cual representa 4055.88 horas de operación durante un año, equivalente a 168.9 días.

Para este proyecto el análisis se planteó para 20 años. Resultó el tiempo de operación para los 20 años de 81117.6 horas. Cabe mencionar que en todas las estimaciones realizadas se utilizó el modelo matemático obtenido en el punto anterior.

Por otro lado, es necesario conocer la capacidad de generación de nuestro parque eólico, esta capacidad de producción depende totalmente de las características técnicas del generador que se utiliza en el parque. Debido a la ubicación geográfica de la estación de medición, se seleccionaron tres aerogeneradores, los cuales pueden trabajar con las velocidades del viento que presenta la zona.

Los aerogeneradores analizados fueron los siguientes:

- Joliet 20kW
- Joliet 10kW
- Joliet 5kW.

Con toda esta información como marco referencial, se procedió a hacer la estimación de la energía eléctrica que potencialmente se podría generar en los próximos 20 años, tiempo promedio de vida de los aerogeneradores.

Con el valor de la producción anual del aerogenerador, el capital generado por la producción se obtiene de la siguiente manera:

Los precios unitarios de cada equipo (aerogenerador), con un tipo de cambio en mayo del 2009 de \$ 18.38 pesos M.N. por Euro, son los siguientes:

Equipo 20K 43,640 Euros = \$ 802 103.2 MN,
 Equipo 10K 26,390 Euros = \$ 485 048.2 MN y
 Equipo 5K 19,050 Euros = \$ 350 139 M.N.

La tarifa por KWH a los usuarios por el consumo de la energía eléctrica, establecido por la Comisión Federal de Electricidad, de acuerdo a las tarifas eléctricas; doméstica, industrial y agrícola, para el año 2008 varió desde \$0.639 hasta \$.661 el KWH., en consumos domésticos, para servicios públicos de \$1.880 a \$1.980 el KWH., en media tensión y de \$2.236 a \$2.357 el KWH., en baja tensión y para consumos agrícolas varió desde \$0.980 a \$1.219 el KWH. [2].

Para poder hacer el análisis económico para cada tarifa, es necesario llevar a tiempo futuro el valor de la tarifa, debido a que se tienen incrementos mensuales y las ganancias se verán a partir del primer año de operación. Se requiere saber el crecimiento anual de la tarifa dado que la evaluación del proyecto está medida en años, con esto, se podrá conocer su comportamiento, así como, manejar mejor las cantidades de capital que involucra el proyecto.

Para tener una idea más clara de este hecho, en la tabla 2 se muestran las características de los ingresos por producción en cada servicio para el equipo de 10K.

Variables	Doméstica Básico 1-75	Servicios Públicos Media Tensión	Servicios Públicos Baja Tensión	Agrícola 1-5000
Valor Futuro de Tarifa Pt \$/Kwh.	0.687	2.112	2.502	1.569
Producción Anual KW	12955	12955	12955	12955
Producción Anual KW con Factor de planta de 46.6%	6037.03	6037.03	6037.03	6037.03
Ingresos por Producción año 1 A1 \$	4149.04	12754.66	15105.08	9472.23

Tabla 2 Características de los ingresos por producción en cada servicio para el Equipo 10K.

Conforme la tarifa se incrementa para cada servicio, las ganancias anuales de igual forma se incrementan. Pero dichas ganancias no pueden ser sumadas directamente, deben ser pasadas a costos actuales para poder realizar una comparación de la inversión inicial contra los correspondientes por producción anual de KWH.

Ahora bien, el criterio de decisión será basado en el Valor Presente Neto, como se muestra a continuación:

$$VPN \begin{cases} > 0 & \text{Si conviene siempre y cuando } VPN \gg \gg 0 \\ = 0 & \text{Ingresos} = \text{Egresos} \\ < 0 & \text{No conviene, Presenta Perdas} \end{cases}$$

Si los datos calculados caen dentro del criterio de decisión del VPN, entonces se podrá tomar una resolución más firme de la evaluación del proyecto en el aspecto económico, además, permitirá saber si el proyecto es rentable a largo plazo.

Para obtener un Valor Presente Neto de cero, el cual indicaría que el proyecto se sostiene solo, respecto a las capacidades de generación de energía eléctrica, las velocidades del viento requeridas para cada tipo de servicio (diferentes tarifas) en cada aerogenerador son las siguientes:

El problema que se genera a vientos mayores a 13 m/s es que se activan los sistemas de protección de los equipos contra vientos violentos; este valor es señalado por el fabricante. Por tal motivo, la generación necesaria para lograr un Valor Presente Neto de cero es casi imposible. Ahora bien, si se analiza el Valor Presente Neto de la inversión en el equipo y de los gastos por mantenimiento para cada uno de ellos con respecto a la generación de KWH en el tiempo de operación de cada aerogenerador, la perspectiva del proyecto cambia comparándolos con los costos por producción de la CFE.

Los gastos de mantenimiento son de \$10 000 pesos y una TREMA del 25 % con un interés del 10% a 20 años el Valor Presente Neto para cada aerogenerador es de:

$$VPN = P + 10000[P/g,]$$

En la tabla 3 se muestra el VPN para cada uno de los aerogeneradores bajo estudio.

	VPN
20 KW	\$ 863 503.2 M.N.
10 KW	\$ 546 448.2 M.N.
5 KW	\$ 411 539 M.N.

Tabla 3. Valor Presente Neto para cada Aerogenerador.

Con la producción de cada aerogenerador de 300 MWH para el de 20 KW, 259 MWH para el de 10 KW y 129 MWH para el de 5 KW.

El costo del KWH para cada aerogenerador es de:

Para el aerogenerador 20 KW = \$ 2.87 el KWH

Para el aerogenerador 10 KW = \$ 2.10 el KWH

Para el aerogenerador 5 KW = \$ 3.19 el KWH

Calculado para mayo del 2009, el costo de producción por cada KWH generado por la Comisión Federal de Electricidad fue de \$1.40

MN. Si se compara este valor con los costos de producción de cada aerogenerador para cada tipo de servicio se siguen teniendo pérdidas monetarias. Sin embargo es importante comentar que en nuestro país la energía eléctrica se encuentra subsidiada, por lo que los usuarios no pagamos la cuota real de energía y que en este estudio la comparación se hace sobre el precio subsidiado.

V. CONCLUSIONES

Los ingresos generados por el recurso eólico con el que cuenta el estado son muy bajos, aún con las consideraciones de que la velocidad del viento puede ser mayor en ciertas horas de día, la producción sigue siendo baja para el costo beneficio del proyecto. Para poder optimizar esta parte se requiere que las capacidades de generación de los aerogeneradores, en velocidades de viento bajas, sean más altas.

Otra consideración importante es que más que buscar la viabilidad de un proyecto con el presente trabajo se pretende contar con una metodología que permita llevar a cabo el estudio; esto es, si se cuenta con nuevos datos simplemente se alimenta al sistema con ellos, se genera el modelo y se procede a hacer la estimación correspondiente, teniendo en cuenta la actualización tanto de los precios como de las tarifas.

Es importante resaltar que no se tiene control alguno sobre la velocidad del viento en la zona donde se realizaron las mediciones para poder aumentar la capacidad de generación de energía eléctrica.

Finalmente, además del beneficio económico se pueden considerar los relacionados con el medio ambiente, ya que es una opción que surge de una fuente de energía limpia, con todo lo que esto conlleva.

VI. REFERENCIAS

- [1] I.Troen, E.Lundtang Petersen, *European Wind Atlas*, Department of Meteorology and Wind Energy, Riso National Laboratory, DK-4000 Roskilde, Denmark.
- [2] M.J. Souto, C.F. Balseiro, V. Pérez-Muñuzuri, M. Xue, K. Brewster, “*Impact of Cloud Analysis on Numerical Weather Prediction in the Galician Region of Spain*”, *Journal of Applied Meteorology* 42, 129-140 (2003)
- [3] C.F. Balseiro, P. Carracedo, B. Gómez, P.C. Leitão, P. Montero, L. Naranjo, E. Penabad, V. Pérez-Muñuzuri, “*Tracking the "Prestige" oil spill. An operational experience in simulation*”, *MeteoGalicia. Weather*, 58 p. 452- 458(2003)
- [4] *Manual de energía eólica investigación, diseño, promoción, construcción y explotación de distintos tipos de instalaciones*, Escudero López, Jose María, Mundiprensa, 2004.
- [5] I. Pajarón, “*Diseño y optimización de la energía producida de un parque eólico en San Javier*”, Proyecto fin de carrera UPCT Cartagena, 2006.

CONSULTAS EN INTERNET:

[http:// www.windpower.org](http://www.windpower.org).

<http://www.conae.gob.mx/>

<http://www.cfe.gob.mx/es/>

[http:// www.jccm.es/agricul/sigpac.htm](http://www.jccm.es/agricul/sigpac.htm)

EMPLEO DE SERIES DE FOURIER EN EL MODELADO DEL CAUDAL DE UN RÍO

J.A. Gordillo Sosa¹, R. Guzmán-Cabrera², A. González Parada², M.A Gómez Martínez², Joel Herrera Cabral¹.

¹UTSOE, Universidad Tecnológica del Suroeste de Guanajuato
Departamento de Tecnologías de la Información y la Comunicación
antgor@antoniogordillo.com

²FIMEE, Universidad de Guanajuato, México
Departamento de Ingeniería Eléctrica
guzmanc@salamanca.ugto.mx

RESUMEN

En este trabajo se presenta la obtención del modelo matemático del caudal del río La Antigua. Este modelo se obtiene a partir de un conjunto de datos proporcionados por la Comisión Nacional del Agua. El modelo es utilizado para predecir el comportamiento futuro del caudal del río. Mediante la predicción del comportamiento futuro del caudal del mismo, se pretende justificar la instalación de una planta de generación de energía eléctrica de tipo mini-hidráulico.

Palabras clave: fuentes alternas de energía, Serie de Fourier, software de optimización.

I. INTRODUCCIÓN

En la actualidad existen muchos países que como medida preventiva para disminuir el problema de la contaminación están haciendo uso de nuevas fuentes alternas de generación eléctrica. Estas fuentes serían una alternativa a otras tradicionales y producirían un impacto ambiental mínimo. Las energías renovables comprenden: la energía solar, la hidroeléctrica (se genera haciendo pasar una corriente de agua a través de una turbina), la eólica (se produce por una presión diferencial del aire), la geotérmica (producida por el gradiente térmico entre la temperatura del centro de la Tierra y la de la superficie), y la procedente de la biomasa (se genera a partir del tratamiento de la materia orgánica).

El objetivo principal de las nuevas fuentes alternas de generación eléctrica aplicadas a la producción

distribuida es generar la energía eléctrica lo más cerca posible al lugar de consumo, precisamente como se hacía en los albores de la industria eléctrica, incorporando ahora las ventajas de la tecnología moderna y el respaldo de la red del sistema eléctrico, para compensar cualquier requerimiento adicional de compra o venta de dicha energía.

Aunque la cantidad de energía generada por estos métodos no es a gran escala comparable con las fuentes convencionales, podemos hacer uso de la generación distribuida, que es la producción de electricidad con instalaciones que son suficientemente pequeñas en relación con las grandes centrales de generación, de forma que se puedan conectar casi en cualquier punto de un sistema eléctrico. Es un subconjunto de recursos distribuidos. En general se acepta que las potencias de los generadores van desde 3 a 5 MW - los más pequeños - y de 10 a 30 MW los de mayor capacidad.

Debido al gran problema existente en México en relación a la contaminación, sería de gran apoyo para la conservación de nuestros recursos naturales la incorporación de las fuentes alternas de generación a nuestro sistema eléctrico. Examinando la gran capacidad de agua en los ríos en los diferentes estados de México, es conveniente pensar en la generación mini-hidráulica como una posible fuente alterna de generación eléctrica.

El potencial hidroeléctrico total nacional, se estima en 53.0 GW de potencia eléctrica aprovechable, del

cual se tienen identificados 541 sitios con un potencial de 19.6 GW. Según los datos proporcionados por la CFE, el potencial hidroeléctrico aprovechado actualmente para la generación de electricidad asciende a los 9,121 MW en 77 centrales con una generación anual de poco más de 20,000 GW-h al año.

El potencial estimado para centrales con capacidades instaladas menores a los 10 MW se sitúa en los 3,250 MW. Actualmente se han instalado 34 centrales de este rango de capacidad, en las que se han instalado una capacidad total de 109.0 MW, generándose anualmente 479.0 GW-h. México con recursos hidráulicos muy importantes y uno de los grandes retos ha sido aprovechar este recurso en forma ordenada, sistemática y óptima. Este recurso puede ser aprovechado por asentamientos humanos, ganadería, agricultura y además aprovecha su potencial como fuente de energía.

El agua como fuente de energía es un recurso renovable que no contamina, no genera subproductos, su utilización como combustible no limita su aprovechamiento para otras aplicaciones, por lo que debería ser explotada con más interés para satisfacer necesidades de electrificación en regiones donde se den las conducciones.

A continuación se presenta una relación de algunos ríos y su capacidad de generación de energía eléctrica. Cabe mencionar que se presentan solamente los principales del país, aunque se podrían analizar en un futuro otros escurrimientos, debido a que existen recursos hidrológicos de baja potencia, donde es factible la instalación de sistemas micro-hidroeléctricos, que por lo regular se encuentran en la sierra y en las montañas, donde también existen muchos asentamientos humanos con escasos recursos y muy limitados servicios.

Se cuenta con un potencial hidrológico de al menos 280,000 millones de metros cúbicos al año. Sólo entre los ríos Grijalva, Balsas, Santiago e Ixtapatongo se tiene el 89.7% de la potencia hidroeléctrica total instalada [7], como se puede apreciar en la tabla 1.

Río	% de Potencia hidroeléctrica total
Grijalva	52.30
Balsas-Santiago	20.60
Ixtapatongo	16.30
Papaloapan	6.40
Yaqui-Mayo	4.40

Tabla 1 Hidroeléctricas en México

Este tipo de generación es a gran escala, pero nos sirve como ejemplo del gran potencial que se tiene para la producción de energía eléctrica utilizando los escurrimientos de agua. En este escenario, escurrimientos de menor tamaño podrían ser utilizados para generar energía eléctrica en una escala menor y precisamente en esta parte se encuentra ubicado el presente trabajo.

En este trabajo se presenta un método para la obtención del modelo matemático del caudal de un río. Con este modelo se pretenden realizar estimaciones futuras que permitan determinar la viabilidad de instalación de una central de generación de energía de tipo mini hidráulico. Este trabajo se encuentra organizado de la siguiente manera: en la sección dos se presenta el método propuesto para llevar a cabo la estimación de caudal; en la sección 3 se presentan los resultados obtenidos y por último en la sección cuatro se presentan las conclusiones.

II. MÉTODO PROPUESTO

El estudio de la factibilidad de implantación es una de las partes principales y más importantes en el desarrollo de sistemas de generación mini-hidráulica, ya que por medio de este estudio podremos observar el comportamiento a largo plazo del río en cuestión. La decisión de desarrollar el proyecto reside en las predicciones realizadas por nuestro modelo.

Con la finalidad de comprender mejor el desarrollo del presente trabajo se muestra a continuación la metodología seguida en el desarrollo del mismo.

Para el caso de sistemas mini-hidráulicos, se establece un procedimiento general para determinar un potencial aprovechable dentro de una zona determinada. Esta metodología se divide en los siguientes pasos:

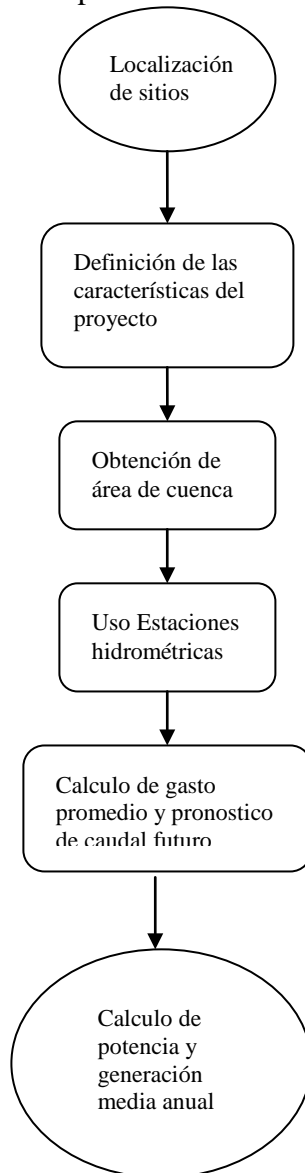


Fig. 1 Metodología para evaluar un sistema mini-hidráulico

La localización de los sitios para proyectos mini-hidráulicos, involucra tres puntos fundamentales sobre el río considerado:

- 1) El sitio donde es viable construir la obra de captación del agua.
- 2) El sitio donde es viable construir el tanque de carga e inicio de la tubería a presión y
- 3) El sitio donde es viable ubicar la casa de máquinas.

Una vez determinados los puntos antes mencionados, prácticamente quedan definidas las características del proyecto a estudiar: las dimensiones de la boquilla determinan la geometría de la presa alpina o vertedora sobre el cauce del río, el trazo y longitud de la conducción, la cota del tanque de carga, el trazo y desarrollo de la tubería a presión, la ubicación y cota de la casa de máquinas. Otro aspecto importante para la localización es el área de la sub-cuenca, de manera que el proyecto propuesto capte un escurrimiento que permita obtener las potencias y generaciones congruentes con el estudio. Las áreas de las cuencas o de escurrimiento representan la posible cantidad de captación de agua hasta el sitio de la boquilla. Se comienza por trazar el parte aguas de la cuenca por aprovechar. Este parte aguas lo define la topografía de las montañas circundantes al sitio de la boquilla. En teoría, una gota de lluvia que caiga sobre el parte aguas se dividirá: 50% se dirigirá hacia la cuenca en estudio y el restante 50% hacia otras cuencas aledañas. El área interior del parte aguas define el área de la cuenca y se mide en kilómetros cuadrados.

Las estaciones hidrométricas son de gran utilidad para este tipo de estudio; registran periódicamente el gasto que llevan los principales ríos de cualquier región de la República Mexicana. La metodología en este caso consiste en localizar la estación hidrométrica más cercana al proyecto o río seleccionado, por lo general, aguas abajo del punto de captación (boquilla).

Con base en los datos obtenidos de la estación hidrométrica, se forma la serie de tiempo que representa la variación de caudal a lo largo del tiempo. A esta serie de tiempo se le aplican técnicas de pronóstico con el propósito de obtener el comportamiento a futuro del caudal y con base en los resultados obtenidos calcular la potencia

aprovechable en el sitio seleccionado. Esto es de manera particular lo que se presenta en la sección de resultados: el modelo matemático con el cual se podrán hacer estimaciones sobre el caudal de un río.

El río empleado para estudiar la factibilidad de implantación es el río La Antigua ubicado en el estado de Veracruz.

El método propuesto para estudiar la factibilidad de implantación del sistema mini-hidráulico se basa en el uso de las series de Fourier. La ecuación de Fourier nos servirá como herramienta alternativa a los métodos estadísticos comunes para realizar predicciones. En la siguiente ecuación, la variable independiente nos describirá el tiempo y la variable dependiente del caudal:

$$y(x) = a_0 + \sum a_i \cos\left(2\pi \frac{x}{c_i}\right) + b_i \sin\left(2\pi \frac{x}{c_i}\right)$$

En el diagrama de flujo de la figura 2 se puede observar el procedimiento empleado para poder obtener un buen ajuste en la dispersión de la variable en cuestión.

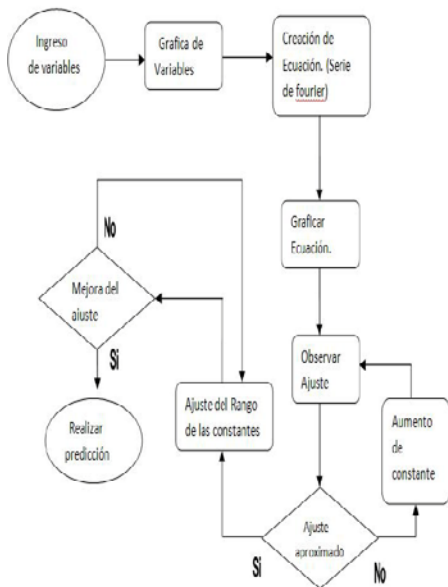


Fig. 2 Diagrama de flujo para optimización de las series de Fourier

III. RESULTADOS OBTENIDOS

Como primera parte para el desarrollo del método propuesto, se calcula la media anual desde el año 1951 hasta 1985, obteniéndose un total de 35 datos. A continuación se procede a graficarlos para observar la dispersión generada. Esto se representa en la figura 3.

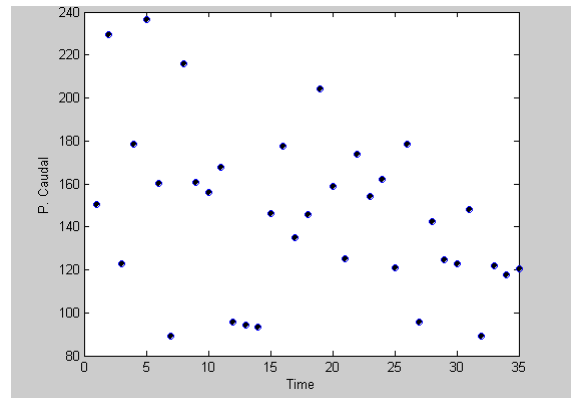
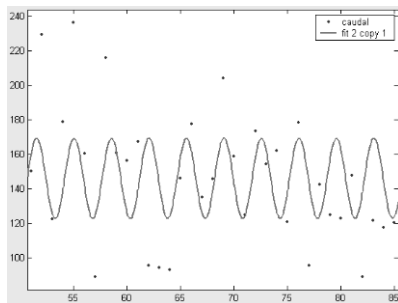


Fig. 3 Dispersión de datos

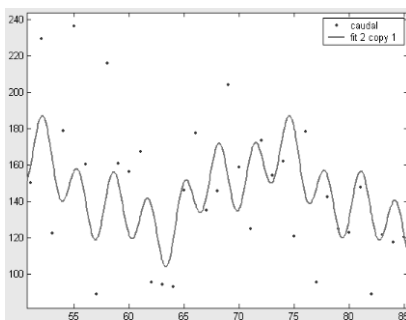
El objetivo es obtener un modelo matemático que permita realizar la representación gráfica de estos puntos en el espacio muestral. Para lograr obtener este modelo se utiliza el método de aproximación utilizando series de Fourier, en el cual se varía el número de constantes utilizadas para ir apreciando gradualmente el mejor modelo que permita representar los puntos de una manera adecuada. Como resultado de esto, en la figura 3 se muestra el proceso con un aumento de hasta cuatro constantes. Aplicando el procedimiento indicado anteriormente en la figura 2, se obtiene la ecuación siguiente, que fue el mejor ajuste que se pudo obtener al variar el rango de las constantes desde C₁ hasta C₅:

$$\begin{aligned}
 F(x) = & 145.1 - 23.71 \cos\left(2\pi \frac{x}{16.54}\right) + \\
 & 31.13 \sin\left(2\pi \frac{x}{16.54}\right) + 7.193 \cos\left(2\pi \frac{x}{2.856}\right) \\
 & - 27.9 \sin\left(2\pi \frac{x}{2.856}\right) - 0.6963 \cos\left(2\pi \frac{x}{7.721}\right) \\
 & + 7.7 \sin\left(2\pi \frac{x}{7.721}\right) + 15.09 \cos\left(2\pi \frac{x}{14.34}\right) + \\
 & 12.18 \sin\left(2\pi \frac{x}{14.34}\right) + 13.11 \cos\left(2\pi \frac{x}{9.038}\right) \\
 & + 7.472 \sin\left(2\pi \frac{x}{9.038}\right)
 \end{aligned}$$

Para ilustrar el impacto de las constantes en el modelo obtenido, se presentan a continuación algunas gráficas utilizando como aproximación diferente número de constantes. En la figura 4, parte (a), se muestra la aproximación a 2 constantes, en ella podemos observar cómo se da una mala aproximación con respecto al diagrama de dispersión ya que la mayoría de los puntos no son tocados por la curva. En la figura 4 (b) se muestra la aproximación con 6 constantes y se puede observar cómo en este incremento de constantes el modelo propuesto tiende a tomar, poco a poco, la forma de la dispersión de los datos.



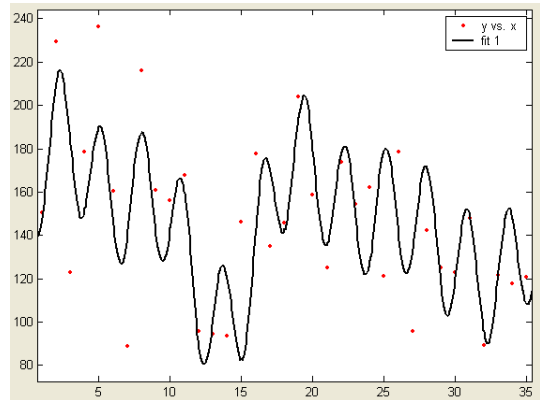
(a)



(b)

Fig. 4 Aproximación a dos constantes (a) y seis constantes (b) de Fourier.

El ajuste que se obtiene al graficar la ecuación 2 para nuestra dispersión de datos se muestra en la figura siguiente con su correspondiente modelo matemático.



$$F(x) = 145.1 - 23.71 \cos\left(2\pi \frac{x}{16.54}\right) + 31.13 \sin\left(2\pi \frac{x}{16.54}\right) + 7.193 \cos\left(2\pi \frac{x}{2.856}\right) - 27.9 \sin\left(2\pi \frac{x}{2.856}\right) - 0.6963 \cos\left(2\pi \frac{x}{7.721}\right) + 7.7 \sin\left(2\pi \frac{x}{7.721}\right) + 15.09 \cos\left(2\pi \frac{x}{14.34}\right) + 12.18 \sin\left(2\pi \frac{x}{14.34}\right) + 13.11 \cos\left(2\pi \frac{x}{9.038}\right) + 7.472 \sin\left(2\pi \frac{x}{9.038}\right)$$

Fig. 5. Mejor aproximación obtenida al realizar el ajuste del rango de constantes de C5

Este modelo es el que se utiliza para hacer la estimación de caudal en el río bajo estudio. En la figura 6 se puede observar el comportamiento que predice nuestro modelo para un periodo de datos de 35 a 60 que corresponde al periodo de años desde 1985 al 2010.

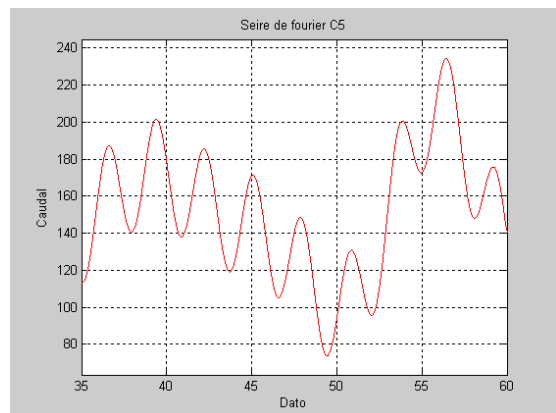


Fig. 6. Comportamiento de caudal teórico dato 35 a 60

La figura 7 muestra el intervalo completo del dato 0 al 60, donde la línea punteada nos indica la predicción realizada.

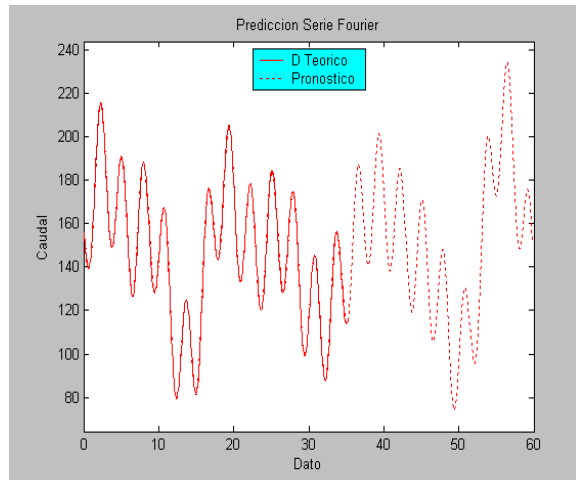


Fig. 7. Comportamiento de caudal teórico del dato 0 al 60

IV. CONCLUSIONES

Para llevar a cabo una medida de la exactitud del modelo matemático encontrado se calculó la diferencia entre el valor real y el simulado del conjunto de datos proporcionado, encontrando que este valor se aproxima a cero. Por esta razón nos atrevemos a mencionar que el modelo es preciso. Estos valores de predicción de caudal pueden entonces ser utilizados con una confianza mayor para sustentar la toma de decisiones futuras sobre posibles proyectos de generación de energía eléctrica que dependen del caudal del río bajo estudio.

V. REFERENCIAS

[1] Barry R.G. *Atmósfera, Tiempo y Clima* Omega. Barcelona, 1980.

- [2] *Contaminación Atmosférica*, Agency for Toxic Substances & Disease Registry, Junio 2007.
- [3] *Review of the national ambient air quality for sulfur oxides*, EPA, Julio 1993.
- [4] Bridgman H.A. *Global air pollution: Problems for the 1990*, Behaven Press, 1990.
- [5] *Convención del marco de las naciones de unidas sobre el cambio climático. Protocolo de Kyoto*, 11 de diciembre 1997. http://es.wikipedia.org/wiki/protocolo_de_kyoto
- [6] Sapag, Chain., *Preparación y evaluación de proyectos*, McGrawHill, 2000.
- [7] R.P.Gallardo *Factibilidad y diseño de sistemas de Generación Distribuida basados en micro generación hidráulica*, Tesis de Maestría Universidad Michoacana de San Nicolás de Hidalgo, 2006.

SISTEMA ELECTRÓNICO PARA EL MONITOREO A DISTANCIA DE PROCESOS DE CORROSIÓN MEDIANTE LA TÉCNICA DE RUIDO ELECTROQUÍMICO

Laria Menchaca Julio, Castán Rocha José Antonio, Malibrán José, Castán Rocha Emilio, Ibarra Martínez Salvador

Facultad de Ingeniería “Arturo Narro Siller”, Universidad Autónoma de Tamaulipas, Centro Universitario Tampico – Madero, Tampico, Tamaulipas, México.
Tel. 833-2402000 ext. 3335, correo electrónico: jlaria@uat.edu.mx

RESUMEN

En el presente trabajo se muestran los resultados obtenidos en el desarrollo de un sistema virtual para la medición a distancia de los procesos de corrosión mediante la técnica de ruido electroquímico. Dicho sistema está compuesto por una celda, un módulo para la medición de corriente de ruido, un módulo para la medición del voltaje de ruido, un módulo de adquisición, dos módems y una computadora portátil. El sistema está controlado por un software escrito en LabVIEW que permite presentar el sistema como un instrumento virtual. El sistema desarrollado permite detectar señales de corriente de ruido superiores a los 10nA y de voltajes de ruido superiores a los 100 μ V.

Palabras clave: Ruido electroquímico, corrosión, sistema virtual.

I. INTRODUCCIÓN

La corrosión en los metales por lo general es de naturaleza electroquímica. Por lo tanto, la medición de los parámetros electroquímicos asociados a los procesos de corrosión se puede utilizar para monitorear los procesos de corrosión. Existen varias técnicas para el monitoreo en tiempo real o en línea de los procesos de corrosión. Entre las más utilizadas están el uso de Probetas del material de interés, la medición de la Resistencia Eléctrica (RE), la técnica de Resistencia a la Polarización Lineal (RPL), la Espectroscopia de Impedancia Eléctrica (EIE) y la técnica de Ruido Electroquímico (RE) [1,2,3].

La utilización de las probetas ha sido una de las técnicas más antiguas y aun se emplean en la actualidad y consiste en determinar la pérdida de peso de la probeta durante su exposición al ambiente corrosivo y a partir de ello determinar la velocidad de corrosión del material.

La técnica de la RE es muy similar a la de la probeta con la diferencia de que en esta se mide la resistencia del material, la cual varía a medida que este pierde peso. Estas dos técnicas tienen el inconveniente de que la medición no se realiza en tiempo real sino que hay que esperar a que ocurra el proceso de corrosión. A diferencia de estas, las técnicas de RPL, EIE y RE brindan una mayor sensibilidad y permiten el monitoreo en tiempo real de la corrosión. Sin embargo, para realizar la medición de la corrosión por las técnicas de RPL y EIE, es necesario polarizar el material lo cual complica la electrónica y su implementación. Además, estas técnicas solo brindan información sobre la velocidad de corrosión pero no permiten determinar el tipo de corrosión que está ocurriendo.

De todas las técnicas para el monitoreo de la corrosión antes mencionadas, la más reciente es la de RE. Entre las ventajas de esta última tenemos que en comparación con las demás, no es necesario la polarización del material y permite el monitoreo en tiempo real de la corrosión así como determinar el tipo de corrosión que está ocurriendo en el metal [4-11]. La técnica de RE se basa en la medición de las fluctuaciones espontáneas de potencial y

corriente que ocurren entre dos o tres electrodos metálicos similares o idénticos.

Para el monitoreo en tiempo real de la corrosión se pueden encontrar sistemas como el descrito por Croall [12] que permite el monitoreo de la corrosión utilizando la técnica de RPL. Por otra parte, uno de los sistemas que se ha empleado con mucha frecuencia en el monitoreo de la corrosión es el SmartCET de la empresa norteamericana Honeywell el cual puede utilizar por separado o en conjunto las técnicas de RPL y RE [13]. Ambos sistemas están diseñados en base a un equipo electrónico que es el encargado de la polarización del material, el registro y análisis de las señales así como de la representación de los resultados.

Con el advenimiento de las computadoras y el desarrollo de sistemas de acondicionamiento y adquisición de señales analógicas, los principios y metodologías para la implementación de los sistemas de monitoreo de la corrosión han cambiado sustancialmente. Todo ello ha propiciado el desarrollo de diferentes tipos de sistemas adaptados a las necesidades concretas del medio a estudiar. Es por ello, que en el presente trabajo proponemos un sistema para el estudio de los procesos de corrosión utilizando la técnica de ruido electroquímico y que por los módulos electrónicos empleados lo convierte en un sistema virtual de fácil adaptación a diferentes condiciones de trabajo y medios de estudios.

II. MÉTODO

Los métodos de implementación así como las características que deben reunir los sistemas para la medición de las señales de corriente y potenciales en la técnica de RE se exponen en la Guía Estándar ASTM G1.11.04 [14]. En esta Guía se plantea que la técnica de RE puede implementarse con celdas de dos o tres electrodos todos del mismo material. La electrónica para la medición del potencial de ruido debe tener una alta resistencia de entrada, en el orden de 10^{11} a $10^{14} \Omega$. Además, el instrumento debe ser capaz de medir con exactitud niveles de potencial superiores a los $10\mu\text{V}$ en presencia de una señal de 1 V de DC en el

rango de frecuencias de 10^{-4} a 10 Hz. Así mismo, la electrónica para la medición de la corriente de ruido debe ser capaz de medir con buena exactitud niveles de corriente en un rango entre 10^{-11} y 10^{-3} A para el rango de frecuencias de 10^{-4} a 10 Hz. Por lo general las frecuencias de muestreo de las señales de potencial y corriente de ruido electroquímico están en el rango entre 0.5 y 2 Hz y los tiempos de medición pueden durar hasta varios días. Para la calibración de la electrónica, en [14] se exponen los pasos a seguir.

Al tratarse de una técnica en desarrollo, la metodología para el análisis de datos de ruido electroquímico sigue siendo en la actualidad un objeto de discusión, ya que, entre otras razones, el propio sistema a estudiar puede condicionar el método de análisis a utilizar. Incluso algunos autores recomiendan que, de cara a obtener la máxima información posible, sean utilizados distintos métodos para cada sistema. Entre los métodos de análisis de los datos que se utilizan están, la inspección directa de los registros experimentales, el análisis estadístico en el dominio temporal, el análisis en el dominio de frecuencias, el análisis basado en la Teoría del Caos, y el análisis basado en la Transformada de Wavelets [15-20]. En el presente trabajo para el análisis de los datos utilizamos solo el análisis estadístico en el dominio temporal de las señales registradas y en específico se determinó la desviación estándar de las señales de corriente (σI) y potencial (σV) de ruido, la determinación de la resistencia de ruido (R_n) y el índice de picadura (PI) según [15, 16].

Sistema de Medición

En la Fig. 1 se muestra el esquema en bloques del sistema electrónico desarrollado. El sistema está conformado por la celda electroquímica en la cual están los electrodos para la medición de los procesos de corrosión. Además el sistema incluye los circuitos para la medición de las señales de corriente y voltaje de ruido proveniente de los electrodos. Estos permiten la detección de esas señales y su acondicionamiento (amplificación y filtrado) para después pasar a la digitalización de

las mismas en la tarjeta de adquisición la cual se conecta al modem para la transmisión de las señales obtenidas a distancia hacia otro modem conectado a una PC que contiene el software del sistema.

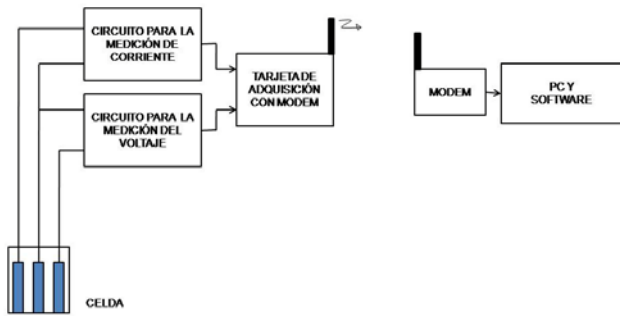


Figura 1. Esquema en bloques del Sistema Electrónico.

El diseño de celda más extendido para implementar la técnica de ruido electroquímico es el denominado sistema de tres electrodos el cual consiste en tres electrodos idénticos, o sea, del mismo material y de las mismas dimensiones. Estos se colocan los más cercano posible uno del otro y se cubren con resina epóxica, dejando solo una de las caras de los electrodos expuesta. Para las pruebas del sistema diseñamos una celda similar con electrodos de acero inoxidable del tipo 316L y el área superficial de la cara expuesta de cada electrodo es de 1cm^2 o sea 0.5cm de ancho por 1cm de largo.

Tomando como base los requerimientos que se exponen en la Guía Estándar ASTM G1.11.04 [14] y que deben cumplir los equipos para la medición de ruido electroquímico, en la Fig. 2 se muestra el esquema del circuito eléctrico para la medición de la corriente y de voltaje de ruido electroquímico que se diseñó. El circuito para la medición de corriente de ruido está constituido por un amplificador electrométrico (IC1) del tipo AD549JH de la ANALOG DEVICES y que entre sus principales características tenemos: corriente de fuga de $\pm 300\text{ fA}$, corriente de offset de 65 fA ,

voltaje de offset de $\pm 260\text{ }\mu\text{V}$, nivel de ruido de voltaje de $92\text{ nV}/(\text{Hz})^{1/2}$ para la frecuencia de 10 Hz , nivel de ruido de corriente de $0.22\text{ fA}/(\text{Hz})^{1/2}$ para el rango de las frecuencias entre 0.1 Hz y 10 Hz , impedancia de entrada de $10^{15}\text{ }\Omega$.

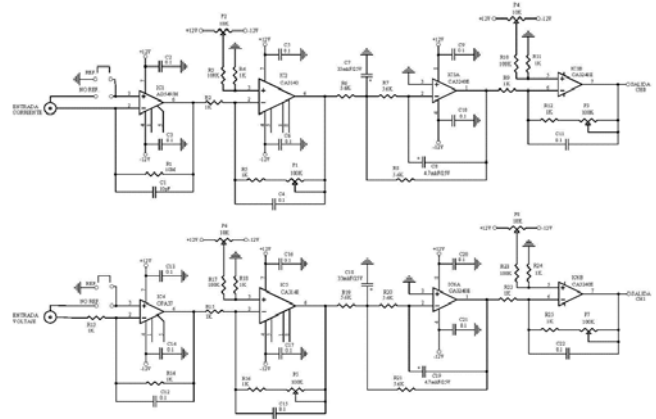


Figura 2. Circuito eléctrico para la medición de la corriente y el voltaje de ruido electroquímico.

Después del circuito electrométrico viene la primera etapa de amplificación en base al amplificador operacional IC2 y las resistencias R2, R5 y el potenciómetro P1. Este último permite variar la ganancia de amplificación de esta etapa desde $\times 1$ hasta $\times 11$. Además en esta etapa se puede ajustar el nivel de offset a la salida de esta etapa con el potenciómetro P2. A la salida de esta etapa de amplificación está conectado un filtro pasa banda de segundo orden conformado con el amplificador operacional IC3a, las resistencias R6, R7 y R8 así como los capacitores C7 y C8. Este filtro tiene una frecuencia de corte de 1 Hz , y una ganancia de $\times 1$. A la salida del filtro está conectada una segunda etapa de amplificación en base al amplificador operacional IC3b conectado de manera similar a la primera etapa de amplificación e incluye además el ajuste del nivel de offset de salida.

El circuito para la medición del voltaje de ruido electroquímico es similar al circuito de medición de corriente de ruido, la única diferencia es que en este, el circuito de entrada posee un amplificador

de bajo ruido (IC4) del tipo OPA37GH de la TEXAS INSTRUMENT.

Las señales de corriente y voltajes de ruido electroquímico amplificado y filtrado que se obtienen a la salida de los circuitos antes descritos, se envían respectivamente a dos de las entradas analógicas de la tarjeta de adquisición cuyo circuito eléctrico se presenta en la Fig. 3. Esta tarjeta está constituida por un conversor análogo digital de 12 bits (U5) cuya salida está conectada a un puerto del microcontrolador (U8) del tipo AT89C4051 de la familia 80C51. Los datos de ambos canales de medición son enviados a través de una interface RS-232, que posee esta tarjeta de adquisición, a un modem transmisor del tipo SST-2450 que envía dichas señales a otro modem receptor conectado a una computadora portátil (ver figura 1) los que pueden estar situados a una distancia no mayor de 300m del modem transmisor.

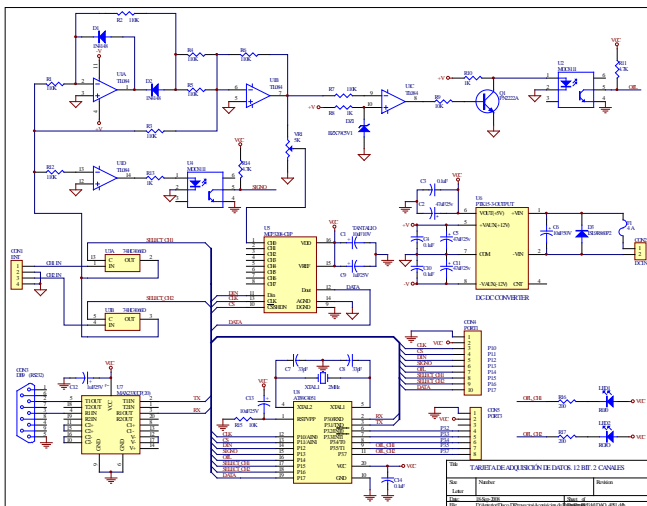


Figura 3. Circuito eléctrico de la Tarjeta de Adquisición.

La tarjeta de adquisición tiene un software propio que tiene la función de leer las señales de cada canal a la salida del conversor cada 200ms y cada 13 segundos enviar por RS-232 un paquete que contiene 40 mediciones de cada canal.

El sistema posee otro software escrito en base a LabVIEW el cual se corre en la computadora y

permite leer la señal recibida por el modem receptor a través de la interface RS-232 de la computadora, procesarla y almacenarla en un archivo para su posterior tratamiento. En la Fig. 4 se muestra la pantalla del panel frontal que vierte este software.

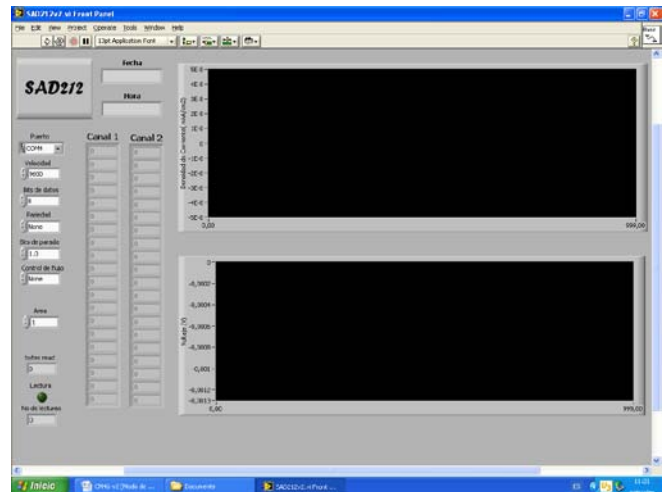


Figura 4. Panel frontal del software de adquisición.

Calibración y Validación del Sistema

La calibración del Sistema se procedió según lo expone la Guía Estándar ASTM G1.11.04 [14]. El nivel de ruido promedio obtenido para el canal de medición de corriente se obtuvo en el orden de los 36.2 pA oscilando entre ± 80 pA y está montado sobre un voltaje de DC de 13.9 mV. Así mismo, el nivel de ruido promedio obtenido para el canal de medición de voltaje esta en el orden de los 196 nV oscilando entre ± 100 μ V y está montado sobre un voltaje de DC de 746 μ V. Por tanto, el Sistema permite detectar señales de corriente que sobrepasen los ± 80 pA y señales de voltaje que sobrepasen los ± 100 μ V lo cual está dentro de los parámetros que deben cumplir los sistemas para la medición de la corrosión mediante la técnica de ruido electroquímico que se recoge en [14].

Para validar el funcionamiento del sistema realizamos la medición de los procesos de corrosión en muestra de acero inoxidable del tipo 316L en una solución al 70% de HAc, 1500ppm

de HBr y 1500ppm de HCl a temperatura ambiente. Para ello se utilizó una celda electroquímica compuesta por tres electrodos de trabajo, todos del mismo material de acero 316L, embebidos en una resina epóxica. Por la parte inferior solo se exponen las caras de los tres electrodos situados a una distancia de 1mm entre ellos y el área superficial de la cara expuesta de cada electrodo es de 0.5 mm^2 .

Las mediciones se realizaron primero utilizando un equipo comercial, un potencióstato modelo Gill AC y con el objetivo de no variar las condiciones experimentales, después de medir con dicho equipo durante un período de tiempo de 2048 segundos, continuamos la medición con el sistema desarrollado. Para el análisis de los resultados utilizamos el método de análisis estadístico en el dominio temporal de las señales registradas y en específico se determinó la desviación estándar de las señales de corriente (σI) y potencial (σV) de ruido y la determinación del índice de picadura (PI) [15, 16].

En ambos sistemas se observó que el acero 316L en la solución empleada muestra una corrosión uniforme fundamentalmente y mixta en pequeños intervalos de tiempo.

III. CONCLUSIONES

Como resultados del presente trabajo, se diseñó, calibró y validó un sistema electrónico que permite medir a distancia los procesos de corrosión electroquímica empleando la técnica de ruido electroquímico. Entre las características técnicas del sistema tenemos que con el mismo podemos medir niveles de corrientes de ruido electroquímico por encima de $\pm 80 \text{ pA}$ y niveles de voltajes de ruido por encima de $\pm 100 \text{ } \mu\text{V}$. Se desarrolló un software en ambiente LabVIEW de fácil manejo y que permite la adquisición y representación gráfica de los datos de corriente y voltaje de ruido electroquímico. El sistema desarrollado puede con muy pequeñas variaciones ser empleado para el monitoreo a distancia de las señales de diferentes tipos de sensores de parámetros físicos.

IV. REFERENCIAS

- [1] *Uhlig's corrosion handbook*, Publ: John Wiley & Sons. Inc, New York (2000).
- [2] *Electrical and electrochemical methods for determining corrosion rates*, NACE Publication 3D170, Publ: NACE, Houston, USA (1984).
- [3] Thompson N. L. and Payer J. H. *DC Electrochemical test methods. Corrosion testing made easy*, Vol. 6. Publ: NACE International, Houston, USA, (1998).
- [4] Iverson W. P. "Transient voltage changes produced in corroding metals and alloys", *Journal of the Electrochemical Society*, vol. 115, No6, pp: 617-618 (1968).
- [5] Hladky K. and Dawson J. L. "The measurement of localized corrosion using electrochemical noise", *Corrosion Science*, vol.21, No4, pp:317-322 (1981).
- [6] Hladky K. and Dawson J. L. "The measurement of corrosion using electrochemical 1/f noise", *Corrosion Science*, vol.22, No3, pp:231-237 (1982).
- [7] Hladky K. *Corrosion monitoring*, United State Patent No4575678, 11 march (1986).
- [8] Cottis R. and Turgoose S. *Electrochemical impedance and noise. Corrosion testing made easy*. Vol. 7. Publ: NACE International, Houston, USA (1999).
- [9] Edgemon G. and Barr E. "Applications of electrochemical noise: real time plant and field challenges", *Proc. Conf. Corrosion 2001*. Paper 01283. Publ: NACE International, Houston, USA (2001).
- [10] DeBruyn H. J., Lawson K. and Heaver E. E. "On-line monitoring using electrochemical noise measurement in CO-CO₂-H₂O system", *Electrochemical noise measurement for corrosion applications*, ASTM STP 1277, pp:214. Publ: ASTM, USA (1996).
- [11] Macdonald D., Liu C. and Manahan M. "Electrochemical noise measurements on

carbon and stainless steels in high subcritical and supercritical aqueous environments”, Electrochemical noise measurement for corrosion applications. ASTM STP 1277, pp:214. Publ: ASTM, USA (1996).

- [12] S. Croall and A. Mc Ilraith “*Remote corrosion monitoring helps reduce operating cost*”, Pipeline & Gas Industry, pp:68- 72 (2001).
- [13] <http://honeywell.com>.
- [14] J.R. Kearns, D.A. Eden, M.R. Yaffe, J.V. Fahey, D.L. Reichert, D.C. Silverman, “*ASTM Standardization of Electrochemical Noise Measurement*”, Electrochemical Noise Measurement for corrosion Applications, ASTM STP 1277, pp 446-470 (1996).
- [15] D.A. Eden, A.N. Rothwell y J.L. Dawson, Proc. Corrosion/91, Paper 444, Houston, EE.UU., 1991, National Association of Engineers.
- [16] D.A. Eden, A.N. Rothwell, “*Electrochemical Noise data: Analysis, Interpretation and Presentation*”, en Corrosion 92, Paper 292, NACE, Houston (1992).
- [17] J. Gollner, I. Garz, K. Meyer, *Korrosion* 17 (1986).
- [18] K. Nachstedt, K.E. Heysler, *Electrochim. Acta* 33 (1988).
- [19] Aballe V., F.J. Botana P., M.M. Barcena; *Ruido Electroquímico. Métodos de Análisis*; Septem Ediciones, Oviedo (2002).
- [20] D.E. Williams, Proc. *Electrochemical Corrosion Testing*, Ferrara, Italy, 1985, Dechema.

especialidad en física, en el Centro de Investigación en Ciencia Aplicada y Tecnología de Avanzada del IPN. Tiene más de 20 años de experiencia en el diseño y desarrollo de instrumentación electrónica.

M. en C. José Antonio Castán Rocha obtuvo el título de Maestría en Ciencias con especialidad en Tecnología Educativa en la Universidad Autónoma de Tamaulipas. Tiene más de 5 años de experiencia en el desarrollo de software.

M. en C. Emilio Castán Rocha obtuvo el título de Maestría en Ciencias con especialidad en Electrónica en el Instituto Tecnológico de Madero. Tiene más de 5 años de experiencia en el desarrollo de equipos electrónicos.

M. en C. José Malibrán obtuvo el título de Maestría en Ciencias con especialidad en Electrónica en el Instituto Tecnológico de Madero. Tiene más de 15 años de experiencia en el desarrollo de equipos electrónicos.

Dr. Salvador Ibarra Martínez obtuvo el título de Doctor en Ciencias con especialidad en robótica en la Universidad de Girona, España. Tiene más de 5 años de experiencia en el desarrollo de equipos electrónicos y robóticos.

Agradecimientos

Los autores del presente trabajo le agradecen al Fondo Mixto CONACYT-Gobierno del Estado de Tamaulipas, México, su apoyo en recursos para el desarrollo de los trabajos así mismo le agradece a la Universidad Autónoma de Tamaulipas los apoyos recibidos.

V. AUTORES

Dr. Julio Laria Menchaca obtuvo su título de Doctor en Tecnología de Avanzada con

SISTEMA DIGITAL DE TEMPORIZACIÓN

E. Montoya-Suárez, I.O. Rodríguez-Ruíz, L. León-Roa

Área de Ciencias Básicas e Ingenierías
Universidad Autónoma de Nayarit
Cd de la Cultura “Amado Nervo”, Tepic, Nayarit, México
correo electrónico: emontoya@nayar.uan.mx

RESUMEN

Se presenta el diseño y desarrollo de un sistema digital de temporización (SiDiTe) para gestionar el tiempo en el control de edificios, escuelas, casas y pequeñas empresas, entre otras. El sistema tiene como núcleo principal un microcontrolador PIC. El SiDiTe implementado muestra ser un sistema de bajo costo, confiable y muy flexible en diversas situaciones reales de control que requieren gestión de tiempo, esto en comparación con un sistema comercial.

Palabras clave: Temporización, microcontrolador PIC, control.

I. INTRODUCCIÓN

La tecnología electrónica atraviesa por un período de rápido desarrollo, en donde el empleo de la electrónica digital para solucionar problemas ha hecho que esta tecnología reemplace a otras, y desarrolle nuevos productos cada vez más económicos y miniaturizados.

Los microcontroladores como parte del diseño en los sistemas de control automatizado están conquistando el mundo. Los encontramos presentes en nuestro trabajo, en nuestra casa y en nuestra vida. En general, se pueden encontrar controlando el funcionamiento de los ratones y teclados de las computadoras, en los teléfonos, en los hornos de microondas y los televisores, juguetes, impresoras, módems, el sistema de arranque de un automóvil, y otras aplicaciones con las que seguramente no estaremos tan familiarizados como son la instrumentación electrónica, control de sistemas en una nave

espacial, entre otros. Además, de que una aplicación típica podría emplear varios microcontroladores para controlar pequeñas partes del sistema total.

Como ya es bien sabido, el actual desarrollo tecnológico toma gran importancia en los múltiples campos científicos, así como en la vida cotidiana para lograr los mejores resultados, es decir, una solución óptima según las necesidades de la industria, edificios, comercios, escuelas, viviendas, entre otros campos de aplicación. Ejemplo de ello es la apertura inmensa de los sistemas de control automático y la importante labor que puede desempeñar un sistema electrónico, capaz de llevar a cabo la automatización a través de diseños electrónicos e informáticos, como soluciones deseables a situaciones específicas, en una diversidad de formas adaptándose a las necesidades de las instalaciones eléctricas.

Sin duda alguna, los beneficios que proporcionan las automatizaciones son, y serán, innumerables y a ello se debe, al gran crecimiento que han tenido las tecnologías electrónicas y el alto impacto que han provocado en las sociedades del mundo. Una de las tecnologías que más se han desarrollado son los sistemas digitales de temporización para diferentes y únicas aplicaciones como soluciones reales. Una gran variedad de productos inundan el mercado actualmente [1-3], con diversas soluciones, a muy diversos precios y complejidades, aumentando la eficiencia en su uso.

Al observar la necesidad que existe de desarrollar sistemas de bajo costo para la automatización y

control de diversas aplicaciones en distintas áreas, aunado con la amplia utilización que se le puede dar a un sistema de temporización, en este trabajo se diseña y construye un sistema digital de temporización que sea capaz de gestionar el tiempo con diferentes funciones controlando situaciones reales, que resulte ser una alternativa sencilla, además de económica y competitiva.

II. PROPUESTA

Se propone diseñar y construir un sistema autómatas, que sea capaz de gestionar el tiempo con diferentes funciones que controlen distintas situaciones reales. Facilitando al usuario el manejo del sistema al programar las funciones; resultando ser una alternativa sencilla, económica y competitiva para algunos procesos, esto debido a que existe mucha competitividad extranjera que constituye un reto [1-3], consiguiendo con esto un paso más en el desarrollo de tecnología nacional enfocada a satisfacer las necesidades del control en nuestro país. La Fig. 1 muestra el diagrama a bloques propuesto para cumplir con lo anterior.

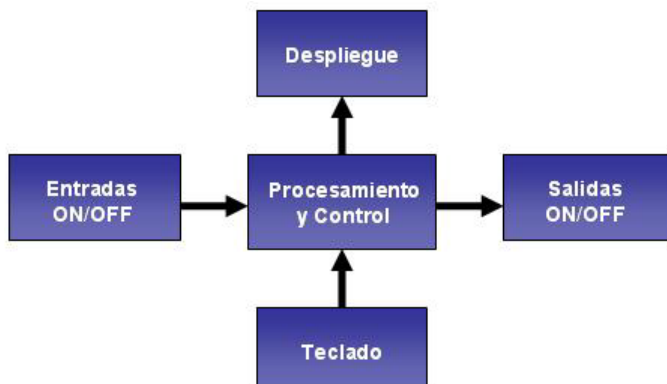


Figura 1. Diagrama a bloques del SiDiTe propuesto.

En las etapas de *Entradas-Salidas ON/OFF* se utilizan circuitos opto-acopladores y dispositivos de potencia para que el sistema pueda interactuar con entradas de diferente voltaje; y con cargas grandes como motores, lámparas, entre otros.

El *Teclado* está conformado por 8 botones (pushbutton) conectados en forma individual y leídos en modo polling.

Para la etapa de *Despliegue* se utiliza un exhibidor de cristal líquido (LCD), el cual es un periférico de salida muy potente y flexible, además de que los dispositivos más adecuados para optimizar sus prestaciones y control son los microcontroladores; en realidad el LCD tiene incrustado un microcontrolador específico para regular su funcionamiento. Con dicho exhibidor se desplegarán los menús para realizar la programación de las funciones de las que consta el SiDiTe, permitiendo la interacción con el usuario.

La unidad de *Procesamiento y Control* está basada en un microcontrolador PIC, y cuyas funciones principales son:

- Configuración del LCD.
- Funcionar como reloj de tiempo real (RTC).
- Programación de las funciones de temporización.
- Procesamiento de datos programados.
- Control del despliegue de la información.
- Activación/desactivación de salidas.

III. HARDWARE

Se optó por seleccionar el PIC16F877 [4,5] debido a su alta capacidad de memoria y de pines de entrada/salida. Este dispositivo es de la compañía Microchip, e incluye todos los componentes de una computadora.

El teclado se compone de 8 pushbutton normalmente abiertos, los cuales se encuentran conectados en los pines PB0-PB4 del puerto B y en los pines PD0-PD2 del puerto D. Al cerrar uno de los botones provoca que se introduzca un “1” lógico al microcontrolador PIC, lo que indica que se ha presionado una tecla.

La pantalla LCD 16x4 [6], 4 líneas con 16 caracteres cada una, se conecta al puerto C (PC0-PC7) y al puerto E (PE0-PE2), con el contraste y la iluminación regulados por medio de resistencias variables con un valor de 10k Ω cada una. Para desplegar los menús en el LCD se emplearon las tres líneas de control manejadas por el puerto E y las ocho líneas de datos conectadas al puerto C.

Las líneas de control sirven para el manejo de las diferentes funciones del LCD como enviar un dato o instrucción, o leer el estado del LCD; es decir, son las líneas que gestionan el uso de las líneas de datos. Mientras que las líneas de datos son usadas para el envío de datos (caracteres) o instrucciones.

Las salidas del sistema están destinadas cinco de ellas al puerto A (PA0-PA5) y sólo una al puerto D (PD3). Como dichas salidas deben de manipular cargas que consumen corrientes elevadas, se utilizan dispositivos de acoplamiento óptico (opto-acopladores) con salida a TRIAC (MOC3011) [7] y dispositivos TRIAC (TIC226D) que soportan hasta 10A [8].

Cuando se aplica un “0” lógico a la entrada del opto-aislador, el LED no emite luz con lo que el optoTRIAC no conduce. Por otra parte, si se aplica un “1” lógico, el LED proporciona la energía suficiente para el optoTRIAC conduzca, permitiendo que el tiristor TRIAC de salida se dispare, con lo cual la carga a controlar se encontrará energizada.

Así mismo, las entradas del sistema se encuentran conectadas al puerto B (PB5-PB7), acopladas con un circuito opto-aislador con salida a transistor (PC817) [9] y sus respectivas resistencias limitadoras de corriente, esto con el fin de que permita trabajar con entradas de voltaje diferente al que trabaja el sistema completo, por ejemplo, 12V ó 24V de cd.

Debido a que el SiDiTe debe de contar con un módulo de respaldo de energía para el funcionamiento del RTC y del LCD, en caso de un corte de energía, se consideró implementar dicho módulo como un circuito de alimentación armado con diodos de respuesta rápida (1N4148) [10], reguladores de voltaje (L7805) [11], capacitores de 100nf y una batería alcalina de 9V. Por lo que, el voltaje suministrado por la fuente de alimentación debe ser mayor al de la batería para que el diodo que está conectado con la batería no entre en conducción y no ocasione la descarga de la batería. Entonces, en el momento que haya una falla en la energía principal, la batería permite que el

microcontrolador PIC y el LCD sigan activos, pero no mantendrá la iluminación del LCD ni alimentará a los otros dispositivos del sistema, ya que estos dependerán del regulador conectado a la energía principal.

La Fig. 2 muestra el diagrama esquemático completo del SiDiTe desarrollado.

IV. SOFTWARE

El código fuente se realizó en lenguaje ensamblador para microcontrolador PIC, basándose en el software MPLAB [12] de Microchip, el cual es una herramienta que permite la simulación en pantalla del funcionamiento de los microcontroladores PIC, así como interactuar con diferentes grabadores para PIC, entre los que destaca el PICSTAR plus.

El programa se encuentra desarrollado en forma modular, es decir, con subrutinas que permiten fácilmente conocer el flujo del código y una rápida localización de errores, así como ahorrar memoria y tiempos de ejecución.

A. Subrutina: Reloj de Tiempo Real (RTC)

Es la rutina básica del sistema, la cual utiliza la interrupción del timer1 para crear la base de tiempo de 100ms (1/10 s). Esta rutina permite un manejo de tiempo en horas, minutos y segundos, en el formato de 24 horas. Dicha base de tiempo permite la ejecución de las funciones del sistema en tiempo real.

B. Subrutina: Calendario

Esta subrutina se basa en el RTC, tomando en cuenta el formato de fecha en día, mes y año. Para obtener el último día del mes, se considera el número de mes y si el año actual es bisiesto, si lo anterior se cumple se hace un incremento en mes; así mismo, si el mes es igual a 13 se incrementa año, partiendo de la condición que la fecha inicial del calendario es a partir del año 2000 con un límite de 60 años.

C. Subrutina: Día de la Semana

Esta subrutina viene asociada a las rutinas del RTC y del calendario. Haciendo que las fechas y días de

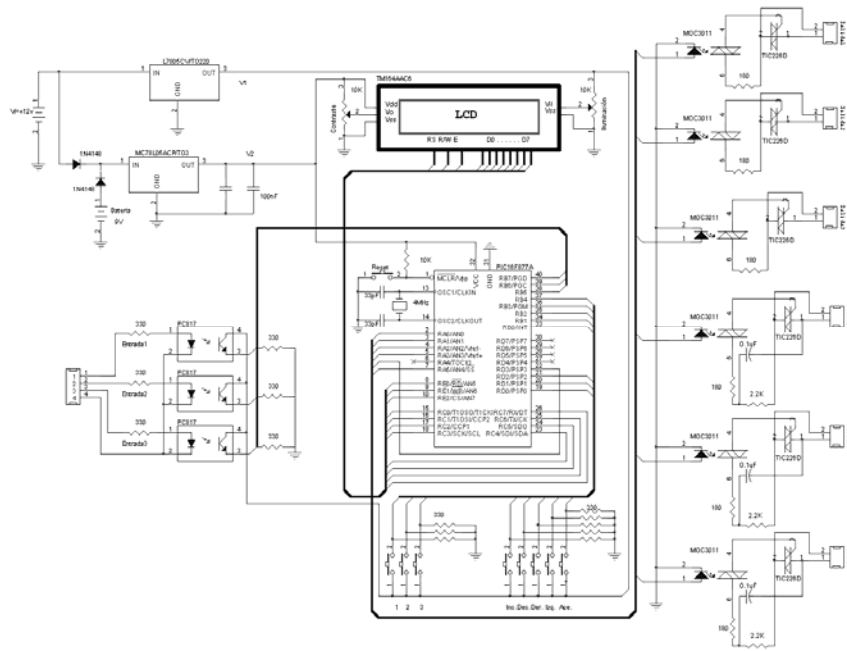


Figura 2. Diagrama esquemático del hardware para el SiDiTe.

la semana cambien con el tiempo. El cálculo del día de la semana se desarrolló considerando el calendario Gregoriano [13], el cual establece que todos los años múltiplos de cuatro son bisiestos, y los años seculares (que acaban en 00), solamente son bisiestos cuando son múltiplos de 400. Para el cálculo del día de la semana se considera una clave por cada mes (EFM AMJ JAS OND = 144 025 036 146) y dependiendo del resultado será el día de la semana (0 sábado, 1 domingo, 2 lunes, 3 martes, 4 miércoles, 5 jueves, 6 viernes).

D. Subrutina: Horario Verano/Invierno

Para el cambio de horario se tomó en cuenta el sistema que se maneja en el país. Para el verano se suma una hora en el reloj en el primer domingo de Abril, y para el invierno en el último domingo de octubre se resta una hora.

E. Subrutina: Teclado

Esta rutina se ejecuta continuamente y tiene la función de verificar el estado de los bits 0, 1, 2, 3 y 4 del puerto B, y los bits 0, 1 y 2 del puerto D. Si se presiona una tecla, se identifica cual ha sido pulsada y se asigna su función de acuerdo al estado interactivo en el que se encuentre.

F. Subrutina: Despliegue

El despliegue es parte básica de la interacción del usuario con el sistema, ya que mediante éste se muestran los mensajes que le indican al usuario las acciones que puede ejecutar. Para realizar el despliegue de los mensajes, se lee el mensaje previamente guardado en una tabla, una vez cargado el mensaje se le indica al LCD que se hará una escritura/lectura enviando carácter por carácter, para finalizar la transmisión se verifica que el carácter a enviar sea un cero.

V. FUNCIONES DEL SISTEMA

El SiDiTe es un nuevo concepto de diseño para el control de edificios, tiendas, escuelas, casas, pequeñas empresas, entre otras; en donde se presenta la gestión del tiempo. Este sistema ofrece cinco funciones básicas y sencillas: Programación horaria/anual, Programación horaria/semanal, Programación impulsos, Temporización al cierre, e Intermitente.

El sistema permite controlar seis canales de salida, y cuenta con tres entradas únicas para condicionar la programación de dos de las funciones (temporización al cierre e intermitente), todo según las necesidades de la aplicación. Por ejemplo, se puede asignar una de las cinco funciones por canal

de salida, o hasta tres diferentes programaciones de función en tiempos distintos en el mismo canal; aunque se pueden realizar otras combinaciones.

A. Programación Horaria/Anual

Esta función permite programar periodos fechados y distribuidos en el año.

B. Programación Horaria/Semanal

Esta función permite programar periodos horarios:

- Diferentes para cada día de la semana.
- Idénticos para varios días de la semana.
- Distribuidos en varios días de la semana.

C. Programación Impulsos

Esta función permite programar impulsos:

- Impulsos (de 1 a 59 segundos) en diferentes instantes de la semana.
- Un impulso puede programarse en uno o varios días de la semana.
- Posibilidad de programar impulsos de duración diferente en un mismo día.

D. Temporización al Cierre

Esta función permite temporizar la activación del canal de salida teniendo una entrada asociada.

- El tiempo de duración de la temporización se puede programar en hr/min/seg.
- La temporización comienza al activar la entrada asociada (es una condición mantenida).
- La desactivación de la entrada anula la temporización o desactivación del canal de salida.
- Se puede programar esta función en un periodo horario programable en uno o varios días de la semana.
- Periodos horarios diferentes e idénticos para cada día de la semana.
- La duración de la temporización puede ser variable distribuida en los días de la semana.

E. Intermitente

Esta función permite:

- Activar un canal de salida durante un periodo de encendido y un periodo de apagado, de manera repetida.

- Da la posibilidad de condicionar la función por una entrada.
- Se puede programar o autorizar esta función en un periodo horario programable en uno o varios días de la semana.
- Los periodos de encendido y apagado pueden programarse en varios días de la semana con un tiempo de duración igual y/o diferente.

VI. PRINCIPIO DE FUNCIONAMIENTO Y PROGRAMACIÓN

La forma de programación es muy sencilla guiada con un lenguaje claro. Las principales características de la programación son:

- El sistema ajusta automáticamente el cambio de horario verano/invierno.
- Respeto año bisiesto.
- Tiene la opción de ajustar fecha y hora.
- Permite realizar tres programaciones para cada una de las funciones horaria/anual, horaria/semanal y programación impulsos.
- Permite realizar dos programaciones para cada una de las funciones temporización al cierre e intermitente.
- Permite modificar parámetros de las funciones programadas.
- Permite borrar completa la función programada.
- Tiene una clave de acceso que protege al sistema y que viene activada por los diseñadores, pero con la opción de poder cambiarla si se desea.

La pantalla Principal y la función de cada tecla se observan en la Fig. 3.

A continuación se describirá paso a paso un ejemplo de programación para tener un panorama aún más extenso del funcionamiento y de cómo guiarse a través de los menús para llevar a cabo las programaciones. Ejemplo: Dentro de la función “Programación Horaria/Anual” con periodos horarios fechados y distribuidos en todo el año, realizar lo siguiente: Activar el canal 1 de salida en el día 30/01/00 a las 16:00 horas y desactivarlo el día 05/02/00 a las 08:00 horas.

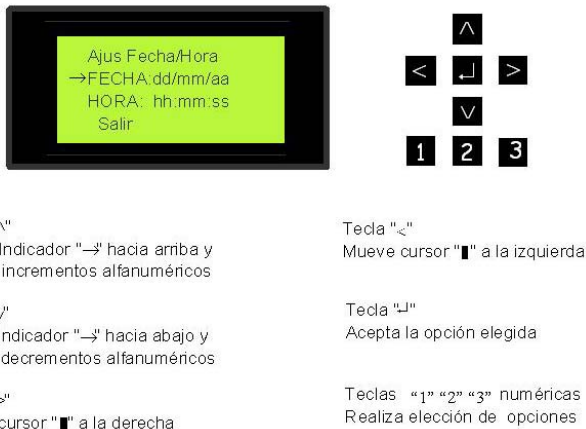


Figura 3. Pantalla Principal y funciones del teclado.

Los pasos a seguir son:

1. Al encender el sistema lo primero es ajustar el tiempo en fecha y hora (Fig. 4.a). Posteriormente se presiona aceptar ↵ en salir para ir a la Pantalla “Menú Principal”.
2. Estando en la Pantalla “Menú Principal” se selecciona → programación y se presiona la tecla aceptar ↵ (Fig. 4.b)
3. En el “Menú Programación” se selecciona → crear función y se presiona la tecla aceptar ↵ (Fig. 4.c).
4. En este “Menú Funciones” se selecciona la función a programar → Horario/Anual presionando la tecla aceptar ↵ (Fig. 4.d).
5. Pasando a la Pantalla “Seleccionar Canal” donde se asigna (1) a canal de salida con la tecla incremento ^ y se presiona la tecla aceptar ↵ (Fig. 4.e).
6. Para entrar a la Pantalla “Activar/Desactivar/HS” donde se posiciona el cursor █ y con la ayuda de las teclas se introduce la fecha/hora de activación y desactivación (Fig. 4.f). Terminada la programación de los parámetros regresa a la Pantalla “Menú Programación”, en donde se pueden hacer cambios, ya sea borrar la programación completa, modificar sólo algunos

parámetros o bien poner en ejecución el programa saliendo a la Pantalla “Estado”.

7. En esta pantalla de estado se observa activo el canal de salida programado (canal 1), permaneciendo en este estado durante el periodo de tiempo programado (Fig. 4.g). Una vez terminado el periodo programado, se observa que el canal de salida se pone inactivo (Fig. 4.h).

8. Ahora bien, si antes de poner en ejecución la programación de dicha función se necesita hacer un cambio, se debe regresar a la Pantalla “Menú Programación” donde se selecciona → Mod. Parámetros y se presiona la tecla aceptar ↵ (Fig. 4.i).

9. Para entrar a visualizar el número de programaciones existentes (en este caso solo aparecerá la programación que hemos creado) presione la tecla 1, y confirme que se desea modificar presionando nuevamente 1 que indica “Sí” (Fig. 4.j-k).

10. Con lo anterior se regresa a la Pantalla “Activar/Desactivar/HA” donde se visualizan los parámetros que se han creado anteriormente. Si se requiere modificar sólo el periodo de desactivación presione la tecla derecha > para pasar directamente a hacer el cambio en la desactivación (Fig. 4.m-n). Si se optó por no hacer la modificación, entonces se elige la opción “No” con la tecla 2 (Fig. 4.l) y el sistema se pone de nuevo en la Pantalla “Menú Programación” y se puede seguir programando.

11. En otro de los casos, si lo que se elige es borrar la programación de dicha función, solo hay que elegir la opción dentro del mismo menú de programación y consecuentemente pasa algo similar a lo de modificar la programación, en donde se hace la pregunta ¿si en realidad se desea borrar?, si es “Sí” se borra completamente la programación, y se puede realizar otra programación completamente distinta a la que se había hecho presentándose la pantalla con la función ya borrada; si se decidió “No” borrar la función entonces se regresa a la Pantalla “Menú de programación” (Fig. 4.o-s).

12. Si se optó por hacer algunos cambios a las programaciones cuando el sistema está en ejecución (Pantalla “Estado”), debe asignar la clave que le pida el sistema (Fig. 4.t). Si la clave fue correcta, se regresa a la Pantalla “Menú Principal”, de lo contrario regresa a la Pantalla “Estado”.

VII. CARACTERÍSTICAS TÉCNICAS

Las características técnicas del sistema SiDiTe y su comparación con los sistemas comerciales DTM-9 de DMX Inc [2] e Ikeos de Schneider [1] se muestran en la tabla 1.

Tabla 1. Características Técnicas del SiDiTe.

Característica	<i>DTM-9</i>	<i>Ikeos</i>	SiDiTe
Núm. De Funciones	1	9	5
Canales de Salida	1	4	6
Canales de Entrada	1	6	3
Tipo de Carga	Resistiva	Resistiva	Resistiva/ Inductiva
Corriente por Salida	5A	10 ^a	10A
Respaldo de Energía	Sí	Sí	Sí
Costo (pesos)	\$2,200.00	\$5,000.00	\$1,200.00

Cabe señalar que se siguen desarrollando funciones, así como se está buscando otra alternativa de microcontrolador que permita incrementar el número de entradas/salidas con lo cual se alcanzará mayor funcionalidad y costo-beneficio, y estar en posibilidad de manejar cargas más grandes. El costo máximo del nuevo prototipo se estima en \$2,500.00.

VIII. CONCLUSIONES

Realizar un sistema de control va más allá de armar y probar circuitería, es necesario que se utilicen todas las herramientas disponibles para llevar a cabo un proyecto que permita satisfacer algunas necesidades en forma confiable y óptima. Para probar el funcionamiento del sistema se tuvieron que realizar diversas pruebas, en donde se conectaron cargas inductivas y resistivas en cada uno de los canales de salida que conforman el

sistema y condicionando alguna de ellas a las entradas del mismo, dichas entradas manejadas a través de pulsadores. Las pruebas arrojaron resultados ampliamente satisfactorios, más allá de lo esperado inicialmente; llegando a un punto en donde el proyecto tiene un gran alcance en la fase de automatización puesto que es un diseño para una gran variedad de aplicaciones y a un bajo costo.

La implementación de este proyecto está basada en la necesidad real, en observaciones en la industria, en el campo, pequeñas empresas, el hogar, comercios, escuelas entre otros. Con los conocimientos de ingeniería electrónica, programación y control automático, se observó que era posible realizar un sistema digital de temporización a bajo costo y con características de confiabilidad y flexibilidad comparables a un sistema comercial. Al término del desarrollo se puede concluir que:

- ✓ Es posible desarrollar un sistema de control automático con parámetros programables basándose en un sistema digital y que contenga operaciones de fácil manejo.
- ✓ Construir un sistema con pantalla de despliegue que muestre diferentes menús, conjuntamente con un teclado para un fácil manejo de los mismos, permite realizar las programaciones para llevar a cabo un control automático en tiempo real. Esta característica lo hace muy atractivo puesto que resulta ser una alternativa sencilla, económica y competitiva, y de esta manera se cumple el objetivo principal del sistema.
- ✓ El sistema tiene la posibilidad de expandirse en cuanto a la programación. Pueden desarrollarse más funciones con mayor número de programaciones por función, más canales de salida, como también más entradas, una interfaz hacia la PC, con lo que se obtendría una mayor proyección del sistema.

Con lo anterior se logra tener una visión más amplia de los alcances que tiene el sistema y de la gran proyección que puede alcanzar.

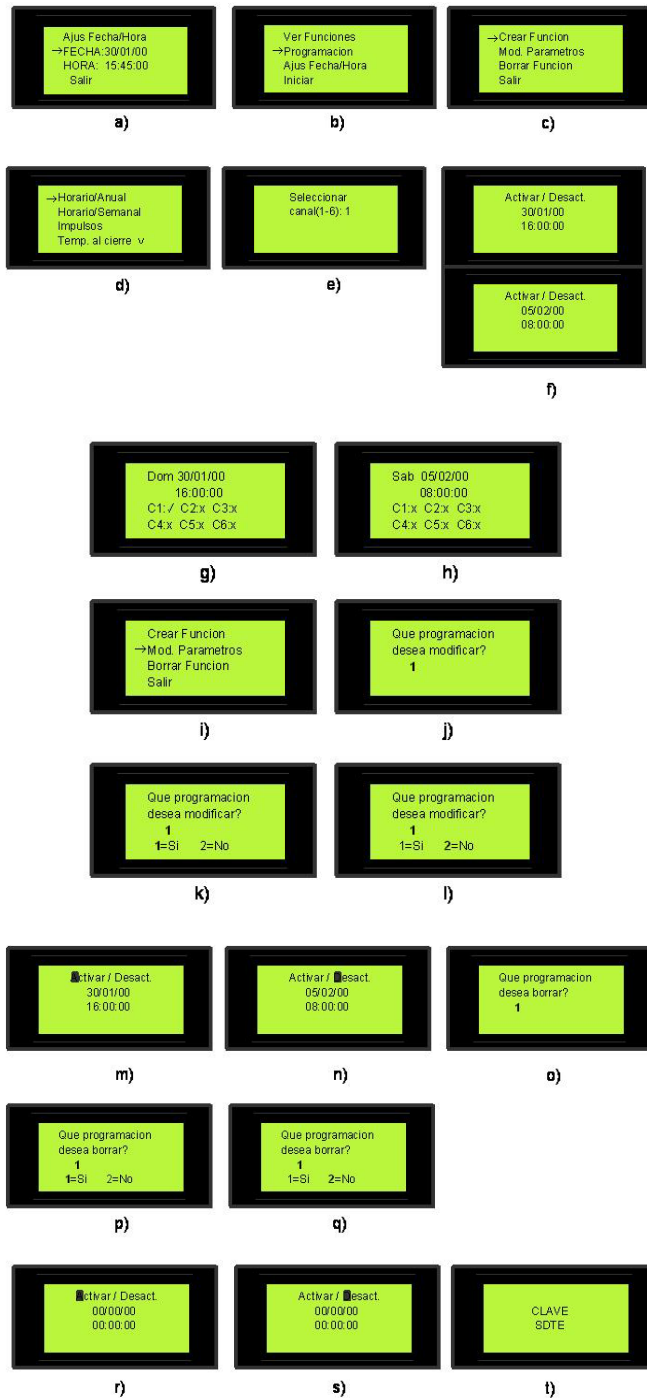


Figura 4. Secuencia para un ejemplo de Programación Horaria/Anual.

IX. REFERENCIAS

[1] *Ikeos Merlin Gerin*, Schneider Electric en <http://www.schneiderelectric.es/html/Ikeos/index.html>.

[2] *Temporizador Electrónico Programable para Siete Días EMX DTM-9*, en <http://www.emxinc.com/dtmnine.html>.

[3] *Temporizador Digital Programable para 7 días*, en http://www.extech.com/instruments/resources/manuals/375475_UMsp.pdf

[4] J. M. Angulo, I. Angulo, *Microcontroladores PIC 16F87X*, McGraw-Hill, 2000.

[5] *PIC16F87XA Datasheet*, Microchip Technology Inc. en <http://www.microchip.com/downloads/en/device.doc/39582b.pdf>.

[6] *HD44780U LCD-II Datasheet*, Hitachi en <http://www.microchip.com/downloads/en/device.doc/39582b.pdf>.

[7] *Random-Phase Optoisolators Triac Driver Output Datasheet*, Fairchild Semiconductor en <http://www.fairchild.com/ds/MOC3011-M.pdf>.

[8] *TIC226A/B/C/D/E/M/S/N Silicon TRIACS Datasheet*, Texas Instruments en <http://www.datasheetarchive.com/download.php?pi=240999>.

[9] *PC817 Series Datasheet*, Sharp en <http://www.spelektroniikka.fi/kuvat/pc817.pdf>.

[10] *L7800 Series Datasheet*, ST en <http://www.st.com/stonline/books/pdf/docs/2143.pdf>.

[11] *1N/FDLL 914/A/B/ 916/A/B/ 4148/ 4448 Small Signal Diode Datasheet*, Fairchild en <http://www.fairchildsemi.com/ds/1N/1N4148.pdf>.

[12] J. M. Angulo, I. Angulo, *Microcontroladores PIC: Diseño práctico de aplicaciones 16F84X*, McGraw-Hill, 1999.

[13] *Cómo calcular el día de la semana mentalmente*, en http://www.cnice.mec-d.es/eos/MaterialesEducativos/mem2000/astromonia/chicos/pensar/calendario_mental.html.

INSTRUMENTO VIRTUAL PARA LA MEDICIÓN Y ANÁLISIS DE LAS VIBRACIONES MECÁNICAS EN MÁQUINAS ROTATIVAS

Laria Menchaca Julio^a, Castán Rocha José Antonio^a, Malibrán^a José, Castán Rocha Emilio^a, Ibarra Martínez Salvador^a, Arias Reyes Luis Augusto^b.

^a Facultad de Ingeniería “Arturo Narro Siller”, Universidad Autónoma de Tamaulipas, Centro Universitario Tampico – Madero, Tampico, Tamaulipas, México.

Tel. 833-2402000 ext. 3335, correo electrónico: jlaria@uat.edu.mx

^b Servi Equipo Industrial S.A. de C.V., Niños Héroes 306, Col. Primero de Mayo, C.P.85410, CD. Madero, Tamaulipas, México

RESUMEN

En el presente trabajo se expone un instrumento o sistema virtual para la medición y análisis de las vibraciones mecánicas. El mismo se diseñó en base a un acelerómetro, un módulo electrónico de adquisición y una computadora portátil, todo controlado por un software desarrollado bajo la plataforma LabVIEW. El sistema permite medir a partir de las señales del acelerómetro, la aceleración, la velocidad, el desplazamiento y la condición de los baleros de los sistemas o máquinas rotativas con un error no mayor al 2%. Además permite el análisis de la Transformada rápida de Fourier para un mejor diagnóstico de los problemas que pueda presentar la máquina. Como resultado del trabajo se tiene un sistema que permite realizar un diagnóstico general y en tiempo real del estado de la máquina que se estudia, similar al resultado que se obtiene con sistemas que se ofertan en el mercado pero cuyos costos son muy superiores al sistema presentado.

Palabras clave: Vibraciones mecánicas, acelerómetros, instrumentación virtual.

I. INTRODUCCIÓN

Debido a los crecientes avances tecnológicos, las industrias han tenido que hacer modificaciones e innovaciones a sus procesos productivos, con el fin de ser altamente competitivos. Como consecuencia de ello, se han diseñado un gran número de equipos rotatorios que giran a

velocidades más altas y son fabricados con nuevos materiales que los hacen más livianos y con alta flexibilidad mecánica. Por tanto, se ha hecho necesario un mayor control de los incrementos en las fuerzas dinámicas y estáticas en los equipos. Estas fuerzas producen vibraciones que conllevan a fatigas y desgastes en los componentes de la máquina y aumentan la posibilidad de que aparezcan fallas anticipadas e imprevistas. La medición y análisis de vibraciones como técnica de prevención y diagnóstico de fallas mecánicas de máquinas en operación, constituye, actualmente una de las técnicas fundamentales de todo plan de Mantenimiento Predictivo.

Entre las causas que provocan las vibraciones en una máquina podemos mencionar la falta de rigidez, fallas de lubricación, desalineación de ejes, desbalances, problemas eléctricos, elementos sueltos, problemas hidráulicos, etc. Varios estudios realizados en máquinas rotatorias han identificado el desbalance como la causa fundamental de las vibraciones producidas en ellas; algunos estudios han establecido que cerca del 40% de los problemas de vibraciones son atribuidos al desbalance y un 30% a la desalineación [1].

El monitoreo de las vibraciones se puede realizar en tres niveles [2]. El primer nivel consiste en medir el promedio de las vibraciones de la máquina lo que permite solo detectar si existe algún problema. El segundo nivel consiste en

realizar un análisis espectral o de frecuencias mediante la Transformada Rápida de Fourier lo cual permite localizar donde existe el problema de la máquina. El tercer nivel consiste en utilizar técnicas especiales como el análisis de picos de energía, el método de pulsos de choques o los diagramas de cascada las que permiten realizar un análisis más detallado de diferentes defectos de la máquina.

Existen varios tipos de equipos para el monitoreo de las vibraciones y que son de amplio uso en la industria [3,4]. Estos equipos por sus características los podemos dividir en dos grupos. En el primer grupo están los equipos portátiles de pequeño tamaño, de precios bajos, por lo general operados con baterías, utilizan un solo sensor y permiten la medición del nivel promedio de las vibraciones en unidades de desplazamiento, velocidad y aceleración (ejemplo el modelo 810), los que además incluyen la medición del promedio de picos de energías (ejemplo el modelo 81) y los que incluyen el análisis espectral de las mediciones (ejemplo los modelos 880 y 890) [2]. En el otro grupo se encuentran los equipos y sistemas que por lo general se instalan en línea y los que pueden trabajar con más de un sensor a la vez y por consiguiente sus precios son más elevados. Estos sistemas por lo general pueden medir los tres niveles de monitoreo de las vibraciones pero tienen el inconveniente que el análisis espectral no se realiza en tiempo real [5].

Con el advenimiento de las computadoras y el desarrollo de sistemas de acondicionamiento y adquisición de señales analógicas, los principios y metodologías para la implementación de los sistemas análisis de vibraciones desde el punto de vista de instrumentación electrónica se han podido simplificar dándole mayor peso al software. Sin embargo, el proceso de monitoreo de las vibraciones consume tiempo y aún los equipos son costosos [5,6]. Es por ello, que en el presente trabajo presentamos el desarrollo de un sistema virtual diseñado en base a un sistema de adquisición de datos, bajo la plataforma del software LabVIEW que permite realizar el

monitoreo de las vibraciones mecánicas por los dos primeros niveles antes expuestos.

II. MÉTODO

Existen varias normas que dan una condición del estado de la maquina en dependencia del nivel promedio de vibraciones medido [2, 4, 6, 7, 8, 9, 10]. Sin embargo estas normas solo se deben utilizar como una guía ya que están basadas en un amplio rango de máquinas y no pueden ser exactas para un tipo específico de máquina y su condición de operación. El sistema aquí presentado ha sido diseñado para realizar el monitoreo del nivel promedio de las vibraciones tomando como base las normas antes referidas. Además se puede realizar el análisis de la severidad de las vibraciones mecánicas según la norma ISO-10816 [11] y el análisis espectral mediante la Transformada Rápida de Fourier (FFT) de la señal del sensor en tiempo real.

Las vibraciones mecánicas pueden ser ocasionadas por una gran variedad de factores tales como: desbalances, falta de alineación, elementos rodantes defectuosos, lubricación inadecuada, aflojamiento mecánico, problemas en las bandas de accionamiento, problemas de engranaje y problemas eléctricos [2, 3, 6], cada uno de estos factores se manifiesta de manera diferente en las señales de vibraciones. Por ejemplo, Las vibraciones debidas al desbalance de la máquina es una de las causas más comunes de la vibración y en ocasiones es producido por problemas de excentricidad o sea cuando la línea central del eje no es la misma que la línea central del rotor. En este caso, los datos arrojados indican que la frecuencia de vibración se manifiesta a 1x las rpm de la pieza desbalanceada y la amplitud es proporcional al nivel de desbalance. La vibración debido a desalineamientos, en la mayoría de los casos indican que la frecuencia de la vibración es 1x, 2x y 3x las rpm y la amplitud es proporcional al nivel de desalineamiento. Las vibraciones debidas a elementos rodantes defectuosos, defectos en las pistas, en las bolas o en los rodillos de rodamientos ocasionan vibraciones de altas

frecuencias y esta no es necesariamente múltiplo de las rpm de la maquina.

Por tanto, es importante realizar el análisis de frecuencias de la señal de vibración mediante la FFT para poder obtener un diagnóstico más detallado y preciso de los problemas que presenta la máquina.

Desarrollo del sistema

El sistema virtual para el análisis de vibraciones mecánicas se diseñó de forma modular donde se incluye el módulo sensor, el módulo de adquisición y el módulo de procesamiento como se muestra en la Fig. 1.



Figura 1. Esquema en bloques del Instrumento virtual.

La estructura modular, así como las características del módulo de adquisición de datos y del módulo de procesamiento permiten que al sistema se le puedan incluir varios módulos sensores. Sin embargo, en el sistema que presentamos solo se incluye un módulo sensor.

El módulo sensor está constituido por un acelerómetro modelo IRD-970 el cual contiene el acelerómetro, su circuito electrónico de polarización y amplificación de la señal, todo integrado en un solo dispositivo construido de acero inoxidable como medida de protección para ambientes industriales. Entre las características fundamentales de este módulo sensor están que permite la medición de las señales de vibración en cualquier posición tanto horizontal, vertical y axial, tiene una frecuencia de resonancia de 25KHz y una sensibilidad de 50mV/g.

El módulo de adquisición está constituido por una tarjeta de adquisición de datos modelo NIDAQ-

6112 que se une al módulo de procesamiento a través del puerto USB. Entre las características fundamentales de la tarjeta de adquisición están que permite 16 entradas analógicas con una velocidad de conteos máxima de hasta 250 KS/s y una resolución de 16 bits.

El módulo de procesamiento está constituido por una computadora portátil y un software diseñado en plataforma LabVIEW. En la Fig. 2 se muestra la pantalla principal del sistema obtenida a partir del software y en la Fig. 3 se muestra un esquema de los bloques que conforman el software del sistema.

La señal como resultados de las vibraciones y que se obtiene a la salida del módulo del sensor se digitaliza en el módulo de adquisición y se pasa al bloque 1 de la Fig3. del software que no es más que un filtro pasa banda al que se le puede variar tanto la frecuencia de corte inferior y superior entre 10Hz y 30KHz en los controles del panel frontal, representados en el área marcada con el número 1 de la Fig. 2. Como el sensor que se utiliza es un acelerómetro la señal de salida del filtro pasa banda se envía directamente al selector de unidad de medida a la selección de aceleración esta unidad se mide en in/s^2 . Para obtener el valor de velocidad de la señal de vibración en unidades de in/s , la señal se integra vez en el bloque 2 de la Fig. 3 y así mismo para obtener la señal de desplazamiento, en unidades de mil, se integra la señal de velocidad en el bloque 3 de la misma figura.

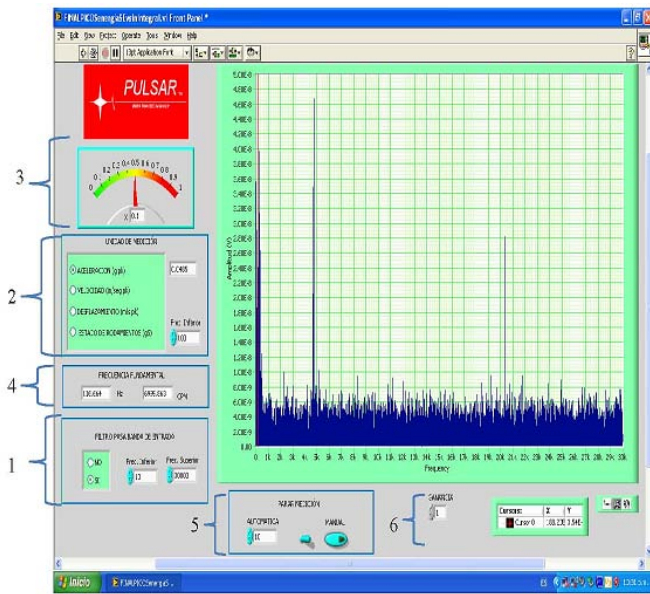


Figura 2. Pantalla principal del software del sistema.

La unidad que se desea leer se define en el selector del panel frontal representado por el área marcada con el número 2. A la señal a medir que sale del selector, se le determina el valor rms en el bloque 4 de la Fig.3 y el resultado se refleja en el indicador de aguja representado por el área marcada con el número 3 en el panel frontal (Fig. 2) cuya escala se regula de manera automática en el bloque 5 de la Fig. 3. En el selector existe una cuarta posición que nos permite determinar el nivel de la señal para conocer el estado de los baleros. Esta unidad se conoce también como picos de energía y solo la miden los equipos IRD-811. Es una unidad muy importante para el análisis de problemas en baleros fundamentalmente y son señales de alta frecuencia hasta un máximo de 30 KHz. Para determinar el estado de los baleros, en el bloque 6 de la Fig. 3, la señal proveniente del sensor se pasa a través de un filtro pasa alto cuyo valor de frecuencia de corte se puede seleccionar en el control del panel frontal que se encuentra dentro del área marcada con el número 2. Posteriormente, a la señal de salida del filtro se le determina el valor pico-pico y se pasa a través de una línea de retardo con una constante de tiempo específica y de ahí al selector.

El sistema permite realizar el análisis de la señal de utilizando la Transformada Rápida de Fourier

(FFT). Para ello, la señal de salida del filtro pasa banda se pasa a través del bloque 7 de la Fig. 3 donde se ajusta la ganancia el cual se puede realizar en el control que se encuentra en el área 6 de la Fig. 2 y de ahí se pasa al bloque 8 donde se realiza la FFT y el espectro obtenido se representa en tiempo real en el bloque 9 que es el gráfico que aparece en la pantalla principal del sistema. En el área marcada con el número 6 en el panel frontal, Fig. 2, hay unos controles para poder mover el cursor en el espectro, así como para poder ampliar zonas determinadas dentro del espectro y moverse por todo el espectro. El bloque 10 en la Fig. 3, permite seleccionar el nombre del archivo en el que se desee guardar los valores de la unidad medida y el espectro de frecuencias obtenido. En el área marcada con el número 5 en la Fig. 2 se encuentra el control del tiempo de medición. Este tiene dos posiciones, manual en donde la medición es detenida por el usuario y automática donde la medición se detiene una vez transcurridos los segundos seleccionados previamente en el selector que se encuentra en la misma área marcada. De esta manera podemos medir las unidades de vibración y realizar mediante la FFT, un análisis en tiempo real de la señal de vibración que estamos monitoreando y poder determinar y localizar el problema que presenta la máquina.

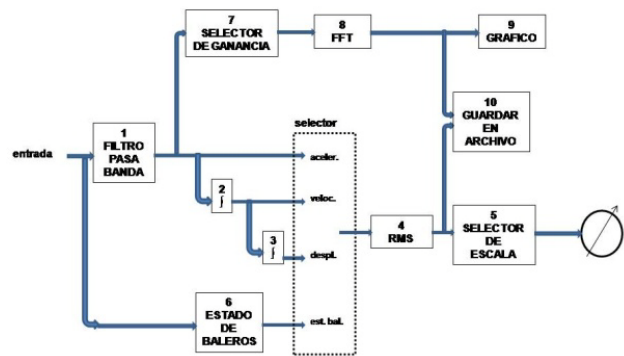


Figura 3. Esquema en bloques del software del sistema

El sistema se calibró en la empresa Servi Equipo Industrial S.A. de C.V. Para la calibración de las unidades de aceleración, velocidad y

desplazamiento, así como la FFT, se utilizó un calibrador de vibraciones modelo 401 de la Tig-Tek INC. La calibración de la unidad de estado de baleros, se realizó por método de comparación utilizando un equipo IRD-811 calibrado con un error del 1%. En la calibración se obtuvo un error en la medición para todas las unidades de medida de no mayor al 2%. Por tanto, la exactitud de medición del sistema se obtuvo muy buena, si se tiene en cuenta, que los equipos que generalmente se emplean para la medición de las vibraciones mecánicas alcanzan hasta un 10% de error en sus mediciones.

III. CONCLUSIONES

Como resultado del trabajo se desarrolló un sistema virtual para el análisis de vibraciones en máquinas rotativas. El sistema virtual se diseñó de forma modular donde se incluye el módulo sensor, el módulo de adquisición y el módulo de procesamiento. La estructura modular, así como las características del módulo de adquisición de datos y del módulo de procesamiento permiten que al sistema se le puedan incluir varios módulos sensores. Sin embargo, en el sistema que presentamos solo se incluye un módulo sensor.

El software del sistema permite el análisis del nivel promedio de las vibraciones en las unidades de aceleración, velocidad, desplazamiento así como medir el estado de los baleros y la FFT de la señal en tiempo real, con un error no mayor al 2%, lo que posibilita realizar un análisis y diagnóstico más preciso de la condición de la máquina.

IV. REFERENCIAS

- [1] G.A. Joshi and C.C. CHEN. “*Vibration Condition and Energy Loss Monitoring: A case Study*”, Proceedings of Vibration Institute. Vol. 17. N° 20. St Louis. USA(2000).
- [2] M. P. Srivastava, *IRD Mechanalysis, Vibration Monitoring for Predictive Maintenance*, Purchase, (1993).
- [3] R. Collacott, *Vibration Monitoring and Diagnosis*. Wiley (1979).

- [4] B. C. Nakra, G. S. Yadava and L. Thuestad, *Vibration Measurement and Analysis*, National Productivity Council, New Delhi, pp. 54-60, (1989).
- [5] S. Booksh. “*Virtual Instrumentation*”, Proceedings of National Technical Training Symposium and 26th Annual Meeting. Pittsburgh. Pennsylvania. USA. (2002).
- [6] J. S. Rao, *Vibratory Condition Monitoring of Machines*, Narosa Publishing House, pp. 354-356, (2000).
- [7] B. K. N. Rao, *Handbook of Condition Monitoring*, 1st edition, Elsevier Advanced technology, UK, pp.76, (1996).
- [8] B. R. Satyan and H. N. Nagarajan, *Predictive Maintenance through Vibration Monitoring*, Technical article, Noise and Vibration Laboratory, CMTI, (1988).
- [9] *Mechanical Vibration of Machines with Operating Speeds from 10 to 200 rev/sec.–Basis for Specifying Evaluation Standards*, ISO 2372, (1974).
- [10] *Vibration Limits for Maintenance*, Canadian Government Specification, DA/MS/NVSH/ 107.
- [11] ISO 10816-3. “*Mechanical Vibration – Evaluation of Machine Vibration by Measurements on Non-Rotating Parts*”. ISO. Switzerland (1998).

V. AUTORES

Dr. Julio Laria Menchaca obtuvo su título de Doctor en Tecnología de Avanzada con especialidad en física, en el Centro de Investigación en Ciencia Aplicada y Tecnología de Avanzada del IPN. Tiene más de 20 años de experiencia en el diseño y desarrollo de instrumentación electrónica.

M. en C. José Antonio Castán Rocha obtuvo el título de Maestría en Ciencias con especialidad en Tecnología Educativa en la Universidad Autónoma de Tamaulipas. Tiene más de 5 años de experiencia en el desarrollo de software.

M. en C. Emilio Castán Rocha obtuvo el título de Maestría en Ciencias con especialidad en

Electrónica en el Instituto Tecnológico de Madero. Tiene más de 5 años de experiencia en el desarrollo de equipos electrónicos.

M. en C. José Malibrán obtuvo el título de Maestría en Ciencias con especialidad en Electrónica en el Instituto Tecnológico de Madero. Tiene más de 15 años de experiencia en el desarrollo de equipos electrónicos.

Dr. Salvador Ibarra Martínez obtuvo el título de Doctor en Ciencias con especialidad en robótica en la Universidad de Girona, España. Tiene más de 5 años de experiencia en el desarrollo de equipos electrónicos y robóticos.

Ing. Luis Augusto Arias Reyes es el gerente de la empresa Servi Equipo Industrial S.A. de C.V. y con la cual tenemos estrecha vinculación debido a la experiencia de dicha empresa en el servicio de medición y análisis de vibraciones mecánicas.

Agradecimientos

Los autores del presente trabajo le agradecen al Fondo Mixto CONACYT-Gobierno del Estado de Tamaulipas, México, su apoyo en recursos para el desarrollo de los trabajos así mismo le agradece a la Universidad Autónoma de Tamaulipas los apoyos recibidos.

MONITOR CARDIACO MÓVIL EN TELÉFONOS CELULARES PARA LA DETECCIÓN DE ANOMALÍAS CARDIACAS

E. Rojas, C. Morales, A. Juárez, F. Mata

Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzada del
Instituto Politécnico Nacional
Avenida Instituto Politécnico Nacional No. 2580, Colonia Barrio la Laguna Ticomán
Delegación Gustavo A. Madero, CP. 07340 México D.F, Tel. 57296000

RESUMEN

Se presenta un sistema para monitorear a un paciente que sufra de anomalías cardíacas, por ejemplo, síncope o fibrilación auricular (FA). El sistema consta de un hardware de adquisición de datos (*monitor cardiaco*) y una aplicación para teléfono celular (*detector cardiaco*). El *monitor cardiaco* obtiene la señal del corazón y la envía vía Bluetooth a la aplicación *detector cardiaco* instalada en el teléfono celular del paciente. El *detector cardiaco* procesa la señal por medio de un algoritmo que detecta la FA.

Adicionalmente, el detector cardiaco permite visualizar en tiempo real el electrocardiograma (ECG) y la frecuencia cardíaca en la pantalla del teléfono celular. También, cuando el paciente sufre un episodio de FA, el *detector cardiaco* manda una alerta al paciente y envía un correo electrónico y/o mensaje SMS dirigido al médico para informar sobre el episodio. Se han obtenido resultados satisfactorios durante las pruebas. El sistema brinda asistencia tecnológica al médico para dar seguimiento a pacientes que padecen FA, permitiendo hacer diagnósticos y tratamientos más efectivos comparados con el enfoque tradicional, siendo esta la principal aportación.

Palabras Clave: Fibrilación auricular, Monitor cardiaco, Bluetooth, Aplicación móvil, Salud a distancia, Telemedicina.

I. INTRODUCCIÓN

Las telecomunicaciones intervienen en diferentes ciencias, una de ellas es la medicina, permitiendo tener grandes beneficios y asistencia tecnológica tanto para pacientes como para el personal médico. Un caso particular ocurre con ciertos padecimientos del corazón, su diagnóstico, seguimiento y monitoreo utilizando dispositivos especiales.

Por ejemplo, los electrocardiógrafos convencionales monitorean la actividad eléctrica del corazón y permiten detectar diferentes anomalías cardíacas [7]. Sin embargo el paciente debe acudir al hospital para una revisión médica y recibir un diagnóstico. El resultado es entregado en cinta de papel térmico el cual tiene grabada la actividad eléctrica del corazón. Este escenario ocurre cuando la anomalía cardíaca puede detectarse en pequeños lapsos de estudio (cuando un paciente asiste a una clínica para ser revisado). Sin embargo, cuando la anomalía no se detecta durante la revisión se requeriría que el paciente estuviera bajo supervisión durante varios días. Entonces, el paciente debe realizarse otro tipo de estudios, como el de Holter [3].

Por lo tanto, para asistir, mejorar y automatizar este proceso de diagnóstico y seguimiento, se pueden dotar a un dispositivo de monitoreo cardíaco de las capacidades de comunicación inalámbrica, por ejemplo Bluetooth. Así como aprovechar que hoy en día la mayoría de los teléfonos celulares pueden comunicarse con otros dispositivos usando esta tecnología. De esta forma se desarrolló un prototipo

de adquisición y transmisión inalámbrica de la señal de ECG de un paciente mediante comunicación Bluetooth, donde una aplicación para teléfono celular analiza una señal biológica para detectar anomalías cardíacas de FA que son enviadas a un médico para su revisión.

II. ANTECEDENTES

La fibrilación auricular es una arritmia que se caracteriza por un ritmo rápido e irregular debido a movimientos convulsivos de fibras musculares individuales del corazón. También, se identifica por la pérdida de la activación cíclica y ordenada de las aurículas, la cual es sustituida por una actividad continua y desordenada donde simultáneamente pequeñas áreas auriculares se despolarizan [14]. El diagnóstico de fibrilación auricular reposa en la demostración electrocardiográfica de las ondas «f», las cuales son caracterizadas por ser oscilaciones irregulares en configuración, amplitud, frecuencia y reproducibilidad de sus ciclos secuenciales [1]. En la Figura 1 se muestra un ECG de fibrilación auricular.

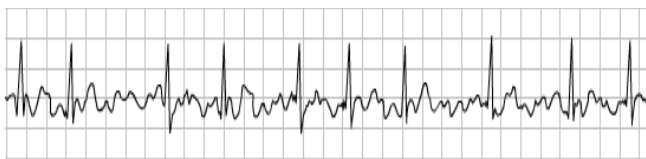


Fig. 1. ECG de una fibrilación auricular

La fibrilación auricular tiene dos variedades clínicas, la paroxística y la crónica.

La fibrilación auricular paroxística se presenta en episodios de minutos, horas o días. La forma crónica es la que persiste por semanas, meses o años. Esto, implica la pérdida de la contracción auricular organizada [1], [7].

La certeza diagnóstica de fibrilación auricular paroxística radica en el número de eventos de FA registrados. Un hallazgo interesante en este sentido fue presentado en un estudio de la farmacéutica Roche [2]. En tal estudio se evaluaron 65 pacientes

con palpitaciones recurrentes con Holter de 24 horas negativo. Se les instaló un dispositivo de registro de eventos con análisis automático de la arritmia, encontrando que en una observación promedio de 77 + 36 horas, 31% de los pacientes tenían fibrilación auricular paroxística, siendo la mitad de ellos asintomáticos durante el registro. Esto sugiere que entre más larga sea la observación mayor es la probabilidad de hacer un adecuado diagnóstico [2].

Es evidente que el electrocardiograma es útil para el diagnóstico de fibrilación auricular y se debe tomar en el momento en que el paciente presente los síntomas (disnea, palpitaciones, síncope, mareo o dolor torácico) o signos que sugieran el diagnóstico. Sin embargo, aún en el paciente asintomático el electrocardiograma puede mostrar anomalías como alteraciones de la onda «P», hipertrofia ventricular, infartos, bloqueos de rama, etc., que sugieren anomalías estructurales que pueden favorecer la aparición de la fibrilación auricular. El registro electrocardiográfico establece fácilmente el diagnóstico en fibrilación auricular permanente o persistente pero no será tan sencillo en el caso de fibrilación auricular paroxística [1].

El monitoreo electrocardiográfico de 24 - 48 horas (Holter) es una herramienta muy comúnmente usada también para evaluación de diferentes tipos de arritmias. Como es obvio, tiene un papel importante en la fibrilación auricular y al igual que el electrocardiograma tiene aplicaciones en el diagnóstico, pronóstico y respuesta al tratamiento [1]. Sin embargo, la utilidad del Holter en el diagnóstico de fibrilación auricular paroxística dependerá de la frecuencia de los eventos para que durante el registro de 24-48 horas se puedan detectar fácilmente. Sólo basado en la probabilidad de detectar un paroxismo se debe tomar la decisión de utilizar el Holter.

Se ha intentado evaluar si ciertas características determinadas en el Holter pueden ayudar a predecir el pronóstico de los pacientes con fibrilación

auricular. Un hallazgo encontrado en los pacientes con fibrilación auricular paroxística es que en la medida que disminuye la frecuencia de las ondas «f» los pacientes terminarán su episodio de fibrilación auricular con muy buena exactitud por lo que este seguimiento en el Holter puede mostrar este comportamiento. Sin embargo, la utilidad clínica de este dato es difícil de reconocer debido a que los estudios no pueden hacerse durante largos periodos por comodidad del paciente [4].

Otros trabajos han utilizado métodos alternos para detectar anomalías cardiacas en un electrocardiograma (ECG). Por ejemplo en [20] se utilizó la transformada wavelet para la detección de estas señales. En donde se reportan 0,15% falsas detecciones de 46 archivos de la base de datos de arritmias MIT-BIH.

Por otra parte, la implementación para la comunicación inalámbrica Bluetooth, se puede realizar con el JSR 82 (Java Service Request) que está implementado en J2ME. En este documento se especifican las clases, interfaces y métodos para brindar esta funcionalidad en un dispositivo móvil como PDA's, celulares o SmartPhones.

III. PLANTEAMIENTO DEL PROBLEMA

En la actualidad existen varias alternativas para poder monitorear las señales cardiacas de un paciente, por ejemplo, el estudio de Holter y el ECG. La más común requiere que el paciente acuda periódicamente a revisiones con el médico. Además, el uso del Holter para el diagnóstico de la fibrilación auricular puede ser un tanto problemático, ya que si es de la clase paroxística los episodios pueden no presentarse en el periodo de grabación de éste, comúnmente 24 horas [1]. Así mismo cuando se hace un estudio de ECG en el hospital es aún menos probable que se presente un episodio ya que la ventana de estudio es considerablemente más pequeña. Lo que dificulta o impide su diagnóstico y seguimiento por parte de un médico. La figura 2 muestra un ECG de ritmo normal.



Fig. 2. ECG ritmo sinusal normal

IV. SOLUCIÓN PROPUESTA

La solución planteada es un sistema con monitoreo constante las 24 horas del día o durante días si es necesario. Para lograrlo, el monitoreo lo realizará un programa instalado en el teléfono celular (*detector cardiaco*), parte de una investigación más amplia, pero se pretende presentar una solución a la dificultad inherente a un monitoreo contante. Este programa grabará y enviará parte de la señal que sea de interés al médico mediante un correo electrónico. Además, el detector cardiaco permite observar el ritmo cardiaco y trazos de la señal cardiaca en tiempo real. El paciente realizará sus actividades cotidianas de forma normal mientras es monitoreado debido a que la comunicación del hardware de adquisición de datos (*monitor cardiaco*) y el teléfono celular es inalámbrica (vía Bluetooth).

Así también, el sistema provee un método automático de interacción con el médico por medio de correos electrónicos. Éste contiene una captura de la pantalla de la señal cuando se presente un episodio de FA cardiaca, además del ritmo cardiaco y datos personales del paciente.

Para describir este enfoque, presentamos en la sección V, la arquitectura general del sistema para llevar a cabo la solución.

IV. ARQUITECTURA GENERAL

En la figura 3 se muestra la arquitectura general de los elementos que componen al sistema de monitor cardiaco para teléfonos celulares.

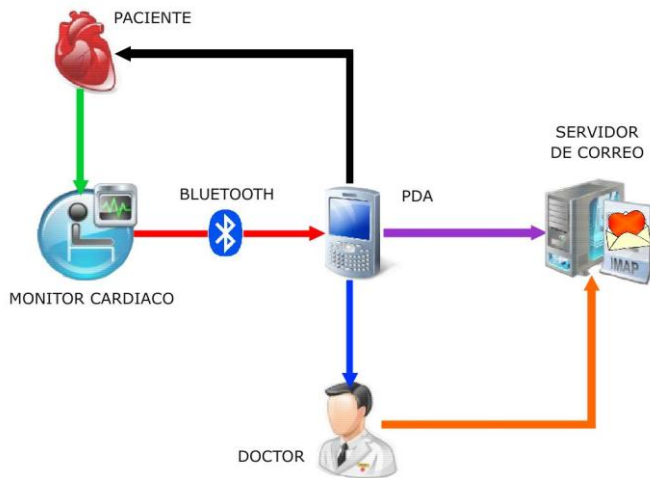


Fig. 3. Arquitectura general del monitor cardíaco móvil.

A continuación, explicaremos cada uno de los módulos que conforman al sistema.

El paciente es la parte central de este proyecto. El sistema está dirigido hacia personas que sufren pérdidas bruscas de conciencia (síncope) y palpitaciones. Por lo cual requieren monitorear el comportamiento de su corazón. De esta forma, se le facilitan al médico las tareas de diagnóstico y tratamiento para un paciente.

Adicionalmente, el paciente puede interactuar con el sistema registrando las actividades que realizó durante el día. De forma similar a cuando un usuario escribe un mensajes corto (SMS). Por ejemplo, si salió a correr, caminar, etc. En esta forma el médico puede tener un mejor panorama de las circunstancias en los cuales se presentaron los episodios de arritmia.

El monitor cardíaco es un método de utilidad diagnóstica basada en el registro de la actividad eléctrica cardíaca. El monitor cardíaco se encargará de recoger la actividad eléctrica cardíaca a través de una serie de cables conectados a la superficie corporal del individuo, una vez tomada esta señal se amplificará la pequeña diferencia de potencial que se ha producido en el músculo cardíaco.

Esta etapa del sistema es crucial, ya que se debe asegurar que la señal que sale del ECG (que es del orden de mV) llegue con una buena resolución al

teléfono celular para que nuestros algoritmos trabajen correctamente.

Esta etapa consistirá de dos módulos: el primero es integrar un transmisor en el ECG y el segundo sincronizar el transmisor con nuestro teléfono celular para iniciar la comunicación entre ambos dispositivos.

En el dispositivo móvil es donde estarán implementados los algoritmos en una aplicación móvil desarrollada en lenguaje Java, particularmente en la plataforma J2ME. Esta aplicación recibe la señal proveniente del ECG vía Bluetooth, después procesa la señal para quitarle ruido. Posteriormente la analiza en busca de patrones de FA [12]. Si encuentra un patrón crítico, envía automáticamente una alarma al usuario, así también se envía vía SMS y por correo electrónico al médico.

El sistema ofrece al médico dos posibilidades para analizar el electrocardiograma: la primera es cuando el paciente tenga cita con él y pueda observar en tiempo real el ECG; la segunda es revisando su correo electrónico y mensajes SMS recibidos en su teléfono celular. El SMS sólo indica que ha recibido un correo electrónico que contiene la imagen del ECG y los datos generales del paciente, así como las actividades realizadas antes del episodio de FA. El médico podrá entonces revisar el análisis en cuanto le sea posible.

El módulo SMS y de correo electrónico se encarga de enviar automáticamente un correo electrónico al médico con toda la información del mismo: hora, duración, datos relevantes del paciente, etc. Además de que se adjuntará en formato de imagen un episodio del ECG. Se utiliza la API de java mail (<http://java.sun.com/products/javamail/>) para la programación del envío automático de correos electrónicos. Mientras que para el envío automático de mensajes SMS se utilizó un Gateway que proporciona una API de Java (<http://www.aunix.com.mx/sms.htm>).

V. TRANSMISIÓN Y ADQUISICIÓN DE DATOS

El monitor cardiaco graba la actividad eléctrica del corazón por medio de una serie de cables conectados a la superficie del cuerpo del paciente. La señal es capturada y se amplifica la pequeña diferencia de potencial que ocurre en los músculos de corazón. El monitor cardiaco está diseñado de acuerdo a la figura 4.

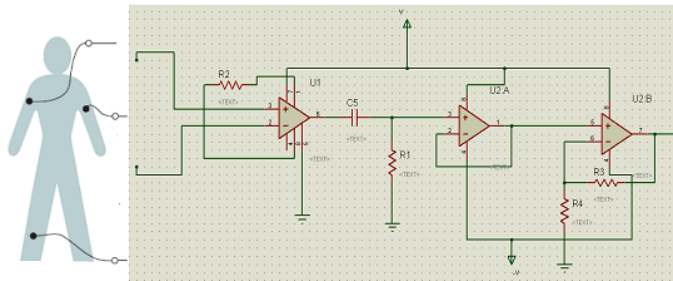


Fig. 4. Diseño de HeartMonitor

Como se muestra en la figura anterior las entradas del circuito AD620 son conectadas directamente al paciente. Durante las pruebas se observó, la señal cardiaca de la línea de alimentación, de 60 Hertz, además del ruido debido a los músculos que controlan la respiración [5], por lo tanto es necesario eliminar este ruido. Posteriormente, un filtro pasa-bajas con una frecuencia de corte de 100Hz y una pendiente de 40dB es diseñado. Los resultados de estas operaciones se muestran en la figura 5.

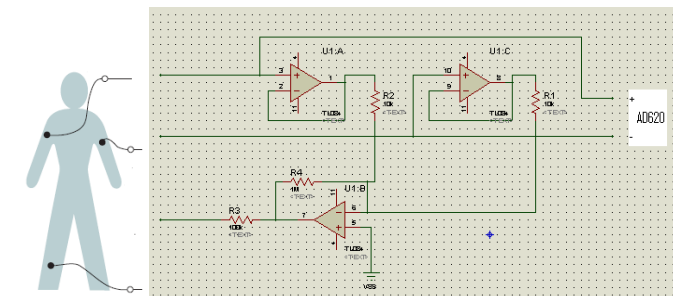


Fig. 5. Filtros usados en HeartMonitor

De esta manera, se eliminan las altas frecuencias del ruido causadas por fuentes musculares y electromagnéticas. Aún más una tierra virtual o referencia se establece en la pierna izquierda del paciente. Esto reduce los voltajes de modo común. La disminución en el voltaje causada por una corriente común filtrada por el paciente es obtenida reduciendo la impedancia del electrodo de referencia.

Además, la conversión analógica-digital se hace con un microcontrolador PIC16F876A, y su respectivo compilador en lenguaje C. Se tiene una resolución de 10 bits, con una tasa de transferencia de 7200 baudios. Esta tasa de transferencia es mayor que el mínimo necesario para asegurar que la información se transmita correctamente, debido a que la tasa de muestreo para digitalizar la señal es de 250 muestras por segundo. La figura 6 muestra el diagrama del circuito utilizado, que interactúa con el modem ESD-200.

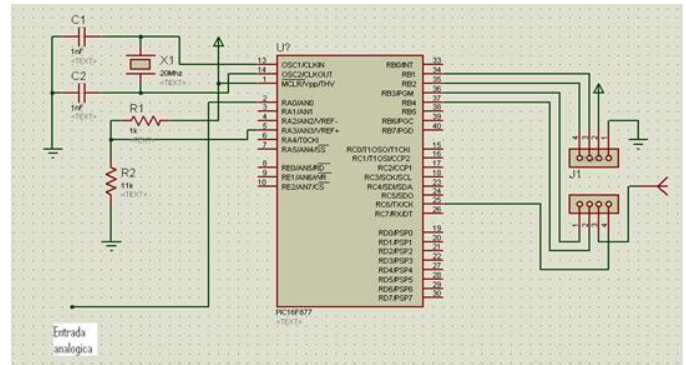


Fig. 6. Diseño del convertidor analógico-digital

A continuación en la siguiente sección se presentan los resultados preliminares obtenidos de nuestra propuesta.

VI. RESULTADOS PRELIMINARES

Para efectuar las pruebas se consideró el escenario que se muestra en la figura 7.

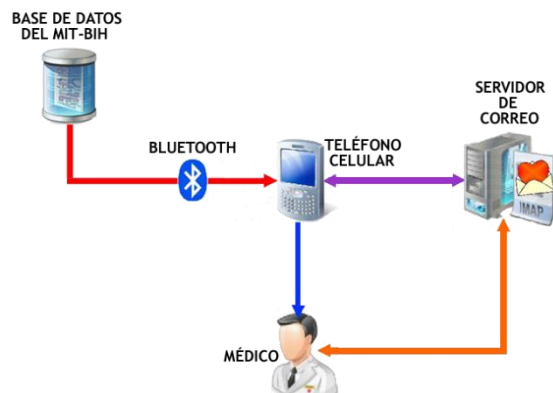


Fig. 7. Escenario de Pruebas

En el escenario de la figura 7 el paciente es sustituido por la base de datos MIT-BIH. La cual contiene datos de fibrilación auricular de acuerdo a los registros digitalizados de estudios de Holter de

pacientes reales con diversas arritmias. En este caso únicamente utilizaremos los datos referidos a FA. La razón para sustituir al paciente es porque es difícil precisar cuándo un paciente sufrirá un episodio de FA. Por ellos se utilizarán datos reales de estudios anteriores. Cabe mencionar que estos registros contienen los ruidos propios de un ECG, como el ruido de línea, el mioeléctrico y el de la respiración.

VII. RESULTADOS ESPERADOS

Se han efectuado pruebas preliminares en la parte del hardware con la instrumentación necesaria para obtener la señal cardiaca con la precisión requerida. Eliminando el ruido presente en la señal. El ECG se envía a través de un modem Bluetooth hacia el celular. Las especificaciones técnicas del teléfono celular son las de una HP iPAQ.

A continuación se muestran pantallas en el teléfono HP iPAQ utilizando el emulador de HP pues el proyecto sigue desarrollándose para perfeccionar el funcionamiento real. Mientras tanto, en la figura 8 se muestra un diagrama general de la navegación requerida en el sistema.

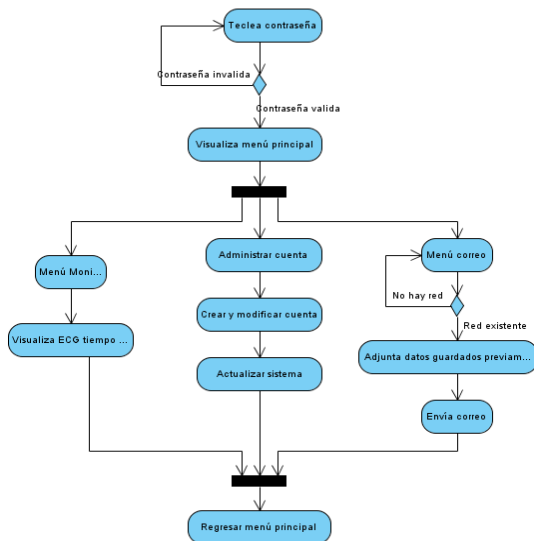


Fig. 8. Diagrama de navegación por el sistema

En la figura 9 como se puede apreciar se muestra la pantalla que permite iniciar sesión al usuario. Mediante la autenticación del mismo.



Fig. 9. Autenticación del usuario

Mientras que, en la figura 10 se visualiza el menú principal de la aplicación con tres opciones: 1) Monitor Cardíaco, 2) Anotaciones y administración de cuentas de correo y 3) Envío automático de SMS y de correos al médico.



Fig. 10. pantalla de menú principal

Entonces, de estas opciones cuando se selecciona la de monitor cardiaco, se visualiza un fragmento del ECG en tiempo real. Este resultado se despliega en la pantalla, la cual se muestra en la figura 11.



Fig. 11. Pantalla de ECG en tiempo real

La imagen mostrada en la figura 11 se obtiene debido al constante monitoreo del corazón. Entonces, cuando el algoritmo detecta una arritmia se visualizará la FA y se enviarán los datos vía correo o SMS al médico para el seguimiento del padecimiento. En el caso del envío de correo electrónico, cuando no existe disponibilidad de red, se almacenan los datos en la aplicación para su envío posterior. La figura 12 muestra la detección de FA.



Fig. 12. Detección de un episodio de FA

VIII. CONCLUSIONES

Se ha diseñado un prototipo capaz de enviar signos vitales de importancia a través de un celular con capacidad *bluetooth*, con el fin de alcanzar grandes beneficios de tiempo, movilidad, tamaño del sistema, y eliminación de cables.

Este proyecto está limitado a la detección de una única arritmia. El caso de estudio es la fibrilación auricular porque es el padecimiento más común y del cual se tienen mayores datos. Además de que es un padecimiento que hasta cierto punto permite al paciente realizar las actividades de su vida cotidiana, con cierta reserva y bajo un tratamiento y supervisión médica y tomando en cuenta que el algoritmo no se ha pensado para sustituir a las pruebas controladas de los hospitales. Debe tomarse en cuenta que este proyecto sigue en fase de desarrollo y perfeccionamiento. Como trabajo a futuro se considera incluir otros algoritmos, además de la FA, que permitan detectar otros tipos de arritmias, así como otros medios para transferir la información.

El sistema cuenta ya con la fase de diseño y análisis terminado. Los módulos de sistema tienen un funcionamiento adecuado de forma separada. Actualmente se trabaja en la integración de los mismos para realizar las pruebas finales, pues como ya se hizo mención anteriormente este trabajo está en las últimas fases de su desarrollo.

IX. AGRADECIMIENTOS

Los autores de este artículo agradecen al Instituto Politécnico Nacional, a la Secretaria de Investigación y Posgrado, a la Comisión de Operaciones y Fomento de Actividades Académicas (COFAA), al ICyTDF, al Instituto Nacional de Cardiología (INC) por su apoyo y financiamiento y especialmente al profesor Raul Memije por su valiosa asesoría.

X. REFERENCIAS

- [1] *Resistance Thermometry: Principles and Applications of Resistance Thermometers and Thermistors*, MINCO Application Aid #18, 861006-2, Minco 08/17/2000.
- [1] Guillermo Mora P., *Evaluación electrocardiográfica y con Holter de la fibrilación auricular: Guías de diagnóstico y tratamiento de la fibrilación auricular*, Vol. 14 Suplemento 3, ISSN 0120-5633
- [2] Rajendra Acharya U., *Advances in Cardiac Signal Processing*, Springer-Verlag, Berlin Heidelberg, 2007.
- [3] Ganong, William F, *Fisiología Médica*, Ed. El Manual Moderno, 20ª Edición, 2006
- [4] Fonseca L., Montiel J., Ramírez A, *Electrocardiógrafo de 12 derivaciones*, Ed. Lampone, 2004
- [5] Jean-Pierre Deschamps, Géry Jean Antoine Bioul, *Synthesis of Arithmetic Circuits: FPGA, ASIC and Embedded Systems*, John Wiley & Sons, 2006
- [6] <http://www.physionet.org/physiobank/database/mitdb/>, última revisión 23/04/09
- [7] Am J Cardiol, *Atrial fibrillation follow-up investigation of rhythm management -- the AFFIRM study design*,. 1997; 79: 1198-202
- [8] Bharucha D, Kowey P., *Management and prevention of atrial fibrillation after cardiovascular surgery*, 2000, pag. 85 (10 Suppl 1): 20-24
- [9] Fabio Guerrero, Oliver Cardona, *Análisis de interferencia entre las tecnología inalámbricas Bluetooth e IEEE 802.11g*, IEEE Transactions, Vol. 34, 2004
- [10] John G. Proakis, *Tratamiento digital de Señales*, Ed. Pearson, 4ª Edición, 2004
- [11] Black IW, Fatkin D, Sagar KB, Khandheria BK, Leung DY, Galloway JM et al. *Exclusion of atrial thrombus by transesophageal echocardiography does not preclude embolism after cardioversion of atrial fibrillation*. A Multicenter study circulation, 1994; 89: 2509-2513
- [12] Luis Díaz Baldevenito, Cesar Castro Rojas, *Desarrollo de un monitor de ritmo cardíaco*, Ingelectra IEEE, año 2006
- [13] John G. Webster, *Medical Instrumentation: Application and Design*, Ed. Wiley, 3a Edición
- [14] Ivaylo I. Christov, *Real time electrocardiogram QRS detection using combined adaptive threshold*, Wiley, 2002
- [15] K. Tateno L. Glass, *Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and ARR intervals*, IEEE Transactions, Vol. 23, 2001
- [16] Marc Gilbert Ginesta, Alvaro Peña Gonzales, *Ingeniería de software*, Formación de posgrado
- [17] Juan José González, *Bluetooth conceptos básicos y nuevas soluciones*, Revista Española de Electrónica, No. 16, Año 2005
- [18] spanish.bluetooth.com, última revisión 12/04/09
- [19] *Boylestad Nashelky*, Electrónica: teoría de los circuitos y dispositivos electrónicos, Ed. Pearson, 9ª Edición
- [20] Tompkins Carl, Shen Luo, *ECG beat detection using filter banks*, Biomedical Engineering, IEEE Transactions, Vol. 46, Issue 2, Feb. 1999

XI. AUTORES

Emmanuel Rojas es actualmente estudiante del decimo semestre de la carrera de Telemática en la UPIITA-IPN. Este artículo es parte de su tesis para obtener el título de ingeniero en telemática.

Cristian Morales es actualmente estudiante del decimo semestre de la carrera de Telemática en la UPIITA-IPN. Este artículo es parte de su tesis para obtener el título de ingeniero en telemática.

Alberto Juárez es actualmente estudiante del noveno semestre de la carrea de Telemática en la Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas (UPIITA-IPN). También, es alumno del programa PIFI del IPN.

Félix Mata, es doctor en ciencias de la computación y profesor investigador en la academia de telemática en la UPIITA-IPN. Actualmente, dirige diversos proyectos relacionados con el cómputo móvil, las telecomunicaciones y sistemas web orientados al conocimiento y los SIG.

DISEÑO E IMPLEMENTACIÓN DE UNA INTERFAZ DE REGISTRO DE DATOS EN TARJETAS DE MEMORIA FLASH SD CARD, PARA LA UNIDAD SÍSMICA SR04

Santiago Cruz Lauro, Hernández Arias Humberto, Hernández López Ana Angélica

Instituto de Ingeniería, Coordinación de Instrumentación,
Universidad Nacional Autónoma de México
Edificio 12, circuito exterior s/n, Ciudad
Universitaria, México D. F.
Tel. 5623 3600 ext. 8831; correo electrónico: lsc@pumas.iingen.unam.mx

RESUMEN

En este artículo se presenta el diseño e implementación de un sistema basado en un microcontrolador que permite el registro de los datos provenientes de la unidad sísmica SR04 en tarjetas de memoria flash *SD Card*. La información almacenada cumple con los sistemas de archivos FAT (File Allocation Table) de mayor uso en la actualidad: FAT16 y FAT32, para lograr compatibilidad con cualquier computadora que soporte estos sistemas de archivos, es decir, que los datos almacenados en las tarjetas SD (Secure Digital) Memory Card puedan ser leídos por una PC.

Palabras clave: memoria, sistema de archivos, SD Card.

I. INTRODUCCIÓN

Ha sido una necesidad humana histórica preservar información a lo largo del tiempo. Para conseguirlo se ha hecho uso de la tecnología presente en la época.

Un dispositivo capaz de retener información es un dispositivo de almacenamiento, le llamamos una memoria. En la actualidad no existe un dispositivo de almacenamiento universal, sino distintos tipos de memorias con sus ventajas y desventajas. Depende de la aplicación particular la elección de la más conveniente.

Las memorias pueden clasificarse de varias formas: por la tecnología que usan, por la

capacidad que tienen, si son o no removibles, etcétera. Si la memoria pierde su contenido en cuanto el sistema electrónico que la incluye se apaga, se le llama memoria volátil, de otro modo son memorias no volátiles, éstas últimas no requieren de alimentación de energía para mantener su contenido.

En el ámbito de las memorias no volátiles, se encuentran las memorias tipo Flash, tecnología dominante en aplicaciones que requieren grandes cantidades de almacenamiento no volátil de estado sólido, a un costo razonable. Esta tecnología es la que se usa en las “*memory cards*” o tarjetas de memoria. Algunas aplicaciones que utilizan tarjetas de memoria incluyen computadoras de escritorio, computadoras portátiles, teléfonos celulares, cámaras digitales, asistentes digitales personales (PDA, por sus siglas en inglés), reproductores de música, consolas de videojuegos, *dataloggers*, etcétera.

En la actualidad las tarjetas de memoria flash han ganado popularidad en aplicaciones portátiles. Su bajo consumo de energía, tamaño reducido, durabilidad, practicidad y bajo costo las hacen una opción muy atractiva para la electrónica de consumo actual.

Algunos formatos de tarjetas de memoria flash que se encuentran comúnmente son: MultiMedia Card (MMC), Secure Digital Card (SD Card), Memory Stick (MS) y Compact Flash (CF). Una de las más

populares y ampliamente usada en dispositivos muy diversos es la SD Card.

Secure Digital (SD), es un formato de tarjeta de memoria flash, creado en colaboración por *Panasonic Corporation*, *SanDisk Corporation* y *Toshiba Corporation*.

Algunas de las características con las que cuentan este tipo de memorias, ver figura 1, son las siguientes:

- Dedicada a aplicaciones portátiles y de escritorio.
- Gran capacidad de memoria:
Capacidad estándar: Hasta 2 GB.
Alta Capacidad: Desde 2 GB hasta los 32 GB.
- Margen de tensión:
Tarjeta de memoria SD de Alto Voltaje
– Tensión de operación: 2.7-3.6 V.
Tarjeta de memoria SD de tensión dual
– Tensión de operación: Nivel bajo de tensión y 2.7-3.6 V
- Tarjetas diseñadas para sólo lectura y de lectura / escritura.
- Modo predeterminado: tasa variable de reloj de 0 - 25 MHz, velocidad de interfaz de hasta 12,5 MB/s (con 4 líneas de datos en paralelo)
- Modo de alta velocidad: tasa variable de reloj de 0 - 50 MHz, velocidad de interfaz de hasta 25 MB/s (con 4 líneas de datos en paralelo).
- Característica de protección contra escritura usando interruptor mecánico.

Las memorias SD soportan tres modos de transferencia de datos.

1. Modo SD de un bit, en el que separa comandos, canales de datos y un formato propietario de transferencia
2. Modo SD de cuatro bits, que utiliza cuatro terminales para soportar transferencias paralelas.
3. Modo SPI: entrada separada serial y salida serial.

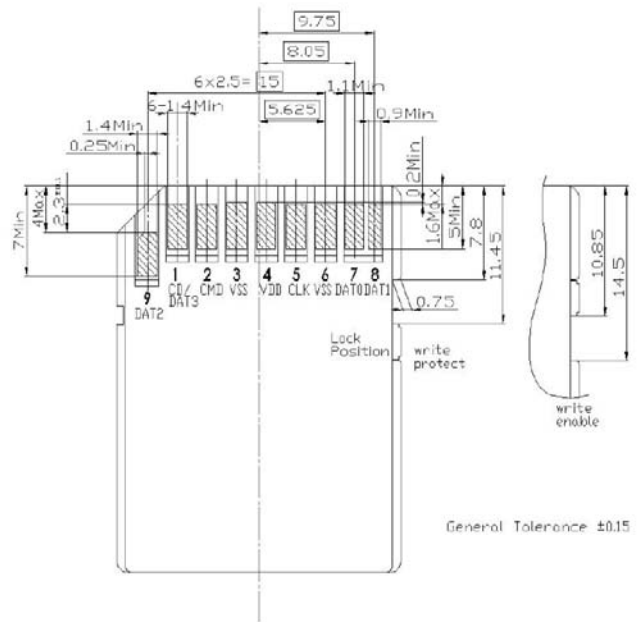


Fig. 1 Características físicas de las tarjetas SD.

El modo SPI es compatible con los puertos SPI de la mayoría de los microcontroladores, ver tabla 1. En este modo la tarjeta sólo permite la entrada de datos por *DATA IN* (2) y la salida de datos por la terminal *DATA OUT* (7). Los comandos se mandan también por la terminal *DATA IN*. La habilitación de la tarjeta se hace por la terminal *CS* (1). La señal de reloj, *CLK*, que se envía desde el *host* o controlador es la que establece la velocidad de la comunicación, es por esto que el diseño del *host* se facilita, ya que no se tienen problemas de desincronización cuando el *host* es más lento que la tarjeta. Sin embargo, como sólo se envían los datos por una terminal se tiene menor utilización del potencial en la velocidad de transferencia de datos.

Las tarjetas de memoria flash, así como los discos duros, los discos compactos y demás dispositivos de almacenamiento, nos proporcionan el medio físico donde se guarda la información. Requieren de un equipo especializado que grabe en ellos datos que luego puedan consultarse. La manera en que la información se graba en los dispositivos de almacenamiento depende de la tecnología que utilicen; sin embargo, es común a todos ellos la necesidad de organizar los datos que guardan para encontrarlos y acceder a ellos cuando se necesite.

<i>Terminal</i>	<i>Nombre</i>	<i>Tipo</i>	<i>Descripción SPI</i>
1	CS	Entrada	<i>Chip Select.</i>
2	DataIn	Entrada	Comandos y datos hacia la tarjeta.
3	VSS1	Alimentación	Tierra.
4	VDD	Alimentación	Fuente de voltaje.
5	CLK	Entrada	Señal de reloj.
6	VSS2	Alimentación	Tierra.
7	DataOut	Salida	Datos y estado desde la tarjeta.
8	RSV	-----	Reservado.
9	RSV	-----	Reservado.

Tabla 1 Descripción de la asignación de terminales de la tarjeta SD, en el modo SPI.

La forma de poner orden a los datos es a través de los sistemas de archivos. Se puede pensar un sistema de archivos como una base de datos de propósito específico, que manipula, organiza y le da una estructura a los datos almacenados, para que puedan ser leídos posteriormente por equipos que conozcan las reglas con las que se guardó la información. Entonces en el grabado de los datos queda implícita la implementación de un sistema de archivos.

Actualmente se han desarrollado muchos sistemas de archivos. La mayoría de los sistemas operativos utiliza su propio sistema de archivos y permiten compatibilidad con otros, además del propio. Algunos ejemplos de sistemas de archivos pueden ser el HFS Plus (Hierarquical File System, para Mac OS), XFS (Linux), FAT (File Allocation Table para MS-DOS) y NTFS (New Technology File System para Windows NT, XP y Vista). Cabe comentar que el sistema operativo Windows es el más usado en las computadoras y el sistema de archivos FAT el más común.

El sistema de archivos FAT tiene sus orígenes entre finales de 1970 y principios de 1980, y fue el sistema de archivos soportado por el sistema

operativo Microsoft ® MS-DOS ®. Originalmente fue desarrollado como un sistema de archivos simple, adecuado para las unidades de disco flexible con capacidad menor a 500 KB. Con el tiempo se ha mejorado para soportar volúmenes de mayor capacidad. Actualmente hay tres tipos de sistemas de archivos FAT: FAT12, FAT16 y FAT32. La diferencia básica en estos subtipos de FAT, y el motivo de los nombres, es el tamaño en bits de las entradas de la estructura real FAT en el disco. Hay 12 bits en una entrada FAT en FAT12, 16 bits en FAT16 y 32 bits en FAT32.

Los sistemas de archivos FAT de mayor uso en la actualidad son FAT16 y FAT32.

Un sistema de archivos FAT está compuesto por cuatro regiones básicas, ver figura 2, que se encuentran en el siguiente orden dentro del volumen:

0. **Región reservada.**
1. **Región FAT.**
2. **Región del directorio raíz (no existe para FAT32).**
3. **Región de datos, directorios y archivos.**

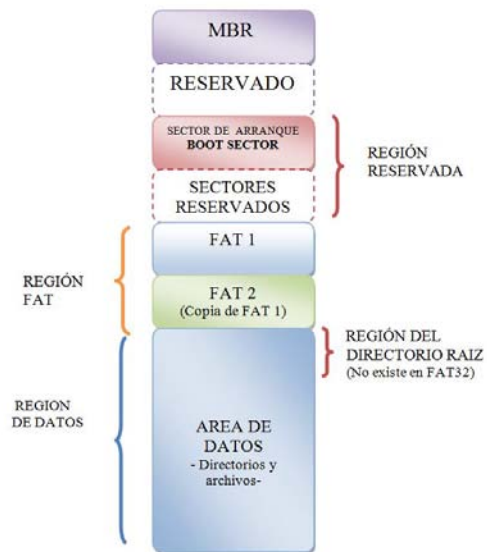


Fig. 2 Estructura típica de un volumen con formato FAT.

El primer sector físico del volumen es el llamado Sector Maestro de Arranque o MBR (del inglés *Master Boot Record*). Los primeros 446 bytes del

MBR contienen código que sirve para *arrancar* a la computadora. Después se encuentran 64 bytes de una tabla de partición, que es una serie de 4 registros de 16 bytes cada uno, conteniendo información sobre la división lógica del medio de almacenamiento.

La primera estructura en un volumen FAT es la llamada *Bloque de Parámetros del BIOS* (BPB: BIOS Parameter Block), la cual se encuentra en el primer sector lógico del volumen en la Región Reservada. Este sector es llamado a veces el Sector de Arranque (*Boot Sector*).

El BPB en el sector de arranque fue definido para MS-DOS 2.x y sólo permitía volúmenes FAT con estrictamente menos de 65,536 sectores (32 MB con sectores de 512 *bytes*). Esta limitación se debe a que el parámetro “Total de sectores”, definido en el BPB, era un campo de sólo 16 bits. Esta limitación fue superada por MS-DOS 3.x, donde el BPB fue modificado para incluir un nuevo campo de 32 bits para el número total de sectores.

El siguiente cambio hecho al BPB ocurrió con el sistema operativo Microsoft® Windows 95®, donde el sistema FAT32 fue introducido.

En la Región FAT se encuentra la estructura FAT, que está definida como una lista ligada de los clústers que componen a los distintos archivos. La FAT mapea la región de datos del volumen por los números de clúster, indicando el número del siguiente clúster de un archivo dado e indicando con una marca especial el fin de dicho archivo. El primer clúster del área de datos es el clúster 2.

Para medios con FAT12 y FAT16, el directorio raíz está localizado en un área fija dentro del disco, en la Región de directorio raíz y es de un tamaño fijo, indicado en uno de los campos del BPB. Para FAT32, el directorio raíz puede ser de tamaño variable y por lo regular se encuentra en el primer clúster del área de datos.

En un volumen FAT, un directorio es simplemente un archivo que indica en sus atributos que es un directorio, y está compuesto por una lista de los archivos que *contiene*. El único directorio especial

y que siempre debe estar presente es el directorio raíz.

El dispositivo que intente leer algún medio de almacenamiento debe saber cuál es el sistema de archivos con el que fue grabada la información.

El tipo de FAT que tiene un medio de almacenamiento se determina calculando la cantidad de clústers en el volumen, esto se realiza usando algunos campos de la estructura BPB del volumen.

En un volumen FAT12 se tienen 4,084 clústers como máximo, con FAT16 el número de clústers permitidos son 65,524. Un volumen FAT32 no puede tener menos de 65,524 clústers.

Las tarjetas de memoria flash que pueden ser integradas a equipos con necesidades de almacenamiento considerable, que requieran de portabilidad y practicidad, son ideales en equipos de adquisición como los *dataloggers*. Cualquier dispositivo electrónico que registre información durante un periodo de tiempo, para luego utilizarla, es considerado un *datalogger*. Esta definición incluye a muchos sistemas de adquisición de datos que usan una computadora como medio para almacenar datos en tiempo real.

En la actualidad, la mayoría de los fabricantes de *dataloggers* los consideran equipos autónomos, característica que lleva implícita el uso de memoria en el sitio para almacenar datos adquiridos por el dispositivo, para luego exportarlos a una computadora. Los primeros *dataloggers* usaban cinta magnética y papel perforado como medio de almacenamiento. Los *dataloggers* están incorporando nuevas tecnologías de almacenamiento y comunicación. La nueva tecnología permite el uso de memorias de estado sólido, algunos ejemplos son: SRAM, EEPROM y Flash, además, su uso hace posible reducir el tamaño de estos equipos, cualidad que los hace aún más atractivos para aplicaciones que requieran de traslados a localidades remotas. Usos típicos de los *dataloggers* incluyen estudios de campo, monitoreo de transportación, estudios de calidad,

investigación y, en general, donde se requiera registrar información de eventos que ocurran alrededor sin tener a alguien necesariamente presente.

El área de Ingeniería Sismológica del Instituto de Ingeniería de la UNAM cuenta con equipos de detección y adquisición de eventos, que son llevados a campo para hacer registro de datos usando computadoras portátiles. Uno de los equipos utilizados por los especialistas del Instituto es la unidad sísmica SR04. Esta unidad es propiamente un digitalizador de datos, con todos los componentes necesarios para convertir una señal sísmica analógica a su forma digital. La unidad consta de una caja metálica donde se integran tres sensores sísmicos y dos tarjetas electrónicas. Los sensores son geófonos y cada uno de éstos mide una de las tres componentes de movimiento (x, y, z). Una de las tarjetas es la SADC20, la cual lleva a cabo la conversión analógica a digital, mientras que la otra es la denominada GPSDCF, la cual consiste en un receptor GPS utilizado para establecer la referencia de tiempo exacto.

La tarjeta GPSDCF convierte las señales de tiempo de un receptor GPS en pulsos codificados de acuerdo con el estándar DCF77. La tarjeta SADC20 es una tarjeta de conversión analógica digital de 24 bits, diseñada para ofrecer altas prestaciones en aplicaciones sísmicas, con una de las mejores relaciones costo-calidad disponible en el mercado. Transmite información en paquetes de 9 *bytes* para los datos del tiempo absoluto y en paquetes de 5 *bytes* para los datos de cada uno de los canales. En estos paquetes, el primer *byte* es un identificador cuyo valor es mayor o igual a 128 decimal (0x80 hexadecimal) y sirve para distinguir si se trata de los datos de algún canal o los datos del tiempo. El último *byte* indica el fin de paquete y siempre tiene un valor mayor a 240 (0xF0). Los demás *bytes* del paquete son valores menores a 128 (0x80) y contienen los datos sin tratar. La SADC20 acepta comandos para obtener la versión del *firmware* y el contenido de la EEPROM con la que cuenta, o para establecer parámetros como el tiempo, la fecha, la frecuencia de muestreo, la

compensación del reloj y correcciones al tiempo GMT. Estos comandos están formados por 6 *bytes*, donde el primer *byte* especifica la operación a realizar, mientras que los *bytes* restantes son los parámetros propios de cada comando.

Un problema importante que presenta la unidad sísmica SR04 es el hecho que para coleccionar, guardar y procesar la información digitalizada siempre debe estar conectada a una computadora, personal o *lap top*. En aplicaciones en donde se requiere únicamente coleccionar y guardar dicha información, para largos periodos de muestreo, esto no sería posible sin la computadora. Dado lo anterior se propuso el desarrollar una unidad de colección y almacenamiento de datos, liberando así a la SR04 del uso de cualquier computadora. El registro de datos deberá hacerse en unidades de memoria con tecnología actual y con la característica de ser portátil, de preferencia *memory cards*. La información adquirida podrá ser leída directamente de la *memory card* o bien a través de un puerto USB de la unidad por desarrollar.

II. DESARROLLO

Hardware

En la figura 3 se presentan los elementos que integrarán la unidad de colección y almacenamiento de datos.

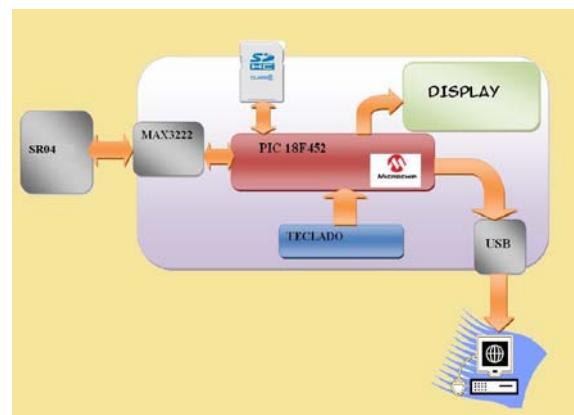


Fig. 3 Diagrama de bloques del sistema propuesto.

El núcleo del sistema es un microcontrolador Microchip® PIC18F452. Las principales

características de este microcontrolador se muestran en la Tabla 2.

CARACTERÍSTICAS PRINCIPALES PIC18F452	
Frecuencia de operación	DC- 40 MHz
Memoria de programa (bytes)	32 K
Memoria de programa (instrucciones)	16384
Memoria de datos eeprom (bytes)	1536
Puertos de entrada/salida	PUERTOS A, B, C, D, E
Temporizadores	4
Módulos de captura/comparación/pwm	2
Comunicación serial	MSSP, USART direccionable
Comunicación paralela	PSP
Módulo analogico-digital ,10 bits	8 Canales
Detección de bajo voltaje programable	SI
Set de instrucciones	75 Instrucciones
Empaques	DIP 40-pin PLCC 44-pin TQFP 44-pin

Tabla 2 Características principales del microcontrolador pic18F452.

Las funciones principales del microcontrolador en el sistema son: la inicialización de la tarjeta de memoria y el registro de datos en la misma, implementado los sistemas de archivo FAT16 ó FAT32; la adquisición de los datos provenientes de la unidad sísmica SR04, así como de la configuración de dicha unidad, decodificar un teclado, manejar un *display* gráfico y comunicación con una PC mediante un puerto USB. Se eligió este microcontrolador por varias razones, entre ellas el que es un dispositivo de fácil adquisición, con herramientas de desarrollo de bajo costo, varios puertos de comunicación, además de que su documentación es amplia, incluyendo ejemplos y notas de aplicación.

Una vez seleccionado el microcontrolador, se construyó el conector para comunicar a la tarjeta SD y el PIC18F452, ver figura 4. Cabe mencionar, que las tarjetas SD trabajan con un voltaje de entre 2.7 y 3.3 V, por lo que todo el sistema será alimentado con 3.3 V, haciendo uso de un regulador de voltaje *LF33CV*, con la configuración recomendada por el fabricante.

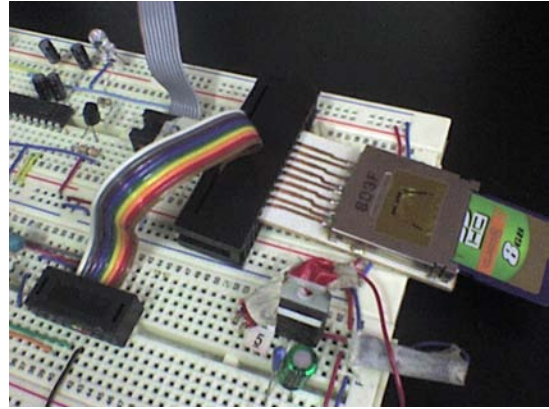


Fig. 4 Detalle del conector para las tarjetas SD.

Para realizar la comunicación entre el microcontrolador y la unidad SR04, mediante el estándar RS -232, se requiere el ajuste de los niveles de tensión, para lo cual se utilizó el circuito integrado *MAX 3222*, que es un dispositivo de bajo consumo, trabajando a 3.3 V.

Como una forma de depuración, se han colocado varios *Leds*, que indicarán un error, o el fin de un proceso, dependiendo de cual encienda.

Entre los elementos que se planea integrar en la unidad de colección y almacenamiento de datos, se encuentran los dispositivos periféricos, que son el display y el teclado. Mediante el uso del teclado y el apoyo del display gráfico, se configurarían en la unidad sísmica SR04 la frecuencia a la que muestreará las señales de entrada. La frecuencia de muestreo se puede ajustar a 10, 20, 25, 50, 100 y 200 muestras por segundo.

Mediante el display gráfico se presentarán los datos de los canales de la unidad SR04, con lo que se podrá saber si la unidad está arrojando datos útiles, y por lo tanto se podrá dejar al sistema guardando los datos en las memorias SD, ver figura 5.

En esta etapa del desarrollo se tiene un módulo adicional, en el que se tiene un display graficando las señales de los tres canales, aunque aún no se integra al resto de la unidad de registro. Así mismo, el teclado todavía se encuentra en la etapa de desarrollo y pruebas.

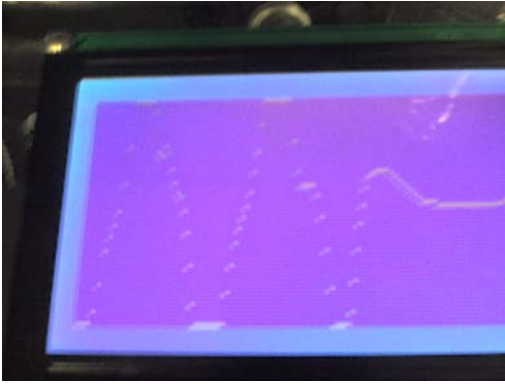


Fig. 5 Gráfica de un canal de la unidad SR04.

Software

El proceso de registro de la información adquirida en las tarjetas SD requiere de las siguientes etapas:

- Inicialización de la tarjeta en el modo SPI.
- Determinación del tipo de FAT en la tarjeta SD.
- Comunicación con la unidad sísmica, con el fin de configurarla.
- Creación del archivo donde se registrarán los datos obtenidos de la unidad sísmica, en las tarjetas de memoria SD, implementando el sistema de archivos correspondiente.
- Registro de los datos provenientes de la unidad sísmica.

En la primera etapa, al energizar el sistema, la tarjeta enciende en el modo nativo SD de un bit. El programa en el microcontrolador deberá entonces enviar a la tarjeta una secuencia de comandos específicos para establecer el modo de transferencia SPI y así poder interactuar de manera correcta. En la figura 6 se observa un diagrama con la secuencia de inicialización de la tarjeta en el modo SPI.

Una vez realizada correctamente la inicialización de la memoria SD, el microcontrolador tiene información sobre la versión de la tarjeta que está manejando, es decir, si se trata de la versión 1.x, que corresponde a las tarjetas de capacidad estándar, o la versión 2.0 o posterior, o si se trata de una tarjeta que no es SD, en cuyo caso enviará un mensaje de error y detendrá el proceso de registro de datos.

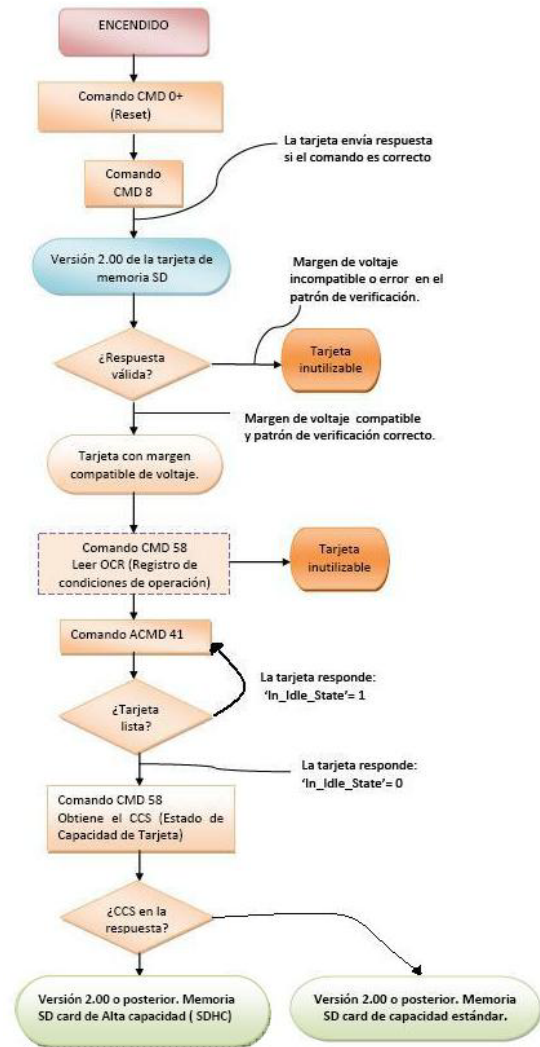


Fig. 6 Secuencia de inicialización en modo SPI.

Inicializada la memoria y lista para trabajar en el modo SPI, se lee el sector de arranque, para conocer los campos de la estructura BPB (Bloque de Parámetros del BIOS), reteniendo en la RAM del microcontrolador esta información. A partir de los campos del BPB, se obtiene el número total de clústers, con lo que se determina el tipo de sistema de archivos FAT de la tarjeta SD.

En la siguiente etapa se establece la comunicación por medio del estándar RS-232 entre la unidad sísmica SR04 y el microcontrolador. Ésta se realiza a una velocidad de transmisión de 38 400 baudios, 8 bits de datos, 1 bit de paro y sin paridad, utilizando el puerto de comunicaciones USART (*Universal Synchronous Asynchronous Receiver Transmitter*) del microcontrolador.

Como se mencionó anteriormente, se utilizará el teclado para establecer la frecuencia de muestreo.

Si no se le envía ningún comando, la unidad SR04 comienza a enviar los datos digitalizados de los sensores, con la última configuración establecida.

Para registrar la información en la memoria SD y posteriormente procesarla en una computadora que soporte los sistemas de archivos FAT, el programa en el microcontrolador graba en la tarjeta la información correspondiente al archivo que se va a generar, como el nombre, la extensión y el tamaño en bytes.

Se empieza por generar y grabar una entrada en el área del directorio raíz, donde se especifica el nombre del archivo, su tamaño en bytes, atributos y el número del primer *clúster* del archivo dentro del medio de almacenamiento.

En el área FAT se generan las entradas, que corresponderán a los clústers que conforman a todo el archivo y que informarán al sistema operativo en donde se encuentra físicamente el archivo dentro de la tarjeta de memoria SD. Considerando que la unidad mínima de asignación es un clúster, se hacen cálculos para conocer el número de entradas correspondientes a este archivo. Lo anterior se realiza, de nuevo, tomando en cuenta los campos leídos del *BPB* de la tarjeta SD. Estos campos son: el número de bytes por sector y el número de sectores por clúster.

Para que un sistema operativo pueda leer correctamente un archivo, primero lee la entrada correspondiente en el directorio raíz, la cual le indica el número del primer clúster que compone al archivo dentro del volumen. Con el número del primer clúster del archivo se entra en la Tabla de Asignación de Archivos (FAT), en donde encontrará una lista con el resto de los números de clúster que componen al archivo. Teniendo todo lo anterior, el microcontrolador comenzará a adquirir los datos enviados por la unidad SR04, enviándolos a su vez a la tarjeta SD que los grabará en el área asignada.

III. PRUEBAS Y RESULTADOS

Las primeras pruebas del conjunto hardware-software desarrollado se realizaron escribiendo una

cadena de caracteres, generados con el programa del microcontrolador, con lo cual se generaron archivos de texto que pudieron ser vistos en un editor de textos en una computadora.

Para poder grabar datos de la unidad SR04 se tuvo que hacer un ajuste en el envío de información hacia la tarjeta SD, debido a que sólo acepta el grabado por bloques de 512 bytes, según las especificaciones generales para este tipo de tarjetas.

Al tener que grabar por bloques, el programa del microcontrolador conforma un buffer de 512 bytes que envía a la tarjeta, y tiene que esperar a que la tarjeta grabe internamente ese bloque, para poder enviarle un nuevo bloque, lo que significaría pérdida de información. Este problema fue solucionado enviando byte a byte la información generada por la unidad SR04, evitando con ello el uso de un buffer y ahorrando tiempo. Con lo anterior, se hicieron varias pruebas de registro, grabando archivos de sólo 512 bytes. Para comprobar la información grabada en el archivo, ésta fue procesada en una PC, por medio de una aplicación hecha en Labview®, desarrollada en el mismo Instituto de Ingeniería. Con este procesamiento se extrajeron los encabezados de tiempo, segundo a segundo, además se graficaron los 3 canales registrados, comprobando que la información fue registrada de manera correcta y sin pérdidas, como puede observarse en las figuras 7 y 8.

Una vez que se determinó que el archivo fue grabado correctamente, con información real, se incremento el tiempo de registro, lo que implica un archivo de mayor tamaño. Las pruebas realizadas, fueron de 10 minutos, 30 minutos, una hora, 6 horas, 12 horas, 24 horas y 48 horas. A los archivos grabados en las tarjetas, producto de estas pruebas, se les aplicó el procesamiento mencionado, para obtener encabezados y gráficas, comprobándose una vez más, que dichos archivos fueron grabados de manera correcta.

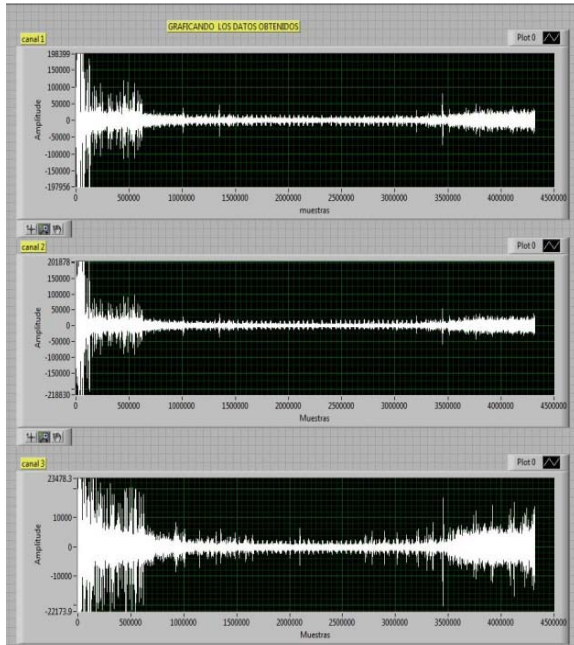


Fig. 7 Gráficas generadas a partir de la información registrada en la memoria SD.

Ruta del Archivo	Tamaño (en bytes)	Encabezados de Tiempo
C:\Users\PHernandez\A\	65188800	//año //mes //dia //SEGUNDO/segundo //hora
1016 00F9 8402 1400 F982 5A49 00F9 820C 1600 F884 5706		129 9 4 24 43 12 1
00F9 8278 4500 F983 6814 00F9 844C 757F F982 1A25 00F8		129 9 4 24 44 12 1
8385 1A07 F984 5267 77F8 827C 00F8 F883 0A27 00F9 840F		129 9 4 24 45 12 1
6A7F F982 6077 77F8 8350 1800 F984 0872 77F8 837A 6C7F		129 9 4 24 47 12 1
F983 0108 00F9 8417 0000 F982 1D65 77F8 827A 727F F884		129 9 4 24 48 12 1
5108 00F8 8205 647F F882 586A 77F8 844C 0000 F882 1764		129 9 4 24 49 12 1
77F8 833C 6D7F F884 7800 00F9 8226 767F F883 7877 77F8		129 9 4 24 50 12 1
847E 0800 F982 4113 00F9 8325 0800 F884 2115 00F9 8225		129 9 4 24 50 12 1
2100 F883 644C 00F8 8443 2000 F982 9022 00F8 8355 1100		129 9 4 24 51 12 1
F884 3F18 00F9 8232 2200 F883 430E 00F8 840A 7C7F F882		129 9 4 24 51 12 1
0A13 00F9 8376 77F8 F884 6C37 77F8 822E 707F F883 1674		129 9 4 24 52 12 1
77F8 8451 4A7F F882 1654 77F8 833D 787F F884 0056 77F8		129 9 4 24 53 12 1
8222 087F F883 0A7D 77F8 8416 747F F982 F84D 77F8 825D		129 9 4 24 53 12 1
947F F984 2319 00F9 8208 683F F882 4162 77F8 F884 2F90		129 9 4 24 54 12 1
F882 6577 711F 8347 727F F184 702D 00F9 8226 0600 F883		129 9 4 24 55 12 1
7310 00F9 8479 1100 F982 7A0D 00F9 835E 3000 F884 0976		129 9 4 24 55 12 1
77F8 8244 0A00 F883 7847 00F8 8463 607F F882 0E7A 77F8		129 9 4 24 56 12 1
8348 4200 F884 7361 77F8 8298 0800 F883 9029 00F9 8438		129 9 4 24 57 12 1
6C7F F882 201A 00F8 8378 77F8 F984 9A75 77F8 8298 0C00		129 9 4 24 57 12 1
F983 2F4A 77F8 841F 7A7F F882 5864 77F8 8304 487F F884		129 9 4 24 58 12 1
7A02 00F9 8248 277F F882 3864 77F8 845C 0000 F882 7C32		129 9 4 24 58 12 1
77F8 8245 727F F884 4E17 00F8 824C 667F F882 786A 77F8		129 9 4 24 59 12 1
8420 1200 F882 6A24 00F9 825C 627F F884 7806 00F9 8266		129 9 4 24 59 12 1
4000 F883 5901 F882 403C 00F8 8222 3A00 F884 2877 77F8		129 9 4 24 0 13 1
8202 2A00 F883 4159 00F9 8434 787F 77F8 7819 00F9 8321		129 9 4 24 1 13 1
3800 F884 4075 77F8 823E 0000 F883 1D69 77F8 8419 737F		129 9 4 24 1 13 1
F882 0775 77F8 8340 177F F884 7876 F882 77F8 8275 537F F883		129 9 4 24 2 13 1
217E 77F8 843F 77F8 F881 090A 182C 0800 00F8 825E 427F		129 9 4 24 3 13 1
F882 390C 77F8 841D 1100 F882 2349 77F8 8241 227F F884		129 9 4 24 4 13 1
7810 00F8 8250 617F F882 245F 77F8 845C 0C00 F882 8224		129 9 4 24 4 13 1
00F9 831C 0800 F884 3C00 00F8 8220 4700 F883 3925 00F8		129 9 4 24 5 13 1
8421 0000 F982 235F 00F8 8358 5200 F984 4473 77F8 822E		129 9 4 24 5 13 1
5F00 F883 4A50 00F8 8438 7A7F F882 5043 00F8 831C 4A00		129 9 4 24 6 13 1
F884 4F71 77F8 823C 1200 F883 3618 00F8 8477 717F F882		129 9 4 24 6 13 1
00A6 77F8 8352 607F F884 2878 77F8 8286 487F F883 4957		129 9 4 24 7 13 1
77F8 8449 0800 F882 773E 77F8 8236 597F F884 380A 00F9		129 9 4 24 8 13 1
8262 767F F883 456E 77F8 844C 0700 F882 2000 00F8 8331		129 9 4 24 8 13 1
0F00 F884 527A 77F8 822A 2C00 F883 112A 00F9 8412 717F		129 9 4 24 8 13 1
F882 1148 00F8 834A 3900 F984 2573 77F8 8276 4800 F883		129 9 4 24 9 13 1
73A2 00F8 8437 727F F882 5228 00F8 8256 2C00 F884 2209		129 9 4 24 9 13 1
00F8 8348 1900 F883 0028 00F8 842F AC00 F882 4A77 77F8		129 9 4 24 10 13 1
8379 0800 F884 7A09 00F8 820C 4D7F F883 3A67 77F8 848F		129 9 4 24 11 13 1
0000 F882 2832 77F8 8363 537F F884 4A00 00F8 8222 337F		129 9 4 24 11 13 1
F882 765D 77F8 8468 0100 F882 4C47 77F8 8259 777F F884		129 9 4 24 12 13 1
0877 77F8 8202 747F F883 3C1A 00F9 8409 607F F882 6828		129 9 4 24 12 13 1

Fig. 8 Marcas de tiempo extraídas del archivo grabado.

Se han hecho pruebas con las cuales se ha podido registrar un archivo con un tamaño máximo de 248MB. Lo anterior implica 48 horas de registro continuo, a una tasa de muestreo de 100 mps.

IV. RESULTADOS Y CONCLUSIONES

En este punto del desarrollo del sistema se cuenta con una interfaz que se comunica con la unidad sísmica SR04 y graba datos en un archivo de tamaño determinado, en una tarjeta de memoria SD. El archivo grabado es leído en una PC y mediante un procesamiento se puede obtener la información que le es útil a los expertos en el área de sismología.

El teclado y el display están en etapa de integración al módulo base, lo que nos permitirá agregar las funciones de configuración de la unidad sísmica, usando un menú en pantalla. En la figura 9 se muestra el módulo con el que se está desarrollando la etapa del display gráfico. En la figura 10 se observa el sistema que se tiene actualmente funcionando, registrando los datos de la unidad sísmica SR04.

Una vez realizadas las tareas descritas procederemos a la realización de un primer prototipo en circuito impreso y con los diferentes elementos que los componen en un gabinete de uso industrial.

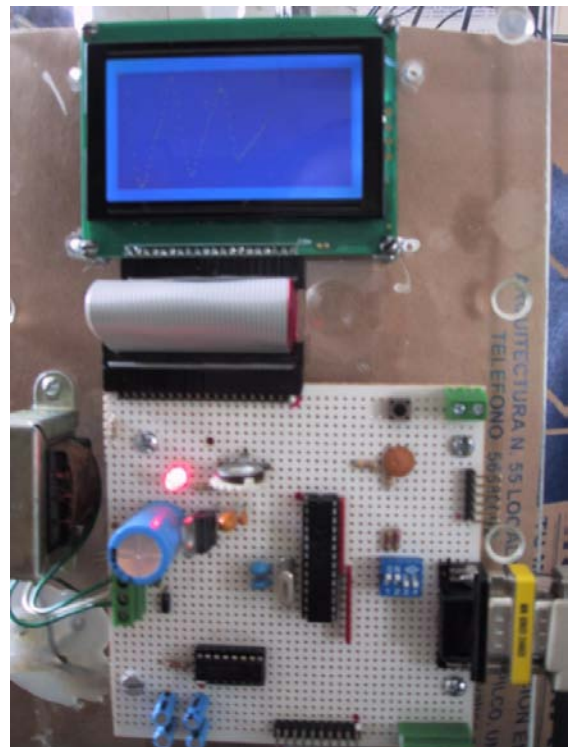


Fig. 9 Módulo del display.



Fig. 10 El sistema en la etapa actual.

V. REFERENCIAS

- [1] *SD Specifications, Part 1 Physical Layer Simplified Specification Version 2.00* September 25, 2006, SD Group.
- [2] *8-bit PIC® Microcontroller (PIC18F452) datasheet.*
<http://ww1.microchip.com/downloads/en/DeviceDoc/39564c.pdf>
- [3] *Microsoft Extensible Firmware Initiative FAT32 File System Specification FAT: General Overview of On-Disk Format.* Version 1.03, December 6, 2000 Microsoft Corporation:
- [4] *SADC10/18/20/30-Communication protocol.* Document revision 23rd March 2004 SARA di Mariotti Gabriele & C snc – Perugia
- [5] *Understanding the FAT32 File system.*
<http://www.pjrc.com/tech/8051/ide/fat32.html>
- [6] *Very low drop voltage regulators with inhibit (LF33CV) datasheet.*
<http://www.datasheetcatalog.org/datasheet/tmicroelectronics/2574.pdf>
- [7] *3.0V to 5.5V, Low-Power, Up to 1Mbps, True RS-232 Transceivers Using Four 0.1µF External Capacitors (MAX3222) datasheet.*

<http://www.datasheetcatalog.org/datasheet/maxim/MAX3222-MAX3241.pdf>

- [8] Kernigham Brian W., Ritchie Dennis M., *El lenguaje de programación C*, segunda edición, Pearson Prentice Hall, Mexico 1991.
- [9] Lázaro Antoni Manuel, Del Rio Joaquin, *LabVIEW 7.1 Programación Grafica para el control de instrumentación*, Thomson, 2005.
- [10] Palacios Enrique, Remiro Fernando, Lopez Lucas, *Microcontrolador PIC16F84 Desarrollo de proyectos*, primera edición, Alfaomega, 2004.
- [11] *MPLAB C18 C Compiler Libraries*, Microchip Technology Inc. 2005.

VI. AUTORES

M. en I. Lauro Santiago Cruz. Ingeniero Mecánico Electricista, egresado de la Universidad Nacional Autónoma de México en 1982. Obtuvo el grado de Maestría en Electrónica en 1986.

Humberto Hernández Arias. Tesista de licenciatura de la carrera de Ingeniero Eléctrico-Electrónico, Módulo de Electrónica. UNAM, Facultad de Ingeniería

Ana Angélica Hernández López. Tesista de licenciatura de la carrera de Ingeniero Eléctrico-Electrónico, Módulo de Electrónica. UNAM, Facultad de Ingeniería.

SISTEMA DE SEGURIDAD PERIMETRAL ÓPTICO “SSPO”

Herrera Galicia Rubén, Becerra Fuentes Olga Mariana, Zepeda Hernández Maria de los Angeles, Moreno Pérez Roque Aarón, Franco Pérez Fernando Francisco.

Área de Control y Automatización, Departamento de Eléctrica y Electrónica,
Instituto Tecnológico de Tuxtla Gutiérrez
Carretera Panamericana Km. 1080, Col. Terán, Tuxtla Gutiérrez, Chiapas, 29070, México
TEL. 6150380, correo electrónico: h_galicia24@hotmail.com, m2m_35@hotmail.com

RESUMEN

Lo que aquí se propone como proyecto es un sistema adecuado a las necesidades específicas de resguardo perimetral de una granja piloto experimental. El proyecto consiste en combinar dispositivos ópticos, comunicación RF, alimentación de circuitos con baterías recargables, celdas solares, y comunicación USB para crear una herramienta útil para el resguardo perimetral de áreas de varias hectáreas en exteriores. El sistema aquí propuesto no requiere el uso de cerca de púas o cerca de malla ciclón. Sin embargo su combinación con algún tipo de cerca mejora la eficacia del sistema.

Palabras clave: Láser, comunicación RF, Microcontroladores, comunicación USB.

I. INTRODUCCIÓN

Debido a los efectos negativos que el cambio climático mundial esta generando, así como a la inestabilidad del mercado petrolero: distintos gobiernos han implementado acciones encaminadas a la promoción del uso de energías que no alteren el equilibrio ecológico y beneficien a las economías locales.

Estas políticas han sido aplicadas, desde 2007 a la fecha. Se ha incursionado en la obtención de combustibles a partir de semillas de Piñón e Higuierilla. Sin embargo, para satisfacer las necesidades de combustibles del mercado es imprescindible incrementar el volumen de

producción, razón por la cual se está experimentando en las especies bioenergéticas de tercera generación, los microorganismos. La obtención de etanol a partir de algas está entre los más promisorios. Actualmente, año 2009, a nivel internacional esta tecnología se encuentra en etapa de estudio y experimentación, previo a su escalamiento a nivel industrial.

El cultivo de microorganismos se hace en “Granjas Piloto Experimentales de Microalgas”. El perímetro de las granjas debe mantenerse vigilado ya que; si entran personas ajenas al personal de la granja, el material biológico que ahí se maneja puede contaminar los cuerpos de agua externos corriendo el riesgo de generar un fuerte impacto ambiental. A su vez, los estanques de almacenamiento de los microorganismos deben estar aislados de elementos externos, tales como las bacterias, ya que estos pueden alterar el proceso de producción.

También, la incursión en el uso de especies bioenergéticas de tercera generación constituye una innovación científica y tecnológica a nivel mundial, de modo que los resultados de las investigaciones que en estas instalaciones se alcancen no pueden estar expuestos a plagio.

El área recomendable para una granja piloto de microalgas desde el punto de vista de su óptima administración cae en el rango de 5 a 10 hectáreas de superficie productiva. El proyecto que aquí se propone es un sistema diseñado en base a las necesidades específicas de resguardo perimetral

de Granjas de microalgas. Pero es adaptable para resguardar gasoductos o subestaciones eléctricas.

Existen en el mundo varias empresas que comercializan sistemas de seguridad perimetral, las más significativas de ellas se presentan a continuación:

Beyton ofrece protección de bardas para detectar una intrusión por medio de luz láser y fotoceldas, las cuales detectan la interrupción del rayo. Esto envía una señal al sistema de seguridad y acciona una sirena avisando el problema. Además ubica el tramo de barda por el que ocurrió la intrusión [7].

Lazer Tripwire basa la detección en la reflexión de un haz láser. El equipo está formado por un cilindro con dos diodos láser orientados en direcciones opuestas que giran a una revolución por segundo obteniendo un área circular de 150 metros de radio. Al lado de cada emisor láser existe un receptor que notará el haz láser reflejado por algún objeto que se encuentre dentro del área [8].

Rasnel Laser vende barreras de seguridad basadas en tecnología láser. El sistema está compuesto por un transmisor láser de luz roja modulada y de un receptor modulado. La máxima distancia de alcance es 60m. Además muestra insensibilidad a la luz solar y ambiental [9].

Sallén ofrece un sistema de seguridad perimetral por fibra óptica; llamado Leopard. Este sistema cambia su respuesta en presencia de movimiento de la fibra óptica, la cual se encuentra montada en una cerca de malla ciclón. La luz emitida por un diodo láser es conducida a través de un cable de fibra óptica, multimodo, de longitud 230 metros. Posteriormente esa luz es captada por un sensor optoelectrónico. Esta señal es analizada espectralmente mediante un microcontrolador, obteniéndose la información deseada; existencia o no de intrusión. El analizador del sistema Leopard se conecta al centro de control del operador siendo posible incorporar sistema de luces, alarma sonora, y activación de un CCTV [10].

Denial System comercializa un sistema basado en tres haces de luz láser colocados a diferentes alturas, cada uno emite una luz azul que al ser interrumpida activa una alarma. Los tres haces no están constantemente encendidos, se van alternando en secuencia. Cada secuencia se reproduce a una velocidad alta. Las secuencias no tienen un orden preestablecido, son generadas de forma aleatoria por un hardware para dicho fin [11].

El equipo E-960-D390Q es un detector infrarrojo de largo alcance con alineación por medio de rayo láser y diseñado para proveer seguridad perimetral. Método de detección: rotura simultanea de 2 rayos para seguridad perimetral confiable, minimizando falsas alarmas causadas por la caída de hojas o aves. Lentes ópticos refuerzan el poder del rayo y prevén excelente inmunidad a falsas alarmas causadas por lluvia o neblina. Máximo alcance exterior: 30, 60, 90, 120m [12].

El sistema DEM-2484 permite el montaje de hasta 4 haces de infrarrojo en una columna. La alarma se activa por corte de los 4 haces. Un microprocesador incluido ajusta los niveles de detección debido a lluvia, con lo cual se evitan falsas alarmas. La alineación se hace mediante guía láser. Alcance en exteriores 200 metros [13].

El proyecto aquí presentado consiste de un sistema de resguardo de un perímetro de $n \times m$ dividido en h sectores, con un máximo de 20 metros de alcance lineal por sector. Para detectar y señalar la presencia de intrusos se emplea: comunicación óptica vía pares emisor-receptor con luz láser, comunicación inalámbrica de RF, comunicación USB con una PC, y un sistema de visualización con lámparas. La PC es usada como unidad de supervisión. El sistema se caracteriza por ser elástico en su diseño y factible de escalamiento permitiendo adaptaciones de nuevos periféricos según las necesidades extras que se presenten; alarma sonora, alerta vía internet o vía comunicación celular. El sistema es una herramienta de apoyo al personal de seguridad de las granjas.

Cuando una granja de cualquier tipo se encuentra protegida solo por un cerco con alambre de púas se sufre de robo furtivo por parte de las personas que habitan los alrededores, debido a la falta de un sistema que impida la entrada a personas ajenas a la granja.

La propagación de las microalgas fuera de las granjas constituye un riesgo ambiental serio; además, el entorno en el que se encuentran debe estar controlado; de manera que se pueda evitar el ingreso o salida de material biológico, situaciones que pueden tener lugar de forma involuntaria por personas que desconozcan las medidas de seguridad biológica o por animales de la zona.

Además, debido a que el estudio de las microalgas como especies productoras de combustibles es una innovación científica y tecnológica a nivel mundial, las investigaciones realizadas y los equipos diseñados e implementados en estas granjas tienen gran valor.

Todas estas son razones por las que es necesario el resguardo del área con un sistema que ofrezca seguridad. El presente proyecto representa una solución a la necesidad de resguardar a las “Granjas Piloto experimentales de Microalgas” con un sistema que apoye al personal de seguridad de la granja para mantener el control biológico, instrumental e intelectual del proyecto. También el proyecto puede ser aplicado a necesidades similares donde se requiera resguardo perimetral y la edificación de una barda perimetral sea costosa, inoperante o insuficiente. Ejemplo: instalaciones de PEMEX o CFE.

II. DESCRIPCIÓN DEL PROTOTIPO

Para su descripción el sistema se puede dividir en cuatro subsistemas identificados a partir de su función: elementos sensores, transmisión RF, recepción RF, y control.

1). *Subsistema de elementos sensores.* Consta de dos partes: emisión y recepción. Para la emisión se usa un circuito con varios diodos láser ubicados

en forma tal que formen un cerco óptico. La etapa de recepción se compone de circuitos receptores cada uno conformado por un arreglo de siete fototransistores modelo genérico, un amplificador seguidor inversor y un comparador, Fig. 1.

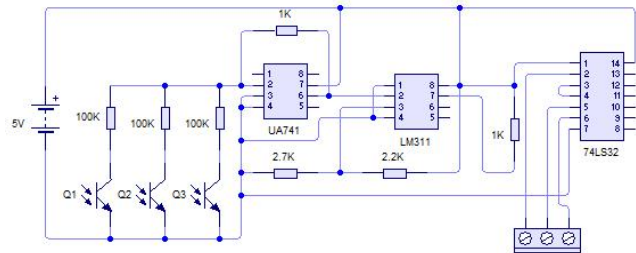


Fig.1 Circuito de recepción LASER.

Con el arreglo de fototransistores se logra una mayor sensibilidad y menor dificultad en el ajuste del rayo de luz láser. Con el comparador se logra la adecuación de la señal de detección a niveles TTL. También se usa una compuerta AND para conectar a todos los receptores de un sector y con la salida de la AND controlar al circuito transmisor. El corte de cualquier rayo de luz láser provoca un cero a la salida de la AND y con eso se inicia la transmisión de paquetes señalizando la presencia de un intruso e indicando el número de sector.

Los módulos de transmisión óptica TO emiten los haces de luz y los módulos de recepción óptica RO están colocados de modo que forman un cerco óptico, Fig. 2. Para que la detección sea efectiva, cada sector del cerco óptico se compone de varios haces de luz láser distribuidos en forma paralela, de manera que la distancia de la superficie del suelo hasta el primer haz es de 20 cm.

Para que los receptores ópticos no se activen al percibir falsas señales de luz que se encuentran en el ambiente, a la vez que se evitan las interferencias generadas por la luz solar, se instalaron los arreglos de fototransistores, dentro de tubos de 15 cm de longitud en cuyo extremo se hizo un corte en un ángulo de 35°.

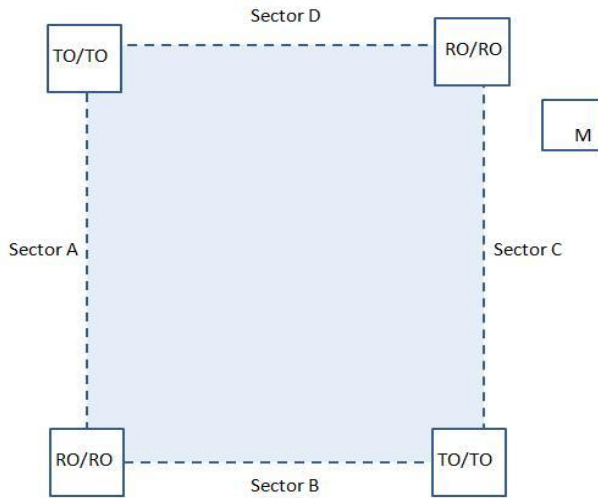


Fig.2 Cerco óptico.

2). *Subsistema de transmisión.* El proceso de transmisión comienza con la detección del corte de un haz de luz, láser. Cada emisor receptor láser esta relacionado con una dirección y un dato de ocho bits únicos, que corresponden al número de sector. La dirección y el dato se configuran en el circuito transmisor a través de interruptores. Una vez detectado el corte del haz de un sector, la dirección y el número de sector son leídos y codificados en un paquete por el dispositivo CIP 8-E. Una vez listo el paquete, un transmisor RF modelo TWS-434a, que opera a una frecuencia de 433.92 MHz de amplitud modulada, se encarga de irradiar el paquete de información generado por el CIP 8-E. Un diagrama a bloques del proceso de transmisión se presenta en la Fig.3. Un diagrama esquemático se presenta en la Fig. 4.

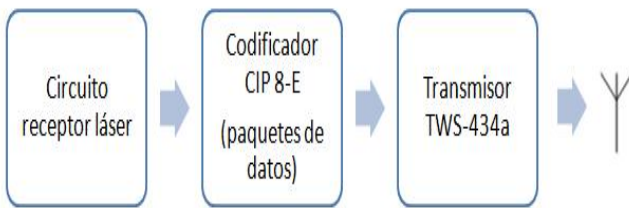


Fig. 3 Diagrama a bloques; sensado y transmisión.

Operación del codificador CIP-8E [6]. Al conectar la alimentación, el codificador entra a modo de bajo consumo de energía. Al poner la terminal TE (habilitación de transmisión) a nivel bajo, el codificador sale del modo de bajo consumo y comienza el proceso de transmisión. Durante una transmisión el codificador lee el estado de las líneas de dirección y las líneas de datos, usa esos valores para formar un paquete de información y lo transmite a través del TWS-434a. La transmisión de paquetes se repite una y otra vez, durante todo el tiempo que la terminal TE este a nivel bajo. Y vuelve a modo de bajo consumo cuando TE vuelve a nivel alto. La dirección y las líneas de datos se actualizan durante la formación de cada paquete y se termina de transmitir el paquete actual, incluso aunque la terminal TE ya este conectada a tierra.

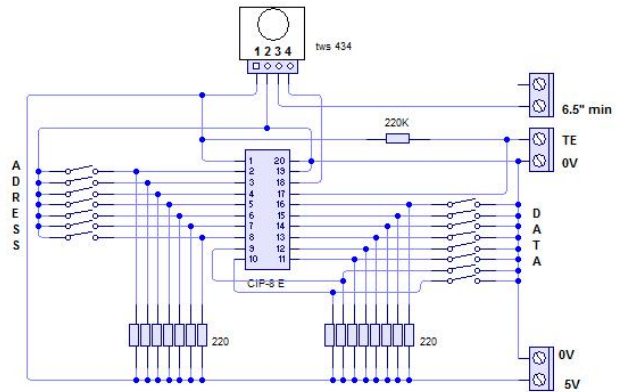


Fig. 4 Subsistema de transmisión RF.

En cada torre existe una fuente de alimentación, que convierte CA a CD, para los circuitos de emisión ó recepción láser. Esto es por el consumo de corriente de estos circuitos. Sin embargo los transmisores RF se alimentan con pilas recargables a base de luz solar. Estas son suficientes para que el circuito transmisor opere correctamente día y noche. Lo anterior se relaciona con el hecho de que durante la mayor parte del tiempo los integrados CIP-8E se encuentran en estado de bajo consumo. También, los transmisores se alimentan de pilas recargables con energía solar, con el objeto de que los

transmisores tengan la capacidad de informar sobre el estado de la línea, en el momento que un diodo láser deje de funcionar por falta de energía eléctrica.

Paquetes de datos codificados. Cada paquete de datos se compone de siete octetos de información: el preámbulo, el byte de sincronización, el byte de dirección, el byte de datos, el byte de dirección, el byte de datos, y un byte de verificación. La dirección y el dato se transmiten dos veces con fines de detección de errores. Y el byte de verificación se obtiene a través de la suma del byte de dirección y el byte de datos; llamado checksum. Este último se usa también con fines de detección de errores.

Después de la transmisión de un paquete se inserta un tiempo de espera de 10 ms. Tiempo suficiente para que el decodificador haga sus tareas: recibir, decodificar, verificar, y procesar el paquete. Durante el tiempo de espera el codificador vuelve al modo de baja potencia.

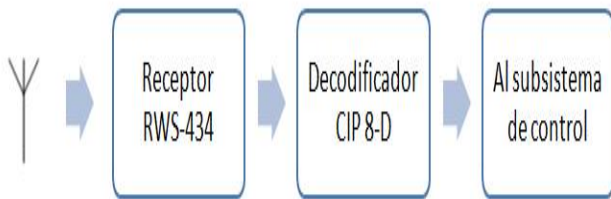


Fig. 5 Diagrama a bloques de la recepción.

3). *Subsistema de recepción.* La etapa de recepción está compuesta por el receptor RF RWS-434, que posee la misma frecuencia de operación que el transmisor TWS-434a. El circuito RWS-434 recibe los paquetes en forma serial, al mismo tiempo que el CIP-8D los

decodifica para entregarlos en formato de ocho bits al subsistema de control. Un diagrama a bloques del proceso de recepción se presenta en la Fig.5. y su correspondiente diagrama esquemático en la Fig. 6.

Operación del decodificador CIP-8D. Al conectar la alimentación, lo primero que hace el decodificador es entrar en un bucle de “recibir datos”, el cual es usado para detectar bytes de sincronización. Si durante 65.5 ms no se ha recibido un byte de sincronización válido, con el desbordamiento del temporizador interno de 16-bits, se actualiza el estado del dispositivo; se provoca la salida del bucle de recibir datos, se restablece la salida de los pines, y se entra nuevamente al bucle de recibir datos.

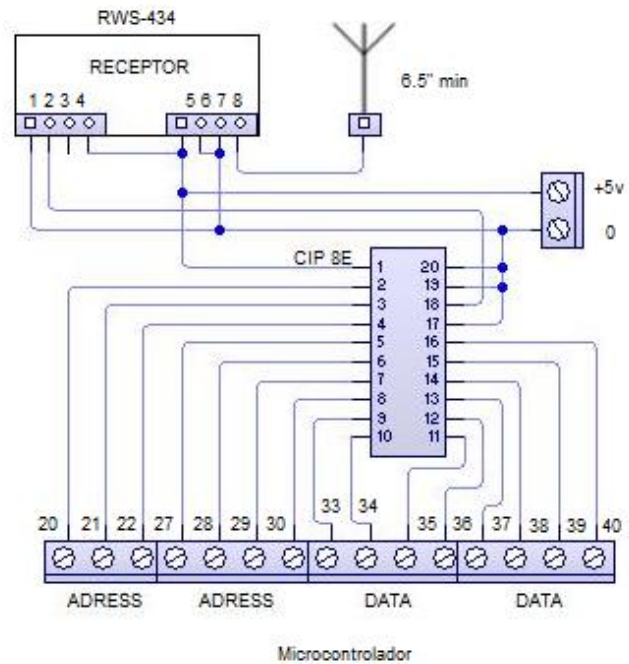


Fig. 6 Subsistema de recepción RF.

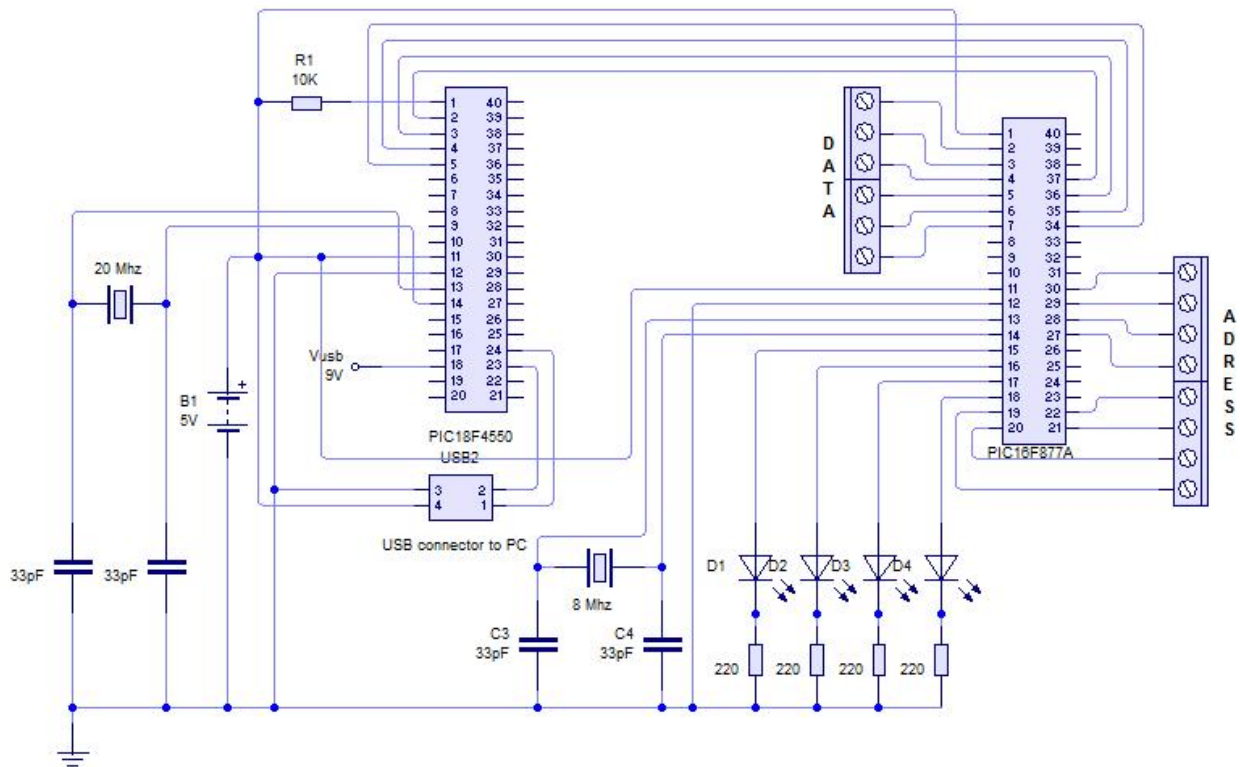


Fig.7 Subsistema de control.

Una vez que un byte de sincronización recibido es válido, se pasa a la etapa dos; el temporizador es desactivado, y el resto del paquete de datos es recibido y almacenado. Después de almacenar el paquete comienza el proceso de verificación de los datos y detección de errores. Una vez que los datos han sido verificados y se ha encontrado que están libres de errores, se pasa a la etapa tres; la información decodificada es colocada en los pines de salida, y el decodificador vuelve a entrar en el bucle de recibir datos para la próxima validación de paquetes. Si los bytes del paquete no están libres de errores, se considera dato no válido; se ignora la información y se regresa al bucle de recibir datos.

Si el decodificador está funcionando en modo fijo $L/M=1$, el último dato válido recibido se mantendrá en la salida hasta que un nuevo dato válido sea recibido. En modo momentáneo $L/M=0$ el decodificador mantendrá el dato de ocho bits

por el tiempo que se mantenga la dirección válida. Cuando se recibe una ráfaga de datos válidos el temporizador se deshabilita y el decodificador responde rápidamente cambiando los valores del bus de datos y manteniendo el patrón binario en sus salidas.

4). *Subsistema de control.* El elemento principal del subsistema de control es un microcontrolador PIC 16F877A [3-5], Fig. 7. Su función consiste en generar las direcciones de cada sector del cerco eléctrico. Una vez generada una dirección el decodificador CIP-8D entrega un dato que corresponde al número de sector. Lo anterior sucede solo si en el aire hay paquetes indicando que el haz de luz de dirección indicada fue cortado. En caso contrario, el CIP-8D no entrega nada y pone las líneas del bus de datos en alta impedancia. Una vez recibido el dato, el PIC 16F877A comprueba que la dirección generada y el dato recibido correspondan con una tabla de

valores preestablecidos: transmisor 1, dirección 0x32, dato 0x01; transmisor 2, dirección 0x54, dato 0x02; transmisor 3, dirección 0xAA, dato 0x03; transmisor 4, dirección 0xCC, dato 0x04. Estas comprobaciones y el sistema de codificación de paquetes usado entre el CIP-8E y el CIP-8D hacen posible la eliminación de falsas alarmas. Y las señales resultantes son empleadas para manipular un tablero de visualización.

El tablero de visualización está construido con lámparas y cada una representa a un sector del perímetro. Cuando algún sector es traspasado, se activa la lámpara correspondiente y se activa una bocina. El objeto de esto es alertar al personal de seguridad e indicarle la localización precisa en que se ha traspasado el perímetro. En el subsistema de control se incluyó un segundo microcontrolador, un PIC 18F4550, cuya función es comunicarse con una PC a través del puerto USB.

III. PROGRAMACIÓN

El modo de funcionamiento del USB usado en el presente trabajo es el modo Human Interface Device HID en versión de baja velocidad [1, 2]. Para comunicar al sistema de control con la computadora se usaron herramientas de software y hardware: EasyHID Wizard, Compilador Basic, Visual Basic 6.0, y un microcontrolador PIC18F4550. Las ventajas que ofrecen estas herramientas son la facilidad de usarlas, y sobretodo facilitan el uso del puerto USB con el EasyHID Wizard. Este software, después de unos pasos de configuración, genera; un archivo fuente en lenguaje BASIC para PIC, con los bloques funcionales básicos necesarios para programar al puerto USB; y un código para Visual Basic, con el código necesario básico para comunicarse a través del puerto USB con el microcontrolador desde la PC.

El EasyHID Wizard genera código para el PIC en Basic ó Protón +, según la selección del usuario. Para el PC genera código en Borland Delphi, Visual Basic ó Visual C++; también según la opción del usuario.

Para obtener los códigos, se tiene que ejecutar el EasyHID Wizard. Una vez ejecutado el programa, aparecen varias ventanas de configuración, aquí se selecciono: Company Name: MyCompany, Product Name: Sistema-seguridad, Vendor ID: 6017, Product ID 2000, Polling Input= 10 ms, Polling Output= 10 ms, Bus Power= 50 x2 ma, Buffer Input= 8 bytes, Buffer Output= 8 bytes, Project Name= Sistema de seguridad, Location= D:\users\Franco\documents\, compiler= microEngineeringLabs PICBASIC PRO, microcontroller= 18F4550, compiler= Microsoft Visual BASIC 5.0. Una vez seleccionadas las opciones mencionadas arriba, EasyHID Wizard esta listo para generar el código.

EasyHID Wizard genera una serie de archivos en los cuales se encuentran todas las funciones para la comunicación USB, los descriptores del microcontrolador, y los archivos de cabecera. Pero también genera una función, donde se introduce código propio, *usbinit*. Es aquí donde se programan las tareas que tiene que hacer el microcontrolador. Hechos estos cambios y si ningún error es detectado durante la compilación, el archivo .hex se creará con éxito y solo restará cargarlo en el microcontrolador. A continuación se presenta el código generado por EasyHID Wizard para la función *usbinit* y en cursiva se presenta el código agregado para las necesidades del presente proyecto.

```

USBBufferSizeMax con 8 ' maximum buffer size
USBBufferSizeTX con 8 ' input
USBBufferSizeRX con 8 ' output
USBBuffer Var Byte[USBBufferSizeMax]
USBBufferCount Var Byte
' *****
reset Var USBBuffer[7]
cl Var reset.BIT0
ADCON1 = 15
CMCON = 7
TRISA = %11111110
TRISD = 0
valor Var Byte
PORTA.0 = 0
' *****

```

```

usbinit
ProgramStart:
  USBService
  ' *****
  valor= PORTA.1*1+ PORTA.2*2+ PORTA.3*4+
        PORTA.4*8
  usbbuffer[6]= valor
  USBService
  gosub DoUSBOut
  pause 50
  USBService
  gosub DoUSBIn
  PORTA.0= cl
  PAUSE 50
  USBService
  PORTA.0= 0
  USBService
  Pause 50
  USBService
goto ProgramStart

```

Con el código que se agregó se hacen diversas tareas; se declaran las variables *reset*, *cl* y *valor*; se configuran las terminales del puerto A, RA0 como salida y el resto como entradas; se leen, dentro de la función *usbinit*, los valores de los pines del puerto A, 1 a 4, para conocer el valor que está enviando el PIC de control 16F877A y determinar cuál de las barreras fue interrumpida; se carga al buffer de envío de datos el número de la barrera interrumpida, si ese fuese el caso; se ejecuta, en repetidas ocasiones, la función *USBService* para mantener el vínculo de conexión entre el PIC y la PC; se manda a llamar a la rutina de escritura del buffer *DoUSBOut*; se hace un retardo de 50ms; se lee al buffer de lectura procedente de Visual Basic con la función *DoUSBIn*; y se envía *cl* a la terminal 0 del puerto A. La variable *cl* depende del bit 0 del buffer de lectura. Si este bit está a uno, ese uno se manda por el pin 0 del puerto A. Esto sirve para generar una interrupción en el PIC16F877A, cuya función es borrar al puerto A de dicho integrado. Este puerto controla a las lámparas que señalizan que un sector fue cortado por un intruso. Esta opción permite apagar a las luces de señalización desde un botón ubicado en la interfase visual de la PC.

El EasyHID Wizard genera código en Visual Basic para comunicación USB desde una PC. El código es presentado con comentarios en inglés. Los comentarios dicen lo que hace cada parte e indican dónde agregar código nuevo para adecuarse a la aplicación. Con los cambios que se introdujeron en ese código se implementó una interfaz gráfica de usuario que sirve para visualizar el estado del cerco óptico. En la Fig. 8 se muestra un ejemplo de la interfaz gráfica de usuario, la cual presenta una imagen que se compone de cuatro apuntadores láser que simulan las barreras de luz construidas físicamente. Cada vez que se corta un haz de luz, se refleja esto en la PC y aparece un número con un círculo rojo indicando el sector que fue cortado. El número desaparece al oprimir el botón *reset*.

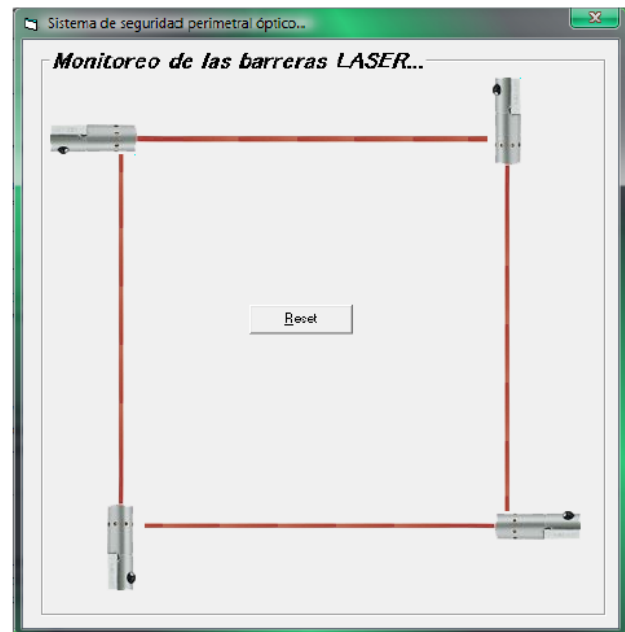


Fig. 8 Interfaz gráfica de usuario.

IV. RESULTADOS

Se implementaron los circuitos diseñados, Fig. 1 a 7, y se montaron sobre torres metálicas. Los sistemas de transmisión fueron conectados a baterías recargables con celdas solares. Esto con el objeto de que en caso de falla por alimentación,

el transmisor siga en capacidad de transmitir. Se desarrollaron e implementaron los algoritmos para los microcontroladores y para la PC. En la etapa de pruebas el sistema funcionó correctamente.

V. CONCLUSIONES

Se diseñó e implementó un sistema de seguridad capaz de alertar la intromisión a un perímetro, cuya detección se hace mediante dispositivos ópticos, su unidad de control se basa en microcontroladores y una PC, y la comunicación entre dispositivos ópticos y el sistema de control es por radiofrecuencia.

VI. REFERENCIAS

- [1] Jan Axelson, *USB Mass Storage: Designing and Programming Devices and Embedded Hosts*, Lakeview Research LLC, Madison, WI, September 2006
- [2] Jan Axelson, *USB Complete Everything You Need to Develop Custom USB Peripherals*, Third Edition, Lakeview Research LLC, Madison, WI, September 2006
- [3] Fernando E. Valdez-Pérez, Ramón Pallas-Areny, *Microcontrollers: Fundamentals and Applications with PIC*, Taylor & Francis Group, 2009.
- [4] Tim Wilmshurts, *Designing Embedded Systems with PIC Microcontrollers*, Elsevier LTD, 2007.
- [5] Steven F. Barrett, Daniel J. Pack, *Microcontrollers Fundamentals for Engineers And Scientists*, Morgan & Claypool Publishers, 2006.
- [6] <http://www.rentron.com/Files/CIP-8.pdf>
- [7] <http://www.beyton.com.mx/perimetral.htm>

- [8] <http://alarmas.teoriza.com/proteccion-perimetral-laser->
- [9] <http://www.directindustry.es/prod/rasnel-laser>
- [10] <http://www.sallen.es/fichas/1.pdf>
- [11] <http://alarmas.teoriza.com/sistema-de-alarma-laser-definitivo-basado-en-el-cine-video-explicativo-del-funcionamiento.php>
- [12] <http://www.seco-larm.com/E960LSpb.htm>
- [13] <http://www.bydemes.com/detalles.asp?id=1995&idsubfamilia=32>

VII. AUTORES

Dr. Rubén Herrera Galicia obtuvo el título de Doctor en Ciencias Técnicas con especialidad en Electrónica por la *Technical University of Warsaw*, Polonia.

Olga Mariana Becerra Fuentes alumna del octavo semestre de la carrera de Ingeniería Electrónica en el *Instituto Tecnológico de Tuxtla Gutiérrez*.

Maria de los Ángeles Zepeda Hernández alumna del octavo semestre de la carrera de Ingeniería Electrónica en el *Instituto Tecnológico de Tuxtla Gutiérrez*.

Roque Aarón Moreno Pérez alumno del octavo semestre de la carrera de Ingeniería Electrónica en el *Instituto Tecnológico de Tuxtla Gutiérrez*.

Fernando Francisco Franco Pérez alumno del octavo semestre de la carrera de Ingeniería Electrónica en el *Instituto Tecnológico de Tuxtla Gutiérrez*.

NOVEDOSO SENSOR REFRACTOMÉTRICO DE FIBRAS ÓPTICAS CON ELEMENTO DE DETECCIÓN SEMICILÍNDRICO PLÁSTICO, PARA MEDIR EL NIVEL EN UN TANQUE DE GASOLINA O DIESEL

Sergei Khotiaintsev, Enrique Llanito Caudillo, Selene Pérez García, Jordi Morales Farah

Departamento de Ingeniería en Telecomunicaciones, Facultad de Ingeniería.
Universidad Nacional Autónoma de México
Av. Universidad 3000, Ciudad Universitaria, Coyoacan, México D. F. CP 04510.
Correo electrónico: sergeik@hotmail.com

RESUMEN

En este artículo se presenta un sensor refractométrico con un elemento óptico de detección semicilíndrico. Un par de fibras ópticas multimodo están acopladas simétricamente al plano del elemento. El acoplamiento de las fibras debido a la reflexión total interna en la superficie del elemento es susceptible al índice de refracción del medio que rodea al elemento. Para obtener un mejor acoplamiento entre las fibras, se realizaron diferentes experimentos cambiando la distancia entre las fibras, con el fin de alcanzar un mayor enfoque de los rayos en la superficie del elemento y así mejorar la sensibilidad al índice de refracción del medio externo al mismo y reducir las pérdidas ópticas intrínsecas en el transductor. Se muestran los resultados teóricos en donde se puede observar que el transductor en particular, puede discriminar entre aire, agua, gasolina y diesel. Al igual se presentan datos experimentales obtenidos con un elemento de detección de plástico, los cuales confirman los resultados teóricos. Un elemento de detección relativamente largo, con un arreglo vertical de fibras ópticas puede usarse como un sensor para medir el nivel de gasolina o diesel en los tanques de automóviles o camiones.

Palabras clave: sensor óptico, refractometría óptica, sensor de nivel para líquidos, gasolina, diesel.

I. INTRODUCCIÓN

Los sensores de fibras ópticas ofrecen muchas ventajas sobre los sensores mecánicos, eléctricos y otros. Anteriormente se usaban fibras ópticas de silicio en los sensores principalmente, recientemente, el avance significativo en la calidad de las fibras ópticas de plástico (POF) ha generado el desarrollo de muchas nuevas aplicaciones para la transmisión óptica y el sensado óptico. Las POFs son atractivas en el mercado debido a que son dispositivos multimodo de un diámetro relativamente grande el cual es fácil de manipular e instalar. Los sistemas de transmisión basados en POF usan componentes simples y baratos como son LEDs (diodos emisores de luz) en el rango de la luz visible o cerca del infrarrojo, conectores, divisores, acopladores simples, etc. Todo el costo del sistema es bajo y su mantenimiento es simple [1]. Actualmente las POFs permiten tasas de transmisión de hasta 10 Gbit/s sobre distancias de alrededor de los 300m [2] y al igual que las fibras de silicio son inmunes a interferencias electromagnéticas y ruido por lo que pueden instalarse junto a cables eléctricos de potencia y las pérdidas por curvatura son muy bajas, permiten un radio de curvatura de hasta 20 mm por lo que facilita su instalación en lugares estrechos y paredes.

Las aplicaciones de POF incluyen distancias relativamente cortas en redes de área local (LANs) en oficinas, edificios, campus pequeños. Otra aplicación emergente es transmisión de datos óptica (analógica y digital) a bordo de carros,

camiones, trenes, barcos, aviones, etc. Una característica distintiva de los vehículos es que necesitan un monitoreo de diferentes cantidades físicas relacionadas a la operación del vehículo (velocidad, temperatura, presión, nivel de diferentes líquidos en depósitos y tanques, etc.). Por tanto, es de un gran interés agregar comunicaciones ópticas con sensado óptico a bordo del vehículo. Por motivos de compatibilidad los sensores deberían operar vía las mismas POFs que el resto de la red óptica a bordo del vehículo.

Como se mencionó anteriormente una de las aplicaciones de estas fibras es en la industria automotriz en la cual se han desarrollado estándares para las redes a bordo de los mismos. Uno de los estándares es el Domestic Digital Bus System (DB2) el cual especifica un sistema de anillo el cual interconecta diferentes dispositivos como es el radio, televisión, cambiadores de CD y sistemas de navegación en vehículos y alcanza unos 20 Mbps en 10 m [3]. La especificación Media Oriented Systems Transport (MOST) es uno de los sistemas más utilizados en aplicaciones de control, seguridad y multimedia con una tasa de hasta 150 Mbps [4], mientras que Byteflight es un sistema que trata con conexiones de sistemas de bolsas de aire y cinturones de seguridad con otros componentes de control a una tasa de 10 Mbps [5].

El sensado de nivel de líquidos es importante en todas las aplicaciones automotrices. Existen un amplio grupo de métodos de sensado disponibles para determinar el nivel de líquidos. Ellos incluyen métodos mecánicos, eléctricos y ópticos y sus combinaciones. Actualmente los sensores de nivel de líquidos eléctricos son empleados para conocer el nivel de gasolina en los tanques de los automóviles. En estos sensores el nivel de la gasolina es medido con un generador de señales eléctrico por una resistencia variable controlada por un flotador dentro del tanque. Estos sensores son simples y baratos, pero su resolución es relativamente baja. También, los cables eléctricos en un ambiente inflamable presentan un peligro potencial. Los sensores de fibra óptica ofrecen distintas ventajas para estas aplicaciones debido a la naturaleza dieléctrica de las fibras ópticas. Las

ventajas de los sensores ópticos es que no están limitados a la ausencia de conductividad eléctrica, ofrecen absoluta inmunidad a campos electromagnéticos y alta resistencia a riesgo por chispas. Estas características hacen a los sensores ópticos ideales candidatos para su uso en tanques de gasolina y diesel.

Hay cuatro principales clases de sensores ópticos: intensidad, fase, polarización y de frecuencia. La clasificación de los sensores en cada tipo, diferencia significativamente su desempeño y costo, el cual casi siempre está relacionado a la complejidad del principio de medición empleado. En las aplicaciones automotrices, los factores más importantes para seleccionar un sensor son su simplicidad, el fácil procesamiento de las señales, el bajo costo y un simple mantenimiento. Entre los cuatro tipos de sensores que mencionamos, el tipo de intensidad (amplitud) es el más simple de implementar debido a la facilidad de la medición de la intensidad, en contraste con las mediciones de fase y polarización. También los sensores ópticos basados en la medición de intensidad son compatibles con POFs, por consiguiente, los sensores de intensidad óptica son los candidatos más prometedores para las aplicaciones automotrices. Los sensores del tipo intensidad están basados en muchos principios. Uno de ellos es la refractometría óptica, la cual se basa en la determinación del índice de refracción de las sustancias, n . El índice de refracción n está relacionado con muchas características cuantitativas de las sustancias. En el caso de líquidos estas características cuantitativas como la temperatura, presión, la concentración de sustancias solubles (como ejemplo, sal, azúcar, etc.) pueden ser obtenidas indirectamente midiendo el índice de refracción. Además, la refractometría permite discriminar entre diferentes sustancias, debido a que cada sustancia o material tienen diferentes índices de refracción.

Anteriormente ya se han desarrollado sensores de POF para aplicaciones automotrices, tal es el caso de un simple y efectivo sensor óptico plástico para la medición de temperatura [6] y un sensor refractométrico plástico que emplea una sección

cónica de fibra plástica de índice gradual [7]. La posibilidad de la medición del nivel de líquidos con el sensor de refractométrico de campo evanescente compuesto de múltiples curvaturas finas en una simple POF extensa o en un arreglo de POFs ha sido exitosamente demostrado en experimentos con agua [8]. Pero, la resolución vertical del mismo sensor es limitada por un tamaño finito de fibras curvadas y podrían ser difíciles de acomodar muchos bucles de fibras en el tanque. Aquí nosotros presentamos un novedoso sensor refractométrico que emplea un elemento de detección semicilíndrico. Puede ser empleado como un sensor puntual, o como un arreglo vertical en cuyo caso puede ser un sensor de nivel de líquidos, como es gasolina o diesel, de una manera discreta en un tanque o contenedor. A continuación se presenta la descripción del sensor y sus características de funcionamiento.

II. SENSOR DE FIBRAS ÓPTICAS REFRACTOMÉTRICO

El sensor que tratado aquí consiste de un elemento óptico de detección semicilíndrico el cual está conectado a un par de fibras ópticas como se muestra en la figura 1. El elemento de detección óptico está hecho de un material transparente dieléctrico. En este trabajo se utilizó un elemento de detección de polimetil metacrilato (PMMA) con un índice de refracción $n=1.4828$ para una $\lambda=940$ nm [9].

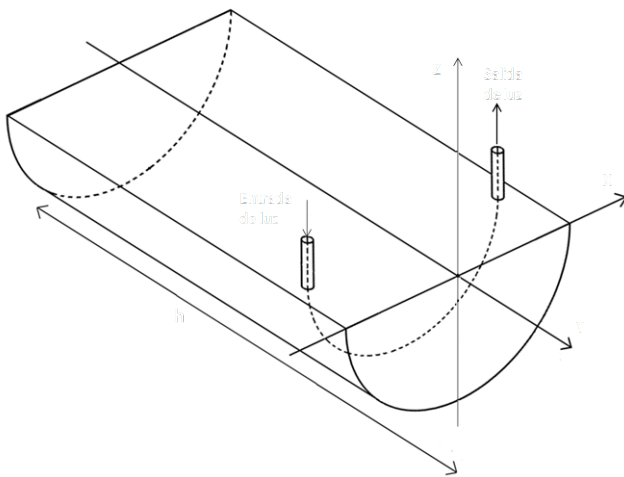


Figura 1. Vista esquemática del sensor.

El principio de operación del sensor es ilustrado en la figura 2, el cual muestra una sección transversal del elemento de detección óptico y las fibras conectadas a él. Una de ellas es la fibra de entrada y la otra la de salida. La luz de una fuente remota (LED no mostrado en la figura) es acoplada vía la fibra de entrada (2) al interior del elemento de detección (1). La luz se propaga por medio de reflexiones totales internas hasta que llega al plano de salida del elemento de detección. En este plano una parte de la luz es acoplada al núcleo de la fibra de salida (3) y llega a través de esa fibra a un fotómetro remoto (no mostrado en la figura).

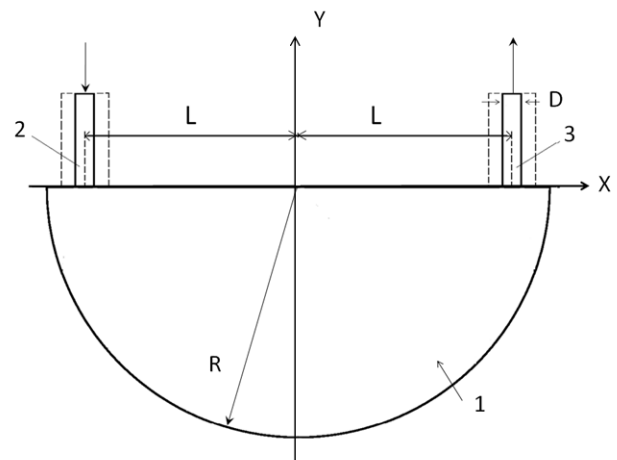


Figura 2. Transductor refractométrico, consiste de un elemento semicilíndrico de dieléctrico transparente (1) integrado con una fibra óptica de entrada (2) y otra fibra óptica de salida (3).

Cuando el elemento está en el aire la reflexión es un poco más potente (bajo cierta posición de las fibras relativa en el elemento de detección), cuando el elemento está inmerso en un líquido la reflexión en la superficie del cilindro disminuye o se desvanece por completo debido a una menor diferencia en los índices de refracción en los dos medios. Por consiguiente la transmisión óptica en el elemento de detección es $T=I_{out}/I_{in}$, I_{out} e I_{in} son la intensidad de la luz de salida y entrada respectivamente, las cuales son una medida de el índice de refracción externo, n .

La diferencia en la transmisión óptica en el elemento de detección en el aire y en un líquido en particular, sirve como discriminación entre los dos medios y es un fundamento para la medición de nivel de líquido con un arreglo vertical de estos sensores (en cuyo caso, su salida es interpretada de una manera binaria).

Todos los parámetros geométricos del sensor están relacionados al radio del elemento de detección R , y se usaron las siguiente cantidades adimensionales para describir la geometría del sensor: el diámetro adimensional del núcleo de las fibras $\Phi=D/R$, la distancia adimensional de las fibras ópticas desde el plano de simetría del elemento $A=L/R$.

La transmisión óptica T es una función de los parámetros ópticos y geométricos del elemento de detección y las fibras ópticas. En vista de obtener una relación entre estos parámetros, se empleó un procedimiento iterativo con el propósito de encontrar la combinación de valores óptimos de los parámetros del sensor. Este procedimiento incluye el análisis de la transmisión del sensor contra el índice de refracción externo, la cual fue obtenida por medio del algoritmo de trazados de rayos y experimentalmente. El análisis arrojó datos para las correcciones del diseño inicial. Este procedimiento fue repetido hasta que se obtuvo una diferencia suficientemente grande de la respuesta del sensor al aire-gasolina y al aire-diesel.

Se simuló el trazado de rayos por medio de un modelo matemático, un algoritmo computacional y un programa de computación en Visual Basic, programa descrito en [10]. El análisis fue simulado para rayos en el plano x - y del elemento de detección para el caso en que el medio externo es aire, este análisis se muestra en la figura 3.

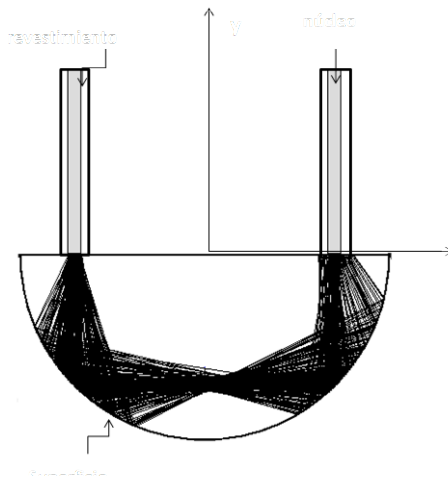


Figura 3. Trayectoria de rayos en el elemento de detección bajo $\Phi=0.075$, $\Lambda=0.7$, $NA_{entrada}=0.485$, $NA_{salida}=0.095$.

Con el análisis numérico se obtuvo también la transmisión óptica contra índice de refracción externo con la geometría del sensor y las constantes del material como parámetros. En práctica, es conveniente caracterizar el funcionamiento del sensor en términos de la transmisión relativa $T^* = I_{out}(n)/I_{out\ air}$, debido a que esta cantidad no incluye la intensidad de la luz de entrada I_{in} la cual no puede ser medida experimentalmente [11]. Las gráficas de la transmisión relativa T^* contra el índice de refracción externo n obtenidas bajo la combinación casi-óptima de los parámetros del sensor para varias distancias adimensionales relativas A son mostradas en la figura 4.

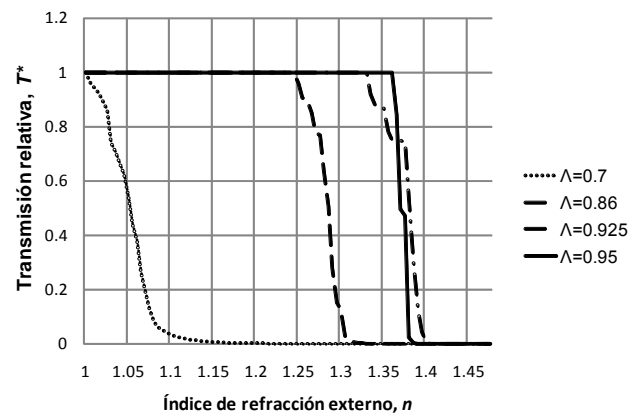


Figura 4. Transmisión relativa calculada T^* vs. Índice de refracción externo, n bajo $\Phi=0.075$, $NA_{entrada}=0.485$, $NA_{salida}=0.095$.

Los datos en la figura 4 muestran el efecto de la distancia adimensional en la respuesta del sensor a diferentes índices de refracción externos. El índice de refracción de la gasolina es alrededor de 1.442 y el del diesel es de 1.461 para una longitud de onda $\lambda = 940$ nm. Para una confiable discriminación entre aire y gasolina o entre aire y diesel es la característica escalonada T^* entre $n=1.3...1.4$ es apropiada. Por tanto, cualquiera de las tres distancias adimensionales Λ ($\Lambda=0.86, 0.925$ y 0.95) pueden ser empleadas en el sensor, debido a que entregan una función de la transmisión relativa en el rango específico de índices de refracción externos n .

Además se realizó un análisis para conocer si es factible fabricar el sensor con algunas fibras ópticas plásticas comerciales, las cuales son mostradas en la tabla 1 [12], todas son fibras comercializadas por Mitsubishi, fabricadas con un núcleo de PMMA, con índice de refracción de 1.49 y cabe mencionar que se seleccionaron estas fibras porque el fabricante las recomienda para el uso de sensado de temperatura, velocidades, niveles de líquidos y posicionamiento, también porque la única característica que cambia entre las mismas es el diámetro del núcleo.

Modelo	Perfil	NA	Diámetro del núcleo [μm]	Diámetro adimensional Φ
SK10	SI	0.5	240	0.012
SK30	SI	0.5	735	0.03675
SK40	SI	0.5	980	0.049
SK60	SI	0.5	1470	0.0735

Tabla 1. Características de algunas POFs comerciales

La transmisión relativa T^* calculada contra en índice de refracción bajo la distancia adimensional $\Lambda=0.95$ y características idénticas de todas las fibras ópticas es graficada en la figura 5. De acuerdo con los datos es posible concluir que la fibra estándar SI POF es la menos adecuada para esta aplicación, debido a la pendiente que presenta en la transmisión relativa.

En la gráfica de la figura 5 se muestra la transmisión relativa T^* calculada contra el índice de refracción con un posicionamiento de las fibras a una distancia dimensional $\Lambda=0.925$ para las fibras mencionadas en la tabla 1. Realizando un análisis de esta gráfica se concluye que el modelo más adecuado para la fabricación del sensor es el SK10 ya que presenta una respuesta escalonada, mientras que el menos adecuado es el SK60.

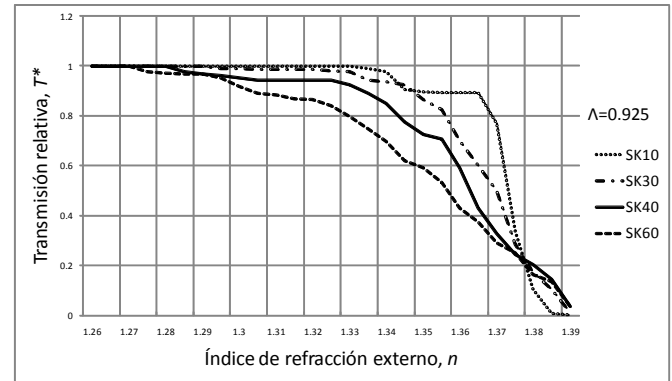


Figura 5. Transmisión relativa calculada T^* vs. Índice de refracción externo, n para algunas fibras POF comerciales

III. EXPERIMENTO.

El experimento fue realizado con un modelo del sensor físico relativamente grande. El diseño esquemático del experimento se muestra en la figura 6. Se utilizó un haz de miles de fibras ópticas de un diámetro pequeño (alrededor de 20 μm cada una). Cada haz cuenta con un diámetro total de 3mm. Se utilizó un diafragma con apertura circular de varios diámetros en orden de modelar físicamente las fibras ópticas con diámetros distintos en su núcleo. La apertura angular efectiva de la fibra de entrada fue controlada cuando fue necesario por el ángulo de incidencia de la misma. Ambas fibras ópticas estaban en contacto con el elemento de detección en puntos simétricos a la superficie del elemento detector. Se empleó un líquido adecuado para acoplar el área de contacto de la superficie del elemento con la fibra con el propósito de evitar reflexiones excesivas en la interfase de contacto.

Las terminaciones de las fibras que estaban en contacto con el elemento de detección fueron montadas en dos unidades de desplazamiento en los planos x-y-z de precisión las cuales permitieron el movimiento de las fibras relativo al plano de simetría del sensor. El elemento semicilíndrico de detección tiene un radio $R = 20\text{mm}$ (en un dispositivo práctico, el radio del elemento de detección puede reducirse y pueden usarse fibras ópticas comunes en el campo de las telecomunicaciones en lugar de un haz de fibras.)

Se utilizó como fuente de luz un LED infrarrojo de alta radiación hecho de GaAlAs con un ángulo de divergencia de emisión de 40° . Dicho LED fue excitado por una onda de corriente triangular que fue generada por un generador de señales, dicha onda tuvo una amplitud de 14 mA a una frecuencia de 100 Hz ; se escogió una onda triangular para identificar una posible saturación en el receptor óptico a la salida del sensor. El generador fue conectado en serie al LED y a una resistencia de $330\ \Omega$ y la fuente de luz fue acoplada a la fibra óptica de entrada.

Como fotodetector se usó un fotodiodo de Ge de gran área de sensibilidad (25 mm^2), el cual se acopló a un amplificador de transimpedancia, el cual se conectó a un fotómetro. Los datos obtenidos de cada medición fueron guardados en una computadora vía una interfase RS232 para después ser procesados.

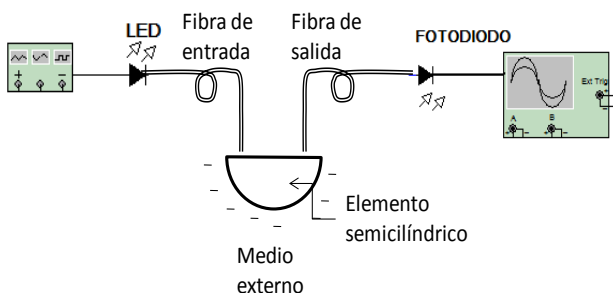


Figura 6. Diagrama esquemático del experimento.

La característica de transmisión del transductor fue obtenida de la siguiente manera: El par de fibras ópticas fueron posicionadas originalmente a la mínima distancia posible alrededor de $0.5 R\Lambda = 0.5$). Así, se incrementó la distancia de las fibras desde el plano de simetría del sensor discretamente con incrementos de $\Lambda=0.005$ hasta el borde del elemento semicilíndrico ($\Lambda = 0.99$). Para cada posición de las fibras la intensidad de la luz fue medida a la salida del transductor; dichas mediciones se realizaron primero como medio externo el aire y después con el elemento semicilíndrico inmerso en diferentes líquidos, como agua destilada, gasolina y diesel. Enseguida la transmisión relativa fue calculada, procesada estadísticamente (se promedió los resultados de 10 mediciones individuales) y graficada.

Alternativamente, la transmisión óptica del sensor fue medida con varios medios externos (el aire y varios líquidos) bajo una posición fija de las dos fibras ópticas. Esto excluyó posibles errores en la posición de la fibra cuando se mide la transmisión óptica bajo diferentes medios externos.

IV. RESULTADOS.

Los datos experimentales de la transmisión del sensor en el aire contra la distancia relativa Λ se graficó en la figura 7. Los picos observados en la gráfica son debidos al acoplamiento de las dos fibras ópticas por dos, tres, cuatro o más reflexiones secuenciales de la luz en la superficie del elemento de detección cilíndrico. Esto corresponde a algunas distancias discretas adimensionales Λ , mientras que para las distancias intermedias no se provee ningún acoplamiento o se provee uno muy reducido.

La transmisión en líquidos es siempre menor que la transmisión en el aire, debido a la reflectividad de la superficie del sensor cuando está inmersa en un medio más denso ópticamente. Por tanto, el cero o las regiones cercanas a cero en la gráfica de la figura 7 son inservibles; los picos de las regiones de transmisión que corresponden a un mayor acoplamiento entre las dos fibras ópticas son las

regiones que sirven para la discriminación entre aire y líquidos y por tanto para la medición de nivel de líquidos.

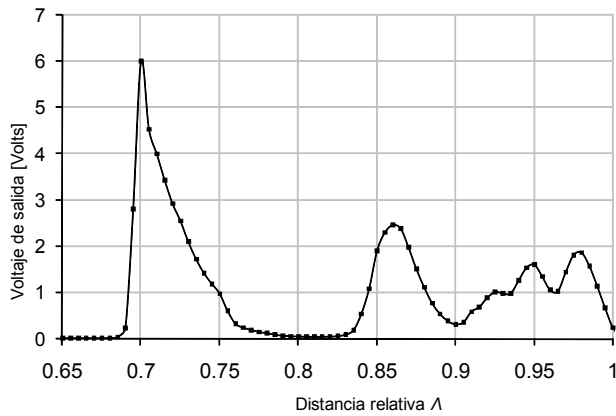


Figura 7. Voltaje de salida del sensor vs posición relativa de las fibras cuando el sensor esta en el aire.

Las distancias adimensionales Λ las cuales corresponden a los picos observados en la transmisión óptica en el aire están dadas en la tabla 2.

Pico de transmisión N.	Distancia Adimensional, Λ
1 °	0.70
2 °	0.86
3 °	0.925
4 °	0.95
5 °	0.975

Tabla 2. Picos de transmisión

Los datos experimentales en la transmisión relativa del sensor para aguas destilada, gasolina y diesel están graficados en la figura 8 para una distancia relativa de $\Lambda=0.925$ y parámetro adimensional de las fibras de $\Phi=0.075$. Para estos datos es posible llegar a la conclusión de que se puede discriminar claramente entre aire-gasolina, aire-diesel, agua-gasolina y agua-diesel a bajo las condiciones previamente descritas.

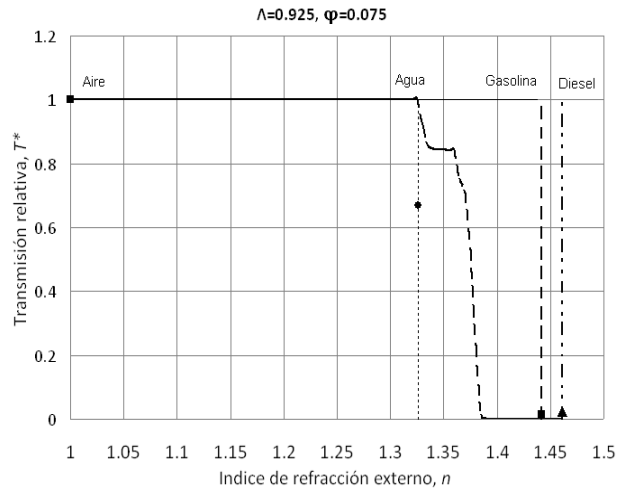


Figura 8. Resultados teóricos (.....) y experimentales bajo $\Phi=0.075$, $N_{Aentrada}=0.485$, $N_{A salida}=0.095$. Los puntos experimentales corresponden a agua destilada (•) gasolina (■) y diesel (▲).

V. CONCLUSIONES.

Como se mostró se investigó teórica y experimentalmente la respuesta de un sensor refractométrico de fibra óptica basado en un elemento de detección semicilíndrico de plástico que mide el índice de refracción de un medio circundante. Particularmente se investigó la transmisión para medios externos como aire, gasolina, agua y diesel con el propósito de integrarlo en el sistema de control de algún vehículo que utilice combustibles como los mencionados para medir el nivel dentro de un tanque de algún vehículo que utilice estos combustibles. Se investigó cuales son los parámetros óptimos en los cuales se puede tener una discriminación confiable entre aire-gasolina y aire-diesel de tal forma que se puede tener en un arreglo vertical de fibras ópticas para poder tener una medición completa de los tanques.

VI. REFERENCIAS.

- [1] O. Ziemann, J. Krauser, P. E. Zamzow, W. Daum, *POF Handbook. Optical Short Range Transmission Systems*, Springer, Berlin Heidelberg, 2008.
- [2] Breyer, F., Lee, J., Randel S., and Hanik, N., “*Comparison of OOK- and PAM-4 Modulation for 10 Gbit/s Transmission over up to 300 m Polymer Optical Fiber,*” Optical Fiber Communications National Fiber Optic Engineers Conference, 2008. OFC/NFOEC 2008. Conference on, 24-28 Feb. 2008, pp. 1–3.
- [3] Ciocan, C., “*The Domestic Digital Bus System (DB2). A Maximum of Control Convenience in Audio Video*”, Consumer Electronics, 1990. ICCE 90. IEEE 1990 International Conference on, 6-8 Jun. 1990, pp. 170 – 171.
- [4] Guglielmetti, L., “*Standardizing Automotive Multimedia Interfaces*”, IEEE Specrum, Apr.-Jun. 2003, pp. 76–78.
- [5] Cena, G. and Valenzano, A., “*Performance Analysis of Byteflight Networks*”, Factory Communication Systems, 2004, Proceedings. 2004 IEEE International Workshop on, 22-24 Sept. 2004, pp. 157–166.
- [6] Arcos M., Chana, D., Contreras, K. et al., “*Design and Fabrication of a Novel Plastic Optical Sensor for Temperature Measurement Using a Chemical Transducer*”, Plastic Optical Fibers, 2007, Proceedings. 16th International Conference on, 10-12 Sept. 2007, p. 189.
- [7] Arrue, J., Jimenez, F., Aldabaldetrekú G. et al., “*Analysis of the Parameters of Tapers in graded-Index POF for the Design of a Refractive-Index Sensor and Other Applications*”, Plastic Optical Fibers, 2007, Proceedings. 16th International Conference on, 10-12 Sept. 2007, pp. 238-241.
- [8] Poisel, H., Meedt, H., and Niewisch, J., “*Fiberoptic Liquid-Level Sensor – FOLLS*”, Plastic Optical Fibers, 2007, Proceedings. 16th International Conference on, 10-12 Sept. 2007, pp. 178-181.
- [9] N. G. Sultanova, S. N. Kasarova, C. D. Ivanov, and I. D. Nikolov, “*Refractive data of optical plastics for laser applications*” International Conference on Holography, Optical Recording, and Processing of Information, vol. 6252 of Proceedings of SPIE, pp. 1–5, Varna, Bulgaria, May 2006.
- [10] Khotiaintsev, K. *Desarrollo de una aplicación de simulación numérica para sensores refractométricos ópticos*, Tesis para obtener el grado, Universidad Autónoma de Puebla, Puebla, Pue., México, 2005 .
- [11] Svirid V., Khotiaintsev S. and Swart, P. “*Linear and Steplike Characteristics in an Optical Fiber Refractometric Transducer with Hemispherical Detection Element*”, OpticalEngineering, vol. 42, no. 5, 2003, pp. 1383-1389.
- [12] <http://www.mitsubishicorp.com/us/en/bg/ps/fo.html>

VII. AUTORES

Dr. Serguei Khotiaintsev recibió el grado de Maestro en Ciencias y el grado de Doctor de la Facultad de Ingeniería en Radio del Instituto Politécnico de Kiev (hoy en día Universidad Nacional Tecnológica de Ucrania). Actualmente es Profesor de Tiempo Completo de la Facultad de Ingeniería de la Universidad Nacional Autónoma de México. Es autor o co-autor de más de 120 artículos, algunos libros y tiene 19 patentes. El Dr. Khotiaintsev es *Senior Member* del IEEE.

Enrique Llanito Caudillo estudió Ingeniería en Telecomunicaciones en la Universidad Nacional Autónoma de México.

Ing. Selene Pérez García obtuvo su título de Ingeniería en Telecomunicaciones en la Facultad de Ingeniería de la UNAM. Actualmente es estudiante de tiempo completo de la maestría en Ingeniería Eléctrica en la UNAM.

Ing. Jordi Morales Farah recibió el grado de ingeniero en la Facultad de Ingeniería de la UNAM.

SISTEMA DE CONTROL Y ADQUISICIÓN DE DATOS PARA UN PROBADOR DE VÁLVULAS CARDÍACAS

Bañuelos Saucedo Miguel Angel, Castillo Hernández José, Monroy Elizondo Jaime.

Grupo de Electrónica, Centro de Ciencias Aplicadas y Desarrollo Tecnológico
Universidad Nacional Autónoma de México
Circuito Exterior S/N Cd. Universitaria, México, D.F. 04510, México
Tel. 56228602, correo electrónico: miguel.banuelos@ccadet.unam.mx

RESUMEN

En este artículo, se presenta el desarrollo de un sistema de control y adquisición de datos para un equipo probador de bioprótesis valvulares cardíacas. El sistema está formado por un programa para una computadora personal y de un módulo electrónico, basado en un microcontrolador PIC16F877, que maneja un motor de pasos y adquiere la señal de dos transductores de presión. El equipo es capaz de mostrar la gráfica de los transductores de presión en tiempo real, registrar los datos de la prueba y exportarlos a una hoja de cálculo. Su diseño está orientado a cumplir con la norma oficial mexicana correspondiente y fue desarrollado para el Instituto Nacional de Cardiología.

Palabras clave: Bioprótesis, válvulas cardíacas, instrumentación, sistema de control.

I. INTRODUCCIÓN

En el Instituto Nacional de Cardiología se han venido desarrollando bioprótesis valvulares cardíacas, a partir de tejido porcino, o pericardio bovino, las cuales se implantan en los pacientes que así lo requieren. El equipo de prueba con que se cuenta, no satisface las necesidades del Instituto, por lo que era necesario desarrollar un nuevo equipo. En particular, era importante desarrollar un equipo que tuviera la capacidad de ser controlado desde una computadora personal, con el propósito de facilitar el despliegue de información y el registro de datos.

El proceso de prueba de una prótesis cardíaca consiste en simular su funcionamiento dentro del corazón, en particular con respecto a los volúmenes desplazados, las frecuencias de operación y el ciclo de trabajo. Estos parámetros se encuentran especificados en la norma oficial mexicana NOM-063-SSA1-1993 [1].

II. ARQUITECTURA DEL SISTEMA

En el CCADET (Centro de Ciencias Aplicadas y Desarrollo Tecnológico) se construyó un equipo de prueba para válvulas cardíacas. Este equipo consiste en una cámara cilíndrica de acrílico transparente en donde se coloca la prótesis. Esta cámara tiene un pistón que es accionado mediante un mecanismo de biela y un motor de pasos. La cámara, el pistón, el motor de pasos y sus engranes fueron diseñados y fabricados por el Grupo de Micromecánica y Mecatrónica y la Sección de Desarrollo de Prototipos del CCADET, respectivamente.

En la Fig. 1 se muestra un esquema de la cámara de prueba. El pistón es el encargado de impulsar el líquido de prueba a través de la válvula, en ciclos pulsantes. Antes y después de la posición de la válvula, se colocan dos conexiones para un par de sensores de presión.

El sistema debe ser capaz de enviar la secuencia de control al motor de pasos para que este efectúe un movimiento de vaivén. De la longitud de la carrera dependerá el volumen de líquido desplazado.

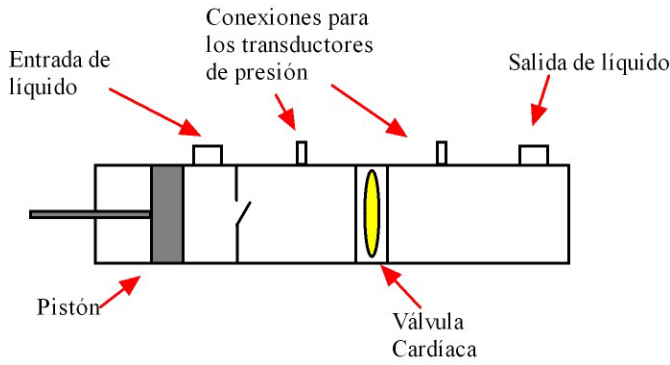


Fig. 1. Esquema de la cámara de prueba.

La arquitectura del sistema consiste en la cámara de prueba, los dos transductores de presión y sus circuitos de acondicionamiento de señal, el microcontrolador y el driver para el motor de pasos, según se muestra en la Fig. 2.

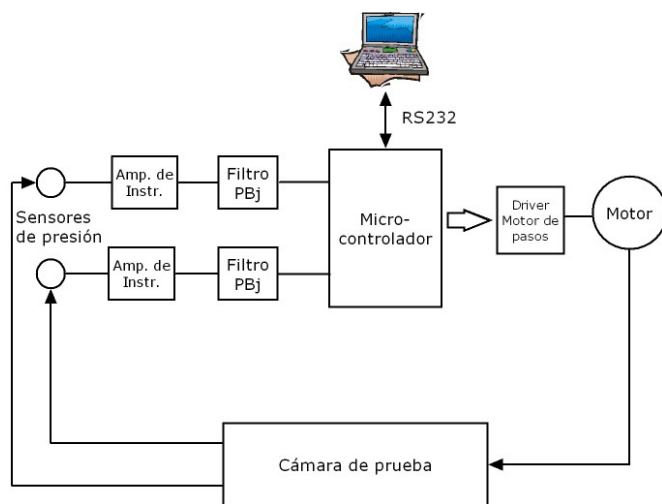


Fig. 2. Arquitectura del sistema.

III. SENSORES DE PRESIÓN

El equipo de prueba debe de reproducir los intervalos de presión sanguínea normal, es decir entre 80 y 120 mm de Hg. Se seleccionó un transductor de presión Honeywell 24PCBFA6G de ± 5 psi [2], que equivalen a 258.6 mm de Hg, lo cual nos da un amplio margen para las pruebas.

El transductor de presión es del tipo resistivo, por lo que para acondicionar su señal se requiere de un amplificador de instrumentación.

En la Fig. 3 se muestra el esquema del amplificador de instrumentación utilizado para adecuar la señal de los transductores de presión. Se utilizaron amplificadores operacionales TL072 de bajo ruido [3]. El amplificador A4 se utiliza para anular los errores de voltaje de offset, y para producir un voltaje de referencia de 2.5 V. De esta manera es posible medir niveles de presión de vacío utilizando el convertidor A/D del microcontrolador, el cual únicamente acepta niveles de voltaje positivos.

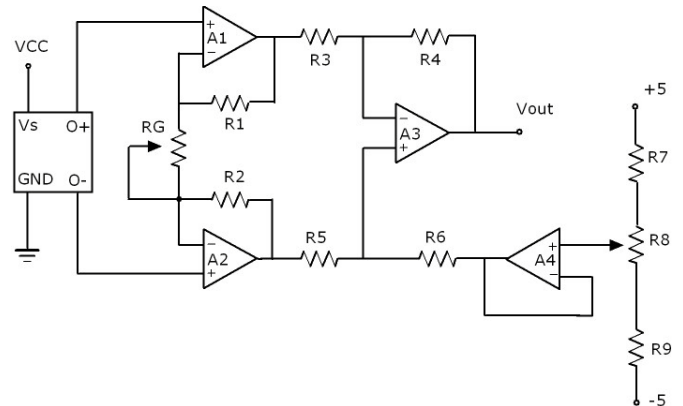


Fig. 3. Acondicionador de señal de los transductores de presión.

Para reducir el nivel de ruido de la señal, se utilizó un filtro paso bajas tipo Sallen-Key con respuesta Butterworth y frecuencia de corte en 40 Hz, a la salida del amplificador de instrumentación, fig.4.

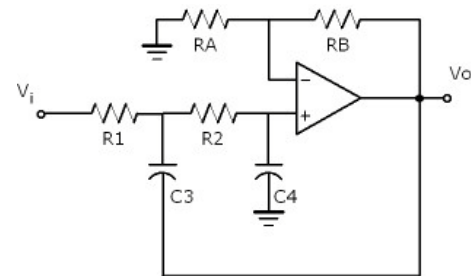


Fig. 4. Filtro paso bajas Butterworth, $f_c=40$ Hz.

Para calibrar los transductores se utilizó una columna de agua, implementada con una manguera transparente y marcando alturas equivalentes a 10 mm de Hg.

IV. SISTEMA DE CONTROL Y ADQUISICIÓN DE DATOS

El control electrónico está basado en un microcontrolador PIC16F877 [4] y fue programado en lenguaje C [5-6]. Este microcontrolador cuenta con un convertidor analógico-digital de 10 bits que se usa para adquirir las señales de los transductores de presión. El microcontrolador establece una comunicación serial RS-232 con una computadora personal para recibir los parámetros de la prueba y para enviar los datos del convertidor A/D.

Para mejorar la precisión del control de velocidad, el microcontrolador es el que genera la base de tiempo para controlar la adquisición de datos y para el envío de pulsos al motor de pasos.

Al inicio de la prueba, la computadora personal le envía al microcontrolador los siguientes parámetros:

- Número de muestras a adquirir
- Periodo de muestreo (en ms)
- Velocidad del motor y el ciclo de trabajo
- El nivel de corriente del motor
- El número de pasos a oscilar
- El sentido de giro

El microcontrolador utiliza un temporizador interno (timer1 16 bits) para generar una base de tiempo (500 μ s). Cada vez que se completa la base de tiempo se genera una interrupción. La rutina de atención a la interrupción del temporizador se encarga de actualizar las señales de control del motor de pasos. La adquisición de datos del convertidor A/D ocurre en múltiplos de la base de tiempo, de manera que cada vez que ocurre la interrupción se calcula si corresponde adquirir datos o no. La velocidad del motor también se controla cambiando el estado de los pulsos en múltiplos de la base de tiempo.

En la Fig. 5. se muestra de manera simplificada el flujo de la rutina de atención a la interrupción del temporizador. Cada vez que se ejecuta la rutina, se actualiza la base de tiempo del temporizador, si se cumple la cuenta de pasos correspondiente, se

manda la siguiente secuencia al motor de pasos, y si se cumple la cuenta de adquisición de datos, entonces se lee el convertidor A/D y se envía el resultado por el puerto serie hacia la computadora personal.

La prueba se da por terminada una vez que se han adquirido todas las muestras, o bien, se recibe el comando para abortar la prueba por el puerto serie. La recepción de un dato por el puerto serie mientras se realiza la prueba dispara una interrupción que es la encargada de ajustar las cuentas para que el motor se detenga al finalizar el siguiente ciclo. Esto es importante debido a que la posición inicial debe conservarse para poder repetir el volumen de líquido requerido en la prueba.

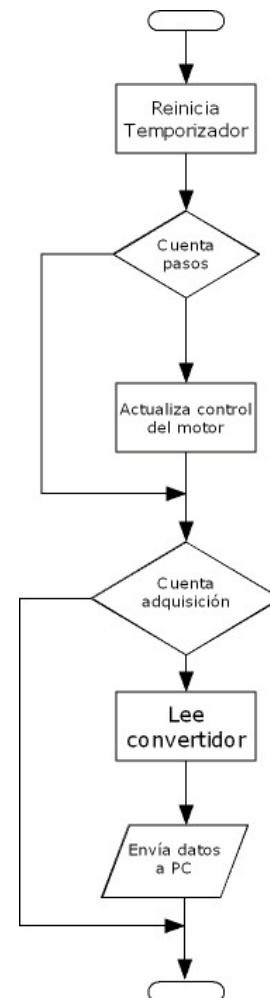


Fig. 5. Diagrama de flujo simplificado de la rutina de atención a la interrupción del temporizador (base de tiempo).

V. INTERFAZ DE USUARIO

La interfaz de usuario fue desarrollada en Visual Basic 6.0 [7] con la biblioteca de componentes Measurement Studio 7.1 de National Instruments. De esta manera se desarrolló una interfaz amigable y fácil de utilizar. Esta interfaz permite enviar los parámetros de configuración de la prueba al microcontrolador, graficar los datos provenientes de la misma en tiempo real, guardar los datos de la prueba y exportarlos en formato compatible con Excel.

En la Fig. 6, se muestra la ventana principal del programa, la cual incluye las ventanas de graficación de los sensores de presión. En la gráfica superior se despliegan las gráficas de la presión de entrada (en rojo) y la presión de salida (en blanco). La gráfica superior despliega la presión diferencial. En la parte superior de la ventana principal se encuentran los botones para:

- Controlar la adquisición de datos
- Controlar el funcionamiento del motor de pasos
- Controlar una electroválvula (versión futura)
- Ajustar el ciclo de trabajo
- Mostrar las gráficas de presión en tiempo real
- Iniciar la prueba

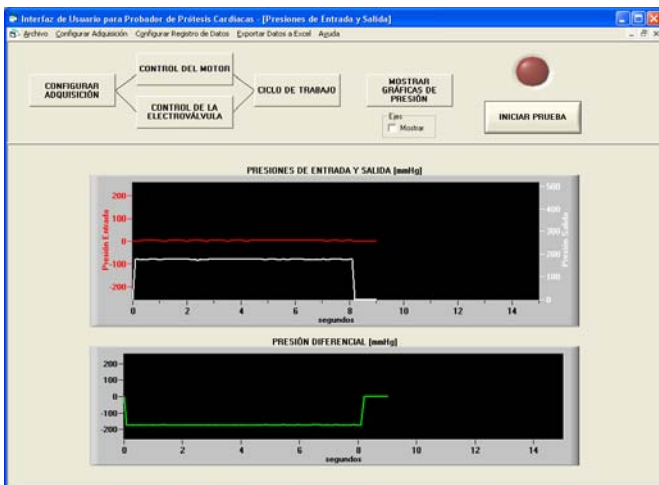


Fig. 6 Ventana principal de la interfaz de usuario.

En la Fig. 7. Se muestra la ventana de configuración de los parámetros de adquisición de datos. Mediante barras de deslizamiento se puede definir el número de datos a desplegar simultáneamente, así como el número de muestras por segundo. También se selecciona el puerto de comunicación serial.

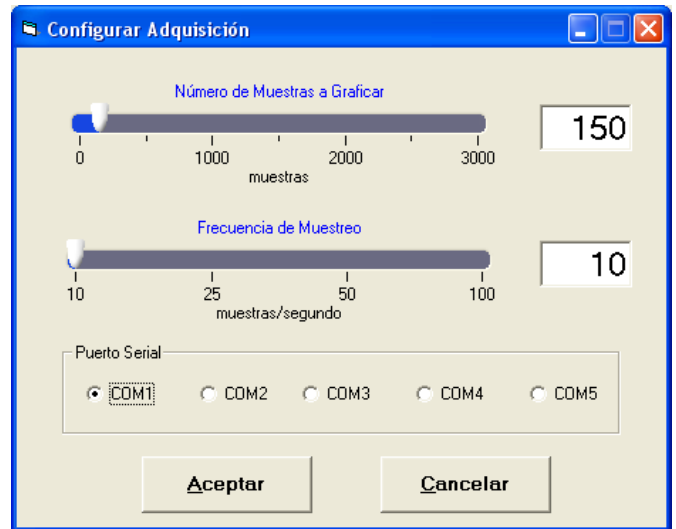


Fig. 7. Ventana para configurar los parámetros de adquisición de datos.

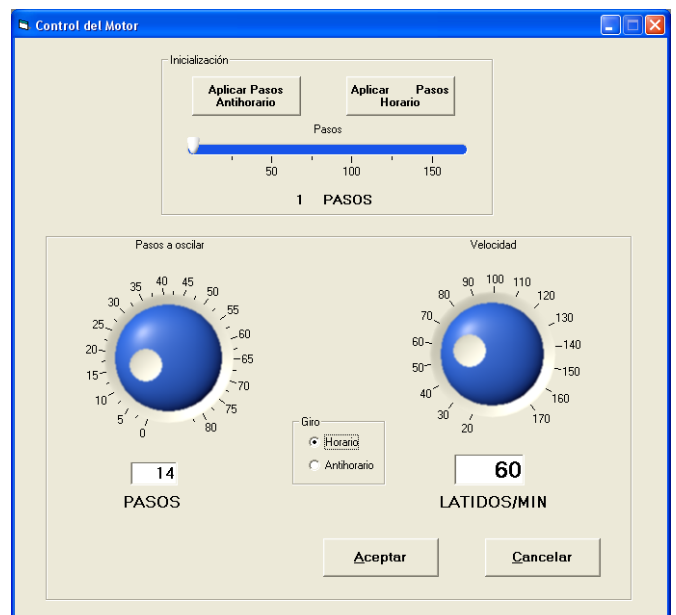


Fig. 8. Ventana de configuración del motor.

La ventana de configuración del motor se muestra en la Fig. 8. Cuenta con controles tipo perilla para ajustar el número de pasos a oscilar y la velocidad de la prueba en ciclos (latidos) por minuto.

También permite realizar un proceso de inicialización de la posición del pistón y seleccionar el sentido del giro.

Otro de los parámetros que se pueden ajustar es el ciclo de trabajo, el cual cuenta con tres opciones según se muestra en la Fig. 9.

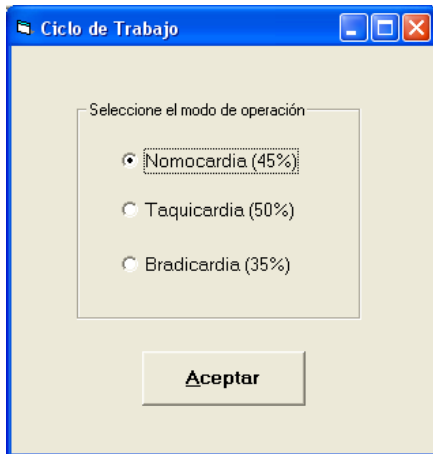


Fig. 9. Ventana para ajuste del ciclo de trabajo.

El sistema cuenta además con una función de registro de datos temporizado. De esta manera, se pueden guardar en un archivo grupos de datos cada 1, 5, 15, 30, 45, o 60 minutos; o bien, de manera continua, tal y como se muestra en la Fig. 10. Esto con la finalidad de no perder datos en el caso de una interrupción en el suministro de energía.

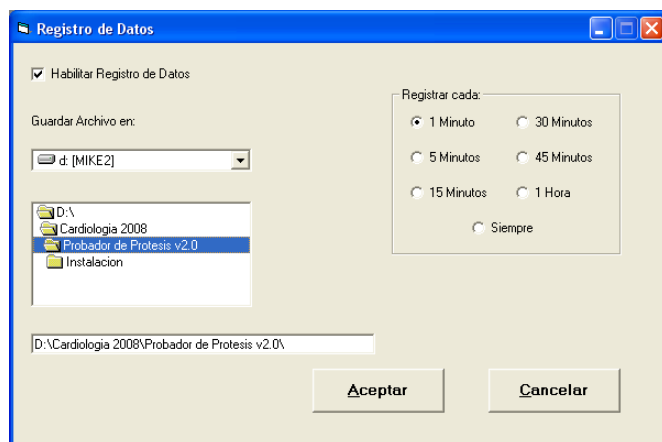


Fig. 10. Ventana de configuración del registro de datos.

Los registros se guardan en archivos de texto. En la Fig. 11 se muestra el encabezado de un archivo de ejemplo.

Tiempo [segundos]	Presiones Entrada	Presiones Salida	Diferencial
0.1	44.9219	50.7812	-5.8594
0.2	44.9219	50.7812	-5.8594
0.3	45.4102	51.5137	-6.1035
0.4	43.9453	51.0254	-7.0801
0.5	43.457	51.0254	-7.5684
0.6	43.9453	51.0254	-7.0801
0.7	46.3867	50.7812	-4.3945
0.8	44.9219	50.7812	-5.8594
0.9	44.9219	50.5371	-5.6152
1	44.9219	50.5371	-5.6152

Fig. 11. Ejemplo del archivo de datos.

VI. RESULTADOS

En la Fig. 12 se muestra una fotografía de las tarjetas electrónicas montadas en su gabinete, y en la Fig. 13 se puede observar el equipo instalado en el Laboratorio de Bioprótesis Cardíacas del Instituto Nacional de Cardiología.



Fig. 12. Interior del gabinete.



Fig. 13. Equipo instalado en el Instituto Nacional de Cardiología.

VII. CONCLUSIONES

Se ha desarrollado un sistema de control y adquisición de datos para un equipo probador de válvulas cardíacas. El equipo es capaz de configurar diferentes parámetros de la prueba, según lo requiere la norma oficial mexicana NOM-063-SSA1-1993. Este equipo representa una mejora con respecto al equipo que se venía utilizando en el Instituto Nacional de Cardiología, ya que este equipo es controlado desde una computadora personal, lo cual facilita su operación y posibilita el registro automático de datos.

Agradecimiento. Este trabajo fue financiado por el Fondo Sectorial CONACYT de Investigación en Salud y Seguridad Social. Desarrollo e innovación de prótesis cardíacas.

VIII. REFERENCIAS

- [1] *NOM-063-SSA1-1993 Que establece las especificaciones sanitarias de las válvulas cardíacas.* Secretaría de Salud, 1994.
- [2] *24PCBFA6G Pressure sensor product page.* Honeywell Sensing and Control en, Linear Technology en http://sensing.honeywell.com/index.cfm/ci_id/140301/la_id/1/pr_id/80889.htm

- [3] *TL071, TL072, TL074 Low-noise JFET-input Operational Amplifiers Datasheet,* Texas Instruments Inc. en <http://focus.ti.com/lit/ds/symlink/tl071a.pdf>
- [4] *PIC16F87X Data sheet,* DS3029, Microchip Technology Inc, 2001.
- [5] *PIC MCU C Compiler,* Reference Manual. Custom Computer Services Inc. 2003.
- [6] N. Gardner, *PICmicro MCU C. An introduction to programming de Microchip PIC in CCS C.* Bluebird Electronics, 2002.
- [7] F. Balena, *Programación avanzada con Microsoft Visual Basic,* McGraw-Hill, 2000.

IX. AUTORES

M en I. Miguel Angel Bañuelos Saucedo obtuvo su título de Maestría en Ingeniería Eléctrica con especialidad en electrónica en la Facultad de Ingeniería de la Universidad Nacional Autónoma de México en 1997. Sus intereses actuales son la instrumentación electrónica y los sistemas embebidos.

M. en I. José Castillo Hernández es Maestro en Ingeniería Eléctrica con especialidad en electrónica por la Universidad Nacional Autónoma de México, 1997. Actualmente trabaja en el desarrollo de sistemas electrónicos de potencia.

Ing. Jaime Monroy Elizondo es Ingeniero Eléctrico-Electrónico por la Facultad de Ingeniería de la Universidad Nacional Autónoma de México, 2006.

DISEÑO Y CONSTRUCCIÓN DE UN SISTEMA CON BASE EN UN MÓDULO XBEE APLICADO EN EL MONITOREO INALÁMBRICO DE DATOS

Jayro Santiago Paz, F. Hugo Ramírez Leyva, Alejandro E. Ramírez González

Instituto de Electrónica y Computación
Universidad Tecnológica de la Mixteca (UTM)
Carretera a Acatlima Km. 2.5 Huajuapán de León, Oaxaca, C.P. 69000, Tel. (01953) 5320399 Ext. (200)
Correo electrónico: jayro.s.p@gmail.com, hugo@nuyoo.utm.mx, ocetxim@mixteco.utm.mx

RESUMEN

En el presente trabajo se muestra el diseño y construcción de un sistema para estudiar el transceptor XBee de la firma Digi International®, que cumple con el estándar IEEE 802.15.4. El sistema consta de dos módulos independientes, a uno de ellos se le pueden conectar hasta cuatro sensores con niveles TTL, así como una entrada y una salida analógica en el rango de 0 a 5V. El segundo módulo tiene la capacidad de conectarse al puerto USB de una computadora personal (PC). Las pruebas de alcance realizadas con ambos módulos determinaron que en línea de vista se tiene un alcance de hasta 100m lo cual es útil para el monitoreo inalámbrico de sensores y/o hacer una red de recolección de datos.

Palabras clave: Sensores inalámbricos, XBee, Redes de instrumentación.

I. INTRODUCCIÓN

En la actualidad, para conocer el estado que guarda una red industrial se requiere de una gran cantidad de sensores, los cuales monitorean el estado de la red y transmiten información hacia el centro de control, en donde es procesada y con base en ella ejecuta la acción de control a realizar [1].

Los sensores de presión, temperatura, humedad, etc., normalmente proporcionan información a una velocidad muy baja. Cuando se requiere transmitir datos hacia una estación central, se tiene dos

opciones: tender cables o utilizar una conexión inalámbrica, para realizar dicha transmisión.

El desarrollo de los sensores inalámbricos es relativamente nuevo. Esto se debe a los avances logrados en la microelectrónica, la computación y las telecomunicaciones. La miniaturización de los componentes electrónicos ha permitido diseñar circuitos que sean capaces de procesar información digital y/o analógica, además transmitirla en ondas de radiofrecuencia en módulos pequeños, que pueden ser fácilmente colocados en espacios reducidos [2].

En el mercado existe una gran variedad de tipos de redes para la adquisición de datos y transmisión de información en forma inalámbrica. Entre las más importantes se pueden mencionar: RFID y ZigBee para el control de sensores [3], Bluetooth para voz y datos, y WiFi para redes locales inalámbricas [4, 5], entre otras.

En la figura 1 se muestra una gráfica con las diferentes tecnologías para comunicaciones inalámbricas, el eje de las abscisas corresponde a la velocidad de transmisión y el eje de las ordenadas a los diferentes tipos de redes: redes de área personal inalámbricas (WPAN), redes de área local inalámbricas (WLAN), redes de áreas metropolitanas inalámbricas (WMAN) y redes de áreas amplias inalámbricas (WWAN). Para las aplicaciones con sensores se usan redes del tipo WPAN.

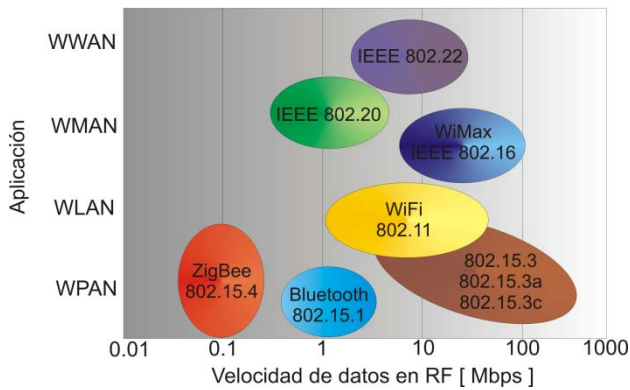


Fig. 1. Tecnologías en la banda de 2.4 GHz.

Cuando es necesario enviar una gran cantidad de información, la alternativa es WiFi o Bluetooth. Sin embargo, para una velocidad de transmisión baja, las alternativas tecnológicas son: RFID y IEEE 802.15.4 (comúnmente denominado ZigBee). El primero es usado para el control de acceso e inventarios; mientras que el estándar IEEE 802.15.4 define el protocolo para la interconexión de módulos inalámbricos de bajo consumo de energía, tasas de transferencia de datos bajas, y facilidad de instalación de los dispositivos, pero con un alcance intermedio (aprox. 30m) [5]. Este estándar es el que se estudiará en el presente trabajo.

El estándar IEEE 802.15.4 es una tecnología inalámbrica para tasas bajas de envío de datos, usa la banda libre ISM (Industrial, Scientific and Medical) de 2.4 GHz, con una velocidad de transferencia máxima de 250 Kbps, modulación M-ario cuasi-ortogonal con M=16 y un alcance aproximado de 30m. Las topologías de red que soporta el estándar es la estrella y malla (figura 2), y cada red puede soportar hasta 65535 nodos distribuidos en subredes de 255 nodos [8].

Los componentes de la red de este tipo se pueden clasificar por su funcionalidad en: Dispositivo de función completa (FFD, *Full-Function Device*) y Dispositivo de función reducida (RFD, *Reduced-Function Device*). El primero puede funcionar como coordinador, router o PAN (*Personal Access Network*) y el segundo es el último eslabón en la red. El FFD puede funcionar como RFD pero éste no puede ser un FFD [8]. Véase la figura 2. El

PAN es el encargado de definir las funciones de los distintos componentes de la red.

El estándar IEEE 802.15.4 define 27 canales en las tres bandas de frecuencia utilizadas, las cuales se distribuyen de la siguiente manera: 1 canal en la banda 868 MHz, 10 canales en la banda de 915 MHz y 16 canales en la banda de 2.4 GHz.

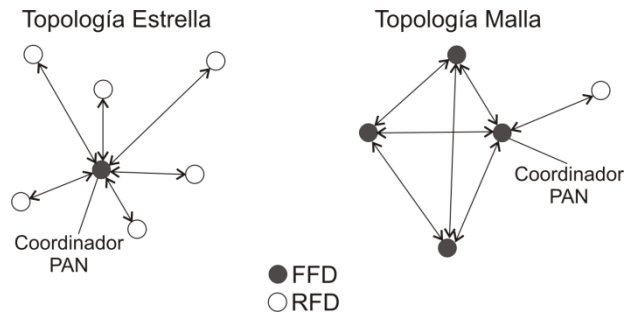


Fig. 2. Topologías de red, soportadas por el estándar IEEE 802.15.4

Entre las principales características del estándar se encuentran [1]:

Cada nodo se le puede asignar una dirección corta de 16 bits (para redes de hasta 65536 dispositivos), o extendidas de 64 bits (para usar direcciones IEEE MAC). La transferencia de datos es de tipo espectro disperso, con asignación opcional de ranuras de tiempo garantizado (GTS, *Guaranteed Time Slot*).

Actualmente se cuenta con una alianza de empresas denominada ZigBee que promueve el desarrollo de esta tecnología, certificando además los productos generados por las empresas pertenecientes a esta alianza. Existen varias empresas que comercializan módulos transceptores que cumplen con parte del estándar y/o productos que hacen uso de ellas, entre las cuales están: Atmel®, Dallas Semiconductors®, Maxim® y Digi International®, entre otras.

Digi International fabrica una familia de transceptores a la cual le denominó XBee, los cuales cumplen con el estándar ZigBee. También distribuye kit's de desarrollo con los cuales se puede empezar a probar el estándar en forma

inmediata, con un precio desde los \$130 US. Dlls [9]. La mayoría de estos kits se conectan vía USB y/o RS-32 a una PC. Estos kits se alimentan directamente del puerto de comunicaciones o con una fuente de voltaje externa. En la figura 3 se muestra una fotografía de ellos.



Fig. 3. Kit de desarrollo de la firma Digi

El objetivo de este trabajo fue construir dos módulos. Uno para el control de sensores con salida TTL, alimentado por pilas o una fuente de voltaje externa, y un segundo módulo para comunicarse con el primero desde una PC. Esto con el fin de probar el rendimiento de los módulos XBee y probar, en forma experimental, algunas de las características del estándar 802.15.4.

II. DESCRIPCIÓN GENERAL DEL SISTEMA

Se desarrolló un sistema para el manejo de sensores digitales y/o analógicos así como una salida digital, que transfiera el estado de las entradas hacia una PC y/u otro módulo que reproduzca la entrada y/o salida como si fuera por un cable. Para el envío y recepción de los datos en forma inalámbrica se usó el circuito integrado XB24-AWI-001 de la firma MaxStream [10]. Las principales características del sistema son:

- Cuatro entradas o salidas digitales con nivel TTL.
- Una entrada analógica en el rango de 0 a 5V.
- Una salida analógica en el rango de 0 a 5V.
- Puerto USB opcional para la conexión de este sistema hacia la computadora.
- Alimentación con un par de pilas tipo AA y de una fuente de voltaje.
- Indicador visual del nivel de RF en la recepción.
- Monitoreo desde una computadora personal.

Para cumplir con los requerimientos se dividió el sistema en dos bloques independientes, uno de ellos servirá cuando se conecte el sistema al puerto USB (que se le denominará XBeeUSB) y otro para el manejo de sensores remotos (el cual se llamará XBeeSensor). En la figura 4 se muestra la forma en que se interconectan.



Fig. 4. Diagrama conceptual del XBeeUSB y el XBeeSensor

El XBeeUSB va a funcionar como coordinador PAN y el XBeeSensor será alimentado por pilas, puede ser usado para monitorear sensores remotos o como repetidor de comunicaciones. Este módulo será explicado en la sección III.

III. SISTEMA XBEESENSOR

Es el encargado de manipular sensores de tipo TTL cumpliendo todas las características del protocolo IEEE 802.15.4. Para hacerlo portable puede ser alimentado por una fuente de voltaje externa o pilas. Este sistema se dividió en seis bloques como se muestra en la figura 5: Módulo XBee, Entradas digitales, Salidas Digitales, Entradas Analógicas, Salidas Analógicas, Indicador de Nivel y Fuente de Alimentación. A continuación se describen cada uno de ellos.

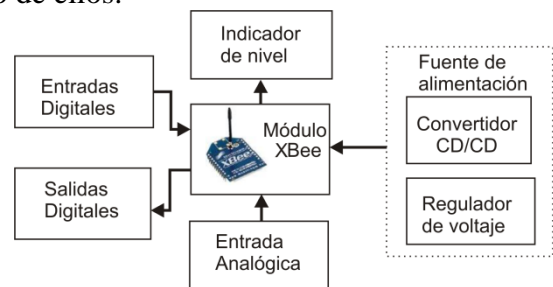


Fig. 5. Diagrama a bloques del sistema XBeeSensor.

Módulo XBee

Es el módulo más importante, ya que es el encargado de recibir y transmitir los datos cumpliendo las normas del estándar IEEE

802.15.4. Los demás bloques sirven de soporte para que este dispositivo trabaje correctamente. El circuito integrado usado es el XB24-AWI-001 de la firma MaxStream por su bajo consumo de energía.

Existen modelos con diferentes capacidades de potencia de RF (desde 1mW a 63mW) en la banda de 2.4 GHz. Se le puede configurar el baud rate en un rango de 1200 a 115200 bps. Tiene modelos con antena de chip, dipolo o con conector para una antena externa (para mejorar el alcance).

El módulo XBee tiene 20 terminales, de las cuales ocho son para entradas o salidas digitales, 4 son de propósito general y 3 para el manejo de la interfaz de comunicaciones seriales. Usa 2 terminales para el voltaje de alimentación de 3.3V. También posee un generador de Modulación por Ancho de Pulso (PWM) con la cual se genera una señal analógica a partir de una salida digital. Internamente tiene un convertidor Analógico Digital (ADC), lo cual permite que directamente se le puedan conectar sensores de tipo analógico, sin necesidad de usar un ADC externo. Estas características hacen posible que el XBee pueda manejar directamente sensores o que funcione como radio módem.

En la figura 6(a) se muestra el esquemático del módulo XB24-AWI-001, el cual posee una antena de chip en la parte superior derecha de la figura. En la figura 6(b) se muestra el diagrama a bloques del módulo XBee, el cual se divide en cinco bloques que son: La antena, El interruptor de antena, El transmisor, El receptor y El procesador.

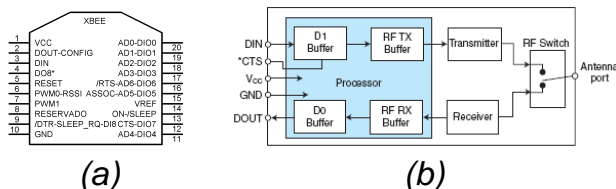


Fig. 6. a) Símbolo del esquemático del CI XBee. b) Diagrama a bloque del control de flujo.

Cuando funciona normalmente puede trabajar como MODEM o se le puede configurar las diferentes opciones (baud rate, la potencia de transmisión, etc.) usando comandos AT (usados

para MODEM. También con estos comandos se lee o se pone el estado de las entradas o salidas digitales, y el valor del ADC.

Entradas y salidas digitales

El módulo de entradas digitales permite conectar entradas digitales con niveles TTL al módulo XBee. Cuando ocurre un cambio de estado en sus entradas, el evento es transmitido hacia la computadora personal o hacia un dispositivo final.

El módulo de salidas digitales acondiciona los voltajes de salida del XBee (3.3V) a TTL. El estado de la salida es función del comando recibido por el módulo XBee, ya sea por una computadora o un módulo remoto.

Como la misma Terminal se puede configurar como entrada o salida, se necesita acondicionar los niveles de voltaje y seleccionar el tipo en forma externa. Para cambiar el nivel de voltaje se usó un arreglo de transistores Darlington en corte y saturación alimentado con 5V y 3.3V (figura 7). La selección de entrada o salida se hace con el conector J1 que es un dipswitch.

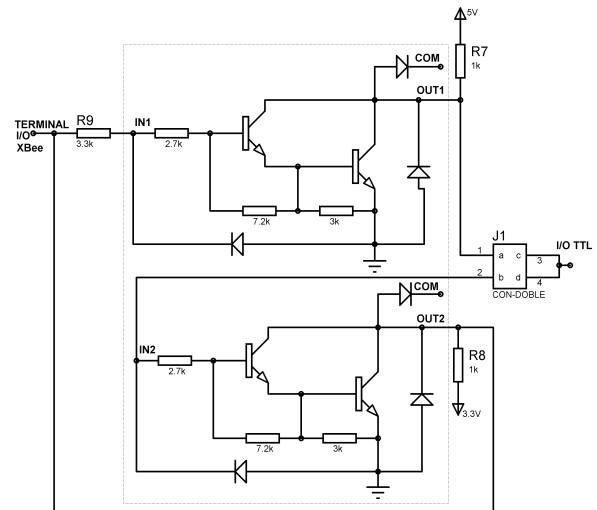


Fig. 7. Arreglo de transistores para las entradas/salidas digitales.

Para minimizar el espacio se usó el circuito integrado ULN2803A, el cual tiene 8 transistores darlington, que son los que se necesitan, ya que cada entrada salida ocupa 2. La ganancia de corriente es de al menos 10000, por lo que con la

resistencia de base de $2.7k\Omega$ se garantiza que se satura con un voltaje superior a 1.5V.

Entrada y salida analógica

Entradas analógicas. Este módulo acondiciona la señal analógica de 0 a 5V al rango de alimentación del convertidor analógico digital del módulo XBee (de 0 a 3.3V), para que el valor digitalizado de la señal analógica sea transmitido.

Para tener una alta impedancia de entrada se utilizó un divisor de voltaje con resistencias de $1.17M\Omega$ y $2.15M\Omega$, de tal manera que cuando el voltaje de entrada sea de 5V a su salida se tiene 3.3V, posteriormente se le pone un Amplificador Operacional (OPAMP) en configuración seguidor. El OPAMP usado es el LM358, que funciona con una fuente de voltaje y posee 2 OPAMPS en su encapsulado SO8. En la figura 8 se muestra la conexión de la entrada analógica, que usa las terminales 19 y 14 del XBee.

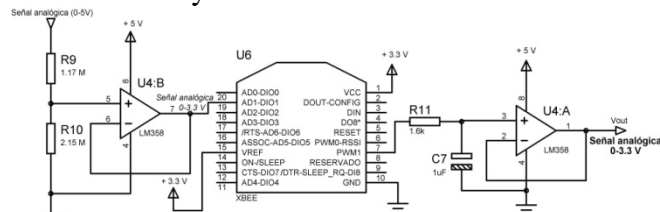


Fig. 8. Diagrama de conexión de la entrada y salida analógica

La salida analógica se genera con uno de los 2 moduladores de PWM que posee el XBee (terminal 7). De tal manera que el valor convertido del ADC se envía hacia el módulo que tiene configurada la salida y genera el modulador de ancho de pulso para conseguir el voltaje deseado. Para obtener un voltaje de corriente directa (CD), se usa un filtro pasa bajas de tipo RC y un amplificador seguidor para tener baja impedancia de salida (figura 8).

Fuente de Alimentación

Es la encargada de proporcionar los voltajes de 5V para las etapas de entrada y salida, y 3.3V para el módulo XBee. La entrada proviene de una fuente de voltaje de 6 a 12V y/o un arreglo de pilas de

2.4V o mayor. En la figura 9 se muestra el diagrama a bloques de este módulo.

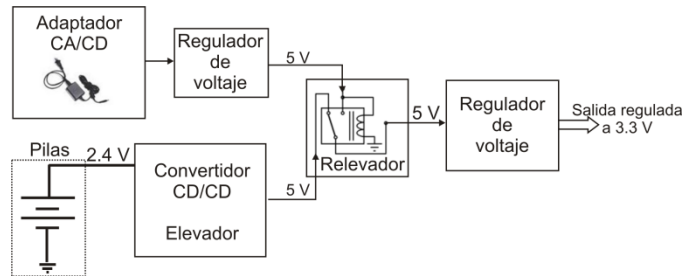


Fig. 9. Diagrama a bloques del módulo de alimentación.

El voltaje externo proviene de un eliminador de pilas. Para generar el voltaje a partir de pilas se utiliza un convertidor CD/CD, para conmutar la fuente se utiliza un relevador de voltaje, cuya bobina es alimentada por la fuente de voltaje externa. Si el relevador no está energizado, la alimentación la proporciona el 5V convertidor, en caso contrario, alimenta la fuente de voltaje externa. Para generar el voltaje de 3.3V que alimenta al XBee se utiliza el regulador de voltaje LF33CV con capacidad de corriente de 500mA.

El convertidor CD/CD usado es el CI MAX1674EUA con encapsulado μ MAX, de montaje superficial de la firma MAXIM. Tiene la capacidad de generar voltajes de 3.3V ó 5V a través de una de sus ocho terminales. En este caso se configuró para que generara 5V, con una capacidad de corriente de 1A. Para su operación requiere de algunos capacitores y un inductor externo de $22\mu H$.

Se caracterizó para ver su funcionamiento ente variaciones del voltaje de las pilas y con diferentes resistencias de carga ($R_L =$ sin carga, 470Ω , 100Ω , 55Ω y 22Ω). En la figura 10 se muestra la respuesta obtenida. Como se puede ver, cuando el voltaje de entrada es mayor a 2V el voltaje de salida es de 5V, lo cual garantiza que con 2 pilas recargables de 1.2V, es suficiente para que funcione correctamente. Cuando el voltaje es bajo, la eficiencia decae mucho.

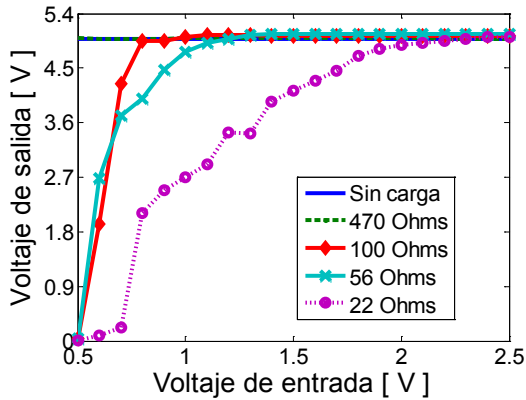


Fig. 10. Voltaje de salida vs voltaje de entrada del MAX1674EUA

Indicador de nivel

La norma del estándar IEEE 802.15.4 especifica que se debe tener un indicador visual para verificar la calidad del enlace de RF. Los módulos XBee tienen una terminal con una salida PWM, cuyo ancho de pulso es función del nivel de energía del último paquete de RF recibido (la terminal seis). Cuando esta señal se hace pasar por un filtro pasabajas, se tiene el voltaje promedio de la señal de PWM. Usando comparadores de voltaje y tres LEDs se tiene el indicador visual del nivel de RF.

El tiempo en alto del PWM del módulo XBee va de 24% a 100% dependiendo del nivel detectado. Un valor inferior al 24% indica que el nivel detectado está por debajo del nivel de sensibilidad del módulo. Si la salida del PWM es cero significa que el PWM se encuentra inactivo.

El circuito integrado que se utilizó para la implementación de los comparadores de voltaje, es el LMV339MTX de la marca National Semiconductor, que cuenta con cuatro comparadores y 14 terminales, se alimenta con un voltaje de 3.3V. Se eligió este integrado por tener el mismo voltaje de alimentación que el módulo XBee. En la figura 11 se muestra el diagrama esquemático de este módulo con el XBee.

En las figuras 12 y 13 se muestran el diagrama esquemático y la placa completa del XBeeSensor.

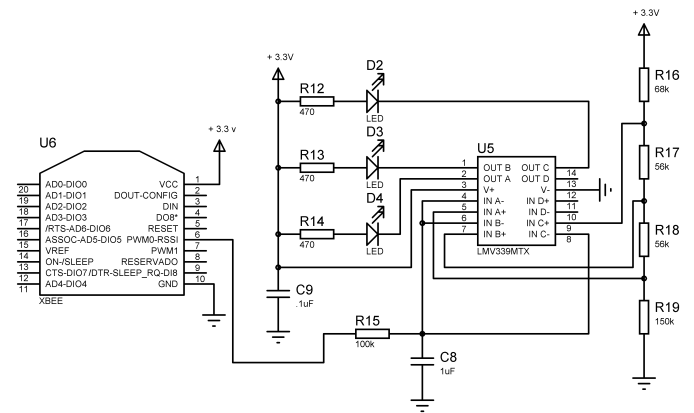


Fig. 11. Diagrama esquemático del detector de nivel de RF

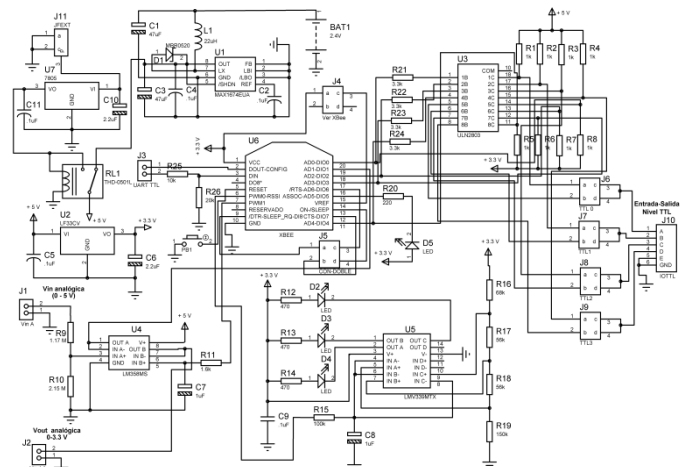


Fig. 12. Diagrama esquemático del XBeeSensor

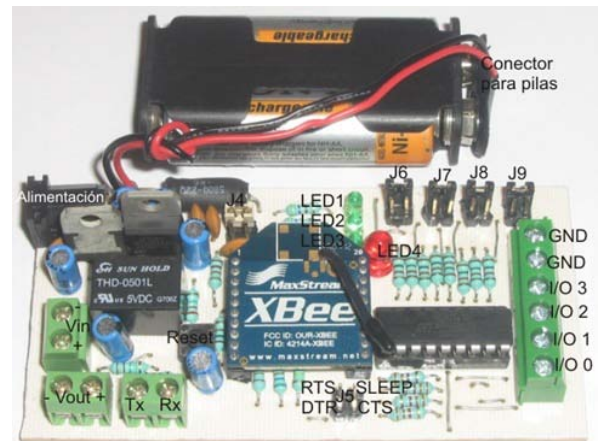


Fig. 13. Placa del sistema XBeeSensor.

Como se puede observar, se alimenta con un par de pilas AA y el conector ‘alimentación’ es el plug por donde se le aplica la fuente de voltaje externa. Se utilizaron bornes con tornillos para conectar las

entradas salidas y la comunicación serial. La placa tiene una base para el módulo XBee, para facilitar su configuración, ya que debe ser programado desde una PC antes de que pueda funcionar adecuadamente.

IV. MÓDULO XBEEUSB

Este sistema tiene la capacidad de transmitir los datos que le lleguen por el puerto USB de una PC, con el fin de dotar de este tipo de interfaz a una o varias PC's y de esta forma generar una red IEEE 802.15.4 con varios componentes. Conceptualmente se divide en 3 bloques que son: El Módulo XBee, El convertidor USB-Serial y los Indicadores luminosos.

El módulo XBee, como se explicó en la sección anterior, es un transceptor bajo el protocolo IEEE 802.15.4. Envía los datos que recibe por RF a su puerto serial, al cual se le conecta el convertidor USB-Serial y de ahí a la PC.

Los indicadores luminosos permiten al usuario conocer de una manera rápida el estado del sistema XBeeUSB. Indicarán si el sistema XBeeUSB está conectado al puerto USB, si se está realizando una transmisión de datos entre la PC y el convertidor de nivel, y también indicarán si el módulo XBee conectado está asociado a una red.

El Convertidor USB-Serial recibe las tramas seriales del módulo XBee, y las envía hacia una computadora en formato USB, permitiendo de esta manera establecer la comunicación con la computadora, además, de este mismo puerto, toma el voltaje de alimentación para que funcione.

El componente principal del módulo Convertidor USB-Serial es el dispositivo FT232BL de la empresa FTDI. El circuito integrado FT232BL tiene 32 terminales y convierte los niveles USB a formato serial. Tiene la capacidad de funcionar con voltajes de 5V ó 3.3V, este último fue el usado en esta aplicación.

Para que el FT232BL sea reconocido por el sistema operativo de la computadora, se requiere de un driver. Para el sistema operativo Windows XP y posteriores, se obtiene gratuitamente de la página web del fabricante [11].

El circuito integrado es compatible con USB 1.1 o USB 2.0. Una vez que se instala el controlador en la computadora, ésta reconoce al FT232BL como un puerto virtual de comunicaciones seriales, de tipo COM. El voltaje del puerto USB de la computadora es de 5V, por lo cual se utilizó un regulador de voltaje para obtener los 3.3V que necesita el dispositivo XBee.

En la figura 14 se muestra el diagrama esquemático del XBeeUSB, y en la figura 15 la placa del circuito construido.

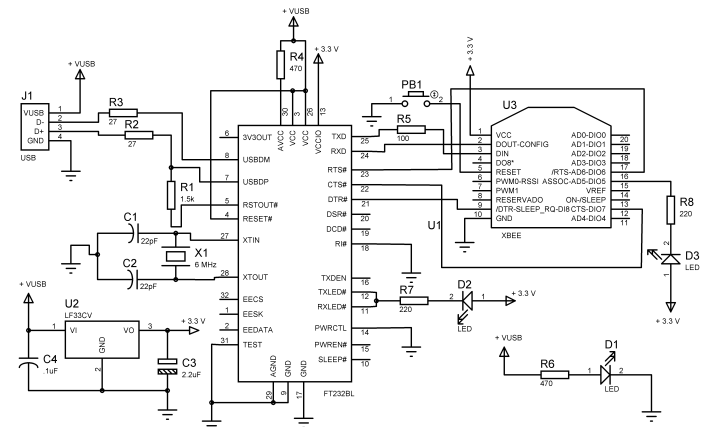


Fig. 14. Diagrama esquemático del XBeeUSB

El sistema XBeeUSB se conecta con la computadora personal mediante el conector USB tipo B. Los componentes marcados en la figura 15 son: El interruptor "Reset" sirve para reiniciar al módulo en caso de requerirse; el LED1 se enciende cuando el sistema XBeeUSB esté conectado en un puerto USB de una computadora; el LED2 se ilumina cuando el módulo XBee conectado al sistema se encuentre asociado a una red; el LED3 se activa cuando se realiza una transferencia de datos entre la computadora y el manejador de USB.

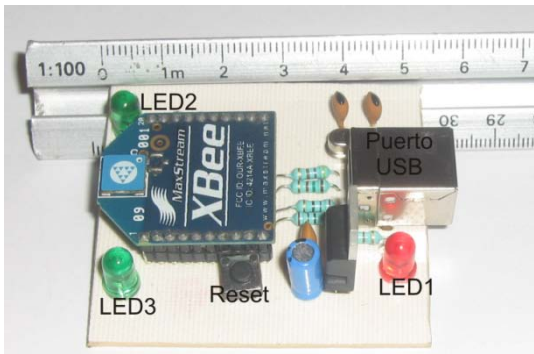


Fig. 15. Placa del sistema XBeeUSB

V. PRUEBAS Y RESULTADOS

Para verificar el correcto funcionamiento del sistema llamado XBeeUSB y XBeeSensor se hicieron pruebas de conexión, de alcance, pruebas de entradas y salidas analógicas y digitales.

Para ver la tasa de error del XBeeUSB se fabricaron 2 placas de este sistema, las cuales se conectan al puerto USB de 2 computadoras, en las cuales se ejecuta un programa realizado en Microsoft Visual Basic 6.0, el cual transmite y recibe una trama de 100 caracteres, la PC transmisora envía la trama, que es capturada por el receptor y a su vez los reenvía a la PC inicial, esto a diferentes velocidades. La distancia entre módulos fue de 10m. Se probaron velocidades de 9600 bps hasta 115200 bps, obteniendo en todos los casos una tasa de error del 0%.

Otra prueba realizada fue verificar la calidad del alcance a diferentes distancias. Para esta prueba se utilizó un XBeeUSB conectado a una PC y un XBeeSensor como repetidor (puenteadas la terminal de transmisión y recepción).

Se uso el programa llamado X-CTU el cual se obtiene gratuitamente de la página de Digi Internacional [12], con la opción “Range test” utilizando el indicador de nivel RF recibido (en dBm) que viene con el programa. Cuando se entra en este modo, se le configura al programa una trama de caracteres a enviar (en este caso fue de 20 caracteres), las envía hacia un módulo remoto, el cual la regresa. Compara la trama enviada y la

recibida y muestra en número de intentos hecho y los recibidos correctamente. Esta prueba se con línea de vista entre ambos módulos.

En la tabla 1 se muestran los resultados a diferentes distancias. Como se puede ver, a 50m el nivel de RF fue de -79dB y la tasa de error del 0% (de un total de 50 intentos), mientras que a 120m la tasa de error fue del 18%. Se hizo una prueba en interior a una distancia de 16.7m con una pared de ladrillos en medio, la tasa de error fue del 0% con un nivel en recepción de -70dBm (igual que la obtenida a 30m con línea de vista).

Tabla 1. Datos de la prueba de alcance

Distancia [m]	Tasa de error [%]	Nivel [dBm]
10	0	-63
20	0	-66
30	0	-70
40	0	-76
50	0	-79
120	18	-94

El módulo XBeeSensor puede trabajar como si fuera un módulo maestro esclavo, es decir a un módulo se le configura la entradas/salidas digitales como entrada y a otro como salidas, de tal manera que cuando en el módulo maestro ocurre un cambio de estado en las entradas, automáticamente se transfiere el estado de ésta hacia el módulo esclavo y cambia su estado. Para esta prueba se fabricaron 2 de estos módulos.

Para medir el tiempo de retardo con un generador de funciones (configurado con una señal de tipo TTL y una frecuencia de 10Hz), se conecta una de las 4 entradas digitales del módulo maestro. Con un analizador de estados lógicos (el E9340A de Agilent Technologies) se mide la entrada y la salida en los módulos XBeeSensor. Véase la figura 16. Posteriormente la misma señal del generador se conecta a las 4 entradas y se median las 4 salidas.

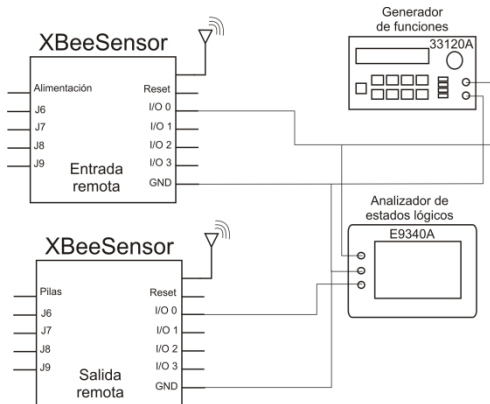


Fig. 16. Figura 9. Prueba de las entradas-salidas digitales.

Cuando sólo estaba conectada una entrada, el tiempo de retardo entre la entrada y la salida fue de 4.922ms, pero cuando se le conectaba a las 4 entradas, el retardo de la 4 salidas fue de 10.397ms. Este tiempo es el que tarda ambos módulos en leer la entrada, codificarla, enviarla serialmente por RF, decodificarla y poner la salida.

Siguiendo el mismo concepto de que un módulo XBeeSensor es el módulo maestro y otro el esclavo, se probaron la entrada y salida analógica. Con una fuente de voltaje (E3646A de Agilent Technologies) se le aplicó un voltaje de 0 a 5V con incrementos de 0.2V. La salida se mide con un osciloscopio (el 54621A de la misma firma). En la figura 17 se muestran los voltajes de entrada vs el voltaje de salida, así como el voltaje teórico esperado.

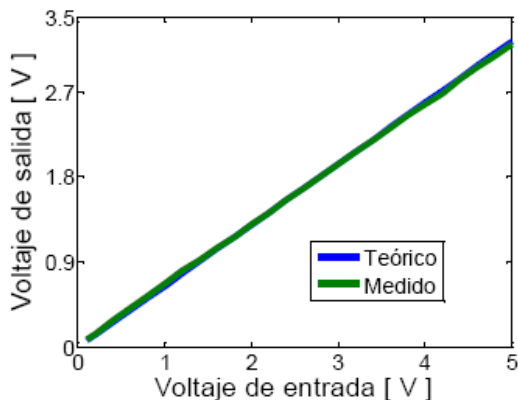


Fig. 17. Resultado de la prueba de la entrada-salida analógica.

Cuando el voltaje de entrada esta en el rango de 0.1V a 1V el error esta en el rango de 32% al 4%.

De 1.2V a 2V el error es de 4% a 0.2%, mayor a 2.2V el error es prácticamente despreciable. Las diferencias se deben a que se usó un divisor resistivo para hacer la reducción del voltaje a la entrada del ADC, y se usaron resistencias de 10% de precisión.

VI. CONCLUSIONES

En el presente trabajo se mostró el desarrollo de un sistema para analizar módulos que cumplen con el estándar IEEE 802.15.4, con el fin de estudiar algunas de las características de este protocolo. El módulo comercial usado es el XBee de la firma Maxstream, que es de bajo costo, se probó que es relativamente fácil su uso, aunque hay varias cuestiones del estándar que no tiene habilitadas completamente. Por ejemplo que no se le puede configurar el canal de RF a usar.

Se desarrollaron con sistemas a los cuales se les llamó XBeeUSB y XBeeSensor. El primero permite habilitar a una PC con este tipo de enlace de comunicaciones, con el fin de probar los módulos remotos, además de que en un futuro hacer más módulos que permitan probar más características del estándar con varios nodos.

El sistema XBeeSensor permite trabajar con sensores con salida digital de tipo TTL, lo cual puede ser usado para detectores de presencia o humo. El que pueda trabajar con pilas lo hace muy versátil y facilita su instalación. Se trabajó y caracterizó un convertidor CD/CD comercial, y se observó que funciona muy bien y es relativamente fácil usarlo. El principal problema para hacerlo funcional fue el conseguir la inductancia requerida.

A pesar de que los módulos XBee no cuentan con salida analógica, se utilizó una salida digital con modulación PWM. Por lo cual una contribución de este trabajo es el haber usado este tipo de salida para generar una señal continua, sin la necesidad de usar un convertidor digital analógico.

Algunos trabajos futuros que se pueden desprender de este trabajo son usar el módulo para automatizar

alguna tarea, como puede ser el control de los parámetros de un invernadero, y/o aplicaciones de seguridad para el control de sensores de presencia inalámbricos, lo que facilitaría enormemente la instalación de sistemas de vigilancia y control de acceso. También sería conveniente interconectar el sistema XBee a una red de área local, con el fin de ampliar las capacidades de este tipo de dispositivos.

VII. REFERENCIAS

- [1] “Automatización industrial”, Wikipedia, en:
<http://es.wikipedia.org/wiki/Automatizaci%C3%B3n>
- [2] López, M., Gutiérrez, M. Marcelín, R.: “*De sensores inalámbricos, mitigación de riesgos y sistemas autoorganizados*”. Contacto S, Número 64 Abril-Junio 2007, UAM-Iztapalapa.
- [3] Roger Smith, *RFID: A Brief Technology Analysis*, CTOnet.org, 2004
- [4] “Tecnología Bluetooth”, Bluetooth, en:
<http://spanish.bluetooth.com/Bluetooth/Tecnology/>
- [5] “Tecnología wi-fi”, wi-fi, en:
http://www.wifi.org/knowledge_center_overview.php
- [6] Honeywell, *Wireless Gauge and Absolute pressure transmitters*, Honeywell, USA 2004
- [7] *Organización Zigbee*, Wikipedia, en:
<http://es.wikipedia.org/wiki/ZigBee>
- [8] IEEE Computer Society, “*IEEE Standard for Information technology – Telecommunications and information exchange between systems. Local and metropolitan area networks. Specific requirements. Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs)*”, USA 2006.
- [9] *Alianza ZigBee*, Página de ZigBee en:
<http://www.zigbee.org/en/>
- [10] MaxStream, *Product Manual v1.xAx, Xbee/Xbee PRO 802.15.4 OEM RF Modules v1.xAx*, USA 2006.
- [11] *Drivers usb serial*, Página del fabricante ftdichip:
<http://www.ftdichip.com/Drivers/VCP.htm>
- [12] *Digi internacional*, Página del fabricante Digi:
<http://www.digi.com/products/wireless/point-multipoint/xbee-series1-module.jsp>

IMPLEMENTACIÓN DE UN SISTEMA DE MEDICIÓN ULTRASÓNICO DE DISTANCIA UTILIZANDO PROGRAMACIÓN EN LABVIEW SOBRE FPGA.

Vargas Zamorano Jorge, Sifuentes Mijares Juan, Aguilar Torres Anna.

Departamento de Posgrado e Investigación del Instituto Tecnológico de la Laguna.
Boulevard Revolución y Calzada Cuauhtémoc, Torreón, Coahuila, C.P. 27000, México.
Tel. 8717051331 Ext130, correo electrónico: juansm@itlalaguna.edu.mx.

RESUMEN

En este trabajo se presenta el desarrollo e implementación de un sistema para la medición de distancia por medio de sensores ultrasónicos. Se ha empleado para su implementación la tarjeta SPARTAN 3E y el programa LabVIEW. En esta aplicación se muestra un ejemplo de la versatilidad de los sistemas con FPGA, así como su facilidad para desarrollo de sistemas digitales, a la vez siendo potencializado por el uso de un sistema de programación de ambiente gráfico como es el LabVIEW.

Palabras clave: Lenguaje G, FPGA, LabVIEW, Instrumento virtual VI.

I. INTRODUCCIÓN

El nombre FPGA viene de su acrónimo en inglés “**Field Programmable Gate Array**”, que se traduce al español como “**Arreglo de Compuertas Programables en Campo**”. Los dispositivos de lógica programable surgen a partir de los años 70’s como una respuesta a las crecientes necesidades de los diseñadores de dispositivos digitales para integrar cada vez más compuertas en un diseño. Un FPGA es un dispositivo semiconductor integrado en un solo chip, el cual está basado en muchos componentes digitales básicos que se pueden interconectar entre ellos para satisfacer un diseño requerido [1].

Existe software especializado que define las conexiones de los interruptores y las

Proyecto apoyado por “DGEST” Dirección General de Educación Superior Tecnológica, y “CONACYT” Consejo Nacional de Ciencia y Tecnología.

funciones de las celdas lógicas de un FPGA tales como el lenguaje de descripción de hardware llamado “VHDL”, o el creciente en popularidad “Verilog” y también el “ABEL” Lenguaje de Expresiones Avanzadas (Advanced Boolean Expression Language). Algunas otras compañías han incorporado sus lenguajes y formas de programación al diseño con FPGA. [2] Como es el caso de LabVIEW “Laboratory Virtual Instrument Engineering Workbench”, el cual es un lenguaje de ambiente gráfico conocido como lenguaje G [3].

Nuestro diseño se centra en el uso de LabVIEW así como de la tarjeta de desarrollo SPARTAN 3E, para implementar un sistema de medición de distancia por medio de un sensor ultrasónico, modelo SRF04.

II. METODOLOGÍA DE DISEÑO EN LABVIEW PARA FPGA.

Creación de Proyectos

Para iniciar a programar, es necesario crear un nuevo proyecto. Una vez creado el proyecto, es imperante añadir el dispositivo o tarjeta FPGA a programar. Para nuestro caso, seleccionaremos la tarjeta Spartan 3E Starter Board, lo cual requiere que este instalado su controlador.

Al seleccionar el modelo de la tarjeta, esta se incluirá a nuestro proyecto y entonces podremos generar código utilizable directamente en la FPGA, teniendo así, acceso a sus puertos de E/S y demás recursos periféricos, inclusive poder crear un instrumento virtual, FPGA VI. Se deben añadir al proyecto los recursos de hardware que vayan a ser utilizados.

Por otra parte los recursos de hardware que no sean usados automáticamente pasan a un estado de alta impedancia (Z). Los pasos para crear un proyecto se enlistan a continuación:

- Ejecutar el programa LabVIEW 8.6 .
- En la pantalla de inicio de LabVIEW, hacer clic en “File” → “New Project”.
- En la ventana del explorador de proyecto, hacer clic secundario en “My Computer” → “New” → “Targets and Devices...” (Figura 1).
- Posteriormente seleccionaremos la opción “New target or device” → “ Spartan-3E Starter Board” → “OK”.
- Para poder crear código descargable directamente a la tarjeta, hacer clic secundario en “FPGA Target (Dev1, Spartan-3E Starter Board)” → “New” → “VI” (Figura 2).
- Para añadir recursos de hardware, hacer clic secundario en “FPGA Target (Dev1, Spartan-3E Starter Board)” → “New” → “FPGA I/O”
- Para seleccionar los recursos de hardware disponibles de nuestra tarjeta, hacemos clic en los elementos del menú “Available Resources” → “Add” → “OK” (Figura 3).

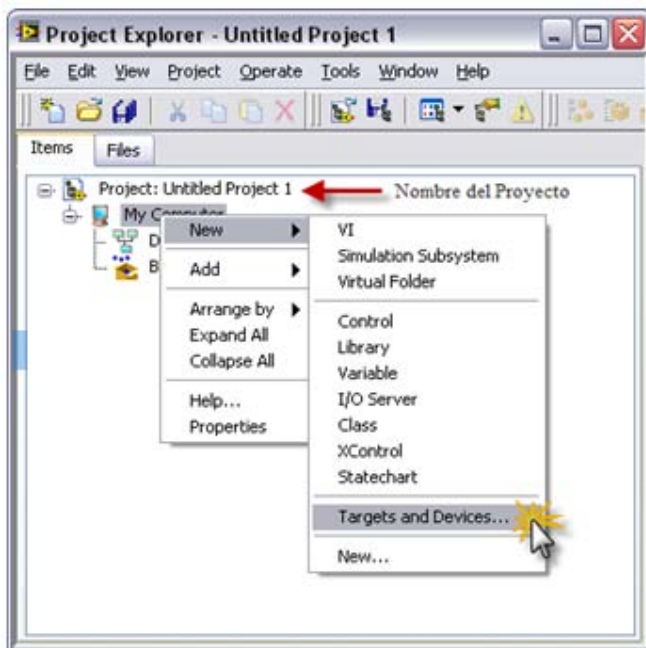


Fig.1 Selección de Tarjeta o Dispositivo.

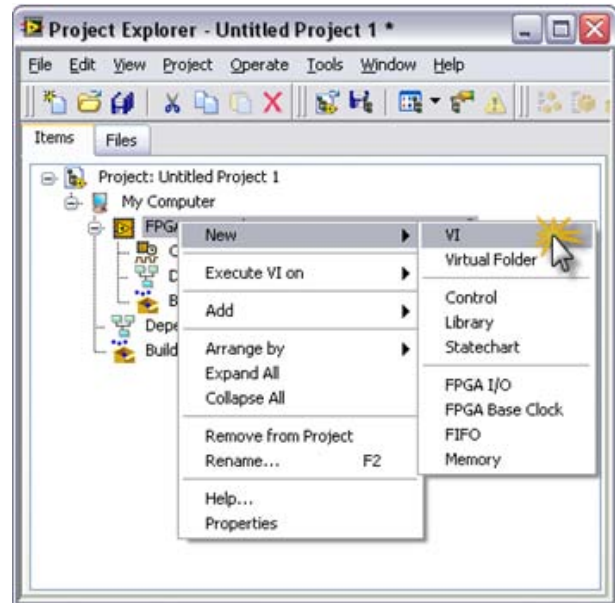


Fig.2 Creación de un FPGA VI.

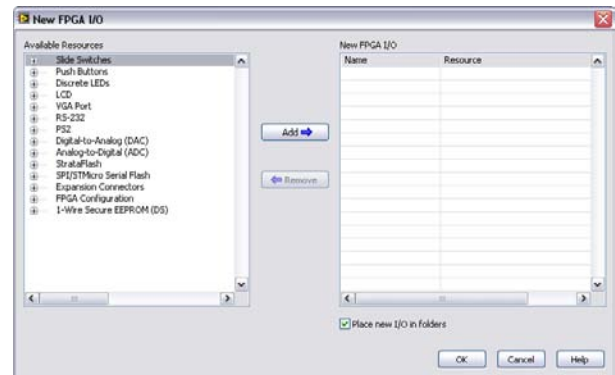


Fig.3 Ventana de Recursos de Hardware.

Una vez creado el proyecto y añadido el FPGA VI, la metodología de programación es la misma que en un proyecto típico de LabVIEW, es decir, tendremos un diagrama de bloques para elaborar código gráfico y un panel frontal en donde se creará la interfaz visual del programa.

Es importante mencionar que un FPGA VI no cuenta con todas las funciones de un VI típico de LabVIEW como puede ser el manejo de datos de punto flotante, controladores PID, manejo de puertos de la computadora, etc. Para solucionar este tipo de inconvenientes, es posible portar bloques E/S del FPGA a un VI genérico que pueda contar con las funciones requeridas por la aplicación y que el FPGA no pueda proporcionar.

III. DESARROLLO DEL SISTEMA DE MEDICIÓN.

En la figura 4 se muestra un diagrama a bloques de las conexiones existentes entre la computadora, la tarjeta de desarrollo y el transductor ultrasónico. En la computadora se encuentra corriendo un instrumento virtual, el cual se encarga del manejo de las señales tanto de visualización como de control. En la tarjeta se encuentra programado un sistema digital para la activación del sensor y la medición de la señal que regresa el transductor ultrasónico.



Fig.4 Diagrama general del sistema.

El transductor ultrasónico proporciona un tren de pulsos ultrasónicos, los cuales rebotan en los objetos sólidos, y mediante la diferencia de tiempos entre la señal enviada y la señal recibida aplicando un factor de conversión se puede obtener la distancia.

El instrumento de medición ultrasónico basado en FPGA tiene dos modos de operación:

- a) **Modo Remoto.** En este modo se requiere de la interacción de la computadora con la tarjeta.
- b) **Modo Autónomo.** En este modo no es necesario que la computadora este presente.

Sensor Ultrasónico SRF04.

Módulo formado por dos capsulas ultrasónicas de 40KHz de la firma DEVANTECH Ltd, el cual emplea un microcontrolador PIC12C508. En la figura 5 se muestra un juego de imágenes del módulo.

El módulo transmite un tren de pulsos de 8 ciclos a 40KHz, tal y como se muestra en la figura 6. En ese momento la señal de salida ECO pasa a "1". Cuando la cápsula receptora recibe la señal transmitida como consecuencia de haber rebotado en un objeto (eco), esta salida pasa de nuevo a "0". El usuario debe medir la duración del pulso de esta

señal, es decir, el tiempo en que la señal eco se mantiene en "1".

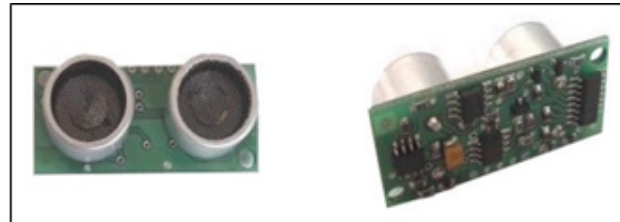


Fig. 5 Vista anterior y posterior del módulo SRF04.

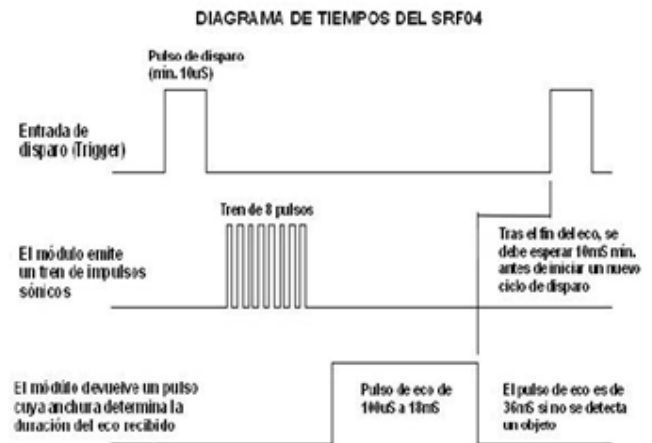


Fig.6 Diagrama de tiempos.

Con objeto de que el módulo se estabilice, se debe dejar un lapso de tiempo de 10mS mínimo entre el momento en que la señal de eco pasa a "0" y un nuevo pulso de disparo inicie el siguiente ciclo o medida. La duración del pulso eco de salida varía entre 100mS y 18mS, en función de la distancia entre las cápsulas del módulo y el objeto. Siendo que la velocidad del sonido en el aire (20°C) es de 343 m/S, se puede establecer que 1 cm es recorrido en 29.15 µS. Debido a que el pulso ultrasónico realiza un recorrido de ida y vuelta, el tiempo de rebote es de 58.30 µS por cada centímetro medido. Por especificación del sensor el rango mínimo que se puede medir es de 1.7 cm y el máximo de 310 cm.

El módulo emplea tan sólo 4 conexiones que se pueden realizar soldando directamente 4 cables o bien mediante un conector de 5 vías. En la figura 7 pueden observarse las conexiones físicas del módulo.



Fig. 7 Conexiones del SRF04.

Unidad de Control y Procesamiento Digital.

Este componente se encuentra implementado con los bloques digitales del FPGA de la tarjeta de desarrollo, su función es procesar la información enviada por el sensor y administrar el disparo al módulo SRF04 para el inicio de su operación. En la figura 8 se presenta un diagrama a bloques de esta unidad.

Dentro de la unidad de control digital se genera un pulso de disparo en alto durante $10\ \mu\text{S}$, activado al presionar un "Push Button" de la tarjeta Spartan 3E (terminal D5 del FPGA) o de la interfaz gráfica de LabVIEW [4]. Este pulso de $10\ \mu\text{S}$ es alimentado a la terminal de disparo del módulo SRF04. Los $10\ \mu\text{S}$ son generados por un contador de 0-500 el cual realiza un conteo cada ciclo de reloj. Al utilizar el reloj de 50 MHz, se obtienen ciclos temporizados de 20 nS y una vez transcurridos los 500 conteos, se habrán generado los $10\ \mu\text{S}$ de tiempo buscados.

La medición de la distancia es función del tiempo en alto de la señal de eco producida por el módulo SRF04, dicho módulo devuelve un pulso de eco con una anchura entre $100\ \mu\text{S}$ y $18\ \text{mS}$ que representa la distancia medida.

Partiendo de que nuestro factor de conversión es $58.3\ \mu\text{S}$ por cada centímetro recorrido, obtenemos su equivalente en milímetros, esto es, $5.83\ \mu\text{S}$ por cada milímetro recorrido, observe en la figura 8 el contador ubicado dentro del ciclo case en color rojo, este contador realizará conteos cada 20 nS, siempre que la terminal de eco del módulo SRF04 se encuentre en nivel alto, debido al uso del ciclo de temporizado con el reloj de 50 MHz de la tarjeta.

Debido a que el programa de LabVIEW para FPGA no permite realizar operaciones de punto flotante en nuestra tarjeta de desarrollo, se recurre a otra forma (un contador) para deducir la distancia sin tener que realizar la división de forma directa. Este contador tiene la tarea de estar habilitando la entrada al ciclo case en color azul cada $5.82\ \mu\text{S}$ ($291 \times 20\ \text{nS}$), véase figura 8, en donde básicamente encontramos contadores por décadas (0-9). Estos contadores irán registrando la representación milimétrica de la distancia medida en los indicadores n1, n2, n3 y n4.

Observe las constantes cableadas a la estructura de ciclo temporizado, las cuales conectan a su vez con los nodos de retroalimentación de los contadores, su tarea es inicializar los nodos de retroalimentación cada vez que se entra al ciclo de temporizado (no entre ejecuciones). La finalización del ciclo temporizado se obtiene cuando la señal de eco es de nivel bajo, esto debido a que esta terminal pasa por un inversor y posteriormente se conecta a la terminal de condición de ciclo, la cual detendrá la ejecución del Ciclo de Temporizado solo si se le conecta un '1' (True) [5].

Para poder tener acceso de lectura y escritura al LCD de la tarjeta Spartan 3E, es necesario deshabilitar la Intel StrataFlash debido a que las líneas de datos del LCD se encuentran compartidas con las líneas de datos de la Intel StrataFlash, esto es conseguido colocando un '1' lógico a la terminal de la Intel StrataFlash.

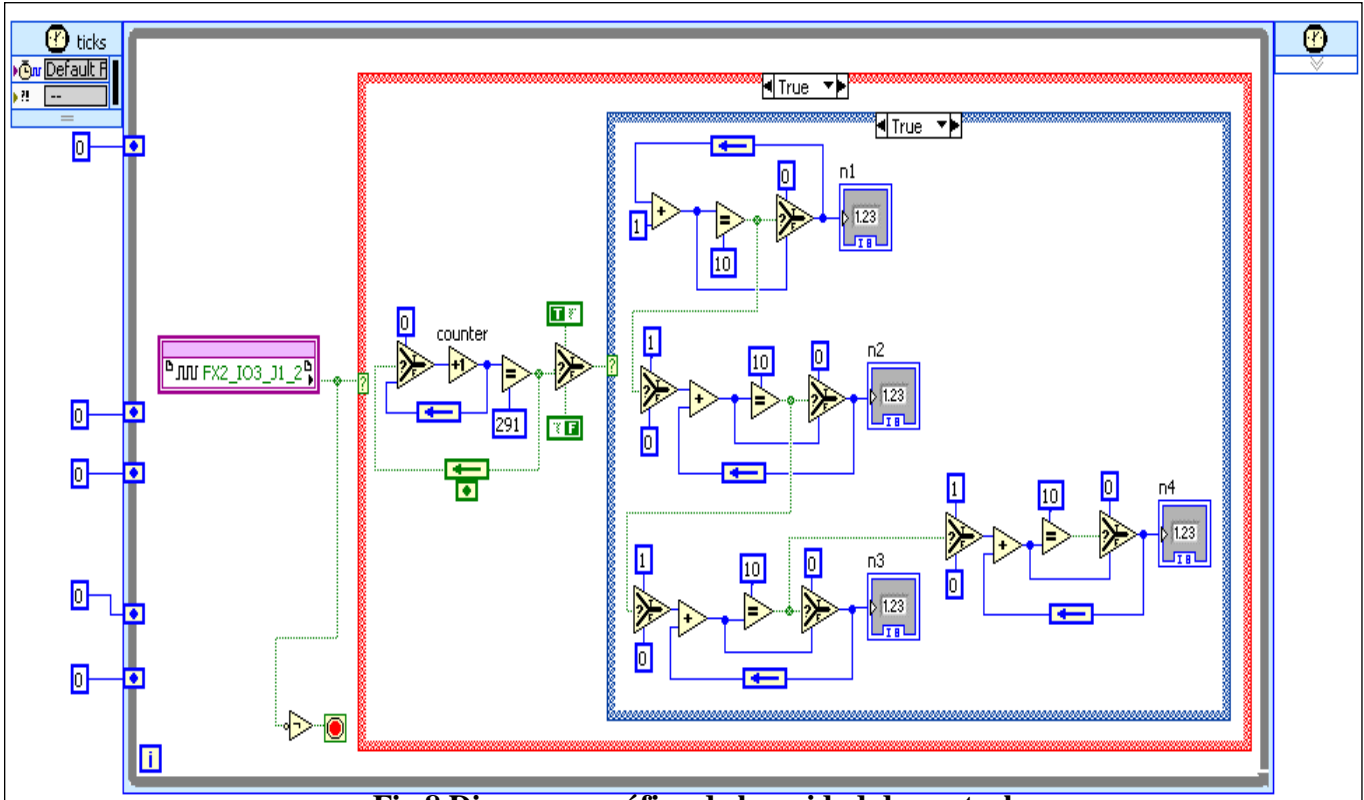


Fig.8 Diagrama gráfico de la unidad de control.

Una vez encendido el LCD, debe ser inicializado para establecer el protocolo de comunicación requerido. Para nuestra tarjeta, el protocolo que es utilizado se basa en la transferencia de 8 bits de datos por medio de una interfaz de 4 bits, en la cual se transfieren primero los 4 bits más significativos y después los 4 bits menos significativos. Realizada la secuencia de inicialización y configuración, como el que se realiza en un LCD común, se procede a enviar el mensaje [6]. En la figura 9 se muestra la escritura de la distancia procesada por la unidad de control, concentrada en un sub-instrumento virtual, al que se le nombró “Escribe Lectura del Sensor Ultrasónico”. Debido a que n4, n3, n2 y n1 únicamente entregan valores decimales entre 0-9, es necesario escalar estos valores a sus respectivos ASCII.

Interfaz Gráfica en LabVIEW.

La interfaz gráfica, o panel frontal es útil para visualizar en la pantalla de la computadora la medición realizada, también permite el control de forma remota del instrumento de medición, al

proveer al usuario de un botón para la puesta en marcha del mismo. Ver figura 10.

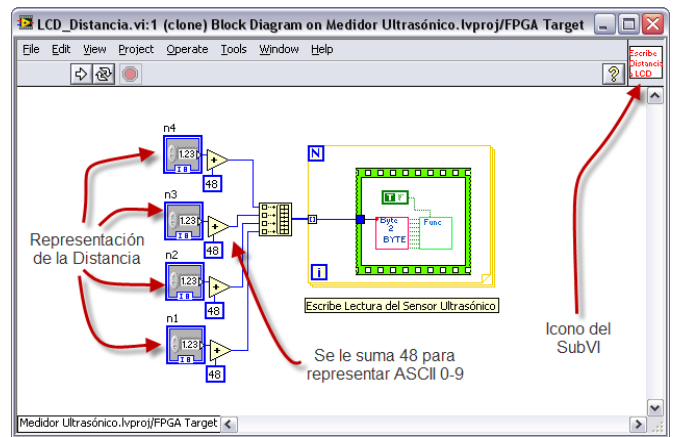


Fig. 9 Sub-instrumento virtual del LCD.

El sistema creado permite la comunicación del instrumento de medición con la computadora, para el control de operación remota del mismo y el despliegue a pantalla de la distancia medida.



Fig. 10 Interfaz gráfica del sistema de medición.

Los elementos requeridos para su funcionamiento son:

- **Interfaz Gráfica.** Construida en LabVIEW. Permite visualizar en la computadora, la información recolectada por el sensor, procesada y transmitida por el FPGA. Así mismo, permite controlar el módulo SRF04 para su operación remota.
- **Conexión USB.** Conexión realizada entre la tarjeta Spartan 3E y en la computadora. No requiere de programación extra.

IV. SISTEMA DE MEDICIÓN IMPLEMENTADO EN VHDL.

El sistema para medición de distancia fue también implementado en VHDL, en la figura 11 se muestra un diagrama a bloques del sistema digital construido. Lo anterior, con la finalidad de tener algún cierto punto de comparación entre los diseños puramente VHDL y diseños LabVIEW.

El código VHDL genera el pulso de activación, adquiere el pulso proporcional a la distancia, lo convierte a distancia con un factor de conversión y mediante comunicación serial transmite el equivalente a la distancia medida a la hyperterminal de la computadora en formato hexadecimal para la visualización de la medición.

Es necesario aclarar que el diseño VHDL tiene menos elementos que el diseño en LabVIEW, aunque los dos realizan la medición de distancia de la misma forma.

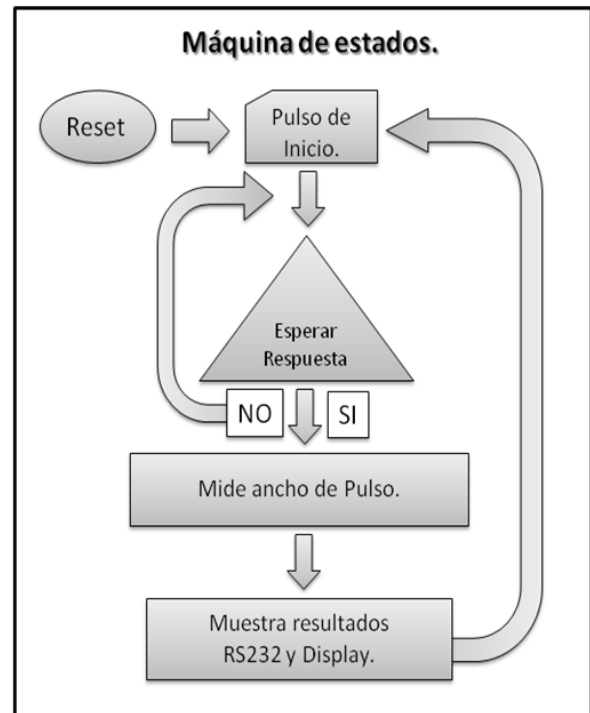


Fig. 11 Diagrama a bloques del sistema de medición en VHDL.

V. RESULTADOS.

En la tabla 1, se muestran los resultados de consumo de recursos del sistema de medición de distancia, implementado en código VHDL en una tarjeta Spartan 3S200. Los recursos utilizados por el código elaborado en LabVIEW, se muestran en la tabla 2. Por último, en la figura 12, se observa la medición de distancia desplegada en el LCD de la tarjeta de desarrollo.

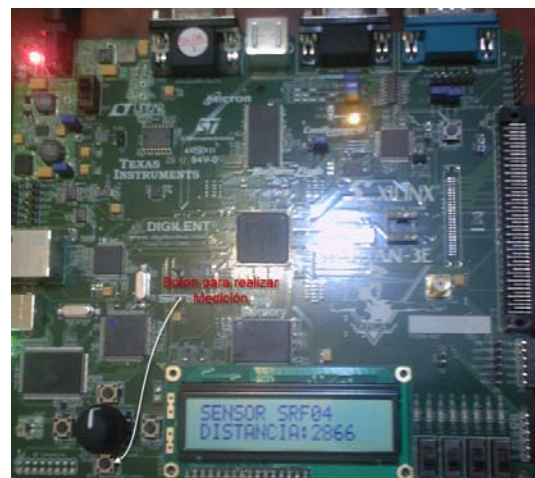


Fig. 12 Tarjeta Spartan 3E mostrando el valor de distancia.

Tabla 1 Consumo de recursos programando en VHDL.

Device Utilization Summary			
Logic Utilization	Used	Available	Utilization
Number of Slice Flip Flops	149	3,840	3%
Number of 4 input LUTs	257	3,840	6%
Logic Distribution			
Number of occupied Slices	159	1,920	8%
Number of Slices containing only related logic	159	159	100%
Number of Slices containing unrelated logic	0	159	0%
Total Number of 4 input LUTs	292	3,840	7%
Number used as logic	257		
Number used as a route-thru	35		
Number of bonded IOBs	7	173	4%
Number of BUFGMUXs	2	8	25%

Tabla 2 Consumo de recursos programando en LabVIEW.

Device Utilization Summary			
Logic Utilization	Used	Available	Utilization
Number of Slice Flip Flops	1,860	9,312	19%
Number of 4 input LUTs	2,291	9,312	24%
Logic Distribution			
Number of external IOBs	90	232	38%
Number of LOCed External Input IBUFs	28	28	100%
Number of external Output IOBs	62	62	100%
Number of BSCANs	1	1	100%
Number of BUFGMUXs	2	24	8%
Number of Slices	1,770	4,656	38%
Number of SLICEMs	18	2,328	1%

VI. CONCLUSIONES.

El módulo LabVIEW para FPGA permite ampliar el campo de desarrollo gráfico de LabVIEW, para crear hardware personalizado sobre el FPGA, en nuestro caso es utilizado para la medición de distancia.

LabVIEW es claramente adecuado para programación de FPGA ya que representa paralelismos y flujo de datos únicos.

El trabajo realizado, muestra la facilidad de integración de LabVIEW al mundo de los FPGA, es importante no perder de vista que habrá casos en los cuales es muy útil su uso, pero en otros será necesario tener que recurrir a las técnicas de diseño tradicionales de bajo nivel.

Por lo tanto se propone la integración de estas dos técnicas de programación para el diseño de sistemas de hardware, dejando abierta la posibilidad de elegir la opción más viable de diseño o inclusive una mezcla de ambas.

VII. REFERENCIAS

- [1] Liang Dong and Massood Z. Atashbar, "An FPGA Experience in ASIC Design", Proceedings of the 2005 ASEE North Central Conference, American Society for Engineering Education, 2005.
- [2] Wakerly John F. *Digital Design*, Prentice Hall, 2000.
- [3] "Tecnología FPGA". Disponible: http://www.ni.com/fpga_technology/esa/
- [4] *Spartan-3E Starter Kit Board User Guide*, UG230 (v1.0) Mar 9, 2006.
- [5] Travis, Jeffrey y Kring, Jim, *LabVIEW for Everyone: Graphical Programming Made Easy and Fun*, Prentice Hall. 2006.
- [6] Garza Garza Juan Ángel, *Sistemas Digitales y Electrónica Digital*, Prentice Hall, 2006.

INTERFASE PARA LA ADQUISICIÓN DE SEÑALES PURAS DEL EQUIPO DE ULTRASONIDO HITACHI EUB-25

Avilés Cruz Carlos, Ferreyra Ramírez Andrés, Ocampo Hidalgo Juan, Vázquez Álvarez Iván, Zúñiga López Arturo y Alducin Castillo Javier.

Departamento de Electrónica. Universidad Autónoma Metropolitana – Azcapotzalco.
Av. San Pablo No. 180 Col. Reynosa. C. P. 02200, México D. F.
caviles@correo.azc.uam.mx

RESUMEN

En el presente trabajo se desarrolló una interfase con el objetivo de obtener señales provenientes de los sensores del equipo de ultrasonido Hitachi EUB-25. Las señales que se obtuvieron están libres de filtrado y no contienen procesamiento alguno realizado por el equipo comercial. Se tuvo la necesidad de revisar los manuales del fabricante y de realizar pruebas en directo sobre el equipo. Los retos principales fueron identificar y caracterizar: los voltajes manejados, niveles de corrientes, multiplexaje y sobre todo la sincronización de las señales, ya que por el mismo transductor se envía y se recibe la señal. La interfase desarrollada por medio de circuitos eléctricos nos permite obtener señales tal cual las entregan los transductores.

Palabras clave: Ultrasonido, HC11, Osciloscopio.

I. INTRODUCCIÓN

Se cuenta con un equipo de ultrasonido HITACHI EUB-25 en uso y buenas condiciones del cual se requiere obtener señales directamente obtenidas de los transductores. Las señales obtenidas sirven para la generación de imágenes ecográficas. Las señales se ubican en Radio-Frecuencia. Primeramente se debe de analizar y entender perfectamente las señales que maneja el equipo. Una vez conocido el funcionamiento del equipo, recuperar por medios electrónicos las distintas señales que participan en la formación de una línea de RF pero antes de ser procesada por el equipo de ultrasonido (señales brutas).

Se analizó, diseñó e implementó circuitería con el fin de obtener las señales brutas del equipo. Una vez obtenidas las señales, estas se desplegaron en un osciloscopio y posteriormente se enviaron multiplexadas a una Computadora Personal (PC). El porqué obtener las señales brutas, se justifica por el hecho de poder modificar los retardos de las señales y poder modular las señales brutas. Tanto los retardos como la modulación servirán para mejorar las representaciones elastográficas de la señal.

El equipo de ultrasonido entrega la información de los sensores, es decir 15 señales de recepción y una de control, en total 16 señales multiplexadas, dichas señales serán digitalizadas para poder capturarlas en un PC, así mismo poder desplegarlas en un osciloscopio de 4 canales, cuatro a la vez.

Para el desarrollo del presente trabajo se utilizó como bibliografía, principalmente el manual de operación y de mantenimiento del equipo HITACHI EUB-25, al igual de los manuales de circuitos eléctricos y electrónicos que sirvieron para el desarrollo de la interfase.

II. GENERACIÓN DE IMÁGENES MÉDICAS POR ULTRASONIDO

Las imágenes médicas de ultrasonido permiten visualizar cortes de un órgano o de alguna parte del cuerpo. Se explota la velocidad de propagación y la cantidad de señal ultrasónica reflejada. La cantidad de señal reflejada es entonces codificada

en niveles de grises permitiendo reconstruir imágenes. En la figura 1 se muestra el equipo

HITACHI EUB-25 usado en el presente proyecto.

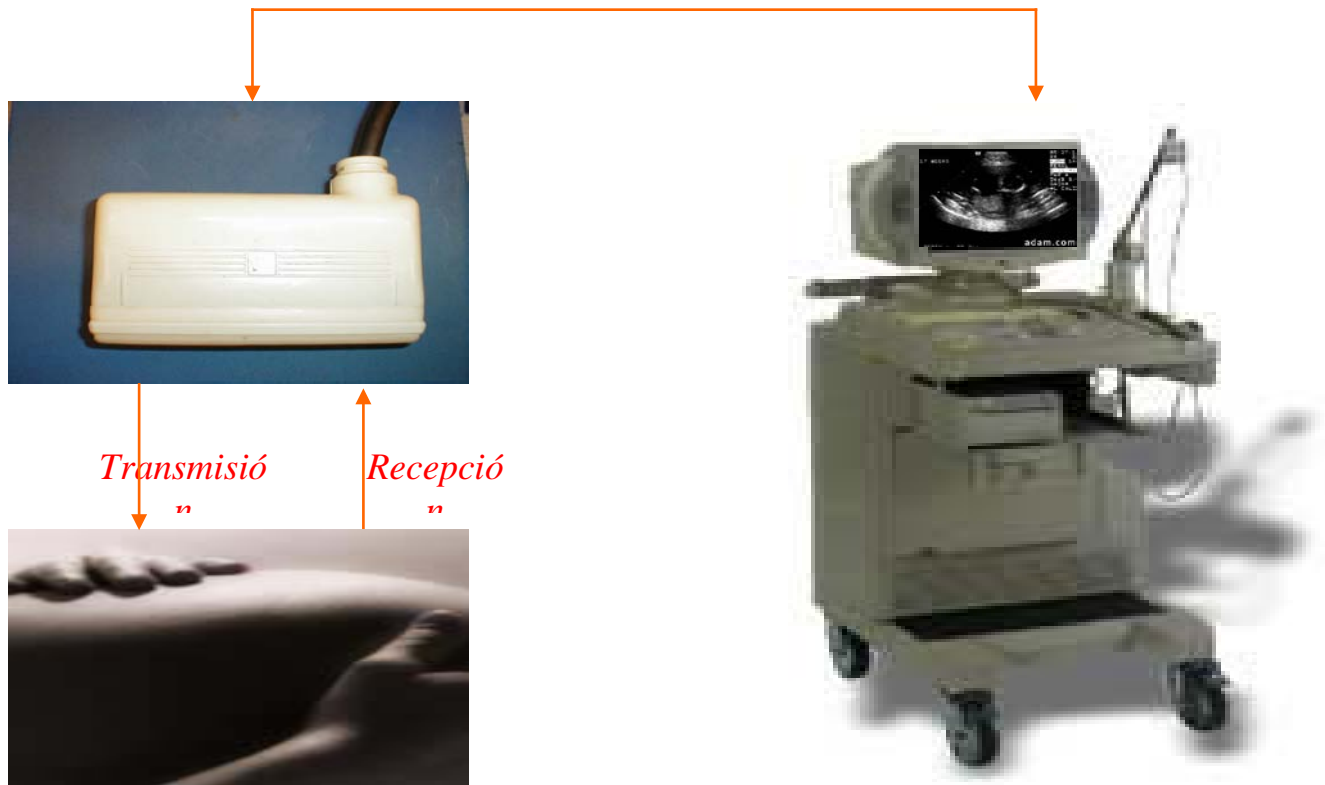


FIGURA 1: Equipo comercial HITACHI EUB-25

Características del sensor:

- Frecuencia de Resonancia: 3.5 MHz.
- Número de elementos: 400 (80 grupos de 5 cristales)
- Longitud: 120 mm.
- Longitud útil: 105mm.
- Peso: 700 grs.
- Dimensiones: 150X34X96 mm.

Un señal acústica es generada a nivel del sensor por medio de cristales piezo eléctricos que vibran de acuerdo a una excitación eléctrica. La señal acústica viaja por un medio que la puede reflejar o difundir. El sensor se pasa ahora a la fase de “escucha” registrando los ecos que van a ser traducidos en señales eléctricas por los mismos

cristales. El ensamble de cristales (el arreglo) va a permitir reconstruir una imagen en función directa de la cantidad de señal recibida. La imagen entonces es desplegada ya sea preprocesada o directamente. Para fines de mejor resolución, las imágenes pueden ser mejoradas vía funciones logarítmicas.

III.1 SEÑALES DE SINCRONIZACIÓN

Una ecografía necesita varias señales de sincronización para poder funcionar correctamente. Dentro de las múltiples señales, dos de ellas son las de interés para nosotros; por un lado la sincronización de inicio de línea y por otro lado, la

sincronización de inicio de imagen. La sincronización de línea permite ubicar el inicio de cada emisión ultrasonora lineal, lo que nos permitirá saber el momento en cual podremos recuperar la señal reflejada por un lado y también llevar a cabo la reconstrucción de la imagen en su totalidad. La sincronización de imagen indica el inicio y el final de cada imagen ecografía.

III.2 LOCALIZACIÓN-IDENTIFICACIÓN DE SEÑALES

Como no se contaba con los planos ni diagramas del equipo, lo primero que se hizo fue identificar los cables del sensor, es decir la relación de cada cable con los cristales que lo componen. Desconectado el equipo se procedió a identificar cada cable con su correspondiente cristal. Dicho trabajo se realizó con la ayuda de un osciloscopio. El experimento anterior nos permitió concluir que de las terminales de 15 alambres corresponde los cristales HP-1 a HP-15. Los otros 16 cables corresponden a señales de control (de LA-8 a LA-23)

El análisis de cables y cristales del sensor, nos permitieron deducir que se pueden utilizar en los dos sentido, tanto para la emisión como para la

recepción. Evidentemente con sus respectivas señales de control.

La emisión en el ecógrafo Hitachi se lleva a cabo de una manera focalizada, hay tres partes principales a nivel de circuitos, los cuales son:

- a) Circuitos de líneas de retardo
- b) Circuito multiplexor y
- c) Circuito generador de líneas de RF.

La figura 2 muestra dichos bloques. Primeramente, un pulso de señal pasa al bloque uno (circuito de líneas de retardo) para obtener 5 señales (de f_0 a f_4) de las cuales una es sin retardo y las otras cuatro con diferentes retardos. A continuación cada señal se duplica en el circuito multiplexor para obtener diez señales de comando. Las señales son multiplexadas de manera simétrica.

Las 10 señales de salida son señales de comando y son tratadas por pares (las dos señales de los bordes no tienen retardo, mientras que las dos señales de en medio tienen el retardo mas grande). Las diez señales comandan los circuitos generadores de RF en emisión. Diez señales de RF de alta tensión de un total de quince. Finalmente, las diez señales de emisión alimentan a diez grupos de elementos diferentes del transductor, lo cual corresponde a un tiro de emisión focalizado.

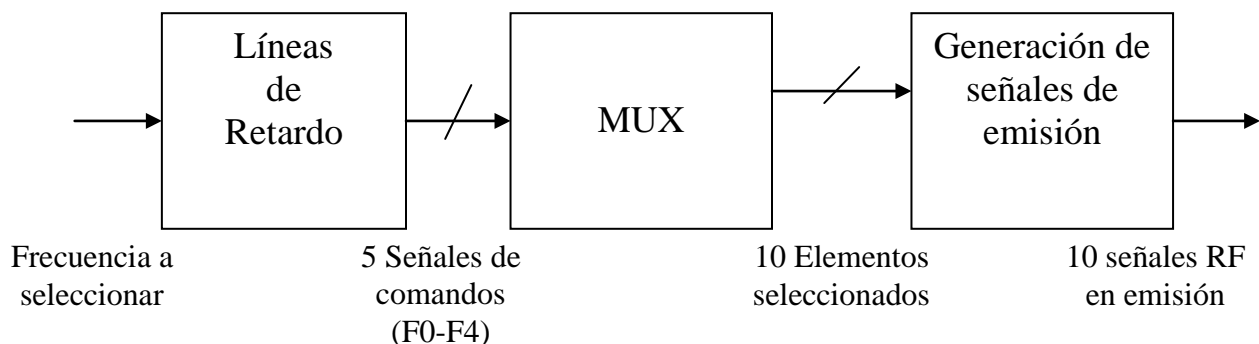


Figura 2: Partes más importantes en el bloque de emisión.

Es importante recordar que los sensores del equipo Hitachi son 400 agrupados en 80 grupos de cinco cristales cada grupo. De ahora en adelante, solo hablaremos de grupos de elementos. Los retardos en emisión son aplicados de manera focalizada hacia un tiro en particular. Los elementos del centro, contienen el retado más grande. La selección de los cristales se hace en dos tiempos, primeramente las señales comando, nos permiten seleccionar 15 grupos dentro de los 80 posibles, posteriormente, 10 elementos de los 15 son excitados por medio del circuito multiplexor.

III.3 SEÑALES DE CONTROL DE EMISIÓN

Las señales de control (CONT1 – CONT16) permiten seleccionar 15 grupos de elementos dentro de los 80. Par llevarlo a cabo, los elementos son aun reagrupados en super grupos de 25 (5 grupos de 5 elementos). Las señales de comandos activan a 75 elementos a la misma vez (5 grupos de

15 elementos). Al término de 10 tiros, se desplazan 25 elementos y son los siguiente 75 que se activan ahora. Las señales que permiten dicha selección son de LA1 a LA16.

III.5 CIRCUITO DE LÍNEA DE RETARDO EN EMISIÓN

Existe una sola señal que llega al circuito de línea de retardo, dicha señal es retardada de cinco formas diferentes para formar 5 señales de salida de esta parte del circuito (f0-f4). Los retardo son arreglados con el objetivo de centrar la energía en una región particular del espacio (lo que permite la focalización de la emisión). Las 5 señales forman la entrada del circuito multiplexor (una señal sin retardo y cuatro con retardo). Los retardos son escogidos en función de la distancia focal a partir de un conjunto de valores predefinidos. Los valores de retardo se pueden apreciar en la tabla 1.

Multiplexor	Retardo FI	Retardo F2	Retardo F3	Retardo F4
0	120	120	180	180
1	130	140	130	140
2	60	120	140	140
3	180	80	100	80
4	120	70	120	70
5	60	100	100	150
6	30	60	80	90
7	15	30	35	40

Tabla 1: Valores de retardo predefinidos (los retardos son en nano segundos)

III.4 CIRCUITO MULTIPLEXOR EN EMISIÓN

Los 15 elementos activos no son emitidos al mismo tiempo se, se tienen que seleccionar 10 de los 15, por medio de un multiplexor en emisión. El control de comando del circuito multiplexor guarda el mismo estado durante dos inicios de línea, lo que significa que hay dos disparos en el mismo lugar. El comando siguiente se desplaza una unidad sobre todas las salidas, lo que corresponde en la realidad a un desplazamiento de 5 cristales. Suponiendo que uno emite sobre los grupos enumerados del 1 al 10 (la señales por pares son

iguales 1-0, 2-9,3-8,4-7 y 5-6) con respecto al retardo. Así hay 5 señales que entran al multiplexor y salen 10. Las señales de comando del multiplexor son de TQA a TAD. Las 10 señales seleccionadas se encuentran a la entrada del circuito generador de señales de emisión.

III.6 ANÁLISIS DE LA RECEPCIÓN

Las señales en recepción son focalizadas de la misma manera que en la emisión. Existen cuatros parte a nivel de circuitos: 1) Circuitos amplificador

de entrada, 2) Circuito multiplexor en recepción, 3) Circuito de líneas de retardo y 4) Circuito amplificador logarítmico; dichos circuitos son descritos a continuación y esquemáticamente se presentan en la figura 3. Hay 15 señales que vienen directamente del transductor, las cuales son amplificadas por el circuito amplificador de entrada UA6-UA20. Las 15 señales son multiplexadas para seleccionar la línea que corresponda.

La salida del circuito multiplexor esta compuesta de 5 señales retardadas por el circuito de retardo de y sumadas a la salida; después de tener las 5 señales en fase, las 5 señales son sumadas para tener una sola y amplificarla logarítmicamente. La señal amplificada será la que se despliegue en una pantalla o monitor.

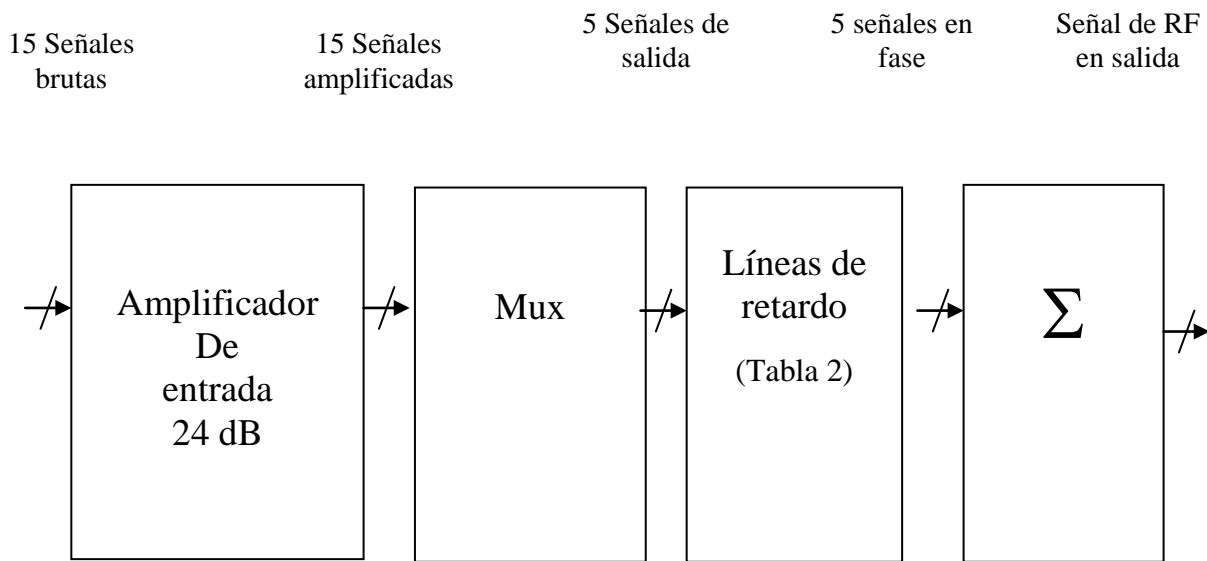


Figura 3: Esquema de recepción del equipo HITACHI

Multiplexor	Retardo 1	Retardo 2	Retardo 3	Retardo 4
0	140	240	300	340
1	90	160	200	220
2	60	100	120	120
3	25	80	180	60
4	80	140	120	200
5	50	90	120	120
6	30	60	70	80
7	15	30	40	40

Tabla 2: Tiempo de retardo en recepción (en nano segundos)

III.7 SEÑALES DE RECEPCIÓN AMPLIFICADAS

Los alambres que tiene las marcas LA1-LA15 son los 15 cables que contienen las señales de recepción brutas, cada alambre lleva la información tanto para la emisión como para la

recepción. Para evitar que las señales de emisión de alta tensión pasen por el circuito amplificador de entrada, se cuenta con una sistema de filtraje de

protección el cual esta ubicado a la entrada de cada circuito amplificador. Los circuitos amplificadores de emisión se encuentran sobre el impreso de “RF AMP Circuit”, la amplificación es de 24 dB. Las señales amplificadas son conectadas a la entrada del circuito multiplexor en recepción,

III.8 CIRCUITO MULTIPLEXOR EN RECEPCIÓN

Las señales de recepción que salen del circuito amplificador de entrada son enumeradas UA6-UA20. Las 15 señales son agrupadas a la entrada de 18 circuitos integrados multiplexores. La selección de los circuitos integrados son comandadas por las señales RQA, RQB, RQC y RQD.

Por cada línea de RF hay 9 grupos de elementos en recepción, los 9 elementos son procesados por pares. El grupo de elementos central tiene el retardo más grande posible, después de los elementos centrales, las señales de recepción, toca a los elementos que están al lado. Los dos grupos de elementos tienen la misma información ya que las señales de recepción llegan al mismo tiempo. Las señales en recepción son procesadas por pares y de forma simétrica.

La figura 4 muestra la técnica del circuito multiplexor en recepción del ecógrafo Hitachi.

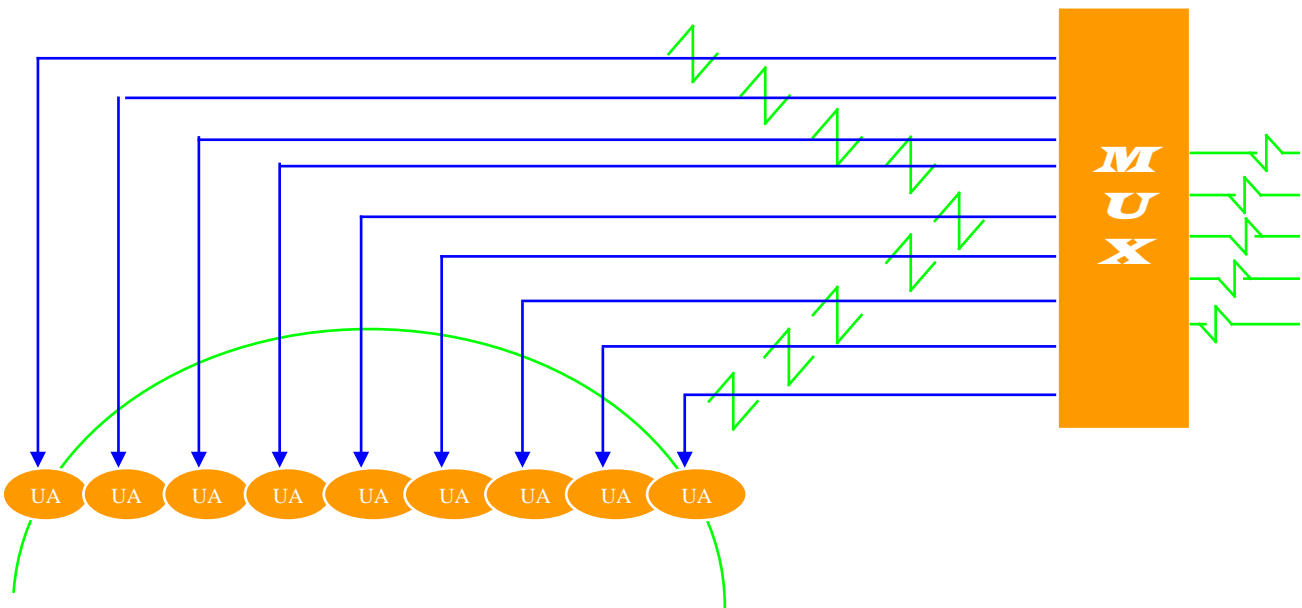


Figura 4: Técnica de recepción del equipo HITACHI

III.9 SOLUCIÓN PARA LA ADQUISICIÓN DE SEÑALES BRUTAS

El objetivo que se persigue en el presente proyecto, es el capturar las señales brutas que provienen del sensor, es decir, captar las señales sin procesamiento alguno. Las señales podrán ser adquiridas por una computadora o bien desplegarlas en un osciloscopio. El problema que

nos enfrentamos fue al manejo de las 15 señales simultáneamente.

Solución: Se propuso utilizar un osciloscopio de 4 canales (disponible en todo laboratorio) para poder desplegarlas.

Ventajas: Utilizar un osciloscopio del laboratorio y una tarjeta simple de multiplexaje diseñada por nosotros mismos.

Inconvenientes: La adquisición de las 15 señales se hace sobre 4 tiempos de imagen.

El osciloscopio utilizado fue HP 9375 bajo las siguientes configuraciones: Sincronización externa hecha por el circuito multiplexor.

Adquisición efectiva:

En nuestro caso, de debe adquirir 5*140 segmentos sean 700 segmentos o 4*140=560 segmentos para el multiplexaje de 4 líneas. Se puede adquirir a 200µS por línea muestreado a 10MHz (2000 puntos) (profundidad de exploración 30.8 cm.), 100µS por línea con muestreo de 25MHz (2500

puntos) (15.4 cm. de profundidad de exploración) y a 50µS por línea a muestreo de 50MHz (2500 puntos) (7.7 cm. de profundidad de exploración).

Tarjeta de multiplexaje

Las 15 señales originarias de los sensores son multiplexadas en los cuatro canales del osciloscopio. Para la sincronización, se tendrá una señal adicional. La tarjeta diseñada maneja todas las señales de control sincronizando sobre el inicio de la línea y el inicio de la imagen, como se puede apreciar en la figura 5, donde se puede ver el cronograma de tiempos.

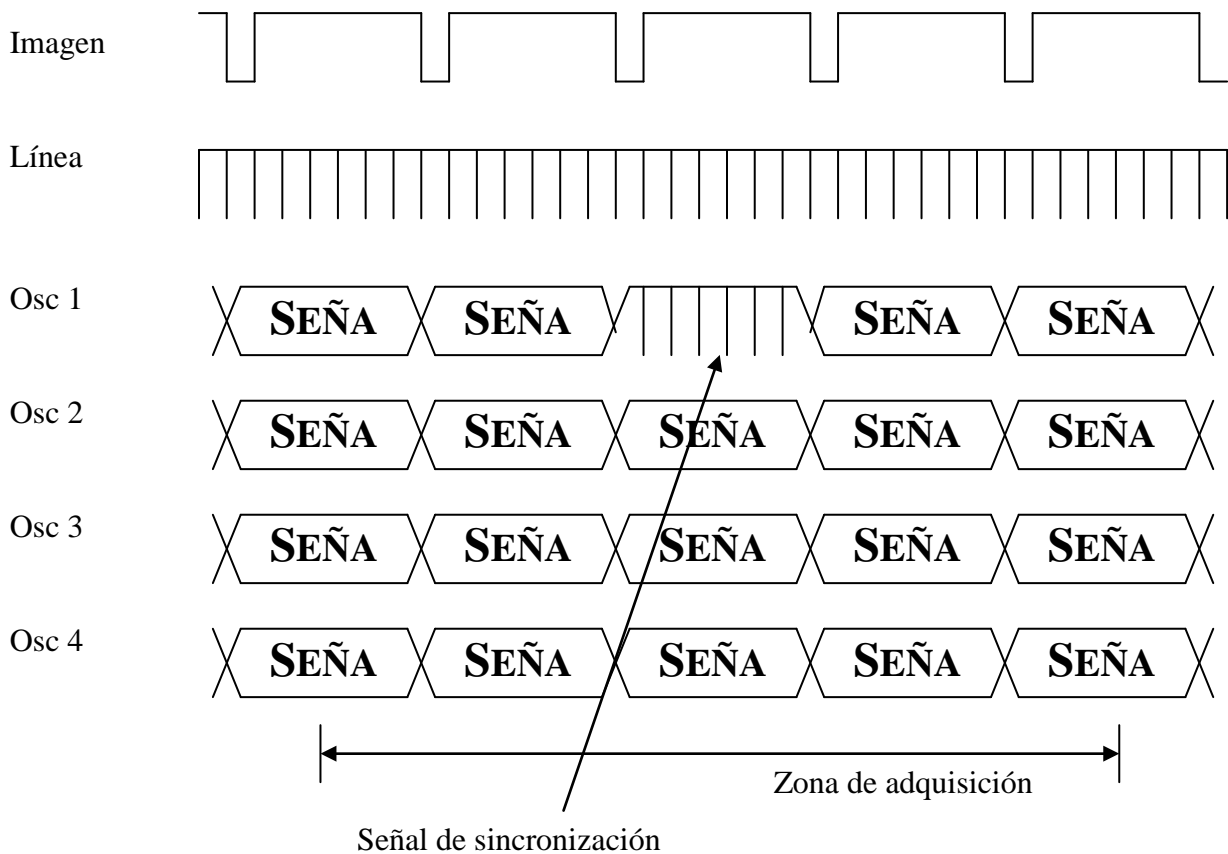


Figura 5: Cronograma de tiempos, así como la sincronización

El osciloscopio adquiere a 4*560 segmentos continuos. Dentro de las señales adquiridas, existe una señal de sincronización la cual permite de encontrar y demultiplexar las 15 señales. El demultiplexaje se hace en base a la sincronización,

encontrando también el orden de la señal. La señal de sincronización sobre la vía uno y la señal sobre la vía dos, tres y cuatro respectivamente. Después de 140 impulsos de sincronización vendrán las señales 5, 6, 7, etc. Utilizando este tipo de

multiplexaje, permite reconstruir una imagen en cuatro tiempos

La tarjeta de multiplexaje recibe en las 15 señales analógicas de RF provenientes de los captosres. El inicio de imagen y el inicio de línea generados por el ecógrafo. A la salida de la tarjeta, se encuentran 4 salidas que se conectan al osciloscopio (también una señal de sincronía).

La tarea de multiplexaje se llevo a cabo con un microcontrolador HC11 pudiendo manejar 14 señales sobre cuatro canales sin ningún problema. El siguiente cronograma muestra la ubicación tanto del inicio de línea como el inicio de imagen, ambas con una señal de sincronización

III. CONCLUSIONES

En el presente trabajo se pudo implementar una interfase con el equipo de ultrasonido HITACHI EUB-25, se respetó la parte de amplificación de las señales del propio aparato. Se logró exitosamente tomar las señales brutas, para que posteriormente éstas sean desplegadas en un osciloscopio y además, pasarlas a un computadora de uso genérico vía una tarjeta de adquisición de señales comercial. Se pudo manipular exitosamente las 15 señales del sensor, así como la señal de sincronización.

Las señales brutas serán utilizadas, por un lado, para trabajar sobre los retardos de la señal, y por otro, sobre las diferentes formas de modulación de ellas. Las representaciones de la señal en forma bidimensional, permiten generar las imágenes. Con el manejo de los retardos se podrá obtener representaciones de las imágenes más nítidas y modularlas para representar otro tipo de información bi-dimensional.

Como perspectivas de éste trabajo, se probaran las señales aplicando algoritmos diferentes a los incluidos en el equipo. Se cuenta con la salida de 15 señales multiplexadas, las cuales se pueden utilizar para reconstruir la imagen tal como lo haría el equipo o bien para hacerles algún preprocesamiento. Como la interfase se llevo a

cabo con componentes analógicos, el sistema puede trabajar en tiempo real controlado desde una computadora personal.

IV. REFERENCIAS

- [1] *Manual de Usuario del microcontrolador HC11*, Motorola, USA, 2001.
- [2] Morris Mano, *Circuitos Lógicos*, Mc Graw Hill, Espeña 2000.
- [3] *Manual del Osciloscopio HP*, Hewlet Pacard, 2000.
- [4] John Pears, *Ultrason Principles*, Mc Graw Hill, 2002.

V. AUTORES

El Dr. Carlos Avilés Cruz, nació el 13 de marzo de 1966 en la Ciudad de México, D.F. Obtuvo el título de Ingeniero en Electrónica especialidad en Sistemas Digitales en la Universidad Autónoma Metropolitana en 1991, el grado Maestro con especialidad en Procesamiento Digital de Señales, Imágenes y Voz en el Instituto Politécnico Nacional de Grenoble en 1993 y el grado de Doctor del Instituto Politécnico Nacional de Grenoble en 1997. El Dr. Avilés tiene el nombramiento de profesor-investigador Titular "C" en la U.A.M. de Azcapotzalco. El Dr. Avilés es autor y coautor de más de 30 artículos en revistas y congresos nacionales e internacionales, además de coautoría de dos libros. Sus intereses son la visión por computadora, el procesamiento digital de imágenes, el procesamiento digital de señales y la estadística de orden superior. El Dr. Avilés es miembro del Sistema Nacional de Investigadores nivel I.

DESARROLLO DE UN SELECTOR DE DERIVACIONES ELECTROCARDIOGRÁFICAS BASADO EN UN ARREGLO DE CMOS COMERCIALES

Cano ME, Mena E, Tavares M, Hernández IG, García J. A, Avila J.

Departamento de Ciencias Básicas, Centro Universitario de la Ciénega Universidad de Guadalajara
 Av. Universidad No. 1115, Col. Linda Vista, Ocotlán, Jalisco, México. CP 47820
 Tel. 3929259400, correo electrónico: eduardo.cano@cuci.udg.mx

RESUMEN

En este artículo, se propone una metodología eficiente para implementar un arreglo de 3 CMOS, los cuales son circuitos integrados multiplexores analógicos de 16/1, que pueden ser utilizados una etapa antes de la etapa de amplificación diferencial del circuito que compone un electrocardiógrafo, y que además son muy comerciales y de bajo costo. Con este arreglo es posible seleccionar cualquier combinación de los 10 electrodos que se encuentran colocados en el cuerpo y que conformarán una derivación de un electrocardiograma, ya sea bipolar, unipolar aumentada o precordial.

Palabras Clave: CMOS, Derivación, Electrocardiógrafo.

I. INTRODUCCIÓN

En el cuerpo humano se generan una amplia variedad de señales eléctricas, provocadas por la actividad química que tiene lugar en los nervios y músculos que lo conforman. El corazón por ejemplo conduce a un patrón característico de variaciones de voltaje. Tiene un sistema de conducción compuesto por fibras de músculo cardíaco especializadas en la transmisión de impulsos eléctricos. Como se muestra en la Fig. 1 [1], el corazón se compone de tres tipos de músculo cardíaco principalmente:

- Músculo auricular
- Músculo ventricular

- Fibras musculares excitadoras y conductoras especializadas.

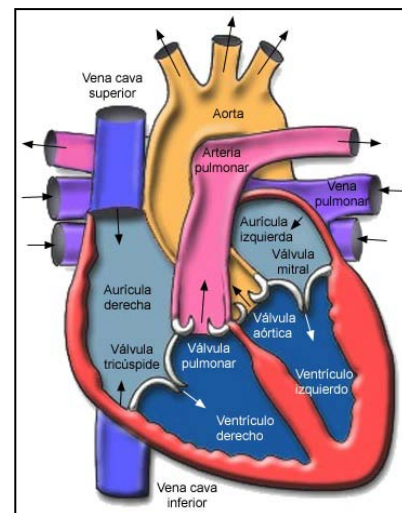


Figura 1: Esquema de las partes del corazón

La contracción comienza en el ápice del corazón, progresando hacia arriba para expulsar la sangre a las grandes arterias. Este sistema de conducción eléctrico explica la regularidad del ritmo cardíaco y asegura la coordinación de las contracciones aurículo-ventriculares.

Esta actividad eléctrica puede ser analizada con electrodos situados en la superficie de la piel, llamándose a esta prueba electrocardiograma. Cuando el impulso cardíaco atraviesa el corazón, la corriente eléctrica también se propaga desde el corazón hacia los tejidos adyacentes que lo rodean. Una pequeña parte de la corriente se propaga hacia la superficie corporal. En la Fig. 2, se muestra un

electrocardiograma normal de dos latidos el corazón.



Figura 2. Gráfica de un electrocardiograma de dos pulsos cardiacos

La Fig. 3, muestra las conexiones eléctricas entre las extremidades del sujeto y el electrocardiógrafo para registrar electrocardiogramas de las denominadas derivaciones bipolares estándar de las extremidades. El término “Bipolar” significa que el electrocardiograma se registra a partir de dos electrodos que están localizados en lados diferentes del corazón, en este caso en las extremidades. Así, una “derivación” no es un único cable que procede del cuerpo sino una combinación de dos cables y sus electrodos para formar un circuito completo entre el cuerpo y el electrocardiógrafo.

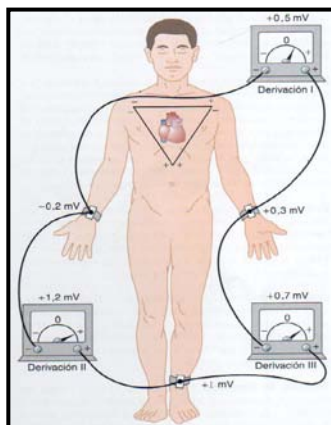


Figura 3: Disposición convencional de los electrodos en un ECG y el triángulo de Einthoven en el tórax.

DI: Cuando se registra la derivación I, la terminal negativa del electrocardiógrafo esta conectado al brazo derecho y la terminal positiva al brazo izquierdo.

DII: Para registrar la derivación II de las extremidades, la terminal negativa del electrocardiógrafo se conecta al brazo derecho y la terminal positiva a la pierna izquierda.

DIII: Para registrar la derivación III de las extremidades, la terminal negativa del electrocardiógrafo se conecta al brazo izquierdo y la terminal positiva a la pierna izquierda.

Así mismo, en la Fig. 3, se dibuja un triángulo, denominado triángulo Einthoven, alrededor de la zona del corazón. Este diagrama ilustra que los dos brazos y la pierna izquierda forman vértices de un triángulo que rodea el corazón. Los dos vértices de las partes superiores del triángulo representan los puntos en los que se conectan eléctricamente a los líquidos que rodean el corazón y el vértice izquierda es el punto en el que la pierna izquierda se conecta a los líquidos. La ley de Einthoven afirma que sí en algún momento dado se conocen los potenciales eléctricos de dos cualesquier de las tres derivaciones electrocardiográficas bipolares de las extremidades, se puede determinar matemáticamente la tercera simplemente sumando las dos.

Habitualmente se registran seis derivaciones estándar del tórax, una cada vez, desde la pared torácica anterior, de modo que el electrodo del tórax se coloca secuencialmente en los seis puntos que se muestran en la Fig. 4, los diferentes registros se conocen como derivaciones V1, V2, V3, V4, V5 Y V6.

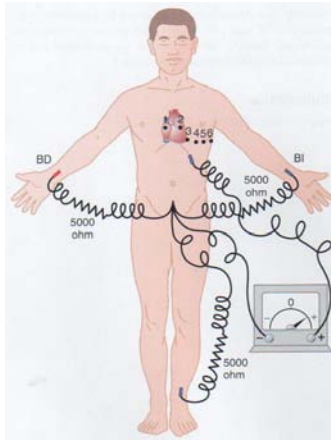


Figura 4: Conexiones de los electrodos para registrar las derivaciones del tórax. BD (Brazo derecho); BI (Brazo izquierdo).

Como las superficies del corazón están próximas a la pared torácica, cada una de las derivaciones del tórax registra principalmente el potencial eléctrico de la musculatura cardíaca que está inmediatamente abajo del electrodo. Por tanto, alteraciones relativamente pequeñas de los ventrículos, particularmente de la pared ventricular anterior, pueden producir grandes alteraciones de los electrocardiogramas que se registran en las derivaciones del tórax. Otro sistema de derivaciones que se utiliza mucho es la derivación unipolar ampliada de las extremidades. En este tipo de registro, dos de las extremidades se conectan mediante resistencias eléctricas a la terminal negativa del electrocardiógrafo, y la tercera extremidad se conecta a la terminal positiva. Cuando la terminal positiva está en el brazo derecho la derivación se conoce como derivación AVR, cuando se encuentra en el brazo izquierdo es la derivación AVL y cuando se encuentra en la pierna izquierda es la derivación AVF, [2].

II. MÉTODO

En la realización de un electrocardiograma con equipo profesional, el médico puede escoger la derivación que le interesa observar, ya sea para visualizarlo gráficamente o bien para imprimirlo, por lo tanto, el diagrama electrónico del electrocardiógrafo debe estar compuesto por una etapa de selección de los electrodos, que formarán

la combinación que entrará a la etapa diferencial para ser amplificada [4,5,6,7].

Desde luego que si se colocarán 12 etapas diferenciales de amplificación y filtraje, una para cada derivación electrocardiográfica, se tendría una buena alternativa para la conformación del circuito. Esta alternativa redundaría en un costo 12 veces mayor en la construcción del circuito además repercutiría de manera directa en el tamaño del circuito completo.

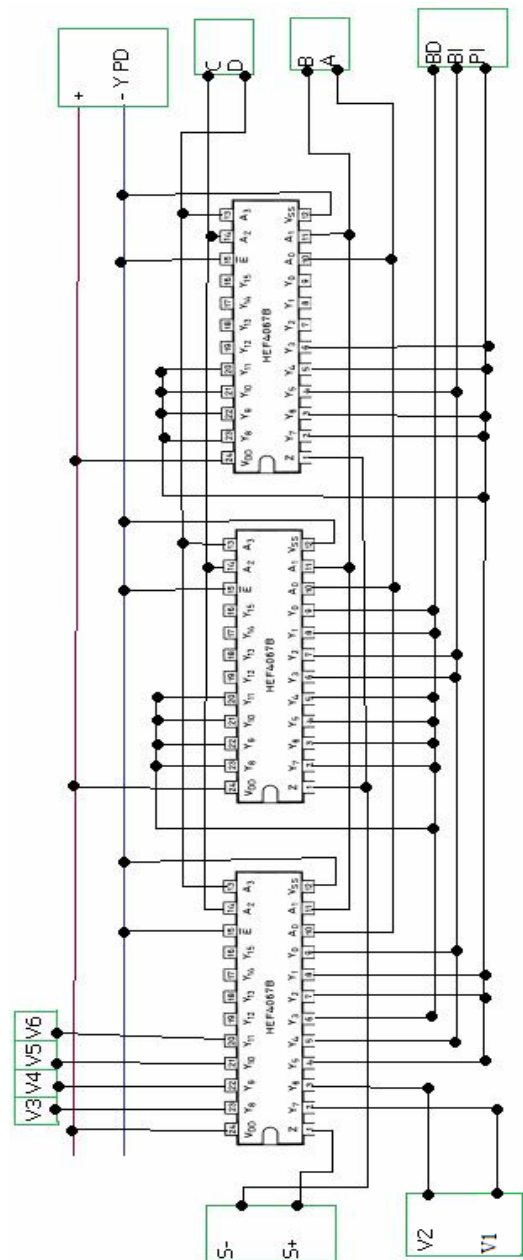


Figura 5: Circuito con 3 CMOS

Debido a que el acondicionamiento de las señales electrocardiográficas no depende del tipo de derivación que se pretenda adquirir, es decir, es idéntica en cualquiera de los 12 casos, es suficiente colocar una sola etapa, acompañada de un buen selector de derivaciones.

Es natural pensar en el uso de un elemento de la familia MOS (semiconductor a base de metal oxido), por aproximarse más a un conmutador ideal. El elemento lógico MOS fundamental es el inversor que se muestra en la figura 4 (a).

El dispositivo superior (subido) es en realidad un transistor en modo de agotamiento fue diseñado para funcionar como un resistor relativamente grande cuando se conecta como se observa en la figura 4 (a) y cuando el conmutador inferior de transistor está cerrado. Para fines de este análisis simplificado, basta considerarlo como una gran resistencia, como se observa en la figura 4 (b). Cuando x es el 0 lógico y V_x es 0 volts, como en la figura 4 (c) el conmutador inferior del transistor (bajado) está abierto. Debido a que en la resistencia R_D no hay corriente, el voltaje de salida es +5 volts y la salida lógica, z , es 1. Cuando $x=1$ y V_x es positivo, el conmutador inferior del transistor está cerrado, como se observa en la figura 4 (d). Ahora la salida está conectada en 0 volts y $z = 0$.

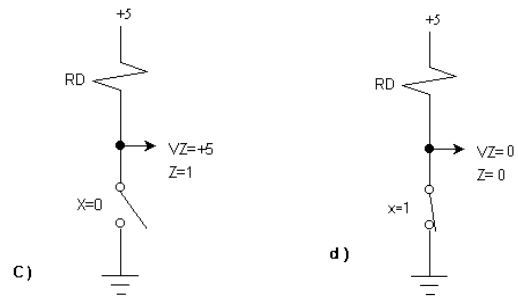
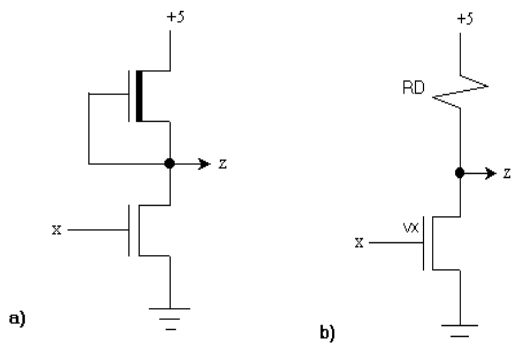


Figura 6, (a), (b), (c) y (d): Elementos MOS.

Para la construcción del selector realizado en este trabajo se utilizó el integrado HEF4067BMSI, [3] basado en la tecnología CMOS, el cual es un Multiplexor que trabaja de la sig: forma:

- i-) Modo A. 16 canales de entrada a 1 canal de salida.
- ii-) Modo B. Con 1 canal de entrada y 16 canales posibles de salidas.

Este integrado cuenta con un pin (control binario) denominado “inhibit” el cual trabaja en modo A, al mantenerlo en estado bajo, consta también de 4 pines (controles binarios) denominados A, B, C y D con los cuales, al asignarles combinaciones valores de valores binarios se puede escoger el canal de salida.

En este trabajo implementó un arreglo de 3 CMOS como se puede ver en la Fig. 5. Donde la notación que se siguió es la siguiente:

PD: pie derecho (aterrizado).

PI: pie derecho.

BD: brazo derecho

BI: brazo izquierdo.

V1: (EI1) 4^{to} espacio intercostal (entre las costillas) inmediatamente a la derecha del esternón.

V2: (EI2) 4^{to} espacio intercostal inmediatamente a la izquierda del esternón.

V3: (C1) directamente en la costilla entre V2 y V4.
 V4: (EI3) 5^{to} espacio intercostal – línea media clavicular izquierda.

V5: (EI4) 5^{to} espacio intercostal – línea axilar anterior izquierda.

V6: (EIM5) 5^{to} espacio intercostal – línea medio axilar izquierda.

En la tabla 1 se muestran las combinaciones de los estados posibles A, B, C y D que recibirá al mismo tiempo cada integrado, para determinar las salidas de los distintos canales y así activar cada una de las derivaciones. Se muestran también sus respectivas extremidades.

Como se mencionó anteriormente, una vez que se tienen seleccionadas las extremidades que serán analizadas, éstas deben pasar a una etapa de diferencial de preamplificación y luego a una etapa de filtrado, para ser desplegadas, almacenadas e impresas en el papel, que mostrará el registro electrocardiográfico.

Esas etapas no son estudiadas a lo largo de este trabajo.

DERIVACION	SELECTOR	CMOS 1		CMOS 2		CMOS 3	
NOMBRE	ABCD	CANAL	EXTREMIDAD	CANAL	EXTREMIDAD	CANAL	EXTREMIDAD
D1	0000	0	BI	0	BD	NC	NC
D2	1000	1	PI	1	BD	NC	NC
D3	0100	2	PI	2	BI	NC	NC
AVR	1100	3	BD	3	BI	3	PI
AVL	0010	4	BI	4	BD	4	PI
AVF	1010	5	PI	5	BD	5	BI
V1	0110	6	EI1	6	BD	6	PI
V2	1110	7	EI2	7	BD	7	PI
V3	0001	8	C1	8	BD	8	PI
V4	1001	9	EI3	9	BD	9	PI
V5	0101	10	EI4	10	BD	10	PI
V6	1101	11	EI5	11	BD	11	PI

*NC (No Conectar).

Tabla 1. Extremidades que conforman la derivación, combinación de ABCD necesaria y canales de salida de cada CMOS

Por ejemplo: para activar la primera derivación D1 se necesita colocar A, B, C, y D en estado bajo en los dos primeros integrados y en el tercero en estado NC, esto activará el canal cero en los dos primeros que corresponden a los electrodos conectados en BI y BD, usando la convención de que la pierna derecha esta acoplada a la tierra del electrocardiógrafo.

III. CONCLUSIONES

Se ha presentado un nuevo método para seleccionar y multiplexar las diferentes derivaciones posibles que son analizadas a lo largo de un estudio completo electrocardiográfico. Este trabajo está basado en el uso de integrados de tecnología CMOS muy comerciales y de bajo costo, los cuales pueden funcionar como multiplexores/demultiplexores de 16/1.

Este integrado es bastante eficiente para este tipo de aplicaciones debido a que presenta una baja resistencia de entrada en estado bajo, así como una alta resistencia por canal de salida en estado alto y una característica principal es el rango de corriente de entrada, el cual oscila entre los 100nA y 1 μ A, por lo que resulta bastante adecuado para su acoplamiento directo a los electrodos.

Este arreglo constituye la primera etapa que conforma un electrocardiógrafo, que puede ser acoplada directamente a la etapa de amplificación diferencial del mismo, que está fuera del alcance de este trabajo y que representa nuestra principal perspectiva.

IV. AGRADECIMIENTOS

Este trabajo fue soportado parcialmente por el proyecto PROMEP/103.5/08/4722.

V. REFERENCIAS

- [1] Arthur C. Guyton, John E. Hall, *Tratado de Fisiología Médica*, Ed.11, Edit. Elsevier España, 2006.
- [2] Dale Davis, *Interpretación del ECG*, Editorial Medica Panamericana, Buenos Aires, 1986.
- [3] *16-channel analogue multiplexer/demultiplexer HEF4067BMSI*, <http://www.chipcatalog.com/Datasheet/F8397C81F65CB241ADF26F164C057EF.htm>.
- [4] Hernández-Ledezma FU, Cordova-Fraga T, et al. “*Instrumentation for a Device to Recording Bioelectrical Signals of the Heart and Stomach*”, International Journal of Bioelectromagnetism Vol. 9, no. 1, pp. 21- 22, 2007.
- [5] Hernández-Ledezma FU, Córdoba-Fraga T, et. al, “*Instrumentación y uso de un dispositivo para medir actividad eléctrica de corazón y estómago*”,

Revista Mexicana de Cardiología, Vol. 20, no. 1, pp. 29-34, 2009.

- [6] Thaddeus R. F. Fulford-Jones, Gu-Yeon Wei, Matt Welsh “*A Portable, Low-Power, Wireless Two-Lead EKG System*”, Proceedings of the 26th Annual International Conference of the IEEE EMBS San Francisco, CA, USA September 1-5, 2004.
- [7] José J. Segura-Juárez, David Cuesta-Frau, Luis Samblas-Pena, and Mateo Aboy, “*A Microcontroller-Based Portable Electrocardiograph Recorder*”, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 51, NO. 9, SEPTEMBER 2004.

VI. AUTORES

Dr. Mario Eduardo Cano González es Físico de formación con especialidad en Biofísica y obtuvo su grado en el Instituto de Física de la Universidad de Guanajuato complementado con una estancia en el grupo de Biomagnetismo de la Facultad de Filosofía Ciencias e Letras de Ribeirao Preto, Sao Paulo, Brasil

Dr. Eri Atahualpa Mena Barbosa es Físico de Formación, especialista en Física Matemática y obtuvo su grado en el Instituto de Física de la Universidad de Guanajuato.

Mauricio Tavares es ingeniero en computación de especialista en Hardware e interfases, actualmente realiza su tesis en el Centro Universitario de la Ciénega de la Universidad de Guadalajara.

Jorge A. García es ingeniero en computación de especialista en Hardware e interfases, actualmente realiza su tesis en el Centro Universitario de la Ciénega de la Universidad de Guadalajara.

Israel Hernández es ingeniero en computación de especialista en Hardware e interfases, actualmente realiza su tesis en el Centro Universitario de la Ciénega de la Universidad de Guadalajara.

M. en C. Jose Avila Paz obtuvo su título de Maestría en Ciencias con especialidad en Computación Aplicada, por la Universidad Central "Marta Abreu" de las Villas, y es profesor de la Universidad de Guadalajara del Centro Universitario de la Ciénega.

SISTEMA DE MONITOREO DE LA PLANTA TRATADORA DE AGUAS RESIDUALES DE LA UAM-AZCAPOTZALCO

Magos Rivera Miguel ⁽¹⁾, Cruz Ramírez Ariel H., Soldevilla Juárez Francisco L., González Aragón Abelardo ⁽²⁾.

Universidad Autónoma Metropolitana—Azcapotzalco.

(1) Departamento de Electrónica, (2) Departamento de Energía
Av. San Pablo No. 180, Col. Reynosa Tamaulipas, México, D.F. 02200, México

(1) Tel. 53189550, correo electrónico: mrm@correo.azc.uam.mx

(2) Tel. 53189080, correo electrónico: gaa@correo.azc.uam.mx

RESUMEN

En este documento se describe la instrumentación y monitoreo de distintas variables realizados en puntos representativos de la planta de tratamiento de aguas residuales de la UAM-Azcapotzalco. El sistema desarrollado permite monitorear las variables en forma local mediante indicadores y/o en forma remota por medio de un programa de cómputo. Para el modo remoto, se emplearon transmisores digitales para enviar las señales a la computadora. La información del comportamiento de las variables se presenta en la pantalla de la computadora mediante instrumentos virtuales elaborados en Labview. Asimismo, el programa permite comparar el comportamiento de las señales, así como almacenar la información para su posterior análisis.

Palabras clave: Tratamiento de agua, Instrumentación, Monitoreo y Adquisición de Datos, Transmisores, Sensores.

I. INTRODUCCIÓN

Uno de los retos de mayor envergadura al cual se enfrentan actualmente las grandes ciudades del mundo es abastecer de agua potable a un número cada vez mayor de habitantes. Según proyecciones de Conapo, entre 2005 y 2025, la población en México se incrementará en 15 millones de habitantes de los cuales más del 90% se ubicará en zonas urbanas donde los recursos hídricos son limitados [1]. Aunado a esta situación, la

producción de aguas residuales tanto de tipo doméstico como industriales tenderá a incrementarse en proporciones similares.

Con el fin de liberar agua potable para el consumo humano, así como reducir la contaminación de ríos y mares, en las últimas décadas el tratamiento de las aguas residuales se ha incrementado en forma considerable [2, 3]. El agua obtenida a partir de plantas de tratamiento puede satisfacer la demanda en aplicaciones que no requieran de agua potable como son: riego de áreas verdes, llenado de lagos y fuentes, recarga de acuíferos y riego agrícola, entre otras.

El plantel Azcapotzalco de la Universidad Autónoma Metropolitana cuenta con una población cercana a 15000 personas, entre estudiantes, profesores y personal administrativo. Parte de las aguas residuales generadas en baños, laboratorios y comedores son tratadas en la *Planta Piloto de Tratamiento de Aguas Residuales* [4]. Esta planta tiene un doble objetivo, ya que además de producir agua tratada para el riego de las áreas verdes del plantel, también está concebida como instalación de demostración y prácticas para los alumnos de las diversas licenciaturas y posgrados que se imparten en la Universidad.

La planta se compone de una combinación de operaciones unitarias que tienen como objetivo disminuir la cantidad de contaminantes en el agua tratada hasta valores que muestren que tiene la

calidad mínima para ser utilizada como agua de riego.

El monitoreo de algunas variables físicas en las distintas etapas del tratamiento permite conocer la eficiencia de los procesos involucrados y comprobar que la calidad del agua a la salida del sistema es correcta. De esta forma, en caso de no cumplir con los requerimientos mínimos, los operadores pueden realizar los ajustes necesarios.

En este trabajo se describen los elementos y la programación que conforman el *Sistema de Monitoreo de la Planta Tratadora de Aguas Residuales de la UAM-Azcapotzalco*. En la segunda sección de este documento se describe en forma sencilla las partes básicas de la planta de tratamiento indicando los puntos de monitoreo. La sección tres presenta los elementos que conforman la instrumentación del sistema: sensores y transmisores tanto analógicos como digitales. El programa desarrollado en Labview se describe en la cuarta sección del artículo. Por último se presentan algunos resultados y conclusiones.

II. DESCRIPCIÓN DEL PROCESO

La *Planta Tratadora de Aguas Residuales de la UAM-Azcapotzalco* tiene una capacidad nominal de procesamiento de 0.5 l/s siendo el máximo 0.75 l/s. Está compuesta por una serie de operaciones que tienen como objetivo el reducir la concentración de algunos constituyentes a niveles previamente establecidos.

Pretratamiento. La finalidad de esta etapa inicial es la de reducir por medio de trituradores el tamaño de la materia orgánica contenida en el agua y permitir, por sedimentación, separar el arena arrastrada.

Coagulación-Floculación. En esta etapa, mediante la adición de coagulantes, se favorece la agrupación de las partículas. Posteriormente se agregan polímeros al líquido con el fin de aumentar el tamaño y densidad de los grumos, lo que facilita la precipitación de la materia orgánica en la siguiente etapa.

Sedimentación. En esta operación se realiza, por medios físicos, la separación de los grumos generados en la etapa anterior. Se emplea para esto, un sedimentador de placas paralelas con tiempo de residencia promedio de 135 minutos.

Filtración. El efluente resultante de la sección anterior, pasa ahora por una torre de filtración en la cual se separan los sólidos finos que aun restan en el flujo. Una unidad de microfiltración final permite eliminar microorganismos así como partículas remanentes.

Desinfección. En la última etapa del proceso se realiza una doble desinfección del agua, la primera consiste en agregar agentes químicos (por ejemplo, hipoclorito de sodio) y la segunda se basa en la exposición del líquido a luz ultravioleta.

Las variables monitoreadas en esta primera etapa del proyecto son: **Potencial de óxido-reducción (ORP), pH, flujo y temperatura.** Los puntos en los cuales se realiza el monitoreo de las variables mencionadas son: entrada de la planta, a la salida de la filtración/oxidación y a la salida de la planta. Cabe mencionar que la variable temperatura no es monitoreada a la salida de la planta.

El diagrama de la figura 1, muestra la secuencia de las operaciones que se llevan a cabo en la planta, así como los puntos de monitoreo.

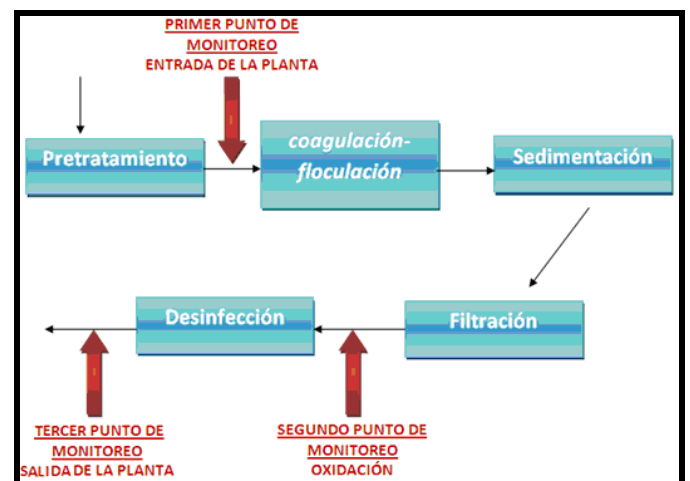


Fig. 1 Secuencia de operaciones en la planta.

III. INSTRUMENTACIÓN DEL SISTEMA

Sensores. Como ya se mencionó, las variables que se supervisan en el proceso son: ORP, pH, flujo y temperatura, en tres puntos del proceso. Todos los sensores seleccionados cuentan con indicadores que permiten monitorear y ajustar en forma local los parámetros del proceso sin necesidad de desplazarse hasta el lugar en el que se encuentra la computadora de supervisión y adquisición de datos. Una característica de estos indicadores es que incluyen transmisores de tipo analógico (salidas de 4 a 20 mA) que permiten el envío de la información a puntos remotos de la planta. Son estas salidas analógicas que nos permitirán mediante el uso de transmisores digitales llevar las señales hasta la computadora. Tanto los sensores como los indicadores son de la compañía GF Signet.

ORP/pH. El sensor utilizado es el modelo: **2757**. Este elemento puede ser empleado para sensar ORP o pH. Se trata de un electrodo que a través de un amplificador externo modelo **2750** de la misma compañía, genera la señal que requiere el indicador modelo **8750**, figura 2. Las principales características del conjunto electrodo-amplificador-indicador se listan a continuación [5, 6, 7]:

- Rango de medición para ORP (redox): ± 1500 mV
- Rango de medición para pH: 0 a 14
- Alimentación: 12 a 30 VDC.
- Salida: 4 a 20 mA.

Flujo. Se empleó un sensor de tipo turbina modelo: **2536**. El rango de medición de la velocidad del fluido es: de 0.3 a 6 m/s, entregando como señal de salida pulsos de 49 Hz por m/s [8]. La alimentación del sensor puede variar entre 3.5 a 24 VDC. El indicador empleado es el modelo **8550** cuya alimentación es de 12 a 30 VDC y la salida es de 4 a 20 mA [9]. Los elementos se muestran en la figura 3.

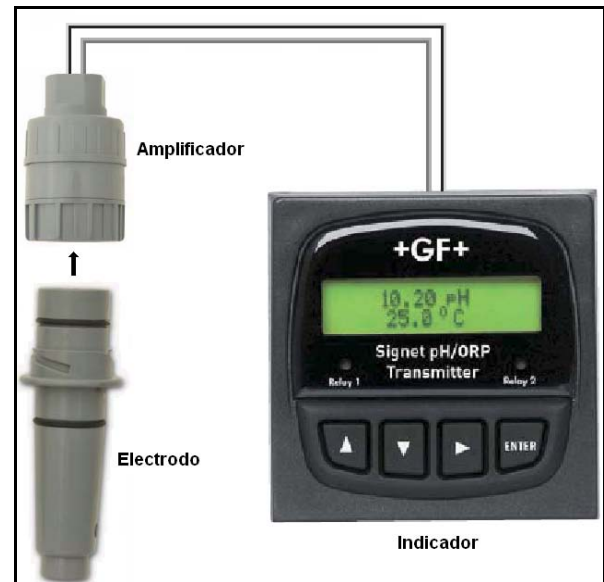


Fig. 2 Conjunto electrodo-amplificador-indicador de ORP/pH.



Fig. 3 Sensor de flujo tipo turbina con indicador.

Temperatura. En este caso se utiliza el sensor modelo **2350**, se trata de un sensor de tipo RTD con un rango de medición de 0 a 100°C [10]. El indicador empleado es el modelo **8350** cuyo voltaje de alimentación es de 12 a 30 VDC y la salida es de 4 a 20 mA [11]. Los elementos se muestran en la figura 4.

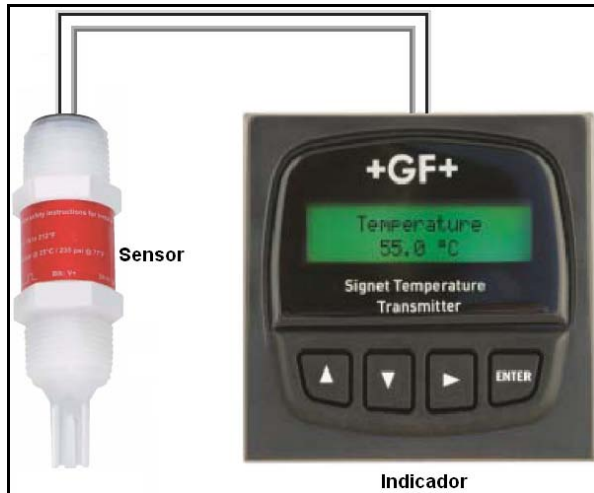


Fig. 4 Sensor de temperatura tipo RTD con indicador.

Módulos de Adquisición de Datos.

Las señales analógicas enviadas por los indicadores son digitalizadas e introducidas a la computadora con ayuda de módulos de adquisición de datos de la compañía Advantech, el modelo seleccionado para esta aplicación es el **ADAM4018**, figura 5. Estos dispositivos digitalizan las señales analógicas conectadas en sus entradas (hasta ocho entradas) para posteriormente transmitir las al puerto serie de cualquier computadora. El envío de la información se realiza bajo un estándar digital denominado RS-485 que emplea únicamente dos hilos para mandar las señales de las ocho entradas. Se trata de transmisión tipo serial parecida a la que manejan la mayoría de las computadoras de escritorio la que a su vez se denomina RS-232. La diferencia radica principalmente en los niveles de voltaje, lo que se refleja en la distancia que es posible transmitir. Bajo RS-232 es posible el envío sin pérdida de información algunas decenas de metros, mientras que bajo RS-485 la distancia puede rebasar los mil metros. [12]

Una característica importante de estos dispositivos es que pueden funcionar en forma modular, esto es, si se requiere enviar más señales se agrega otro módulo de ocho entradas el cual se conecta al mismo par de cables, pudiéndose conectar hasta 32 módulos.



Fig. 5 Módulo de adquisición de datos ADAM 4018.

La modularidad de estos dispositivos, así como su distancia de transmisión, posibilitó el envío hasta la computadora bajo formato digital de las señales de los 11 sensores empleando sólo dos cables, en lugar de 22 (2 por señal) que hubiesen sido necesarios bajo un estándar de transmisión analógico.

En este proyecto se propuso la colocación de un módulo de ocho entradas en cada punto de monitoreo por lo que queda abierta la posibilidad de monitorear variables adicionales sin necesidad de realizar mayores modificaciones al sistema.

Finalmente, es necesaria la incorporación de un convertidor que haga compatibles las señales digitales transmitidas por los módulos RS-485 y las señales que el puerto serial de una computadora puede recibir, RS-232. El módulo convertidor que se utiliza es el: ADAM-4520, este elemento se colocó cerca de la computadora ya que en su salida tenemos las señales bajo estándar RS-232 y como ya se mencionó, esta no puede transmitirse a distancias considerables. Los módulos de adquisición de datos así como el convertidor RS-485 a RS-232, requieren de una alimentación de voltaje en el rango de 10 a 30 VDC.

Tableros de Conexiones. Los indicadores locales así como los módulos de adquisición de datos fueron colocados en tres gabinetes, uno por punto de monitoreo. Se usaron gabinetes con protección contra salpicaduras, marca Himel modelo: CRN-

64250M, con las siguientes dimensiones: Alto: 60 cm, Ancho: 40 cm, Prof: 25 cm.

Se tiene un tablero adicional, colocado al lado de la computadora de monitoreo, el cual contiene el módulo de conversión RS-485 a RS-232 así como la fuente de alimentación para todos los dispositivos. Se trata de un gabinete marca Himel, modelo CRN-2025150 KT -M con dimensiones: Alto: 30 cm, Ancho: 25 cm, Prof: 15 cm.

Considerando el consumo de energía para cada uno de los dispositivos que conforman el sistema, se determinó que era necesaria una corriente máxima de 2.5 A., por lo que se eligió una fuente de voltaje con una capacidad ligeramente superior: 3 A. Se trata del modelo: DR 75-24 del fabricante Moxa Inc., para montaje en riel DIN con voltaje de salida de 24 VDC.

En este último tablero se encuentra el interruptor general del sistema. A su vez cada uno de los tableros locales cuenta con su propio interruptor. Si el interruptor general está encendido se activa uno de los indicadores de los tableros locales. El segundo indicador se activará, junto con los elementos del tablero en cuestión, al momento de encender localmente éste. Lo anterior permite tener en operación sólo algunos de los tableros. Esto es útil cuando se realizan pruebas de funcionamiento en alguna sección específica del proceso.

La figura 6, muestra en forma esquemática la conexión entre los tableros y la computadora de monitoreo.

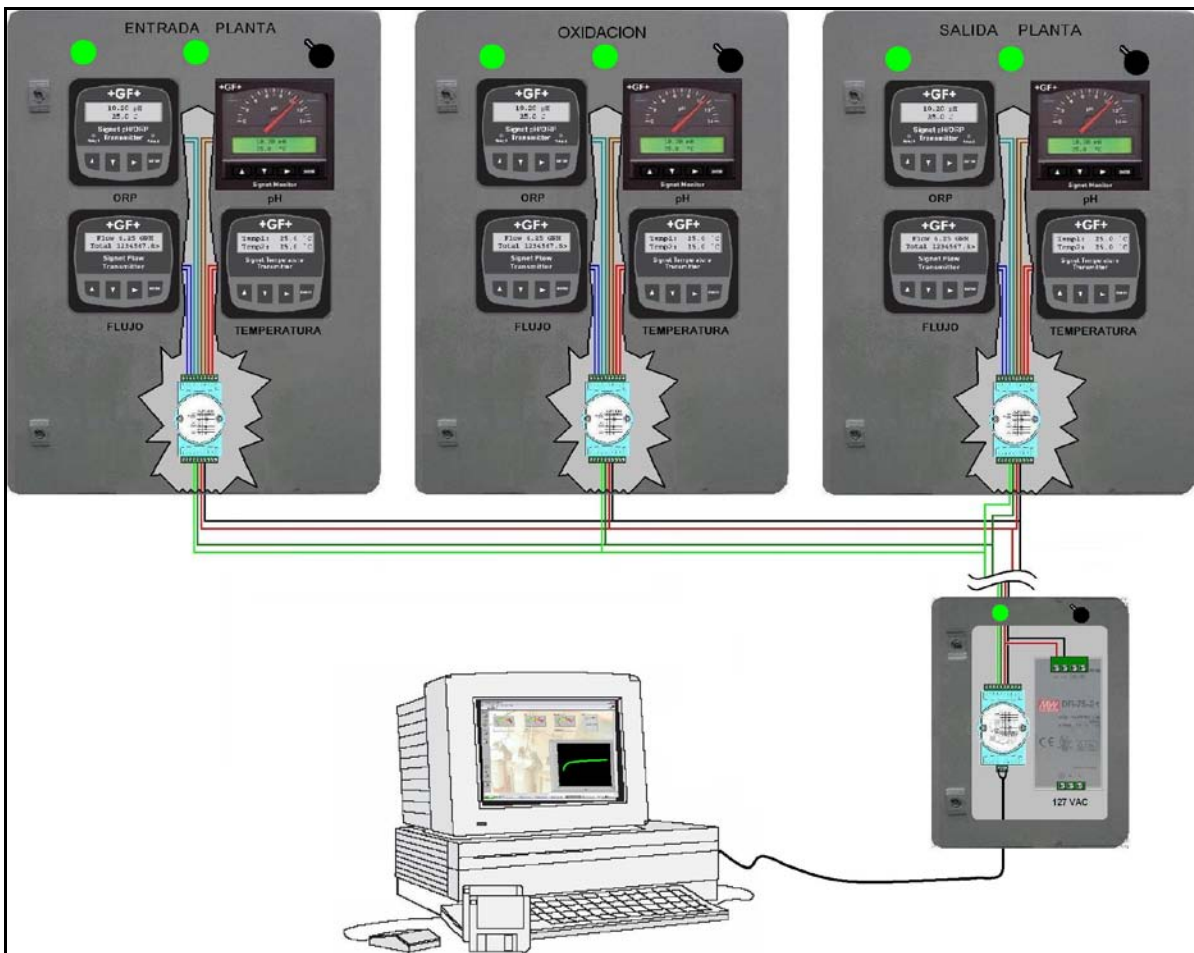


Fig. 6 Diagrama de conexiones del sistema.

IV. PROGRAMA EN LABVIEW

Los elementos descritos en la sección anterior, permiten llevar hasta la computadora las señales de cada uno de los sensores instalados en el sistema. A continuación describiremos el programa de cómputo elaborado para leer, desplegar, y almacenar la información.

La herramienta utilizada para realizar dicha tarea fue **LabView** de National Instruments. Se trata de un paquete de cómputo que cuenta con funciones para la adquisición, el análisis y la presentación de datos, así como funciones que permiten la comunicación con instrumentos de medición y control. Por medio de una interfase gráfica orientada a objetos, el usuario programa su aplicación conectando entre sí iconos seleccionados de una serie de cajas de herramientas. La elaboración de las pantallas de presentación se realiza eligiendo de entre una variedad de dispositivos de despliegue, los cuales son colocados en la posición deseada en la pantalla de la computadora [13, 14].

Pantallas de Presentación.

Al ejecutar el programa, aparece una pantalla que permite al usuario indicar cuáles son los tableros que se encuentran encendidos y por lo tanto qué puntos del proceso serán monitoreados, figura 7.

Una vez seleccionados los puntos de monitoreo se entra directamente a la pantalla correspondiente. Se describe a continuación el caso en el que se tienen los tres tableros funcionando.

El sistema base consta de tres pantallas, una para cada punto de monitoreo, en las cuales se despliega el comportamiento de las variables.

La figura 8, muestra la imagen de la pantalla de la entrada de la planta.

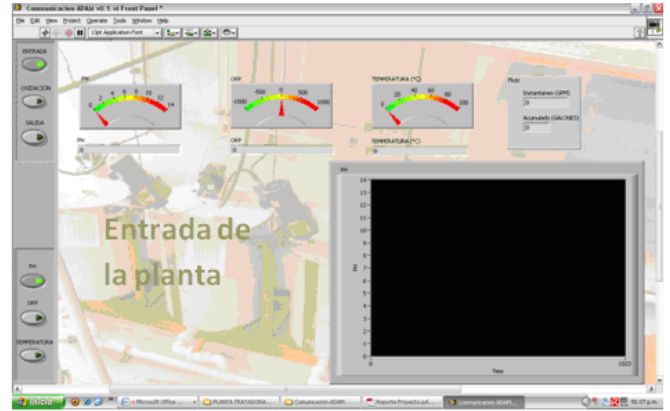


Fig. 8 Formato de las pantallas.

Se observa que para las variables: pH, ORP y temperatura, se colocaron indicadores analógicos en combinación con indicadores numéricos. Para la variable flujo sólo se colocaron indicadores numéricos. Pero para esta última variable, además de la indicación del valor instantáneo, se tiene un indicador con el valor acumulado de fluido que ha circulado en el punto de monitoreo. Finalmente, se tiene un graficador que muestra el comportamiento en el tiempo de una de las variables del proceso, excluido el flujo. Cabe recordar que la variable temperatura no es monitoreada a la salida de la planta por lo que los indicadores correspondientes no aparecen en esta pantalla, ni el botón asociado de graficación.

La selección del punto de monitoreo, así como de la variable a graficar se realiza mediante los botones colocados en el lado izquierdo de las pantallas, figura 9. La combinación que se muestra, indica que se despliegan los valores de las variables a la salida de la operación de oxidación y que se grafica el comportamiento de la variable pH en este punto.

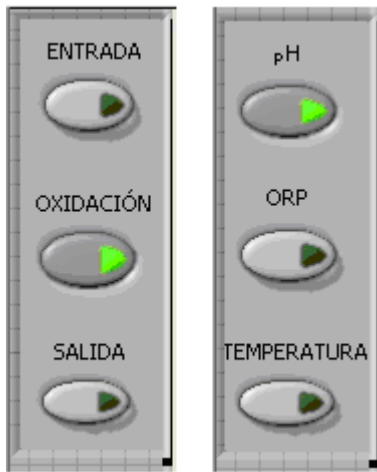


Fig. 9 Botones de selección de punto de monitoreo y variable a graficar.

Todos los datos de cada una de las variables son almacenados en archivos compatibles con Excel de tal forma que sea posible analizar el comportamiento de la planta posteriormente.

Programa de la Aplicación.

El programa de la aplicación consta de diversos bloques que realizan las acciones necesarias para solicitar información a los módulos ADAM respecto de los valores de las señales enviadas por los sensores. Se ocupa del procesamiento de la información recibida para su presentación y posterior almacenamiento. El diagrama de bloques de la figura 10, muestra la estructura del programa desarrollado, el cual se describe a continuación.

Bloque de Almacenamiento de los Comandos para los Módulos ADAM: Este bloque almacena los comandos que controlan a los módulos ADAM. Dichos comandos están en ASCII.

Bloque de Comunicaciones: Contiene todo lo referente a la comunicación.

Configuración del Puerto: Permite realizar la configuración del puerto que se utilizará para las comunicaciones (selección del puerto, velocidad, paridad, etc.).

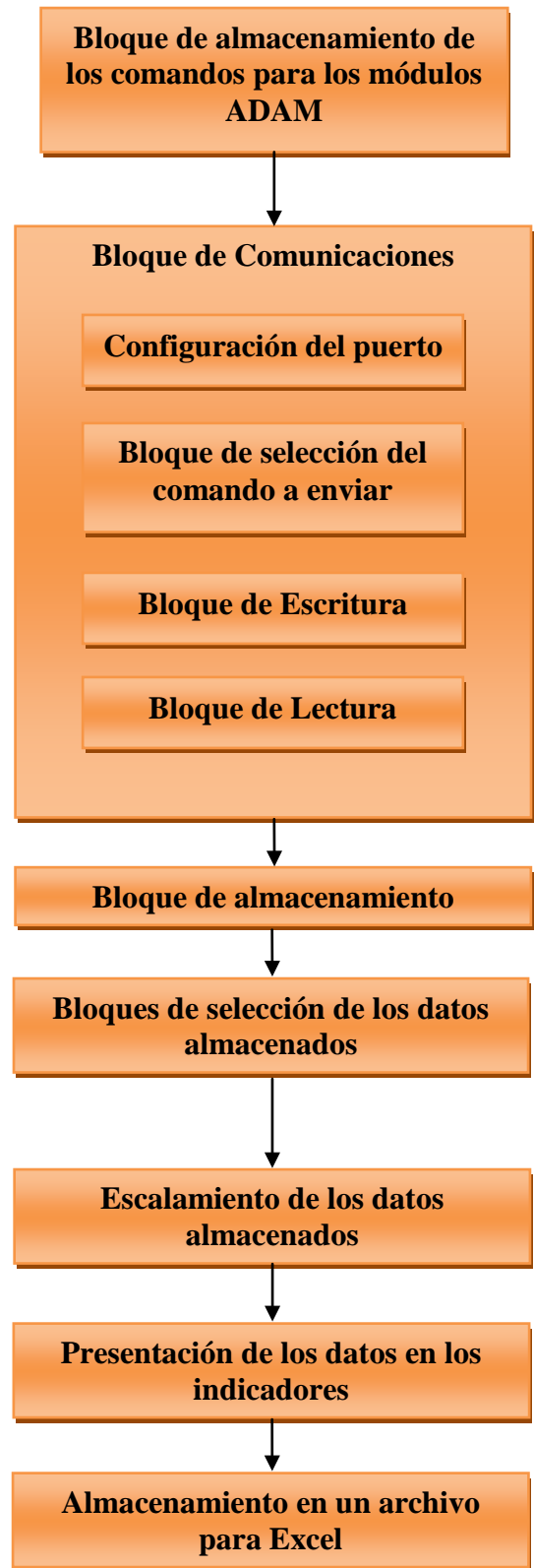


Fig. 10 Diagrama de bloques que muestra la estructura del programa.

Bloque de Selección del Comando a Enviar:

Este bloque va seleccionando cada uno de los comandos del bloque de almacenamiento, que serán enviados posteriormente.

Bloque de Escritura: El comando seleccionado en el bloque anterior, es enviado a través del puerto por este bloque.

Bloque de Lectura: Este bloque recibe la información que llega al puerto como respuesta al comando enviado.

Bloque de Almacenamiento: Los datos que llegan por el puerto se van almacenando en una lista. El primer dato de la lista corresponde a la lectura del canal cero del primer módulo ADAM. El segundo dato corresponde a la lectura del canal 1 del primer módulo ADAM. Así sucesivamente, hasta completar todos los canales de los tres módulos ADAM.

Bloque de Selección de los Datos Almacenados: Del bloque de almacenamiento se van desglosando los datos que corresponden a cada sensor, para de esta forma, tener separadas las lecturas de cada punto de monitoreo.

Escalamiento de los Datos Almacenados: Las señales que llegan por el puerto corresponden a señales de 4 a 20mA de los transmisores. Para que muestren las lecturas correspondientes éstas necesitan ser escaladas.

Presentación de los Datos en los Indicadores: Este bloque contiene las terminales asociadas a todos los indicadores de las pantallas de presentación del programa. Las señales ya escaladas se conectan con el indicador correspondiente.

Almacenamiento en un Archivo para Excel: Las lecturas mostradas en cada indicador, se almacenan en un archivo con extensión lvm, el cual puede ser abierto por Excel.

En las figuras 11 y 12 se muestran dos bloques representativos del programa desarrollado. El primero muestra la forma como se almacenan los comandos para los módulos ADAM en un arreglo. Posteriormente, otra sección del programa selecciona cada uno de los datos almacenados en el arreglo anterior, para enviarlos por el puerto serie.

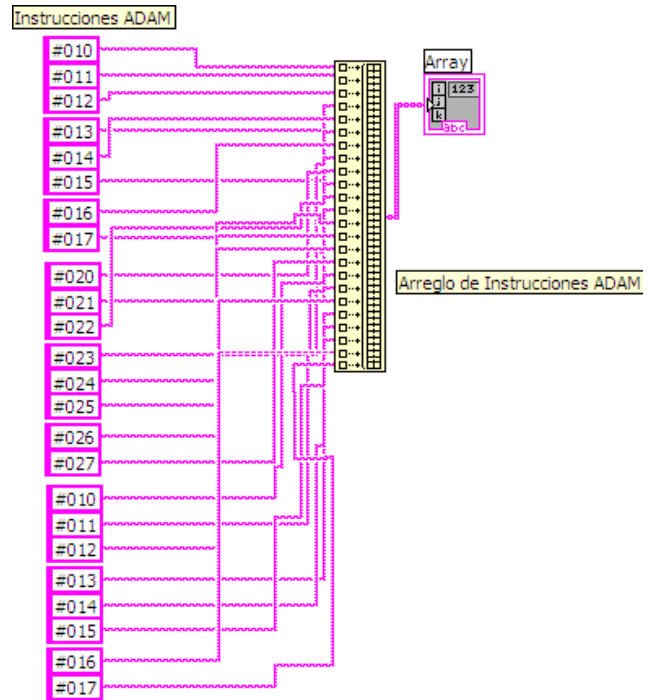


Fig. 11 Almacenamiento de los comandos de comunicación.

La figura 12, muestra los elementos empleados para realizar el intercambio de información entre los módulos de transmisión digital, colocados en los tableros, y el programa de la computadora.

Para que se pueda enviar y recibir información por el mismo puerto se debe seguir una secuencia, es decir, primero se transmite la información por el puerto, y después se espera a recibir algo. Esto se hace empleando una estructura llamada **Flat Sequence Structure**, la cual hace que el programa se ejecute secuencialmente, dependiendo de los eventos que se configuren. Para nuestro caso, solo necesitamos dos eventos, unos para mandar información por el puerto y otro para recibir información.

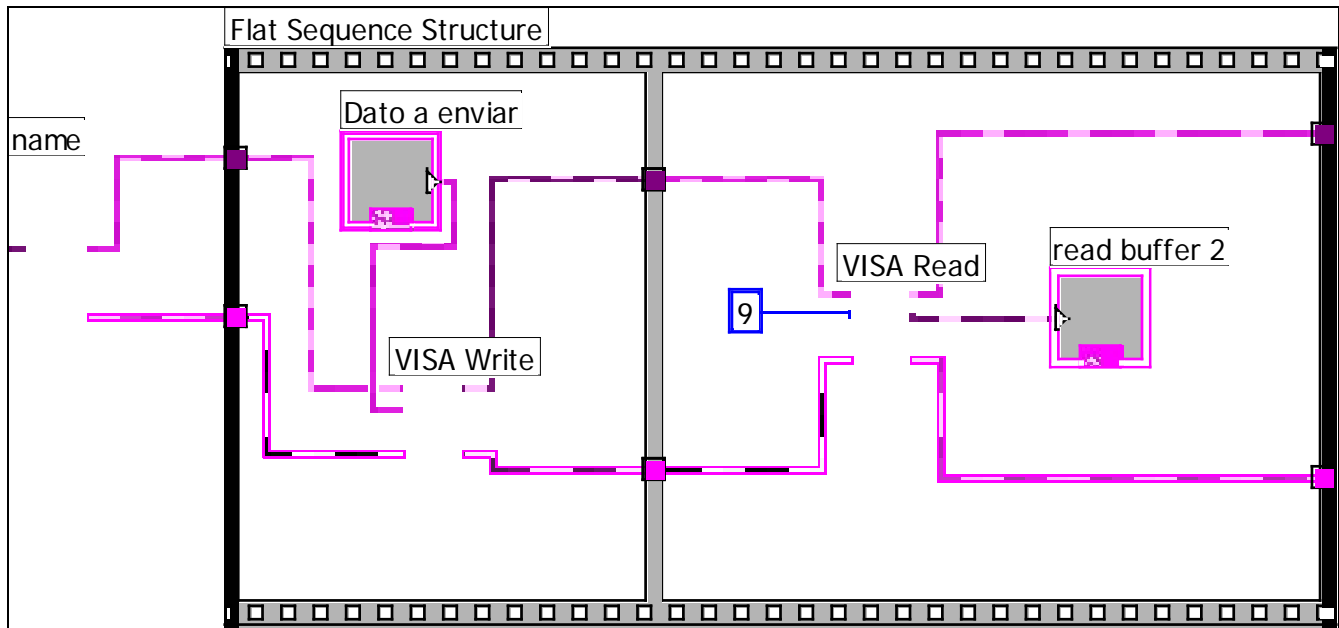


Fig. 12 Arreglo básico para la comunicación serial.

En la figura 12, se puede apreciar como el bloque Visa Configure Serial Port, que es el que configura el puerto, se encuentra fuera del ciclo, ya que es necesario que en todo momento en puerto este configurado.

En el primer cuadro del ciclo, se encuentra el bloque Visa Write, que es el que manda la información por el puerto. Después en el segundo ciclo, se ejecuta la segunda casilla, en la cual se encuentra el bloque Visa Read, que es el que se encarga de leer lo que llegue por el puerto.

V. CONCLUSIONES

Se ha presentado la secuencia de diseño y construcción de un sistema de monitoreo y adquisición de datos para la Planta Tratadora de Aguas Residuales de la UAM-Azcapotzalco. Desde el punto de vista electrónico el sistema funciona correctamente. Los usuarios están en la etapa de aprendizaje y familiarización con el manejo del sistema. A futuro se tienen contemplados diversos trabajos, desde el punto de vista de los operadores de la planta, este sistema les permitirá realizar diversas pruebas que anteriormente no se realizaban debido a lo

extenuante que resultaba tomar datos del comportamiento del proceso por periodos largos de tiempo. Desde el punto de vista del control e instrumentación, el sistema implementado abre las puertas a la posibilidad de controlar algunas de las variables del proceso. Asimismo, empleando los elementos instalados, se plantea el uso del proceso para fines académicos pudiéndose, por ejemplo, realizar el monitoreo y la adquisición de datos con programas elaborados en lenguajes y/o paquetes distintos a Labview. El sistema descrito permitirá, a estudiantes de las diversas ingenierías que se imparten en la UAM-Azcapotzalco, tener disponibles en una computadora las señales de un proceso real con las cuales poder realizar experimentación que complementen su formación en temas como Control, Sensores, Instrumentación Industrial e Interfaces Hombre-Máquina, por mencionar algunos.

VI. REFERENCIAS

- [1] Partida Bush, V., "Proyecciones de la Población de México 2005-2050". Consejo Nacional de Población. México D.F. 2006.
- [2] Restrepo, I., *et al.*, "Avances en Investigación y Desarrollo en Agua y

- Saneamiento para el Cumplimiento de las Metas del Milenio”. Universidad del Valle. Cali, Colombia. 2005.
- [3] Braulio Villalobos, M. A., *et al.*, “Operación y Rediseño de una Tecnología para el Tratamiento de Aguas Residuales en Cuemanco”, *Revista Mexicana de Ingeniería Química* Vol. 5, Supl. 1, pp. 5-9, Nov. 2006.
- [4] González Aragón, A., *et al.*, “Operación de la Planta Piloto de Tratamiento de Agua de la UAM-A. Antecedentes y Perspectivas”. 8ª Semana del Medio Ambiente. UAM-Azcapotzalco, Octubre 2008. México D.F.
- [5] *Signet DryLoc pH and ORP Electrodes*, GF Signet 2757 Operation Manual, George Fisher Signet Inc. 2001. USA.
- [6] *Signet 2750 DryLoc pH/ORP Sensor Electronics*, GF Signet 2750 Operation Manual, George Fisher Signet Inc. 2001. USA.
- [7] *Signet 8750 pH/ORP Transmitter*, GF Signet 8750 Operation Manual, George Fisher Signet Inc. 2001. USA.
- [8] *Signet 2536 Rotor –X Flow Sensors*, GF Signet 2536 Operation Manual, George Fisher Signet Inc. 2003. USA
- [9] *Signet 8550 Flow Transmitter*, GF Signet 2750 Operation Manual, George Fisher Signet Inc. 2002. USA.
- [10] *Signet 2350 Temperature Sensor*, GF Signet 2350 Operation Manual, George Fisher Signet Inc. 2003. USA
- [11] *Signet 8350 Temperature Transmitter*, GF Signet 8350 Operation Manual, George Fisher Signet Inc. 2001. USA.
- [12] *ADAM 4000 Data Acquisition Modules User’s Manual*, Advantech Co. Ltd. 2007. USA.
- [13] Robert H. Bishop, *Labview 8 Student Edition*, Prentice Hall, 2006.
- [14] Jose R. Lajara, Jose Pelegrí, *Labview: Entorno Gráfico de Programación* Marcombo, 2007.

ACTUADOR LINEAL ULTRASÓNICO PIEZOELÉCTRICO PARA PANTALLAS TÁCTILES

Hernández Mücke Hermes, Preza Hernández Enrique, Velázquez Guerrero Ramiro

Mecatrónica y Control de Sistemas - Universidad Panamericana,
Josemaría Escrivá de Balaguer 101, Fracc. Rústicos Calpulli,
20290, Aguascalientes, Ags., México

Tel: (449) 9106200, web: www.robotica-up.org, correo electrónico: rvelazquez@ags.up.mx

RESUMEN

Este artículo presenta el modelado y la caracterización de un actuador lineal ultrasónico piezoeléctrico para el diseño e implementación de pantallas táctiles. El actuador consiste de un vástago, un tambor deslizante y un transductor cerámico, y produce movimiento lineal basado en el principio de desplazamiento inercial. A diferencia de otros mecanismos, este actuador ofrece un diseño compacto y ultraligero que pretende hacer posible la realización de pantallas táctiles altamente portátiles para los invidentes y débiles visuales. Principio de accionamiento, caracterización, diseño y el estado actual del primer prototipo de pantalla táctil portátil se presentan y discuten en este artículo.

Palabras clave: actuador lineal miniatura, actuador piezoeléctrico, modelado y simulación, pantallas táctiles Braille.

I. INTRODUCCIÓN

Las aplicaciones para actuadores miniatura abundan en diversos campos como la medicina, la bio-ingeniería, la electrónica, la robótica, la aeronáutica y la industria automotriz.

Las soluciones de accionamiento han sido tan diversas con diseños que varían desde las tecnologías electromagnética, neumática e hidráulica, hasta las nuevas tecnologías llamadas inteligentes: piezoeléctrica, aleaciones con memoria de forma (AMF), polímeros y fluidos electrorheológicos.

De hecho, el interés en los materiales inteligentes ha crecido en la última década debido a sus propiedades únicas: esta clase de materiales tiene la capacidad de cambiar su forma, tamaño y dureza, entre otras propiedades, al aplicárseles un campo eléctrico, electro-magnético o térmico.

El presente artículo se enfoca particularmente en la necesidad de utilizar actuadores miniatura para pantallas táctiles. Las pantallas táctiles son dispositivos dedicados a estimular la piel, particularmente las yemas de los dedos, para crear sensaciones de contacto. Dichos dispositivos han sido desde hace mucho tiempo la base para el desarrollo de herramientas de transmisión de información para los invidentes, tanto para la lectura (en Braille) como para la representación gráfica de algunas imágenes en 2D. Recientemente las pantallas táctiles se han utilizado en aplicaciones de realidad virtual [1], realidad aumentada [2], reemplazo sensorial [3], tele-cirugía [4] y robótica [5], entre otras.

Las pantallas táctiles son todo un reto en el campo de los actuadores miniatura debido a que se requiere integrar una gran cantidad de actuadores en un espacio muy pequeño.

Braille, el estándar mundial para desplegar texto a los invidentes, es por mucho el sistema táctil más utilizado actualmente. Cada letra se representa por medio de una celda de 6 pines de contacto. El estándar especifica que los pines deben estar a 2.54 mm de separación entre ellos y sobresalir al menos 0.7 mm de la superficie. Así mismo, cada pin debe producir al menos 170 mN de fuerza de empuje a una frecuencia de 6 Hz.

La mayoría de los dispositivos Braille disponibles comercialmente están basados en vigas piezoeléctricas bicapa y en mecanismos tipo relevadores. Las vigas piezoeléctricas bicapa producen vibraciones que mueven el pin de contacto arriba/abajo cuando se les aplica un voltaje (Fig. 1(a)). Por su parte, los relevadores impulsan el pin de contacto mediante una pequeña palanca al aplicarse un voltaje. Cuando se retira el voltaje, el peso de la propia palanca hace que ésta descienda volviendo a su posición original (pin abajo).

Ambos mecanismos comparten un problema de diseño: requieren de un aparatoso módulo que contiene los 6 pines de contacto. Este módulo compromete la capacidad de integración de los actuadores. Actualmente los dispositivos Braille más portátiles tienen una masa de 1 kg y dimensiones de 25 x 15.5 x 5 cm³ [6].

Este artículo presenta un novedoso dispositivo Braille utilizando actuadores lineales piezoeléctricos ultrasónicos o *TULA* (Tiny Ultrasonic Linear Actuator). Los actuadores TULA presentan propiedades muy innovadoras como su reducido tamaño, su poca masa y permiten una gran capacidad de integración.

El resto del documento está organizado de la siguiente manera: la Sección 2 introduce el actuador TULA y evalúa sus propiedades de accionamiento. La Sección 3 describe el diseño de una celda Braille a base de actuadores TULA y presenta el primer prototipo desarrollado. Finalmente, la Sección 4 concluye resumiendo los puntos importantes de esta investigación y el trabajo futuro a realizar.

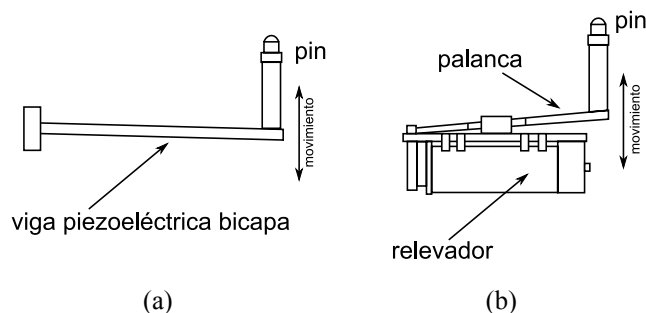


Fig. 1. Ejemplo de mecanismos Braille existentes: (a) viga piezoeléctrica bicapa y (b) relevador.

II. ACTUADOR PIEZOELÉCTRICO

A. Estructura y principio

El actuador TULA que se propone en este trabajo está constituido de un vástago, un tambor deslizante y un disco cerámico piezoeléctrico (Fig. 2(a)). Este actuador es miniatura y ultraligero: 2.4 mm de diámetro y 11.6 mm de largo; su masa es de apenas 370 mg (Fig. 2(b)). Nótese que todo el actuador ocupa un diámetro de 2.4 mm sin necesidad de algún otro acoplamiento o modulo externo.

El disco cerámico consiste de una parte rígida y un transductor. Cuando se aplica un voltaje, el transductor vibra generando fluctuaciones hacia arriba/abajo las cuales son transferidas al vástago y finalmente al tambor deslizante (Fig. 3).

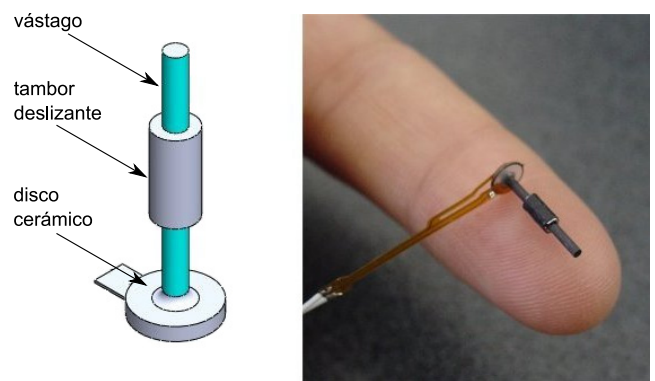


Fig. 2. TULA - Actuador lineal ultrasónico piezoeléctrico miniatura: (a) diseño y (b) prototipo [7].

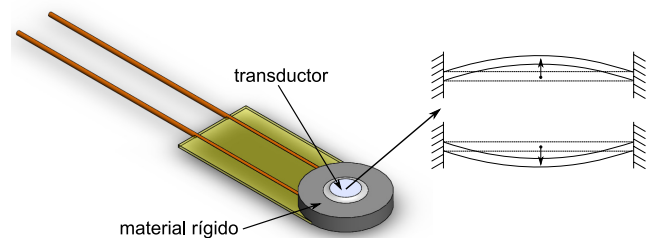


Fig. 3. Estructura del disco cerámico piezoeléctrico y vibración del transductor.

Su movimiento lineal está basado en el principio de desplazamiento inercial que se genera por el potencial eléctrico de una señal tipo diente de sierra [8].

El principio de operación se muestra en la fig. 4(a): (1) En la posición inicial, el actuador no está energizado, el vástago se encuentra en la posición A y el tambor deslizante en la posición x_0 . (2) Cuando se aplica el voltaje máximo, ambas partes, tambor y vástago, se mueven juntos una distancia A (por lo que la nueva posición del vástago es $2A$ y la del tambor x_0+A). (3). Conforme cae el voltaje a su mínimo, el vástago regresa una distancia $2A$. Sin embargo, el tambor se mantiene en su posición debido a su inercia. (4) Cuando el voltaje se incrementa de nuevo a su valor máximo, tanto el tambor como el vástago se mueven de nuevo juntos una distancia $2A$ (el vástago a la posición $2A$ y el tambor a la posición x_0+3A). El ciclo se repite: (5) el voltaje decrece, el vástago regresa a 0 mientras que el tambor se mantiene en x_0+3A . (6) El voltaje se incrementa, el vástago va a la posición $2A$ y el tambor a la posición x_0+5A .

La fig. 4(b) muestra la señal diente de sierra correspondiente.

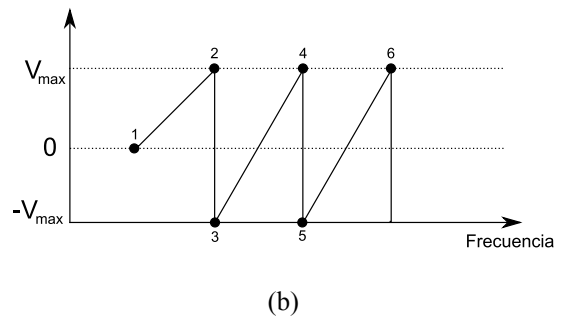
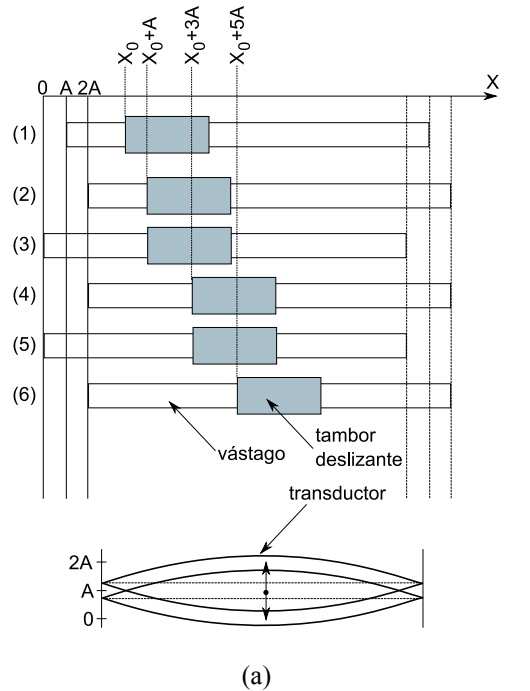


Fig. 4. (a) Principio de desplazamiento inercial y (b) señal de control tipo diente de sierra.

B. Actuador táctil

Los actuadores TULA han sido utilizados para diseñar y construir un nuevo actuador táctil (taxel¹) para pantallas táctiles. Éste se muestra en la fig. 5. Note que se le ha agregado un pin plástico al tambor deslizante para mejorar el contacto con los dedos.

Los taxels son bi-estables: debido a la fricción entre el tambor deslizante y el vástago, el tambor es capaz de mantener su posición sin utilizar energía. La energía sólo se utiliza cuando se cambia de una posición a otra. La bi-estabilidad evita el consumo inútil de potencia y la degradación gradual del desempeño del taxel.

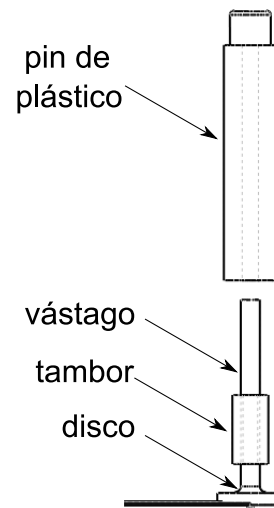


Fig. 5. Estructura del taxel.

¹ Taxel: tactile pixel (Píxel táctil).

C. Modelo y simulación del actuador TULA

Los cerámicos piezoeléctricos son un tipo especial de materiales dieléctricos que son conocidos por lo que se conoce como los efectos piezoeléctrico y piezoeléctrico inverso. El efecto piezoeléctrico provoca que un dieléctrico produzca un potencial eléctrico cuando éste está sujeto a una vibración. En cambio, el efecto piezoeléctrico inverso provoca una vibración en el dieléctrico cuando se le somete a un potencial eléctrico.

La fig. 6 muestra el diagrama de bloques de ambos efectos. Nuestro interés se concentra naturalmente en el efecto piezoeléctrico inverso. En éste, un voltaje se produce en el dieléctrico cuando se somete a un campo eléctrico. Debido al efecto piezoeléctrico, el voltaje genera una fuerza, la cual por una relación mecánica, se convierte en una elongación.

Goldfarb y Celanovic propusieron en [9] un modelo electromecánico bastante certero de ambos efectos. La fig. 7 muestra el modelo para el efecto piezoeléctrico inverso.

Aquí, la parte eléctrica se representa básicamente por un circuito RC. El efecto piezoeléctrico se representa por T_{em} , el cuál es un transductor electromecánico con una tasa de transformación T_{em} . El voltaje u_p se debe al efecto piezoeléctrico. La fuerza F_p se induce del lado eléctrico y mueve mecánicamente el material piezoeléctrico. La elongación resultante del actuador piezoeléctrico se define como y . La relación entre F_p e y está representada por un sistema masa-resorte-amortiguador. Para el taxel, la totalidad de las masas m desplazadas por el transductor es la suma de las masas del vástago, el tambor y el pin plástico de contacto. La tasa de amortiguamiento b es la fricción entre el vástago y el tambor mientras que k es la dureza equivalente.

El conjunto de ecuaciones electromecánicas que representan el comportamiento del actuador piezoeléctrico son las siguientes:

$$\frac{U_p(s)}{U(s)} = \frac{1}{RCs + 1} \dots\dots\dots(1)$$

$$F_p(s) = T_{em} U_p(s) \dots\dots\dots(2)$$

$$\frac{Y(s)}{F_p(s)} = \frac{1}{ms^2 + bs + k} \dots\dots\dots(3)$$

Los valores de los distintos parámetros del taxel fueron obtenidos experimentalmente y se especifican en la tabla 1.

Parámetro	Valor	Unidades
Resistencia R	340	Ω
Capacitancia C	1.7×10^{-9}	F
Tasa de transducción T_{em}	0.0075	N/V
Masa m	140×10^{-6}	kg
Amortiguamiento b	10^{-9}	N-s/m
Dureza k	4.5×10^{-9}	N/m

Tabla 1. Resumen de los parámetros involucrados en el modelado del taxel.

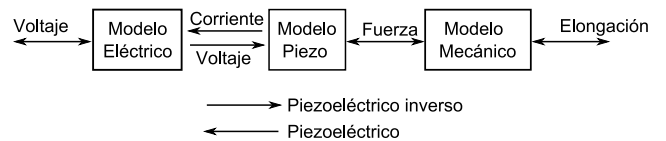


Fig. 6. Diagrama de bloques de los efectos piezoeléctrico y piezoeléctrico inverso.

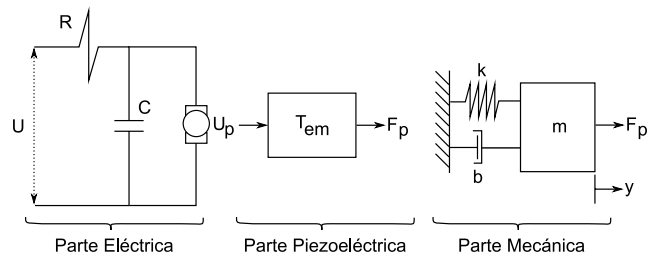


Fig. 7. Modelo electromecánico para el efecto piezoeléctrico inverso.

La fig. 8 muestra las relaciones electromecánicas para el taxel piezoeléctrico. La fig. 8(a) muestra la evolución del voltaje u_p cuando se le aplica una señal diente de sierra de 15 V a la entrada. Note que el comportamiento eléctrico del motor piezoeléctrico es similar al de un capacitor de alta frecuencia.

La fig. 8(b) muestra la relación entre el voltaje u_p y la fuerza transducida F_p . Note que el actuador es capaz de entregar una fuerza de empuje de hasta 200 mN.

La fig. 8(c) muestra la relación entre la fuerza transducida F_p y el desplazamiento y . Para esta gráfica, el transductor es capaz de mover el vástago, el tambor deslizante y el pin plástico de contacto una distancia de $0.18 \mu\text{m}$ en un tiempo de 0.1 ms . Contrariamente a lo que se muestra en la fig. 4(a), el tambor deslizante tiene de hecho un pequeño retroceso cuando el vástago regresa a la posición θ . Sin embargo, esta distancia es despreciable comparada con su desplazamiento positivo hacia delante.

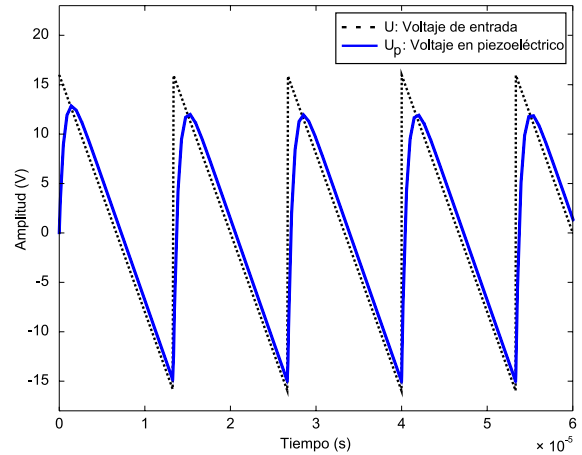
Un análisis de elemento finito se realizó a la estructura del taxel para evaluar si ésta era capaz de soportar la fuerza de los dedos o de la palma de la mano durante el proceso de exploración o lectura.

Típicamente, una fuerza de 50 a 100 mN se aplica por los dedos durante una exploración suave de un objeto [10]. Fuerzas más allá de esa magnitud pueden ser consideradas como conscientes e intencionalmente violentas. Los actuadores convencionales de Braille están diseñados para soportar fuerzas de hasta 170 mN.

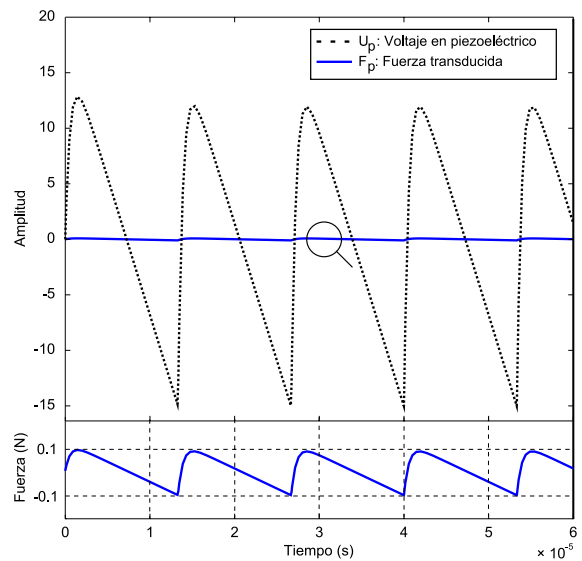
Dos casos se pueden esperar durante la exploración: (a) la fuerza externa F_e es normal a la superficie de contacto. (b) La fuerza externa F_e es tangente a la superficie de contacto.

La fig. 9 muestra el resultado del análisis. La fig. 9(a) muestra que la estructura del taxel que soporta la fuerza F_e cuando ésta es normal a la superficie de contacto es el tambor deslizante. Nótese que el tambor es perfectamente capaz de soportar la fuerza y que no hay puntos críticos que puedan indicar un posible daño de la estructura.

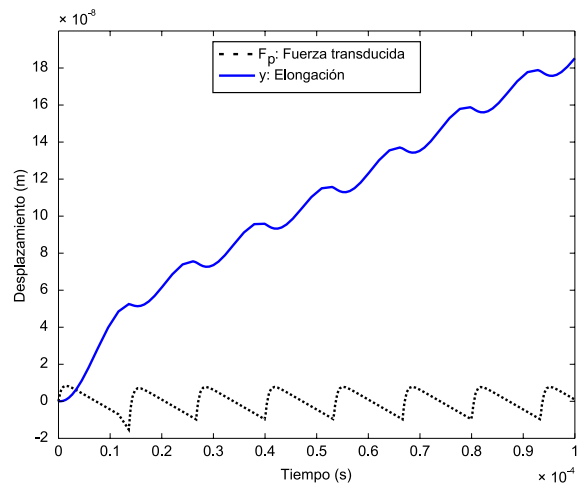
La fig. 9(b) muestra los resultados cuando F_e es tangente a la superficie de contacto. Nótese que el vástago es la estructura del taxel solicitada en este caso. Los resultados de la simulación muestran que el vástago falla al tener que soportar una fuerza tangente de más de 170 mN y



(a)



(b)



(c)

Fig. 8. Relaciones (a) eléctricas, (b) piezoeléctricas y (c) mecánicas para el taxel.

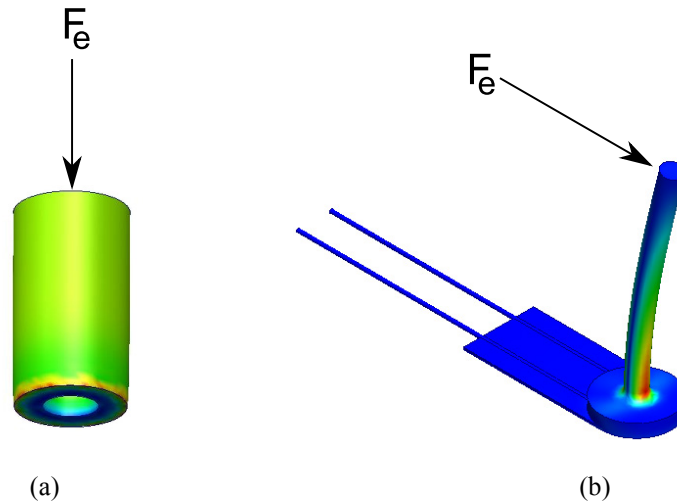


Fig. 9. Análisis de elemento finito con una fuerza de contacto de 170 mN: (a) fuerza normal a la superficie de contacto y (b) fuerza tangente a la superficie de contacto.

que podría ceder doblando así al taxel. Para evitar esta situación, una posible solución es reforzar el taxel con estructuras guía.

III. PANTALLA TÁCTIL

Los actuadores TULA que se caracterizaron en la sección anterior están siendo utilizados actualmente para construir un prototipo de pantalla táctil portátil tipo Braille.

Un dispositivo Braille debe de cumplir con los siguientes requisitos de diseño:

- El espacio entre los pines debe ser de 2.54 a 3 mm (el estándar Braille).
- Los taxels deben de ser controlados independientemente y su movimiento no debe limitar el movimiento de otros taxels vecinos.
- Cada taxel debe estar colocado adecuadamente en la celda Braille y tener acceso a la alimentación sin alterar otros taxels.
- Un taxel debe poder ser removido y remplazado fácilmente en caso de ser necesario.
- El diseño de la pantalla táctil debe permitir tener al final un prototipo robusto, fácil de dar mantenimiento y sencillo/rápido de construir.

El diseño que se utilizó para construir la pantalla táctil está basado en un diseño multi-capas: si se colocaran los actuadores TULA uno junto al otro en una sola capa resultaría en un proceso de ensamble bastante complejo. Utilizando un diseño multi-capas, los taxels se pueden colocar de tal manera que se facilite su manejo/ensamble/remplazo. De igual manera, el diseño multi-capas permite acercarse al espaciamiento Braille deseado (2.54 - 3 mm).

La fig. 10(a) muestra la representación conceptual de una celda Braille utilizando el diseño multi-capas. Este primer diseño contiene 7 capas: la primera es la superficie de contacto, la segunda sirve como estructura guía para los pines de contacto. Las capas 3 a 6 contienen los taxels. Las capas 3 y 4 contienen 2 taxels cada una, mientras que las capas 5 y 6 sólo uno. La capa 7 sirve como base de la celda.

La fig. 10(a) muestra en detalle la estructura de las capas 3 a 6. Note que 6 taxels se encuentran colocados en 4 capas diferentes.

La fig. 10(b) muestra el primer prototipo de celda Braille construida usando el concepto de multi-capas. El espaciamiento entre actuadores es de 4 mm en este primer prototipo. Su desempeño general es muy similar al obtenido con las celdas Braille convencionales: 5 Hz de frecuencia, 200 mN de fuerza de empuje y posiciones

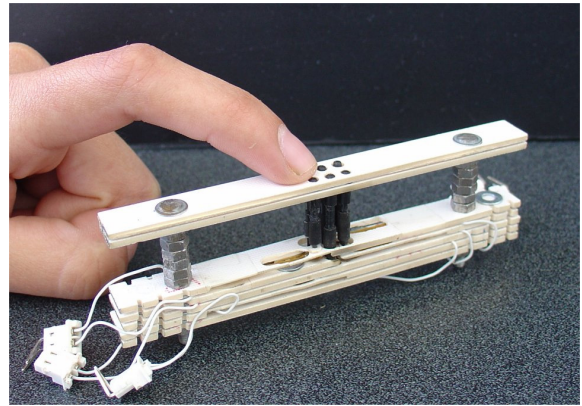
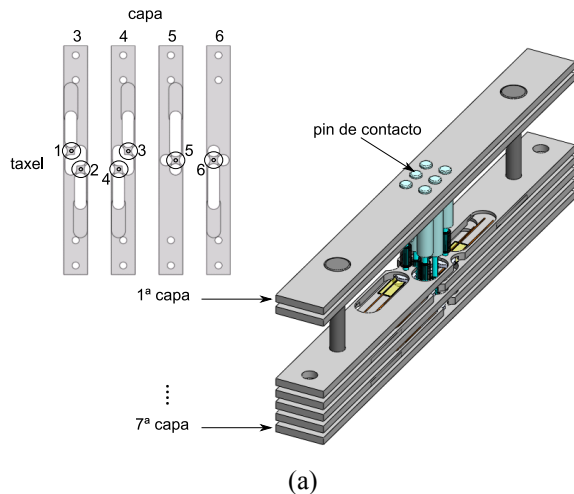


Fig. 10. Celda Braille con un diseño multi-capas: (a) detalle de la estructura de las capas y ensamblaje. (b) Primer prototipo de celda Braille usando actuadores TULA y el concepto multi-capas.

controlables a lo largo de los 11.6 mm de longitud del vástago. La ventaja principal de este prototipo es su tamaño compacto y peso ligero.

Actualmente el desarrollo se enfoca en la manufactura de 10 celdas Braille para desplegar una línea de texto completa.

IV. CONCLUSIÓN

Este artículo presentó el principio de operación y caracterización de un novedoso actuador lineal ultrasónico piezoeléctrico (TULA).

Este actuador presenta avances significativos comparado con los actuadores actuales: tamaño, compactibilidad y estructura simple. Estas ventajas lo hacen interesante para un gran número de aplicaciones, particularmente en las escalas meso/micro.

Este artículo describió el diseño, desempeño e implementación de un actuador táctil (taxel) para pantallas táctiles.

Los resultados de su desempeño son bastante similares a los obtenidos con actuadores Braille convencionales como lo son las vigas piezoeléctricas bicapa y los relevadores. Sin embargo, su compactibilidad permite el desarrollo de pantallas táctiles altamente portátiles para personas con discapacidad visual.

Un prototipo de celda Braille se ha implementado utilizando los actuadores TULA y actualmente se trabaja en la fabricación de 10 celdas más para desplegar una línea de texto.

Paralelamente se está trabajando en integrar una fuente de alimentación compacta y efectiva para poder mantener la portabilidad del sistema.

V. AGRADECIMIENTOS

Este trabajo fue financiado en su totalidad por el CEPii (Centro Panamericano de Investigación e Innovación - México).

VI. REFERENCIAS

- [1] D. Kontarinis and R. Howe, "Tactile display of vibratory information in tele-operation and virtual environments", *Presence*, 4(4), pp 387-402, 1995.
- [2] S. Ino, S. Shimizu, T. Odagawa, M. Sato, M. Takahashi, T. Izumi and T. Ifukube, "A tactile display for presenting quality of materials by changing the temperature of skin surfaces", *Proc. IEEE International Workshop on Robot and Human Communication*, pp 220-224, 1993.

- [3] R. Velazquez, E. Pissaloux, J.C. Guinot and F. Maingreud, “*Walking using touch: design and preliminary prototype of a non-invasive ETA for the visually impaired*”, Proc. of Annual International Conference of the IEEE-EMBS, pp 6821-6824, 2005.
- [4] Y. Bar-Cohen, C. Mavroidis, M. Bouzit, B. Dolgin, D. Harm, G. Kopchok and R. White, “*Virtual reality robotic telesurgery simulations using MEMICA haptic system*”, Proc. of SPIE Annual International Symposium on Smart Structures and Materials, 2001.
- [5] E. Petriu and W McMath, “*Tactile operator interface for semiautonomous robotic applications*”, Proc. of Artificial Intelligence, Robotics and Automation, pp 77-82, 1992.
- [6] HumanWare. Updated info available at: www.humanware.com.
- [7] Piezoelectric Technology Co. Updated information available at: <http://www.piezotech.com>
- [8] H. Ko, S. Kim, S. Borodinas, P. Vasiljev and C. Kang, “*A novel tiny ultrasonic linear motor using the radial mode of a bimorph*”, Sensors and Actuators A: Physical, 125, pp 477-481, 2006.
- [9] M. Goldfarb and N. Celanovic, “*Modeling piezoelectric stack actuators for control of micromanipulation*”, IEEE Control Systems Magazine, 17, pp 69-79, 1997.
- [10] J. Loomis, “*On the tangibility of letters and Braille*”, Perception and Psychophysics, 29(1), pp 37-46, 1981.

DESARROLLO DE UN PROTOTIPO PARA CORTE DE PLACA METALICA

Francisco Javier López Jáquez¹, María Magdalena Hernández Ramos².

Departamento de Ingeniería Industrial y Manufactura, Instituto de Ingeniería y Tecnología
Universidad Autónoma de Ciudad Juárez
Av. Del Charro 450 Nte. Col. San Lorenzo, Ciudad Juárez Chihuahua, México
Tel. 656-6884843, correo electrónico: ¹frlopez@uacj.mx ²magdys_hr@hotmail.com

RESUMEN

Este artículo trata sobre los resultados obtenidos durante la realización de un equipo donde se integran diferentes áreas del conocimiento. En especial se discute la integración del equipo necesario para realizar movimiento coordinado en un plano XY. La tarea de crear una máquina de movimiento en un plano se debe a que se desea tener un equipo para corte de hoja o placa metálica utilizando la tecnología de plasma, una antorcha debe moverse coordinadamente para realizar el corte de trayectorias complejas. La utilización de equipo de prototipos rápidos para la manufactura directa de varios componentes del equipo de corte, hechos a la medida, desde el diseño hasta el prototipo funcional fue de gran ayuda para la realización de la máquina de corte.

Palabras clave: Automatización, Prototipos Rápidos, Control, Trayectoria, Corte.

I. INTRODUCCIÓN

La integración de equipo tanto mecánico como electrónico y de software es lo que ahora se aglutina como equipo mecatrónico. El equipo que se integra básicamente consta de una cortadora de plasma, dos módulos para control de motores a pasos (uno para cada eje), dos motores a pasos, software para realizar secuencias de movimiento y diferentes componentes mecánicos para convertir el movimiento giratorio, de los motores, a lineal. Además de elementos de sujeción y conexión. El control de movimientos se efectúa por medio de la caja de control que se incorporó y que contiene los módulos de control de los motores. El sistema de

control es de ciclo abierto y la comunicación es a través de una red RS-485 [1][2][3].

Un aspecto del proceso es la generación de trayectorias a través de software propio el cual es capaz de convertir una trayectoria, de un dibujo en formato bmp, a una secuencia de comandos de acuerdo a las especificaciones del software de los controladores, el cual se utiliza para cargar la secuencia y ejecutar el movimiento. Este aspecto se ha tratado con mayor detalle en [4][5]. Sin embargo, en este artículo se incluye la validación de la trayectoria mediante el resultado real obtenido al realizar el corte. La validación se realizó mediante los puntos obtenidos conectando decodificadores lineales y una pantalla donde se muestra la posición de la herramienta de corte, antorcha.

La finalidad del proyecto inicial fue crear un prototipo funcional en donde se utilizara tanto equipo de maquinado convencional, CNC y de prototipos rápidos disponibles en las instalaciones propias y poder así validar las secuencias generadas con el software.

El uso de equipo de prototipos rápidos fue un catalizador para obtener en forma tangible varios de los componentes del equipo, en un tiempo corto, que si se manufacturaran en máquinas convencionales, permitiendo decidir respecto al diseño final y su funcionalidad en el equipo. El equipo de prototipos usado durante este proyecto fue una impresora 3D Spectrum Z510 y otra impresora Dimension elite.

Las piezas que se manufacturaron en la Spectrum Z510 fueron remplazadas por las piezas hechas en la Dimension elite, una vez que se comprobó su

funcionalidad ya que el material con el que se realizan en la primera impresora es más frágil que el material utilizado en la segunda impresora, más resistentes para la tarea que se desea.

Los prototipos de los diseños que se han utilizado durante las pruebas han resultado resistentes. Han resistido las pruebas efectuadas en el funcionamiento final de equipo.

De ser necesario, si algún diseño hechos en los prototipos rápidos no resistiera su función se reemplazará por uno manufacturado en la máquinas convencionales para darle mayor resistencia. Aunque, por el momento no se ha dado el caso de reemplazar algún diseño ya que estos han soportado el esfuerzo a que se han sometido.

Se realizaron pruebas de corte, utilizando la antorcha de plasma, los componentes diseñados y procesados en los prototipos rápidos han resultado de gran utilidad ya que han permitido efectuar las pruebas sin necesidad de reemplazarlos. El material utilizado es resistente a los movimientos efectuados por el equipo, los cuales son: movimiento en un plano X, Y.

II. MÉTODO

Manufactura de componentes

Una parte importante del equipo es que los ejes presenten un error mínimo o nulo al reproducir la trayectoria que se desea. Esto requiere que los componentes del equipo de corte tales como ejes X e Y tengan una sujeción adecuada para mantener los ejes perpendiculares, entre si, y paralelos a la superficie de corte. Además de que tengan buena apariencia.

Se utilizó el software de AutoCad, SolidWord y MasterCam durante la creación de los diseños y manufactura de los mismos. Los materiales de construcción fueron Delrin y plástico ABS. De igual manera se llevó a cabo la construcción de la secuencia de ensamble para evitar problemas al momento de ensamblar el equipo.

Con la utilización de los software antes mencionados se lleva a cabo la realización de los diseños ya que estos programas permiten crear figuras en 2D para posteriormente crearlos en una

visión tridimensional 3D de igual manera MasterCam permite crear el código alfanumérico que es utilizado en la fresadora CNC donde se manufacturó una de estas piezas (Figura 1-b).

La Fig. 1, muestra las piezas de sujeción finales de los diseños realizados para el equipo. Las cuales fueron manufacturadas usando máquinas convencionales y fresadora CNC. Con apoyo del código que se generó en Mastercam. Cabe mencionar que el código generado en este software fue modificado manualmente, esto con la finalidad de que lo reconociera el procesador de la fresadora. Algunos de los cambios que se realizaron fueron: el cambio de herramientas por las que tiene la fresadora, la compensación de las mismas, la eliminación de la letra A, parámetro que se genera al crear el código para la CNC.

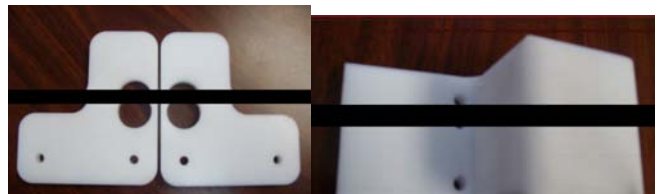


Figura 1- Sujetadores a) eje Y y b) eje X.

El herramental utilizado durante la manufactura Fig. 1-a se usaron brocas y cortadores planos, el maquinado fue simulado en la computadora previamente utilizando el programa Mastercam, también en la máquina de fresado es posible realizar la simulación para asegurar el maquinado final. Las operaciones fueron orificios, cavidades y contornos.

La manufactura de la Fig. 1-b se realiza en la fresadora convencional, para lo cual la máquina debe estar bien calibrada con el propósito de que la pieza quede a las medidas requerida de lo contrario no quedarían a las especificaciones necesarias. Las operaciones en esta pieza fueron: escariado, escuadre, orificios y contorno. En este caso el orificio donde va el balero para deslizarse sobre el eje X fue necesario utilizar una broca y luego un cortador para ensanchar el diámetro del orificio a la medida final ya que no se contaba con un broca de la dimensión requerida.

El material utilizado para estos componentes es el Delrin, este material tiene alta resistencia mecánica, rigidez, dureza, buena estabilidad dimensional, fácil de maquinarse y sobre todo mantiene sus propiedades en diversos ambientes de operación.

Para el uso de prototipos rápidos sólo es necesario utilizar un programa de CAD que permita crear los archivos del diseño con una extensión STL para poder imprimir la pieza. El archivo STL se transporta al programa correspondiente de la impresora, que en este caso es el Zprint 7.3 para la impresora de Spectrum Z510, que imprime capas de 0.1mm de grosor y el área de trabajo de esta impresora es de (254x356x203 mm). De esta forma permite realizar cualquier impresión que este en 3D, de forma rápida.



Figura 2- Sujetador tornillo de fuerza.

En la Fig. 2 se muestra el sujetador que conecta el tornillo de fuerza del eje X a la base del eje Y. Además tiene la opción para conectarse al decodificador de posición. El diseño se realizó con el software Rhinoceros y se guardó con la extensión STL para poder ser usado en el programa de Zprint, el software de la máquina de prototipos.

Una vez que se verificó la funcionalidad de esta pieza. Se llevó a cabo la manufactura de la misma en la impresora Dimension elite, esto con la finalidad de hacer la pieza más resistente a las fuerzas ejercidas al operar el equipo.

Cabe mencionar que esta última impresora trabaja con el programa de CatalystEX, que tiene características similares al programa Zprint, en

cuanto al proceso de impresión. Tiene un área de trabajo de (254x254x306 mm).

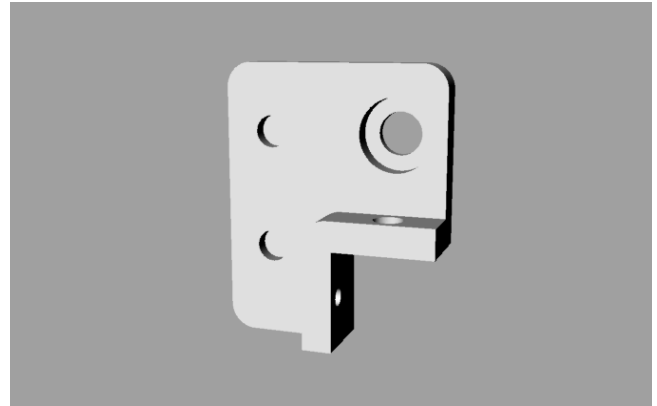


Figura 3- Sujetador de guías y tornillo eje Y.

La Fig. 3 muestra otra pieza, de igual manera que la Figura 2, fue primero impreso en la Spectrum y una vez que se verificó su funcionalidad se imprimió en la Dimension elite. Imprimiéndola dos veces para sujetar de ambos extremos las guías y el tornillo sin fin tipo ACME. Las piezas son espejo y tienen una pestaña con orificios donde se sujeta al perfil, base del eje.

Los diseños que se ilustran en las Fig. 4 y 5 se imprimieron directamente en la máquina de prototipos rápidos Dimension elite.

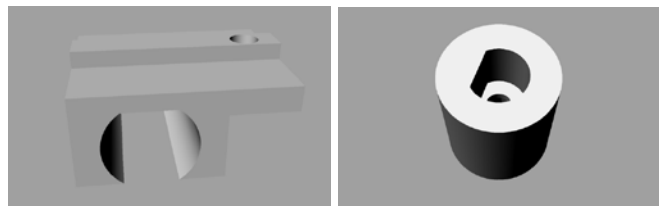


Figura 4- Diseños guía menor y cople motor y tornillo de fuerza.

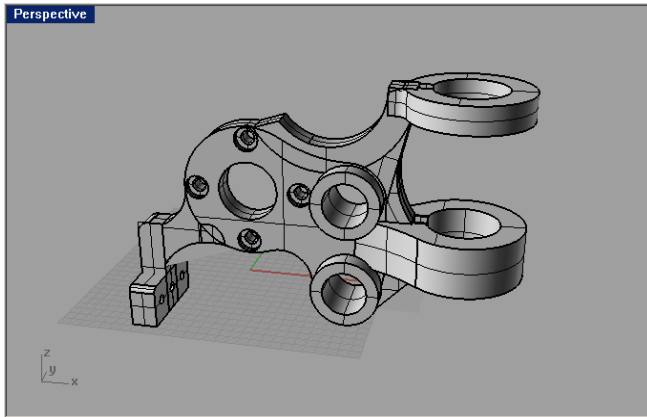


Figura 5- Sujetador de antorcha.

La pieza de las Figuras 5, fue diseñada para sujetar la antorcha así como la guía sin fin y de igual manera el eje de coordenadas.

El eje Y quedo en la parte superior de la mesa y se desplaza sobre el eje X por medio de dos baleros, en el otro extremo sólo tiene un balero que se desplaza sobre un guía. Los dos baleros tienen el propósito de dar firmeza y mantener la perpendicularidad entre los ejes X-Y en toda la extensión, al momento de que el eje Y se desplaza sobre el eje X.

Se utilizó la máquina de prototipos rápidos con la finalidad de manufacturar algunos diseños de los componentes para poder verificar el ensamble de los mismos, así como su funcionamiento. Ya que las piezas se realizan con más rapidez que si se manufacturaran en un torno o fresadora convencional. Lógicamente, después de ver la funcionalidad de las piezas se procederá a la manufactura en las máquinas convencionales, si es necesario. El uso de equipo de prototipos rápidos permite producir un diseño en unas cuantas horas, por muy complejo que éste sea.

Secuencia de ensamble

Ensamble del eje Y. dos guías, dos barras que son ensambladas en los perfiles con tornillos. Cuatro perfiles. Un perfil de 12cm de largo que se coloca en la parte superior usando tornillos y un perfil de 94cm, un balero de 12.7mm que va en el extremo del eje. Tanto las guías como los baleros entran a presión.

Ensamble del eje X. dos sujetadores para eje, guía de $\varnothing 25.4\text{mm}$ y 980mm de largo, dos baleros de $\varnothing 25.4\text{mm}$. Todos los componentes entran a presión.

Ensamble de la vista trasera de la mesa, dos perfiles de 89cm y tres perfiles de 98cm, para este ensamble se utilizan tornillos. El ensamble de las guías.

La vista lateral y la vista derecha contienen los mismos componentes. Consta de tres perfiles de 94cm y un perfil de 98cm. En este caso sólo se ensamblando dos perfiles de 98cm porque los perfiles de las esquinas son parte del ensamble anterior.

Uno de los motores va ensamblado en la parte de la vista derecha por medio de tornillos. Mediante una placa se ensambla al perfil y, del otro extremo, al el motor. El eje del motor se acopla al tornillo sin fin.

Generación de trayectorias

Las siguientes tablas se muestran los valores de las lecturas del diseño, en la Tabla 1a corresponden al equipo de corte y en la Tabla 1b las lecturas correspondientes a el programa, con un total de datos de 194.

X	Y
140.1	79.8
144.3	83.9
144.2	89.8
144.2	95.4
148.8	96.9
153.3	100.8
153.3	106.6
153.9	112.3
155.8	117.9
161.3	116.6
.	.
.	.
.	.
140.4	80.9

X	Y
143.7	80.4
147.9	84.6
146.7	90.3
146.4	96
152.1	97.5
156.6	101.4
156	107.1
157.2	112.8
159	118.5
164.7	116.7
.	.
.	.
.	.
143.7	80.4

Tabla 1- a) Corrida equipo de corte, b) Corrida programa.

En la Fig. 6 se muestra el diseño de prueba, la línea solida corresponde a la lectura del software y las

líneas punteadas correspondientes a las lecturas tomadas del lector acoplado al equipo de corte. Se pueden observar variación que existe entre ambos datos. Sin embargo, el resultado produce una aproximación buena para el caso de piezas donde no se requiere alta precisión, como pueden ser figuras artísticas o logotipos. Aunque lo ideal sería tener una pieza donde no se requiera un proceso extra, por el momento esta variación puede ser eliminada con un proceso extra, si así se desea.

La secuencia de comandos para los controladores se genera al simular la trayectoria en la computadora. El primer movimiento, asumiendo que la herramienta de corte está en la posición 0,0 se calcula en base al primer punto de la trayectoria, se obtienen las componentes XY. Se convierten las distancias de las componentes a número de pasos que se requieren en ambas componentes. Los módulos son capaces de generar 9600 pasos por cada 25.4 mm. Esta información es la base para calcular los pasos que se requieren en ambos ejes. En el eje que requiere más pasos se elige como movimiento principal y si sobrepasa más de 800 pasos se asocia a una velocidad máxima (V_{max}) de 1875 pasos/s con una aceleración de 1250 pasos/s² así que requerirá $t=1.5Pasos/V_{max}$. La velocidad y aceleración para el otro eje se deducen usando el mismo tiempo t .



Figura 6- Diseño aguilá.

El movimiento se realiza con el comando PPM, movimiento trapezoidal por posición que lleva como parámetros los pasos a recorrer, la velocidad máxima y la aceleración. El sentido de movimiento

se indica agregando un signo negativo a los pasos para indicar dirección contraria a la positiva.

La Fig. 7, muestra el interior de la caja de control, se indican los componentes: a) fuente de poder, b) módulos de control ejes X e Y, c) bloque de contactos, d) conectores para red rs485, e) cables para conexión de motores, f) interruptor de seguridad y cables para la interconexión de los componentes.

La Fig. 8 es una vista de la máquina de corte donde se indican los componentes principales del sistema: a) caja de control b) motores c) mecanismos para convertir movimiento giratorio a lineal d) Pistola de corte e) oruga para cableado f) porta herramienta g) estructura h) baleros y sujetadores.

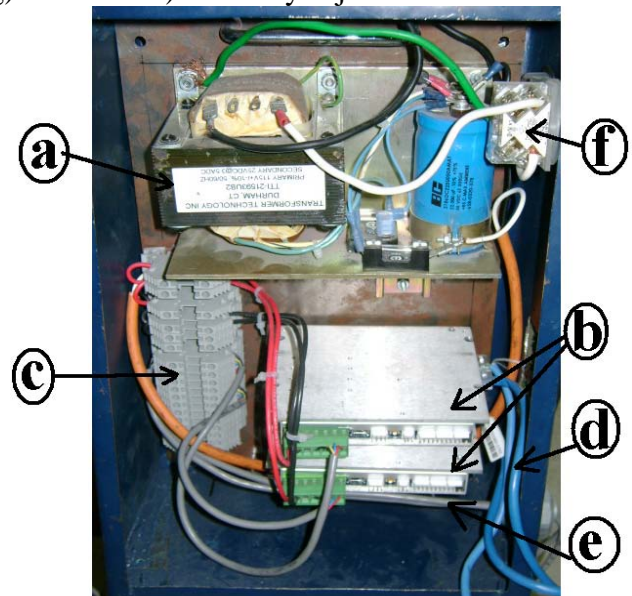


Figura 7- Caja de control de motores.

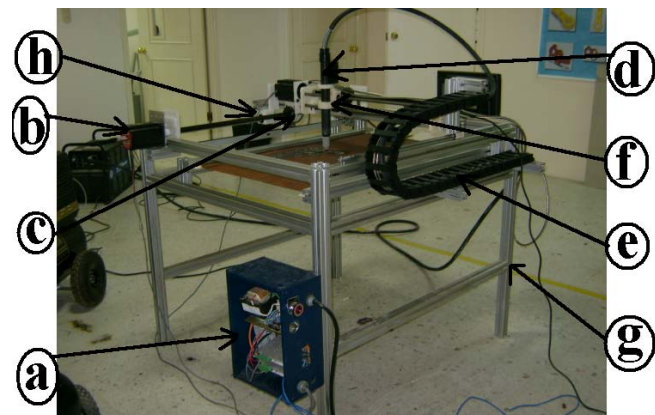


Figura 8- Máquina XY.

El software provee una opción para indicar la distancia mínima permitida, selección del archivo que contiene la trayectoria, opción para rapidez de simulación y opción para iniciar la simulación de la trayectoria. En esta última opción es cuando se generan los comandos para la secuencia de movimiento. En el caso de la opción de distancia mínima permitida, ésta debe especificarse antes de cargar el archivo ya que al cargar el archivo es cuando se genera la trayectoria de corte en coordenadas XY, el valor por omisión es de 30 que significa que la distancia mínima permitida para los segmentos es de 30 mm. La figura muestra un ejemplo de una trayectoria donde la distancia mínima permitida es de 10, el valor mínimo es de 1 mm que es el valor que produce la máxima resolución en la trayectoria.

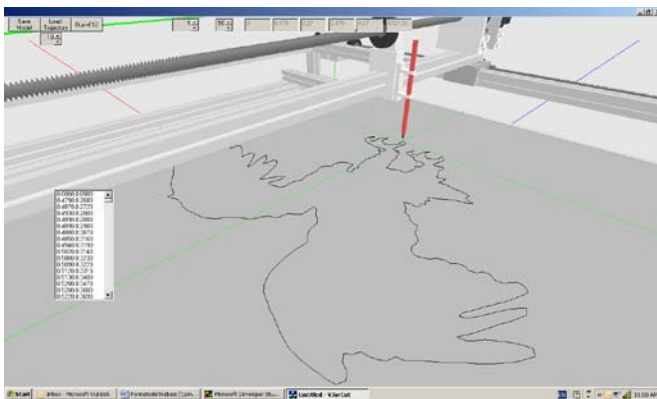


Figura 9- Simulador de trayectoria de corte.

En la Fig. 10 se puede observar la interfase que provee el fabricante de los controladores que permite la ejecución de una secuencia de movimientos. La secuencia se carga y se activa la red de control. Al presionar el botón de inicio los comandos que se van ejecutando aparecen en el recuadro derecho donde se puede observar que la secuencias se componen de un inicio a cero en ambos ejes (A1 y A2) con el comando POS, seguido de un comando de movimiento por posición (PPM) donde se especifica la cantidad de pasos, la velocidad máxima y la aceleración.

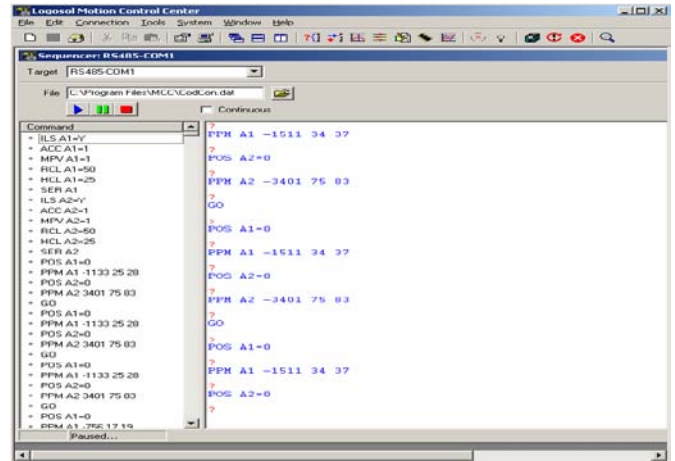


Figura 10- Interface para correr secuencias.

En la Fig. 11 se ilustra el resultado de un corte utilizando la máquina XY, En este caso la trayectoria se dibujó en formato bmp, se convirtió a comandos usando el software de simulación de trayectorias y después de unas dos horas se obtuvo el corte, al cual fue pulido y se le agrego un base de cartón con fondo de color.

En realidad es posible crear dos piezas a la vez ya que el metal que se corta se puede usar para crear la figura complementaria, en este caso se trata de un logotipo.



Figura 11- Logotipo del SPAIA.

III. CONCLUSIONES

Se ha presentado la integración de un equipo para corte de trayectorias en un plano. Que es capaz de convertir directamente de un dibujo en formato

electrónico a una secuencia de comandos para los módulos de control independientes y realizar el movimiento coordinado.

Una restricción al momento de calcular la velocidad máxima y aceleración para la secuencia de movimiento es que deben redondearse a múltiplos de 25 por lo que ésta es una fuente de error al calcular los parámetros del otro eje. Habrá que investigar tanto en la estrategia como en el comportamiento del error para minimizarlo o compensarlo. Aunque los resultados muestran diferencias, estas son del orden de los milímetros que dan una buena aproximación a la pieza final, aunque en caso de mayor precisión se requiere un proceso de acabado, dejando de antemano una distancia para trabajar en el acabado.

La experiencia en este proyecto ha permitido fincar la base para proyectos futuros, hay un gran campo para mejorar desde el software hasta sistema tanto mecánico como eléctrico/electrónico. Así como integrar otros componentes como servomotores, encoders, sistemas de compensación para auto corrección de trayectoria, entre otras cosas.

IV. REFERENCIAS

- [1] M. Minhat, V. Vyatkin, X. Xu, S. Wong, Z. Al-Bayaa, "A novel open CNC architecture based on STEP-NC data model and IEC 61499 function blocks", Robotics and Computer-Integrated Manufacturing. No. 25, pp. 560-569, 2009.
- [2] A. S. Ghafari, M. Bezhad, "Investigation of a micro-step control positioning system performance affected by random input signals", MECHATRONICS. Vol. 15, pp. 1175-1189, 2005.
- [3] J. R. Rogers, K. Craig, "On-hardware optimization of stepper-motor system dynamics", Mechatronics, No. 15, pp. 291-316, 2005
- [4] F. J. Lopez-Jaquez, S. A. Noriega-Morales, "Generating 3D Digital Objects from 2D

Bitmap Files", ICNAAM- 2005. pp. 628-630

- [5] F. J. Lopez-Jaquez, Bernard Hon, "Computer modeling of an XY Table", Advances in Manufacturing Technology-XX, pp. 313-318, 2006

V. AUTORES

Dr. Francisco Javier López Jáquez obtuvo su título de Doctorado en filosofía en el área de Manufactura por la Universidad de Liverpool en el Reino Unido, Inglaterra. Actualmente está adscrito como profesor investigador al departamento de Ingeniería Industrial y Manufactura de la Universidad Autónoma de Ciudad Juárez, es líder del cuerpo académico de Diseño del Producto y Materiales para Manufactura.

Ing. María Magdalena Hernandez Ramos es Ingeniero en Manufactura por la Universidad Autónoma de Ciudad Juárez, Chihuahua, México. Actualmente cursa estudios de posgrado en el área de Manufactura.

ANÁLISIS FRACCIONAL PARA LA REDUCCIÓN DE ORDEN DEL MODELO MATEMÁTICO DE UN AUTOBÚS

Vázquez Álvarez Iván, Ocampo Hidalgo Juan Jesús, Ferreyra Ramírez Andrés, Avilés Cruz Carlos.

Área de Instrumentación, Departamento de Electrónica, División de CBI
Universidad Autónoma Metropolitana--Azcapotzalco
Av. San Pablo No. 180, Col. Reynosa Tamaulipas, México, D.F. 02200, México
Tel. 53189550, correo electrónico: {iva,jjoh,fra,caviles}@correo.azc.uam.mx

RESUMEN

El presente trabajo muestra el modelo completo de movimiento longitudinal de un vehículo automotor, se parte de los sistemas de referencia para cada uno de los subsistemas, se considera que se está trabajando en una zona cercana a la fuerza de contacto máxima, lo cual físicamente corresponde a una condición cercana al bloqueo de una rueda.

Posteriormente se obtienen los tiempos característicos para cada uno de los subsistemas principales, se evalúan para valores de un autobús de transporte urbano. Finalmente se toma como referencia el tiempo T_3 , con lo cual se reduce el modelo a un sistema considerablemente menor que el sistema original, y que resulta válido para las condiciones que se consideran.

Palabras clave: Modelado matemático, análisis fraccional, automóvil, movimiento longitudinal, contantes de tiempo.

I. INTRODUCCIÓN

Actualmente las empresas automotrices se orientan a la modernización de los automóviles; dicha modernización incluye no solo la potencia en el motor o mejoras en el diseño interior y/o exterior del vehículo, también es muy importante mejorar las propiedades de los materiales que se utilizan para garantizar la seguridad de sus ocupantes.

Hay dos tipos de seguridad en los vehículos, la seguridad pasiva y la activa, la primera está encaminada a minimizar las consecuencias sobre el pasajero en caso de un accidente (Cinturones de Seguridad, AirBag, interiores ergonómicos, etc) y

la segunda engloba los dispositivos sobre los que el conductor puede influir directamente (Sistema de Frenos ABS, Suspensión, Dirección, etc.).

En el caso de los sistemas de seguridad activos, se han desarrollado ideas sobre los sistemas de frenos ABS, los cuales existen en una gran diversidad y son divididos de acuerdo a su funcionamiento: de tambor, de disco, neumáticos, hidráulicos, con sistema antibloqueo (ABS) e incluso sistemas de freno inteligentes.

En el presente trabajo se busca obtener un modelo matemático de un vehículo automotor, que permita trabajar con el control del sistema de frenos ABS, por eso se trabaja en la zona de máxima fuerza de contacto, situación en la cual resulta efectivo el uso de los frenos ABS.

El análisis fraccional [8] nos permite analizar sistemas de acuerdo a parámetros específicos que nos permiten separar variables de acuerdo a los intervalos dentro de los cuales suceden, en este caso particular las variables se dividen de acuerdo a los intervalos de tiempo en que ocurren, con lo cual se puede hablar de variables rápidas o lentas, tal como sucede al trabajar con perturbaciones singulares.

II. MODELO COMPLETO DEL SISTEMA

II-A Sistemas de coordenadas y traslaciones

Suponemos que el movimiento del automóvil ocurre sobre una superficie horizontal, los diferentes sistemas de coordenadas se muestran en fig. 1 y fig. 2.

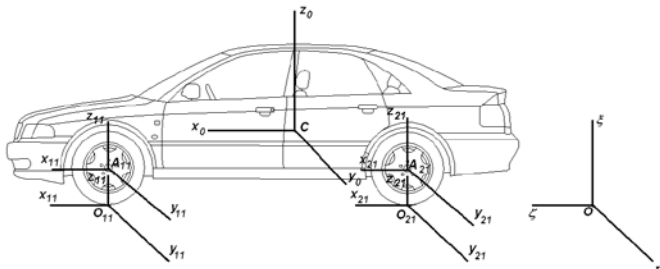


Fig. 1 Vista lateral de los ejes

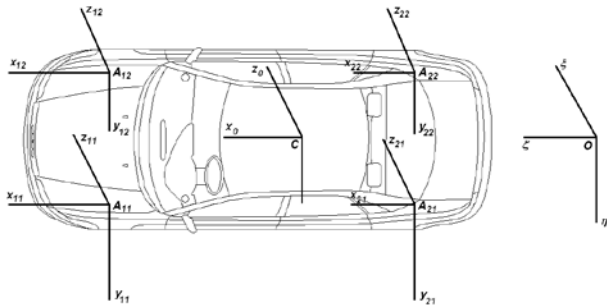


Fig. 2 Vista superior de los ejes

Introducimos los siguientes sistemas de coordenadas: $O\xi\eta\zeta$ - Fijo. El plano $O\xi\eta$ coincide con la superficie de movimiento. El eje $O\xi$ es vertical, $O\eta$ - va en la dirección horizontal.

La posición del centro de masa C del automóvil se da en coordenadas cartesianas X,Y,Z en el sistema $O\xi\eta\zeta$.

El sistema $Cx_0y_0z_0$ - con origen en el centro de masa, el eje Cz_0 es vertical, el eje Cx_0 se encuentra en el plano longitudinal de simetría del automóvil. El sistema $Cx_0y_0z_0$ puede obtenerse de $O\xi\eta\zeta$ como resultado de una traslación en el punto C y una rotación del mismo en ψ sobre el eje ζ . El ángulo ψ es el ángulo de curso del automóvil.

El sistema $Cxyz$ - relacionado con el cuerpo del automóvil coincide con el eje de inercia central. Un cambio del sistema $Cx_0y_0z_0$ a $Cxyz$ está dado por los ángulos de rotación, γ sobre el eje x y θ el ángulo de ataque sobre el eje y .

Los sistemas $A_{ij}x_{ij}y_{ij}z_{ij}$ - relacionados con la rueda. El índice "i" toma los valores $i=1$ delantero, $i=2$ trasero, en tanto que el índice "j" tiene los valores $j=1$ derecha y $j=2$ izquierda, de acuerdo a la dirección del movimiento. El punto A_{ij} se encuentra en la intersección del eje de giro de la ij -ésima rueda y el plano de simetría longitudinal. El eje $A_{ij}y_{ij}$ coincide con el eje de rotación de la rueda, el eje $A_{ij}x_{ij}$ es horizontal.

El sistema $O_{ij}x_{ij}y_{ij}z_{ij}$ con origen en el punto O_{ij} .

II-B Ecuaciones dinámicas de movimiento del vehículo completo.

Proponemos que el movimiento ocurre solo en el plano x_0y_0 . Las ecuaciones de movimiento del centro de masa del automóvil en el sistema tridimensional $Cx_0y_0z_0$ se escriben de la siguiente forma [4]:

$$M \frac{dV_x}{dT} = \sum_{i,j=1}^2 P_{ijx} + MV_y \Omega_z + F_{ax} \quad (1)$$

$$M \frac{dV_y}{dT} = \sum_{i,j=1}^2 P_{ijy} - MV_x \Omega_z + F_{ay} \quad (2)$$

$$M \frac{dV_z}{dT} = \sum_{i,j=1}^2 N_{ij} - Mg \quad (3)$$

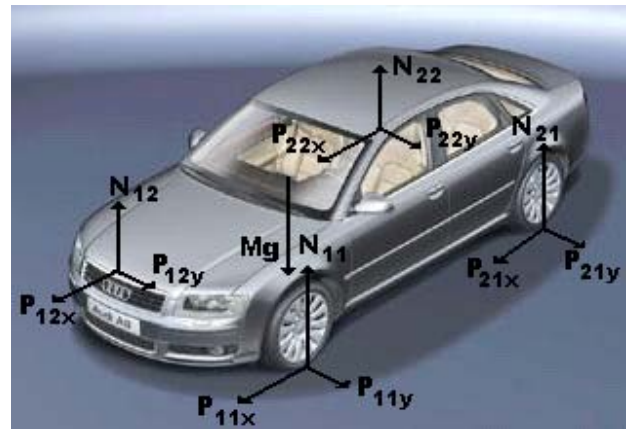


Fig. 3 Fuerzas presentes

Aquí T - tiempo, M - masa total del automóvil (incluyendo las ruedas), - (V_x, V_y, V_z) proyecciones sobre el sistema $Cx_0y_0z_0$ del vector de velocidad absoluta del punto C , Ω_z - proyección sobre el eje vertical de la velocidad angular absoluta, Mg - peso, $(P_{ijx}, P_{ijy}, N_{ij})$ - proyecciones de la fuerza de contacto y de la reacción normal. $(MV_y \Omega_z, -MV_x$

$\Omega_z, 0)$ - proyecciones de la fuerza de inercia, (F_{ax}, F_{ay}) - proyecciones de la fuerza aerodinámica, las fuerzas que actúan sobre el sistema se muestran en fig. 3.

Las ecuaciones de cambio en el momento cinético del automóvil completo relativas al punto C en proyección tridimensional $Cx_0y_0z_0$, considerando ausencia de saltos a causa de irregularidades en el camino tienen la forma:

$$I_x \frac{d\Omega_x}{dT} = [(N_{11} + N_{21}) - (N_{12} + N_{22})B] - Z \sum_{i,j=1}^2 P_{ijy} - \sum_{j=1}^2 I_{1j} [\dot{\Omega}_{1j} \sin \Theta + \Omega_{1j} \dot{\Theta} \cos \Theta - \Omega_{1j} \Omega_z \cos \Theta] \quad (4)$$

$$I_y \frac{d\Omega_y}{dT} = (N_{21} + N_{22})A_2 - (N_{11} + N_{12})A_1 - Z \sum_{i,j=1}^2 P_{ijx} - \sum_{j=1}^2 I_{1j} [\dot{\Omega}_{1j} \cos \Theta + \Omega_{1j} \dot{\Theta} \sin \Theta - \Omega_{1j} \Omega_z \sin \Theta] \quad (5)$$

$$I_z \frac{d\Omega_z}{dT} = [(P_{12x} + P_{22x}) - (P_{11x} + P_{21x})]B + (P_{11y} + P_{12y})A_1 - (P_{21y} + P_{22y})A_2 \quad (6)$$

donde I designa el momento de inercia tanto del vehículo como de cada una de las ruedas de acuerdo a los subíndices [3], [5], [11].

II-C Ecuaciones cinemáticas del vehículo

Escribimos las ecuaciones cinemáticas del sistema en la forma:

$$\frac{dX}{dT} = V_x \cos \psi - V_y \sin \psi \quad (7)$$

$$\frac{dY}{dT} = V_x \sin \psi - V_y \cos \psi \quad (8)$$

$$\frac{dZ}{dT} = V_z \quad (9)$$

$$\frac{d\psi}{dT} = \Omega_z \quad (10)$$

Y las ecuaciones para la velocidad angular en una aproximación lineal son:

$$\frac{d\gamma}{dT} = \Omega_x \quad (11)$$

$$\frac{d\vartheta}{dT} = \Omega_y \quad (12)$$

II-D Relaciones para la fuerza de contacto entre la rueda y el camino

Se tienen las siguientes expresiones:

$$P_{11x_0} = P_{11x} \cos \Theta_{11} - P_{11y} \sin \Theta_{11} \quad (13)$$

$$P_{11y_0} = P_{11x} \sin \Theta_{11} - P_{11y} \cos \Theta_{11} \quad (14)$$

$$P_{12x_0} = P_{12x} \cos \Theta_{12} - P_{12y} \sin \Theta_{12} \quad (15)$$

$$P_{12y_0} = P_{12x} \sin \Theta_{12} - P_{12y} \cos \Theta_{12} \quad (16)$$

$$P_{21x_0} = P_{21x}, \quad P_{21y_0} = P_{21y} \quad (17)$$

$$P_{22x_0} = P_{22x}, \quad P_{22y_0} = P_{22y}$$

La fig. 4 muestra las fuerzas que aparecen en las expresiones anteriores [13], [6]

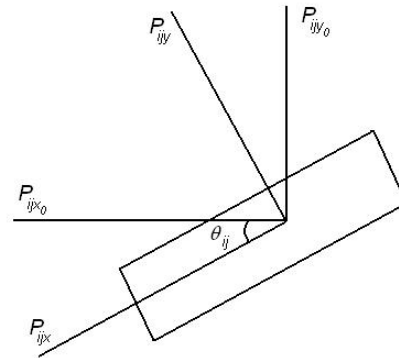


Fig. 4 Relaciones de la fuerza de contacto

III-E Modelo de movimiento del elemento de contacto

Las ecuaciones de movimiento se escriben como proyecciones en $O_{ij}x_{ij}y_{ij}z_{ij}$:

$$M_c \frac{d}{dT} (V_x + \Omega_{ij} R + \xi_{ij}) = -C_x \dot{\xi}_{ij} - K_x \xi_{ij} + P_{ijx} \quad (18)$$

$$M_c \frac{d}{dT} (V_y + \eta_{ij}) = -C_y \dot{\eta}_{ij} - K_y \eta_{ij} + P_{ijy} \quad (19)$$

$$0 = -C_z \dot{\zeta}_{ij} - K_z \zeta_{ij} + N_{ij} \quad (20)$$

Donde P_{ijx}, P_{ijy}, N_{ij} - proyecciones de la fuerza de contacto correspondientes al elemento de contacto, $-K_x \xi_{ij}, -K_y \eta_{ij}, -K_z \zeta_{ij}$ - proyecciones del vector de fuerzas elásticas correspondientes al elemento de contacto, considerando la parte no deformable del

neumático, K_x, K_y, K_z - coeficientes de elasticidad, $-C_x \xi'_{ij}, -C_y \eta'_{ij}, -C_z \zeta'_{ij}$ - proyecciones del vector de fuerzas de amortiguamiento, C_x, C_y, C_z - coeficientes de amortiguamiento. En esta parte se supone que no existen traslaciones verticales del elemento de contacto, la fig. 5 muestra el elemento de contacto [9].

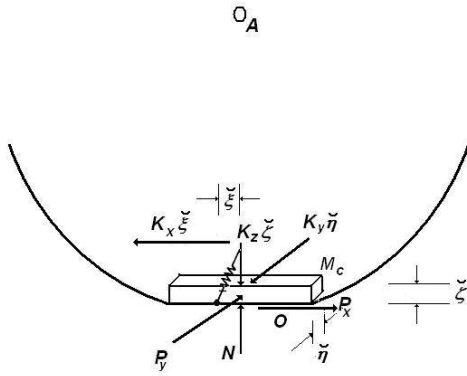


Fig. 5 Elemento de contacto

III-F Ecuación de rotación de la rueda

La ecuación de cambio de momento en la cantidad de movimiento describe la rotación de la rueda:

$$I_j \frac{d\Omega_{ijy}}{dT} = -P_{ijx}(R - \xi_{ij}) + L_{ij} \quad (21)$$

Aquí I_j - momento de inercia de la rueda, L_{ij} - momento de frenado.

El sistema de rotación de la rueda se muestra en la fig. 6.

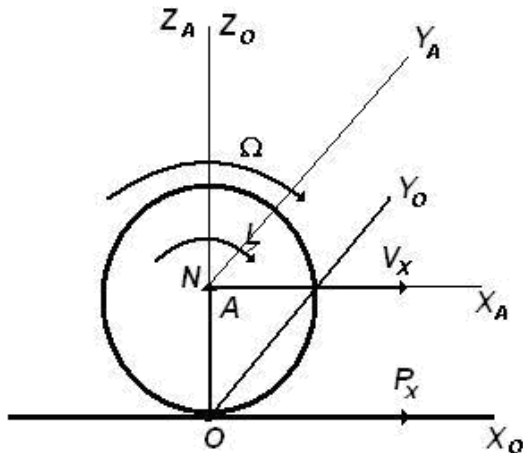


Fig. 6 Rotación de la rueda

III-G Modelo de la fuerza de contacto

Partimos del modelo obtenido en [7]. El deslizamiento relativo es representado por s

$$s_{xij} = \frac{V_{Oxij}}{V_{Axij}} \quad (22)$$

$$s_{yij} = \frac{V_{Oyij}}{V_{Axij}} \quad (23)$$

Donde V_{Oxij} - velocidad del centro de la sombra de contacto, V_{Axij} - Velocidad longitudinal del centro de masa de la rueda, así tenemos las expresiones siguientes:

$$V_{Oxij} = V_{Axij} - \Omega_{ij} R + \xi_{ij} \quad (24)$$

$$V_{Oyij} = V_{Ayij} + \eta_{ij} \quad (25)$$

Sustituyendo en las ecuaciones anteriores tenemos:

$$s_{xij} = \frac{V_{Axij} - \Omega_{ij} R + \xi_{ij}}{V_{Axij}} \quad (26)$$

$$s_{yij} = \frac{V_{Ayij} + \eta_{ij}}{V_{Ayij}} \quad (27)$$

Para la fuerza de contacto tenemos:

$$P_{xij} = -vN \frac{s_{xij}}{s_{ij}} \varphi(s_{ij}) \quad (28)$$

$$P_{yij} = -vN \frac{s_{yij}}{s_{ij}} \varphi(s_{ij}) \quad (29)$$

Donde $s_{ij} = \sqrt{s_{xij}^2 + s_{yij}^2}$

La función $\varphi(s)$ se define experimentalmente y será descrita más adelante.

Cuando el automóvil se mueve solo a lo largo del eje x , tenemos $s_{yij} = 0$ y $P_{yij} = 0$, entonces la ecuación anterior queda:

$$P_{xij} = -vN \varphi(s_{ij}) \text{sign}(s_{xij}) \quad (30)$$

III-H Modelo de vibraciones verticales de la rueda

Estudiamos el sistema mecánico compuesto por los elementos de la suspensión, del cuerpo del vehículo al elemento de apoyo. Escribimos las ecuaciones de movimiento del centro de masa de dicho sistema [1], [2].

$$m_A \frac{d^2 z_{Aij}}{dT^2} = N_{ij} + C_{zj} \Delta z_{ij} + D_j \frac{d\Delta z_{ij}}{dT} \quad (31)$$

Aquí m_A - masa equivalente del sistema,

$z_{Aij} = R \sqrt{\zeta_{ij} + \text{const}}$ - coordenadas del centro de masa, Δz_{ij} - deformación de los resortes, D_j , C_{jz} - coeficientes equivalentes de amortiguamiento y de dureza del resorte, en general diferentes para las ruedas delanteras y traseras.

III-I Modelo del sistema de frenos

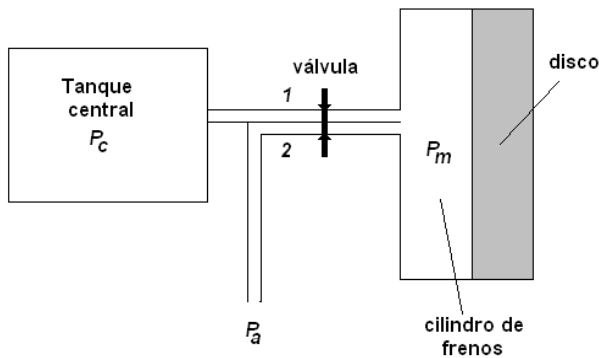


Fig. 7 Modelo del sistema de frenos

El sistema de frenos neumático mostrado en la figura anterior incluye dos tipos de tanques de reserva: el tanque central y el cilindro de frenos, de cada uno de ellos sale un conducto; del tanque central al cilindro de frenos (1) y del cilindro de frenos a la atmósfera (2). Analizamos el sistema en el cual se tiene instalada una válvula que abre y cierra ambos conductos, cuando 1 se encuentra abierto 2 está cerrado (llenado) y viceversa (expulsión), fig. 7 [12].

Suponemos que el momento de la fuerza de frenado es proporcional a la presión en el cilindro de frenado y la dirección es en contra de la velocidad angular de la rueda.

$$L_{ij} = -K_L P_{mij} \text{sign} \Omega_{ij}$$

Utilizamos una aproximación de primer orden para el modelo de cambio de presión en el cilindro de frenado.

$$\begin{cases} T_{e_{in}} \frac{dP_{mij}}{dT} + P_{mij} = P_c (T + \Delta t_b) & \text{llenado} \\ T_{e_{in}} \frac{dP_{mij}}{dT} + P_{mij} = P_a & \text{expulsión} \end{cases} \quad (32)$$

Donde P_{mij} - presión en el final del conducto que llega al cilindro de frenado, T_e - Tiempo característico del conducto y Δt_b - retardo de tiempo debido al movimiento de la válvula.

III-J Aproximación de Pade para la característica $\varphi(s)$

Estudiamos el comportamiento del sistema en la región de valor máximo del momento de frenado, o sea, en la región del máximo de la función $\varphi(s)$, en el rango $0.12 \leq s \leq 0.52$, este rango corresponde a los valores de $\varphi(s)$ de hasta un 10% menores al valor máximo. Para esto proponemos la siguiente función de aproximación [10], [13]:

$$\varphi(s) = \frac{a_1 s^2 + a_2 s + a_3}{s^2 + a_4 s + a_5} \quad (33)$$

Para definir el valor de los coeficientes se escribe la ecuación en la siguiente forma:

$$a_1 s_i^2 + a_2 s_i + a_3 - \varphi(s_i)(a_4 s_i + a_5) = \varphi(s_i) s_i^2 \quad (34)$$

Donde i indica el valor de pruebas experimentales.

Entonces podemos construir el siguiente sistema:

$$PA = B$$

Donde

$$P = \begin{pmatrix} s_1^2 & s_1 & 1 & -\varphi(s_1)s_1 & -\varphi(s_1) \\ \vdots & \dots & \dots & \dots & \vdots \\ s_i^2 & s_i & 1 & -\varphi(s_i)s_i & -\varphi(s_i) \end{pmatrix}$$

La matriz P es no cuadrada y su tamaño depende del número de pruebas experimentales.

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_5 \end{pmatrix}$$

$$B = \begin{pmatrix} \varphi(s_1)s_1^2 \\ \vdots \\ \varphi(s_i)s_i^2 \end{pmatrix}$$

Calculamos A mediante el método de mínimos cuadrados y tenemos:

$$A = (P^T P)^{-1} P^T B$$

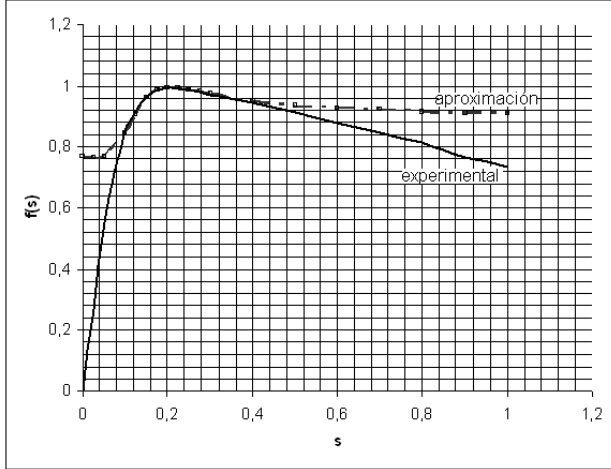


Fig. 8 Aproximación para $\varphi(s)$

III-K Modelo completo

Ecuaciones dinámicas

$$\begin{aligned} M \frac{dV_x}{dT} &= \sum_{i,j=1}^2 P_{ijx} + MV_y \Omega_z + F_{ax} \\ M \frac{dV_y}{dT} &= \sum_{i,j=1}^2 P_{ijy} - MV_x \Omega_z + F_{ay} \\ M \frac{dV_z}{dT} &= \sum_{i,j=1}^2 N_{ij} - Mg \end{aligned} \quad (34)$$

$$\begin{aligned} I_x \frac{d\Omega_x}{dT} &= [(N_{11} + N_{21}) - (N_{12} + N_{22})B] - Z \sum_{i,j=1}^2 P_{ijy} - \\ &- \sum_{j=1}^2 I_{1j} [\Omega_{1j} \sin \Theta + \Omega_{1j} \Theta \cos \Theta - \Omega_{1j} \Omega_z \cos \Theta] \end{aligned}$$

$$\begin{aligned} I_y \frac{d\Omega_y}{dT} &= (N_{21} + N_{22})A_2 - (N_{11} + N_{12})A_1 - Z \sum_{i,j=1}^2 P_{ijx} - \\ &- \sum_{j=1}^2 I_{1j} [\Omega_{1j} \cos \Theta + \Omega_{1j} \Theta \sin \Theta - \Omega_{1j} \Omega_z \sin \Theta] \\ I_z \frac{d\Omega_z}{dT} &= [(P_{12x} + P_{22x}) - (P_{11x} + P_{21x})]B + \\ &+ (P_{11y} + P_{12y})A_1 - (P_{21y} + P_{22y})A_2 \end{aligned} \quad (35)$$

Ecuaciones cinemáticas

$$\begin{aligned} \frac{dX}{dT} &= V_x \cos \psi - V_y \sin \psi \\ \frac{dY}{dT} &= V_x \sin \psi - V_y \cos \psi \\ \frac{dZ}{dT} &= V_z \\ \frac{d\psi}{dT} &= \Omega_z \\ \frac{d\gamma}{dT} &= \Omega_x \\ \frac{d\vartheta}{dT} &= \Omega_y \end{aligned}$$

Relaciones de la fuerza de contacto

$$\begin{aligned} P_{11x_0} &= P_{11x} \cos \Theta_{11} - P_{11y} \sin \Theta_{11} \\ P_{11y_0} &= P_{11x} \sin \Theta_{11} - P_{11y} \cos \Theta_{11} \\ P_{12x_0} &= P_{12x} \cos \Theta_{12} - P_{12y} \sin \Theta_{12} \\ P_{12y_0} &= P_{12x} \sin \Theta_{12} - P_{12y} \cos \Theta_{12} \\ P_{21x_0} &= P_{21x}, \quad P_{21y_0} = P_{21y} \\ P_{22x_0} &= P_{22x}, \quad P_{22y_0} = P_{22y} \end{aligned} \quad (36)$$

Vibraciones verticales

$$m_A \frac{d^2 z_{Aij}}{dT^2} = N_{ij} + C_{jz} \Delta z_{ij} + D_j \frac{d\Delta z_{ij}}{dT} \quad (37)$$

Ecuación de rotación de la rueda

$$I_j \frac{d\Omega_{ijy}}{dT} = -P_{ijx} (R - \zeta_{ij}) + L_{ij} \quad (38)$$

Movimiento del elemento de contacto

$$\begin{aligned} M_c \frac{d}{dT} (V_x + \Omega_{ij} R + \zeta_{ij}) &= -C_x \dot{\zeta}_{ij} - K_x \zeta_{ij} + P_{ijx} \\ M_c \frac{d}{dT} (V_y + \eta_{ij}) &= -C_y \dot{\eta}_{ij} - K_y \eta_{ij} + P_{ijy} \\ 0 &= -C_z \dot{\zeta}_{ij} - K_z \zeta_{ij} + N_{ij} \end{aligned} \quad (39)$$

Fuerza de contacto

$$\begin{aligned} P_{xij} &= -\nu N \frac{s_{xij}}{s_{ij}} \phi(s_{ij}) \\ P_{yij} &= -\nu N \frac{s_{yij}}{s_{ij}} \phi(s_{ij}) \\ s_{xij} &= \frac{V_{Axij} - \Omega_{ij} R + \zeta_{ij}}{V_{Axij}} \\ s_{yij} &= \frac{V_{Ayij} + \eta_{ij}}{V_{Ayij}} \\ s_{ij} &= \sqrt{s_{xij}^2 + s_{yij}^2} \end{aligned} \quad (40)$$

Modelo del sistema de frenos

$$L_{ij} = -K_L P_{mij} \text{sign} \Omega_{ij} \quad (41)$$

$$\begin{cases} T_{e_{in}} \frac{dP_{mij}}{dT} + P_{mij} = P_c (T + \Delta t_b) & \text{llenado} \\ T_{e_{in}} \frac{dP_{mij}}{dT} + P_{mij} = P_a & \text{expulsión} \end{cases} \quad (42)$$

III. ANÁLISIS FRACCIONAL

III-A Descripción de las constantes de tiempo

III-A.1 Tiempo característico del movimiento longitudinal del automóvil

Estudiamos el movimiento longitudinal del automóvil, se considera que el sistema es simétrico y no existen fuerzas laterales.

De (1) tenemos

$$M \frac{dV_x}{dT} = \sum_{i,j=1}^2 P_{ijx}$$

Analizamos el movimiento del vehículo en el caso en que $P_{ijx} = P^*$ para todo i, j . La ecuación toma la forma

$$\frac{dV_x}{dT} = \frac{4P^*}{M}$$

y

$$V_x = \frac{4P^*}{M} (T - T_0) - V_{x_0}$$

Entonces el tiempo característico será en el cual V_x crece de 0 a un valor dado V_{x^*} , tomando en cuenta $V_{x_0} = 0, T_0 = 0$ y $V_x = V_{x^*}$ tenemos:

$$T_0 = \frac{MV_{x^*}}{4P^*} \quad (43)$$

T_0 - Tiempo característico en el cual la velocidad del vehículo cambia en una magnitud de orden V_{x^*} , debido a una fuerza proporcional a su peso $4P^* = Mg$.

III-A.2 Constante de tiempo de vibraciones verticales del cuerpo del vehículo

Estudiamos las vibraciones elásticas verticales del cuerpo del vehículo en los resortes, para lo cual tomamos el modelo mostrado en la fig. 9. Se desprecia la fuerza de amortiguamiento de los amortiguadores tanto delanteros como traseros.

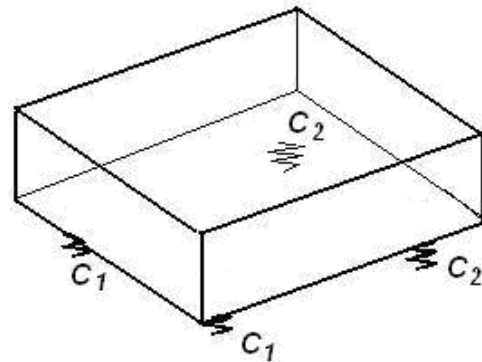


Fig. 9 Modelo para vibraciones verticales

$$M \frac{d^2 z}{dT^2} = 2(C_1 + C_2)z - Mg$$

Reescribiendo tenemos:

$$\frac{M}{2(C_1 + C_2)} \frac{d^2 z}{dT^2} = z - \frac{Mg}{2(C_1 + C_2)}$$

Tomamos como constante de tiempo el periodo de vibraciones del chasis sobre el resorte

$$T_1 = \sqrt{\frac{M}{2(C_1 + C_2)}} \quad (44)$$

T₁ - constante de tiempo de las vibraciones del cuerpo del automóvil debidas a los resortes de la suspensión.

III-A.3 Constante de tiempo parcial de vibraciones de la masa no deformable del neumático y la suspensión

Estudiamos las vibraciones verticales debidas al movimiento longitudinal sobre la superficie de contacto del automóvil.

Escribimos las ecuaciones de vibraciones verticales de la rueda como en (28) para un cuerpo en reposo con una superficie de contacto plana.

$$m_A \frac{d^2 z_{Aij}}{dT^2} = K_z \zeta_{ij} + C_{jz} \Delta z_{ij}$$

Como z_{Aij}=R- ζ_{ij} y Δz_{ij}= z_{Aij}- z_{Aij0} podemos escribir de la siguiente forma

$$m_A \frac{d^2 \zeta_{ij}}{dT^2} = K_z \zeta_{ij} - C_{jz} \dot{\zeta}_{ij}$$

Entonces el periodo de las vibraciones verticales es igual a

$$T_2 = \sqrt{\frac{m_A}{K_z + C_{jz}}} \quad (45)$$

T₂ - Periodo de vibraciones verticales de la masa no deformable.

III-A.4 Constante de tiempo de cambios en la velocidad angular de la rueda, debida a una fuerza de contacto longitudinal

Estudiamos el movimiento de la rueda en dirección horizontal. Suponemos que no existen bloqueos en la rueda ni derrapes, también consideramos que el tamaño de la parte deformable del neumático es pequeño en comparación con el tamaño total del neumático, el automóvil se mueve a una velocidad

constante V_{x*}.

De las ecuaciones (4), (28) y (30) tenemos

$$I_j \frac{d\Omega_{ijy}}{dT} = -P_{ijx} R + L_{ij}$$



donde P_{ijx} = vN_{ij}K₀ε_{ijx}

Reescribiendo tenemos

$$\frac{I_j V_{x*}}{RT} \frac{d\Omega_{ijy}}{dT} = -R v N_* K_0 \varepsilon_{ijx} \sum_{i,j=1}^2 N_{ij} + L_{ij}$$

Reacomodando algunos términos

$$\frac{I_j V_{x*}}{R^2 v N_* K_0 T} \frac{d\Omega_{ijy}}{dT} = -\varepsilon_{ijx} \sum_{i,j=1}^2 N_{ij} + L_{ij}$$

finalmente obtenemos

$$T_3 = \frac{I_j V_{x*}}{R^2 v N_* K_0}$$

T₃ - Estimación de la constante de tiempo de cambios en la velocidad angular de la rueda debida a la acción de una fuerza de contacto longitudinal.

III-A.5 Constante de tiempo parcial de la deformación del neumático

Estudiamos los cambios de la deformación del neumático, suponemos que no existen vibraciones verticales, tampoco existen fuerzas laterales y el movimiento ocurre solo en dirección longitudinal.

Consideramos V_{ijx} y Ω_{ijy} constantes.

De las ecuaciones (15), (28) y (30) tenemos

$$M_c \frac{d^2 \xi_{ij}}{dT^2} = -K_x \xi_{ij} - C_x \frac{d\xi_{ij}}{dT} + P_{ijx}$$

Donde la fuerza de contacto para deslizamientos pequeños se supone lineal.

$$P_{ijx} = -v N_{ij} K_0 \left(\frac{V_{ijx} + \Omega_{ijy} R + \frac{d\xi_{ij}}{dT}}{V_{ijx}} \right)$$

Entonces

$$M_c \frac{d^2 \xi_{ij}}{dT^2} + K_x \xi_{ij} + C_x \frac{d\xi_{ij}}{dT} = -\nu N_{ij} K_0 \left(\frac{V_{ijx} + \Omega_{ijy} R + \frac{d\xi_{ij}}{dT}}{V_{ijx}} \right)$$

Reescribiendo

$$M_c \frac{d^2 \xi_{ij}}{dT^2} + \left(C_x + \frac{\nu N_{ij} K_0}{V_{ijx}} \right) \frac{d\xi_{ij}}{dT} + K_x \xi_{ij} = -\nu N_{ij} K_0 \left(\frac{V_{ijx} + \Omega_{ijy} R}{V_{ijx}} \right)$$

Entonces el periodo de cambio de ξ es igual a

$$T_{4*} = \frac{1}{\sqrt{\frac{K_x}{M_c} - \left(\frac{C_x V_{ijx} + \nu N_{ij} K_0}{2M_c V_{ijx}} \right)^2}}$$

Si $\frac{K_x}{M_c} \gg \left(\frac{C_x V_{ijx} + \nu N_{ij} K_0}{2M_c V_{ijx}} \right)^2$ podemos escribir

$$T_4 = \sqrt{\frac{M_c}{K_x}} \quad (47)$$

Tomamos T_4 como estimación de la constante de tiempo parcial de deformaciones en el neumático.

III-A.6 Constante de tiempo de parcial de deformación del neumático en la dirección lateral

Estudiamos la deformación del neumático, suponemos que no existen vibraciones verticales, tampoco fuerzas laterales externas. El movimiento del automóvil ocurre en dirección longitudinal.

De las ecuaciones (16), (28) y (30) tenemos

$$M_c \frac{d^2 \eta_{ij}}{dT^2} = -K_y \eta_{ij} - C_y \frac{d\eta_{ij}}{dT} + P_{ijy}$$

Para la parte de dependencia lineal de $\varphi(s)$ tenemos

$$P_{ijy} = -\nu N_{ij} \frac{s_{ijy}}{s_{ij}} \varphi(s_{ij}) = -\nu N_{ij} K_0 \frac{d\eta_{ij}}{dT}$$

Entonces

$$M_c \frac{d^2 \eta_{ij}}{dT^2} + C_y \frac{d\eta_{ij}}{dT} + K_y \eta_{ij} = -\nu N_{ij} K_0 \frac{d\eta_{ij}}{dT}$$

El periodo de cambio de η es

$$T_{5*} = \frac{1}{\sqrt{\frac{K_y}{M_c} - \left(\frac{C_y + \nu N_{ij} K_0}{2M_c} \right)^2}}$$

Como $\frac{K_y}{M_c} \gg \left(\frac{C_y + \nu N_{ij} K_0}{2M_c} \right)^2$ podemos escribir

$$T_5 = \sqrt{\frac{M_c}{K_y}} \quad (48)$$

T_5 - Estimación de la constante de tiempo parcial de la deformación lateral del neumático.

III-A.7 Constante de tiempo de vibraciones laterales del cuerpo del vehículo

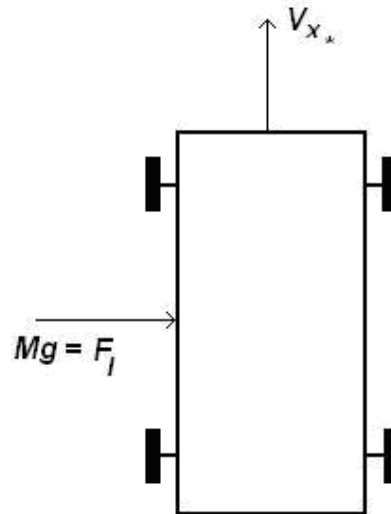


Fig. 10 Modelo para vibraciones laterales

Estudiamos el movimiento lateral del automóvil para el caso en que las fuerzas laterales toman un valor en el umbral de $\varphi(s)$, tal como se muestra en la fig. 10, donde podemos modelar esta característica con la siguiente función lineal

$$\varphi(s) = K_0 s \quad (49)$$

Suponemos que el automóvil se mueve en dirección horizontal debido a una fuerza constante $F_1 \sim Mg$.

Esto se cumple cuando $V_y, \Omega z (A_1 + A_2) \ll V_x$ entonces tenemos

$$M \frac{dV_y}{dT} = \sum_{i,j=1}^2 P_{ijy} + F_1$$

Tomando en cuenta (51), $P_{ijy} = \nu N K_0 s_y$ o sea

$$M \frac{dV_y}{dT} = -\nu K_0 \frac{V_y}{V_{x^*}} \sum_{i,j=1}^2 N_{ij} + Mg$$

Reordenando

$$\frac{V_{x^*}}{\nu g K_0} \frac{dV_y}{dT} + V_y = \frac{V_{x^*}}{\nu K_0}$$

la constante de tiempo de esta ecuación es

$$T_6 = \frac{V_{x^*}}{\nu g K_0} \quad (50)$$

III-A.8 Constante de tiempo del sistema de frenos

Suponemos que la presión en el cilindro de frenos $P^* = \text{const}$, para la fase de llenado $P^* = P_c$ y para la fase de expulsión $P^* = P_a$

Entonces elegimos T_{ein} como tiempo característico para el llenado y T_{eout} como tiempo característico de expulsión.

Debido a que estos tiempos se diferencian más o menos en 3 veces elegimos el menor de ellos como tiempo característico del sistema de frenos.

$$T_7 = T_{e_{in}} \quad (51)$$

III-B Relación de tiempos característicos

En adelante analizaremos los parámetros correspondientes a un autobús de transporte público metropolitano.

Se cuenta con los siguientes valores numéricos:

$M = 9584\text{kg} \sim 10^4$, $K_0 = 10$, $\nu \sim 1$ (varía de acuerdo a las características del camino, para asfalto mojado 0.5 y para asfalto seco 0.8), $C_1 = 248487\text{Ns/m} \sim 2.5 \times 10^5$, $C_2 = 407115\text{Ns/m} \sim 4 \times 10^5$, $m_A = 390\text{kg}$, $K_z = 841960\text{N/m} \sim 8.5 \times 10^5$, $I_j = 18.9\text{kgm}^2 \sim 20$, $R = 0.53\text{m} \sim 0.5\text{m}$, a una velocidad $V_x = 20\text{m/s}$

Entonces

$$T_0 \sim 1$$

$$T_1 \sim \frac{1}{10}$$

$$T_2 \sim \frac{1}{50}$$

$$T_3 \text{ de } \frac{1}{300} \text{ hasta } \frac{1}{60} \text{ dependiendo de } \nu$$

$$T_4 \sim \frac{1}{10000}$$

$$T_5 \sim \frac{1}{10000}$$

$$T_6 \sim \frac{1}{10}$$

$$T_7 \sim \frac{1}{30}$$

III-C Análisis fraccional de las ecuaciones de movimiento

Introducimos variables adimensionales (el símbolo “*” significa el valor característico dimensional de la cantidad elegida).

$$V_x = V_{x^*} v_x \quad (52)$$

Aquí V_{x^*} - velocidad característica en la dirección longitudinal. Para V_{x^*} elegimos la velocidad inicial del automóvil en la dirección longitudinal.

Si el tiempo analizado no es grande entonces ésta velocidad no alcanza a cambiar significativamente y $v_x \sim 1$

$$V_y = V_{y^*} v_y \quad (53)$$

Donde V_{y^*} - velocidad característica en la dirección lateral. Para V_{y^*} elegimos la velocidad inicial del automóvil en la dirección lateral. Suponemos que el movimiento en la dirección lateral es tal, que la velocidad lateral no sobrepasa $0.1V_{x^*}$

$$V_z = V_{z^*} v_z \quad (54)$$

V_{z^*} - Velocidad característica en la dirección vertical. Para V_{z^*} elegimos la amplitud característica del cambio en la velocidad debido a vibraciones verticales de la suspensión en la

frecuencia característica. $V_{z^*} = \frac{Z_* 2\pi}{T_1}$

$$N_{ij} = N_* n_{ij} \quad (55)$$

N_* - valor característico de la reacción normal,

para N_* elegimos el valor $N_* = \frac{Mg}{4}$

$$P_{ijx} = P_* P_{ijx} \quad (56)$$

$$P_{ijy} = P_* P_{ijy} \quad (57)$$

Aquí P_* - valor característico de la fuerza de contacto. Suponemos movimiento en la dirección longitudinal del automóvil y ausencia de vibraciones verticales. Entonces $P_* = v_* N_*$, donde $v_* = 1$

$$\Omega_x = \Omega_{x^*} \omega_x \quad (58)$$

Aquí Ω_{x^*} - velocidad angular característica de viraje del automóvil, en ausencia de vibraciones verticales. El giro ocurre alrededor del centro de masa. Elegimos $\Omega_{x^*} = \frac{V_{z^*}}{B}$

$$\Omega_y = \Omega_{y^*} \omega_y \quad (59)$$

Ω_{y^*} - velocidad angular característica de inclinación del automóvil, en ausencia de vibraciones verticales. El giro ocurre alrededor del centro de masa. Tenemos $\Omega_{y^*} = \frac{V_{z^*}}{A_2}$

$$\Omega_z = \Omega_{z^*} \omega_z \quad (60)$$

Ω_{z^*} - velocidad angular característica de cambio en el curso del automóvil, el automóvil gira alrededor del eje vertical del centro de masa. Tenemos

$$\Omega_{z^*} = \frac{V_{y^*}}{A_1}$$

$$\Omega_{ij} = \Omega_* \omega_{ij} \quad (61)$$

Donde Ω_* - velocidad angular característica de rotación de la rueda. Para Ω_* escogemos la velocidad angular en ausencia de de deslizamientos y derrapes. El automóvil se mueve en dirección longitudinal con velocidad V_{x^*} , por lo que

$$\Omega_* = \frac{V_{x^*}}{R}$$

$$\Theta_{ij} = \Theta_* \theta_{ij} \quad (62)$$

Aquí Θ_* - ángulo característico de vuelta en las ruedas delanteras alrededor del eje vertical. Seleccionamos como valor característico el valor máximo de éste ángulo.

$$\Psi = \Psi_* \psi \quad (63)$$

Ψ_* - ángulo característico de viraje del automóvil. Seleccionamos como Ψ_* , el cambio en el ángulo Ψ debido a la rotación del chasis con una velocidad

angular Ω_{z^*} en el tiempo T_6 , entonces $\Psi_* = \Omega_{z^*} T_6$
 $\Gamma = \Gamma_* \gamma \quad (64)$

Γ_* - ángulo característico de inclinación del automóvil. Escogemos como Γ_* el cambio en el ángulo Γ debido a la inclinación del chasis durante el tiempo característico de vibraciones verticales, tenemos $\Gamma_* = \Omega_{x^*} T_1$

$$\Xi = \Xi_* \vartheta \quad (65)$$

Ξ_* - ángulo característico de curso del automóvil. Escogemos el cambio en el ángulo después del tiempo de vibraciones verticales, tenemos $\Xi_* = \Omega_{y^*} T_1$

$$\xi_{ij} = \xi_* \xi_{ij} \quad (66)$$

ξ_* - valor característico de deformación de la llanta en dirección longitudinal. Seleccionamos para la deformación de la llanta debida a una velocidad longitudinal durante el tiempo característico T_3 , tenemos $\xi_* = V_{x^*} T_3$

$$\eta_{ij} = \eta_{ij^*} \eta_{ij} \quad (67)$$

η_* - valor característico de deformación de la llanta en dirección lateral. Seleccionamos para η_* la deformación de la llanta debida a una velocidad lateral V_{y^*} durante el tiempo característico T_5 , tenemos $\eta_* = V_{y^*} T_5$

$$\zeta_{ij} = \zeta_* \zeta_{ij} \quad (68)$$

ζ_* - valor característico de deformación de la llanta en dirección vertical. Seleccionamos como ζ_* la deformación de la llanta debida a la reacción

normal, entonces $\zeta_* = \frac{Mg}{4K_z}$

$$T = T_* t \quad (69)$$

T_* - tiempo característico, el cual será seleccionado dependiendo de la tarea a resolver, para nuestro caso se utilizará T_3 .

Como se mencionó anteriormente nos interesa el comportamiento del vehículo alrededor de la fuerza de contacto máxima, que es cuando entra en funcionamiento el sistema ABS, para este caso el tiempo característico es T_3 , los cambios de

velocidad longitudinal se consideran muy pequeños en comparación a los cambios en la velocidad angular de la rueda, por lo que podemos considerar $V_x = \text{const.}$

El primer paso para la reducción del modelo consisten en sustituir las variables descritas en el sistema de ecuaciones por su equivalente que contiene un valor característico y una variable adimensional, después se requiere sustituir el valor del tiempo característico T_3 , después de la manipulación algebraica de las ecuaciones con los valores sustituidos se llega al siguiente sistema

$$\frac{d^2 \xi_{ij}}{dt^2} = -q \frac{d\xi_{ij}}{dt} - p \xi_{ij} - v k n_{ij} \varphi(s_{ij}) - \frac{d\omega_{ij}}{dt} \quad (70)$$

$$\frac{d\omega_{ij}}{dt} = -p_{ijx} + l_{ij} \quad (71)$$

$$\begin{cases} \frac{dp_{m_{ij}}}{dt} = \frac{T_3}{T_{e_{in}}} \left(-p_{m_{ij}} + \frac{P_c}{P_{m^*}} \right) \\ \frac{dp_{m_{ij}}}{dt} = \frac{T_3}{T_{e_{out}}} \left(-p_{m_{ij}} + \frac{P_{ac}}{P_{m^*}} \right) \end{cases} \quad (72)$$

$$s_{ij} = 1 + \omega_{ij} + \frac{d\xi_{ij}}{dt}$$

El sistema resultante es adimensional y describe solo las variables que influyen directamente en la dinámica del vehículo para intervalos de tiempo dentro del rango de T_3 .

IV. CONCLUSIONES

Se presentó el modelo completo de un vehículo automotor, tomando en cuenta las diferentes dinámicas que afectan su movimiento, se consideró primero el vehículo como un cuerpo rígido con 6 grados de libertad, posteriormente se presentaron algunos subsistemas que fueron tomados como casos aislados.

El modelo completo resultante tomó valores propios de un autobús de transporte metropolitano, con lo cual se tuvieron los elementos necesarios para realizar el análisis fraccional, dado que el interés principal es la obtención de un modelo adecuado para el estudio de sistemas ABS se consideró T_3 , la constante de tiempo de cambios en la velocidad angular como constante de tiempo característico, con lo cual se obtuvo un sistema de

solo 4 ecuaciones, teniendo como variables la velocidad angular de la rueda, la deformación longitudinal del elemento de contacto, la presión en el cilindro de trabajo y el deslizamiento relativo de la rueda.

Es importante mencionar que el modelo obtenido es adimensional y normalizado, con lo cual es posible hacer una comparación directa entre los cambios que se producen para diferentes variables con respecto al tiempo.

V. REFERENCIAS

- [1] Bulgakov, B., *Vibraciones*, Ed. Estatal de Literatura Técnica Teórica, 1949 (en ruso)
- [2] Dixon, J., *Tires, Suspension and Handling*, SAE International, 1996
- [3] Gantmájér, F., *Mecánica Analítica*, Editorial URSS, 2003
- [4] Genta, G., *Motor Vehicle Dynamics, Modeling and Simulation*, World Scientific, 2006
- [5] Gillespie T. *Fundamentals of Vehicle Dynamics*, SAE International, 1992
- [6] Kruchinin, P., Mogamedov, M. y Novozhilov I., “*Modelo matemático de una rueda de automóvil para régimen de movimiento antibloqueo*”, *Mekhanika Tverdogo Tela*, Moscú, 2001, No. 6, pp 63-69 (en ruso)
- [7] Kruchinin, P., Mogamedov, M. y Makarov, L. “*Rechazo activo de vibraciones parásitas en el trabajo de sistema ABS*”, *Vestnik Moskovskogo Universiteta*, serie 1, matemática y mecánica, Moscú, 2003, No. 5, pp 25-29
- [8] Novozhilov, I., *Fractional Analysis, Methods of motion decomposition*, Birkhäuser, 1997
- [9] Olson, B., Shaw, S., Stépán, G., “*Nonlinear Dynamics of Vehicle Traction*”, *Vehicle Systems Dynamics*, Vol 40, No. 6, 2003, pp 377-399
- [10] Pacejka H.B., *Tire and Vehicle Dynamics*, SAE International, 2002
- [11] Rajamani, R., *Vehicle Dynamics and Control*, Springer, 2006

- [12] Vázquez, I., Acosta, C., Rubio J., Valentín, J., “*Modelado Matemático para el Sistema Suspensión-Frenos de un Autobús*”, 6th International Conference on Electrical and Electronics Engineering Research, Noviembre de 2008, pp 251 – 255
- [13] Vázquez, I., Kruchinin, P., “*About wheels vibrations during the operation on automobile antilock brake system*”, IX International Workshop on Stability and Oscillations of Nonlinear Control Systems, Mayo 31 de 2006, Moscú, Rusia

DISEÑO Y PRIMER PROTOTIPO DE MANO ROBÓTICA A BASE DE MATERIALES CON MEMORIA DE FORMA

Farías López Valerie, Solís Paredes Lucía, Meléndez Rasgado Lorena,
García Gallegos Claudia, Velázquez Guerrero Ramiro

Mecatrónica y Control de Sistemas - Universidad Panamericana,
Josemaría Escrivá de Balaguer 101, Fracc. Rústicos Calpulli,
20290, Aguascalientes, Ags., México

Tel: (449) 9106200, web: www.robotica-up.org, correo electrónico: rvelazquez@ags.up.mx

RESUMEN

Este artículo presenta el diseño e implementación de una mano robótica de 4 dedos para aplicaciones de prensión y manipulación diestra y precisa. El primer prototipo desarrollado es de tamaño humano y tiene 3 grados de libertad (GdL) por dedo y 2 GdL en el pulgar para un total de 11 GdL. Para accionar las articulaciones de la mano robótica, se emplean materiales inteligentes como las aleaciones con memoria de forma (AMF). Los materiales en AMF presentan ciertas ventajas: son compactos, ultra-ligeros, de bajo costo y su accionamiento es limpio y silencioso. Para incrementar la eficiencia y disminuir la disipación de potencia, los micro-actuadores en AMF incluyen estructuras magnéticas bi-estables. Diseño mecánico, principio de accionamiento y primer prototipo desarrollado serán presentados y discutidos.

Palabras clave: actuador bi-estable, micro-bobinas, mano robótica, aleaciones con memoria de forma (AMF).

I. INTRODUCCIÓN

La mano humana es una estructura altamente funcional que investigadores han intentado imitar desde hace mucho tiempo. Con sus 20 GdL, 19 músculos, 17 juntas, 19 huesos, además de sus ligamentos, nervios y numerosos sensores, la mano humana es una estructura muy compleja y difícil de reproducir mecánicamente.

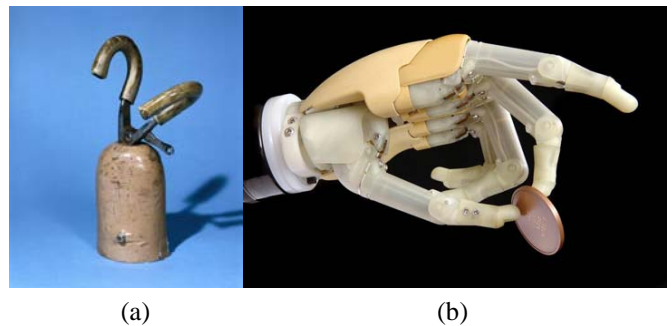


Fig. 1. Prótesis de mano: (a) mano tipo gancho (1979) y (b) mano i-Limb (2007), de [6] y [7] respectivamente.

Más aún, ésta es diestra, precisa y estable, pero al mismo tiempo es fuerte, flexible y capaz de efectuar movimientos rápidos.

A lo largo de los últimos 20 años, un sinnúmero de manos robóticas han sido desarrolladas para aplicaciones de manipulación en diversas áreas como la médica, espacial, industrial, ambientes virtuales y el bienestar común [1]-[5].

Sin duda alguna, el campo de aplicación donde más se han desarrollado manos artificiales es en el diseño y fabricación de prótesis para personas amputadas. En 30 años, la emblemática prótesis tipo gancho ha devenido en una prótesis con características inclusive superiores a las de la mano real (Fig. 1).

Tradicionalmente los diseños de manos robóticas han tendido a ser pesados, voluminosos y ruidosos debido al uso de actuadores convencionales basados en las tecnologías neumáticas y electromecánicas. Como resultado de los problemas inherentes a este tipo de actuadores,

los diseñadores han adaptado otro tipo de tecnologías de accionamiento. Se cree que el uso de materiales inteligentes como las aleaciones con memoria de forma (AMF), los cerámicos piezoeléctricos, los polímeros electro-activos (EAPs), y los fluidos electro-rheológicos (ER), proveerá nuevas metodologías de diseño y paradigmas para construir manos robóticas ligeras y de mejor desempeño.

La tecnología AMF parece ser una de las candidatas más prometedoras para el desarrollo de meso/micro actuadores: su tamaño compacto, la buena relación potencia entregada/peso, la alta resistencia a la fatiga en operación cíclica y su desempeño limpio (sin grasas, chispas, etc.) y silencioso los convierte en un principio interesante de accionamiento [8]. Sin embargo, también presentan 2 grandes desventajas: baja velocidad de respuesta y un comportamiento no-lineal que los hace difíciles de controlar.

En este artículo se presenta el diseño y la implementación de una mano robótica a base de materiales en AMF. El tamaño del primer prototipo desarrollado es el de una mano humana, con 4 dedos y 11 GdL. Se propone un micro-actuador basado en la tecnología AMF para accionar las articulaciones de los dedos. El actuador consiste en par antagonico de micro-bobinas fabricadas con la aleación Nickel-Titanio (NiTi) junto con estructuras magnéticas bi-estables para incrementar su eficiencia general. Este diseño permite tener todos los actuadores incluidos intrínsecamente dentro de la estructura de la mano, de manera que el prototipo resulta compacto, ligero, de bajo consumo de potencia y económico.

El resto del artículo está organizado como sigue: La Sección 2 describe las características esenciales de la tecnología AMF. La Sección 3 presenta el diseño, la caracterización y la implementación de un micro-actuador a base de micro-bobinas en AMF. La Sección 4 introduce el diseño de una mano robótica mientras que la Sección 5 presenta el prototipo desarrollado así como su desempeño en prensión de objetos. Finalmente la Sección 6 resume las contribuciones y perspectivas a futuro de esta investigación.

II. CARACTERÍSTICAS ESENCIALES DE LA TECNOLOGÍA AMF

El efecto de memoria de forma (EMF) es la habilidad de cierto grupo de materiales de “recordar” una forma específica cuando están sujetos a un proceso térmico. Los materiales que exhiben esta propiedad son en su mayoría aleaciones metálicas como Au-Cd, Au-Cd, Cu-Al-Ni, Cu-Sn, In-Ti, Ni-Al y el popular Ni-Ti [9].

El EMF ocurre como resultado de una transformación que depende de la temperatura y que sucede entre dos fases cristalinas: la *austenita* (fase que tiene lugar a altas temperaturas) y la *martensita* (fase que ocurre a bajas temperaturas). La austenita es la fase que contiene la forma predefinida o memorizada de la estructura cristalina del material. Cuando el material se enfría hasta la fase martensita, su estructura cristalina presenta poca rigidez; es bastante maleable y es capaz de adquirir y retener nuevas formas. Al calentar el material, la estructura cristalina regresa a su configuración original en la fase austenita (Fig. 2 superior).

A nivel microscópico, la estructura cristalina en fase austenita es muy ordenada y altamente simétrica. Una vez enfriada, ésta colapsa inclinándose en direcciones opuestas a través de capas subsecuentes que se auto-acomoda de tal forma que no se aprecia una deformación macroscópica en el material. El aplicar un esfuerzo externo al material en fase martensita causa que las capas se inclinen en la misma dirección. Una vez que todas las capas se han inclinado en un solo sentido, se dice que el material en AMF se ha alineado. Al calentarse, las capas cristalinas se alienan nuevamente para recobrar su simetría original (Fig. 2 inferior).

El EMF es repetible y se considera un fenómeno de transformación de energía térmica a trabajo mecánico, que por lo general produce un desplazamiento y una fuerza. Por ello, los materiales en AMF son utilizados como elementos activos en actuadores. Las AMF pueden tomar casi cualquier forma. Las formas más populares son alambre recto, bobina, tubo, plana y lámina u hoja.

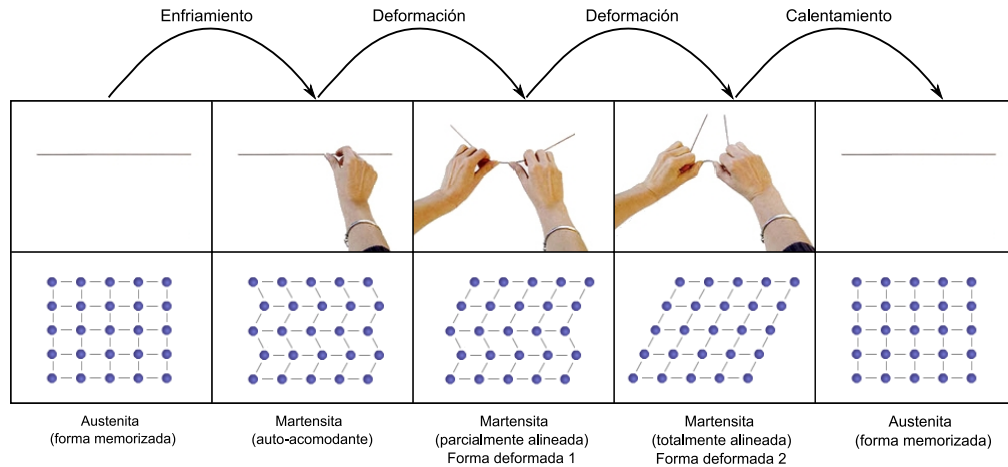


Fig. 2. El EMF en AMF a nivel macroscópico y cristalográfico.

Pruebas experimentales realizadas en alambres rectos de NiTi han revelado que la carrera del material está limitada a 4-5% de su longitud original [10]. Comparado con el alambre, los resortes o bobinas poseen una carrera mayor. Carreras impresionantes de 300% pueden obtenerse fácilmente con bobinas helicoidales. Sin embargo, debido a que su desempeño es en torsión en lugar de tensión, no son capaces de desarrollar la misma fuerza.

Este artículo sugiere que bobinas helicoidales en AMF pueden ser diseñadas para ejercer fuerzas significativas para accionar las articulaciones de manos robóticas, ofreciendo además, carreras que los alambres rectos son incapaces de proporcionar.

III. UN MICRO-ACTUADOR EN AMF BI-ESTABLE

A. Material

Micro-bobinas helicoidales en NiTi fueron fabricadas con alambre recto Flexinol (Ni₅₂Ti₄₈) con las siguientes características geométricas: 200 μm de diámetro de alambre, 1.5 mm de diámetro externo de la bobina, 12 espiras y 7 mm de largo. La masa de cada micro-bobina es de tan solo 30 mg y el costo de laboratorio es de 1 USD.

La forma helicoidal se logró enrollando el alambre alrededor de un tornillo y posteriormente calentando ambos a 600°C por 5 min.

Posteriormente se procedió a enfriar o templar la micro-bobina con agua, concluyendo de esta manera el procedimiento (Fig. 3).

Pruebas realizadas con calorímetro de escaneo diferencial (DSC) revelaron que la fase austenita se consigue totalmente a 65°C mientras que la fase martensita se alcanza totalmente a los 36° C [11].

B. Evaluación de desempeño

Los parámetros geométricos de la micro-bobina se ajustaron para que ésta desarrollara una fuerza apropiada de acuerdo con su deformación y temperatura. La fig. 4(a) muestra la relación experimental carga-deformación bajo diferentes condiciones isotérmicas para una deformación máxima de 10 mm.

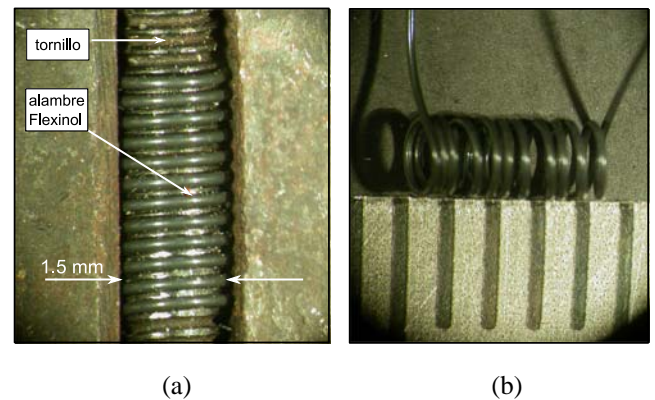


Fig. 3. Fabricación de una micro-bobina en AMF: (a) alambre Flexinol enrollado en un tornillo y (b) forma helicoidal después de tratamiento térmico.

Se aprecian 2 comportamientos diferentes: *elástico* antes que la carga alcance un valor crítico y *plástico* después de alcanzar este valor. La curva indica que la fuerza máxima de la micro-bobina en AMF a una deformación de 10 mm es aproximadamente de 180 mN a 20°C, 410 mN a 85°C y 720 mN a 105°C.

A partir de estos datos es posible formular un modelo computacional del comportamiento de la micro-bobina en el espacio carga-deformación-temperatura (Fig. 4(b)) [11]. Note que, al igual que la fig. 4(a), la fuerza de tensión desarrollada por la micro-bobina contra una carga, se incrementa al aumentar la temperatura.

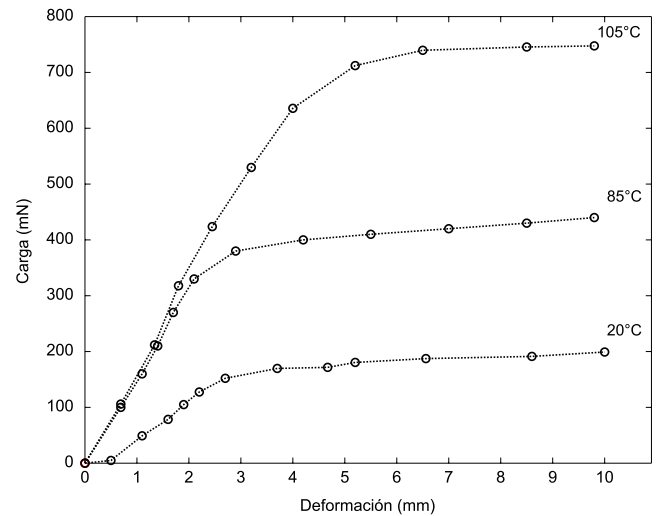
C. Diseño de actuadores

Como se mencionó en la Sección 2, los materiales en AMF logran su accionamiento a través del EMF que es un fenómeno repetible pero no reversible. Esto implica que por sí mismo, un AMF no puede ser considerado como actuador dado que no es capaz de proveer movimiento reversible.

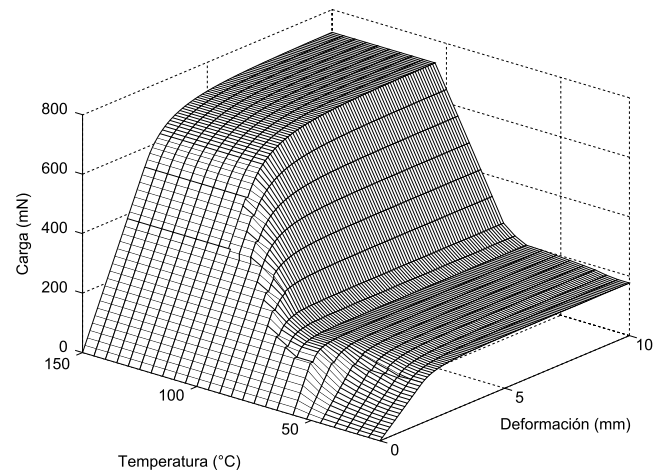
Para aplicaciones de actuadores, existen 2 tipos de materiales en AMF: de simple efecto y de doble efecto.

El efecto simple describe la capacidad o habilidad de los AMF para recobrar una forma memorizada cuando éstos se calientan hasta la fase de austenita, pero retienen esta forma cuando se enfrían a la fase martensita. Para que el AMF pueda ser empleado en aplicaciones de accionamiento cíclico, es necesario introducir una fuerza que induzca una deformación en la fase martensita.

El efecto doble describe un proceso de memoria con 2 formas estables: una en austenita y la otra en martensita. En consecuencia, el efecto doble no requiere de ningún mecanismo externo para lograr el accionamiento cíclico. El uso de AMF de doble efecto implica menos elementos, es más simple y compacto. Las desventajas son que requiere procesos extras de manufactura, es difícil de lograr correctamente y



(a)



(b)

Fig. 4. Relación experimental carga-deformación de la micro-bobina a temperatura constante (20, 85 y 105°C) y (b) modelo computacional en el espacio carga-deformación-temperatura. Se observó experimentalmente que a partir de 105°C el comportamiento carga-deformación era prácticamente el mismo.

la fuerza desarrollada es solo la mitad de la que se observa en AMF de simple efecto [12].

Aún cuando los AMF de simple efecto pueden ser diseñados de manera que ejerzan una fuerza en tres dimensiones (cuando se deforman en 3 direcciones a partir de la configuración de memoria), la gran mayoría solo aplican una fuerza de tensión direccional y no pueden aplicar directamente una fuerza de compresión. En muchas aplicaciones en robótica, ésto se

consigue mediante el empleo de una masa, un elemento elástico o un segundo AMF en arreglo antagonista [13]. Los mecanismos de la fig. 5 muestran como estas configuraciones pueden ser usadas para diseñar actuadores de movimiento lineal.

Note que todas estas configuraciones proveen movimiento esencialmente entre 2 puntos definidos por los estados activos y pasivos del AMF. Sin embargo, solo el estado pasivo puede ser considerado estable ya que no necesita potencia para mantenerse. En contraste, el estado activo no sólo requiere energía para alcanzarse sino también para mantenerse; el AMF debe permanecer activo, lo que lleva a una eficiencia pobre e implica serios inconvenientes cuando el AMF se calienta eléctricamente.

En principio, un AMF puede ser calentado arbitrariamente rápido haciendo circular una corriente suficientemente grande a través de él. Sin embargo, corrientes excesivamente grandes tienen la capacidad de sobrecalentar el AMF causando fatiga por esfuerzo térmico y una degradación gradual de su desempeño. Se requiere por tanto, una estrategia para retener la posición activa del AMF en fase martensita, de manera que se evite el consumo de potencia y el sobrecalentamiento inútil.

Un método para retener las 2 posiciones estables consiste en el empleo de mecanismos de bloqueo. De este modo, el actuador necesita energía únicamente al cambiar de un estado a otro y no requiere de energía en ninguna de las posiciones finales.

Si consideramos que las articulaciones de la mano permanecen la mayor parte del tiempo en una posición fija, se puede concluir que los actuadores que consumen una gran cantidad de energía para mantener su estado activo no son aceptables. Los actuadores bi-estables son la mejor opción para este tipo de aplicación [14].

En la práctica, el empleo de estructuras bi-estables ofrece ciertas ventajas: éstas proveen un movimiento preciso, repetitivo y en presencia de alteraciones, perturbaciones y variaciones en el ambiente, mantienen la posición deseada [15].

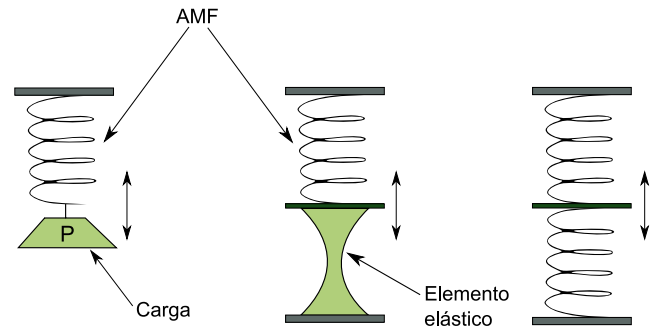


Fig. 5. Configuraciones principales de actuadores con AMF de simple efecto.

IV. DISEÑO DE UNA MANO ROBÓTICA CON AMF

El diseño de una mano robótica antropomórfica de 4 dedos se muestra en la fig. 6 izquierda. Este diseño tiene 3 GdL por dedo y 2 GdL en el pulgar para un total de 11 GdL. Los movimientos posibles de la mano se muestran en la fig. 7. Note que la flexión máxima de cada articulación es de 90° . Así, si las 3 articulaciones se flexionan 90° , el espacio de trabajo de las puntas de los dedos es de 360° .

La fig. 6 derecha muestra la estructura mecánica equivalente. Esta estructura es del tamaño de una mano humana adulta. La base y los dedos fueron enteramente fabricados de aluminio ligero, mientras que las articulaciones se elaboraron a partir de plástico endurecido. La masa de cada unión del dedo es estimada en 35g, por lo que el dedo completo se estima en 105g. La masa de estructura total es de alrededor 600g.

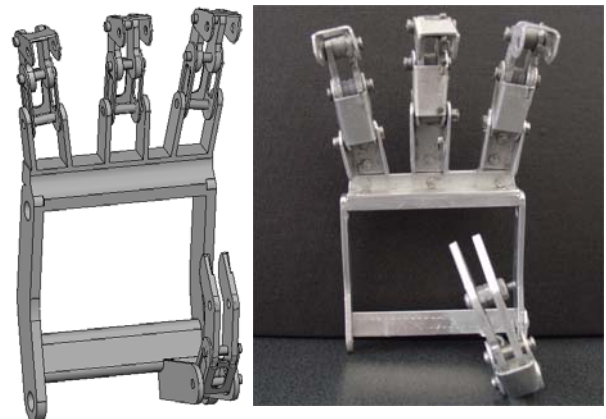


Fig. 6. Diseño de una mano robótica de 4 dedos y su estructura mecánica equivalente.

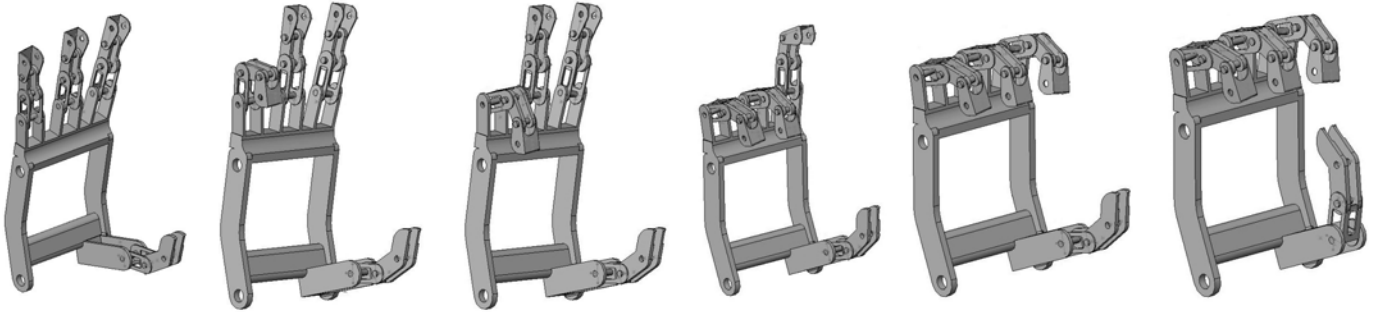


Fig. 7. Movimientos posibles de la mano: 3GdL por dedo y 2 GdL en el pulgar.

V. PROTOTIPO

La estructura mecánica de la fig. 6 es accionada por un grupo de micro-bobinas en AMF de características descritas en la Sección 3. Note en la fig. 4(a), que la micro-bobina es perfectamente capaz de desarrollar fuerzas apropiadas para accionar las uniones de los dedos de 35 g de masa cuando las micro-bobinas se calientan a su fase austenita. Más aún, las micro-bobinas son muy pequeñas por lo que pueden integrarse dentro de la estructura de la mano. No es necesario contar con un módulo externo de accionamiento.

Las micro-bobinas en AMF fueron dispuestas en configuración antagonista. Como se puede apreciar en la fig. 5, el principio antagonista consiste en calentar alternadamente un elemento en AMF, de tal forma la fase de austenita de uno produzca una fuerza y un desplazamiento sobre el otro que se encuentra en fase martensita. Para retener la posición activa, el actuador integra estructuras bi-estables.

Muchos mecanismos exhiben bi-estabilidad: switches o interruptores, cierres, bisagras, tapas de botellas de shampoo, soportes de bicicletas, cintas métricas, plumas retráctables, etc. Una solución magnética pareció adecuada para incorporarse en la mano robótica.

El principio de operación se muestra en la fig. 8: (1) en la posición inicial, ambas micro-bobinas se encuentran en fase martensita. La posición se retiene por la atracción magnética de un par de imanes (Fig. 8(a)). (2) Cuando la

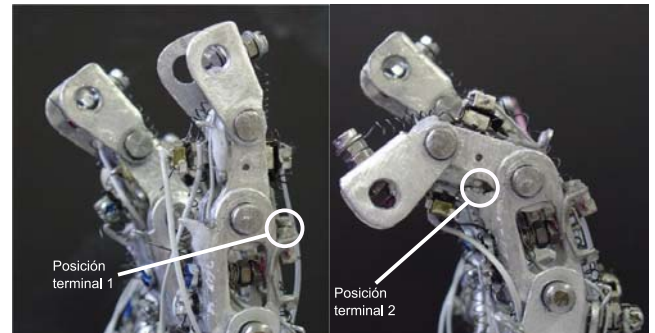


Fig. 8. Actuador AMF en configuración antagonista. Cuando una micro-bobina en AMF se calienta, se comprime estirando la articulación en una dirección. Cuando la otra micro-bobina se calienta, mueve la articulación en sentido contrario. Las posiciones terminales se aseguran por la atracción magnética de un par de imanes.

corriente eléctrica fluye a través de uno de los AMF, su forma memorizada comprimida de austenita desbloquea los imanes, estira el segundo AMF en fase martensita y mueve la articulación hasta que ésta alcanza la otra posición terminal. Otro par de imanes se atraen asegurando la posición por lo que no se requiere aplicar energía hasta que se requiera mover de nuevo la articulación (Fig. 8 (b)).

El primer prototipo de la mano robótica se muestra en la fig. 9. Se sabe que los AMF no pueden soldarse, por lo que tuvieron que asegurarse mecánicamente a través de uniones de aluminio. Esta solución facilita la tarea de reemplazo en caso de falla puesto que las micro-bobinas simplemente se desatornillan de la unión y se retiran. Un total de 6 micro-bobinas en AMF fueron colocadas en cada dedo, mientras que el

pulgar posee solo 4. La mano completa contiene 22 micro-bobinas en AMF u 11 pares antagonistas. Cada micro-bobina en AMF esta conectada a un par de cables para ser controlarlos eléctricamente desde un dispositivo electrónico. Se incluyen guías para los cables a lo largo de la estructura para evitar el riesgo de que entren en contacto y generen un corto circuito o atasco.

La fig. 10(a) muestra el diagrama esquemático del sistema electrónico empleado para controlar la mano robótica. El sistema consiste de una interfaz de usuario que permite elegir las articulaciones a mover al dar un click con el ratón sobre ellas (Fig. 10(b)). Empleando el protocolo RS232, la computadora transmite esta información a un módulo electrónico donde un controlador interpreta el comando y coloca la articulación en la posición correspondiente.

Algunos ejemplos de manipulación y prensión llevados a cabo con este prototipo se muestran en la fig. 11. Los objetos en las figuras 11(a)-(c) pueden ser clasificados como objetos “grandes” o manipulación de potencia como define [16] mientras que la fig. 11(d) se puede clasificar como un objeto “pequeño” o manipulación de precisión. Trabajo actual se enfoca en manipulación de precisión para asistir al humano con objetos difíciles de manipular.

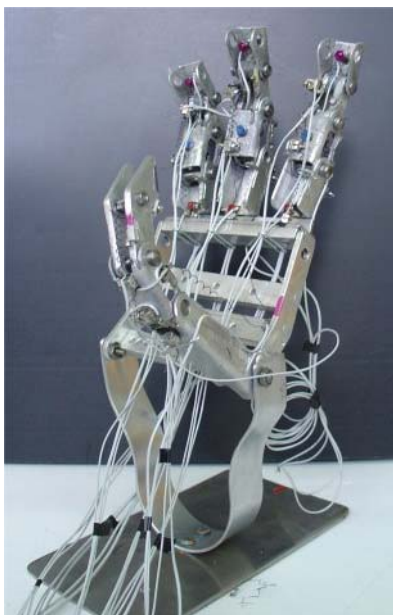
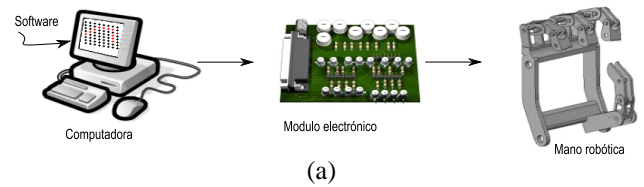


Fig. 9. Prototipo de mano robótica con materiales AMF.



(a)



(b)

Fig. 10. (a) Sistema electrónico de control de la mano robótica. (b) Interfaz gráfica del usuario: las articulaciones a activar se indican en rojo.

VI. CONCLUSIÓN

Este artículo ha presentado tanto el diseño como la implementación de una mano robótica antropomórfica de 4 dedos creada para la manipulación y la prensión de objetos para diferentes aplicaciones. El prototipo propuesto emplea actuadores no convencionales como es el caso de los AMF que permiten un desempeño eficiente, compacto, ligero y económico.

Una micro-bobina en AMF controlada eléctricamente se ha propuesto como actuador para accionar las articulaciones de la mano. Este actuador emplea un arreglo antagonista de micro-bobinas de NiTi cuyo comportamiento termomecánico ha sido caracterizado. Para incrementar la eficiencia, asegurar la repetibilidad de movimiento y disminuir la disipación de potencia, este micro-actuador integra estructuras magnéticas bi-estables.

El primero prototipo desarrollado tiene 3 GdL por dedo y 2 GdL en el pulgar para un total de 11 GdL. La mano integra 11 actuadores

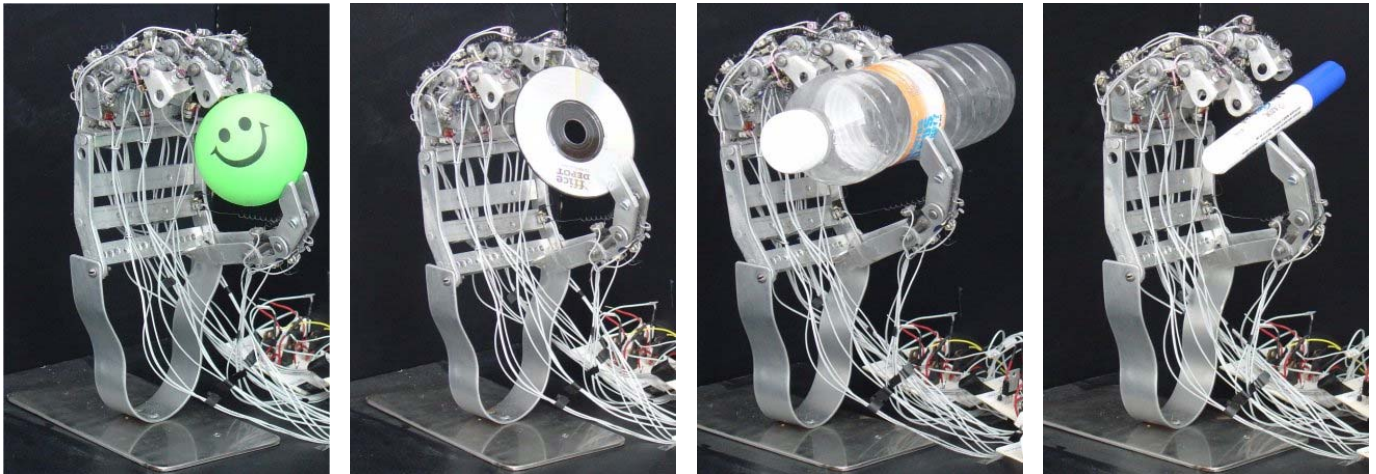


Fig. 11. Algunos ejemplos de manipulación de potencia y de precisión:
(a) esfera, (b) disco, (c) cilindro “grande” y (d) cilindro “pequeño”.

antagonistas en AMF dentro de la estructura y es capaz de sujetar diversos objetos.

Experimentos en curso comprenden la manipulación de precisión. Las perspectivas a futuro de esta investigación incluyen la telemanipulación basada en señales mio-eléctricas. Dichas señales se medirán desde los músculos en el antebrazo del operador para detectar el movimiento en los dedos. Se espera que el prototipo se mueva de igual forma.

VII. REFERENCIAS

- [1] I. Yamano and T. Maeno, “Five-fingered robot hand using ultrasonic motors and elastic elements”, Proc. IEEE International Conference on Robotics and Automation, pp 2684-2689, 2005.
- [2] K. DeLaurentis, C. Mavroidis and C. Pfeiffer, “Development of a shape memory alloy actuated robotic hand”, Proc. 7th International Conference on New Actuators, pp 281-285, 2000.
- [3] J. Butterfass, G. Hirzinger, S. Knoch and H. Liu, “DLR’s multisensory articulated hand, part I: hard and software architecture”, Proc. IEEE International Conference on Robotics and Automation, pp 2081-2086, 1998.
- [4] J. Zhao, Z. Xie, L. Jiang, H. Cai, H. Liu and G. Hirzinger, “A five-fingered underactuated prosthetic hand control scheme”, Proc. IEEE International Conference on Biomedical Robotics and Biomechatronics, pp 995-1000, 2006.
- [5] H. Kawasaki, T. Komatsu and K. Uchiyama, “Dexterous anthropomorphic robot hand with distributed tactile sensor: Gifu hand II”, IEEE/ASME Transactions on Mechatronics, 7(3), pp 296-303, 2002.
- [6] Science Museum/Science and Society. Information available at: <http://www.scienceandsociety.co.uk/>
- [7] Touch Bionics Inc. Information available at: <http://www.touchbionics.com/>
- [8] K. Ikuta, “Micro/Miniature shape memory alloy actuator”, Proc. IEEE International Conference on Robotics and Automation, pp 2156-2161, 1990.
- [9] C. Mavroidis, C. Pfeiffer and M. Mosley, “Conventional actuators, shape memory alloys, and electrorheological fluids”, Automation, Miniature Robotics and Sensors, pp 189-214, 2000.

- [10] N. Chaillet, “*Contribution à la modélisation, la commande et la réalisation de microactionneurs et de composants pour la microrobotique*”. HDR Thesis, Franche-Compte University, 1999.
- [11] R. Velazquez, M. Hafez, E. Pissaloux and J. Szewczyk, “*A computational-experimental thermomechanical study of shape memory alloy microcoils and its application to the design of actuators*”, *Journal of Computational and Theoretical Nanoscience*, 3(4), pp 538-550, 2006.
- [12] J. Gill, K. Ho and G. Carman, “*Three-dimensional thin-film shape memory alloy microactuator with two-way effect*”, *Journal of Microelectromechanical Systems*, 11(1), pp 68-77, 2002.
- [13] T. Waram, “*Actuator design using shape memory alloys*”, T.C. Waram Publishing, Ontario, 1993.
- [14] R. Velazquez, “*Contribution à la conception et à la réalisation d’interfaces tactiles portables pour les déficients visuels*”, PhD Thesis, Paris 6 University, 2006.
- [15] M. Hafez and M. Benali-Khoudja, “*3D tactile rendering based on bi (multi) stable SMA monolithic systems*”, *Proc. of IEEE International Symposium on Micro-mechatronics and Human Science*, pp 93-98, 2004.
- [16] J. Napier, “*The prehensile movements of the human hand*”, *Journal of Bone and Joint Surgery*, 38-B(4), pp 902-913, 1956.

DISEÑO Y CONSTRUCCIÓN DE LOS SISTEMAS DE POTENCIA Y CONTROL PARA UN ROBOT INDUSTRIAL TIPO CARTESIANO

Hernández Treviño Josué, M. A. Vargas, S. Vergara, O. Félix, F. Reyes.

Facultad de Ciencias de la Electrónica, Área de Robótica y Control
Benemérita Universidad Autónoma de Puebla
Av. San Claudio y 18 sur C.U. edificio 109 A. Puebla, Puebla. C.P. 72570, México.
Correo electrónico: josue_ht1@ece.buap.com

RESUMEN

En la actualidad existen robots industriales de muchos tipos con tecnología que desconocemos, ya que no se informa a detalle la forma del diseño y construcción por parte de los investigadores que desarrollan este tipo de avances, son los llamados secretos industriales. Actualmente el crecimiento de un país se ve reflejado en la tecnología que éste genera y particularmente en el ámbito industrial, por lo que es muy importante crear tecnología propia con la cual se puedan automatizar procesos. Dichos procesos elevan la calidad de los productos, permitiendo así entrar en el terreno competitivo que el mercado mundial demanda. Lo que proponemos es desarrollar el diseño y construcción de los sistemas de potencia y control para un robot industrial tipo cartesiano, con lo cual se pretende generar las herramientas necesarias para la manipulación de diversos procesos tanto industriales como del tipo didácticos mediante el manejo de todas las partes que componen al robot. Un robot con todas estas características permitirá desarrollar estrategias de control, así como lograr la automatización de procesos como puede ser: corte, pintura, paletizado, fresado, soldadura, etc. El proceso de diseño y construcción representa un gran desarrollo de hardware, firmware y software. En este artículo mostramos una descripción del sistema de potencia y control.

Palabras clave: Diseño y construcción de sistemas de potencia y control, tarjeta PCI, instrumentación, automatización, firmware.

I. INTRODUCCIÓN

Es bien sabido que en la actualidad la robótica es un elemento clave para el desarrollo industrial de un país, casi todos los procesos automatizados utilizan algún tipo de robot. Por lo que representa el desarrollo de la robótica, es de gran importancia crear tecnología propia que nos acerque a los avances que se desarrollan en los países tecnológicamente avanzados, por esta razón se propone el estudio, desarrollo y construcción de los sistemas de potencia y control de un robot industrial tipo cartesiano, el cual tendrá enormes aplicaciones con fines tanto académicos como en aplicaciones en la industria en procesos automatizados. En la actualidad, la tecnología avanza rápidamente y a cada momento nos sorprende con nuevas técnicas de diseño y medición, estamos en el tiempo de optimizar todas las herramientas que nos permiten continuar avanzando tecnológicamente.

Una de las herramientas más utilizadas por sus características es la computadora, que está en constante desarrollo y por tanto poco a poco incrementa su velocidad, capacidad de almacenamiento, etc. Por tanto es una herramienta que nos podría abrir la puerta para desarrollar en ella muchas otras aplicaciones. Por lo anterior consideramos que un robot industrial controlado por computadora y desarrollado con la tecnología que tenemos a la mano permitirá el crecimiento económico y tecnológico, además de reducir la diferencia en cuanto al retraso tecnológico que en la actualidad presenta el país, con esto se puede

ayudar tanto a investigadores como a estudiantes.

Se pretende crear una herramienta que permita la automatización de procesos industriales en forma económica desarrollando tecnología propia. Con los primeros estudios desarrollados hemos diseñado la etapa de potencia y la forma de manejar motores trifásicos de inducción, además se cuenta con una tarjeta PCI que funciona como interfaz entre la computadora y el robot, la idea es ir desarrollando cada uno de los pasos necesarios para el adecuado control del robot. En las siguientes secciones se describe el funcionamiento y partes que constituyen nuestro robot cartesiano.

II. DISEÑO DEL CONTROL

Para lograr el control de un robot industrial en toda su estructura, es necesario ser capaz de desarrollar la arquitectura de estos robots y esto representa trabajos de un alto grado de dificultad. Este tipo de robots utilizan tres sistemas: (1) electromecánico; referente a la estructura mecánica y actuadores, (2) electrónico: referente al tratamiento de las señales y (3) software de control: referente a las rutinas que seguirá el robot. Cada uno de estos sistemas se acopla adecuadamente con lo cual se logra el control del robot. En este artículo se describe el diseño de cada una de las etapas del robot, donde se cuenta con el diseño y construcción de: la etapa de potencia, la lógica para controlar motores de inducción de corriente alterna, diseño de fuentes de alimentación, sistema sensorial, la lógica del accionamiento, comunicación del sistema con la computadora mediante el bus PCI utilizando un FPGA de la marca ALTERA.

El robot posee 3 grados de libertad. El control se efectúa desde la computadora cerrando el lazo mediante el monitoreo de las señales por medio del software LabView 8.0 como interfaz de usuario. Las fuentes necesarias para controlar los motores trifásicos del robot serán de alta corriente y voltaje variable de un rango de 4-100volts a 20 amperes.

El diagrama a bloques del sistema completo se muestra a continuación.

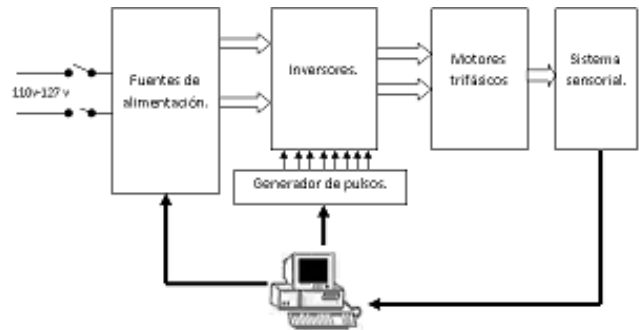


Fig. 1 Diagrama a bloques general.

La primera parte del diagrama a bloques corresponde a la fuente controlada por computadora conectada a la entrada de los circuitos inversores, necesarios para trabajar con los motores del robot. Para lograr enviar los pulsos adecuados que generen las fases necesarias para mover a los motores trifásicos se utiliza un FPGA y el bus PCI de la computadora, esto se representa con el bloque denominado generador de pulsos. La fuente de alimentación diseñada cuenta con circuitos electrónicos para la comunicación controlada desde la computadora a través del bus PCI, la razón de utilizar el bus PCI es por las prestaciones de velocidad de procesamiento con la que se permite trabajar, ya que se desea una respuesta rápida y dicho bus es capaz de ejecutar una instrucción cada 30 nanosegundos. Después, en la etapa en la que se envían los pulsos hacia los circuitos de potencia o "inversores" se controla la frecuencia con la que trabajan los motores, una vez que analizamos esto, nos encontramos con la etapa de los sensores de posición, que se encargan de informar a través de niveles de voltaje la posición de cada motor a partir del número de las vueltas que éste a dado, se da la posición por niveles de voltaje debido a que los sensores de posición son analógicos (resolvers), por lo cual es necesario usar circuitos convertidores de resolver/digital (R/D) para comunicarse con la computadora.

La etapa del FPGA, en donde se concentra el accionamiento y herramientas de control del robot consta de varios módulos, en primer lugar tenemos las secuencias necesarias para mover a los motores trifásicos en ambas direcciones, la variación de la

frecuencia para cada motor, el acoplamiento de las señales de posición provenientes de los resolvers, el control digital del voltaje proveniente de las fuentes que alimentan a los motores, además la tarjeta PCI junto con el FPGA sirven como interfaz entre el control a utilizar y los efectos que se produzcan en el robot ya que desde aquí se monitorea el funcionamiento del robot.

III. ACTUADORES

El motor de inducción de corriente alterna, tiene como principal inconveniente la complejidad del convertidor electrónico necesario para variar su velocidad. La aparición en un principio del Tiristor (SCR) y recientemente del transistor IGBT (Insulated Gate Bipolar Transistor) creado para trabajar en conmutación de potencia a altas tensiones, intensidades y frecuencias, a facilitado la regulación.

Los motores de inducción pueden ser monofásicos o trifásicos y dentro de estos últimos, pueden ser a su vez: de anillos rozantes o de jaula de ardilla. Para que un motor de inducción con jaula de ardilla produzca par o haga girar su carga, los circuitos de barras de rotor deben sentir un cambio de flujo respecto al tiempo. Básicamente el funcionamiento es el siguiente: al aplicar un sistema trifásico de tensiones de frecuencia al devanado estático, se produce una fuerza electromotriz (f.e.m) móvil distribuida senoidalmente en el entrehierro, que da origen a un campo magnético giratorio, esto se fundamenta en el teorema de ferraris.

El procedimiento para generar las fases que demanda el motor trifásico para su funcionamiento, se basa en estudiar el comportamiento de un sistema trifásico, el cual consiste en fuentes del mismo valor de voltaje pero desfasadas 120° entre sí. Mediante el circuito inversor se logra este hecho de la siguiente forma: a partir de una fuente de DC y mediante interruptores de potencia se generan cada una de las fases que se conectarán directamente al estator del motor (ver figura 2).

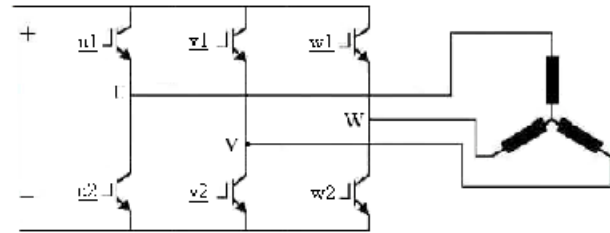


Fig. 2 Inversor trifásico conectado en estrella.

Ahora bien, como ya se mencionó anteriormente se definen las secuencias con las cuales trabaja el motor trifásico basados a una forma de señales idealizadas (ver figura 3) con lo cual se estudian los niveles de voltaje que se requieren en el estator del motor de inducción.

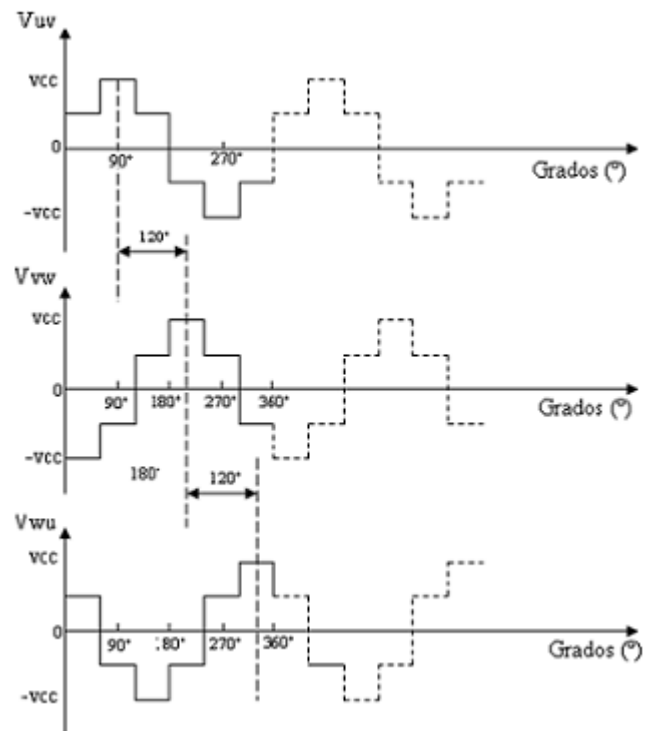


Fig. 3 Formas de onda idealizadas para el circuito de disparo.

Del estudio de los niveles de voltaje y de todas las posibles secuencias se obtienen 6 diferentes combinaciones en un determinado orden, lo que provoca el giro en 360° del motor en una dirección. Para lograr mover al motor en dirección contraria se intercambian dos fases del motor a nivel de firmware, es decir se dicta dentro del FPGA la

forma de intercambiar dos fases modificando las secuencias y el orden en que se transmiten mediante el firmware. El FPGA a utilizar es el llamado CICLONE II, su matrícula es EP2C8Q208C7, es un dispositivo muy robusto en el que se puede implementar hasta un DSP, cuenta con 208 terminales, el modo de programar dichos dispositivos es por medio del software de diseño creado por ALTERA Quartus II versión 5.0.

Las combinaciones encontradas para mover al motor en una dirección se muestran en la tabla 1, las cuales siguen a las formas de onda idealizadas mostradas en la figura 2. Los niveles de voltaje para los devanados del estator se analizan estudiando el flujo de corriente a través de cada uno de los devanados de acuerdo a las combinaciones posibles que el circuito inversor puede proporcionar. La figura 4 nos muestra la forma de los devanados del estator y las ecuaciones (1), (2) y (3) definen los niveles de voltaje. Estos niveles de voltaje encontrados se muestran en la tabla 1.

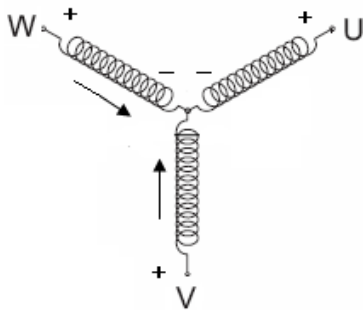


Fig. 4 Devanados y polaridades del motor de inducción.

El funcionamiento cíclico del circuito de disparo se divide en seis intervalos de igual duración de tiempo. Para encontrar las secuencias por las cuales se logre mover al motor se establecen las diferencias de potencial de cada devanado de acuerdo con la figura 4, de donde se obtienen las ecuaciones que describen el nivel de voltaje de cada devanado, de acuerdo a las seis combinaciones que se pueden presentar, dichas ecuaciones son:

$$V_{UV} = V_U - V_V \tag{1}$$

$$V_{VW} = V_V - V_W \tag{2}$$

$$V_{WU} = V_W - V_U \tag{3}$$

Las combinaciones necesarias para el giro del motor en una dirección se muestran a continuación en la tabla 1, donde se indican los niveles de voltaje encontrados para cada combinación de interruptores o secuencias del circuito de disparo.

Tabla 1. Combinaciones necesarias para el giro.									
Secuencia	u1	u2	v1	v2	w1	w2	Vuv	Vvw	Vwu
1	0	0	0	1	1	0	Vcc/2	-Vcc	Vcc/2
2	1	0	0	1	0	0	Vcc	-Vcc/2	-Vcc/2
3	1	0	0	0	0	1	Vcc/2	Vcc/2	-Vcc
4	0	0	1	0	0	1	-Vcc/2	Vcc	-Vcc/2
5	0	1	1	0	0	0	-Vcc	Vcc/2	Vcc/2
6	0	1	0	0	1	0	-Vcc/2	-Vcc/2	Vcc

IV. VARIACIÓN DE LA VELOCIDAD EN EL MOTOR ASÍNCRONO

Poder variar la velocidad de un proceso o de un sistema, es una necesidad que se plantea obligatoriamente en la mayoría de los sectores de la industria, la aparición del control electrónico de los motores eléctricos, para variar y regular la velocidad, ha aportado a los procesos industriales grandes ventajas, aumentando sus posibilidades y prestaciones, además reduciendo su mantenimiento de energía y consumo entre otras [4].

Los motores de ca. no se adaptan tan bien a las aplicaciones con velocidad variable como los motores de cd, por que no se puede controlar su velocidad en forma satisfactoria con una variación sencilla de voltaje de suministro. Esto es, un motor de inducción de ca. funcionando a un voltaje reducido no es capaz de mantener una velocidad de eje razonablemente estable en presencia de pequeños cambios en demanda de par, impuesto por la carga mecánica [3].

Se puede obtener un control satisfactorio de la velocidad de un motor de inducción de ca. sólo cuando se varía la frecuencia de suministro mientras se varía el voltaje de suministro. Por lo anterior se debe variar la frecuencia al mismo tiempo que se varia el voltaje, de ahí la necesidad de construir fuentes de alta corriente y voltaje variable para manipular el voltaje aplicado, además se debe considerar la rapidez de conmutación de la fuente para lograr un adecuado control del robot, mientras que la frecuencia se varia controlando el tiempo en que se envían los pulsos al inversor encargado de generar las fases del motor, de esta forma controlamos el par de los motores trifásicos.

V. CONTROL POR VARIACIÓN DE FRECUENCIA

Se basa en el control de la velocidad a través de la variación de la frecuencia de la tensión de alimentación. Este método permite una regulación de velocidad con un buen rendimiento en una amplia gama de frecuencias de trabajo, que abarca desde valores muy superiores a la velocidad nominal del motor hasta velocidad nula y en ambos sentidos, todo ello manteniendo un elevado par motor [4].

Por lo que respecta a los modernos equipos de regulación que existen actualmente en el mercado, las técnicas de control que utilizan principalmente se pueden resumir en:

- Control V / f : Tensión/frecuencia.
- Control I / f : Intensidad/frecuencia.
- Control vectorial de flujo magnético.

Las condiciones óptimas de funcionamiento de un motor de inducción se consiguen manteniendo constante el flujo eficaz en el entrehierro, sea cual sea la frecuencia y cualquier par transmitido. Por lo que si se mantiene constante el flujo (ϕ), el par motor también lo será. Este hecho se consigue manteniendo constante la relación: V_1 / f_1 , de tal forma que:

$$\frac{V_1}{f_1} = K_1 = cte. \tag{4}$$

Del estudio del motor asíncrono o motor de inducción se llega a la siguiente relación:

$$M_i = V_1^2 \tag{5}$$

$$\phi_{max} = \frac{V_1}{\pi\sqrt{2}N_1f_1} \tag{6}$$

donde:

V_1 es el voltaje aplicado en rms.

f_1 es la frecuencia de la fuente de alimentación.

N_1 es el número de vueltas de los devanados del estator.

M_i es el par interno del motor.

VI. FUENTE DE ALTA CORRIENTE CONTROLADA POR COMPUTADORA

Para llevar a cabo un control satisfactorio de los motores de inducción trifásicos es necesario contar con fuentes de alta corriente y voltaje variable que responda a gran velocidad, las cuales no están disponibles en el mercado nacional y sus costos son muy grandes fuera del país, por lo que se tiene la necesidad de construirlas. El diagrama a bloques de la fuente es como se ilustra en la figura 5, donde se cuenta con un transformador con cuatro secundarios independientes a 18 V cada uno y una corriente máxima de 20 A, posee un circuito de selección de los secundarios para lograr utilizar el máximo voltaje de la fuente, es decir, el voltaje que puede ofrecer la fuente es de alrededor de 4 - 100 V a 20 A.

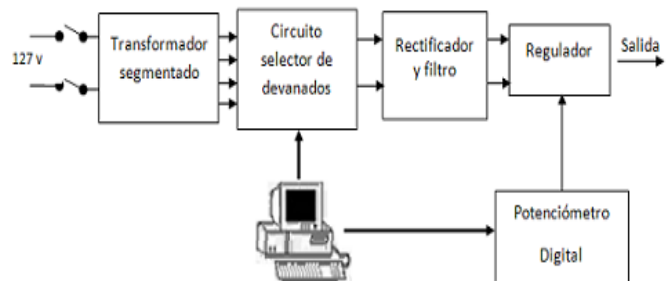


Fig. 5 Diagrama a bloques de la fuente controlada.

VII. GENERADOR DE PULSOS

A continuación se muestran los diferentes diagramas de flujo que gobiernan dentro del circuito de disparo del generador de pulsos, en dichos diagramas se muestra la inversión del giro del motor mediante el intercambio de dos de sus fases, se indica para cada instante de tiempo la nueva secuencia en ambos sentidos de dirección utilizando los saltos adecuados.

El siguiente diagrama de flujo (figura 6) representa la forma de arranque para un motor trifásico utilizado dentro del firmware, el cual inicia con una secuencia de alto total y carga un valor en la variable N para realizar la siguiente secuencia que provoca el movimiento del motor en una dirección.

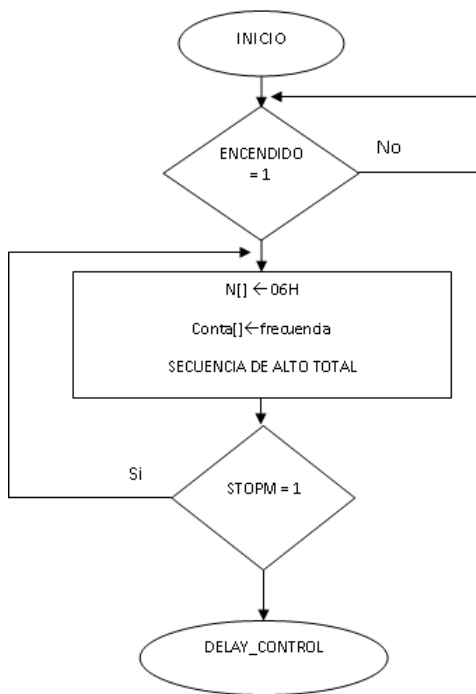


Fig. 6 Diagrama de flujo para el arranque del motor de inducción.

A continuación se muestra la forma que tienen los diferentes estados necesarios para mover al motor trifásico en una dirección con opción a cambiarla, para lo cual nos apoyamos con diagramas de flujo. Todos los estados que permiten el movimiento en una dirección son

similares solo cambian los valores de la variable N y las secuencias en cada estado son diferentes (no se presentan todos los diagramas de flujo), en el orden que muestra la tabla 1. Si se desea invertir el giro del motor se deben buscar las nuevas secuencias que permitan intercambiar dos de las fases, la tabla 2 muestra este nuevo orden encontrado bajo el estudio de las fases.

Tabla 2. Combinaciones necesarias para invertir dos fases

Secuencia	u1	u2	v1	v2	w1	w2
1	1	0	0	0	0	1
2	1	0	0	1	0	0
3	0	0	0	1	1	0
4	0	1	0	0	1	0
5	0	1	1	0	0	0
6	0	0	1	0	0	1

La figura 7 muestra la lógica utilizada para lograr el desplazamiento del motor en una dirección con opción a invertir el giro.

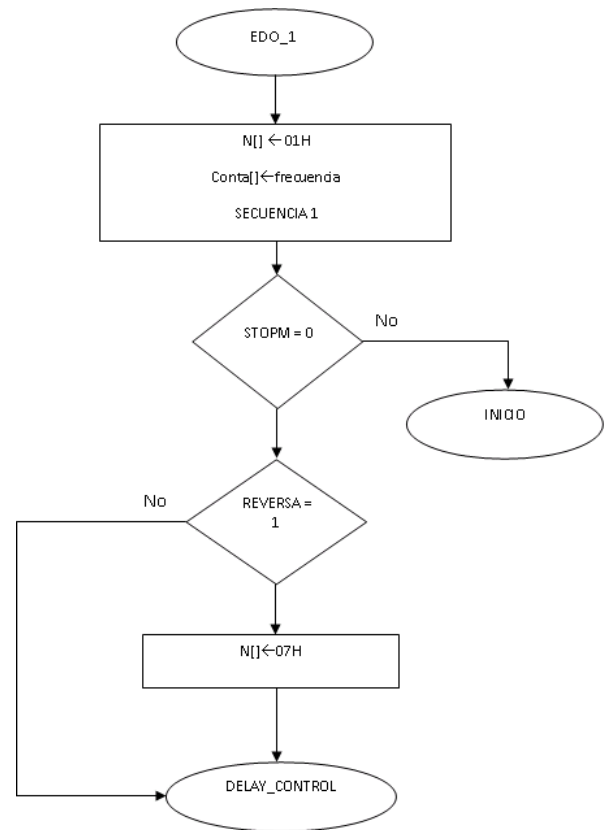


Fig 7. Diagrama de flujo para el movimiento en una dirección durante la primera secuencia.

Para lograr mover al motor en el sentido inverso se requieren de más estados, los cuales de igual forma incluyen la posibilidad de cambiar de dirección de giro en cualquier momento (no se presentan todos los diagramas de flujo), el cambio que se presenta con el diagrama anterior es la forma de activar la señal de reversa ya que en este caso se activará con un 0 lógico. A continuación se muestra el diagrama de flujo base para estos estados, de los cuales solo cambia el valor de la variable N y la secuencia en cada estado, dichas secuencias y el orden en que se deben presentar se mencionan en la tabla 2.

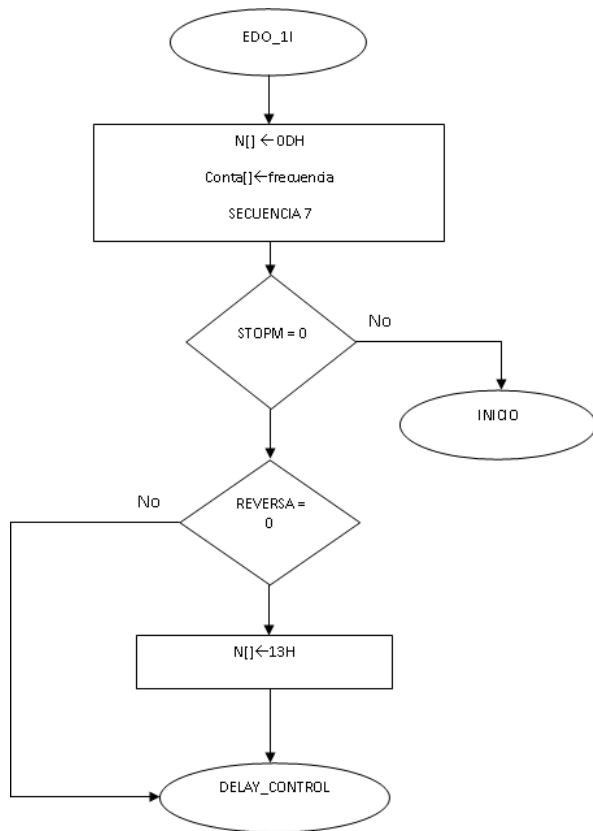


Fig 8. Diagrama de flujo para el movimiento dirección contraria durante la primera secuencia.

Para el circuito de disparo del generador de pulsos se desarrolló el código descrito por los diagramas de flujo anteriormente expuestos dentro del programa Quartus II versión 5.0.

VIII. PROTOCOLO PCI

Tenemos el protocolo PCI, sabemos que se les llama protocolo de red o protocolo de comunicación al conjunto de reglas que controlan la secuencia de mensajes que ocurren durante una comunicación entre entidades que forman una red. En este contexto, las entidades de las cuales se habla son programas de computadora o automatismos de otro tipo, tales y como dispositivos electrónicos capaces de interactuar en una red. Así es que por medio del protocolo de comunicaciones PCI que será programado en el FPGA lograremos comunicación con la computadora mediante el programa LabView 8.0.

IX. TARJETA ELECTRÓNICA PCI

Se cuenta con una tarjeta electrónica diseñada y desarrollada en la Facultad de Ciencias de la Electrónica, Departamento de Robótica y Control como interfaz entre la computadora y el sistema completo, es decir, dentro de esta tarjeta especial se coloca el FPGA, donde se indican las entradas y salidas del robot, dentro del firmware que se carga al FPGA se establece el protocolo de comunicación del bus PCI de la computadora y se interconecta en la tarjeta. La tarjeta electrónica se muestra a continuación.



Fig 9. Tarjeta Electrónica en 3D.

Con la tarjeta electrónica se comunican los siguientes sistemas: fuentes de alimentación, control por computadora, sistema sensorial, y circuito de disparo para los inversores.

X. INVERSOR

Del inversor propuesto y mencionado anteriormente se cuenta con el prototipo armado en una tablilla de conexiones, donde se incluye la forma de conectarse a la tarjeta electrónica del bus PCI. La figura 10 muestra la tablilla de conexiones con el inversor implementado con IGBTs.

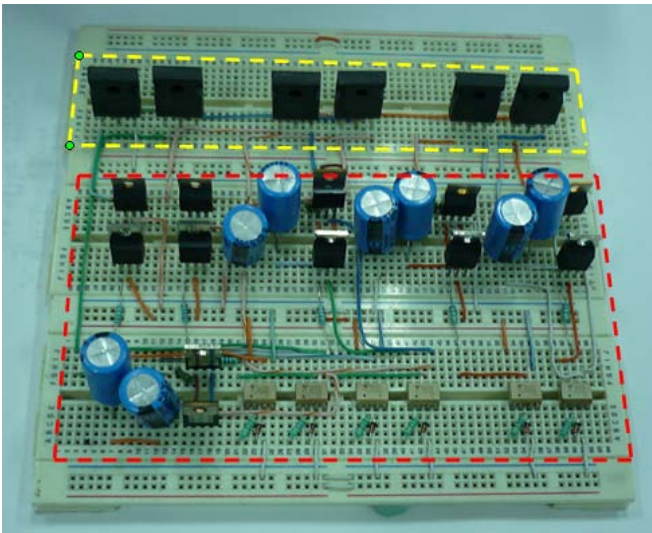


Fig 10. Inversor en tablilla de conexiones.

De la figura 10, dentro de recuadro rojo se encuentra la etapa de acoplamiento de las señales provenientes de la tarjeta electrónica, es decir, las señales que envía la tarjeta se adaptan y aíslan de la etapa de potencia mediante circuitos optoacopladores y con el apoyo de transistores se depuran las señales con el fin de activar los interruptores (IGBTs) de forma adecuada.

Del recuadro en amarillo se tiene la etapa de potencia implementada con IGBTs los cuales suministran las fases necesarias para mover a los motores de inducción trifásicos.

Finalmente se menciona la infraestructura con la que se cuenta, integrada por la parte mecánica y actuadores, es decir, se cuenta con la mecánica robot industrial tipo cartesiano y los motores trifásicos que lo mueven como se muestra en la figura 11. La contribución del trabajo esta en el diseño y construcción de los dispositivos de potencia y control que permitan mover y realizar

un proceso con el robot y que además con ello se logre tener total control en las distintas etapas del robot de forma independiente, a fin de lograr varios y diversos procesos automatizados donde se apliquen las teorías de control.



Fig 11. Robot industrial tipo cartesiano a controlar.

XI. CONCLUSIONES

Se ha presentado la forma en que se pretende generar el movimiento de los motores trifásicos que componen al robot industrial tipo cartesiano, así como la forma de controlar el par de dichos motores, además se presentó de forma general, la forma de comunicar al robot con la computadora por medio del bus PCI y el porque de la necesidad de construir fuentes de alta corriente y voltaje variable para el adecuado control de los motores mediante circuitos inversores.

XII. REFERENCIAS

- [1] B Antonio, PL Felipe, A Rafael, *Fundamentos de robótica*, Primera edición, McGraw Hill, España 1997, pp. 327
- [2] Timothy J. Maloney, *Electrónica Industrial Moderna*, Quinta edición, Pearson Education, México 2006, pp. 972
- [3] Velasco B. Joaquín, Oriol V. Miguel, Otero A. José., *Sistemas Electrotécnicos de Potencia, Electrónica de regulación y control de*

potencia, Primera edición, Paraninfo, España 1998, pp. 143.

[4] Sanz Feito Javier, *Máquinas Eléctricas*, Primera edición, Prentice Hall, Madrid 2002, pp. 491

[5] Merino Azcárraga José María, *Convertidores de frecuencia para motores de corriente alterna. Funcionamiento y Aplicaciones*”, Primera edición, McGraw Hill, Madrid 1998, pp. 408

[6] Kelly Rafael, Santibáñez Víctor, *Control de Movimiento de Robots Manipuladores*, Primera edición, Prentice Hall , 2003, pp. 323

Automatización de la Facultad de Ciencias de la Electrónica de la BUAP, perteneciente al grupo de robótica de la misma institución, e-mail: freyes@ece.buap.mx

XIII. AUTORES

Lic. Josué Hernández Treviño obtuvo su título de Licenciatura en Ciencias de la Electrónica con especialidad en Robótica y Control por la Benemérita Universidad Autónoma de Puebla (BUAP).

Dra. A. Vargas Treviño es Doctora en Optoelectrónica por la BUAP, profesor investigador de la Maestría en Ciencias de la Electrónica Opción en Automatización de la Facultad de Ciencias de la Electrónica de la BUAP, perteneciente al grupo de robótica de la misma institución, e-mail: mavargas@ece.buap.mx

Dr. S. Vergara Limón es Doctor en Optoelectrónica por la BUAP, profesor investigador de la Maestría en Ciencias de la Electrónica Opción en Automatización de la Facultad de Ciencias de la Electrónica de la BUAP, perteneciente al grupo de robótica de la misma institución, e-mail: svergara@ece.buap.mx

Dra. O. Felix Beltran es Doctora en Física por el Instituto de Física de la Benemérita Universidad Autónoma de Puebla (BUAP), perteneciente al grupo de Robótica de la misma institución, e-mail: olga_flix@yahoo.com.mx

Dr. F. Reyes es Doctor en Ciencias de la Electrónica y Telecomunicaciones por el Centro de Investigación Científica y de Estudios Superiores de Ensenada (CICESE), profesor investigador de la Maestría en Ciencias de la Electrónica Opción en

SILLA DE RUEDAS ROBÓTICA PARA PERSONAS CON DISCAPACIDAD MOTORA SEVERA

Camacho Montoya Abraham, Romero Gutiérrez Benjamín, Martínez Ramírez Juan-Salvador, Varona Salazar Jorge, Velázquez Guerrero Ramiro

Mecatrónica y Control de Sistemas - Universidad Panamericana,
Josemaría Escrivá de Balaguer 101, Fracc. Rústicos Calpulli,
20290, Aguascalientes, Ags., México

Tel: (449) 9106200, web: www.robotica-up.org, correo electrónico: rvelazquez@ags.up.mx

RESUMEN

Este artículo presenta el diseño, implementación y plataforma experimental de una silla de ruedas robótica a bajo costo. El primer prototipo desarrollado es una silla de ruedas eléctrica tres veces más económica que las disponibles comercialmente y de funcionamiento tanto manual como autónomo. La operación manual depende de un joystick y esta destinada a personas que, a pesar de su discapacidad motora, son capaces de maniobrar la silla de ruedas. En operación autónoma, la silla depende de un sistema de visión por computadora para navegar de manera autónoma a través de corredores, pasillos y diversos entornos exteriores/interiores y está destinado para personas con severos problemas motrices como los tetraplégicos. Además de ser un dispositivo accesible económicamente, la aplicación actual es una plataforma guía para museos: la silla de ruedas guía al usuario a través de los espacios y salas del museo mientras reproduce información por audio.

Palabras clave: navegación autónoma, visión por computadora, detección de bordes, silla de ruedas eléctrica (SRE), asistencia a la movilidad.

I. INTRODUCCIÓN

Según la Organización Mundial de la Salud (OMS), entre el 7 y 10% de la población mundial sufre de alguna discapacidad física.

En Latinoamérica los discapacitados físicos son cerca de 55 millones de personas, los cuales representan el 9% de la población total [1].

Este censo indica que la discapacidad más común es la motora, seguida por la ceguera, sordera, intelectual y el habla. Tan solo los discapacitados motores alcanzan 20 millones en América Latina y se anticipa un crecimiento en esta cifra debido al envejecimiento, la longevidad y las lesiones relacionadas con accidentes (la mayoría automovilísticos y por deporte).

Sin duda, las sillas de ruedas representan una parte significativa de los dispositivos para la asistencia motora. Desde 1930, el diseño de una silla de ruedas ha sido el mismo: una estructura principal o marco, 2 ruedas traseras grandes y 2 ruedas delanteras pequeñas (Fig. 1). Este diseño ha sido utilizado para desarrollar las sillas de hoy en día aplicando pequeñas modificaciones para producirlas más ligeras, durables y confortables para las personas que las utilizan a diario.

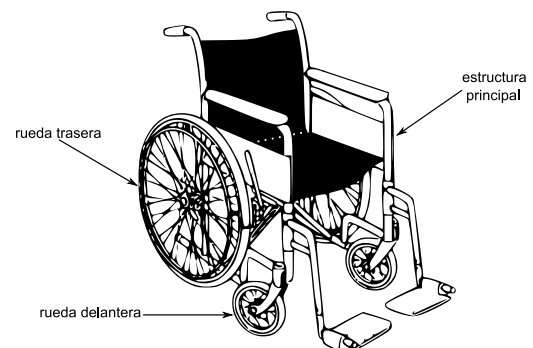


Fig. 1. Diseño clásico de una silla de ruedas estándar para los discapacitados motores.

Mientras las sillas de ruedas estándares son utilizadas principalmente por quienes tienen la fuerza necesaria para moverlas y maniobrarlas, las sillas de ruedas eléctricas (SRE) están destinadas para quienes no pueden: personas con excesivo peso corporal, mala salud, debilidad de las extremidades superiores y quienes llevan largo tiempo viviendo con la discapacidad [2].

Desafortunadamente, las SRE convencionales no son siempre suficientes para compensar las discapacidades motrices: personas que padecen de coordinación y destreza física, espasticidad cerebral, tetraplejía/paraplejía, temblores, espasmos o traumatismo craneoencefálico no pueden usar una SRE y se les prescribe una silla de ruedas estándar cuya movilidad se ve proporcionada por una persona auxiliar. En este caso, la robótica y mecatrónica podrían ser relevantes para desarrollar sillas de ruedas con cierto grado de autonomía.

Este artículo presenta el diseño, implementación y primer prototipo de una SRE que tiene como objetivo responder a 2 criterios: ofrecer una SRE verdaderamente accesible para el entorno económico de América Latina y ser flexible para el usuario: manual para aquellos que son capaces de maniobrarla y autónomas para los severamente discapacitados con la opción, por supuesto, de ser operada de manera estándar (i.e. con la propulsión de las manos) en todo momento.

El sistema propuesto está basado en una silla de ruedas comercial la cual ha sido equipada con 2 motores de DC de alta potencia y de bajo costo, y un sistema de visión para un control inteligente y una navegación autónoma.

El resto del documento está organizado de la siguiente manera: la Sección 2 presenta de manera general el diseño, implementación y plataforma experimental de una SRE a bajo costo. La Sección 3 describe sus modos de operación mientras que la Sección 4 aborda una aplicación en proceso para la SRE: plataforma guía para museos. Finalmente, la Sección 5 concluye con un resumen de los principales conceptos y perspectivas de trabajo futuro.

II. DISEÑO Y PROTOTIPO

El primer prototipo desarrollado se muestra en la fig. 2. El sistema está basado en una silla de ruedas estándar plegable de aluminio disponible comercialmente por 180 USD.

Dos motores de DC de alta potencia fueron incorporados en las ruedas traseras. Éstos son alimentados por 2 baterías de 12 V y son controlados a través de una unidad electrónica de control. La silla de ruedas es operada en modo manual por un joystick de alta sensibilidad. La silla alcanza una velocidad máxima de 10km/h y tiene 2 h de carga en las baterías.

Su costo de fabricación es de sólo 1000 USD, lo que representa un tercio del costo de la mayoría de las SRE comerciales disponibles [3]. Además, una computadora portátil y una videocámara han sido incorporadas al sistema. El costo total de este prototipo asciende a 2100 USD, lo cual representa todavía un costo menor al de una SRE comercial (3000 USD).

Dos modos de operación están disponibles para la silla de ruedas: manual o eléctrica y autónoma y, por supuesto, estándar o de propulsión con las manos en todo momento.

La fig. 3 muestra un diagrama de bloques del funcionamiento de los 2 modos de operación.

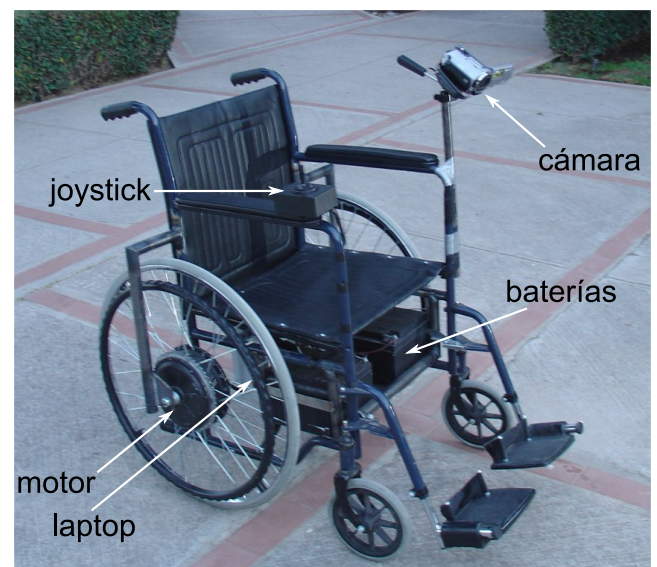


Fig. 2. Primer prototipo de SRE robótica a bajo costo.

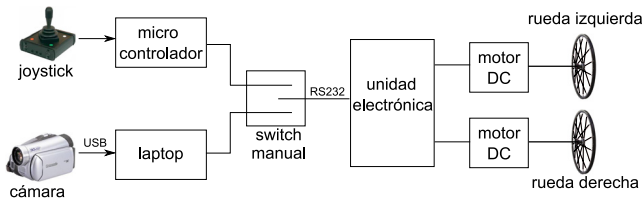


Fig. 3. Esquemático del sistema de control de la SRE.

III. MODOS DE OPERACIÓN

A. Operación manual

Un joystick es comúnmente la interfaz principal entre una persona con discapacidad motora y una SRE [4]. Éste proporciona una señal de control proporcional a la fuerza ejercida sobre la palanca para controlar la silla de ruedas en dirección y velocidad. El usuario generalmente maneja la silla de ruedas usando un joystick de 2 ejes. La velocidad depende del movimiento adelante-atrás del joystick mientras que la dirección cambia con los movimientos de derecha-izquierda.

En este modo, un micro-controlador recibe la señal analógica del joystick y la traduce a señal digital utilizando su ADC interno (Fig. 3). Al mover el joystick hacia adelante-atrás, el micro-controlador envía una señal digital a la unidad electrónica de control para tener la misma velocidad en ambos motores. Por otro lado, la magnitud de la velocidad es proporcional a la inclinación hacia delante o hacia atrás del joystick.

Si el joystick es movido hacia delante pero con cierta inclinación a la derecha-izquierda, la silla de ruedas girará dependiendo del lado al que fue movido el joystick. Al mover el joystick a la derecha o izquierda sin inclinación hacia adelante o atrás, la silla de ruedas girará sobre su propio eje (Fig. 4).

Contrariamente a la mayoría de las SRE en las cuales el usuario tiene que adaptar y moderar su fuerza ejercida sobre el joystick, el sistema permite optimizar la sensibilidad del joystick de acuerdo a las capacidades del usuario. Un

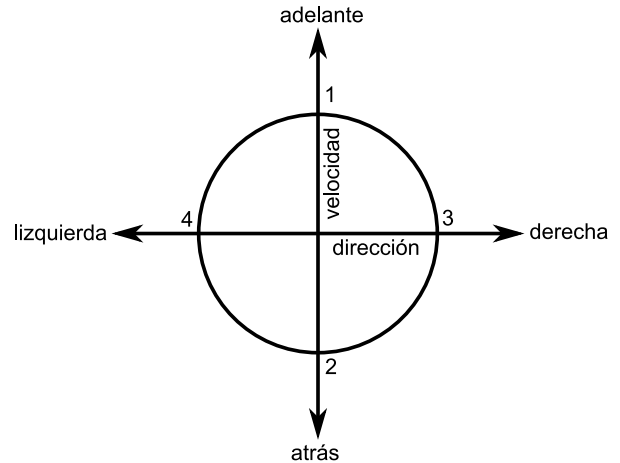


Fig. 4. Interpretación de los ejes del joystick: 1) velocidad máxima hacia delante, 2) velocidad máxima reversa, 3) velocidad máxima giro en sentido de las manecillas del reloj y 4) velocidad máxima giro en sentido contrario a las manecillas del reloj.

software permite ajustar fácil y rápidamente los 2 parámetros del joystick (velocidad y dirección) para acomodarse al usuario.

La optimización de la sensibilidad del joystick por el usuario suprime movimientos erráticos de la mano y extrae los movimientos pertinentes [5]. Esto facilita y hace accesible el control de la SRE a más personas, además que también mejora la seguridad y eficiencia.

B. Operación autónoma

En este modo, la SRE emplea un sistema de visión para navegar autónomamente a través de corredores, pasillos y una amplia gama de entornos interiores y exteriores.

Una videocámara a color obtiene imágenes y las transmite a una computadora portátil (laptop) a bordo de la silla de ruedas. La computadora toma la mejor decisión basada en las características de la imagen procesada con algoritmos de visión y envía las señales correspondientes a la unidad electrónica de control de los motores. El sistema de visión se basa en una investigación previa con el robot móvil IVWAN [6] (Fig. 5).



Fig. 5. Robot móvil omnidireccional IVWAN (Intelligent Vehicle With Autonomous Navigation).

Al igual que en IVWAN, el sistema de visión de la SRE se basa en una videocámara Panasonic que proporciona imágenes de 120x160 píxeles. El procesamiento de imágenes se realiza en LabVIEW¹ mientras que las decisiones se toman mediante un algoritmo desarrollado en JAVA².

B.1. Capacidades de visión

Una de las aplicaciones más populares de visión en robots autónomos móviles es el seguimiento de línea. Esta aplicación requiere que un vehículo siga una línea sólida y/o línea discontinua a lo largo de un camino evitando los obstáculos, superando los cambios del terreno y tratando de maximizar la velocidad. El desafío consiste en aplicar algoritmos confiables que funcionen bajo condiciones impredecibles y variantes de luminosidad, sombras, etc.

Como en los robots móviles, esta aplicación puede ser transferida a vehículos para personas con discapacidad motora.

La fig. 6 muestra un ejemplo representativo del seguimiento de línea bajo ciertas condiciones de luminosidad y sombras.

La fig. 6(a) muestra la imagen original obtenida por la videocámara. Nótese la presencia de sombras y un nivel de luminosidad que hace

problemática la discriminación automática entre el suelo (pasto) y la línea blanca.

La fig. 6(b) muestra la misma imagen con las sombras atenuadas. Esta imagen se obtuvo por una transformación de RGB a HSI, seguido de una rotación de 60° en el plano HS (de modo que el amarillo se convierta en rojo) y luego una transformación de vuelta a RGB.

Para realzar los colores brillantes (azul) en la imagen, una transformación de RGB a gris se realizó usando $R+G+3B$ en lugar de la típica fórmula $(R+G+B)/3$ (Fig. 6(c)). A continuación, un umbral (threshold) se aplicó para identificar mejor los colores brillantes (Fig. 6(d)).

Para minimizar el número de objetos en la imagen y retener los de mayor densidad de masa, las vecindades aisladas de 3x3 se removieron así y los agujeros de 3x3 se rellenaron (Fig. 6(e)). En esta etapa, la imagen ya es aceptable y la transformada de Hough puede aplicarse confiablemente (Fig. 6(f)).

El seguimiento de línea es sin duda útil para la navegación en entornos preestablecidos pero es impráctico para entornos nuevos y desconocidos donde una línea a seguir simplemente no existe.

Otra posibilidad para la navegación autónoma es la detección de corredores o caminos. Esta aplicación requiere que el vehículo detecte ambas orillas o bordes del corredor o camino, calcule el espacio disponible y navegue entre ellas en todo momento. Nuevamente, bajo condiciones impredecibles y variantes de luminosidad y sombras.

La fig. 7 muestra 3 ejemplos representativos de la detección de corredor/camino: (izquierda) la videocámara obtiene una imagen del entorno, (centro) usando detección de bordes de Canny ambas orillas del camino son encontradas, (derecha) la transformada de Hough muestra el borde más cercano a la silla de ruedas. Esta es la línea a seguir por el vehículo.

¹ National Instruments.

² Sun Microsystems.

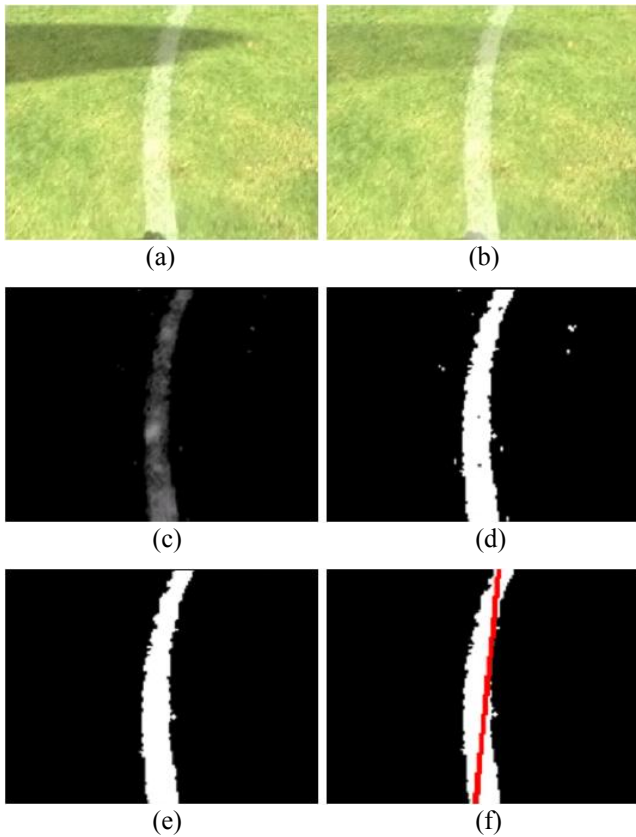


Fig. 6. Procesamiento de imágenes para seguimiento de línea: (a) imagen original, (b) remoción de sombras, (c) realce de colores brillantes, (d) imagen con umbral, (e) remoción de partículas y rellenado de agujeros y (f) la transformada de Hough.

La fig. 7(a) muestra una situación ideal mientras que las figs. 7(b) y (c) muestran las variaciones del entorno: detección de bordes bajo condiciones de sombra y obstáculos. Note la robustez del algoritmo de detección.

B.2. Toma de decisiones

La imagen resultante de la transformada de Hough proporciona una serie de coordenadas de líneas encontradas en la imagen.

Para ambos casos (seguimiento de línea y detección de bordes), los datos proporcionados por la transformada de Hough son procesados por un algoritmo de toma de decisiones implementado en JAVA en donde se remueven de la imagen las líneas innecesarias y se dejan solo las líneas importantes.

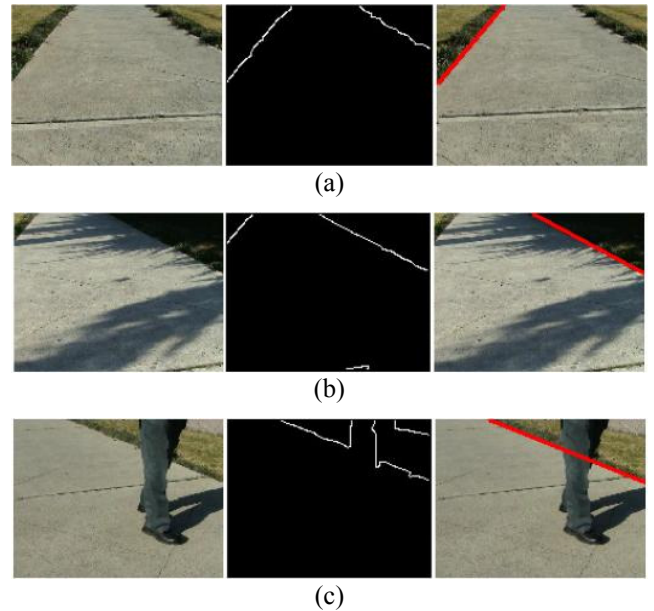


Fig. 7. Procesamiento de imágenes para navegación en corredores/caminos: (a) caso ideal y detección de bordes con (b) sombras y (c) obstáculos.

Como ejemplo, considere un grupo de 4 líneas (L1-L4) encontradas en la etapa de procesamiento de imagen (Fig. 8).

Una función probabilística de distribución Gaussiana es utilizada para asignar la importancia al grupo de líneas respecto a su ángulo [7]. Esto es: es más probable esperar una línea a $\theta = 0^\circ$ (i.e. paralela a la silla de ruedas y a su desplazamiento) que a otros ángulos. Líneas con $\theta \geq 70^\circ$ o perpendiculares a la silla que aparecen súbitamente en la imagen son omitidas.

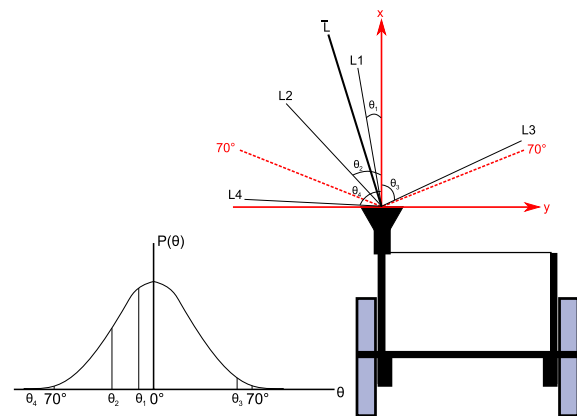


Fig. 8. Navegación autónoma: procedimiento gaussiano para la toma de decisiones para un grupo de líneas en la imagen. \bar{L} es calculada y es la línea a seguir.

En cada imagen tomada por la cámara, la siguiente línea a seguir por la SRE está determinada por la simple función de promedio de la ecuación (1).

$$\bar{L} = \frac{\sum_{i=1}^n L_i * P(\theta_i)}{\sum_{j=1}^n P(\theta_j)} \quad (1)$$

La ecuación (1) se puede descomponer en coordenadas x , y para determinar el siguiente movimiento de la SRE.

IV. APLICACIÓN DE LA SRE COMO PLATAFORMA-GUÍA EN MUSEOS

En colaboración con el Museo Descubre [8] en Aguascalientes, México, se está trabajando para hacer el museo accesible a personas con dificultades motrices severas.

La idea principal es utilizar las capacidades de visión de la silla de ruedas para desplazarse con autonomía dentro del museo, mientras se transporta con toda seguridad al visitante.

Con este fin, la modalidad de detección de bordes en corredores/caminos fue elegida para evitar la necesidad de pintar líneas que perturbarían la estética del museo. Dos sistemas se incorporaron a la SRE: (1) un par de altavoces para ofrecer información sobre pinturas, esculturas y exposiciones en mensajes de audio pregrabados y (2) un sistema inalámbrico de banda ancha que transmite el estado de la SRE a un servidor localizado en una estación central.

En cuanto al sistema de explicación por audio, un conjunto de mensajes de voz fue registrado en la computadora abordo. Cuando la SRE se aproxima a un punto predeterminado del recorrido, ambos, punto y SRE, se comunican por tecnología RFID y la SRE reproduce el mensaje correspondiente.

En relación con el sistema de monitoreo, es importante conocer la ubicación y el estado de la SRE en todo momento. Por seguridad del

visitante, un operador remoto puede tomar el control de la SRE en cualquier momento. Con este objetivo, se ha implementado una plataforma de vigilancia activa (Fig. 9).

Usando el sistema inalámbrico de banda ancha, la SRE transmite 4 parámetros en tiempo real a la estación central: (1) la imagen obtenida por la cámara, (2) la velocidad de los motores, (3) la ubicación de la SRE en el museo y (4) el estado de la batería.

Localmente en el servidor de la estación central, (5) la imagen es procesada como en la silla de ruedas, (6) las señales de control enviadas al motor son reconstruidas y desplegadas al operador así como la ruta seguida por la SRE.

Por otro lado, el operador es capaz de establecer de manera remota (8) la velocidad máxima de los motores y (9) detener la SRE.

Aun cuando la SRE ha pasado con éxito un número de pruebas de laboratorio, es necesario aún añadirle un sistema para evitar colisiones. Si bien es cierto que la gente en el museo se aparta al ver acercarse la SRE, no se puede dar por hecho que siempre sucederá así. Aparte, colisiones con objetos olvidados en el suelo como mochilas o bolsas son probables.

Dos soluciones están en curso de evaluación: sensado ultrasónico y detección de obstáculos con el sistema de visión. Si bien es cierto que éste último permitiría conservar un prototipo más simple con menos elementos y a menor costo, representa ciertamente un reto mayor de implementación.

V. CONCLUSION

En este artículo, se presentó el diseño, implementación y primer prototipo de una SRE inteligente y de bajo costo para personas con discapacidad motora. Esta plataforma tiene por objetivo responder a 2 criterios: ser flexible y accesible económicamente para cualquiera.

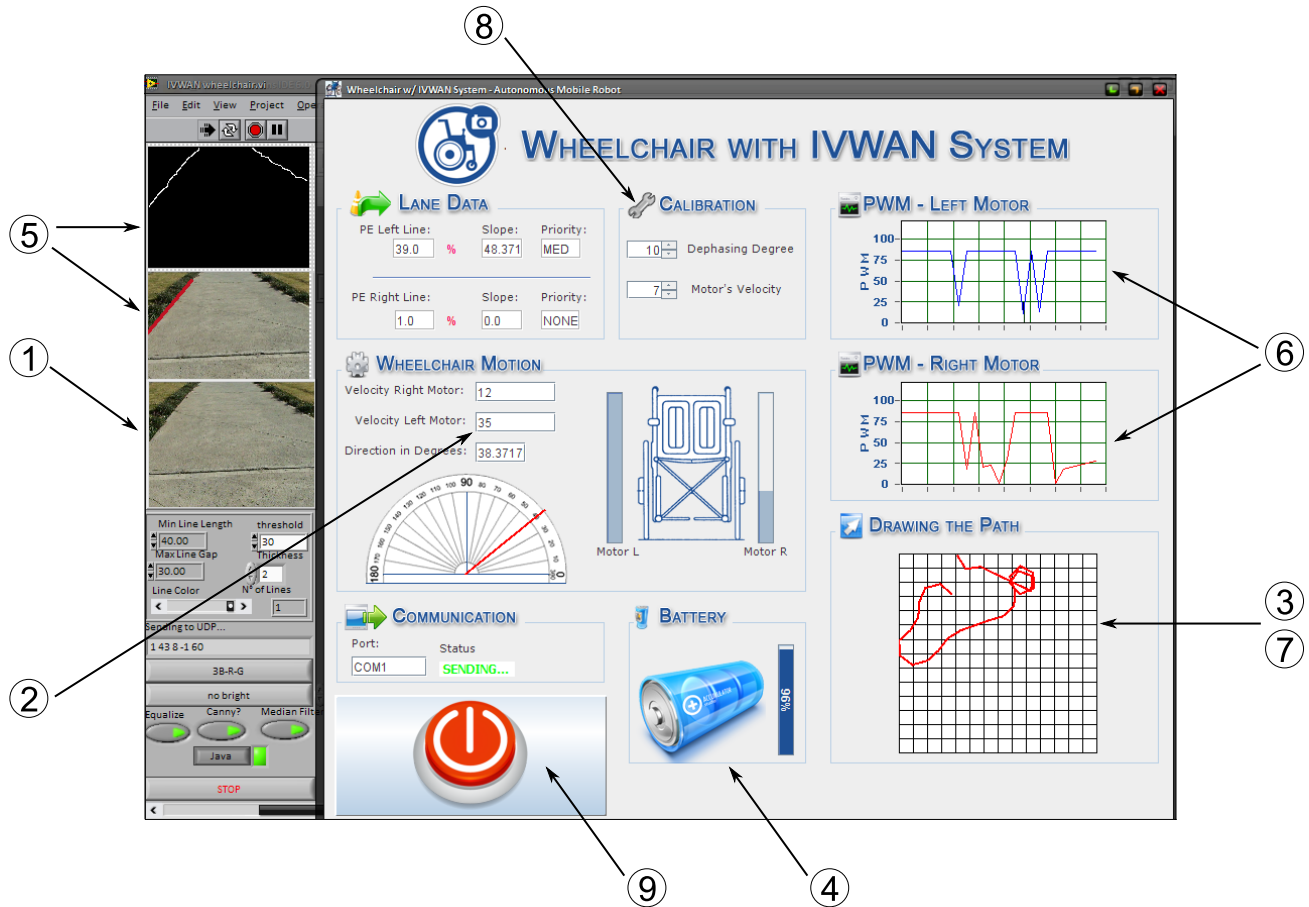


Fig. 9. Plataforma de monitoreo remoto para la SRE dentro del museo: vista de la interfaz gráfica del operador. Los parámetros 1-4 se transmiten de la SRE a la estación central. Los parámetros 5-7 se procesan localmente en la estación central. Los parámetros 8-9 se envían de la central a la SRE a decisión del operador.

El primer prototipo desarrollado puede ser operado en modo estándar (i.e. con las manos) y en modo manual para aquellos que son capaces de maniobrarla, y en modo autónomo para los severamente afectados en sus capacidades motoras.

El costo del prototipo de SRE es atractivo: una tercera parte de los comercialmente disponibles y, en producción en masa, se puede esperar definitivamente un precio mucho menor.

Ambos modos, manual y autónomo, se presentaron y describieron. En particular, el modo autónomo se basa en el sistema de visión por computadora de un robot móvil y permite a la SRE desplazarse por pasillos, corredores, centros comerciales, aeropuertos, entre otros entornos tanto interiores como exteriores.

Este artículo también presentó el procesamiento de imágenes y la toma de decisiones para 2 aplicaciones: seguimiento de línea y detección de bordes en corredores o caminos.

El prototipo desarrollado está siendo adaptado como una plataforma guía en un museo para visitantes severamente afectados en su capacidad motora junto con un sistema remoto de vigilancia en tiempo real.

VI. AGRADECIMIENTOS

Este trabajo fue financiado en su totalidad por el CEPii (Centro Panamericano de Investigación e Innovación - México).

VII. REFERENCIAS

- [1] World Health Organization, “*The World Health Report 2002: Reducing Risks, Promoting Healthy Life*”, Geneva, Switzerland, 2002.
- [2] A. Koontz, J. Pearlman, B. Impink, R. Cooper and M. Wilkinson, “*Wheelchairs*”, Chapter 8 in *An Introduction to Rehabilitation Engineering*, Taylor & Francis, pp 129-155, 2006.
- [3] B. Woods and N. Watson, “*A Short History of Powered Wheelchairs*”, *Assistive Technology*, 15, pp 164-180, 2003.
- [4] R. Cooper, “*Wheelchair Selection and Configuration*”, Demos Medical Publishing, Inc., New York, 1998.
- [5] D. Ding, R. Cooper and D. Spaeth, “*Optimized Joystick Controller*”, Proc. of Annual International Conference of the IEEE-EMBS, pp 4881-4883, 2004.
- [6] J. Martinez, G. Moran, B. Romero, A. Camacho, D. Gutheim, J. Varona and R. Velazquez, “*Multifunction All-Terrain Mobile Robot IVWAN: Design and First Prototype*”, Proc. of 2nd Israeli Conference on Robotics, Herzlia, Israel, 2008.
- [7] D. Hunt, L. Nolte and W. Ruedger, “*Performance of the Hough Transform and its Relationship to Statistical Signal Detection Theory*”, *Computer Vision, Graphics, and Image Processing*, 43(2), pp 221-238, 1988.
- [8] Museo Descubre – Aguascalientes, Ags. Updated information available at: www.descubre.org.mx

DISEÑO DE UN CONTROL ÓPTIMO EN TIEMPO CONTINUO PARA LA ESTABILIZACIÓN DE UN ROBOT MÓVIL

Morín Castillo María Montserrat, Hernández Angulo Lucio,
Gutiérrez Arias José Eligio Moisés

Facultad de Ciencias de la Electrónica.

Benemérita Universidad Autónoma de Puebla.

Av. San Claudio y 18 Sur, Col. San Manuel, C.P. 72570, Puebla, Pue., México.

Correo electrónico: mmorin@ece.buap.mx / lucio_y6@hotmail.com / jmgutierrez@ece.buap.mx

RESUMEN

Se plantea el problema de encontrar un control óptimo lineal para la estabilización de trayectorias en un robot móvil provisto de dos ruedas activas y una pasiva, la programación dinámica es el fundamento para la síntesis del control. Se obtiene las ecuaciones dinámicas no lineales de movimiento del robot móvil, y posteriormente para una trayectoria deseada se obtienen las ecuaciones lineales. En la síntesis del control óptimo, es importante la determinación de una matriz, llamada solución de la ecuación matricial diferencial de Riccati. En este trabajo, al diseñar el control óptimo en un tiempo finito, obtenemos un sistema lineal variante en el tiempo, cuando se considera una trayectoria de tipo circular del robot móvil. Las matrices de peso del indicador de desempeño se pueden elegir para mejorar la calidad de la estabilización. Se presentan gráficas del comportamiento obtenido para cada uno de los estados del sistema linealizado.

Palabras clave: Robot móvil, control óptimo, programación dinámica, ecuación matricial de Riccati.

I. INTRODUCCIÓN

Un robot móvil es una máquina autónoma capaz de moverse en un ambiente dado utilizando como herramienta alguna configuración mecánica que permita algún tipo de desplazamiento, combinado con algunos componentes computacionales. De acuerdo a su forma de locomoción se clasifican en: robots móviles con locomoción mediante orugas, mediante patas y mediante ruedas. Aunque la

mayoría de los robots móviles actualmente en operación son experimentales, ya están empezando a desarrollarse para fines industriales; las cuales, entre otras características, pueden ser inhóspitas, peligrosas o inaccesibles para seres humanos.

El caso más simple de robots móviles son los robots con ruedas; estos robots constan de una o más ruedas de tracción, tienen ruedas pasivas opcionales, y posiblemente ruedas de dirección. Considerando físicamente a un robot móvil como un conjunto de varios cuerpos sólidos con diferentes grados de movilidad, un robot móvil puede constar de 5 grados de libertad como en este caso; sin embargo, los robots móviles propulsados por ruedas a su vez se clasifican de acuerdo al número y al tipo de grados de libertad. Los grados de libertad considerados en robótica móvil son de dos tipos, de movilidad y de direccionabilidad. El grado de movilidad es un grado de libertad asociado a un variable de velocidad del robot y el grado de direccionabilidad es un grado de libertad asociado a una variable de dirección del robot [1]. La configuración del robot móvil que se utilizará a lo largo de este trabajo, es el más comúnmente usado y es llamado “control diferencial”; esta configuración consta dos ruedas para tracción y dirección, y una rueda pasiva para estabilidad de la plataforma del robot, con lo anterior, se puede decir que se trata de un robot móvil de tipo (2,0). El diseño del control diferencial tiene dos motores montados en posiciones fijas al lado izquierdo y derecho de el robot, independientemente conducen una rueda cada uno [2].

A causa de que se adiciona la complejidad del control para la coordinación del movimiento del robot, a la complejidad de una posible trayectoria en un ambiente posiblemente hostil, es importante la elección de un control que pueda minimizar los tiempos de estabilización a los posibles cambios de trayectoria que puede presentar el robot al recorrer el entorno dado.

En la literatura, el estudio y diseño de los sistemas de control para los robots móviles es muy amplio, así como también la cantidad de algoritmos de control con los que se puede encontrar una aplicación interesante. Aunque el robot móvil es un sistema con movimientos físicamente simples, matemáticamente no lo es, debido a que su modelo dinámico es no lineal. La aplicación de algoritmos de control para un robot móvil es basta, pasando desde algoritmos clásicos de control como en [3] donde se plantea un control PID y de linealización retroalimentada para el control de la parte dinámica (que involucra a la velocidad de los motores) y la parte cinemática (que involucra a la trayectoria deseada) respectivamente, hasta controles de tipo Difuso como en [4] donde el control se realiza en base a conjuntos de decisión y de respuesta, cabe mencionar que para este caso solo se utiliza el modelo cinemático del robot móvil. Trabajos más recientes como [5] presenta un algoritmo de control para un robot móvil, basado en información visual para la estabilización del sistema. En [6] se presenta el modelo matemático del robot móvil en un sistema de referencia polar de 2 dimensiones, planteando el problema de un regulador o algoritmo de control basado en los teoremas de estabilidad de Lyapunov. En materia de control óptimo para robots móviles, en [7] se plantea el problema de encontrar esta ley de control considerando el modelo matemático a partir del error de orientación y de posición para una trayectoria deseada obteniendo una ecuación lineal de cuatro variables de estado, además haciendo una transformación sobre los valores característicos se obtiene un segundo sistema de dos variables de estado las cuales representa los modos lentos del sistema, el control óptimo es calculado solo para este segundo sistema.

II. ESTABLECIMIENTO DEL PROBLEMA

Considerando el siguiente proceso controlable que representa las ecuaciones dinámicas del robot móvil

$$\begin{aligned} \dot{y} &= f(y, u), \\ u(\cdot) &\in U = \{u : u(t) \in \Omega \subseteq \mathfrak{R}^r\}; \end{aligned} \quad (1)$$

donde $y \in \mathfrak{R}^n$ contiene los estado del sistema, $u \in \mathfrak{R}^m$ representa los controles de entrada. Las ecuaciones diferenciales lineales que gobiernan las desviaciones $x(t) = \tilde{y}(t) = y(t) - y^d(t)$ para algún movimiento alrededor de alguna trayectoria deseada $y^d(t)$ y un control deseado $u^d(t), \in [t_0, t_1)$ pueden escribirse como:

$$\dot{x} = A(t)x + B(t)\Delta u; \quad (2)$$

donde

$$A(t) = \frac{\partial f[y^d(t), u^d(t)]}{\partial y}, B(t) = \frac{\partial f[y^d(t), u^d(t)]}{\partial u};$$

y considerando también el criterio de desempeño

$$\int_{t_0}^{t_1} [x^T(t)Q(t)x(t) + u^T(t)G(t)u(t)] dt + x^T(t_1)Sx(t_1), \quad (3)$$

donde S y Q son matrices simétricas definidas no negativas ($S = S^T \geq 0$; $Q(t) = Q(t)^T \geq 0$); y $G(t)$ es una matriz simétrica definida positiva ($G(t) = G(t)^T > 0$) para $t_0 \leq t \leq t_1$. Entonces surge el problema de determinar un vector de entrada $\Delta u(t)$, donde $t_0 \leq t \leq t_1$, tal que minimice el criterio de desempeño (3) [9,11].

III. ECUACIONES DE MOVIMIENTO DEL ROBOT MÓVIL

El tipo de diseño del robot móvil que se utilizará a lo largo de este trabajo será el de *control diferencial*, el cuál como ya se mencionó, consta de dos ruedas activas para tracción y dirección, y una rueda pasiva (ver figura 1).

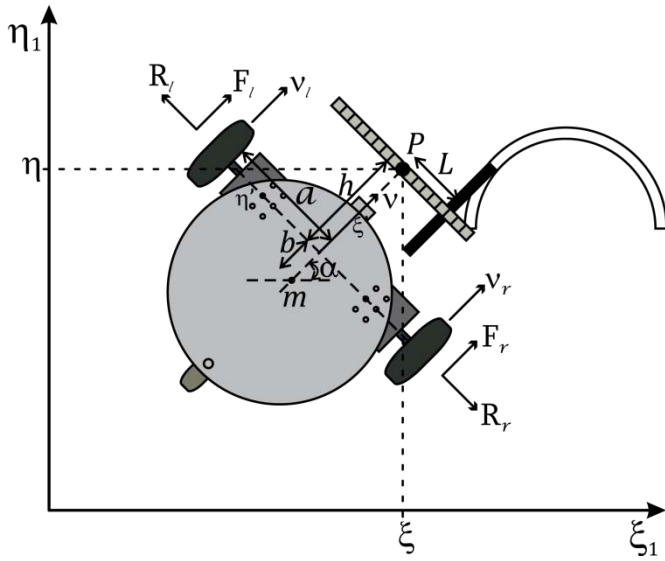


Figura 1. Diagrama físico del Robot Móvil.

Considerando que el robot sólo presenta movimientos planares traslacionales y rotacionales, y que no presenta ningún tipo de deslizamientos lo que lleva a las restricciones no holonómicas (las velocidades de las ruedas activas deben tener la dirección P_ξ); podemos, utilizando principios básicos de la mecánica, obtener las siguientes ecuaciones que describen la cinemática del robot móvil dentro del marco de referencia fijo ξ_1, η_1 :

$$\dot{\xi} = v \cos(\alpha) - h\omega \sin(\alpha), \quad (4)$$

$$\dot{\eta} = v \sin(\alpha) + h\omega \cos(\alpha), \quad (5)$$

$$\dot{\alpha} = \omega, \quad (6)$$

Utilizando el *principio del equilibrio dinámico de D'Alembert* [8], y sustituyendo las restricciones por las fuerzas reactivas \vec{R}_r y \vec{R}_l , se obtienen las ecuaciones dinámicas que involucran las fuerzas de empuje \vec{F}_r y \vec{F}_l :

$$m(\dot{v} + b\omega^2) = F_l + F_r, \quad (7)$$

$$\tilde{J}\dot{\omega} - mbv\omega = (F_r - F_l)a. \quad (8)$$

Es necesario además, asociar las fuerzas F_r y F_l a las coordenadas φ_r y φ_l que determinan el giro de cada uno de los motores del robot móvil; para esto,

se sabe que el torque o momento de un motor está dado por

$$\tau = F\rho; \quad (9)$$

donde τ y F son el torque y la fuerza aplicadas por el motor y ρ es el radio de la rueda acoplada el motor. Es necesario también, un modelo matemático que describa el movimiento del motor, para esto el modelo más simple conocido para un motor de DC es:

$$\tau = \chi V - \sigma\dot{\phi}; \quad (10)$$

donde τ es el torque del motor (Nm); V es el voltaje aplicado a las terminales del motor (V); $\dot{\phi}$ es la velocidad angular (rad/s); χ es la fricción viscosa; y σ es la fuerza contra-electromotriz del motor [13]. Considerando las expresiones para la velocidad angular de cada una de las ruedas

$$\dot{\phi}_r = \frac{v_r}{\rho} = \frac{v + a\omega}{\rho}, \quad (11)$$

$$\dot{\phi}_l = \frac{v_l}{\rho} = \frac{v - a\omega}{\rho}; \quad (12)$$

y considerando las ecuaciones (9) y (10) se pueden obtener una expresión de fuerza asociada a las ruedas derecha e izquierda

$$F_r = \frac{\chi V_r \rho - \sigma(v + a\omega)}{\rho^2}, \quad (13)$$

$$F_l = \frac{\chi V_l \rho - \sigma(v - a\omega)}{\rho^2}; \quad (14)$$

sustituyendo las expresiones (13) y (14) en las ecuaciones de movimiento (7) y (8), se tiene

$$m\dot{v} + mb\omega^2 = \frac{\chi}{\rho}(V_r + V_l) - \frac{2\sigma v}{\rho^2}, \quad (15)$$

$$\tilde{J}\dot{\omega} - mbv\omega = \frac{a\chi}{\rho}(V_r - V_l) - \frac{2\sigma a^2 \omega}{\rho^2}. \quad (16)$$

Por lo tanto, las ecuaciones (4), (5), (6), (15) y (16)

$$\dot{\xi} = v \cos(\alpha) - h\omega \sin(\alpha),$$

$$\dot{\eta} = v \sin(\alpha) + h\omega \cos(\alpha),$$

$$\dot{\alpha} = \omega, \tag{17}$$

$$\dot{v} = -b\omega^2 - \frac{2\sigma v}{m\rho^2} + \frac{\chi}{m\rho}(V_r + V_l),$$

$$\dot{\omega} = \frac{mbv\omega}{\tilde{J}} - \frac{2\sigma a^2\omega}{\tilde{J}\rho^2} + \frac{a\chi}{\tilde{J}\rho}(V_r - V_l).$$

describen el movimiento del robot móvil.

IV. TRAYECTORIAS DESEADAS Y ECUACIONES LINEALES

Ya que cualquier trayectoria puede ser aproximada mediante segmentos circulares y rectos, por medio del análisis de los posibles movimientos del robot en el plano, se pueden obtener las trayectorias deseadas o programadas que este puede realizar. Considerando que el robot presenta movimientos horizontales y verticales a través de movimientos paralelos a los ejes $0, \xi$ y $0, \eta$ tenemos en la tabla 1 las siguientes trayectorias deseadas (ver figura 2).

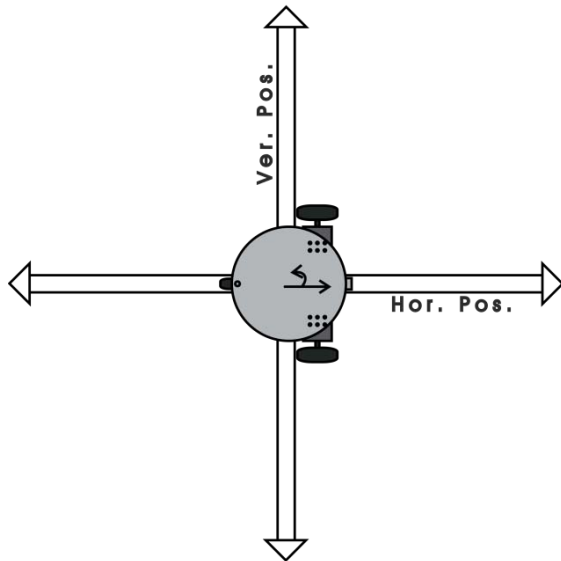


Fig. 2. Trayectorias Rectas del robot móvil.

Ahora se consideran los desplazamientos circulares; para esto se supondrá que el estado inicial o posición de casa del robot es paralelo al eje ξ y que el movimiento se describe sobre un círculo de radio R , el cual está seccionado en cuatro cuadrantes (ver figura 3); además se tomará en cuenta el sentido en que se recorre cada

segmento de circunferencia; las trayectorias de muestran en el tabla 2.

Trayectoria	ξ^d	η^d	α^d	v^d	ω^d
1. Línea Horizontal sentido Positivo	$\xi_0 + v_0 t$	0	0	v_0	0
2. Línea Horizontal sentido Negativo	$\xi_0 + v_0 t$	0	π	v_0	0
3. Línea Vertical sentido Positivo	0	$\eta_0 + v_0 t$	$\frac{\pi}{2}$	v_0	0
4. Línea Vertical sentido Negativo	0	$\eta_0 + v_0 t$	$-\frac{\pi}{2}$	v_0	0

Tabla 1. Trayectorias deseadas Rectas.

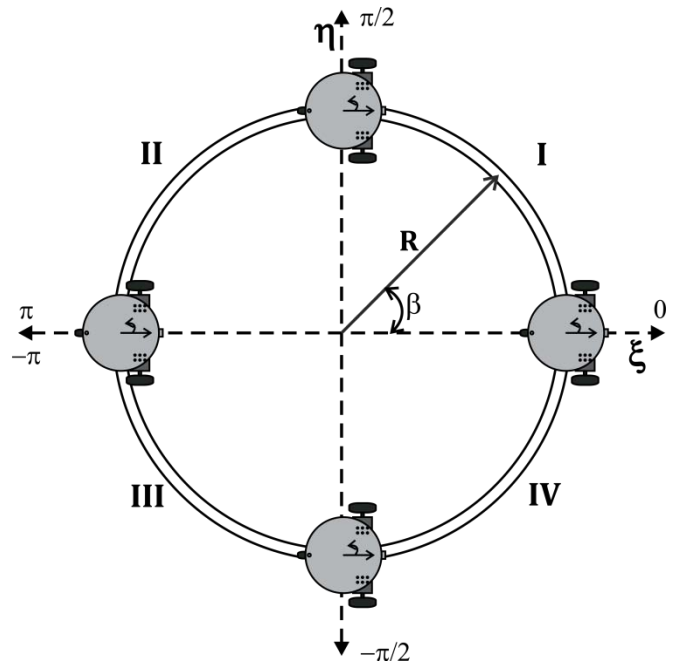


Fig. 3. Trayectorias Circulares del robot móvil.

Siendo las ecuaciones de movimiento del robot móvil el conjunto de ecuaciones (17), se pueden aproximar linealmente a vecindades cercanas a las trayectorias deseadas utilizando la expresión (2). Siendo la trayectoria deseada $y^d = (\xi^d \ \eta^d \ \alpha^d \ v^d \ \omega^d)^T$ perteneciente a un movimiento en línea horizontal en sentido positivo (tabla 1, número 1), se tienen las ecuaciones lineales

Trayectoria	ξ^d	η^d	α^d	v^d	ω^d
5. Cuadrante I (sentido reloj)	$R \cos(\beta)$	$R \sin(\beta)$	$-\omega_0 t$	$\omega_0 R$	ω_0
6. Cuadrante I (contra reloj)	$R \cos(\beta)$	$R \sin(\beta)$	$\frac{\pi}{2} + \omega_0 t$	$\omega_0 R$	ω_0
7. Cuadrante II (sentido reloj)	$R \cos(\beta)$	$R \sin(\beta)$	$\frac{\pi}{2} - \omega_0 t$	$\omega_0 R$	ω_0
8. Cuadrante II (contra reloj)	$R \cos(\beta)$	$R \sin(\beta)$	$\pi + \omega_0 t$	$\omega_0 R$	ω_0
9. Cuadrante III (sentido reloj)	$R \cos(\beta)$	$R \sin(\beta)$	$\pi - \omega_0 t$	$\omega_0 R$	ω_0
10. Cuadrante III (contra reloj)	$R \cos(\beta)$	$R \sin(\beta)$	$-\frac{\pi}{2} + \omega_0 t$	$\omega_0 R$	ω_0
11. Cuadrante IV (sentido reloj)	$R \cos(\beta)$	$R \sin(\beta)$	$-\frac{\pi}{2} - \omega_0 t$	$\omega_0 R$	ω_0
12. Cuadrante IV (contra reloj)	$R \cos(\beta)$	$R \sin(\beta)$	$\omega_0 t$	$\omega_0 R$	ω_0

Tabla 2. Trayectorias deseadas Circulares.

$$\begin{aligned}
 \dot{\xi} &= \tilde{v}, \\
 \dot{\eta} &= -v_0 \tilde{\alpha} + h \tilde{\omega}, \\
 \dot{\alpha} &= \tilde{\omega} \\
 \dot{v} &= -\frac{2\sigma}{m\rho^2} \tilde{v} + \frac{\chi}{m\rho} u_1, \\
 \dot{\omega} &= \frac{mbv_0}{\tilde{J}} \tilde{\omega} - \frac{2\sigma a^2}{\tilde{J}\rho^2} \tilde{\omega} + \frac{a\chi}{\tilde{J}\rho} u_2;
 \end{aligned} \tag{18}$$

Si la trayectoria deseada y^d pertenece a alguno de los movimientos circulares de la tabla 2, se pueden obtener las siguientes ecuaciones lineales que dependen de α^d , la cual representa la trayectoria circular que se desea obtener

$$\begin{aligned}
 \dot{\xi} &= -\left(\omega_0 R \sin(\alpha^d) + h\omega_0 \cos(\alpha^d)\right) \tilde{\alpha} + \\
 &\tilde{v} \cos(\alpha^d) - h\tilde{\omega} \sin(\alpha^d),
 \end{aligned}$$

$$\begin{aligned}
 \dot{\eta} &= \left(\omega_0 R \cos(\alpha^d) - h\omega_0 \sin(\alpha^d)\right) \tilde{\alpha} + \\
 &\tilde{v} \sin(\alpha^d) - h\tilde{\omega} \cos(\alpha^d), \\
 \dot{\alpha} &= \tilde{\omega}
 \end{aligned} \tag{19}$$

$$\begin{aligned}
 \dot{v} &= -\frac{2\sigma}{\rho^2 m} \tilde{v} - 2b\omega_0 \tilde{\omega} + \frac{\chi}{m\rho} u_1, \\
 \dot{\omega} &= \frac{mb\omega_0}{\tilde{J}} \tilde{v} + \frac{mb\omega_0 R}{\tilde{J}} \tilde{\omega} - \frac{2\sigma a^2}{\tilde{J}\rho^2} \tilde{\omega} + \frac{a\chi}{\tilde{J}\rho} u_2;
 \end{aligned}$$

V. DEDUCCIÓN DEL CONTROL MEDIANTE LA PROGRAMACIÓN DINÁMICA

Considerando el problema de control óptimo

$$\min J = \int_{t_0}^{t_1} F(x, u, t) dt + S[x(t_1)]$$

$$\text{sujeto a : } \dot{x}(t) = f(x, u, t),$$

$$\text{con : } x(t_0) = x_0,$$

$$u(t) \in \Omega(\tau);$$

se define la *función valor* $W^*(x, t)$, que proporciona el mínimo valor del funcional objetivo del problema de control

$$W^*(x, t) = \min_{u(\tau)} \int_{t_0}^{t_1} F(x, u, t) dt + S[x(t_1)]$$

$$\text{sujeto a : } \dot{x}(t) = f(x, u, \tau), \text{ para } t \leq \tau \leq t_1,$$

$$\text{con : } x(t_0) = x_0,$$

$$u(t) \in \Omega(\tau).$$

Suponiendo que $W(x, t)$ es una solución de la ecuación de Hamilton-Jacobi-Bellman, se verifica que:

$$\begin{aligned}
 0 &= \min_{u \in \Omega} \{F(x, u, t) + \nabla_x W(x, t) f(x, u, t)\} + \\
 &W_t(x, t), \quad \forall x, t,
 \end{aligned} \tag{20}$$

$$\text{con : } W(x, t_1) = S(x), \text{ para todo } x.$$

Suponiendo también que $u(x, t)$ alcanza el mínimo en la ecuación (20), para todo x y t . Entonces la trayectoria de control $u(t)$ para $t_0 \leq t \leq t_1$, es óptima [11, 12].

Para solucionar este problema de control óptimo, se escribe la ecuación de Hamilton-Jacobi-Bellman de acuerdo al indicador de desempeño (3) sujeto al sistema (2) y se establece la suposición de que la solución a la ecuación es de la forma:

$$W(x, t) = x^T P(t)x, \quad (21)$$

donde $P(t)$ es una matriz simétrica n -dimensional. Por lo que se puede obtener

$$0 = \min_{u(t)} \left\{ x^T Qx + u^T Gu + 2x^T P(t)[Ax + Bu] \right\} + x^T \frac{\partial P(t)}{\partial t} x; \quad (22)$$

derivando parcialmente con respecto a u la expresión entre llaves e igualando a cero se obtiene:

$$u = -G^{-1}B^T P(t)x. \quad (23)$$

Sustituyendo (23) en la ecuación (22), se deduce:

$$-\frac{\partial P(t)}{\partial t} = Q - P(t)BG^{-1}B^T P(t) + P(t)A + A^T P(t), \quad (24)$$

con : $P(t_1) = 0$,

que es una *ecuación matricial diferencial de Riccati*, donde $P(t)$ se encuentra resolviendo esta ecuación.

Por lo tanto, el *control óptimo* del problema es:

$$u(t) = -\rho Kx(t), \quad K = G^{-1}B^T P(t); \quad (25)$$

la solución $P(t) \geq 0$ de la ecuación matricial diferencial de Riccati (24) es llamada *solución de estabilización*; y la matriz K definida en (25) es la *matriz de ganancia del control óptimo* o la *matriz de estabilización óptima* [10].

Finalmente, la ecuación del *sistema del control óptimo en lazo cerrado* está dada por

$$\dot{x}(t) = (A - BK)x(t). \quad (26)$$

VI. SOLUCIÓN Y RESULTADOS

El valor de cada uno de los parámetros del sistema necesarios para obtener los resultados se presenta en la tabla 3.

Variable	Valor	Descripción
v_0	1,5	Velocidad deseada [m/s]
ω_0	0.5	Velocidad angular deseada [rad/s]
a	0,40	Distancia entre las ruedas [m]
b	0,40	Distancia del centro de masa al eje de las ruedas [m]
h	0,10	Distancia del eje de las ruedas al arreglo de sensores [m]
m	4,5	Masa del Robot Móvil [Kg]
ρ	0,08	Radio de las Ruedas [m]
R	0,35	Radio de inercia del Robot Móvil [m]
χ	0,01	Fricción viscosa del motor
σ	0,009	Fuerza contra-electromotriz del motor
\tilde{J}	0,2868	Momento de inercia del Robot Móvil

Tabla 3. Valor de los parámetros del Robot Móvil.

El algoritmo de solución para la obtención del control óptimo para las trayectorias del robot móvil es el siguiente:

1. Obtener el sistema lineal (2) alrededor de alguna trayectoria deseada (tablas 1 y 2).
2. Verificar si el par de matrices (A, B) del sistema lineal (2) es completamente controlable, si no, no es posible obtener el control óptimo.
3. Desarrollo y solución de la ecuación diferencial matricial de Riccati con condiciones finales (24), una vez que se han escogido valores adecuados para las matrices de peso Q y G . Para la solución de la ecuación diferencial de Riccati se realiza un

cambio de variable para el tiempo ($\tau = t_f - t$), para convertir este problema en uno de condiciones iniciales.

4. Obtener mediante polinomios una aproximación a cada una de las curvas de solución de la matriz diferencial de Riccati; posteriormente se hace la inversión de estas curvas en el tiempo para obtener el resultado a condiciones finales buscado.
5. Obtener el control óptimo (25).
6. Se obtienen las gráficas del comportamiento del sistema bajo el control óptimo diseñado mediante la simulación de la solución de la ecuación del sistema en lazo cerrado (26).

A continuación, se presenta la solución del control óptimo para el robot móvil, para una trayectoria correspondiente a un segmento de círculo en el cuadrante IV para $\beta \in (-\pi/2, 0)$ en sentido de las manecillas del reloj ($\alpha^d = \omega_0 t$). En esta situación las matrices $A(t)$ y B son:

$$A(t) = \begin{pmatrix} 0 & 0 & -0,175 \sin(0,5t) - 0,05 \cos(0,5t) \\ 0 & 0 & 0,175 \cos(0,5t) - 0,05 \sin(0,5t) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} \cos(0,5t) & -0,1 \sin(0,5t) \\ \sin(0,5t) & 0,1 \cos(0,5t) \\ 0 & 1 \\ -0,625 & -0,4 \\ 3,14 & -0,471 \end{pmatrix},$$

$$B = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0,0278 & 0 \\ 0 & 0,1743 \end{pmatrix};$$

las cuales hacen que el sistema lineal sea variante en el tiempo, ya que la matriz de parámetros A depende del tiempo.

Para asegurar que existe una ley de control óptima para este sistema, es necesario afirmar que el sistema lineal variante en el tiempo es completamente controlable; para esto se utilizará el siguiente criterio.

Proposición 1. [14] Asumiendo que las funciones matriciales $A(t)$, $B(t)$ son analíticas en R^+ . El sistema $[A(t), B(t)]$ es completamente controlable en algún tiempo finito si

$$\text{rango}[M_1(t_0), M_2(t_0), \dots, M_k(t_0)] = n \quad (27)$$

$$\exists t_0 > 0.$$

donde

$$M_1(t) = B(t),$$

$$M_k(t) = -A(t) + \frac{d}{dt} M_{k-1}(t), \quad k=2, \dots, n-1.$$

Teniendo como resultado un rango de 5, se afirma que el sistema es completamente controlable, por lo tanto es posible aplicar el algoritmo para encontrar una ley de control óptima.

En la solución y diseño de un control óptimo, la matriz K , se calcula con una apropiada selección de las matrices de peso Q y G . Cuando se trabaja con sistemas grandes es común elegir a estas matrices como diagonales. Se tomará la matriz de peso $R^{2 \times 2}$ como una matriz identidad, con lo cual se asigna igual peso a las señales de control y se logra que cada una de ellas actúe en forma independiente.

Los resultados de la simulación de cada variable de estado del sistema linealizado para el procedimiento descrito anteriormente, se muestran en la figura 4, obteniendo la siguiente matriz de peso Q que mejora la estabilidad del sistema; cabe mencionar que esta matriz no es única. Las condiciones iniciales para la simulación del sistema en lazo cerrado son $x(0) \ni [-1 \ -1 \ 1 \ 1 \ -1]$; esta simulación se realizó en un intervalo de tiempo de 0 a 20seg. En la figura 4, se puede observar que el error de cada uno de los estados del sistema es llevado a cero; esto indica que el robot móvil parte de una

posición arbitraria y posteriormente se estabiliza en la trayectoria deseada.

$$Q = \begin{pmatrix} 9,5 & 0 & 0 & 0 & 0 \\ 0 & 1000 & 0 & 0 & 0 \\ 0 & 0 & 500 & 0 & 0 \\ 0 & 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 0 & 900 \end{pmatrix}$$

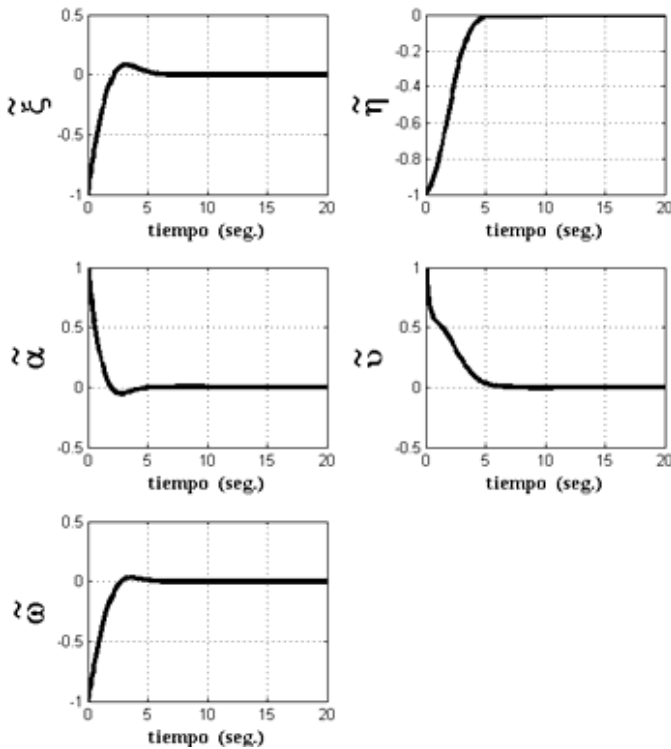


Fig. 4. Simulación del sistema en lazo cerrado con matriz de peso Q determinada iterativamente a prueba y error.

VII. CONCLUSIONES

Estudiamos el modelo dinámico del robot móvil, el cual consta de 5 ecuaciones de primer orden; se consideran la cinemática y la dinámica del robot, esta última resulta de considerar el movimiento rotacional de las ruedas del robot móvil. Mediante la programación dinámica se sintetizó la ley de control óptimo para una trayectoria circular del robot móvil. La dinámica del control óptimo se obtuvo solucionando una ecuación diferencial matricial de Riccati, elaborando un programa en el

software MATLAB que da la solución a dicha ecuación matricial, que complementada con una buena sintonización de las matrices de peso proporciona un buen desempeño de la ley de control.

VIII. REFERENCIAS

[1] C. Canudas, B. Siciliano, G. Bastin, B. Brogliato, G. Campion, B. D’Andrea-Novel, A. De Luca, W. Khalil, R. Lozano, R. Ortega, C. Samson y P. Tomei, *Theory of Robot Control*, 1996, London.

[2] T. Bräunl, *Embedded Robotics*, Springer, 2° Ed., 2006, págs. 3-5, 41-44, 97-105.

[3] E. Aranda, T. Salgado, M. Velasco, *Control no Lineal Discontinuo de un Robot Móvil*, Computación y Sistemas, Número Especial, 2002, págs. 042-049.

[4] R. Carrasco, A. Cipriano, *Sistema de Guiado para un Robot Móvil basado en Lógica Difusa*, XV Congreso de la Asociación Chilena de Control Automático, 2002.

[5] D. P. Tsakiris, C. Samson, P. Rives, *Vision-Based Time-Varying Mobile Robot Control*, Dans Final European Robotics Network (ERNET) Workshop, Germany, 1996.

[6] T. Y. Wang, C. C. Tsai, J. L. Pang, *Nonlinear Regulation and Path Tracking of a Wheeled Mobile Robot in Polar Coordinates*, Journal of the Chinese Institute of Engineers, Vol. 28, No. 6, 2005, págs. 925-933.

[7] A. Hemami, M. G. Mehrabi, R. M. H. Cheng, *Synthesis of an Optimal Control Law for Path Tracking in Mobile Robots*, Automatica, Vol. 28, No. 2, 1992, págs. 383-387.

[8] F. Beer, R. Johnston, W. Clausen, *Mecánica Vectorial para Ingenieros: Dinámica*, McGraw Hill, 8° Ed., 2007.

[9] H. Kwakernaak, R. Sivan, *Linear Optimal Control Systems*, Wiley-Interscience, 1972, págs. 24-65.

- [10] V. Dragan, A. Halanay, *Stabilization of Linear Systems*, Birkhäuser, 1999, págs. 19-40.
- [11] E. Cerdá, *Optimización Dinámica*, Prentice Hall, 2001, págs. 109-132, 249-252.
- [12] D. E. Kirk, *Optimal Control Theory*, Prentice Hall, 1970.
- [13] S. Cetinkunt, *Mechatronics*, John Wiley & SONS INC., 2007.
- [14] P. Niamsup, V. N. Phat, *H_∞ control problem of linear time-varying systems via ontrollability approach*, Vietnam Journal of Mathematics.

IX. AUTORES

Dra. María Montserrat Morín Castillo obtuvo su grado de Doctor en Ciencias Matemáticas con especialidad en Modelación Matemática por la Facultad de Físico-Matemáticas de la Benemérita Universidad Autónoma de Puebla (BUAP).

Ing. Lucio Hernández Angulo es Ingeniero en Mecatrónica por la Universidad Popular Autónoma del Estado de Puebla (UPAEP), actualmente realiza estudios para obtener el grado de Maestro en Ciencias de la Electrónica en la Benemérita Universidad Autónoma de Puebla (BUAP).

Dr. José Eligio Moisés Gutiérrez Arias obtuvo su grado de Doctor en Ciencias Matemáticas con especialidad en Teoría de Control Óptimo por la Benemérita Universidad Autónoma de Puebla (BUAP).

MINI-ROBOT TETRÁPODO DE INSPIRACIÓN BIOLÓGICA: DISEÑO, PROTOTIPO Y MODELADO

Lorenzana Gutiérrez Américo, Juárez Miramontes Sergio, Muñoz Zermeño Mauricio,
Herrera Espinoza Juan José, Velázquez Guerrero Ramiro

Mecatrónica y Control de Sistemas - Universidad Panamericana,
Josemaría Escrivá de Balaguer 101, Fracc. Rústicos Calpulli,
20290, Aguascalientes, Ags., México

Tel: (449) 9106200, web: www.robotica-up.org, correo electrónico: rvelazquez@ags.up.mx

RESUMEN

Este artículo presenta el diseño, implementación y modelado de un mini-robot tetrápodo tipo cangrejo accionado mediante servomotores. El primer prototipo desarrollado es compacto ($23 \times 33 \times 16 \text{ cm}^3$), ligero (1100 g), de bajo costo de realización (250 USD) y es capaz de desplazarse por diversos terrenos irregulares y confinados. La robustez del diseño y la optimización de su locomoción aseguran la estabilidad del prototipo bajo perturbaciones externas. En primera instancia, se presenta el diseño mecánico del robot usando herramientas de diseño asistido por computadora (CAD). Posteriormente, se describe la implementación física del prototipo así como la arquitectura de su sistema electrónico de control. Finalmente, se presenta un análisis cinemático para caracterizar su locomoción.

Palabras clave: mini-robot móvil, tetrápodo, servomotor, modelado y simulación.

I. INTRODUCCIÓN

La robótica y los sistemas inteligentes, incluyendo sensores, microprocesadores, micro-mecanismos y micro-electrónica, han avanzado al punto en que es ahora posible desarrollar robots móviles inteligentes altamente funcionales para uso en el hogar, el trabajo y lugares públicos. El rango de aplicaciones de los robots móviles es amplio: desde realizar trabajos

repetitivos o peligrosos para el hombre hasta simplemente educar o entretener.

Los sistemas de locomoción para robots móviles se pueden dividir básicamente en robots con ruedas y robots con patas.

Los sistemas de locomoción basados en ruedas o cadenas son relativamente simples, confiables, rápidos y eficientes. Desde el punto de vista de la movilidad, los vehículos con ruedas ofrecen la mejor solución cuando se requiere operar sobre terreno agreste. En cualquier condición, los vehículos con ruedas ofrecen el mejor desempeño en términos de velocidad, capacidad de carga, economía de potencia y rango de operación. Las principales líneas de investigación en este campo son la optimización de la maniobrabilidad del vehículo y el diseño de sistemas de suspensión [1]-[2].

Los robots con patas ofrecen ciertas ventajas cuando se trata de desplazarse sobre terrenos particularmente difíciles. La principal característica de las máquinas caminadoras es su capacidad de adaptar su postura para sortear las discontinuidades del terreno. Comparados con los robots con ruedas, los robots con patas son articulados y en consecuencia mecánicamente más complejos. Sin duda, requieren de un control más elaborado. La principal línea de investigación en robots con patas concierne el control de las estructuras cinemáticas para asegurar la postura y el equilibrio [3]-[4]. El gran reto para los robots con patas es conseguir un desempeño similar al de los robots con ruedas.

Desde hace algunos años, los investigadores en robots móviles se han dado cuenta de las ventajas de utilizar principios biológicos para el diseño de robots con patas. Inspirados en insectos o animales (Fig. 1), se han desarrollado un sinnúmero de prototipos de robots tetrápodos, hexápodos y octópodos [5], [6], [7].

Este artículo presenta el diseño y primer prototipo de un mini-robot tetrápodo tipo cangrejo desarrollado en la Universidad Panamericana. Este prototipo presenta ciertas ventajas: es compacto, ligero, de bajo consumo de potencia, bajo costo de realización y es capaz de desplazarse a velocidad razonable por diversos terrenos irregulares y discontinuos.

El resto del artículo está organizado de la siguiente manera: la Sección 2 presenta el diseño de un mini-robot tetrápodo mientras que la Sección 3 describe su implementación física y el primer prototipo desarrollado. La Sección 4 presenta el modelo matemático que describe el comportamiento dinámico del robot. Finalmente, la Sección 5 resume los conceptos principales y perspectivas de trabajo futuro.

II. DISEÑO

Para robots caminadores, se ha demostrado que a mayor número de patas, mayor velocidad y mejor estabilidad durante la locomoción [8], [9]. Sin embargo la complejidad del diseño mecánico y del control también se incrementa.

La locomoción tetrápoda ha demostrado ser la más simple manteniendo las características de omni-direccionalidad, adaptación al terreno y estabilidad tanto estática como dinámica.

La fig. 2(a) muestra una vista general del diseño completo del mini-robot tetrápodo propuesto en este artículo. Esta estructura está básicamente constituida por un cuerpo y 4 patas. Cada pata está conformada por 4 eslabones, unidos entre ellos por juntas rotacionales operadas por 2 servomotores. Cada pata tiene 2 grados de libertad (GdL) dando un total de 8 GdL a todo el mini-robot (Fig. 2(b)).

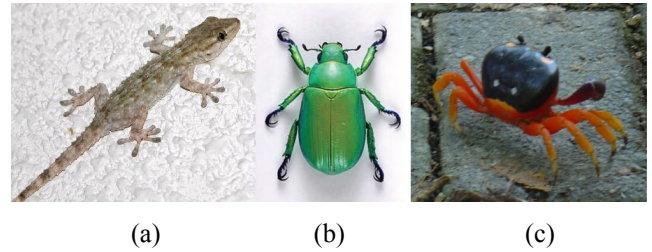
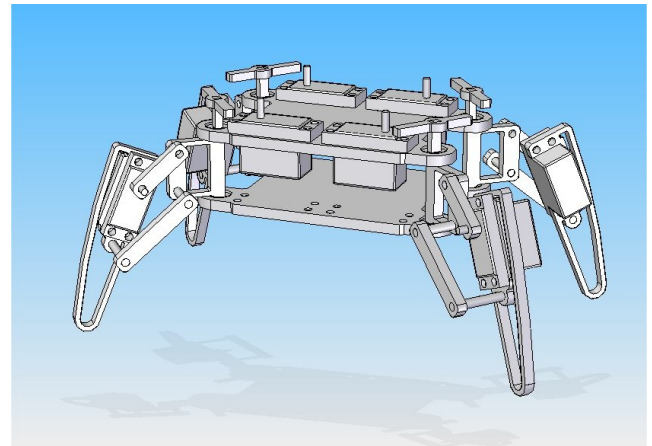
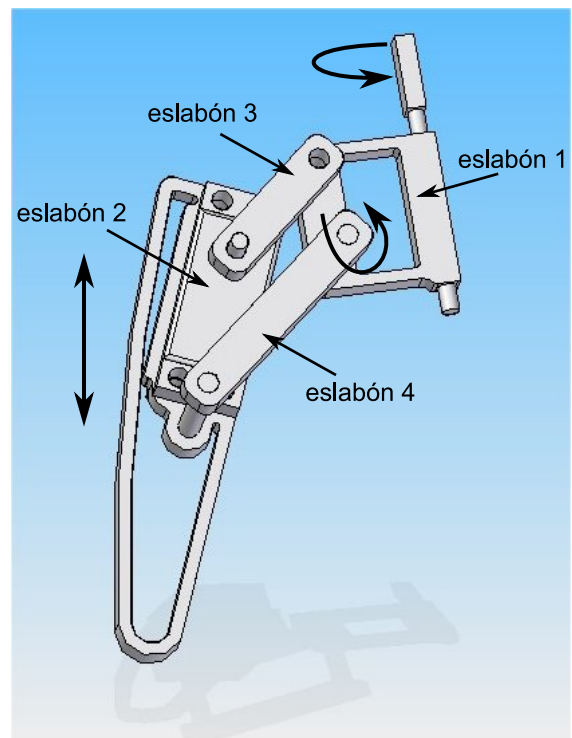


Fig. 1. Ejemplo de animales/insectos (a) tetrápodos, (b) hexápodos y (c) octópodos.



(a)



(b)

Fig. 2. Diseño mecánico: (a) mini-robot tetrápodo y (b) pata de 2 GdL.

Una plataforma cuadrada o marco, que constituye la base estructural del tetrápodo, incorpora las 4 patas y los dispositivos electrónicos de control. Las patas están dispuestas en cada esquina de la estructura de forma simétrica de manera que el centro de gravedad se encuentre en todo momento en el eje central del cuerpo y que los esfuerzos realizados por cada pata se distribuyan uniformemente.

La locomoción tetrápoda, desde los primeros anfibios, se basa en sostener el peso del cuerpo siempre sobre 2 puntos de apoyo diametralmente opuestos. Esto es: pata anterior derecha y pata posterior izquierda o pata anterior izquierda y pata posterior derecha.

De inspiración biológica, este diseño usa la rotación del eslabón 1 y el movimiento arriba/abajo del eslabón 2 de la pata (Fig. 2) para posicionar las patas en el proceso de locomoción.

El proceso de locomoción del tetrápodo se muestra en la fig. 3. En la posición inicial (1) las 4 patas están en contacto con el suelo. (2) Para dar un paso, las patas *a* y *d* permanecen en contacto con el suelo mientras que *b* y *c* se levantan e impulsan el cuerpo. (3) El siguiente movimiento es inverso al anterior: *b* y *c* sostienen al cuerpo y *a* y *d* se levantan e impulsan el cuerpo.

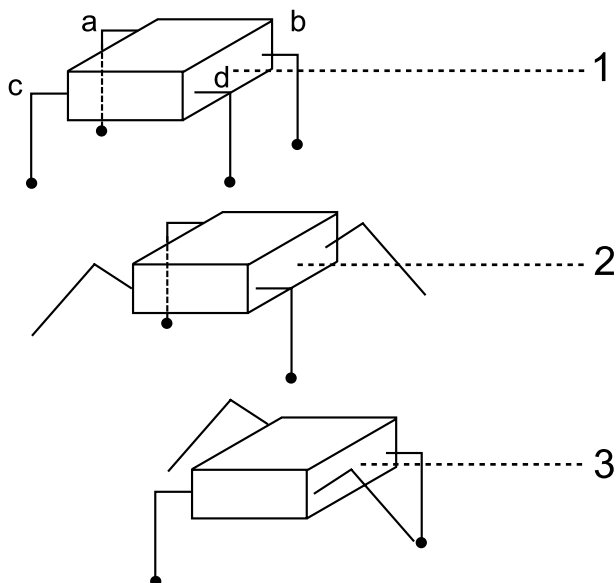


Fig. 3. Secuencia de locomoción de un tetrápodo.

III. PROTOTIPO

El diseño conceptual de mini-robot tetrápodo de la fig. 2(a) fue implementado enteramente en aluminio lo que permite tener una estructura ligera, robusta y de bajo costo. Todas las piezas fueron cortadas con máquina láser de precisión y soldadas con TIG (Tungsten Inert Gas).

La fig. 4 muestra la pata del robot con 2 servomotores incorporados a su estructura para el accionamiento de los 4 eslabones de la fig. 2(b). Note que los servomotores son compactos, se adaptan fácilmente a la estructura y no incrementan significativamente el volumen del prototipo. Cada pata se eleva 3 cm de la superficie de contacto lo que da, dentro de las dimensiones del prototipo, la posibilidad de desplazarse sobre distintos obstáculos y superficies irregulares.

La fig. 5 muestra el ensamble final. Este primer prototipo es de dimensiones compactas: $23 \times 33 \times 16 \text{ cm}^3$, ligero: 1100 g y de bajo costo de realización: 250 USD. Cuenta con un total de 8 servomotores que accionan los 8 GdL del tetrápodo y proporcionan una locomoción estable a una velocidad máxima de 0.076 m/s (i.e. 1 m en 13 s). Cada servomotor requiere 1.1 W de la fuente de alimentación para operar a esta velocidad. Como fuente de alimentación se prefirió una batería compacta que permitiera ser instalada a bordo del mini-robot. Actualmente una pila de 7.4 V a 800 mA-h permite una operación continua de 32 min.

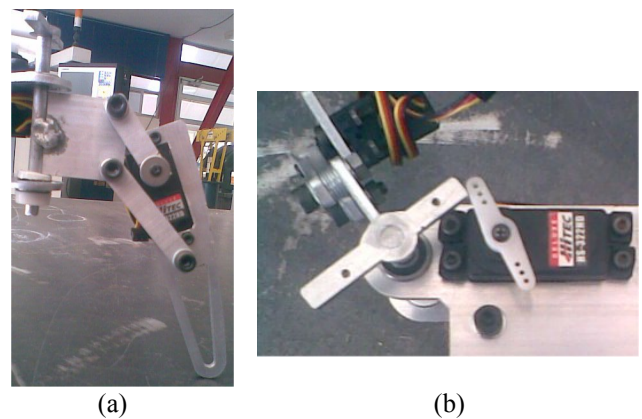


Fig. 4. Estructura y accionamiento de los 2 eslabones de la pata del tetrápodo.



Fig. 5. Prototipo de mini-robot tetrápodo con servomotores.

El sistema de control del mini-robot tetrápodo se muestra en la fig. 6. Éste está constituido por una PC encargada de enviar comandos de dirección al robot (avance, retroceso, giro y alto total). La conexión entre la PC y el prototipo es inalámbrica; mediante un modulo de radio frecuencia X-Bee es posible transmitir datos a 250 kbps [10]. Los datos recibidos se procesan en un microcontrolador, el cual genera una señal PWM (Pulse-width modulation) que se acondiciona en una etapa de potencia para controlar los servomotores en posición y velocidad. Note que un solo microcontrolador gobierna el movimiento de todas las patas del robot lo que permite, además de conservar el sistema compacto y a bajo costo, simplificar el diseño evitando sincronización y multiplexeo entre varios micro-controladores.

Por otro lado, también es posible programar en el microcontrolador una rutina específica a seguir. En este caso el modulo transmisor-receptor X-Bee está deshabilitado.

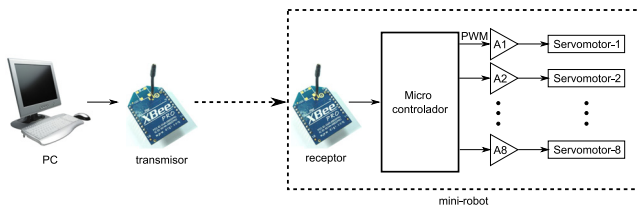


Fig. 6. Arquitectura del sistema electrónico de control del mini-robot tetrápodo.

IV. MODELADO Y SIMULACIÓN

Consideremos un sistema de locomoción tetrápodo con 4 patas idénticas distribuidas simétricamente a lo largo de ambos lados del cuerpo (Fig. 7(a)).

Cada pata esta constituida básicamente por una cadera y una rodilla con 2 eslabones {d1, d2} y 2 juntas rotacionales {j1, j2}. El movimiento de la pata puede describirse por un sistema universal de coordenadas como lo muestra la fig. 7(b).

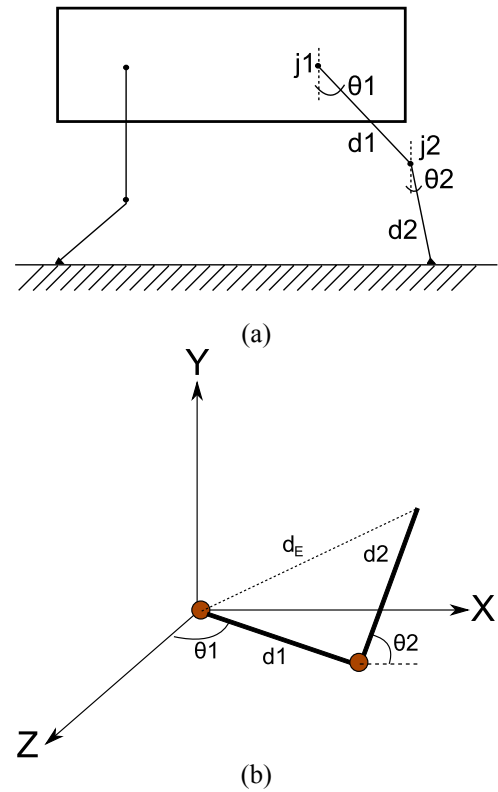


Fig. 7. (a) Modelo cinemático general de un robot tetrápodo y (b) esquemático del mecanismo de la pata.

Aquí, d_E es el vector resultante que indica la posición de la punta de la pata respecto al origen del plano coordenado. Vectorialmente, d_E y sus componentes pueden ser escritos como:

$$\begin{aligned} \vec{d}_E &= \vec{d}_1 + \vec{d}_2 \\ x_E &= d_1 \cos \theta_1 + d_2 \sin \theta_2 \\ y_E &= d_2 \cos \theta_2 \\ z_E &= d_1 \sin \theta_1 \end{aligned} \tag{1}$$

Para el prototipo de la fig. 5, la tabla 1 especifica los parámetros de la pata obtenidos por simple medición.

Parámetro	Valor
$\theta_{1_{max}}$	55°
$\theta_{2_{max}}$	15°
d1	4 cm
d2	11.5 cm

Tabla 1. Resumen de los parámetros de la pata del prototipo de mini-robot tetrápodo.

Las ecuaciones en (1) permiten obtener el movimiento o trayectoria de la pata para $0 \leq \theta_1 \leq 55^\circ$, $0 \leq \theta_2 \leq 15^\circ$ (Fig. 8).

Las ecuaciones en (2) muestran los resultados de la primera derivada respecto al tiempo de los tres componentes de d_E . Estas ecuaciones son las expresiones de la velocidad de la pata.

$$\begin{aligned} x'_E &= -\theta'_1 d_1 \text{sen}\theta_1 + \theta'_2 d_2 \text{cos}\theta_2 \\ y'_E &= -\theta'_2 d_2 \text{sen}\theta_2 \\ z'_E &= \theta'_1 d_1 \text{cos}\theta_1 \end{aligned} \quad (2)$$

Aquí, θ'_1 y θ'_2 son evidentemente las velocidades angulares producidas por los 2 servomotores sobre los 2 eslabones del mecanismo. La fig. 9 muestra por ejemplo la velocidad lineal resultante en x de la pata a 3 velocidades distintas del servomotor. Note que la velocidad máxima de 3340 rpm en ambos servomotores permite tener una frecuencia de operación de la pata de 1 Hz.

En el contexto de simulación también es útil analizar la aceleración experimentada por la pata. Las ecuaciones en (3) muestran las expresiones de aceleración para las 3 componentes espaciales de movimiento obtenidas de la primera derivada respecto al tiempo de las ecuaciones en (2).

$$\begin{aligned} x''_E &= -\theta''_1 d_1 \text{cos}\theta_1 - \theta''_1 d_1 \text{sen}\theta_1 - \theta''_2 d_2 \text{sen}\theta_2 + \theta''_2 d_2 \text{cos}\theta_2 \\ y''_E &= -\theta''_2 d_2 \text{cos}\theta_2 - \theta''_2 d_2 \text{sen}\theta_2 \\ z''_E &= -\theta''_1 d_1 \text{sen}\theta_1 + \theta''_1 d_1 \text{cos}\theta_1 \end{aligned} \quad (3)$$

La fig. 10 muestra la aceleración resultante experimentada por la pata en el plano x.

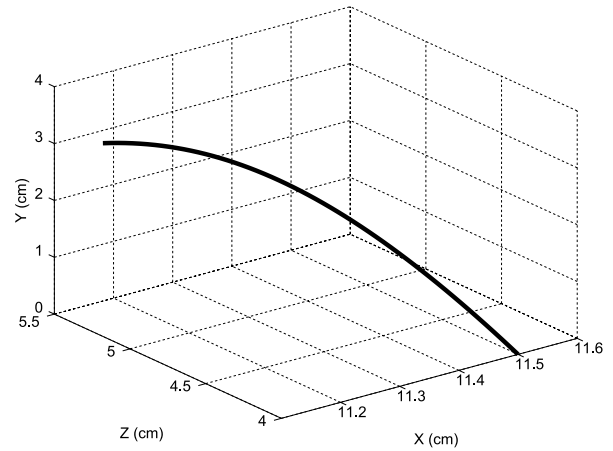


Fig. 8. Trayectoria de la pata del mini-robot tetrápodo en el espacio XYZ.

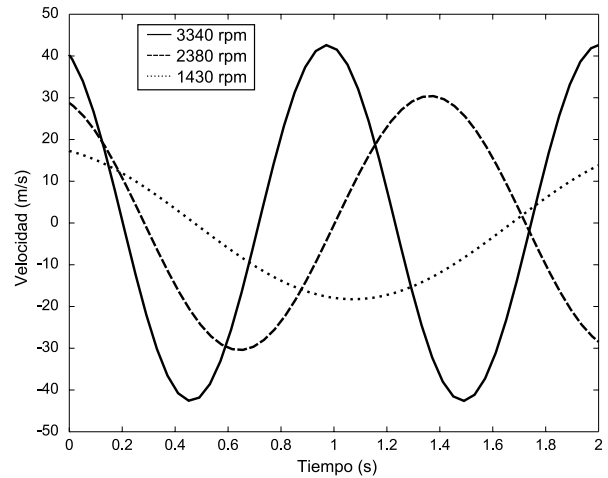


Fig. 9. Velocidad lineal en x experimentada por la pata a diferentes velocidades del servomotor.

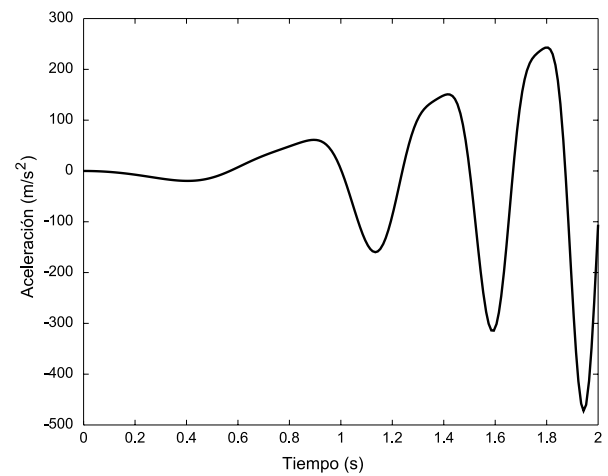


Fig. 10. Aceleración lineal en x de la pata durante la simulación a aceleración constante.

Note que la aceleración no es de tipo senoidal lo que sugiere que la pata no presenta movimiento armónico simple.

V. CONCLUSIÓN

Este artículo presentó el diseño, implementación y modelado de una plataforma móvil miniatura de bajo costo basada en un robot caminador de 4 patas. El diseño propuesto permite gran omni-direccionalidad, robustez y estabilidad para desplazarse sobre obstáculos, superficies y terrenos irregulares.

El uso de servomotores permite una gran eficiencia tanto de locomoción como energética, al brindar la potencia necesaria para accionar los eslabones del robot requiriendo un bajo consumo de energía. Para optimizar su operación, se han realizado pruebas de locomoción para definir las configuraciones y secuencias de las patas que permiten una mayor velocidad y menor gasto energético.

Se ha diseñado un controlador electrónico sencillo, compacto y ligero que permite ser integrado a bordo del mini-robot. Este controlador permite la operación tanto remota como programada del prototipo. A la vez, una fuente de alimentación compacta e igualmente integrada a la estructura del robot permite una autonomía de alrededor de 32 min.

Como perspectivas de trabajo futuro se pretende incorporar una videocámara a la estructura del mini-robot que permita obtener imágenes para fines de reconocimiento, exploración y vigilancia.

VI. REFERENCIAS

- [1] E. Rollins, J. Luntz, A. Foessel, B. Shamah and W. Whittaker, “*Nomad: a demonstration of the transforming chassis*”, Proc. of IEEE International Conference on Robotics and Automation, 1998.
- [2] R. Volpe, “*Rocky 7: A next generation mars rover prototype*”, Journal of Advanced Robotics, 11(4), 1997.
- [3] R.B. McGhee, “*Finite state control of quadruped locomotion*”, Proc. of International Symposium on External Control of Human Extremities, 1996.
- [4] C. Grand, F. BenAmar, F. Plumet and P. Bidaud, “*Stability and traction optimization of reconfigurable vehicles: application to an hybrid wheel-legged robot*”, The International Journal of Robotics Research, 23, 2004.
- [5] E. Calle, I. Ávila, J. Zambrano, “*Diseño e implementación de un robot móvil cuadrúpedo*”, Revista Tecnológica de la ESPOL, 20(1), 2007.
- [6] B. Han, Q. Luo, X. Zhao and Q. Wang, “*Research on hexapod walking bio-robot's workspace and flexibility*”, Bioinspiration and Robotics: Walking and Climbing Robots, 2007.
- [7] H. Ma, L. Wang, D. Chen, X. Hao and H. Luo, “*Design of a crab-like octopod robot*”, Applied Mechanics and Materials, 10, 2007.
- [8] S. Hirose and K. Kato, “*Quadruped walking robot to perform mine detection and removal task*”, Proc. of 1st International Conference on Climbing and Walking Robots, 1998.
- [9] J.A. Gálvez, J. Estremera, P. Gonzalez de Santos, “*SIL04: a versatile quadruped robot for research in force distribution*”, Proc. of 3rd International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines, 2000.
- [10] Digi International Inc., *X-Bee datasheet*. Información actualizada disponible en: <http://www.digi.com/>

DISEÑO DE UN CONTROL ÓPTIMO DISCRETO PARA UN ROBOT MÓVIL

Gutiérrez Arias J. E. Moisés, Galindo Téllez Aline.

Facultad de Ciencias de la Electrónica, Maestría con Opción en Automatización
Benemérita Universidad Autónoma de Puebla
Av. San Claudio y 18 Sur, Col. Jardines de San Manuel, Puebla, Pue. 72570, México
Tel. (222)2295500, correo electrónico: jmgutierrez@ece.buap.mx, galindotellez@yahoo.com.mx

RESUMEN

Presentamos el modelo dinámico de un robot móvil provisto de dos ruedas activas y una pasiva, el principio del máximo discreto de Pontryagin es el fundamento para sintetizar la ley de control. Damos la descripción de las trayectorias deseadas mediante las cuales podemos alcanzar cualquier trayectoria real. Determinadas las ecuaciones lineales en las desviaciones, procedemos a discretizarlas para aplicar el principio del máximo. La solución de un sistema de ecuaciones en diferencias proporciona la ley de control óptima, objetivo principal de este trabajo. Presentamos la simulación de dicha ley de control.

Palabras clave: Robot móvil, modelo dinámico, linealización, discretización, principio del máximo, problema de optimización, control óptimo, simulación por computadora.

I. INTRODUCCIÓN

Los sistemas actuales de ingeniería, tienen una tendencia hacia una mayor complejidad debido a los requerimientos de tareas complicadas y de elevada precisión. En cualquier sistema que cuente con un control, siempre es deseable mejorar al máximo el desempeño de dicho proceso, es decir que se convierte en un problema de optimización, donde se trata entonces de tomar la mejor decisión para maximizar o minimizar algún criterio de desempeño.

Actualmente, la mayoría de los robots móviles que se encuentran en operación son experimentales; sin embargo se están empezando a desarrollar para configuraciones industriales,

donde las condiciones de trabajo pueden ser inhóspitas, peligrosas o inaccesibles para los seres humanos.

Anteriormente, los sistemas mecatrónicos y de robótica solo contaban con sistemas de control en tiempo continuo; en la actualidad, se tratan diferentes técnicas de control y de optimización para sistemas discretos (digitales). Lo anterior implica una gran ventaja por aplicación directa que se tiene con los sistemas digitales y por la versatilidad que dichos sistemas tiene sobre los sistemas analógicos.

La mayor aportación del presente artículo es la obtención de un control para un sistema de con cinco ecuaciones de movimiento, lo cual representa la solución de un sistema de diez ecuaciones en diferencias simultáneas para mejorar el control de un robot. Es decir que, aunque la teoría presentada fue desarrollada en los años 60's es ahora cuando podemos retomarla, ya que se cuentan con poderosas herramientas de Software y que pueden ser directamente implementadas en el Hardware actual, además que se puede implementar en cualquier sistema actual de control donde sea requerida una optimización.

II. ECUACIONES DINÁMICAS DEL ROBOT MÓVIL

Consideramos la clase de robots móviles autónomos que corresponden al diseño de control o conducción diferencial, consta de dos ruedas de tracción y una rueda pasiva. Cuenta con restricciones no-holonómicas, que aparecen como consecuencia de la hipótesis de no deslizamiento.

Las velocidades del centro de las ruedas son denotadas como v_r y v_l (Ver Fig. 1).

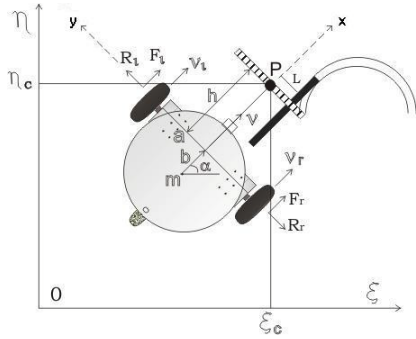


Figura 1. Diagrama del sistema coordinado del robot móvil.

La posición del robot con respecto al sistema de referencia fijo está dada por las siguientes relaciones:

$$\begin{aligned} \dot{\xi} &= v \cos \alpha - h\omega \sin \alpha \\ \dot{\eta} &= v \sin \alpha - h\omega \cos \alpha \\ \dot{\alpha} &= \omega \end{aligned} \quad (1)$$

Para obtener las ecuaciones dinámicas, se sustituyen las restricciones no-holonómicas por las fuerzas reactivas \bar{R}_l y \bar{R}_r . Se tienen también las fuerzas de empuje \bar{F}_l y \bar{F}_r . Se describen las ecuaciones en el sistema no fijo mediante los teoremas principales de la mecánica, obteniendo así las siguientes relaciones dinámicas:

$$\begin{aligned} m(\dot{v} + b\omega^2) &= F_r + F_l, \\ \tilde{J}\dot{\omega} + mb\omega v &= (F_r - F_l)a. \end{aligned} \quad (2)$$

Si sustituimos las fuerzas activas por los torques de los motores con la relación:

$$\tau = F\rho, \quad (3)$$

donde ρ es el radio de las ruedas, y τ es el torque del motor. También se asocia el modelo más simple de un motor:

$$\tau = \chi u - \sigma\dot{\phi}, \quad (4)$$

donde $\sigma\dot{\phi}$ es la suma de la fricción viscosa y la fuerza contra electromotriz. Entonces, de (3) se despeja la fuerza y se sustituye (4) para cada rueda, y sabiendo la relación:

$$\dot{\phi} = \frac{v}{\rho}; \quad (5)$$

se puede sustituir directamente en (2), para así obtener las ecuaciones dinámicas en función del voltaje. Finalmente, se obtienen las ecuaciones de movimiento:

$$\begin{cases} \dot{\xi} = v \cos \alpha - h\omega \sin \alpha, \\ \dot{\eta} = v \sin \alpha - h\omega \cos \alpha, \\ \dot{\alpha} = \omega, \\ m\dot{v} = -mb\omega^2 - \frac{2\sigma}{\rho^2} + \frac{\chi}{\rho}(u_r - u_l), \\ \tilde{J}\dot{\omega} = mb\omega v - \frac{2a^2\omega\sigma}{\rho^2} + \frac{\chi a}{\rho}(u_r + u_l), \end{cases} \quad (6)$$

que permiten simular la dinámica del robot.

III. TRAYECTORIAS DESEADAS.

Se puede intuir que cualquier trayectoria real para el robot puede ser descrita por un conjunto de líneas rectas y curvas. Consideremos las trayectorias posibles ilustradas en la Fig. 2,

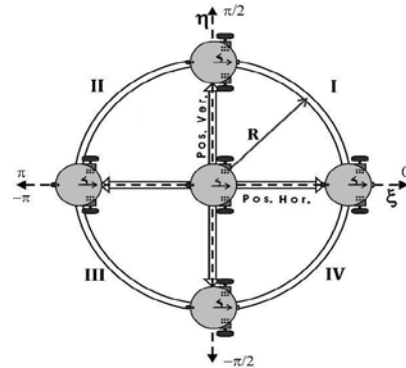


Figura 2. Diagrama del sistema coordinado del robot móvil.

donde se observa que el robot puede moverse por una línea horizontal o vertical en sus dos sentidos, en donde se cuentan con 4 trayectorias, son descritas en la Tabla 1.

Trayectoria	ξ^d	η^d	α^d	v^d	ω^d
1. Línea horizontal sentido negativo	$v_0 t + \xi_0$	0	0	v_0	0

2. Línea horizontal sentido positivo	$v_0 t + \xi_0$	0	π	v_0	0
3. Línea vertical sentido negativo	0	$v_0 t + \eta_0$	$\frac{\pi}{2}$	v_0	0
4. Línea vertical sentido positivo	0	$v_0 t + \eta_0$	$-\frac{\pi}{2}$	v_0	0

Tabla 1. Trayectorias programadas para líneas.

Además, también es posible que se mueva en trayectoria de semicírculo, en donde se tienen ocho posibles configuraciones, como podemos observar en la tabla 2.

Trayectoria	ξ^d	η^d	α^d	v^d	ω^d
5. Cuadrante I (dirección reloj)	$R \cos \varphi$	$R \sin \varphi$	$-\omega_0 t$	$\omega_0 R$	ω_0
6. Cuadrante I (contra reloj)	$R \cos \varphi$	$R \sin \varphi$	$\omega_0 t + \frac{\pi}{2}$	$\omega_0 R$	ω_0
7. Cuadrante II (dirección reloj)	$R \cos \varphi$	$R \sin \varphi$	$-\omega_0 t + \frac{\pi}{2}$	$\omega_0 R$	ω_0
8. Cuadrante II (contra reloj)	$R \cos \varphi$	$R \sin \varphi$	$\omega_0 t + \pi$	$\omega_0 R$	ω_0
9. Cuadrante III (dirección reloj)	$R \cos \varphi$	$R \sin \varphi$	$-\omega_0 t + \pi$	$\omega_0 R$	ω_0
10. Cuadrante III (contra reloj)	$R \cos \varphi$	$R \sin \varphi$	$\omega_0 t - \frac{\pi}{2}$	$\omega_0 R$	ω_0
11. Cuadrante IV (dirección reloj)	$R \cos \varphi$	$R \sin \varphi$	$-\omega_0 t - \frac{\pi}{2}$	$\omega_0 R$	ω_0
12. Cuadrante IV (contra reloj)	$R \cos \varphi$	$R \sin \varphi$	$\omega_0 t$	$\omega_0 R$	ω_0

Tabla 2. Trayectorias programadas para curva.

IV. LINEALIZACIÓN.

Si $u^d(t)$ es una entrada dada para el sistema (6) y, $y^d(t)$ es una trayectoria deseada, se encuentran aproximaciones a las soluciones vecinas. Utilizando el proceso de linealización con las trayectorias programadas, obtenemos los sistemas de ecuaciones lineales en desviaciones:

- Sistema lineal para una trayectoria de línea horizontal en sentido negativo:

$$\begin{aligned} \dot{\xi} &= -v, \\ \dot{\eta} &= -v_0 \alpha - h\omega, \\ \dot{\alpha} &= \omega, \end{aligned} \tag{7}$$

$$\begin{aligned} \dot{v} &= -\frac{2\sigma v}{m\rho^2} + \frac{\chi}{m\rho} (u_r + u_\ell), \\ \dot{\omega} &= \left(\frac{mbv_0}{\tilde{J}} - \frac{2\sigma a^2}{\tilde{J}\rho^2} \right) \omega + \frac{\chi a}{\tilde{J}\rho} (u_r - u_\ell). \end{aligned}$$

V. PRINCIPIO DISCRETO DEL MÁXIMO.

El problema de diseño puede plantearse de la siguiente manera: encontrar el control óptimo $u(k)$ sobre $[0, N]$, tal que el índice de desempeño J , de la ecuación (9) sea mínimo.

$$J = G[x(N), N] + \sum_{k=0}^{N-1} F[x(k), u(k), k] \tag{8}$$

El índice de desempeño debe estar sujeto a la restricción de igualdad:

$$x(k+1) = f[x(k), u(k), k]. \tag{9}$$

Considere el vector $p(k)$ de $n \times 1$ como el vector de *coestado*, que es incorporado a las ecuaciones (8) y (10). El problema de optimización equivale a minimizar:

$$\begin{aligned} J_c &= G[x(N), N] + \sum_{k=0}^{N-1} F[x(k), u(k), k] + \\ &\langle p(k+1), [x(k+1) - f(x, u, k)] \rangle. \end{aligned} \tag{10}$$

Se define la función escalar $H[x(k), u(k), p(k+1), k]$ como el Hamiltoniano, tal que:

$$\begin{aligned} H[x(k), u(k), p(k+1), k] &= F[x(k), u(k), k] + \\ &\langle p(k+1), f(x(k), u(k), k) \rangle. \end{aligned} \tag{11}$$

El Hamiltoniano forma la base del principio del máximo, y debe tener la propiedad de ser máximo a lo largo de una trayectoria óptima.

A continuación, se muestran las ecuaciones que representan la condición necesaria para que J_c sea un extremo:

$$\frac{\partial H(k)}{\partial p(k+1)} = x(k+1), \quad (12)$$

$$\frac{\partial H(k)}{\partial x(k)} = p(k), \quad (13)$$

$$\frac{\partial H(k)}{\partial u(k)} = 0. \quad (14)$$

Las ecuaciones (12) y (13) representan un sistema de $2n$ ecuaciones de diferencias de primer orden, conocidas como las ecuaciones canónicas de estado. La ecuación (14) proporcionará el control óptimo $u(k)$.

VI. DISEÑO DEL CONTROL MEDIANTE EL PRINCIPIO DISCRETO DEL MÁXIMO.

Discretizamos el sistema lineal (7), con la ayuda del Software Wolfram Mathematica 7.0 e introduciendo el algoritmo de diferenciación de Euler, y se presenta a continuación de la forma $x_{k+1} = Ax_k + Bu_k$, con un periodo de muestreo de 0.1 segundos:

- Matrices del sistema discreto correspondiente a la trayectoria de línea:

$$A = \begin{pmatrix} 1 & 0 & 0 & -.0941 & 0 \\ 0 & 1 & -.15 & 0 & -.116 \\ 0 & 0 & 1 & 0 & .464 \\ 0 & 0 & 0 & .9411 & 0 \\ 0 & 0 & 0 & 0 & 4.6407 \end{pmatrix}; \quad (15)$$

$$B = \begin{pmatrix} -.0002 & -.0002 \\ -.002 & .002 \\ .008 & -.008 \\ .002 & .002 \\ .0809 & -.0809 \end{pmatrix}.$$

Se procede a armar la función del Hamiltoniano descrita en (12), donde:

$$F[x(k), u(k), k] = x^T(k)I_{5 \times 5}x(k) + u^T(k)I_{2 \times 2}u(k), \quad (16)$$

para cada uno de los sistemas anteriores.

Posteriormente, se aplica el principio del máximo, donde de la ecuación (11) se debe obtener el sistema original. De las ecuaciones (11) y (12) se obtienen las ecuaciones canónicas, las cuales se presentan a continuación.

- Sistema canónico correspondiente a la trayectoria de línea:

$$\begin{aligned} x_1(k+1) &= x_1(k) - .3809x_4(k) - .000025p_1(k+1) \\ &\quad + .0000525p_4(k+1) \\ x_2(k+1) &= x_2(k) - .75x_3(k) - .1454x_5(k) - .000015876p_2(k+1) \\ &\quad + .00018774p_3(k+1) + .00037548p_5(k+1) \\ x_3(k+1) &= x_3(k) - .171x_5(k) + .00018774p_2(k+1) \\ &\quad - .00022201p_3(k+1) - .00044402p_5(k+1) \\ x_4(k+1) &= .7619x_4(k) - .00011025p_4(k+1) + .0000525p_1(k+1) \\ x_5(k+1) &= -.3421x_5(k) + .00037548p_2(k+1) \\ &\quad - .0000444408p_3(k+1) - .00088804p_5(k+1) \\ p_1(k) &= p_1(k+1) + 2x_1(k) \\ p_2(k) &= p_2(k+1) + 2x_2(k) \\ p_3(k) &= -.75p_2(k+1) + p_3(k+1) + 2x_3(k) \\ p_4(k) &= -.3809p_1(k+1) + .7619p_4(k+1) + 2x_4(k) \\ p_5(k) &= .1454p_2(k+1) + .171p_3(k+1) + .3421p_5(k+1) + 2x_5(k) \end{aligned} \quad (17)$$

Y de (14) se obtuvieron los siguientes controles:

$$u_r = \frac{1}{2} \begin{pmatrix} .0002p_1(k+1) + .002p_2(k+1) - .008p_3(k+1) \\ -.0026p_4(k+1) - .0809p_5(k+1) \end{pmatrix} \quad (18)$$

$$u_\ell = \frac{1}{2} \begin{pmatrix} .0002p_1(k+1) - .002p_2(k+1) + .008p_3(k+1) \\ -.0026p_4(k+1) + .0809p_5(k+1) \end{pmatrix}$$

VII. SOLUCIÓN DEL SISTEMA EN DIFERENCIAS Y SIMULACIÓN DE LA LEY DE CONTROL.

La solución del sistema tratado aquí, fue resuelta con la ayuda del programa realizado en Wolfram Mathematica 7.0, es decir que se simuló la solución del sistema descrito en (18) el cual ya contiene los controles óptimos mostrados en (19). A continuación se presentan las simulaciones del comportamiento de las variables de coestados.

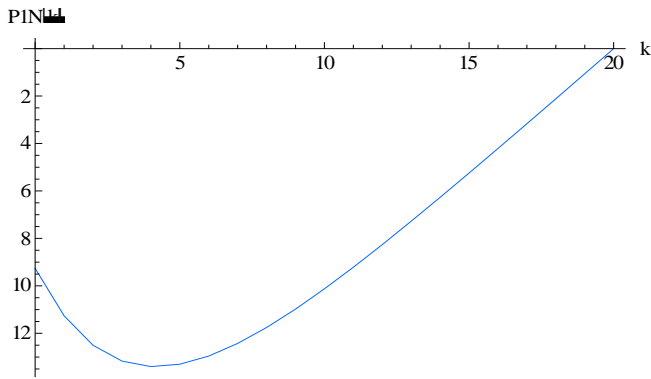


Figura 3. Comportamiento de la primera variable de coestado.

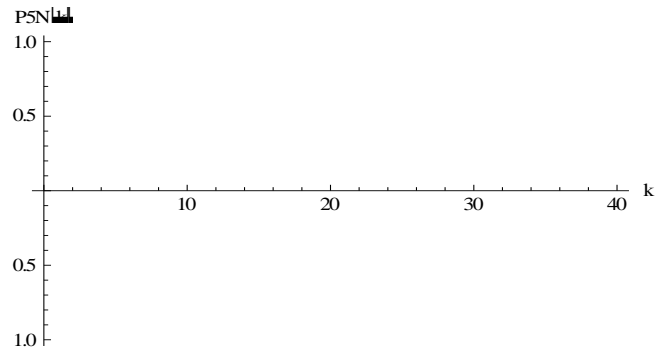


Figura 6. Comportamiento de la cuarta variable de coestado.

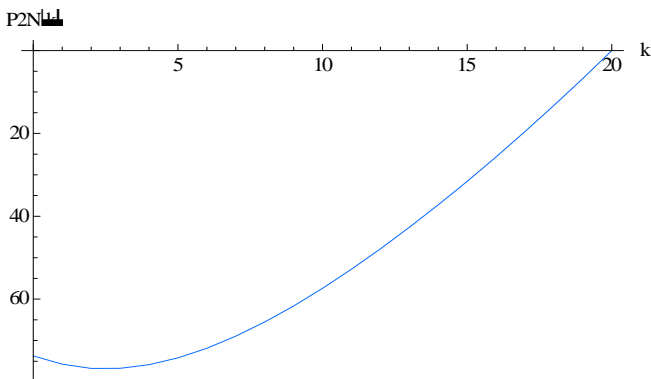


Figura 4. Comportamiento de la segunda variable de coestado.

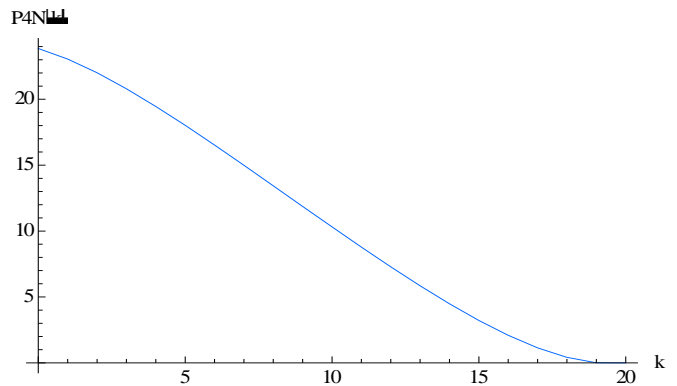


Figura 7. Comportamiento de la quinta variable de coestado.

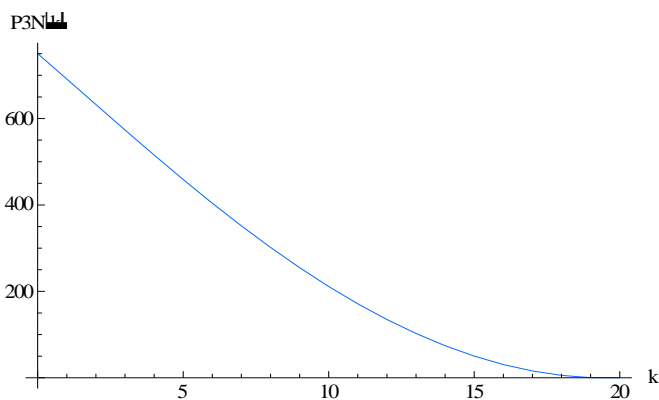


Figura 5. Comportamiento de la tercera variable de coestado.

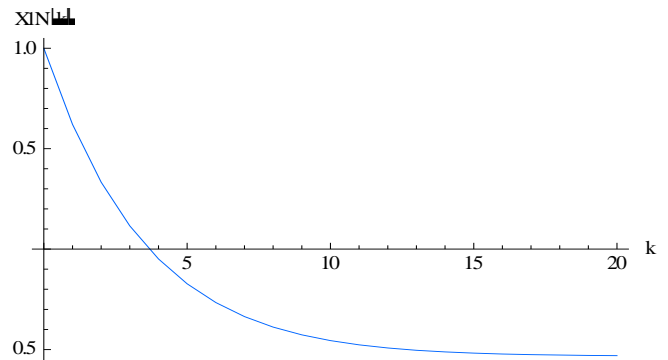


Figura 8. Comportamiento de la primera variable.

Se puede observar en las variables de coestado una clara convergencia a cero en todas las graficas. En la Fig. 6 es posible que la quinta variable de coestado sea igual a cero. A continuación, se presentan las simulaciones del comportamiento de las variables de estado en tiempo discreto.

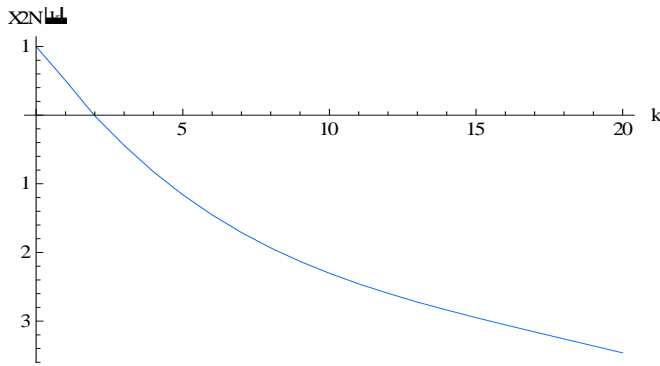


Figura 9. Comportamiento de la segunda variable.

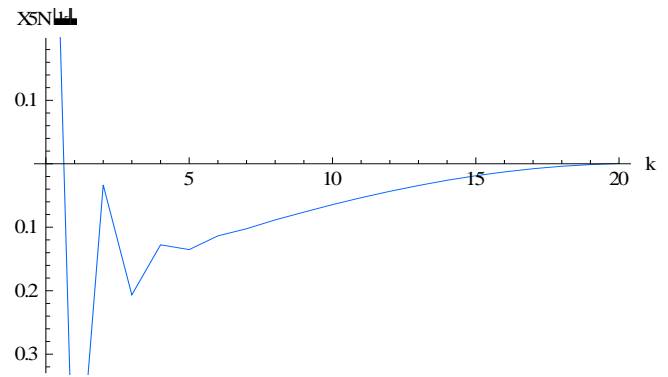


Figura 12. Comportamiento de la quinta variable.

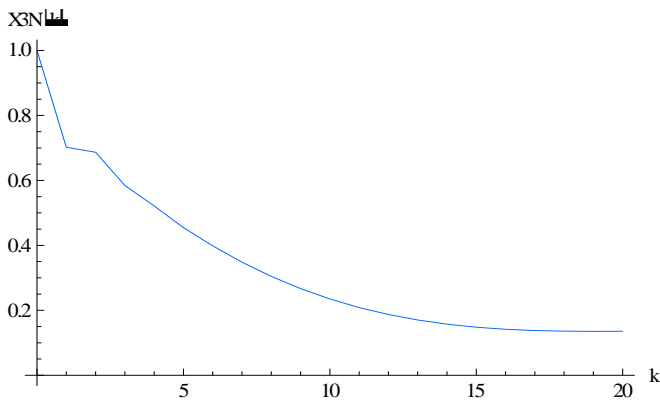


Figura 10. Comportamiento de la tercera variable.

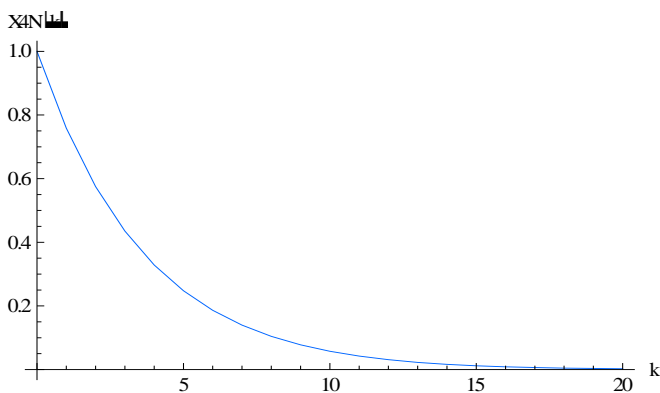


Figura 11. Comportamiento de la cuarta variable.

Se puede observar en las gráficas de la Fig. 8 a la Fig. 12, una clara convergencia de las variables hacia cero, lo cual es lo más deseable y lo esperado; además de la condición establecida de que comiencen en el valor 1.

VIII. CONCLUSIONES.

Se estudiaron las ecuaciones dinámicas del robot móvil, de donde se obtuvieron cinco ecuaciones diferenciales de primer orden. Se linealizó y discretizó el sistema para una trayectoria de línea. Se aplicó el principio del máximo para el sistema de la trayectoria estudiada. Se obtuvieron los controles óptimos para el sistema de línea. Se obtuvieron las simulaciones deseadas que nos muestran el comportamiento del sistema a través del tiempo y que comprueban que la ley de control realiza su función, además de la obtención de las variables de coestados necesaria para su implementación en trabajos futuros. También queda abierta la posibilidad de poder optimizar cualquiera de las trayectorias aquí mencionadas, realizando la misma metodología utilizada.

IX. AGRADECIMIENTOS.

Al CONACYT, por el apoyo brindada y al proyecto

X. REFERENCIAS.

- [1] Donald E. Kirk, *Optimal Control Theory: an Introduction*. Dover Publications (1970).
- [2] R. Vinter, *Optimal Control*. Birkhäuser Boston (2000).

- [3] R. Siegwart, I. Nourbakhsh, *Introduction to Autonomous Mobile Robots*, Springer, 2da edición (2006).
- [4] T. Braunl, *Embedded Robotics*, MIT Press, Cambridge, Mass (2004).
- [5] H. Goldstein, *Classical Mechanics*. Editorial Addison Wesley, Barcelona (2000).
- [6] F. Beer, R. Johnston, W. Clausen, *Mecánica vectorial para Ingenieros: Dinámica*, Mc. Graw Hill, 8va edición (2007)
- [7] L. S. Pontryagin, *The mathematical theory of optimal processes*. Vol. 4. Gordon and Breach Science Publishers. New York (1987).
- [8] F. R. Marotto, *Introduction to mathematical modeling using discrete dynamical systems*, Brooks/Cole Publishing Co. Belmont (2006).
- [9] L. Stephen, *Dynamical systems with applications using mathematica (r)*. Birkhauser Verlag. Boston (2007).

UNA BIBLIOTECA PARA EL MODELADO Y SIMULACIÓN DE LA LOCOMOCIÓN HUMANA EN ROBOTS HUMANOIDES

Orozco Aguirre Héctor Rafael, Ramos Corchado Félix Francisco.

Centro de Investigación y de Estudios Avanzados del I.P.N, Unidad Guadalajara
Av. Científica 1145, Col. El Bajío 45015 Zapopan, Jal., México
Correo electrónico: {horozco, framos}@gdl.cinvestav.mx

RESUMEN

El modelado y simulación de la locomoción humana es una tarea compleja y apasionante que desde sus orígenes siempre ha mantenido la atención de los investigadores en Robótica. En éste artículo se presenta una biblioteca para el modelado y simulación de la locomoción humana. Esta biblioteca proporciona un conjunto de algoritmos útiles en aplicaciones de movimiento en robots humanoides móviles. El modelado y la simulación fueron implementados para el robot humanoide Bioloid de la industria coreana Robotics. Esta biblioteca de locomoción proporciona una serie de algoritmos que son usados para proveer al robot con las habilidades de caminar, de agacharse, de gatear, de incorporarse, de girar a la izquierda y a la derecha, de detectar obstáculos y de detectar sonidos. Estos algoritmos fueron implementados usando el lenguaje Java.

Palabras clave: Robot Bioloid, Humanoide, Locomoción Humana.

I. INTRODUCCIÓN

Actualmente existe un gran deseo de construir robots humanoides que puedan coexistir en armonía con los seres humanos. Éste deseo hace pensar a los investigadores en Robótica y Ciencias Computacionales que la industria de la construcción y el diseño de robots será una industria líder en éste siglo [1]. Este deseo puede ser posible, partiendo del hecho de que la tecnología actual con la que se dispone está casi lista para realizar un sueño de tales dimensiones. Supercomputadoras, microprocesadores mucho más rápidos, sensores de alta precisión, nuevos

avances en técnicas de control, inteligencia artificial y reconocimiento artificial de sonido y visión computacional, todos en conjunto conllevan a un mejor y cada vez más perfecto diseño de máquinas inteligentes que pueden hacer posible éste sueño en un futuro muy cercano.

Hay que tener presente que los robots humanoides no sólo deberían ser capaces de socializar e interactuar con los seres humanos, sino ser capaces de comportarse como tal y moverse como tal [2], además de que puedan ser empleados para asistir y reemplazar a los seres humanos en tareas tediosas, difíciles o peligrosas tales como rescate en situaciones de desastre o exploración del espacio exterior [3].

De hecho, la locomoción representa una de las principales características estudiadas en los robots humanoides por dos grandes razones: primero, un robot humanoide sólo podrá parecerse a un ser humano si es capaz de moverse como tal; segundo, la locomoción proporciona a un robot humanoide la única manera de interactuar con su entorno y con otras entidades tales como otros robots e incluso los seres humanos [4].

Sin embargo, los robots humanoides disponibles comercialmente actualmente están diseñados para realizar movimientos usando un sistema de control de lazo abierto proveyendo al usuario un paradigma simple para crear formas de andar o caminar predeterminadas para múltiples grados de libertad. Por lo general, estos robots no son capaces de moverse sobre terrenos desiguales y es muy difícil o prácticamente imposible que estos

realicen movimientos que requieran de una reacción inmediata a la inestabilidad. Una manera de compensar estas dificultades es incorporando y utilizando grandes pies, así como mantener un bajo centro de masa y proporcionar una mejor absorción de choque o de impacto, lo cual resulta en robots humanoides con poca semejanza física con el ser humano.

Los objetivos a largo plazo, son sin embargo, permitir a los robots humanoides que puedan ser capaces de caminar, correr, saltar y en general realizar toda clase de movimientos y habilidades de movimiento complejas para reaccionar en una manera dinámica y físicamente similar a la del ser humano en situaciones inestables y terrenos desiguales. Estas metas pueden ser alcanzadas por medio del uso y aplicación de sistemas de control de lazo cerrado a las técnicas de control de los servos de los robots. Así, el flujo de datos de entrada debe constar de una multitud de sensores incluyendo posición y torque de los servos, aceleración y momento de inercia.

El presente trabajo está organizado de la siguiente manera: la siguiente sección presenta una discusión clara y detallada del trabajo relacionado sobre los mecanismos de control de robots móviles y sus aplicaciones; la tercera sección describe de manera clara el kit Bioloid que fue el utilizado en éste trabajo; la cuarta sección presenta una aplicación del modelado y de la simulación de la locomoción humana en el robot humanoide Bioloid, al cual le fueron implementados las habilidades de caminar, de agacharse, de gatear, de incorporarse, de girar a la izquierda y a la derecha, de detectar obstáculos y de detectar sonidos; finalmente en la última sección se presentan las conclusiones obtenidas de éste trabajo.

II. TRABAJO RELACIONADO

La Robótica Situada es el estudio de los robots embebidos en complejos y a menudo ambientes dinámicamente cambiantes. La complejidad del problema de control de los robots en estos ambientes está directamente relacionada a que tan impredecible e inestable es el medio ambiente y a la rapidez con la que los robots deben reaccionar

acorde a que tan complejas sean las tareas que deben de realizar [5]. Los robots en estos ambientes deben poseer un cuerpo físico para sentir su entorno y así poder actuar y avanzar en él [6], de esta manera se dice que cada robot está situado en el ambiente. Así, el problema que debe resolverse es el de manipular de manera correcta el cuerpo del robot de una manera rápida, natural y estable, con lo cual el problema de control del robot se vuelve una tarea compleja que a menudo demanda mucho tiempo.

El problema de control de un robot es el proceso de tomar información del ambiente a través de sensores y procesarla tanto como sea necesario para lograr que el robot pueda ser capaz de tomar decisiones, de actuar y de ejecutar acciones en el ambiente. La estructura del ambiente y el tipo de acciones que deba realizar un robot dictan la complejidad del sistema de control que se requiere y que tan dura y laboriosa en su implementación, ya que a menudo en ambientes dinámicamente cambiantes se requiere que los robots reaccionen rápidamente e inteligentemente.

Aunque si bien hay comercialmente disponible una amplia variedad de programas de control de robots, sólo existe por desgracia un conjunto finito y limitado de diferentes clases de metodologías de control de robots, embebidas por lo general en arquitecturas específicas de control de robots. Las cuatro clases fundamentales son las siguientes:

- Control reactivo (no pensar, reaccionar): esta es una técnica de acoplamiento de entradas sensoriales y efectores de salida que permiten a un robot responder rápidamente a ambientes cambiantes no estructurados [7][8]. Este es un buen mecanismo de control a menudo llamado como su equivalente biológico, estímulo-respuesta.
- Control deliberativo (pensar, entonces actuar): aquí el robot usa toda la información sensorial disponible y todo el conocimiento almacenado para razonar acerca de la siguiente acción que debe de

ejecutar [9]. El razonamiento es por lo general un mecanismo de planeación.

- Control híbrido (pensar y actuar independiente en paralelo): esta técnica de control combina los mejores aspectos del control reactivo y del control deliberativo para lograr una mejor respuesta del robot [10].
- Control basado en comportamiento (pensar la manera más adecuada de actuar): éste mecanismo está inspirado biológicamente e intenta modelar y simular la manera en que los seres vivos reaccionan a su entorno en base a su comportamiento [11][12][13][14].

Cada uno de los métodos mencionados anteriormente tiene sus puntos fuertes y débiles, y todos desempeñan importantes roles de éxito en determinados problemas y aplicaciones. Diferentes enfoques pueden ser adecuados tomando en cuenta la naturaleza de las tareas y del ambiente y las capacidades del robot, en términos del hardware y de la computación. Así pues, los mecanismos de control de robots deben de considerar lo siguiente:

- Se debe consumir poco tiempo para pensar las acciones y así actuar rápidamente.
- Pensar de una manera correcta permite mejores planeaciones, pero invertir mucho en esta tarea es poco conveniente.
- Los robots para pensar necesitan disponer de información actual y relevante, pero como el ambiente cambia conforme los robots piensan, se deben tomar en cuenta ciertas consideraciones para evitar que los robots tomen malas decisiones, ya que la información de la que disponían antes de pensar a menudo no es la misma al momento de reaccionar.

Algunos robots que se han diseñado no piensan del todo y ejecutan acciones preprogramadas, mientras que otros piensan demasiado y actúan muy poco. Sin embargo, el robot ASIMO de la empresa Honda [15] es uno de los robots humanoides más avanzados que se han creado hasta el momento. Éste robot humanoide cuenta con voz, visión y

reconocimiento de gestos [16], además de que le fue implementado un avanzado sistema dinámico de control de locomoción que lo hace capaz de caminar, correr, subir escaleras y evitar obstáculos estáticos y dinámicos mientras camina [17][18].

III. KIT BIOLOID

El kit Bioloid [19] es una colección de pequeñas partes robóticas intercambiables. Con estas partes, se pueden construir sofisticados robots de varias formas para que cumplan con tareas en específico. Las partes más importantes que lo componen son las siguientes:

Unidad de Control y Bus de Comunicación (Módulo CM-5):

la placa controladora CM-5 (ver figura 1) está basada en el procesador de 16Mhz Atmega 128 de Atmel. Esta CPU tiene 128 KBytes de memoria flash. La placa controladora viene dentro de un revestimiento con leds y botones que le permiten algunas operaciones activadas por tacto y luces de estado que indican varios estados de los modos del robot.

El Bioloid viene instalado con un firmware predeterminado y un cargador de arranque. Éste firmware permite al software de la PC (editor de movimientos y de comportamiento) operar por interfaz serial. El firmware puede ser reemplazado por software propietario que el usuario puede crear usando herramientas de compilación estándar Atmel.



Figura 1. Módulo CM-5 y su placa controladora interna

El CM-5 también tiene la habilidad de comunicarse inalámbricamente usando una interfaz ZigBee [20], la cual se conecta dentro del CM-5. El Bioloid usa una comunicación serial UART estándar (como la usada en RS232) a 1 mbps. Sin

embargo el cableado del bus serial es tal que sólo una línea es usada tanto para RX y para TX. Esto es también conocido como RS485.

Módulo AX-12+: cada motor servo-actuador Dynamixel AX-12+ (ver figura 2) tiene un micro controlador que entiende 50 comandos, la mayoría de los cuales fijan o leen parámetros que definen su comportamiento.

Cada servo tiene la capacidad de controlar su velocidad, temperatura, posición, tensión y carga soportada. El algoritmo de control utilizado para mantener la posición de cada servo puede ajustarse individualmente, permitiendo un control de retroalimentación de la velocidad y de la carga que soporta cada actuador, logrando así un control sofisticado del robot. Esto tiene aplicaciones por ejemplo en los robots bípedos, ya que sin necesidad de inclinómetros o acelerómetros, se pueden conseguir efectos de equilibrio. El motor tiene la capacidad de hacer giros en un rango limitado de 0 a 300 grados.



Figura 2. Servo-controlador AX-12+

Módulo AX-S1: el Dynamixel AX-S1 (ver figura 3) es un módulo sensor preparado para redes inteligentes que contiene principalmente:

- 3 sensores de dirección infrarrojos (izquierda, derecha y frontal) que pueden ser usados para medir distancia y luminosidad.
- Un micrófono para detectar sonidos.
- Un zumbador electrónico que puede ser usado para reproducir escalas musicales o simplemente un sonido de “beep”.



Figura 3. Módulo AX-S1

Las tablas 1 y 2 muestran las principales especificaciones técnicas del motor AX-12+ y del sensor AX-S1 respectivamente.

Peso	55 g
Relación de transmisión	01/01/54
Voltaje de entrada	7V ~ 10V
Torque máximo	12 ~ 16.5 kgf.cm
Resolución	0.35°
Ángulo de operación	300° y modo sin fin
Voltaje	7V~10 V (recomendado 9.6V)
Corriente máxima	900 mA
Temperatura de operación	-5 ~ +85 °C
Tipo de protocolo	Comunicación serial asincrónica half dúplex (8bit,1stop,No Parity)
Conexión física	TTL Level Multi Drop (daisy chain type Connector)
ID	254 ID (0~253)
Velocidad de comunicación	7343bps ~ 1 Mbps
Retroalimentación	Posición, temperatura, voltaje de entrada, etc.

Tabla 1. Especificaciones técnicas del AX-S12+

Peso	37g
Resolución	10bit (1024)
Voltaje	7V~10V (recomendado 9.6V)
Corriente de alimentación	40 mA
Temperatura de Operación	-5 ~ +85 °C
Tipo de protocolo	Comunicación serial asincrónica half duplex (8bit,1stop,No Parity)
Conexión física	TTL (Transistor-Transistor Logic) Level Multi Drop (daisy chain type Connector)
ID	254 ID (0~253)
Velocidad de comunicación	343bps ~ 1 Mbps
Retroalimentación	Sensor infrarrojo, temperatura, voltaje de entrada, etc.

Tabla 2. Especificaciones técnicas del AX-S1

Módulo edición de comportamientos (BCP): el control del comportamiento es un concepto muy útil en robótica. El kit Bioloid provee un software que permite al programador hacer la transición de estados del robot cuando un estímulo particular o evento lo dispara (ver figura 4).

Una serie de comandos que son introducidos permiten al programador interrogar todas las funcionalidades del AX-12+, del AX-S1 y del módulo CM-5. La accesible funcionalidad dentro del CM-5 incluye entre otras, reproducción de secuencias de movimientos, botones externos que se pueden presionar y un temporizador del CM-5. Los comandos suministrados con el BCP incluyen:

- Comandos de control del programa (START, END).
- Comandos de bifurcación condicional (IF, ELSE IF, ELSE, CONT IF) con operaciones condicionales (=, >, >=, <, <=).
- Comandos de programación secuencial (JUMP, CALL/RETURN).
- Comandos numéricos (COMPUTE).
- Comandos de asignación (LOAD).

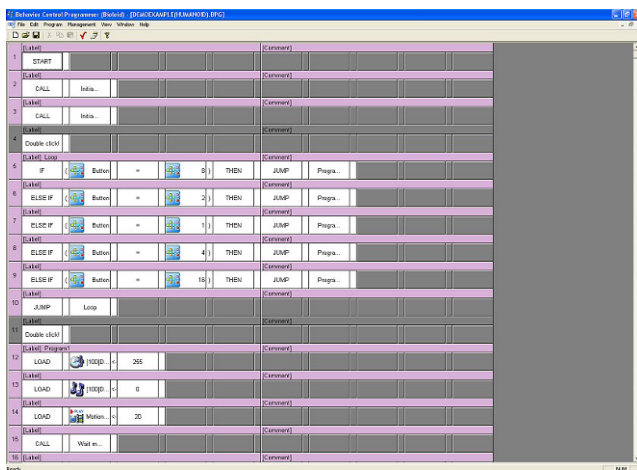


Figura 4. Programa editor de comportamiento del kit Bioloid

Módulo de edición de movimientos (MEP): el kit Bioloid ofrece una GUI para la edición de movimiento (ver figura 5). Esta GUI permite al usuario editar movimientos para robots de múltiples articulaciones (AX-12+). El usuario

puede crear y editar movimientos moviendo manualmente las articulaciones y guardando cada postura generada usando el editor de movimientos. El usuario también puede conectar o repetir movimientos editados.

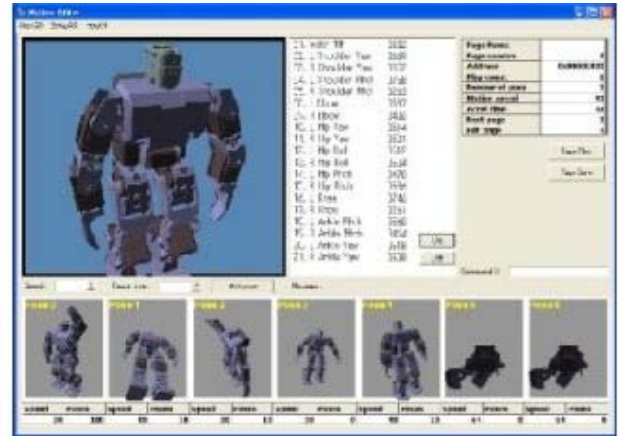


Figura 5. Programa editor de movimientos del kit Bioloid

IV. MODELADO Y SIMULACIÓN DE LA LOCOMOCIÓN HUMANA EN EL ROBOT BIOLOID

En esta sección se expone una biblioteca de movimientos y algoritmos de fácil manejo para controlar al robot humanoide Bioloid [19], con lo cual se sienta una base sobre la cual a futuro se podrán diseñar nuevos movimientos de mayor complejidad. Los principales objetivos para la creación de esta biblioteca son:

- Que el robot Bioloid ejecute con fluidez seis movimientos básicos (caminar, agacharse, gatear, girar a la izquierda y a la derecha e incorporarse) y dos mecanismos de detección, uno de obstáculos y el otro de sonido respectivamente.
- Los movimientos que el robot será capaz de ejecutar deberán ser relativos a un humanoide (imitar los movimientos naturales del ser humano).
- Usarla en aplicaciones de locomoción de robots móviles fomentando así el uso de los robots Bioloid.
- Sentar las bases para incorporar nuevas rutinas de movimiento a futuro.

La biblioteca expuesta en este trabajo fue implementada en lenguaje Java usando el paquete

de clases de la biblioteca BRBrain [21], ya que esta brinda un conjunto de clases sencillo de manejar para manipular los servo-actuadores AX-S12+ y el sensor AX-S1 del robot Bioloid.

Antes de implementar la biblioteca presentada, se hizo un análisis de la cinemática del robot (ver figura 6) para conocer la dependencia entre sus articulaciones (relación padre-hijo) y el eje y sentido de rotación de cada una de ellas (ver figura 7), así como sus valores mínimos y máximos de rotación (ver tabla 3).

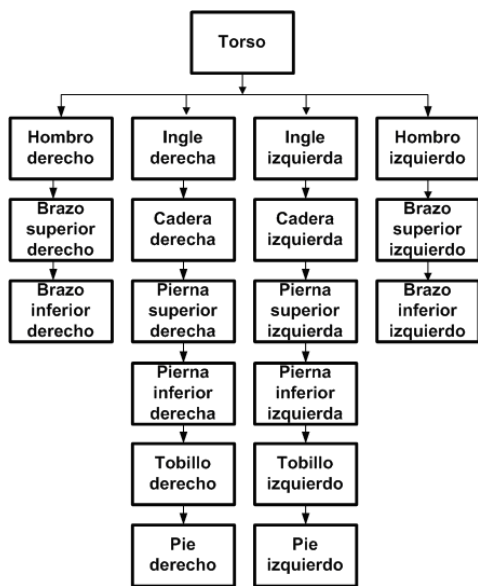


Figura 6. Relación padre-hijo y estructura jerárquica de las articulaciones del robot Bioloid

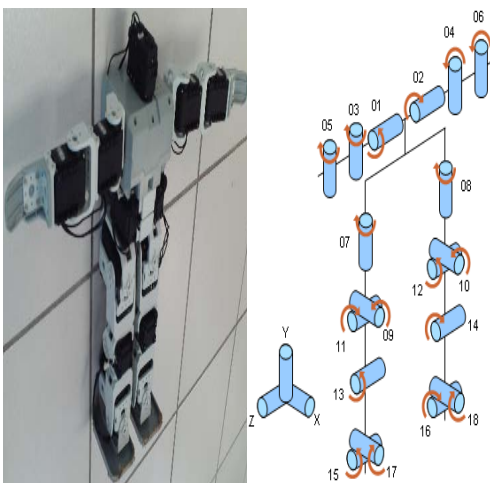


Figura 7. Sentido de rotación de las articulaciones del robot Bioloid

ARTICULACIÓN	EJE DE ROTACIÓN	VALOR MÍNIMO	VALOR MÁXIMO
1. Hombro derecho	z	0°	300°
2. Hombro izquierdo	z	0°	300°
3. Brazo superior derecho	y	50°	250°
4. Brazo superior izquierdo	y	50°	250°
5. Brazo inferior derecho	y	45°	250°
6. Brazo inferior izquierdo	y	45°	250°
7. Ingle derecha	y	0°	300°
8. Ingle izquierda	y	0°	300°
9. Cadera derecha	x	75°	225°
10. Cadera izquierda	x	75°	225°
11. Pierna superior derecha	z	25°	175°
12. Pierna superior izquierda	z	125°	275°
13. Pierna inferior derecha	z	5°	150°
14. Pierna inferior izquierda	z	150°	295°
15. Tobillo derecho	z	120°	250°
16. Tobillo izquierdo	z	45°	175°
17. Pie derecho	x	85°	225°
18. Pie izquierdo	x	85°	225°

Tabla 3. Valores mínimos y máximos para el eje de rotación de cada articulación del robot Bioloid

A continuación se explica el modelado e implementación de cada uno de las habilidades motrices implementadas, aunque cabe mencionar que estas son habilidades prediseñadas estáticamente, es decir siempre van ejecutadas por el robot de la misma manera, pero se tiene pensado diseñar mecanismos para generarlas de manera dinámica acorde a las necesidades del robot y a las condiciones de su entorno. Estas habilidades son:

Habilidad de caminar: para implementarla se partió de una posición neutra inicial (robot de pie y en reposo) y posteriormente se realizó la siguiente secuencia de posturas (ver figura 8):

- Ladear el robot de tal manera que su peso caiga sobre su pierna derecha manteniendo el equilibrio, logrando que la pierna contraria quede al aire.
- Mover hacia delante la pierna izquierda del robot con una ligera inclinación del cuerpo del mismo para lograr el efecto de un paso hacia delante.
- Impulsar la pierna derecha hacia delante para nivelar las dos piernas del robot y

enderezar a la par el cuerpo del robot para llegar a la posición neutra inicial.

- Hasta aquí se ha ejecutado la mitad de la habilidad, a partir de aquí se hace lo mismo pero ahora empezando con la otra pierna del robot.



Figura 8. Habilidad de caminar del robot Bioid

Para lograr un efecto continuo de caminado se efectúa el proceso anterior tantas veces como sea necesario.

Habilidad de agacharse: esta fue una habilidad sencilla de implementar, para hacerlo se partió de igual manera de la posición neutra inicial ya conocida y posteriormente se realizó mediante una serie de interpolaciones la siguiente secuencia de posturas (ver figura 9):

- Agachar secuencial al robot sobre sus piernas de tal manteniendo el equilibrio hasta que el robot quede por completo agachado y mantenerlo un pequeño instante en esa postura.
- Levantar al robot hasta llegar a la posición neutra inicial.



Figura 9. Habilidad de agacharse del robot Bioid

Habilidad de gatear: de igual manera se partió de la posición neutra inicial ya conocida y posteriormente se realizó la siguiente secuencia de posturas (ver figura 10):

- Poner al robot en una posición de cuatro pies y mantenerlo un pequeño instante en esa postura.
- Generar el efecto de gateo mediante un efecto similar al del caminado sin preocuparse tanto por el equilibrio, sino de la fricción sobre el suelo de tal manera que siempre se avance hacia delante y nunca se llegue a retroceder, esto se repite tantas veces como sea necesario.
- Para incorporar al robot se regresa a la posición inicial a partir de la posición de cuatro pies.

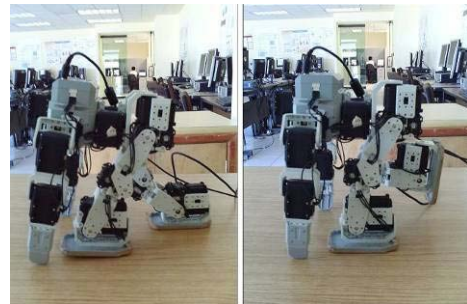


Figura 10. Habilidad de gatear del robot Bioid

Habilidad de incorporarse: para esta habilidad se partió de una posición en reposo inicial neutra con el robot acostado boca abajo y posteriormente se realizó mediante la siguiente secuencia de posturas (ver figura 11):

- Teniendo al robot recostado, se extienden sus brazos y se impulsa hacia arriba el cuerpo del robot manteniendo el equilibrio y se le mantiene un instante en esa postura.
- Se doblan las piernas del robot para llevarlo a una posición a cuatro pies y se le mantiene un pequeño instante en esa postura.
- Para incorporar al robot se le levanta para llevarlo a una posición de pie neutral en reposo.

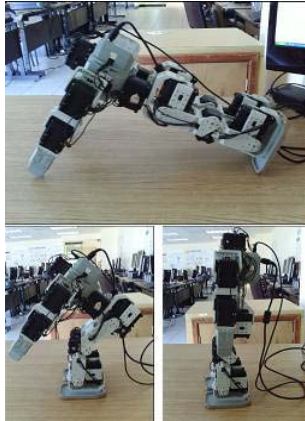


Figura 11. Habilidad de incorporarse del robot Bioloid

Habilidad de girar: para esta habilidad según se requiera que se gire hacia la izquierda o hacia la derecha se partió de igual manera de la posición en reposo inicial neutra con el robot de pie y posteriormente se realizó mediante la siguiente secuencia de posturas (ver figura 12):

- Según el sentido de giro deseado (izquierda o derecha) se inclina un poco al robot sobre la pierna contraria manteniendo el equilibrio.
- Se extienden los brazos del robot.
- Se abre un poco la pierna que está libre hacia el sentido del giro deseado y se la deja descansar sobre el suelo con esto se reparte el peso del robot sobre las dos piernas.
- Se gira un poco el cuerpo del robot en el sentido del giro deseado.
- Se baja el brazo contrario al sentido del giro y se levanta la pierna contraria al sentido del giro manteniendo el equilibrio.
- Se gira la pierna contraria en el sentido del giro deseado y se la deja descansar sobre el suelo manteniendo el equilibrio.
- Se gira por completo el cuerpo del robot en el sentido del giro deseado manteniendo el equilibrio.
- Se baja el otro brazo y se inclina un poco al robot sobre la pierna contraria al sentido del giro manteniendo el equilibrio.
- Por último se regresa al robot a la posición neutral de inicio.

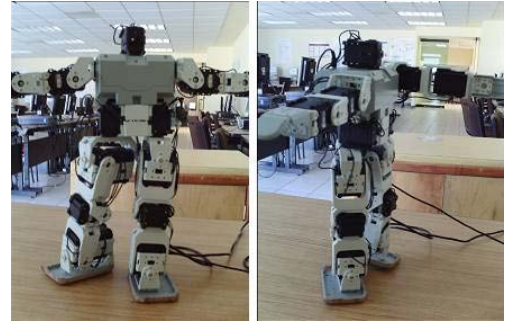


Figura 12. Habilidad de girar del robot Bioloid

Habilidad de detectar obstáculos: para esta habilidad se usó la información sensorial dada por el sensor infrarrojo frontal del AX-S1 para detectar obstáculos que puedan ocasionar colisiones con el robot y se usaron las habilidades de caminar y girar del robot. Para esta habilidad el robot inicia caminando y cuando detecta un obstáculo gira ya sea hacia la izquierda o hacia la derecha y repite el movimiento de caminar y si encuentra de nuevo un obstáculo vuelve a girar y así sucesivamente.

Habilidad de detectar sonidos: para implementar esta habilidad se usó la información dada por el sensor de sonido de la unidad CM-5 del robot sonidos mediante variaciones y repeticiones. Mientras el robot se encuentra caminando el CM-5 está registrando variaciones en los sonidos y sus repeticiones, cuando se detecta un fuerte sonido y una variación que sobrepase un umbral de repetición el robot deja de caminar y permanece un momento detenido y si se registra una estabilidad de nuevo en los sonidos registrados el robot vuelve a caminar, de lo contrario permanece inmóvil. Éste proceso siempre se está llevando a cabo sucesivamente.

V. CONCLUSIONES

En este artículo se presentó una biblioteca implementada en lenguaje Java para proporcionar una serie de habilidades de locomoción (caminar, agacharse, gatear, incorporarse y girar a la izquierda y a la derecha) que simulan la forma de moverse del ser humano y un conjunto de mecanismos de detección de obstáculos y de sonidos para el robot humanoide Bioloid de la empresa coreana Robotics. Esta biblioteca resulta muy útil para aplicaciones donde se requiera que

un robot tome decisiones y ejecute acciones de manera autónoma en entornos que son dinámicos.

Las habilidades de movimiento presentadas en éste trabajo por el momento son habilidades prediseñadas, lo cual quiere decir que el robot siempre las ejecutará de la misma manera. Sin embargo, se tiene contemplado en un futuro añadir habilidades dinámicas de tal manera que el robot las ejecute de distintas formas acorde a las características del ambiente donde se encuentra interactuando, tomando en cuenta por ejemplo, la fricción al desplazarse, la irregularidad del terreno, entre otras consideraciones.

Se tiene pensado además utilizar el mecanismo de comunicación Zigbee para lograr una comunicación rápida y a bajo costo para poder usarse en aplicaciones donde se requiera intercambiar y comunicar información de robot a robot y de la PC al robot y viceversa.

VI. AGRADECIMIENTOS

Este trabajo está parcialmente apoyado por el CoECyT-Jal proyecto No. 2008-05-97094, mientras que su autor Héctor Rafael Orozco Aguirre cuenta con beca de CONACYT No. 203060.

VII. REFERENCIAS

- [1] J. Kim, I. Park and J. Oh. *Walking Control Algorithm of Biped Humanoid Robot on Uneven and Inclined Floor*. Journal of Intelligent and Robotic Systems, Vol. 48, no. 4, pp. 457–484, 2007.
- [2] M. Matarić. *Getting Humanoids to Move and Imitate*. IEEE Intelligent Systems, Vol. 15, no. 4, pp. 18-24, 2000.
- [3] S. Ramamoorthy. *Task Encoding, Motion Planning and Intelligent Control using Qualitative Models*. PhD thesis, The University of Texas at Austin, 2007..
- [4] D. Wolpert, Z. Ghahramani and J. Flanagan. *Perspectives and Problems in Motor Learning*. Trends in Cognitive Sciences, Vol. 5, no. 11, pp. 487–494, 2001.
- [5] M. Matarić. *Designing and Understanding Adaptive Group Behavior*. Adaptive Behavior, Vol. 4, no. 1, pp. 51-80, 1995.
- [6] W. Burgard, A. Cremers, D. Fox D et al. *Experiences with an Interactive Museum Tourguide Robot*. Artificial Intelligence, Vol. 114, pp. 32-149, 2000..
- [7] R. Brooks. *A robust Layered Control System for a Mobile Robot*. IEEE Journal of Robotics and Automation, Vol. 2, pp. 14-23, 1986.
- [8] A. Brooks. *Intelligence without Reason*. Proceedings of the International Joint Conference on Artificial Intelligence Sydney, Australia, pp.569-595. Cambridge, MA. MIT Press, 1991.
- [9] G. Giralt, R. Chatila and M. Vaisset M. *An Integrated Navigation and Motion Control System for Autonomous Multisensory Mobile Robots*. Proceedings of the First International Symposium on Robotics Research, pp. 191-214, Cambridge, MA: MIT Press, 1983.
- [10] J. Connell. *SSS: A Hybrid Architecture Applied to Robot Navigation*. Proceedings of the International Conference on Robotics and Automation, Nice, France, pp. 2719-2724. Los Alamitos, CA: AAAI/MIT Press, 1991.
- [11] M. Matarić. *Integration of representation into Goal-Driven Behavior-Based Robots*. IEEE Transactions on Robotics and Automation, Vol. 8, no. 3, pp. 304-312, 1992.
- [12] M. Matarić. *Behavior-Based Control: Examples from Navigation, Learning, and Group Behavior*. Journal of Experimental and Theoretical Artificial Intelligence, Vol. 9, pp. 323-336, 1997.
- [13] M. Matarić. *Behavior-Based Robotics as A Tool for Synthesis of Artificial Behavior*

and Analysis of Natural Behavior. Trends in Cognitive Science, Vol. 2, no. 3, pp. 82-87, 1998.

de Tecnología de Compiègne Francia, tiene 15 años de experiencia en la investigación, la docencia y el desarrollo tecnológico.

- [14] R. Arkin. *Behavior-Based Robotics*. Cambridge, MA: MIT Press, 1998.
- [15] ASIMO: The Honda Humanoid Robot. <http://world.honda.com/ASIMO/>.
- [16] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki and K. Fujimura. *The Intelligent ASIMO: System Overview and Integration*. IEEE/RSJ International Conference on Intelligent Robots and System, Vol. 3, pp. 2478–2483, 2002.
- [17] J. Chestnutt, M. Lau, G. Cheung, J. Kuffner, J. Hodgins and T. Kanade. *Footstep Planning for the Honda ASIMO Humanoid*. Proceedings of the 2005 IEEE International Conference on Robotics and Automation, pp. 629-634, ICRA 2005.
- [18] T. Takenaka. *The Control System for the Honda Humanoid Robot*. Age and Ageing, Vol. 35, no. 1, 2006.
- [19] *The Bioloid Robotic Kit*. http://www.robotis.com/zbxe/bioloid_en.
- [20] *The Zigbee Alliance*. <http://www.zigbee.org/>.
- [21] *Robotis Bioloid Remote Brain Firmware and Host Library*. <http://www.mit.edu/~vona/BRBrain/BRBrain-info.html>.

VIII. AUTORES

M. en C. Héctor Rafael Orozco Aguirre es Maestro en Ciencias en Ingeniería Eléctrica con especialidad en computación por el CINVESTAV, unidad Guadalajara, desde el 2005 es profesor de asignatura del Centro Universitario de Ciencias Exactas e Ingenierías de la Universidad de Guadalajara y en el 2006 ingresó a estudiar el Doctorado en Ciencias en Ingeniería Eléctrica en el CINVESTAV, unidad Guadalajara.

Dr. Félix Francisco Ramos Corchado obtuvo su título de Doctor en Computación de la Universidad

SIMULADOR 3D PARA ROBOTS MANIPULADORES DE 3 GRADOS DE LIBERTAD

Bravo Hernández Paola, Reyes Cortes Fernando, Vergara Limón Sergio, Ríos Silva Eduardo.

Facultad de Ciencias de la Electrónica, Posgrado en Ciencias de la Electrónica, División de Robótica.
Benemérita Universidad Autónoma de Puebla
Av. San Claudio y 18 Sur Edificio 182-CU, Col. Jardines de San Manuel, Puebla, Pue. 72570, México
Tel. 229-55-00 ext. 7411, correo electrónico: ukiro@hotmail.com, huron_cito@yahoo.com.mx

RESUMEN

Este artículo expone el desarrollo de un simulador (Simrom) con ambiente gráfico tridimensional para la simulación de algoritmos de control en el robot manipulador experimental de 3 grados de libertad, capaz de simular el movimiento real del robot en base a la posición deseada y el algoritmo aplicado. El robot experimental es usado para investigación dentro del Laboratorio de Robótica en la Facultad de Ciencias de la Electrónica de la BUAP, tratándose de un Robot de Transmisión Directa en donde el simulador hasta ahora utilizado es bajo el sistema operativo de DOS en consola dificultando el análisis del comportamiento dinámico del robot. El desarrollo del nuevo software tiene como finalidad y objetivo mejorar la visibilidad del espacio para el usuario y la fácil comprensión sobre su comportamiento dinámico lográndose con la introducción del elemento gráfico en tres dimensiones. El desarrollo del software del simulador está basado en el sistema operativo de Microsoft Windows destacando que la codificación es en arquitectura abierta bajo el esquema de la Win32 y la programación orientada a objetos (POO) proporcionando facilidad de mantenimiento y comprensión de la programación. El motor gráfico 3D se implemento con los recursos de la librería de DirectX, con el análisis de las ecuaciones de movimiento de Lagrange se obtuvo el modelo dinámico del robot con el fin de lograr la trayectoria de movimiento para llegar al punto deseado y así lograr la manipulación del prototipo real por la solución de la ecuación dinámica del robot con el método Runge-Kutta.

Palabras clave: Gráficos 3D, Robot 3GDL, Lagrange, Modelo dinámico.

I. INTRODUCCIÓN

Los simuladores permiten la construcción de modelos virtuales que interaccionen con el usuario de igual forma o transmitiéndole las mismas sensaciones que el modelo real. Desde los grandes simuladores utilizados en aeronáutica para el entrenamiento militar o de riesgos desarrollados en los años 80.

El desarrollo del software para la simulación de un robot manipulador de tres grados de libertad donde la principal característica del simulador es el ambiente gráfico tridimensional es el objetivo al cual se desea llegar otorgándole la ventaja de ser amigable al usuario, y adoptarlo como elemento fundamental para tareas de diseño y evaluación de algoritmos experimentales de control. El simulador 3D nació de la necesidad de poseer una visión espacial, anticipar sucesos, predecir hallazgos o analizar fenómenos sin la necesidad de la plataforma física.

Una manera fácil y sencilla de introducir un motor gráfico tridimensional es con la librería de DirectX. DirectX es una API que permite mostrar escenas de mundos tridimensionales utilizando aceleración gráfica por hardware. Los últimos avances en la API han hecho la plataforma de Microsoft muy potente y estable. DirectX es el estándar actual para programación gráfica sobre Windows.

El simulador resuelve el problema de control de posición que consiste en llevar el efector final del extremo del robot de una posición inicial hacia una posición deseada, mediante la obtención de la ecuación dinámica de movimiento de Lagrange, y su solución con el método de Runge-Kutta de cuarto orden [1], para obtener resultados reales con el simulador, y que facilite el análisis de su comportamiento al visualizarlo en el ambiente tridimensional logrando la definición de trayectorias libres de colisiones y análisis de las características dinámicas del robot manipulador.

II. MÉTODO

Sistema físico: En general un robot manipulador es una estructura mecánica que consiste en una secuencia de cuerpos rígidos (links) conectados por medio de articulaciones (joints), un manipulador se caracteriza por un brazo que asegura movilidad, una muñeca que confiere destreza y un efector final que se encarga de realizar las tareas requeridas por el robot [2].

Rotradi: es un robot experimental de *arquitectura abierta*. Este brazo robot compuesto por eslabones rígidos de aluminio cuenta con tres ejes accionados con servomotores, lo que suministra tres grados de libertad, además de contar con la característica de “transmisión directa”, es decir, que los motores se encuentran directamente acoplados a los ejes, evitando con esto la utilización de transmisores o reductores de par o fuerza (figura 1).



Fig. 1 Robot manipulador de transmisión directa de 3 grados de libertad.

El simulador gráfico permite visualizar la ejecución del robot conforme el tiempo transcurre

evitando dañar al robot físicamente con colisiones o pares elevados de fuerza e incluso elevando la seguridad del usuario, porque de presentarse una anomalía del sistema solo se necesita reiniciar la sesión del simulador y empezar de nuevo el análisis.

Ecuación dinámica del robot: En esta sección se da a conocer la regla matemática que vincula las variables de entrada y salida del sistema. Esta caracterización matemática se describe por medio de ecuaciones diferenciales. El modelo matemático del sistema a controlar se realiza de forma analítica es decir basado en las leyes de la física que rigen el comportamiento del sistema. Las ecuaciones dinámicas de un robot manipulador pueden obtenerse a partir de las ecuaciones de movimiento de Newton. El inconveniente que presenta este método es que el análisis se complica notablemente cuando aumenta el número de articulaciones del robot. En estos casos es conveniente emplear las ecuaciones de movimiento de Lagrange.

En este planteamiento, la problemática de control de posición consiste en llevar el extremo del robot a un punto en el espacio independientemente de su posición inicial, generando una trayectoria, es por esa razón que necesitamos el modelo dinámico y uno de los procedimientos más empleados para la obtención del modelo dinámico del robot manipulador de tres grados de libertad, en forma cerrada, está basado en las ecuaciones de movimiento de Lagrange.

El uso de las ecuaciones de Lagrange para el modelado dinámico del robot manipulador de tres grados de libertad se reduce a cuatro etapas:

1. Cálculo de la energía cinética.
2. Cálculo de la energía potencial.
3. Cálculo del lagrangiano.
4. Desarrollo de las ecuaciones de Lagrange.

Para el estudio y simulación del Rotradi (Robot de transmisión directa) de 3 grados de libertad se realizó el análisis para la obtención de las ecuaciones dinámicas de Lagrange para la base

(τ_1), hombro (τ_2) y codo (τ_3), y se logró como resultado respectivamente las siguientes ecuaciones de movimiento.

$$\begin{aligned} \tau_1 = & \left[l_1 + m_2 l_{c2}^2 \text{sen}^2(q_2) + m_3 [l_2 \text{sen}(q_2) + l_{c3} \text{sen}(q_2 + q_3)]^2 + l_{z2} \cos^2(q_2) + l_{y2} \text{sen}^2(q_2) \right. \\ & \left. + l_{z3} \cos^2(q_2 + q_3) + l_{y3} \text{sen}^2(q_2 + q_3) \right] \ddot{q}_1 \\ & + 2 \left[m_2 l_{c2}^2 \text{sen}(q_2) \cos(q_2) + (l_{y2} - l_{z2}) \text{sen}(q_2) \cos(q_2) \right. \\ & \left. + (l_{y3} - l_{z3}) \text{sen}(q_2 + q_3) \cos(q_2 + q_3) \right. \\ & \left. + m_3 [l_2 \text{sen}(q_2) + l_{c3} \text{sen}(q_2 + q_3)] [l_2 \cos(q_2) + l_{c3} \cos(q_2 + q_3)] \right] \dot{q}_1 \dot{q}_2 \\ & + 2 \left[(l_{y3} - l_{z3}) \cos(q_2 + q_3) \text{sen}(q_2 + q_3) \right. \\ & \left. + m_3 l_{c3} \cos(q_2 + q_3) [l_2 \text{sen}(q_2) + l_3 \text{sen}(q_2 + q_3)] \right] \dot{q}_1 \dot{q}_3 \end{aligned}$$

$$\begin{aligned} \tau_2 = & [l_2 + l_3 + m_3 l_2^2 + m_2 l_{c2}^2 + m_3 l_{c3}^2 + 2m_3 l_2 l_{c3} \cos(q_3)] \ddot{q}_2 \\ & + [l_3 + m_3 l_{c3}^2 + m_3 l_2 l_{c3} \cos(q_3)] \ddot{q}_3 \\ & - [m_2 l_{c2}^2 \text{sen}(q_2) \cos(q_2) \\ & + m_3 [l_2 \text{sen}(q_2) + l_{c3} \text{sen}(q_2 + q_3)] [l_2 \cos(q_2) + l_{c3} \cos(q_2 + q_3)] \\ & + [l_{y2} - l_{z2}] \cos(q_2) \text{sen}(q_2) + [l_{y3} - l_{z3}] \cos(q_2 + q_3) \text{sen}(q_2 + q_3)] \dot{q}_2^2 \\ & - m_3 l_2 l_{c3} \text{sen}(q_3) \dot{q}_3^2 - 2m_3 l_2 l_{c3} \text{sen}(q_3) \dot{q}_2 \dot{q}_3 + [m_2 l_{c2} + m_3 l_2] g \text{sen}(q_2) \\ & + m_3 g l_{c3} \text{sen}(q_2 + q_3) \end{aligned}$$

$$\begin{aligned} \tau_3 = & [l_3 + m_3 l_{c3}^2] \ddot{q}_3 + [l_3 + m_3 l_{c3}^2 + m_3 l_2 l_{c3} \cos(q_3)] \ddot{q}_2 \\ & - [m_3 [l_2 \text{sen}(q_2) + l_{c3} \text{sen}(q_2 + q_3)] l_{c3} \cos(q_2 + q_3) \\ & + [l_{y3} - l_{z3}] \cos(q_2 + q_3) \text{sen}(q_2 + q_3)] \dot{q}_1^2 + m_3 l_2 l_{c3} \dot{q}_2^2 \text{sen}(q_3) \\ & + m_3 l_{c3} g \text{sen}(q_2 + q_3) \end{aligned}$$

Variable	Descripción
q_1	Posición angular del primer eslabón.
q_2	Posición angular del segundo eslabón.
q_3	Posición angular del tercer eslabón.
l_1	Longitud del primer eslabón.

l_2	Longitud del segundo eslabón.
l_3	Longitud del tercer eslabón.
l_{c1}	Distancia al centro de masa del primer eslabón.
l_{c2}	Distancia al centro de masa segundo eslabón.
l_{c3}	Distancia al centro de masa del tercer eslabón.
m_1	Masa del primer eslabón.
m_2	Masa del segundo eslabón.
m_3	Masa del tercer eslabón.
I_1	Inercia del primer eslabón respecto al centro de masa.
I_2	Inercia del segundo eslabón respecto al centro de masa.
I_3	Inercia del tercer eslabón respecto al centro de masa.
g	Aceleración de la gravedad.

Las expresiones anteriores representan las ecuaciones dinámicas del robot manipulador las cuales pueden expresarse de la forma:

$$\tau = M(q)\ddot{q} + C(q, \dot{q})\dot{q} + g(q) + f(\dot{q})$$

Donde $\tau \in \mathbb{R}^{3 \times 1}$ llamado vector de fuerzas externas, siendo generalmente los pares y fuerzas aplicados por los accionadores en las articulaciones; $M(q) \in \mathbb{R}^{3 \times 3}$ es una matriz simétrica definida positiva denominada matriz de inercia, $C(q, \dot{q})\dot{q} \in \mathbb{R}^{3 \times 1}$ es un vector llamado vector de fuerzas centrífugas y de Coriolis, y $g(q) \in \mathbb{R}^{3 \times 1}$ es el vector de fuerzas o pares gravitacionales y la fricción es modelada por un vector $f(\dot{q}) \in \mathbb{R}^n$ dependiente solamente de la velocidad articular \dot{q} , y que combina los denominados fenómenos de fricción viscosa y de fricción de Coulomb:

$$f(\dot{q}) = F_{m1}(\dot{q}) + F_{m2} \text{sign}(\dot{q})$$

donde F_{m1} y F_{m2} son matrices de $n \times n$ diagonales definidas positivas. Los elementos de la diagonal de F_{m1} corresponden a los parámetros de fricción viscosa (b) y los de F_{m2} corresponden a los parámetros de fricción de Coulomb (f_c) en donde se encuentra que:

$$f(\dot{q}) = b\dot{q} + f_c \text{sign}(\dot{q})$$

Desarrollo del software. Simrom es una aplicación C++ desarrollada en el entorno MS Visual C++, con una representación basada en la librería gráfica DirectX y bajo el sistema operativo de Windows.

El proceso de inicialización de Direct3D para una aplicación de gráficos se puede separar en los siguientes pasos:

1. Obtener un puntero al interface **IDirect3D9**. Esta interface se usa para conseguir información acerca de los dispositivos físicos presentes en el sistema y para crear la interfaz **IDirect3DDevice9**, que representa el dispositivo físico que se utilizará para mostrar los gráficos 3D.

2. Enumerar los adaptadores de pantalla disponibles y escoger el deseado para visualizar. Enumerar los modos gráficos disponibles en el adaptador escogido y sus características (resolución, formato del píxel, etc.). Es necesario crear el dispositivo indicando un modo gráfico soportado por éste.

3. Obtener las capacidades del dispositivo (**D3DCAPS9**) para comprobar si el adaptador de pantalla escogido soporta procesamiento de vértices por hardware. Es necesario saberlo para crear el interface **IDirect3DDevice9**.

4. Inicializar los datos de la estructura **D3DPRESENT_PARAMETERS**. Esta estructura determina las características del interface **IDirect3DDevice9** que se va a crear.

5. Crear la interface **IDirect3DDevice9** utilizando la estructura **D3DPRESENT_PARAMETERS** inicializada.

La geometría, las matemáticas y la física son los fundamentos teóricos sobre los que descansa la programación de entornos 3D. En un mundo virtual distinguimos, al menos, tres tipos de objetos: las mallas (listas de polígonos), las cámaras y las luces.

La necesidad de mejorar la infraestructura del Laboratorio de Robótica en el Posgrado de la

BUAP, con propósitos de investigación científica y aplicaciones industriales, se logro con el desarrollo del software del presente simulador, debido a que las simulaciones que se realizaban para el análisis de controladores experimentales era un ambiente en MS Dos modo de consola, dificultando al usuario predecir el movimiento de la trayectoria del robot ocasionando daños físicos en su estructura mecánica.

Simulador con robot modelado en 3D: Los gráficos por computadora siempre han sido considerados un asunto de gran complejidad, principalmente debido a la necesidad de conocimientos matemáticos y de programación, con el surgimiento de las librerías gráficas se pueden desarrollar gráficos interactivos en tres dimensiones de una forma sencilla así como la facilidad de su programación [3].

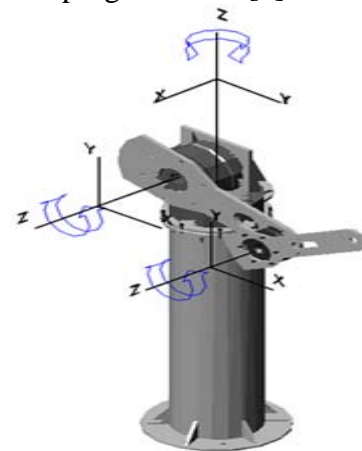


Fig. 2 Robot en el espacio tridimensional.

El modelado en 3D consiste en el desarrollo de una representación visual de un objeto o conjunto de objetos mediante un ordenador para observar el modelo final desde cualquier ángulo (figura 2). La primera necesidad para el desarrollo del simulador es el dibujar el robot manipulador de manera gráfica y en tres dimensiones por lo que se utilizo el software de Solidworks para el diseño del robot manipulador y posteriormente ser exportado a DirectX. La figura 3 muestra la reproducción del brazo, con la ayuda de Solidworks.

La importancia del modelado en 3D se encuentra en la facilidad que da el programador al usuario

para ver el comportamiento dinámico del robot a estudiar antes de llevarlo a la práctica, logrando detectar anomalías o mejoras que serían difíciles de observar con solo datos numéricos o gráficas.

La necesidad que debe cumplir la interfaz es el requerimiento de ser amigable y de fácil manipulación para el usuario, como lo es la manipulación de ventanas en Windows, siendo la mejor parte que la programación se hizo en el entorno de Win32 y de esta forma obtener un software de arquitectura abierta sin la necesidad de casarse con una librería dinámica comercial como lo es la MFC de Visual C, dando la ventaja de poder acceder al código para mantenimiento o manipulación del mismo para los que el usuario convenga.



Fig. 3 Robot manipulador diseñado en Solidworks.

Aplicación Windows y librería gráfica DirectX.

Una ventana es la forma interactiva ante el usuario en que se presentan las aplicaciones de Windows. Cada aplicación tiene al menos una ventana principal, y todas las comunicaciones entre usuario y aplicación se canalizan a través de una ventana. Siempre se ha dicho que Windows tiene la ventaja de ser fácil para el usuario y complicado para el desarrollador. El motivo es porque esta manera de programar y codificar es bastante peculiar. Otra de las ventajas que se encuentra al programar bajo la api win32 es desechar código innecesario que no sabemos el porqué de su existencia, facilitando su mantenimiento y accesibilidad al código para un mejor entendimiento.

El entorno de Microsoft Windows está orientado a los gráficos, lo que significa que proporciona una Interfaz Gráfica de Usuario (GUI) con funciones

para la gestión de ventanas y manejo de mensajes [4], siendo responsable de implementar y facilitar a DirectX la facultad de dibujar en una ventana interactiva para el usuario a través de controles para el intercambio de datos. Dado que los elementos de interfaz gráfico se pueden describir y utilizar mucho más fácilmente como objetos, se utilizó la programación orientada a objetos (POO) para el desarrollo del software del simulador de robots manipuladores.

DirectX surgió en septiembre de 1995 con la aparición del Sistema Operativo Microsoft Windows 95, en un principio fue creada para mejorar los gráficos y sonidos en la plataforma de Microsoft Windows, en la actualidad es un conjunto de aplicaciones de interfaces de programación (API) que permite construir tanto videojuegos como tipos de aplicaciones multimedia de alto rendimiento para un modelado portátil, rápido y elegante cuyos componentes principales son:

1. Direct3D para dibujar gráficos en 3D permitiendo optimizar el resultado al mostrarlo en pantalla tomando en cuenta las ventajas del hardware que existe en la computadora, y si no se encuentra hardware lo hace a nivel software.
2. DirectInput, es la librería que facilita el acceso a los dispositivos de entrada (teclado, ratón, etc.).

El programa principal incluye las funciones necesarias para la interacción con el robot manipulador, basándose en una librería que se crea para manipular variables en la ejecución de la simulación y en la fase de la evaluación del modelo dinámico para el movimiento y posición deseada.

La figura 4 ilustra la interfaz gráfica del simulador de robots manipuladores, al ejecutar el código fuente con DirectX y Visual C++.

Con la obtención de las ecuaciones dinámicas de movimiento de Lagrange, del robot de tres grados de libertad, se logró la ejecución de los movimientos del robot al evaluar el algoritmo

dentro del simulador comprobando su funcionalidad con el algoritmo de control.

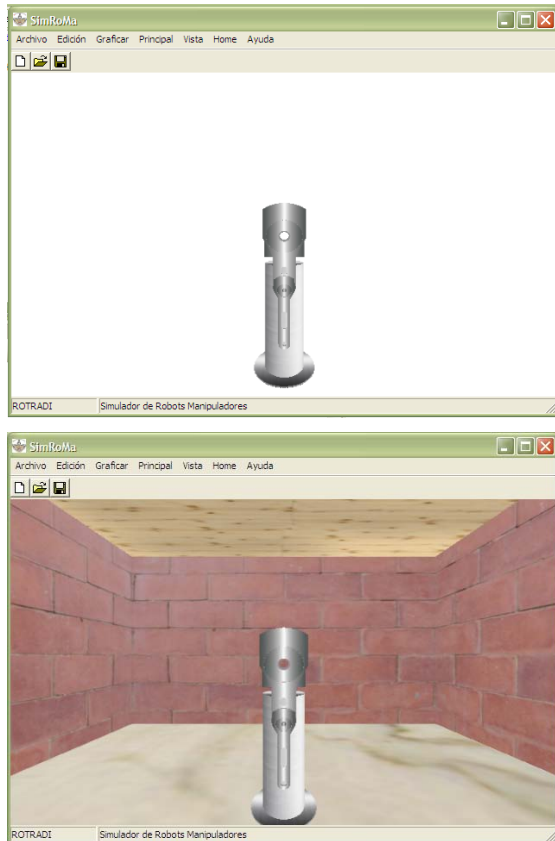
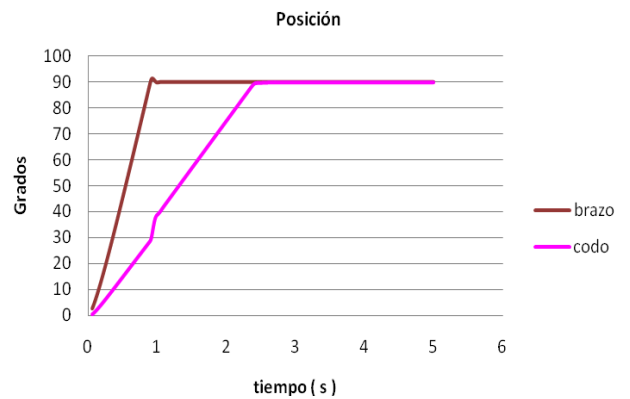
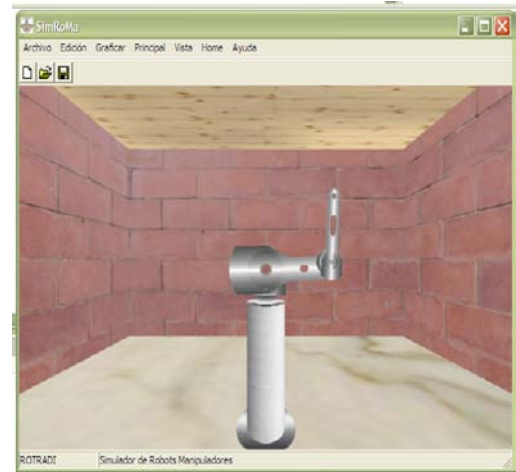


Fig. 4 Interfaz Gráfica del simulador.

En la figura 5 se ilustra el resultado cuando se ejecuto el simulador con la evaluación de un controlador de tipo Tangente hiperbólico [5] y una posición deseada de 90° tanto para el codo como el hombro

Finalmente en el simulador se puede escoger la modalidad de tal manera que la escena tenga 3 vistas como si se tuviera 3 cámaras con diferentes ángulos seleccionando diferentes colores y espacios donde tendrá la manipulación el robot, la ventaja que encontramos con esta escena al dividir la ventana en 4 mini-escenas es la de poder manipular la cámara de la vista superior izquierda mientras que las otras 2 se mantienen fijas y a su vez visualizar la creación de la gráfica en tiempo real mientras se ejecuta el movimiento mostrado en la figura 6.



Parámetros	Notación	Eslabón 1	Eslabón 2
Masa	m	24.25	3.87
Longitud	l	0.45	0.45
Longitud centro de masa	l_{cm}	0.9028	0.045
Inercia respecto al centro de masa	I	1.267	0.0925
Coefficiente de fricción viscosa	F_v	2.288	0.175
Coefficiente de fricción de Coulomb	F_c	7.17	1.734
Ganancia proporcional	k_p	55	35
Ganancia derivativa	k_v	40	25
Posición deseada	q_d	90	90

Fig. 5 Ejemplo de simulación del codo-brazo.

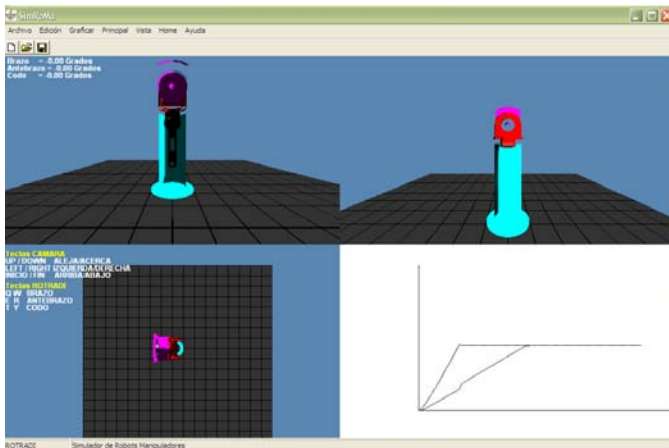


Fig. 6 Simulación con diferentes vistas.

La simulación en tiempo real consiste en obtener esta imagen lo más rápidamente posible, de manera que sea factible generar un número mínimo de imágenes por segundo para que un observador humano no aprecie la transición entre ellas. La velocidad para la visualización de la simulación de un sistema de mide en frames (imágenes) por segundo. El número de fps partir del que podemos considerar una aplicación gráfica como interactiva o en tiempo real es aproximadamente 15. A partir de unos 72 fps, el incremento de velocidad de visualización es inapreciable para la percepción visual humana.

III. CONCLUSIONES

En el presente artículo se ha descrito el desarrollo del software de un Simulador de Robots Manipuladores con ambiente gráfico tridimensional bajo el entorno de Windows realizado el programa con el ambiente integrado de desarrollo de Visual C++, permitiendo compilar en lenguaje C++ la programación del simulador debido a su gestión automatizada de memoria, además de ser un lenguaje estructurado y orientado a objetos.

Dentro de las aportaciones de la investigación se proporciono el desarrollo del software para la creación del simulador gráfico en 3D, destacando que no solo es una animación sino la simulación de los fenómenos físicos reales. Con la simulación del comportamiento del robot, se satisface la necesidad

de controlar las variables y analizar diversos controladores con el fin de calcular cual debe ser la señal de salida para llegar a la posición deseada.

Este proyecto se puede continuar, para mejorarlo y lograr un sistema lo bastante robusto para aplicación a otros proyectos.

IV. REFERENCIAS

- [1] Teukolsky, P., Vetterling. 2002. *Numerical Recipes in C++*. Cambridge University Press, USA.
- [2] Kelly R., Santibañez V., 2003. *Control de Movimiento de Robots Manipuladores*. Prentice Hall, Madrid.
- [3] Landa, A. 2006 *DirectX Programación de Gráficos 3D*. Users .code. Banfield.
- [4] Schildt H. 1995. Programación en C y C++ en Windows 95. McGraw-Hill, México.
- [5] Reyes F., Espinosa L. & Barahona J., *Family of saturated controllers for robot manipulators*, Proc. of the IASTED International Conf. on Control and Applications, Cancún, México, 2002, 250-255.

V. AUTORES

Ing. Paola Bravo Hernández

Ingeniero en Electrónica egresada del Instituto Tecnológico de Puebla en 2005.

Candidata a obtención de grado de Maestría en Facultad de Ciencias de la Electrónica en la Benemérita Universidad Autónoma de Puebla.

Dr. Fernando Reyes Cortés

Profesor investigador de la Facultad de Ciencias de la Electrónica en la Benemérita Universidad Autónoma de Puebla.

Obtuvo el grado de Doctor en Electrónica por el Centro de Investigación Científica y Educación Superior de Ensenada, B.C.

Es responsable del posgrado en Automatización de la Facultad de Ciencias de la Electrónica de la BUAP.

Es miembro del Sistema Nacional de Investigadores (SNI) nivel I.

Dr. Sergio Vergara Limón

Profesor Tiempo completo titular “A”, Postgrado de Automatización, Facultad de Ciencias de la Electrónica de la BUAP.

Licenciado en Electrónica en la Universidad Autónoma de Puebla (1989-1994).

Obtuvo el grado de Maestría en Ciencias en la Facultad de Ciencias Físico Matemáticas, Universidad Autónoma de Puebla (1995-1997).

Obtuvo el grado de Doctor en ciencias Facultad de Ciencias Físico Matemáticas, Universidad Autónoma de Puebla (1997-2000).

Dr. Eduardo Ríos Silva

Profesor Investigador de tiempo completo.

Obtuvo el grado de Doctor en ciencias Facultad de Ciencias Físico Matemáticas, Universidad Autónoma de Puebla

ANÁLISIS DEL COMPORTAMIENTO MECÁNICO DE ACELERÓMETROS PIEZORESISTIVOS UNIAXIALES

A R Cortés-Pérez¹, A L Herrera-May^{1,2}, L A Aguilera-Cortés¹,
y M.A. González-Palacios¹

¹Depto. Ingeniería Mecánica, Campus Irapuato-Salamanca, Universidad de Guanajuato,
Carretera Salamanca-Valle de Santiago km 3.5+1.8km, Salamanca, Gto., México

²Centro de Investigación en Micro y Nanotecnología, Universidad Veracruzana, 94292,
Boca del Río, Ver., México

RESUMEN

Los acelerómetros tienen un amplio rango de aplicaciones en el sector automotriz, en sistemas de navegación, en monitoreo de actividad biomédica y en sistemas de realidad virtual. Dos modelos analíticos para determinar el ancho de banda y los esfuerzos longitudinales máximos en acelerómetros piezoresistivos son presentados. Los modelos son obtenidos con la teoría de viga de Euler-Bernouilli y el método de Rayleigh. Además, un modelo de elementos finitos (EF) de un acelerómetro piezoresistivo es desarrollado. Una carga de aceleración de 10g fue aplicada en dirección perpendicular a la estructura del acelerómetro. Los resultados de los modelos analíticos concuerdan con los obtenidos por medio del modelo EF. Los modelos propuestos pueden ser usados fácilmente por diseñadores para predecir el ancho de banda y los esfuerzos máximos en acelerómetros piezoresistivos basados en vigas tipo cantilever.

Palabras clave: Acelerómetro, elemento finito, esfuerzos, método de Rayleigh.

I. INTRODUCCIÓN

La tecnología de sistemas microelectromecánicos (MEMS) ha permitido el desarrollo de sensores eficientes tales como acelerómetros [1], sensores de presión [2] y sensores de campo magnético [3]. Los sensores basados en esta tecnología presentan importantes ventajas que incluyen el bajo consumo de potencia, alta sensibilidad y bajo costo. Por ejemplo, los acelerómetros son usados

comercialmente en la industria automotriz para la activación de sistemas de seguridad [4], en sistemas de navegación, en monitoreo de actividad biomédica, sismología y aplicaciones electrónicas y de realidad virtual [5-6]. Dos principales requisitos de los acelerómetros incluyen una alta sensibilidad y ancho de banda. Además, es necesario garantizar una operación segura de los acelerómetros, es decir, que no se generen esfuerzos superiores al esfuerzo de ruptura del silicio. De esta forma, los diseñadores e investigadores tienen importantes desafíos para obtener altos valores de sensibilidad y ancho de banda sin la generación de esfuerzos dañinos a la estructura de los acelerómetros.

Desde la década de los ochenta, una variedad de acelerómetros han sido desarrollados basados en técnicas de detección piezoresistiva [7], capacitiva [8], térmica [9], tuneleo [10], piezoeléctrica [11] y óptica [12]. Entre estas técnicas, los acelerómetros piezoeléctricos presentan una rápida respuesta, buena linealidad y mínimo consumo de potencia, pero no responden a aceleraciones constantes [13]. En cambio, acelerómetros de tuneleo tienen alta sensibilidad; sin embargo, requieren un proceso de fabricación complejo y presentan problemas de inestabilidad [14]. Mientras, los acelerómetros térmicos no requieren masa de prueba sólida y su funcionamiento está basado en la convección térmica; aunque, tienen variaciones de offset para condiciones de operación prolongadas causadas por cambios en la temperatura ambiente [15]. Además, nuevos diseños involucran a acelerómetros ópticos que tienen inmunidad a interferencia electromagnética (EMI, por sus siglas

en inglés) y reducen el peso total y el ruido electrónico; sin embargo, presentan pérdidas intrínsecas debido a imperfecciones estructurales y necesitan complicados procesos de fabricación [16]. Otra clase de acelerómetros son los capacitivos que tienen baja dependencia de la temperatura, pero presentan desventajas tales como capacitancias parásitas que dificultan las mediciones de la señal, requieren circuitos electrónicos complicados y necesitan empaquetamiento a muy bajas presiones [17]. En cambio, los acelerómetros piezoresistivos tienen un sencillo y maduro proceso de fabricación, circuitos simples, bajo costo y alta sensibilidad; aunque, presentan dependencia a la temperatura que puede ser controlada con un cuidadoso empaquetado y circuitos de compensación [18-19]. Por lo tanto, los acelerómetros piezoresistivos son una importante alternativa para su uso en aplicaciones convencionales y nuevas de medición de la aceleración.

II. DESCRIPCIÓN DE LOS MODELOS ANALÍTICOS

En esta sección se describe los modelos analíticos usados para determinar el ancho de banda y los esfuerzos normales máximos en un acelerómetro piezoresistivo. Estos modelos fueron aplicados a un acelerómetro con una configuración estructural similar al desarrollado por Plaza *et al.* [20].

Configuración estructural

La Fig. 1 muestra el acelerómetro piezoresistivo basado en vigas tipo cantilever. El acelerómetro tiene dos vigas soportes, una masa de prueba y un puente de Wheatstone con cuatro piezoresistores tipo-p. Las parámetros geométricos de éste acelerómetro son mostradas en la Fig. 2. La configuración estructural del acelerómetro es dividido en tres vigas elásticas con tres diferentes longitudes (L_1 , L_2 y L_3) y dos espesores (s_1 y s_2). La masa de prueba es mayor que las vigas soportes.

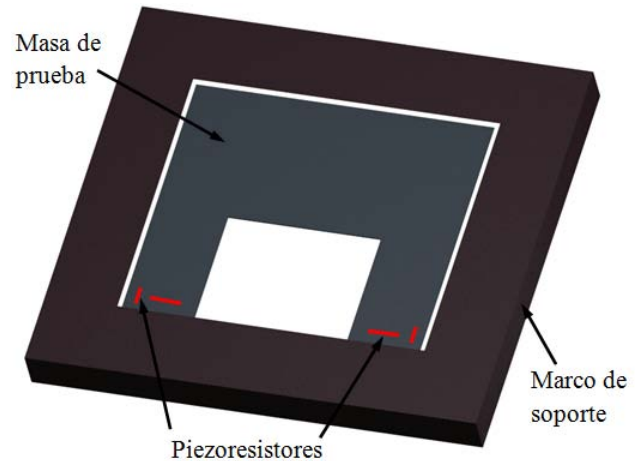


Fig. 1. Vista en 3D del acelerómetro piezoresistivo.

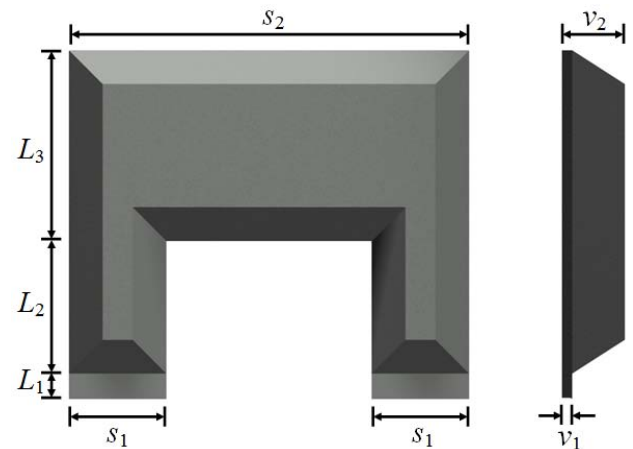


Figura 2. Parámetros geométricos del acelerómetro piezoresistivo.

En cada viga soporte se colocan dos piezoresistores perpendicularmente uno con respecto al otro. Así, cuando una aceleración vertical es aplicada al acelerómetro se obtiene una deformación de las vigas soportes en la dirección longitudinal y transversal. Esta deformación causa un cambio en la magnitud de las resistencias de los cuatro piezoresistores (R_1 , R_2 , R_3 y R_4), lo cual provoca una variación del voltaje de salida (ΔV_{out}) del puente de Wheatstone (ver Fig. 3). De esta forma, la magnitud de la aceleración es medida en función del voltaje de salida del Puente de Wheatstone.

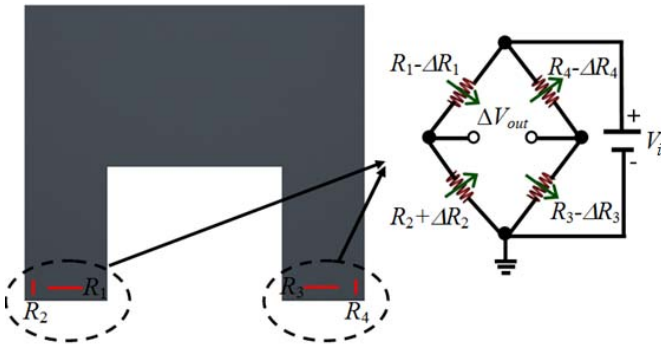


Fig. 3. Puente de Wheatstone ubicado sobre las vigas soportes.

Modelo del ancho de banda

La predicción del ancho de banda del acelerómetro piezoresistivo es importante para establecer el potencial rango de aplicaciones del acelerómetro. Este ancho de banda es determinado con la primera frecuencia fundamental de la estructura del acelerómetro [21]. La frecuencia fundamental puede ser encontrada usando el método de Rayleigh. El método de Rayleigh calcula la frecuencia fundamental de un sistema continuo por medio de la razón de la energía potencial respecto a la energía cinética del sistema [22]. Este método es muy simple y fácil de utilizar para estructuras con rigidez y masa variantes [23]. Por ende, el método de Rayleigh es usado para determinar el ancho de banda del acelerómetro piezoresistivo. Para este caso, una carga de aceleración de 1 g en dirección vertical al acelerómetro es considerada. La aceleración causa una deflexión en la estructura del acelerómetro, la cual es determinada por medio de la teoría de viga de Euler-Bernoulli. Mediante esta teoría son obtenidos los siguientes momentos flexionantes sobre la estructura del acelerómetro sin considerar las pendientes de las vigas.

$$2EI_1 \frac{\partial^2 z_1(x)}{\partial x^2} = -M_1(x), \quad 0 < x < L_1, \quad (1)$$

$$2EI_2 \frac{\partial^2 z_2(x)}{\partial x^2} = -M_2(x), \quad L_1 < x < L_2, \quad (2)$$

$$EI_3 \frac{\partial^2 z_3(x)}{\partial x^2} = -M_3(x), \quad L_2 < x < L_3, \quad (3)$$

donde I_i es el momento de inercia de la sección transversal de la viga, $I_1 = s_1 v_1^3 / 12$, $I_2 = s_1 v_2^3 / 12$, $I_3 = s_2 v_2^3 / 12$, E es el módulo de elasticidad, $M_i(x)$ son las funciones de los momentos flexionantes sobre i -ésima viga y $z_i(x)$ son las funciones de la deflexión de la i -ésima viga de la estructura del acelerómetro.

Las condiciones de frontera de cada viga son determinadas por

$$\begin{aligned} z_1(0) &= 0, & \frac{\partial z_1}{\partial x}(0) &= 0, \\ z_2(0) &= z_1(L_1), & \frac{\partial z_2}{\partial x}(0) &= \frac{\partial z_1}{\partial x}(L_1), \\ z_3(0) &= z_2(L_2), & \frac{\partial z_3}{\partial x}(0) &= \frac{\partial z_2}{\partial x}(L_2). \end{aligned} \quad (4)$$

Las funciones de los momentos flexionantes de cada viga, considerando una carga de aceleración de 1 g, son determinadas de la siguiente forma

Para $0 < x < L_1$:

$$\begin{aligned} M_1(x) &= g \rho s_1 v_1 x^2 - g \rho s_2 v_2 L_2 x - 2g \rho s_1 v_2 L_2 x - 2g \rho s_1 v_1 L_2 x + \\ &+ \frac{1}{2} g \rho s_2 v_2 L_3^2 + g \rho s_2 v_2 L_2 L_3 + g \rho s_2 v_2 L_1 L_3 + g \rho s_1 v_2 L_2^2 + \\ &+ 2g \rho s_1 v_2 L_1 L_2 + g \rho s_1 v_1 L_1^2 \end{aligned} \quad (5)$$

Para $L_1 < x < L_2$:

$$\begin{aligned} M_2(x) &= g \rho s_1 v_2 x^2 - g \rho s_2 v_2 L_3 x - 2g \rho s_1 v_2 L_2 x + \\ &+ \frac{1}{2} g \rho s_2 v_2 L_3^2 + g \rho s_2 v_2 L_2 L_3 + g \rho s_1 v_2 L_2^2 \end{aligned} \quad (6)$$

Para $L_2 < x < L_3$:

$$M_3(x) = \frac{1}{2} g \rho s_2 v_2 x^2 - g \rho s_2 v_2 L_3 x + \frac{1}{2} g \rho s_2 v_2 L_3^2 \quad (7)$$

El ancho de banda del acelerómetro, f_{res} , es determinada mediante el método de Rayleigh [23].

$$f_{res} = \frac{1}{2\pi} \sqrt{\frac{U_{max}}{T_{max}}}, \quad (8)$$

donde U_{max} es la máxima energía potencial y T_{max} es la máxima energía cinética de la estructura del acelerómetro.

$$U_{max} = \int_0^{L_1} I_1 E \left(\frac{\partial^2 z_1(x)}{\partial x^2} \right)^2 dx + \int_{L_1}^{L_2} I_2 E \left(\frac{\partial^2 z_2(x)}{\partial x^2} \right)^2 dx + \int_{L_2}^{L_3} \frac{I_3 E}{2} \left(\frac{\partial^2 z_3(x)}{\partial x^2} \right)^2 dx, \quad (9)$$

$$T_{max} = \int_0^{L_1} \rho A_1 \dot{z}_1^2(x) dx + \int_{L_1}^{L_2} \rho A_2 \dot{z}_2^2(x) dx + \int_{L_2}^{L_3} \frac{\rho}{2} A_3 \dot{z}_3^2(x) dx, \quad (10)$$

donde A_i es el área de la sección transversal de la i -ésima viga.

Las funciones de deflexión $z_1(x)$, $z_2(x)$ and $z_3(x)$ son obtenidas con la sustitución de las ecuaciones (1), (2), and (3) en la ecuación (4). Así,

$$z_1(x) = -\frac{g\rho}{2s_1v_1^3E} (s_1v_1x^4 - 2s_2v_2L_3x^3 - 4s_1v_2L_2x^3 - 4s_1v_1L_1x^3 + 3s_2v_2L_3^2x^2 + 6s_2v_2L_2L_3x^2 + 6s_1v_2L_2^2x^2 + 12s_1v_2L_1L_2x^2 + 6s_2v_2L_1L_3x^2 + 6s_1v_1L_1^2x^2), \quad (11)$$

$$z_2(x) = -\frac{g\rho}{2s_1v_1^3v_2^2E} (s_1v_1^3x^4 - 4s_1v_1^3L_3x^3 - 2v_1^3s_2L_3x^3 + 6s_1v_1^3L_2^2x^2 + 6s_2v_1^3L_2L_3x^2 + 3s_2v_1^3L_3^2x^2 + 4s_1v_1v_2^2L_1^3x + 12s_1v_1^3L_2L_1x + 12s_1v_2^3L_1L_2^2x + 6s_2v_2^3L_1L_3x + 6s_2v_2^3L_1L_2^2x + 4s_2v_2^3L_1^3L_3 + 3s_1v_1v_2^2L_1^4 + 8s_1v_2^3L_1^3L_2 + 6s_1v_2^3L_1^2L_2^2 + 6s_2v_2^3L_1^2L_2L_3 + 3s_2v_2^3L_1^2L_3^2), \quad (12)$$

$$z_3(x) = -\frac{g\rho}{2s_1v_1^3v_2^2E} (s_1v_1^3x^4 - 4s_1v_1^3L_3x^3 + 6s_1v_1^3L_3^2x^2 + 4s_1v_1^3L_2^2x + 12s_1v_2^3L_1L_2^2x + 12s_1v_2^3L_1^2L_2x + 4s_1v_1v_2^2L_1^3x + 6s_2v_1^3L_2L_3^2x + 6s_2v_2^3L_1L_3^2x + 6s_2v_2^3L_1^2L_3x + 6s_2v_2^3L_2L_3x + 12s_2v_2^3L_1L_2L_3x + 3s_1v_1v_2^2L_1^4 + 18s_1v_2^3L_1^2L_2^2 + 4s_1v_1v_2^2L_1^3L_2 + 4s_2v_1^3L_2^3L_3 + 3s_2v_1^3L_2^2L_3^2 + 3s_1v_1L_3^4 + 4s_2v_2^3L_1^3L_3 + 12s_2v_2^3L_1^2L_2L_3 + 8s_1v_2^3L_1^3L_2 + 12s_2v_2^3L_1L_2^2L_3 + 6s_2v_2^3L_1L_2L_3^2 + 3s_2v_2^3L_1^2L_3^2 + 12s_1v_2^3L_1L_2^2), \quad (13)$$

Para simplificar las expresiones anteriores, se utiliza los siguientes parámetros geométricos adimensionales

$$s = \frac{s_2}{s_1}, \quad v = \frac{v_2}{v_1}, \quad L = \frac{L_2}{L_1}, \quad w = \frac{L_3}{L_1}, \quad (14)$$

De esta forma, el ancho de banda del acelerómetro está determinado por

$$f_{res} = \frac{1}{2\pi} \sqrt{\frac{126v^3v_1^2E}{\rho L^4} \frac{12L^5 + A_1L^4 + A_2L^3 + A_3L^2 + A_4L + A_5}{1456L^9 + B_1L^8 + B_2L^7 + B_3L^6 + B_4L^5 + B_5L^4 + B_6L^3 + B_7L^2 + B_8L + B_9}}, \quad (15)$$

donde

$$A_1 = 60v^3 + 30sw$$

$$A_2 = 120v^3 + 20sw^2 + 120sv^3w + 20s^2w^2$$

$$A_3 = 60s^2v^3w^2 + 180sv^3w + 80v^3 + 30s^2w^3 + 60sv^3w^2 + 40v^2$$

$$A_4 = 60s^2v^3w^2 + 40sv^2w + 60v^2 + 15s^2w^4 + 80sv^3w + 60s^2v^3w^3 + 60sv^3w^2$$

$$A_5 = 12v + (30sv^3w^2 + 15sv^3w^3 + 30v^2 + 20v^2w + 20sv^3w + 6w^4)sw$$

$$B_1 = 6552sw + 13104v^3$$

$$B_2 = (6336 + 9936s)sw^2 + (52416sw + 30240v^3 + 22176)v^3$$

$$B_3 = (1680 + 19656s + 5040s^2)sw^3 + (69552sv + 44352v)sv^2w^2 + (77616 + 105840sv^3)sv^3w + (105840v^4 + 12096v + 4368)v^2$$

$$\begin{aligned}
 B_4 &= (15624+15120s)s^2w^4 + (10080+30240s^2+117936s)sv^3w^3 + \\
 &+ (75600v^3+87696s+57456+120960sv^3)sv^3w^2 + (317520v^4+ \\
 &+ 36288v+13104)sv^2w + (158760v^4+20160v^3+4536)v^2 \\
 B_5 &= (19845s^2+5040s+2268)sw^5 + (75600s+78120)s^2v^3w^4 + \\
 &+ (30240s^2+45360s^2v^3+128520s+15120v^3+166320sv^3+ \\
 &+ 10080)sv^3w^3 + (189000v^4+20160v+35280sv+302400sv^4+ \\
 &+ 10080s+12600)sv^2w^2 + (396900v^4+50400v^3+11340)sv^2w + \\
 &+ (120960v+62496)v^5 \\
 B_6 &= (13230s^2+3024s+2184)sw^6 + (79380s^2+20160s+9072)sv^3w^5 + \\
 &+ (90720sv^3+60480s+83160v^3+73080)s^2v^3w^4 + (30240v^4+ \\
 &+ 10080s^2v+3360+332640sv^4+20160s+40320sv+90720s^2v^4)sv^2w^3 + \\
 &+ (7560+30240sv^3+7560s+40320v^3+173880v^4+309960sv^4)sv^2w^2 + \\
 &+ (241920v+124992)sv^5w + (40320v^2+78624v+3360)v^4 \\
 B_7 &= (124740+136080s)s^2v^6w^4 + (75600s^2+238140s+15120)sv^6w^3 + \\
 &+ 60480sv^6w + (151200s+60480)sv^6w^2 + (71820s+15120)s^2v^6w^5 + \\
 &+ (72576+57456s)v^5w^2 + 117936sv^5w + 39744v^5 + (10080+45360s)sv^5w^3 + \\
 &+ (12672+5040sw)v^4 + (39690s^2+9072s+6552)sv^3w^6 + (15120s+13608+ \\
 &+ 47250*s^2)sv^3w^5 + (15120+15120s)s^2v^3w^4 + (11340+16380w)s^2w^3v^2 + \\
 &+ (5544+3780s)s^2w^7 \\
 B_8 &= (71820s+15120)s^2v^6w^5 + 30240s^2v^6w^2 + (60480+30240*s)s^2v^6w^3 + \\
 &+ 26460s^3v^6w^6 + (75600s+47250)s^2v^6w^4 + (57456s+10080)sv^5w^3 + 39744sv^5w + \\
 &+ 25200s^2v^5w^4 + (39312s+39312)sv^5w^2 + (12672sw+5040sw^2+13104)v^4 + \\
 &+ (7560s+11088)s^2v^3w^7 + (13230s^2+9072s+6552)sv^3w^6 + (7560s^2+6048)sv^3w^5 + \\
 &+ 5670s^2v^2w^4 + (3024+5040s)sv^2w^5 + 3276s^2w^8 \\
 B_9 &= (13230w^3+15120w+3780w^4+19845w^2+5040)s^3v^6w^3 + \\
 &+ (3276w^2+5544w+3024)s^2v^3w^6 + (15624w^2+9936+5040w^3+ \\
 &+ 19656w)s^2v^5w^2 + (2268+2184w^6)sv^2w^5 + (6336w+1680w^2+ \\
 &+ 6552)sv^4w + 728sw^9 + 1456v^3
 \end{aligned}$$

Modelo del esfuerzo longitudinal

Gran parte de los acelerómetros piezoresistivos son fabricados en obleas de silicio. Considerando una carga de aceleración vertical de n veces g , se obtiene un momento flexionante total sobre la estructura del acelerómetro. Este momento genera un esfuerzo longitudinal máximo sobre las vigas soportes, el cual está dado por

$$\sigma_{I_{max}} = \frac{nM_0c}{I_0} , \quad (16)$$

donde M_0 es el momento flexionante máximo sobre las vigas soportes, c es la distancia desde el plano neutro de la viga hasta su superficie superior y I_0 es el momento de inercia de la sección transversal de la viga soporte.

El momento flexionante máximo sobre las vigas soportes es obtenido en el extremo de las vigas. Por ende, el momento máximo se determina considerando $x=0$ en la ecuación (5). De esta forma

$$\begin{aligned}
 M_0 &= \frac{1}{2} g \rho s_2 v_2 L_3^2 + g \rho s_2 v_2 L_2 L_3 + g \rho s_2 v_2 L_1 L_3 + \\
 &+ g \rho s_1 v_2 L_2^2 + 2g \rho s_1 v_2 L_1 L_2 + g \rho s_1 v_1 L_1^2 . \quad (17)
 \end{aligned}$$

Sustituyendo la ecuación (17) en (16), se obtiene el esfuerzo longitudinal máximo.

$$\begin{aligned}
 \sigma_{I_{max}} &= \frac{3ng\rho}{2v_1^3} (s_2 v_2 L_3^2 + 2s_2 v_2 L_2 L_3 + 2s_2 v_2 L_1 L_3 + \\
 &+ 2s_1 v_2 L_2^2 + 4s_1 v_2 L_1 L_2 + s_1 v_1 L_1^2) . \quad (18)
 \end{aligned}$$

Usando los parámetros geométricos adimensionales (L , s , v y w) en la ecuación (18), se tiene

$$\sigma_{I_{max}} = \frac{3ng\rho L_1^2}{2v_1} (2+4vL+2vL^2+2svw+2svwL+svw^2) . \quad (19)$$

En base a las ecuaciones (15) y (19) se puede predecir el ancho de banda y el esfuerzo longitudinal máximo del acelerómetro piezoresistivo basado en vigas tipo cantilever. De esta forma, el diseñador puede realizar modificaciones en las dimensiones del acelerómetro y conocer sus efectos en el ancho de banda y el esfuerzo longitudinal máximo. Así, el diseñador podrá adaptar las dimensiones necesarias para alcanzar el ancho de banda necesario para una aplicación específica y garantizar una operación segura (esfuerzos pequeños) del acelerómetro.

III. DESCRIPCIÓN DE LOS MODELOS DE ELEMENTOS FINITOS

En esta sección se presentan un modelo EF usando el software ANSYS para determinar el ancho de banda y distribución de esfuerzos del acelerómetro piezoresistivo. Este modelo fue mallado usando elementos tipo solid95. Estos elementos tienen tres grados de libertad (traslación en las direcciones x , y y z) por cada nodo y en total se utilizaron 8792 elementos. En los extremos de las vigas soportes, donde se ubicarán los piezoresistores, se realizó un mallado más fino. A diferencia del modelo analítico, el modelo EF si considera las pendientes de las vigas soportes y la masa de prueba. Estas pendientes son obtenidas del proceso de grabado del silicio. La Tabla 1 muestra las dimensiones de los parámetros geométricos utilizados en el modelo EF. La Fig. 4 muestra el modelo EF del acelerómetro piezoresistivo. Los resultados obtenidos con el modelo FE serán comparados con los determinados por el modelo analítico.

Tabla 1. Dimensiones de los parámetros geométricos del acelerómetro piezoresistivo.

s_1 (μm)	v_1 (μm)	L_1 (μm)	s_2 (μm)	v_2 (μm)	L_2 (μm)	L_3 (μm)
200	10	200	2500	350	500	1000

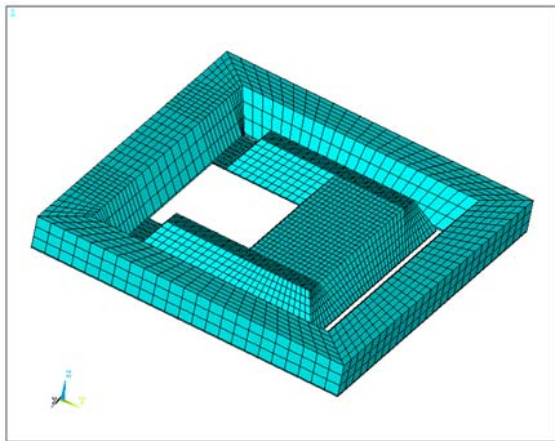


Fig. 4. Modelo EF del acelerómetro piezoresistivo.

IV. RESULTADOS Y DISCUSIONES

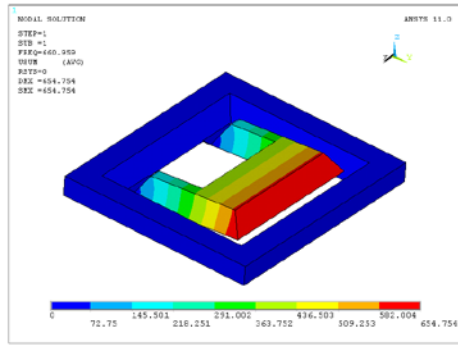
En esta sección se muestran los resultados del ancho de banda y los esfuerzos longitudinales de un acelerómetro piezoresistivo obtenidos mediante los modelos analíticos y el modelo EF. Las dimensiones geométricas del acelerómetro son dadas por la Tabla 1. Además, se consideró una carga de aceleración vertical de 10 g aplicada sobre la estructura del acelerómetro. El material de la estructura del acelerómetro es basado sobre silicio, el cual presenta las siguientes propiedades: módulo de elasticidad de 130 GPa, razón Poisson de 0.28 y densidad de 2300 kg/m³. Mediante el modelo analítico se obtuvo un ancho de banda de 569.23 Hz y un esfuerzo longitudinal máximo de 20.31 MPa, como se muestra en Tabla 2. En cambio, el modelo EF registró un ancho de banda y un esfuerzo longitudinal máximo con una diferencia relativa del 16.1% y -13.3% con respecto a los obtenidos con el modelo analítico. Esta diferencia es ocasionada porque el modelo EF considera las pendientes de las vigas de soporte y la masa de prueba, las cuales se despreciaron en el modelo analítico.

Tabla 2. Resultados del modelo analítico y EF.

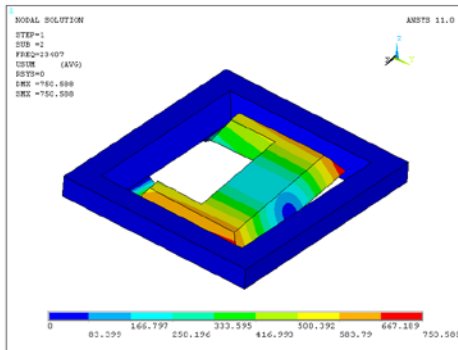
Modelo	Ancho de banda (Hz)	Esfuerzo longitudinal máximo (MPa)
Analítico	569.23	20.31
EF	660.90	17.61

La Fig. 5 muestra los primeros cuatro modos de vibración del acelerómetro. En donde el primer modelo es del tipo flexionante, el segundo torsional y el tercero de tipo lateral. En la práctica, sólo la frecuencia resonante del primer modo de vibración es la que interesa.

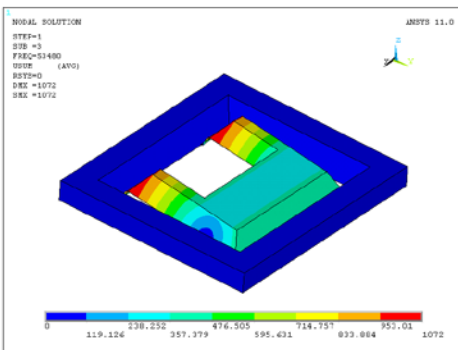
La máxima deflexión del acelerómetro obtenida con el modelo EF fue 7.99 μm , la cual está localizada en el extremo de la masa de prueba. La Fig. 6 muestra la distribución de las deflexiones en el acelerómetro, las cuales son inferiores a la mínima dimensión (10 μm) del acelerómetro.



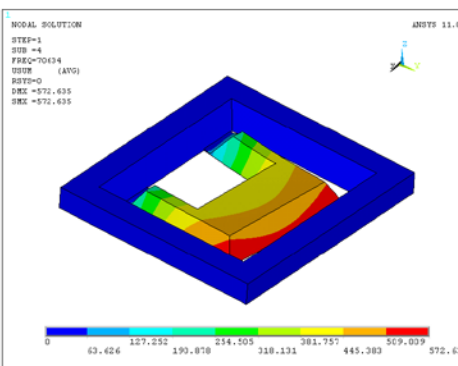
(a)



(b)



(c)



(d)

Fig. 5. (a) Primer, (b) segundo, (c) tercer y (d) cuarto modo de vibración del acelerómetro.

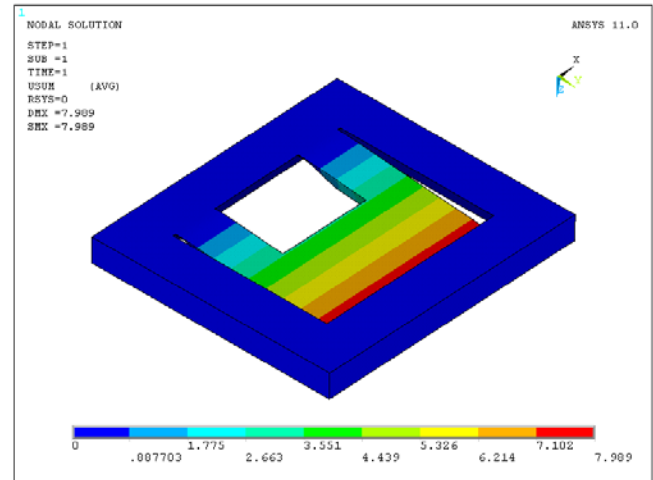
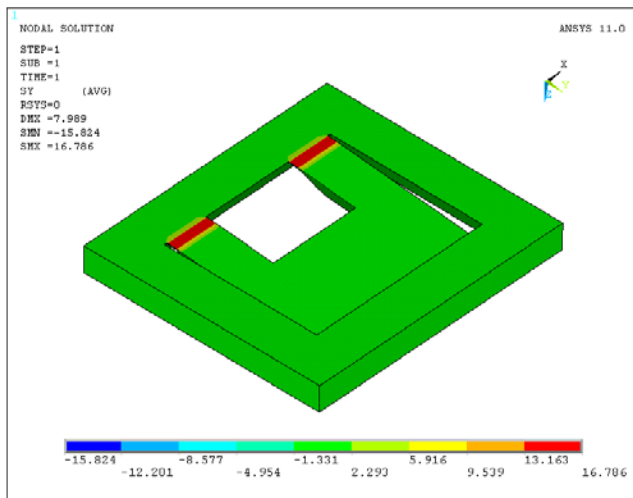


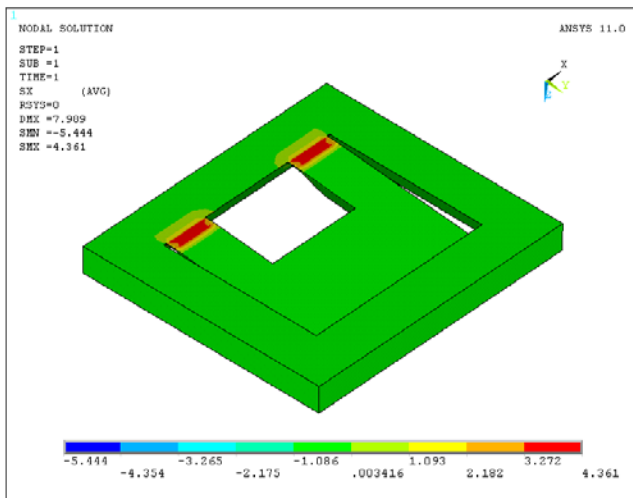
Fig. 6. Distribución de las deflexiones del acelerómetro piezoresistivo.

La Fig. 7 muestra las distribuciones de los esfuerzos longitudinales y transversales sobre la estructura del acelerómetro piezoresistivo. En donde los esfuerzos longitudinales son superiores a los esfuerzos transversales. Esto es adecuado para aumentar la sensibilidad del acelerómetro debido a que mayores esfuerzos causan una elevada deformación en la zona donde se colocarán los elementos piezoresistivos (conectados en un puente de Wheatstone). Así, la resistencia de los piezoresistores tendrá un incremento más pronunciado y el voltaje de salida del puente de Wheatstone será incrementado. Sin embargo, aceleraciones superiores al rango de diseño del acelerómetro puede provocar esfuerzos y desplazamientos que dañen su estructura. Un mecanismo para impedir grandes desplazamientos del acelerómetro, es la colocación de pequeños barreras a una distancia adecuada entre el empaquetado y el acelerómetro. Otro mecanismo para evitar esfuerzos superiores al crítico, es el rediseño de la estructura del acelerómetro. Por ejemplo basado en la ecuación (19), una reducción del esfuerzo longitudinal es producida por el incremento del espesor (v_1) de las dos vigas soportes. Si éste parámetro no puede ser modificado debido al proceso de fabricación, entonces se puede disminuir el ancho (s_2) o longitud de la masa de prueba (L_3). La distribución de estos esfuerzos en la zona de ubicación de los

piezoresistores se muestra en la Fig. 8. Los máximos esfuerzos longitudinales y transversales, sobre los elementos 3408 y 3355, fueron 17.61 MPa y 1.61 MPa, respectivamente. Los esfuerzos obtenidos son inferiores al esfuerzo de falla de 360 MPa para obleas de silicio <100> [24]. Por ende, la configuración estructural del acelerómetro puede tener una operación segura para medir aceleraciones de 10 g para un rango de frecuencias de hasta 569.23 Hz.

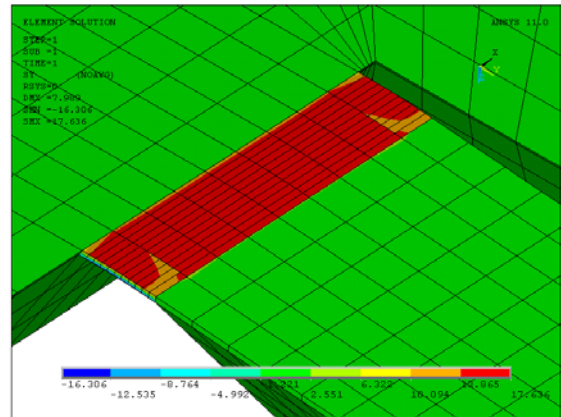


(a)

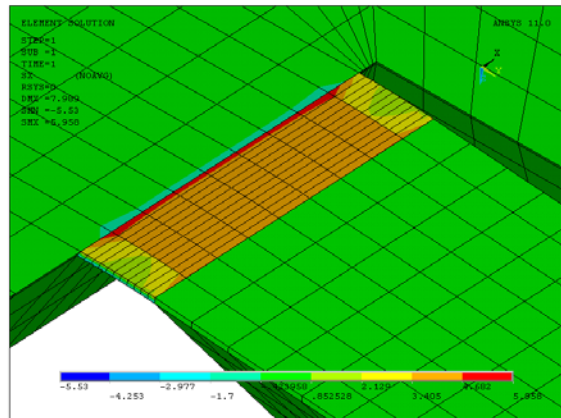


(b)

Fig. 8. Distribución de los esfuerzos (a) longitudinales y (b) transversales del acelerómetro piezoresistivo.



(a)



(b)

Fig. 9. Distribución de los esfuerzos longitudinales y transversales en la zona de ubicación de los piezoresistores.

El modelo analítico y el modelo EF propuesto pueden ayudar a diseñadores en la predicción del comportamiento mecánico de acelerómetros piezoresistivos basados en vigas tipo cantilever. Con estos modelos, el diseñador puede ajustar fácilmente las dimensiones de los parámetros geométricos del acelerómetro para una aplicación específica. Además, el modelo analítico permite un rápido diagnóstico de las magnitudes de la primera frecuencia de resonancia (ancho de banda) y el máximo esfuerzo longitudinal de la estructura del acelerómetro, lo cual puede reducir el tiempo de diseño de esta clase de sensores.

V. CONCLUSIONES

Dos modelos analíticos para la predicción del ancho de banda y el esfuerzo longitudinal máximo

de acelerómetros piezoresistivos han sido presentados. Estos modelos fueron obtenidos usando la teoría de viga de Euler-Bernoulli y el método de Rayleigh. Además, un modelo EF de un acelerómetro piezoresistivo, usando el software ANSYS, fue desarrollado. Los resultados del modelo EF concuerdan con los obtenidos mediante el modelo analítico. Los modelos propuestos son fáciles de utilizar y pueden predecir aceptablemente el ancho de banda y los esfuerzos máximos en la estructura del acelerómetro piezoresistivo basados en vigas tipo cantilever.

VI. AGRADECIMIENTOS

Este trabajo fue apoyado parcialmente por la Universidad de Guanajuato mediante el proyecto DAIP 099/2008 y por CONACYT con el proyecto 84605.

VII. REFERENCIAS

- [1] A. L. Herrera-May, A. R. Cortés-Pérez y Aguilera-Cortés, “*Los microacelerómetros en la actualidad*”, Acta Universitaria Vol. 18, pp. 24-32, 2008
- [2] A. L. Herrera-May, B. S. Soto-Cruz, F. López-Huerta, and L. A. Aguilera-Cortés, “*Electromechanical analysis of a piezoresistive pressure microsensors for low-pressure biomedical applications*”, Rev. Mex. Fis. Vol. 55, pp. 14-24, 2009
- [3] A. L. Herrera-May, P. J. García-Ramírez, L. A. Aguilera-Cortés, J. Martínez-Castillo, A. Saucedo-Carvajal, L. García-González, and E. Figueras-Costa, “*A resonant magnetic field with high quality factor at atmospheric pressure*”, J. Micromech. Microeng. Vol. 19 pp. 15016, 2009.
- [4] N. Yazdi, F. Ayazi, and K. Najafi, “*Micromachined inertial sensors*”, Proc. of the IEEE Vol. 86 pp. 1640-1659, 1998.
- [5] S. Beeby, G. Ensell, M. Kraf and N. White, *MEMS Mechanical Sensors*, Norwood, MA: Artech House Inc, 2004.
- [6] G. K. Fedder, J. Chae, H. Kulah, K. Najafi, T. Denison, J. Kuang, and S. Lewis, *Advanced Micro & Nanosystems. Vol. 2. CMOS-MEMS*. Baltes H, Brand O, Fedder G K, Hierold C, Korvink J, Tabata O (Eds.) Monolithically integrated inertial sensors, Weinheim, Germany: Wiley-VCH, 2005.
- [7] A. R. Sankar, S. K. Lahiri, and S. Das, “*Performance enhancement of a silicon MEMS piezoresistive single axis accelerometer with electroplated gold on a proof mass*”, J. Micromech. Microeng. Vol. 19, pp. 025008, 2009.
- [8] J. Chae, H. Kulah, and K. Najafi, “*A monolithic three-axis micro-g micromachined silicon capacitive accelerometer*”, J. Microelectromech. Syst. Vol. 14, pp. 235-42, 2005.
- [9] U. A. Dauderstädt, P. M. Sarro, and P. J. French, “*Temperature dependence and drift of a thermal accelerometer*”, Sensors Actuators A Vol. 66, pp. 244-9, 1998.
- [10] C. H. Liu and T. W. Kenny, “*A high-precision, wide-bandwidth micromachined tunneling accelerometer*”, J. Microelectromech. Syst. Vol. 10, pp. 425-33
- [11] C. C. Hindrichsen, N. S. Almind, S. H. Brodersen, O. Hansen and E. V. Thomsen, “*Analytical model of a PZT thick-film triaxial accelerometer for optimum design*”, IEEE Sensors J. Vol. 9, pp. 419-29, 2009.
- [12] S. J. Lee and D. W. Cho, “*Development of a micro-opto-mechanical accelerometer based on intensity modulation*”, Microsyst. Technol. Vol. 10, pp. 147-54, 2004
- [13] M. Zhu, P. Kirby, and M. Y. Lim “*Lagrange’s formalism for modeling of a triaxial microaccelerometer with piezoelectric thin-film sensing*”, IEEE Sensors J. Vol. 4, pp. 455-63, 2004.
- [14] C. H. Liu, A. M. Barzilai, J. K. Reynolds, A. Partridge, T. W. Kenny, J. D. Grade, and H. K. Rockstad, “*Characterization of a high-sensitivity micromachined tunneling accelerometer with micro-g resolution*”, J.

- Microelectromech. Syst. Vol. 7, pp. 235-44, 1998.
- [15] G. Kaltsas, D. Goustouridis, and A. G. Nassiopoulou, “A *thermal convective accelerometer system based on a silicon sensor-study and packaging*”, *Sensors Actuators A* Vol. 132, pp. 147-53, 2006.
- [16] A. Llobera, V. Seidemann, J. A. Plaza, V. J. Cadarso and S. Büttgenbach, “*SU-8 optical accelerometers*”, *J. Microelectromech. Syst.* Vol. 16, pp. 111-21, 2007.
- [17] Q. Zou, W. Tan, E. S. Kim, and G. E. Loeb, “Single-and triaxis piezoelectric-bimorph accelerometers”, *J. Microelectromech. Syst.* **17** 45-57, 2008.
- [18] A. Partridge *et al.* “A *high-performance planar piezoresistive accelerometer*”, *J. Microelectromech. Syst.* Vol. 9, pp. 58-66, 2000.
- [19] S. Huang, X. Li, Z. Song, Y. Wang, Yang H, L. Che and J. Jiao, “A *high-performance micromachined piezoresistive accelerometer with axially stressed tiny beams*”, *J. Micromech. Microeng.* Vol. 15, pp. 993-1000, 2005.
- [20] J. A. Plaza, A. Collado, E. Cabruja and J. Steve, “*Piezoresistive accelerometers for MCM package*”, *J. Microelectromech. Syst.* Vol. 11, pp. 794-801, 2002.
- [21] J. Fraden J, *Handbook of Modern Sensors*, New York: Springer, 1996.
- [22] C. F. Beards, *Engineering Vibration Analysis with Application to Control Systems*, United Kingdom: Edward Arnold, 1995.
- [23] S. S. Rao, *Mechanical Vibrations*, 4th ed. Upper Saddle River, New Jersey: Pearson Education, 2004.
- [24] J. M. Borky, *Silicon diaphragm pressure sensors with integrated electronics*, Ph. D. dissertation, University of Michigan, 1997.

VIII. AUTORES

M. en I. Ángel Roberto Cortés Pérez. Se graduó como Ingeniero Mecánico Electricista en la Universidad Veracruzana en 2006. En 2008 obtuvo la Maestría en Ingeniería Mecánica en la Universidad de Guanajuato. Áreas de interés; Vibraciones mecánicas, Optimización, MEMS y Materiales Celulares. Actualmente es estudiante de Doctorado en Ingeniería Mecánica en la Universidad de Guanajuato.

M. en I. Agustín Leobardo Herrera May. Ingeniero mecánico eléctrico, Universidad Veracruzana en 2000. Obtuvo la maestría en Ingeniería Mecánica de la Universidad de Guanajuato en 2002. Ha publicado 13 artículos en revistas indizadas nacionales e internacionales. Investigador del Centro de Investigación en Micro y Nanotecnología de la Universidad Veracruzana desde el 2005. Actualmente estudiante de Doctorado en Ingeniería Mecánica de la Universidad de Guanajuato.

Luz Antonio Aguilera Cortés. Obtuvo la licenciatura, maestría y doctorado en ingeniería mecánica en la Universidad de Guanajuato (UG) en 1988, 1990 y 1995, respectivamente. Áreas de interés; Análisis y Simulación de Sistemas Mecánicos y Mecatrónicos y MEMS. Tiene publicados 24 artículos en revistas indizadas y 60 artículos en congresos (nacional e internacional). Miembro del Sistema Nacional de Investigadores de 1992-1996 y de 2006 a la fecha. Fundador de la carrera de ingeniero en mecatrónica en la UG. Actualmente Profesor Titular A de la UG. Miembro ASME.

Max Antonio González Palacios. Ingeniero mecánico eléctrico, Universidad Iberoamericana-León en 1986. Maestro en ingeniería mecánica en 1989, Universidad de Guanajuato. Ph. D. en 1992 por la Universidad McGill, Canada. De 1997 a 2000 dirigió el Departamento de Investigación y Desarrollo de la Placage Unique, Inc., en Québec. Desde el 2000 es director de Integración de Procesos Industriales, donde ha desarrollado diversos proyectos de investigación aplicada. Actualmente es profesor de la Universidad de Guanajuato y pertenece al SNI (nivel I).

MICROPINZA MEMS BASADA EN ACTUACIÓN ELECTROESTÁTICA

Varona Salazar Jorge^{1,2}, Sáenz Irigoyen Emilio^{1,2}, Fiscal Woodhouse Santiago², Hamoui Anas³

¹Universidad Panamericana campus Bonaterra

²Snowbush Microelectronics Mexico

³McGill University

jvarona@up.edu.mx, esaenz@up.edu.mx, fiscal@snowbush.com, hamoui@mcgill.ca

RESUMEN

Este artículo presenta una micropinza electroestática de placas paralelas fabricada en una tecnología estándar de micromaquinado de superficie. La simplicidad del diseño hace que la pinza sea útil para una amplia gama de aplicaciones en biomedicina y microrobótica. También se presentan modelos analíticos y resultados experimentales para validar el diseño propuesto.

Palabras clave: Microsistemas, MEMS, actuador, micropinza, electrostático.

I. INTRODUCCIÓN

Con el creciente desarrollo de la micro y nanotecnología, existe una gran demanda de microherramientas adecuadas para la manipulación de objetos a pequeña escala. Las aplicaciones de las micropinzas mecánicas son muy diversas, por ejemplo destacan el equipo de microensamblaje de estructuras MEMS complejas, micropinzas biológicas y aplicaciones aeroespaciales [1,2,3].

El objetivo de este trabajo es desarrollar una micropinza capaz de sujetar objetos y partículas menores a 6- μm con aplicaciones probables en biomedicina y microrobótica autónoma. Se ha elegido el accionamiento electroestático debido al bajo consumo de potencia y funcionamiento a temperatura ambiente.

II. DISEÑO CONCEPTUAL

Básicamente, una buena micropinza debe ser capaz de sostener firmemente el objeto de interés con una fuerza que sea lo suficientemente grande como para sujetarlo entre las tenazas y, al mismo tiempo,

lo suficientemente pequeña para evitar dañar el material.

Las micropinzas tradicionales se basan en la topología de accionamiento tipo “peine” descrita en [4, 5] la cual, consiste en un arreglo de capacitores de dedos interdigitados donde un grupo está fijo al sustrato y el otro se encuentra libre y es móvil suspendido con resortes. En un actuador típico de tipo peine, la capacitancia es lineal con respecto al desplazamiento y la fuerza motriz se asume independiente de la posición de los dedos móviles. Mientras estos actuadores son excelentes para aplicaciones de sensado como acelerómetros, resonadores y giroscopios [6], su aplicación como micropinzas presenta algunas limitaciones. Por ejemplo, la configuración tipo peine requiere una gran cantidad de área y requiere del uso de algún tipo de transmisión para acoplar las pinzas. Un ejemplo de implementación de una micropinza basada en la topología tipo peine se muestra en [7] y se presenta en la Figura 1.

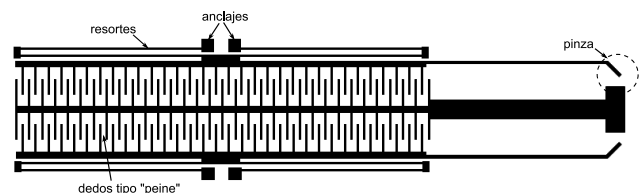


Figura 1: Actuador típico tipo peine.

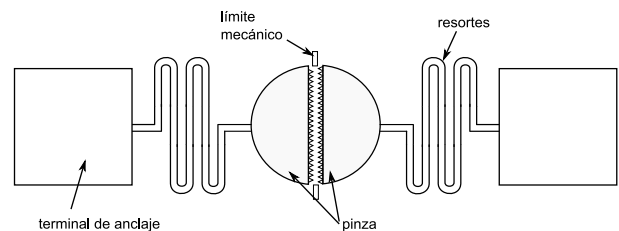


Figura 2: Micropinza propuesta basada en capacitancia de placas paralelas.

El diseño de la micropinza propuesto en este trabajo consiste en dos estructuras de polisilicio libres y separadas por un pequeño espacio entre ellas semejante a un capacitor de placas paralelas. Las dos estructuras están soportadas por resortes tipo serpentina que se encuentran sujetos a una terminal de anclaje en sus extremos. Cuando se aplica una diferencia de potencial entre las dos estructuras, éstas tenderán a colapsarse debido a la atracción electrostática, cerrando así, el espacio que las separa. Los objetos que se encuentren en el espacio serán sujetos por las estructuras de la pinza. Las caras de las estructuras están provistas con “dientes” triangulares para cerciorarse que el objeto será asegurado y que no resbalará. Un diagrama de la micropinza propuesta se muestra en la Figura 2.

Este diseño ofrece numerosas ventajas sobre la pinza electrostática basada en actuadores tipo peine. Primero, la pinza forma parte del actuador por lo que no se necesita una transmisión mecánica de acoplamiento adicional. Segundo, la capacitancia (y por tanto la fuerza motriz) se incrementa con el desplazamiento de las estructuras por lo que se requieren menores voltajes por unidad de área para cerrar la pinza en comparación con los actuadores tipo peine. Aunque el control de la posición y de la estabilidad se vuelve más complicado, esto no constituye un problema en donde las aplicaciones de interés requieren solo un estado completamente abierto o cerrado del actuador. Tercero, el diseño de la micropinza basado en la capacitancia de placas paralelas puede ser liberado del sustrato utilizando técnicas *flip-chip* más fácilmente que los actuadores de peine para poder construir micromáquinas más complejas.

III. ANÁLISIS Y MODELADO

En el actuador típico tipo peine de la Figura 1, la fuerza motriz se obtiene mediante los campos eléctricos de los bordes de los dedos del peine. Aunque la capacitancia total se incrementa linealmente con el desplazamiento, la capacitancia lateral que se forma donde se traslapan los dedos fijos y los dedos móviles tiene una contribución mínima a la fuerza de accionamiento [8]. En tal

caso, los campos eléctricos de los bordes se asumen constantes considerando que los dedos son muy largos con respecto a su desplazamiento. Así, la fuerza de accionamiento resultante es aproximadamente fija dentro de un rango de voltajes de entrada. Esta es una propiedad deseable ya que se obtiene una relación estable entre el desplazamiento y el voltaje aplicado sobre un rango amplio de x de acuerdo con (1):

$$x = \frac{n\varepsilon V^2 t}{k_s g} \quad (1)$$

donde n es el número de dedos del peine, ε es la permitividad del dieléctrico, V es el voltaje aplicado, t se refiere al grosor de los dedos, g es el espacio entre los dedos del peine y k_s es la constante del resorte del sistema.

En contraste con el actuador típico de peine basado en capacitancias de bordes y fuerza constante, la micropinza electrostática aquí propuesta y descrita en la Figura 2, emplea un método más directo de accionamiento por placas paralelas. Tradicionalmente, los actuadores de placas paralelas se han ignorado debido a la inestabilidad del desplazamiento en función del voltaje de entrada. Cuando se aplica voltaje, la fuerza electrostática entre las placas hace que se cierren, sin embargo, debido a que la capacitancia también incrementa con el desplazamiento, la fuerza de atracción crece geométricamente y, finalmente, las placas tienden a colapsarse. Así, el rango útil del desplazamiento de las estructuras del actuador está limitado a una fracción del total del espacio entre las placas para prevenir que se colapsen. Ciertamente esto es un problema en aplicaciones donde se requiere un control preciso de la posición, sin embargo, es de poca importancia en aquellas aplicaciones donde la micropinza se utiliza como interruptor. También, se pueden agregar bordes mecánicos para regular el máximo desplazamiento y evitar que las placas se toquen entre sí como se indica en la Figura 2.

El modelo que representa al actuador de placas paralelas se muestra en la Figura 3. La capacitancia total entre las placas es:

$$C = \epsilon_o \epsilon_r \frac{A}{d_o} \quad (2)$$

Donde ϵ_o y ϵ_r son las permitividades del espacio libre y del aire respectivamente, A es el área del capacitor formado por las placas, y d_o representa la apertura de la micropinza con un voltaje de entrada igual a cero.

La fuerza electroestática entre las placas está dado por:

$$F_e = \frac{1}{2} \frac{C}{d_o} V^2 \quad (3)$$

También:

$$d = d_o - x \quad (4)$$

donde x es el desplazamiento neto de las placas del capacitor y d es el espaciamento instantáneo o apertura de la pinza.

La fuerza de accionamiento como función del desplazamiento se obtiene substituyendo (2) y (4) en (3)

$$F_e = \frac{1}{2} \epsilon_o \epsilon_r \frac{A}{d^2} V^2 \quad (5)$$

Así mismo, la fuerza mecánica de restauración debido a los resortes que soportan las placas está dada por la ley de Hook:

$$F_s = -k_s (d_o - d) \quad (6)$$

donde k_s es la constante total de resorte del sistema.

El estado de equilibrio se obtiene cuando $F_e = -F_s$. Sin embargo, cuando se incrementa el voltaje de entrada, las placas paralelas se atraen entre sí hasta que llegan a un “punto de atracción” donde la fuerza electroestática es demasiado grande haciendo que las placas se colapsen. Esto es, $F_e \rightarrow \infty$ cuando $d \rightarrow 0$. Así, el punto de atracción representa el máximo desplazamiento que se puede obtener antes de que el sistema se vuelva inestable y las placas se colapsen. Dicho punto también puede describirse como el mínimo espaciamento estable entre las placas:

$$d_{\min} = d_o - x_{\max} \quad (7)$$

El punto de atracción se alcanza cuando las fuerzas electroestática y de reacción del resorte son tangenciales e iguales en magnitud. Igualando (5) y (6) y tomando la derivada con respecto a d se obtiene:

$$\epsilon_o \epsilon_r \frac{A}{d^3} V^2 = k_s \quad (8)$$

Así:

$$V^2 = \frac{k_s d^3}{\epsilon_o \epsilon_r A} \quad (9)$$

Substituyendo (9) dentro de $F_e = -F_s$ para determinar d_{\min} se obtiene:

$$d_{\min} = \frac{2}{3} d_o \quad (10)$$

La constante k_s se calcula basándose en la geometría del resorte. Existen diferentes diseños de resortes que se pueden aplicar en actuadores MEMS, algunos ejemplos incluyen los de tipo pata de cangrejo, suspensiones flexibles, vigas desdobladas, espirales Arquideanas y los resortes de serpentinatas [9, 10]. Los resortes deben diseñarse para ser lo suficientemente flexibles a lo largo del eje x (dirección de accionamiento), pero rígidos en el eje y para oponerse a cualquier movimiento potencial en otra dirección. Sin embargo, incrementar la rigidez de las suspensiones implica fuerzas electroestáticas más grandes y, por tanto, voltajes más grandes para obtener el desplazamiento deseado.

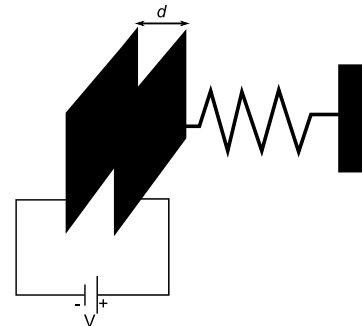


Figura 3: Modelo de un actuador electrostático de capacitor de placas paralelas.

Para el diseño de la micropinza de este trabajo, se eligió un resorte de serpentina curva como se muestra en la Figura 2. Para la optimización del diseño, el modelo analítico de la constante del resorte se derivó basándose en [11]. Primero, el resorte de serpentina se divide en dos segmentos como se ilustra en la Figura 4 y la constante total del resorte es la combinación de la rigidez de todas las flexuras. La constante total de resorte para un solo segmento se aproxima según:

$$k_{seg} = \frac{hG}{b^2 \left(\frac{hG}{EI} b + 3a \right)} \quad (11)$$

donde b es la longitud de la viga, a es la longitud horizontal o viga de conexión, I representa el momento de torsión inercial para una viga rectangular, G es el módulo de elasticidad para materiales isotrópicos, E es el módulo de Young (162-GPa para el polisilicio) y h es una constante que considera el factor de forma.

Los parámetros necesarios para calcular k_{seg} se estiman como:

$$I = \frac{tw^3}{12} \quad (12)$$

$$G = \frac{E}{2(1+\nu)} \quad (13)$$

$$h = tw^3 \left[\frac{1}{3} - 0.21 \frac{w}{t} \left(1 - \frac{w^4}{12t^4} \right) \right] \quad (14)$$

donde ν es la relación de Poisson para el silicio ($\nu=0.3$), w es el ancho y t es el grosor de la viga. La ecuación (10) es la constante del resorte de un segmento por lo que la constante total del sistema, k_s , se obtiene mediante la sumatoria de los N segmentos calculados usando (10).

$$\frac{1}{k_s} = \sum_{n=1}^N \frac{1}{k_{seg(n)}} \quad (15)$$

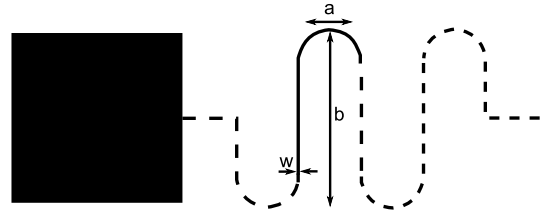


Figura 4: Modelo del resorte de serpentina indicando los parámetros geométricos de un segmento.

Tabla I: Dimensiones del actuador.

Geometría	Valor
Apertura de la pinza con voltaje cero (d_0)	6- μm
Longitud de las placas electroestáticas (L_p)	100- μm
Grosor de las placas electroestáticas (t_p)	3.5- μm
Longitud de las vigas verticales (b)	126- μm
Longitud de las vigas de conexión (a)	18- μm
Ancho de las vigas (w)	6- μm
Grosor de las vigas (t)	2- μm

Un resumen de las dimensiones utilizadas en el diseño de la micropinza propuesta se presentan en la Tabla I.

Usando los parámetros geométricos de la Tabla I y basándose en (9) y (4), se puede obtener el voltaje mínimo requerido para cerrar la pinza. En este caso, el punto de atracción ocurre cuando el espacio entre las placas es de $d_{min}=4\mu\text{m}$ y el voltaje requerido para cerrar la pinza es de 63-V.

Los resultados para el análisis de operación de la micropinza de placas paralelas dentro del rango de atracción se muestran en las gráficas de las Figuras 5 y 6.

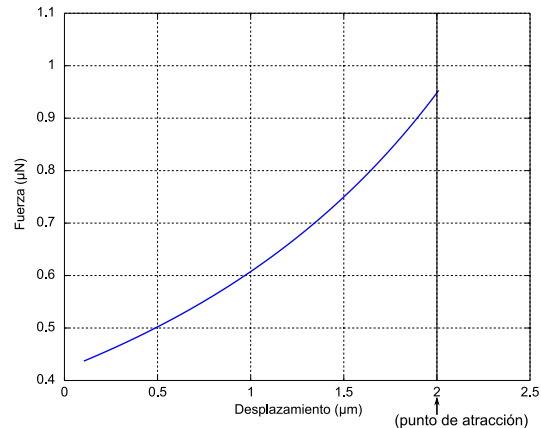


Figura 5: Variación de la fuerza de accionamiento como función del desplazamiento de las placas dentro del rango de atracción.

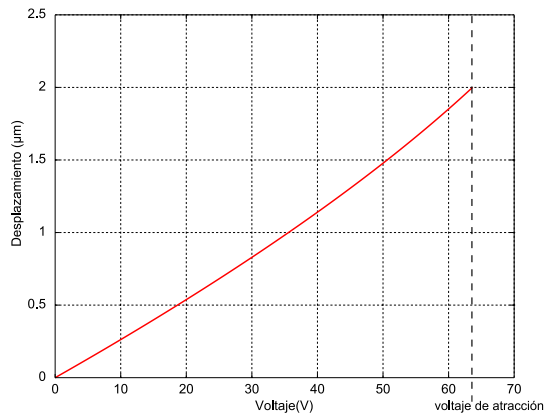


Figura 6: Desplazamiento de las placas vs cambios de voltaje dentro del rango de atracción.

Si el voltaje aplicado excede el valor del punto de atracción de 63-V las placas colapsarán y la pinza se cerrará por completo.

IV. RESULTADOS EXPERIMENTALES

La micropinza presentada en este trabajo fue fabricada usando el procesos estándar MEMS Multi-Usuario (MUMPs) [12]. MUMPs ofrece tres capas de polisilicio y dos capas sacrificables de fosfosilicato sobre un aislante de nitruro de silicio. Las dos últimas capas de polisilicio pueden liberarse. Una capa de oro puede evaporarse sobre la superficie al final del proceso mediante el uso de la deposición química de vapor de baja presión. Después de la construcción, las capas sacrificables se remueven utilizando una solución de ácido fluorhídrico.

La micropinza fabricada se muestra en la Figura 7. Para facilitar las pruebas, el microchip se alambrió y encapsuló dentro de un paquete plástico DIP-40. Para controlar la aplicación del voltaje de entrada durante las pruebas, se diseñó una tarjeta de circuito impreso (PCB) que incluye manejadores de alto voltaje. Una cámara digital CCD fue montada en la estación de prueba para capturar el desplazamiento de la pinza. La configuración inicial de la estación de prueba se muestra en la Figura 8.

En los experimentos, el voltaje de entrada fue incrementado gradualmente para encontrar el punto de atracción. Los resultados experimentales obtenidos se compararon con los resultados

predichos por el modelo analítico como se muestra en la Figura 9.

Convenientemente, el voltaje de atracción resultó alrededor de un 27% más bajo que el estimado analíticamente. Esta diferencia puede explicarse por el hecho de que los modelos utilizados para calcular la constante del resorte no toman en cuenta las características no ideales del proceso de fabricación. Los modelos convencionales ignoran los esfuerzos residuales que se inducen durante la fabricación de la mayoría de las microestructuras y la deformación que ocurre en las áreas donde se unen dos o más elementos. En resortes con vigas que forman dobleces, la interacción bi-axial de los esfuerzos residuales y los puntos de acoplamiento en áreas del resorte sometidas a esfuerzos externos es muy compleja por lo que el modelado es difícil de representar con precisión.

Así, con base en los datos experimentales obtenidos, una señal con una amplitud mayor a 46-V es suficiente para operar la pinza electroestática de la Figura 7.

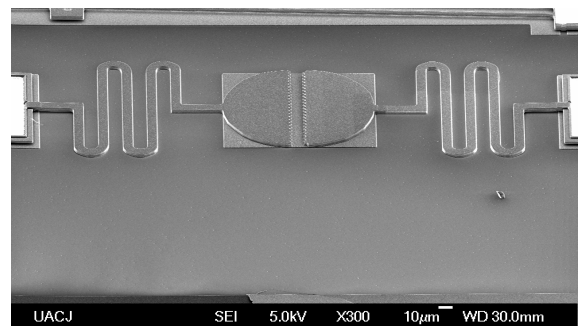


Figura 7: Imagen de microscopio electrónico de la pinza fabricada.

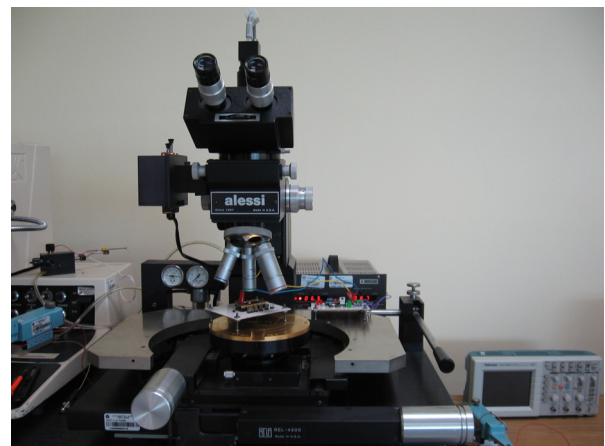


Figura 8: Configuración y equipo de pruebas.

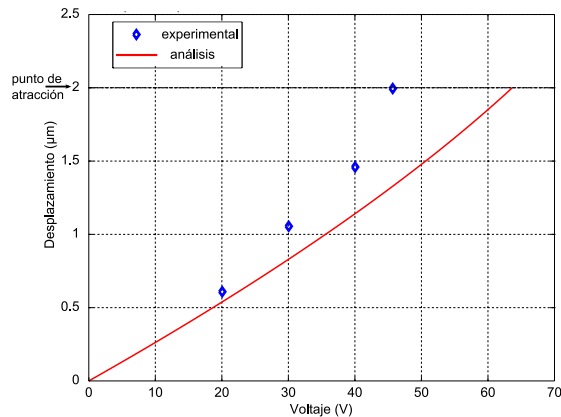


Figura 9: Comparación de los resultados analíticos y experimentales.

V. CONCLUSIONES

Se validó el diseño de una micropinza basada en accionamiento electrostático de placas paralelas. La configuración de placas paralelas ofrece una buena alternativa para realizar actuadores de dos posiciones donde el control de posiciones intermedias no es de interés. Los actuadores de placas paralelas ocupan una menor área que la que se requiere en un diseño tipo peine para una fuerza de salida dada y son más adecuados para el microensablaje y post-procesamiento requerido en la realización de micromáquinas complejas. El dispositivo fabricado requiere un voltaje $>45\text{-V}$ para cerrar la pinza y puede aplicarse en brazos microrobóticos. Un menor valor de voltaje de atracción puede lograrse mediante la optimización de los resortes del sistema.

VI. AGRADECIMIENTOS

Esta investigación fue patrocinada parcialmente por el Instituto Panamericano de Alta Dirección de Empresas (IPADE) a través del Centro Panamericano de Investigación e Innovación (CEPii).

VII. REFERENCIAS

[1] G. Greitmann and R. Buser, "Tactile microgripper for automated handling of microparts," 8th Intl. Conf. Solid-State Sensors and Actuators (Euroensors IX), Stockholm, Sweden, pp. 372-375, 1995.

- [2] C. J. Kim, A. P. Pisano, R. S. Muller, and M. G. Lim, "Polysilicon microgripper", in Tech. Dig. IEEE Solid-State Sensor and Actuator Workshop, pp. 48-51, 1990.
- [3] P. Dario, M. Carrozza, C. Stefanini, Dapos, S. Attanasio, "A Mobile Microrobot Actuated by a New Electromagnetic Wobble Micromotor", IEEE/ASME Transactions on Mechatronics, Vol. 3, No. 1, pp.9-16, March 1998.
- [4] A. P. Pisano, "Resonant-structure micromotors," in Proc. IEEE Micro-Electro-Mechanical Systems Workshop, pp. 1-8, 1989.
- [5] W.C. Tang, T.H. Nguyen, M.W. Judy, R.T. Howe, "Electrostatic-comb Drive of Lateral Polysilicon Resonators", J. Sensors and Actuators, Vol. A21-A23, pp. 328-331, 1990.
- [6] R. T. Howe, R.S. Miller, K.J. Gabriel, W.S.N. Trimmer, "Silicon Micromechanics: Sensors and Actuators on Chip", IEEE Spectrum, pp. 29-35, July 1990.
- [7] I.P.F. Harouche, C. Shafai, R. Gordon, "Design and simulation of a microtweezers using a controlled displacement comb drive", IEEE, pp. 341-343, 2006.
- [8] T. Kuendiger, G.M. Howard, P. Mokrian, M. Ahmadi, W.C. Miller, "Design and Analysis of Planar and Lattice Electrostatic Comb Drive Actuators", Proc. of IEEE-NEWCAS, pp. 19-22, 2005.
- [9] R. Legtenberg, A.W. Groeneveld, M. Elwenspoek, "Comb-Drive Actuators for Large Displacements", J. Micromech. Microeng., Vol. 6, pp. 320-329, 1996.
- [10] W.C. Tang, T.H. Nguyen, R.T. Howe, "Laterally Driven Polysilicon Resonant Microstructures", J. Sensors and Actuators, Vol. 20, pp. 25-32, 1989.
- [11] H.A. Rouabah, C.O. Gollasch, M. Kraft, "Design and Optimisation of an Electrostatic MEMS Actuator with Low Spring constant for an 'Atom Chip'", Proc. NSTI Nanotech, Vol.3, pp. 489-492, May. 2005.
- [12] J. Carter, et al., PolyMUMPS Design Handbook, MEMSCAP Inc., Revision 11.0, 2005.

DISEÑO DE UN SISTEMA ACONDICIONADOR DE SEÑAL PARA SENSORES MEMS DE REACTANCIA VARIABLE

Sosa Torres Miguel Ángel, Navarro Alvarado José Marcos, Montiel Uribe Rodrigo, Pérez Márquez Heber Bernabe

Centro de Investigación y Diseño en Micro y Nano Tecnología (CIDEMyNT), Departamento de Electrónica
Instituto Tecnológico superior de Irapuato
Km. 12.5 Carretera Irapuato-Silao, Irapuato, Gto. 36821, México
Tel. (462)6067900, correo electrónico: misosa@itesi.edu.mx, jonavarro@itesi.edu.mx, romontiel@itesi.edu.mx,
heperez@itesi.edu.mx

RESUMEN

En la actualidad, cada vez es más común el uso de sensores MEMS (Sistemas Microelectromecánicos por sus siglas en inglés Microelectromechanical Systems) en diversas aplicaciones. Estos sensores pueden ser de diversos tipos, capacitivos, inductivos, resistivos, etc., y también se pueden clasificar de acuerdo a su aplicación, como son sensores de presión, posición, giro, etc.

En este trabajo se presenta una breve descripción de los sensores del tipo capacitivo, la caracterización de un circuito acondicionador de señal para sensores MEMS de reactancia variable utilizando el software de diseño y simulación "National Instruments Electronics Workbench 10", el cual tiene la característica de utilizar Modelos de dispositivo desde proveedores líderes como Dispositivos Analógicos™ y Texas Instruments™.

Se presentan de manera gráfica los resultados obtenidos en la respuesta a la frecuencia de un circuito utilizado como acondicionador de señal para sensores capacitivos, además de la respuesta del sistema a diferentes valores del dispositivo de retroalimentación R. Observándose que se tiene mejor respuesta dependiendo de la combinación de el valor del dispositivo R y frecuencias altas.

Se describen además las consideraciones que se deben de tomar para la realización de estos sistemas de acuerdo a las características eléctricas de sus dispositivos.

Área: Microelectrónica.

Palabras clave: Reactancia capacitiva, Sensor capacitivo MEMS, Respuesta a la capacitancia.

I. Introducción

La caracterización de sistemas de acondicionamiento es el conjunto de elementos de un sistema eléctrico o electrónico, aunque también pueden involucrar partes mecánicas [6], estos surgen a partir de un sistema de medida que se produce en la señal de salida, una señal apta para ser presentada o bien que permitirá un procesamiento de señal posterior mediante un equipo o instrumento estándar. El Acondicionamiento de señales ofrece muchas ventajas de mejoras sobre los circuitos, como lo son la amplificación de la señal, el filtrado y la modulación ó demodulación de la señal con que se trabaja [1].

Un sistema de medición consiste de un elemento sensor, un acondicionador de señal y un sistema de visualización. El elemento sensor es aquel dispositivo donde un cambio en la variable física proporciona un cambio de una propiedad eléctrica. El acondicionador de señal es aquel elemento que se encarga de convertir ese cambio de una propiedad eléctrica a un cambio de una

señal eléctrica, de tipo y magnitud adecuada para el elemento de visualización. El elemento de visualización se encarga de que esa señal eléctrica proveniente del acondicionador de señal sea entendible para un operador o para un sistema de control. En la figura 1, se observa un diagrama a bloques de los sistemas de medición.

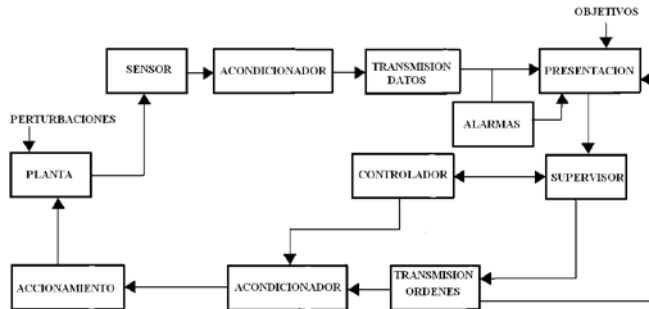


Figura 1. Diagrama general de un sistema de medida y control [1]

Una aplicación en desarrollo es la medición de la temperatura de fluidos a través de su reactancia capacitiva [5].

Acondicionamiento de señales

Los acondicionadores de señal, adaptadores o amplificadores, en sentido amplio, son los elementos del sistema de medida que ofrecen, a partir de la señal de salida un sensor electrónico, una señal apta para ser presentada o registrada o que simplemente permitirá un procesamiento posterior mediante un equipo o instrumento estándar. Consiste normalmente en circuitos electrónicos que ofrecen entre otras funciones, las siguientes:

- Amplificación
- Filtrado
- Adaptación de Impedancias
- Modulación o Demodulación.

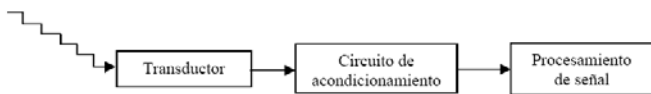


Figura 2. Etapas del acondicionamiento de señal.

Un transductor no se conecta directamente a la parte de procesamiento o de despliegue de un sistema, ya que la señal que nos envía nuestro transductor por lo general es muy débil o contiene ruido u otros componentes (elementos transitorios o acoplamiento de señales no deseadas), por eso se realizan etapas de acondicionamiento de señales.

El circuito de acondicionamiento puede proporcionar una impedancia de entrada adecuada para no demandar mucha corriente al transductor [2] como se ilustra en la figura 3.

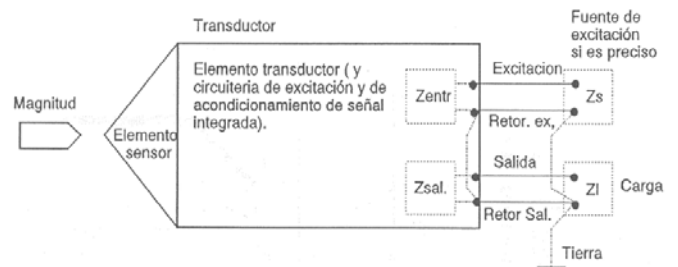


Figura 3. Sistema de Acondicionamiento de señal para un sensor desde el punto de vista eléctrico [1].

II. Método

Un condensador eléctrico consiste en dos conductores separados por un dieléctrico (sólido, líquido y gaseoso), o el vacío como se ve en la figura 4. La relación entre la carga, Q, y la diferencia de potencial, V, entre ellos viene descrita por su capacidad, $C= Q/V$. Esta capacidad depende de la disposición geométrica de los conductores y del material dieléctrico dispuesto entre ellos.

Para un condensador formado por n placas planas paralelas iguales, con área A, distancia d entre cada par, y un material entre ellas con constante dieléctrica relativa ϵ_r , la capacitancia aproximada es:

$$C = \epsilon_0 \epsilon_r A/d (n-1) \tag{1}$$

Donde $\epsilon_0 = 8.85 \text{ pF/m}$ es la constante dieléctrica del vacío.

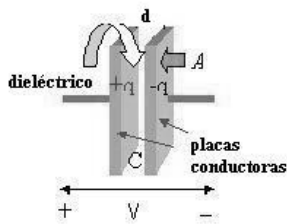


Figura 4. Capacitor de placas paralelas [1]

Así pues, cualquier fenómeno o magnitud que produzca una variación en ϵ_r , A o d, provocará un cambio en la capacidad C, y en principio puede ser detectado mediante el dispositivo anterior. En general cualquier cambio en el dieléctrico o en la geometría puede ser considerado para la detección del fenómeno que lo provoca.

Los sensores capacitivos pueden ser lineales o no lineales en sí mismos. Su linealidad depende del parámetro que varía y de si se mide la admitancia o la impedancia del condensador. En un condensador plano, la salida es lineal si se mide la admitancia, pero no es lineal si se varía la separación entre placas, para esto se tiene la relación:

$$C = \epsilon \frac{A}{d(1+\alpha)} \quad (2)$$

Dónde $\alpha = \frac{x}{d}$.

Una alternativa para tener una salida lineal en un sensor capacitivo plano basado en una variación de la distancia entre placas, es medir, en vez de su admitancia, su impedancia:

$$Z = \frac{1}{j\omega C} \quad (3)$$

El empleo de condensadores diferenciales, permite también obtener una salida lineal con el parámetro que se detecta. Una última consideración relativa a las limitaciones de los sensores capacitivos es su alta impedancia de

salida. Si bien la impedancia decrece cuanto mayor sea la frecuencia de alimentación, la impedancia de las capacidades parásitas también decrece al aumentar su frecuencia. Una solución es poner la electrónica de acondicionamiento de señal cerca. Frente a sus limitaciones, los sensores capacitivos presentan una serie de ventajas que los hacen atractivos en muchas aplicaciones: como sensores de desplazamiento tienen un error por carga mecánica mínimo, al no haber contacto mecánico directo, no hay errores de fricción ni de histéresis y no hay que hacer mucha fuerza para desplazar al elemento móvil [1].

Los sensores capacitivos son a menudo utilizados exitosamente en las aplicaciones que no pueden ser resueltas por otras técnicas de sensado. Estos responden a un cambio de dieléctrico en el medio que rodea la zona activa y, por medio de la regulación incorporada, permite sensar prácticamente cualquier sustancia. Además pueden detectar materiales a través de paredes de vidrio, plástico o láminas de cartón.

Los sensores capacitivos como su nombre indica, están basados en la detección de un cambio en la capacidad inducido por una superficie que se lleva cerca del elemento sensor. Son diseñados normalmente por medio de platos paralelos con electrodos porosos o con filamentos entrelazados en el sustrato. El material dieléctrico absorbe o elimina vapor de agua del ambiente con los cambios del nivel de humedad. Los cambios resultantes en la constante dieléctrica causa una variación en el valor de la capacidad del dispositivo por lo que resulta una impedancia que varía con la humedad. Esto permite a el vapor de agua entrar y salir fácilmente y el secado rápido para la sencilla calibración del sensor.

El elemento sensor es un condensador constituido por un electrodo sensible y un electrodo de referencia. Estos electrodos pueden ser, por ejemplo, un disco y un anillo metálicos separados por un material dieléctrico. Una cavidad de aire seco se suele colocar detrás del elemento capacitivo para proporcionar aislamiento. El resto del sensor está constituido por circuitos

electrónicos que pueden incluirse como una parte integral de la unidad, en cuyo caso suelen estar embebidos en una resina para proporcionar soporte mecánico y sellado.

La figura 5 muestra un diagrama esquemático de un capacitor MEMS. Consiste de dos placas móviles con una capa dieléctrica en la parte superior de la placa inferior.

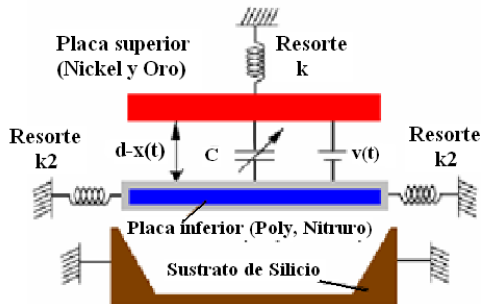


Figura 5. Diagrama esquemático de un capacitor MEMS [2].

Los capacitores MEMS de este tipo con dimensiones de alrededor de 1000 μm tienen sus capacitancias alrededor de 10 pF, dependiendo de las distancias de separación entre las placas.

Un circuito propuesto para realizar las mediciones del sensor capacitivo, es el siguiente:

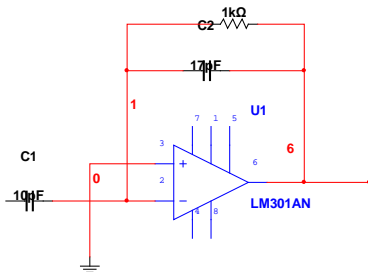


Figura 6. Circuito acondicionador de reactancia variable [3].

Para este circuito se debe de considerar las siguientes características eléctricas del amplificador operacional:

- Inmunidad al ruido

- Alta ganancia
- Ancho de banda
- Alimentación de voltaje
- Nivel de offset de entrada de voltaje o corriente

Los diagramas esquemáticos y simulaciones se realizaron utilizando el software NI Multisim 10, ya que las características de los dispositivos simulados, son datos proporcionados directamente por el fabricante de estos. Utilizando un voltaje de entrada senoidal de 1 Vp con el cual se caracterizo el circuito para diferentes valores de la frecuencia de entrada y diferentes valores del capacitor de retroalimentación, para con esto cambiar la reactancia de este y determinar a qué frecuencias presenta una mejor respuesta el diseño.

En la figura 7 se muestra el circuito realizado en simulación con el software NI Multisim 10.

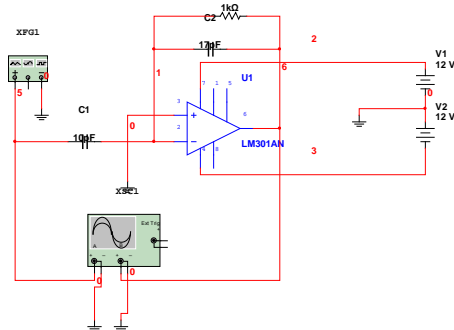


Figura 7. Simulación del circuito de reactancia variable.

En esta figura se observa que la señal de entrada está conectada al canal A del osciloscopio y la salida al canal B de este mismo. En la figura 8 se muestra una gráfica con la señal de salida con un valor en C2 de 17 pF y una frecuencia de entrada de 100 Hz, y en la tabla 1 se presenta el recorrido en la frecuencia.

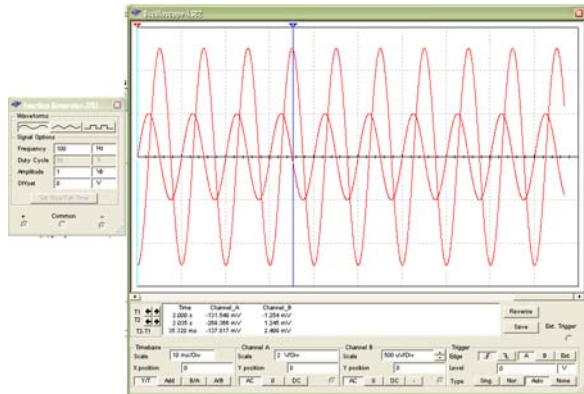


Figura 8. Graficas como resultado de la simulación a 100 Hz y $C_x=C_2=17$ pF. Se muestra la forma de onda seleccionada.

Las tablas 1, 2 y 3 muestran un recorrido en frecuencia con estos valores de reactancia y voltaje de entrada. Se obtiene el valor del voltaje de salida, tomando en cuenta que esta anulado el offset de salida en la medición.

Frec (Hz)	Vin (Vp)	C2 (pF)	Vout (Vp)
10	1	100	1.3 u
100	1	100	13 u
1000	1	100	130 u
10k	1	100	1.3 m
100K	1	100	14 m
1 M	1	100	80 m
10 M	1	100	110 m

Tabla 1. Salida en voltaje para un valor de C2 de 17pF.

Frec (Hz)	Vin (Vp)	C2 (pF)	Vout (Vp)
10	1	50	1.3 u
100	1	50	13 u
1000	1	50	130 u
10k	1	50	1.3 m
100K	1	50	13 m
1 M	1	50	90 m
10 M	1	50	110 m

Tabla 2. Salida en voltaje para un valor de C2 de 50pF.

Frec (Hz)	Vin (Vp)	C2 (pF)	Vout (Vp)
10	1	17	1.25 u
100	1	17	12.5 u
1000	1	17	125 u
10k	1	17	1.25 m
100K	1	17	12.5 m
1 M	1	17	92 m
10 M	1	17	125 m

Tabla 3. Salida en voltaje para un valor de C2 de 100pF.

Con estos datos, nos damos cuenta que se obtiene un voltaje de salida mayor a frecuencias altas.

Las figuras 9, 10 y 11 muestran la respuesta de la reactancia con diferentes valores de frecuencia en el voltaje de entrada. Se obtiene el valor del voltaje de salida, tomando en cuenta que esta anulado el offset de salida en la medición.

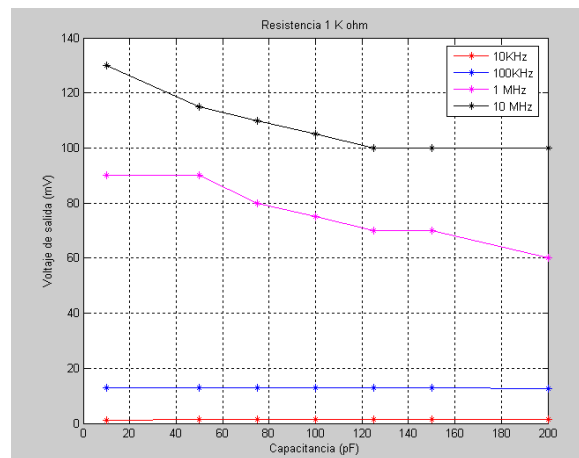


Figura 9. Respuesta a la capacitancia para un valor de R=1KΩ.

De acuerdo con los resultados obtenidos, se observa que el circuito responde mejor a las frecuencias de 1 y 10 MHz, ya el voltaje de salida varía 30 mVp en ambas frecuencias cuando el valor del capacitor de retroalimentación varía de 10 a 200 pF.

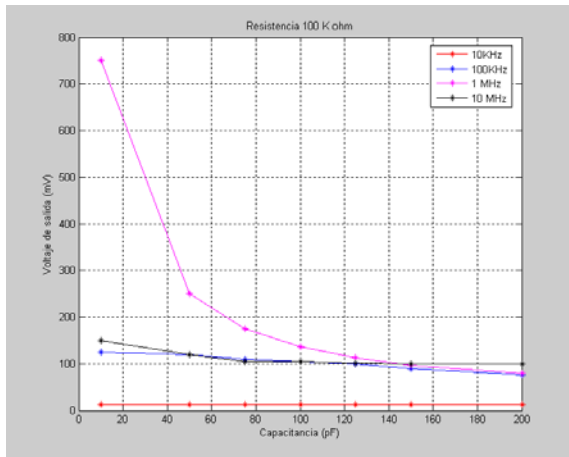


Figura 10. Respuesta a la capacitancia para $R=10K\Omega$.

De acuerdo a la gráfica y a los datos obtenidos en la tabla 2 se puede observar que con una resistencia de $10K\Omega$ se obtiene una mejor respuesta del circuito a cambios del capacitor a una frecuencia de $1MHz$, ya que el voltaje de salida cambia de 750 mVp a 80 mVp cuando el capacitor varía de 10 pF a 200 pF .

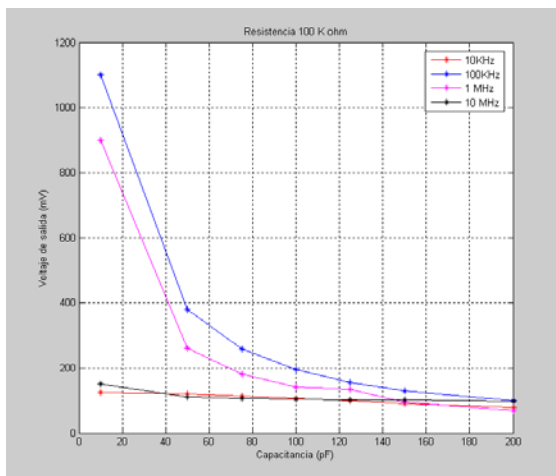


Figura 11. Respuesta a la capacitancia para $R=100K\Omega$.

En la figura 11 se puede observar que la salida presentada por el circuito acondicionador con una resistencia de $100K\Omega$ es mucho mejor en respuesta a los cambios del capacitor a una frecuencia de $100KHz$, ya que el voltaje de salida cambia de 1.1 Vp a 100 mVp cuando el capacitor varía de 10 pF a 200 pF .

III. Conclusiones

Debido a que los cambios en la salida del circuito acondicionador son muy reducidos, ya que se obtienen variaciones de microvolts o milivolts hasta volts en altas frecuencias, se debe de tener cuidado con utilizar amplificadores operacionales que sean inmunes al ruido y además tengan un ancho de banda amplio para un mejor desempeño del circuito.

Otra de las observaciones que se deben de tener presente, es que los cables que inyectan la señal de entrada deben de estar “apantallados” a tierra así como la salida de estos, con el fin de evitar la introducción de señales no deseadas.

En este trabajo, se realizaron las simulaciones con tres valores diferentes de resistencia de polarización, tratando de cubrir un rango amplio de valores que son los más utilizados en los circuitos.

De esta manera se puede determinar a qué valores de los componentes se puede obtener mejor respuesta, dependiendo de la frecuencia a utilizar

Con este diseño se pretende utilizar cualquier sensor MEMS del tipo capacitivo, independientemente de la forma o estructura con la que este fabricado. Se recomienda que para este tipo de circuitos se deba tener en consideración que el capacitor sea en el orden de los pico-Farads.

IV. Referencias

- [1] Pallás, A. Ramón. (2007) *Sensores y Acondicionadores de Señal*, 4ª. Edición. Editorial Marcombo, México. D.F
- [2] Maher Bakri-Kassem and Raafat R. Manssur. *A Wide Tuning Range MEMS Variable Capacitor*. University of Waterloo, Waterloo, Ontario, Canada. CWMEMS Abstract (2003).

- [3] *Quad Precision Op Amp (LT1014), Dual Precision Op Amp (LT1013) Datasheet*, Linear Technology en <http://www.linear-tech.com/prod/datasheet.html?datasheet=155>
- [4] Stephen D. Senturia, *Microsystem Design*, Cap. 19, Kluwer Academic Publisher, USA, 2002
- [5] Behzadi G., Golnabi H., *Monitoring Temperature Variation of reactance Capacitance of Water using a Cylindrical Cell Probe*, Journal of Applied Sciences, 2009, ISSN 1812-5654
- [6] Golnabi H. *Measuring Small Displacements by Using Capacitive Transducer System* SCIENTIA IRANICA, Vol. 7, No. 1

CARACTERIZACIÓN DE UN ACELERÓMETRO MICROMAQUINADO CON TECNOLOGÍA POLYMUMP'S

Rincón Jara Rodolfo Iván, Míreles José, Roberto Ambrosio.

Centro de Investigación y Tecnología Aplicada (CICTA).
Universidad Autónoma de Ciudad Juárez.
Tel. (656)688 48 00, correo electrónico: ivancatedral@yahoo.com.mx

RESUMEN

Una caracterización experimental, consiste en la integración de los elementos necesarios para obtener los parámetros y resultados de operación de un dispositivo. El acelerómetro MEMS diseñado en el CICTA, tomó como referencia los parámetros de operación de un acelerómetro comercial para su caracterización con el fin de buscar la respuesta dinámica y eléctrica de ambos acelerómetros. La obtención de los parámetros de operación, se realizó con la integración de un sistema de prueba formado por un instrumento vibratorio llamado Shaker, un lector de capacitancia universal y además de aparatos de medición como el osciloscopio, multímetro y fuentes de alimentación. La sensibilidad obtenida es de 1.4 V/g y el rango dinámico es de (3.4-12.4 G's).

Palabras clave: MEMS (Sistemas Microelectromecánicos); acelerómetro, masa sísmica, frecuencia de resonancia.

I. INTRODUCCIÓN

Los acelerómetros MEMS, se encuentran a escalas de décimas de micrones con niveles de sensibilidad y error muy pequeños, la unidad de medida son la gravedad (g's). Estos dispositivos son utilizados para sensar los cambios de velocidad con respecto al tiempo y para el control de mecanismos de vibración. Los acelerómetros deben ser analizados y caracterizados, con el fin de poder desarrollar interfaces que ayuden a implementarlos en futuras aplicaciones confiables en su operación y funcionalidad.

II. MÉTODO

Diseño y modelado.

El acelerómetro que se diseñó en el CICTA, es un acelerómetro del tipo capacitivo, el cuál cuenta con un conjunto de peines fijos y otro con peines móviles que tienen las mismas dimensiones de 3.5 μ m de grosor, 2 μ m de ancho y 30 μ m de longitud. El conjunto de peines fijos se encuentra anclado a una sección tipo barra de *Poly1*, *Poly2* y *Metal* de tal manera que conforman un mismo conductor. El conjunto de peines móviles, es aquél que se encuentra unido a los extremos laterales de la placa llamada masa sísmica, compuesta de la misma manera por *Poly1*, *Poly2* y *Metal*. La utilización de estos 3 materiales en el diseño de la masa sísmica y de los peines, ayudará a producir un mejor desplazamiento de la masa sísmica al momento en que el acelerómetro se encuentra bajo una aceleración y además ayudará a tener una frecuencia de resonancia más baja. La descripción de los elementos del acelerómetro se muestra en la figura 1.

El diseño de este acelerómetro utiliza como medio de transducción, un sistema de sensado del tipo capacitivo, cuya función es convertir eventos mecánicos en señales eléctricas y con éstas poder analizar la respuesta del acelerómetro bajo una aceleración de entrada. Existen tres tipos de esquemas para interfaces capacitivas, y para fines de este acelerómetro el esquema de interface que se utilizó se muestra en la figura 2.

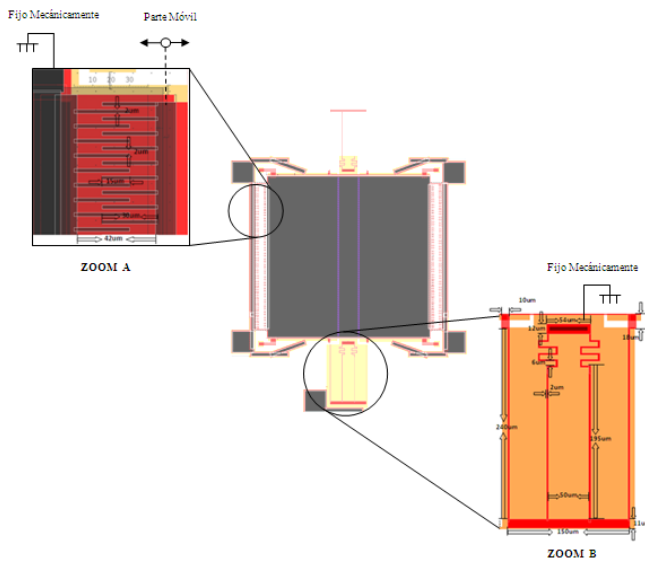


Figura 1. Diseño completo del acelerómetro y detalle de sus peines y resortes.

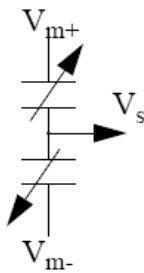


Figura 2. Esquema de interface capacitiva del acelerómetro. [5]

La descripción de cada uno de los elementos que conforman el acelerómetro se presenta en la figura 3. El sistema de suspensión, que se encuentra tanto en la parte superior como en la inferior, tiene como objetivo amortiguar el movimiento de la masa sísmica, cuando es sometido a una fuerza y esta anclado a la placa de *poly0*. Los otros elementos son los peines fijos y los peines móviles, que tienen como objetivo sensor las variaciones de capacitancia, así como los soportes móviles que ayudaran a tener una mejor flexibilidad en el movimiento.

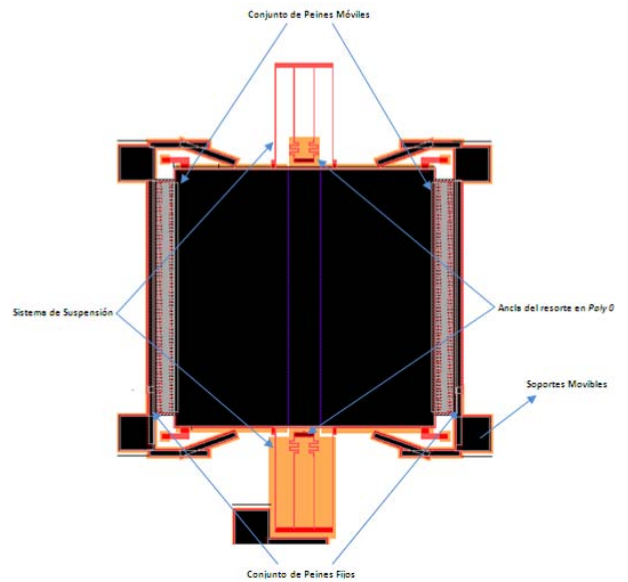


Figura 3. Descripción del diseño final del acelerómetro.

El acelerómetro MEMS, cuenta con un número de condiciones iniciales, cuando éste se encuentra en reposo, debido principalmente a las capacitancias existentes entre las diferentes placas del dispositivo. Cabe mencionar que todas las variables fueron calculadas, tomando en cuenta la distribución de las placas en el dispositivo y las dimensiones que se utilizaron para realizar su diseño. Las condiciones más significativas, son la masa sísmica y la capacitancia inicial del dispositivo.

Masa sísmica: Está compuesta por la unión de varias placas, las cuales son *poly1*, *poly2* y *metal* con un grosor de $2\ \mu\text{m}$, $1.5\ \mu\text{m}$ y $0.5\ \mu\text{m}$ respectivamente. La masa sísmica es muy importante, debido a que cuando el acelerómetro es sometido a una fuerza, esta masa y los peines móviles unidos a ella se desplazan produciendo un cambio de capacitancia entre los peines fijos y los móviles, así mismo entre el traslape que existe con la placa *poly0*.

Masa Específica [3]

$$m = \left[(a \cdot l \cdot g_m)_{placa} + 2 \cdot n_{peines} (a \cdot l \cdot g_p)_{peine} \right] \rho \quad (1)$$

La masa de la placa móvil, es la siguiente:

Masa Especifica

$$m = \left[(656\mu\text{m} \cdot 655\mu\text{m} \cdot 4\mu\text{m})_{\text{placa}} + 2 \cdot 75_{\text{peines}} (2\mu\text{m} \cdot 30\mu\text{m} \cdot 3.5\mu\text{m})_{\text{peine}} \right] (2330\text{Kg} / \text{m}) = 4.07\mu\text{g}$$

Capacitancias iniciales C_1 y C_2 . Una vez que ya se obtuvieron las capacitancias de traslape entre las placas y entre los peines móviles y fijos del acelerómetro, se determinaron las capacitancias iniciales producidas en cada uno de los segmentos. La suma de todas las capacitancias forman C_1 y C_2 .

La importancia de conocer estas capacitancias, es acondicionar una señal que sea capaz de sensar tales valores de capacitancia, que por lo general son valores tan pequeños como los fF .

Capacitancia inicial segmento 1:

$$C_1 = C_{TS1} + C_{TP1} + C_{pp} = 820\text{fF} + 69\text{fF} + 35\text{fF} = 924\text{fF}$$

Capacitancia inicial segmento 2:

$$C_2 = C_{TS2} + C_{TP2} + C_{pp} = 826\text{fF} + 69\text{fF} + 35\text{fF} = 930\text{fF}$$

Condiciones Dinámicas. Cuando el acelerómetro MEMS es sometido a una aceleración, se presentan diferentes parámetros que son importantes para su funcionamiento, como la frecuencia angular, frecuencia de resonancia, desplazamientos, etc. Tales parámetros causan cambios en las capacitancias C_1 y C_2 , debido principalmente al desplazamiento que sufre la masa sísmica del acelerómetro con el área de traslape de las placas.

Algunos de los parámetros dinámicos, se obtuvieron simulando el acelerómetro en el software Coventor. En la simulación, se aplicó una serie de fuerzas con el fin de analizar los desplazamientos producidos por ésta. En la figura 4 se muestra una de las simulaciones y en la tabla 1 se dan los resultados del desplazamiento.

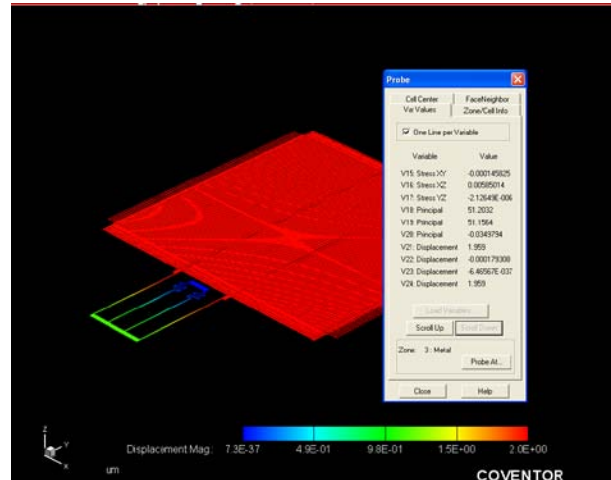
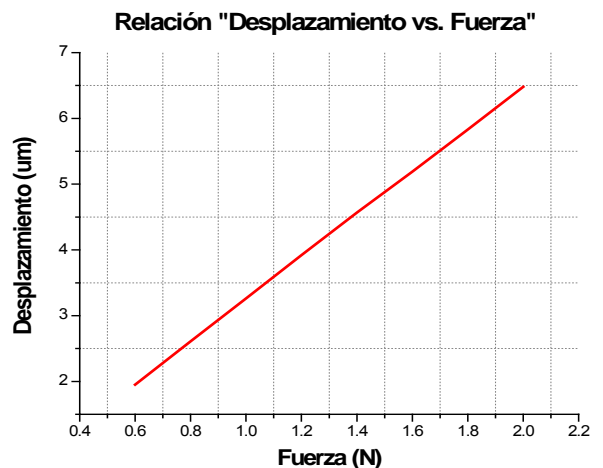


Figura 4. Simulación del acelerómetro sometido a una fuerza de $0.6\mu\text{N}$ y tabla de resultados.

Tabla 1. Resultados de los desplazamientos producidos por una fuerza aplicada.

Fuerza μN	Desplazamiento μm	Aceleración requerida m / s^2
0.6	1.95	147.42
0.8	2.61	196.56
1	3.26	245.70
1.2	3.92	294.84
1.4	4.57	343.98
1.6	5.19	393.12
1.8	5.83	442.26
2	6.48	491.40



Gráfica 1. Relación de desplazamiento por una fuerza aplicada.

Los parámetros obtenidos en el modelado del acelerómetro se muestran en la tabla 2, en ésta se dan los diferentes rangos de la frecuencia, fuerza, aceleración, desplazamiento y la variación de capacitancia.

Tabla 2. Parámetros modelados del acelerómetro.

Parámetro	Rango Característico
Frecuencia f (Hz)	(8.21-152.80)
Frecuencia angular ω (Hz)	(51.61-960)
Fuerza (μ N)	(0.0112-4.04)
Aceleración (m/s^2)	(2.76-995.02)
Desplazamiento (μ m)	(0.0376-13.2)
Variación de Capacitancia(ff)	(0.3-107.77)

III. RESULTADOS EXPERIMENTALES

Con el modelado del acelerómetro MEMS, se realizaron las diferentes caracterizaciones experimentales, para obtener los valores de las capacitancias reales de cada uno de los segmentos del acelerómetro. Éstas se obtuvieron con la utilización de un lector de capacitancias MS3110 [12]. Un esquemático de la medición de la capacitancia se muestra en la figura 5.

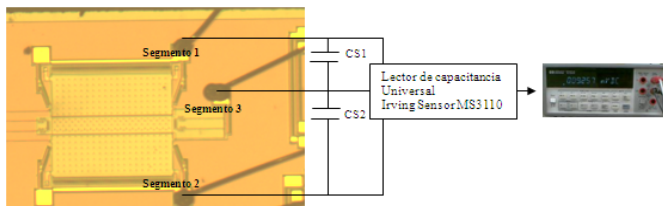


Figura 5. Medición de capacitancia, en cada segmento del acelerómetro utilizando el MS3110.

Con las mediciones realizadas y utilizando las formulas 2 y 3, se obtuvieron los diferentes valores de capacitancia para cada uno de los segmentos del acelerómetro. En la figura 6 se muestran las capacitancias obtenidas.

$$CS2 = \frac{(V_0 - V_{REF}) \cdot C_F}{Gain \cdot (V2P25)(1.14)} + 10pF \quad (2)$$

$$CS1 = 10pF - \frac{(V_0 - V_{REF}) \cdot C_F}{Gain \cdot (V2P25)(1.14)} \quad (3)$$

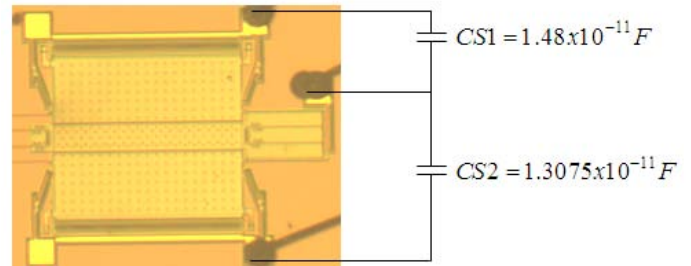


Figura 6. Valores de las capacitancias iniciales CS1 y CS2 del acelerómetro MEMS.

Las frecuencias del acelerómetro MEMS, fueron obtenidas utilizando un equipo PolyTech, realizando un barrido de frecuencias y polarizando cada uno de los segmentos del acelerómetro. El dispositivo presentó movimiento a tres diferentes frecuencias (359, 718, 12318Hz), los resultados se muestran en las siguientes figuras.

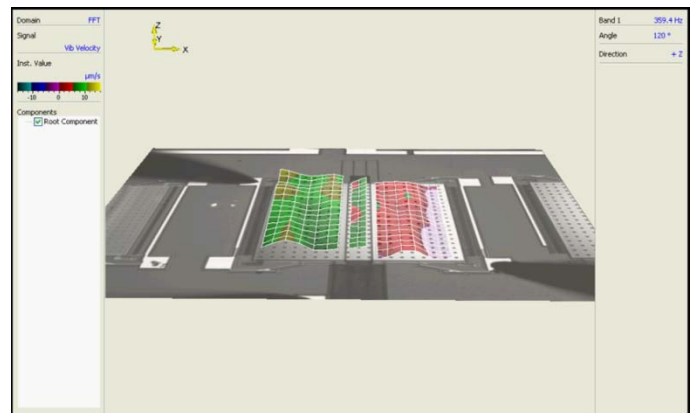
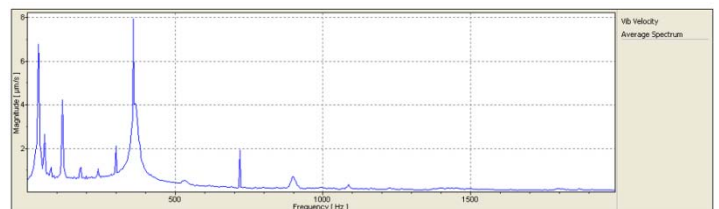


Figura 7. Resonancia a 359Hz.



Gráfica 2. Frecuencia vs velocidad a 359Hz.

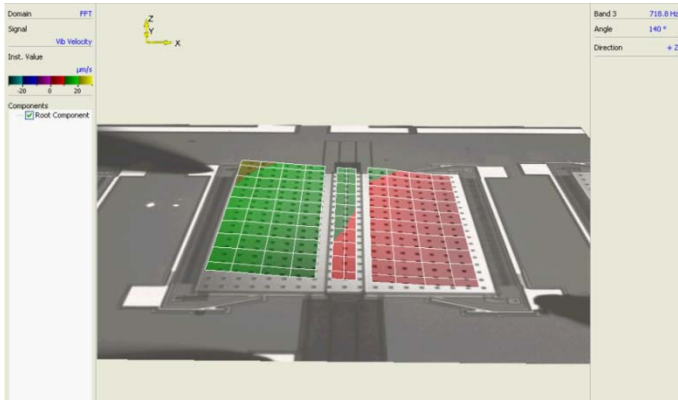


Figura 8. Resonancia a 718Hz

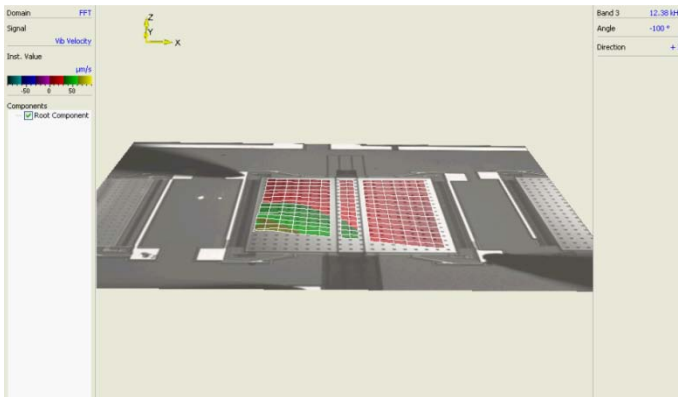
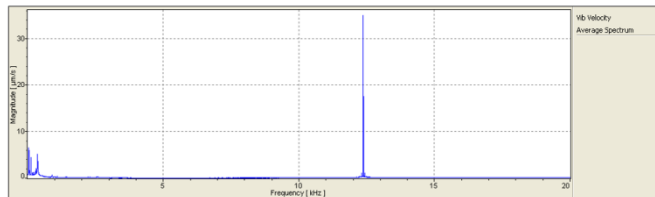


Figura 9. Resonancia 12.38KHz



Gráfica 3. Frecuencia vs velocidad a 12.38KHz.

Para el paso más importante de la caracterización experimental, el acelerómetro fue puesto bajo movimiento debido a la vibración de un shaker, con el fin de obtener un desplazamiento de su masa sísmica, la cual causaría una variación de la capacitancia inicial en cada uno de los segmentos del mismo. Esta variación se debe a que la capacitancia de traslape existente entre los peines fijos y móviles cambia por un desplazamiento en el conjunto de peines móviles. En cada uno de los segmentos la variación se hace de manera diferencial.

Algunas de las respuestas a la frecuencia, que el acelerómetro presentó bajo movimiento vibratorio se presentan en las siguientes figuras. En éstas se puede observar la variación de voltaje a través de una onda senoidal y su correspondiente transformada de Fourier.

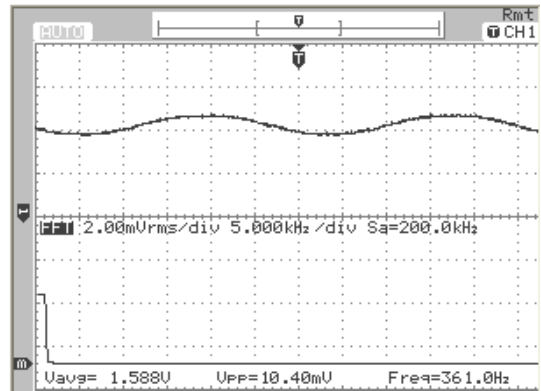


Figura 10. Respuesta del acelerómetro a 359Hz.

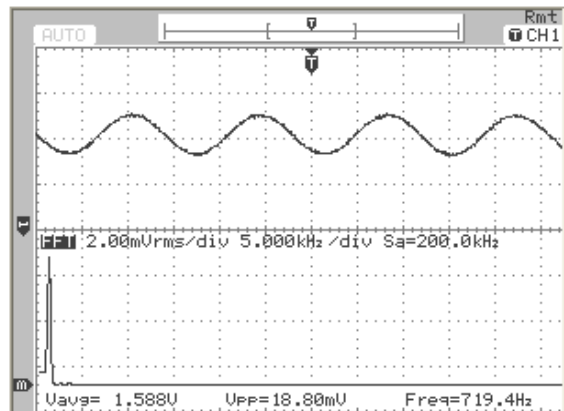


Figura 11. Respuesta del acelerómetro a 718Hz.

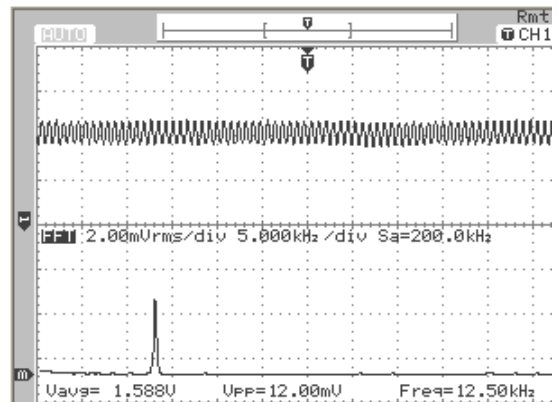


Figura 12. Respuesta del acelerómetro a 12.38KHz

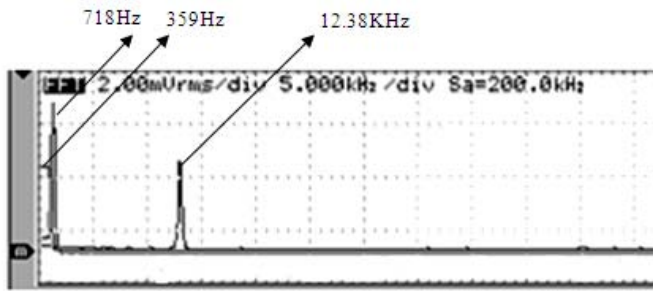
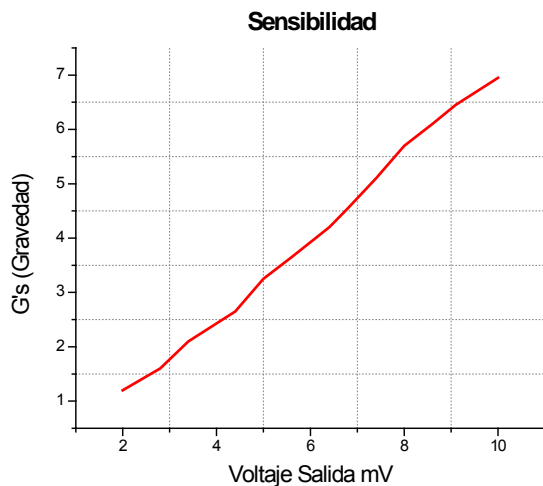


Figura 13. Comportamiento de las 3 frecuencias en su FFT.

IV. CONCLUSIONES

El acelerómetro MEMS, presentó parámetros satisfactorios a la aplicación de aceleraciones en G's por medio de un shaker. La respuesta a la frecuencia nos determinó, que el acelerómetro tenía una buena sensibilidad a bajas frecuencias, obteniendo un ancho de banda de 359Hz-4000KHz. La sensibilidad a la que el acelerómetro MEMS reaccionó bajo el sistema de vibración es de 1.4mV/g.



Gráfica 4. Sensibilidad del acelerómetro.

V. REFERENCIAS

[1] Julian W. Gardner, Vigía K. Varadan, “*Microsensors MEMS and smart devices*”, Ed. Wiley, Pennsylvania USA, pp.
 [2] [Http://www.memsnet.org/mems/what-is.html](http://www.memsnet.org/mems/what-is.html)

[3] Chang Liu, “*Foundations of MEMS*”, Illinois Ece Series (Prentice Hall), New Jersey USA, pp. 530.
 [4] Vijay K. Varadan, K. J. Vinoy, K. A. Jose, “*RF MEMS and their applications*”, Ed. Wiley, Pennsylvania USA, pp. 394
 [5] Stephen Bebi, Graham Ensell, “*MEMS mechanical Sensors*”, Ed. Artech house, Norwood USA, pp. 269
 [6] Jim Carter, Allen Cowell, “*PolyMUMPs Design Handbook*”, MEMS CAP, pp. 40
 [7] M. Elwenspoek, R. Wiegerink, “*Mechanical Microsensors*”, Ed. Springer, Germany, pp. 294
 [8] Mohamed Gad el Hak, “*Mems applications*”, Ed. Taylor and Francis, Florida USA, pp 547
 [9] Stephen D. Senturia, “*Microsystems Design*”, Ed. Springer, New York USA, pp 689
 [10] Mohamed Gad el Hak, “*Mems: Introduction and fundamentals*”, Ed. Taylor and Francis, Florida USA, pp 547
 [11] Gabriel R. Revéis, “*RF Mems*”, Ed. Wiley, New Jersey, pp 495
 [12] “*MS3110BDCP Operating Specifications and User Manual*”, Irving Sensor Corporation, Costa Mesa, California USA, 2004.
 [13] <http://www.analog.com/library/analogDialogue/archives/30-4/acccel.html>.
 [14] “*ADXLI50 Single Accelerometers datasheet*”, Analog Devices, USA.

VI. AUTORES

Dr. José Míreles Jr. por la Universidad Autónoma de Cd. Juárez.

Dr. Roberto Ambrosio por la Universidad Autónoma de Cd. Juárez.

Ing. Rodolfo Iván Rincón Jara es Ingeniero Eléctrico por la Universidad Autónoma de Ciudad Juárez.

DISEÑO DE MICRO Y MACRO ESPEJOS DE ACTUACIÓN ELECTROSTÁTICA

R. Campos-Rodríguez, M. Alcaraz-Mejía, S. Rosales-Sánchez, S. González-Santibáñez, J. Camacho-Espinoza

Departamento de Ciencias Tecnológicas,
Centro Universitario de la Ciénega, Universidad de Guadalajara
Av. Universidad 1115, Col. Lindavista, Ocotlán, Jalisco, 47820, México
Tel. +52(392)9259400, ext. 8369, correo electrónico: {raul.campos,mildreth.alcaraz}@cuci.udg.mx

RESUMEN

En este trabajo se presenta el diseño de micro y macro espejos de actuación electrostática. Las dimensiones del micro-espejo se encuentran en el rango de los micrómetros, mientras que las del macro-espejo se encuentran en el rango de los milímetros. Existe un factor de escalamiento de 1:250 entre las dimensiones de los espejos. Se muestra su diseño electro-mecánico y se explica su principio de funcionamiento. Se proporcionan modelos matemáticos y simulaciones numéricas que predicen el comportamiento de los espejos. Se discuten las implicaciones en el mecanismo de actuación como resultado del escalamiento en las dimensiones. Finalmente, se presentan potenciales aplicaciones de estos sistemas.

Palabras clave: Microsistemas, Espejos, Fuerza Electrostática, MEMS.

I. INTRODUCCIÓN

El término *Microsistema* ó *MEMS* (MicroElectroMEchanical Systems por sus siglas en inglés) describe a un conjunto de dispositivos microscópicos capaces de realizar actividades útiles en entornos específicos [1]. El término *Tecnología MEMS* se utiliza para describir el conjunto de técnicas, métodos y procesos para la fabricación de estos dispositivos [2], [3]. Así, la Tecnología MEMS o Tecnología de los Microsistemas se refiere a la ciencia o arte de miniaturizar los sistemas, en varios órdenes de magnitud, a fin de mantener al mínimo variables de operación como el volumen, peso y consumo de energía de los dispositivos. La minimización de

estas variables de operación no es la única razón para miniaturizar los sistemas. De hecho, la conceptualización de sistemas de mínimas dimensiones abre la posibilidad a nuevas aplicaciones que resultan imposibles mediante técnicas de fabricación tradicionales [4], [5].

A fin de que la creación de microsistemas sea posible, es necesario utilizar técnicas novedosas para la construcción de micro-estructuras, la mayoría de ellas adoptadas de la fabricación de Circuitos Integrados [7]. Por ejemplo, los procesos de micro-maquinado de superficie basados en fotolitografía permiten con toda facilidad la creación de engranes de involuta con un diámetro base de 80 micrómetros [6]. Esto es imposible de realizar mediante técnicas de fabricación basadas en tornos convencionales.

Este trabajo es una extensión del trabajo presentado por el autor principal en [11]. Se presenta el diseño micro y macro espejos de actuación electrostática con dimensiones en los micrómetros y milímetros, respectivamente. Existe un factor de escalamiento de 1:125 entre las dimensiones de ambos espejos. La finalidad de realizar dos prototipos con este factor de escalamiento es estudiar el efecto que produce el cambio de dimensiones en la fuerza electrostática que se genera entre las placas paralelas, los electrodos de actuación y de torsión.

El resto de este trabajo se organiza de la siguiente manera. En la Sección II., se presentan los fundamentos de la fuerza electrostática y sus principales ecuaciones. En la Sección III., se muestra el diseño mecánico de los espejos y en la Sección IV., se desarrolla la estimación de

parámetros para los modelos. En la Sección V., se presentan algunos resultados experimentales y finalmente en la Sección VI., se proporcionan las conclusiones de este trabajo.

II. FUERZA ELECTROSTÁTICA

La fuerza electrostática de atracción se presenta cuando cargas eléctricas de signos opuestos, se colocan una cerca de otra, ya sea en el espacio libre o ante la presencia de algún dieléctrico [8]. Tal fuerza se describe mediante la Ley de Coulomb, cuya versión escalar tiene la forma:

$$F_e = k_e \frac{q_1 q_2}{r^2} \quad (1)$$

donde q_1 y q_2 representan las cargas, r la distancia entre ellas y k_e es una constante de proporcionalidad conocida como la *Constante de Coulomb*. La constante k_e está relacionada con las propiedades eléctricas del espacio libre. Para fines prácticos, se considera que $k_e = 8.987 \text{ Nm}^2 \text{ C}^{-2}$.

En situaciones reales, la fuerza electrostática es un vector y no un escalar como en (1). Sin embargo, el estudio exhaustivo de este fenómeno queda fuera del alcance de este trabajo. Un lector interesado puede referirse a [8].

La fuerza electrostática representada por (1), se puede aprovechar para realizar movimiento por deformación mecánica mediante diversas configuraciones. La más común consiste en un par de electrodos alineados geoméricamente, como se muestra en la Figura 1. Al electrodo superior se le conoce como *Electrodo de Actuación*, mientras que al inferior se le denomina *Electrodo de Polarización*. En esta configuración, el Electrodo de Polarización se encuentra anclado a una

superficie de trabajo, mientras que el Electrodo de Actuación esta libre en uno de sus extremos. Al polarizar los electrodos con cargas de signos opuestos, se forman líneas de campo eléctrico, creando una fuerza de atracción entre ellos, la cual trata de juntarlos. Al desaparecer las cargas, desaparecen también las líneas de campo eléctrico y por consecuencia la fuerza de atracción, con lo que el electrodo de actuación regresa a su posición original. Esto siempre que no exceda los límites de su deformación elástica.

El área efectiva de polarización entre los electrodos en la Figura 1 forma un capacitor de placas paralelas con núcleo de aire, cuya capacitancia se puede aproximar como:

$$C = \frac{\epsilon_0 A}{d} \quad (2)$$

A partir de (1) y (2), y tomando la derivada de la capacitancia respecto a la separación entre placas, se encuentra que la fuerza que se puede generar con una configuración de electrodos como los de la Figura 1 es [8]:

$$F_e = -\frac{1}{2} \frac{\epsilon_0 A V^2}{d^2} \quad (3)$$

Al igual que en (1), en (2) solo se representa la versión escalar de la fuerza electrostática en un capacitor de placas paralelas con núcleo de aire. En la realidad, como se puede deducir de la Figura 1, esta fuerza se presenta en tres dimensiones. Más aun, debido al doblamiento que sufre la placa superior por la fuerza electrostática, la capacitancia, y por consecuencia la carga eléctrica, no es uniforme. De hecho, es de esperarse que la distancia entre placas sea menor en la punta del electrodo de actuación que en la base, por lo que la fuerza que experimente esta placa debe ser también mayor en dicha región. Más complicado aún resulta el hecho de que la condición inicial de la separación entre placas cambia respecto al tiempo de aplicación de la carga, retroalimentando a cada instante a la fuerza electrostática según se puede deducir en (3). Todo esto complica encontrar una ecuación analítica que describa razonablemente bien este comportamiento por lo que generalmente se utilizan técnicas numéricas, como simulación de elemento finito. En la Sección IV., se explicará las simulaciones numéricas utilizadas en este trabajo para la estimación de parámetros.

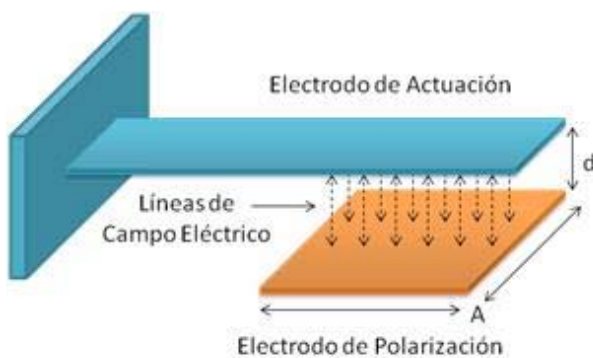


Figura 1. Fuerza Electrostática

III. DISEÑO MECÁNICO

Una estructura tipo trampolín, anclada en ambos extremos, como se muestra la Figura 2, permite aprovechar la fuerza electrostática para generar movimiento. El arreglo de las placas en la forma mostrada permite generar una torsión del trampolín. La barra central sirve de soporte a la estructura, al mantener suspendida a una gran placa sobre un par de electrodos de actuación. Los electrodos están colocados de manera simétrica como se ilustra. Esta gran placa es también el electrodo de actuación, ya que forma un par de capacitores de placas paralelas con los electrodos. Al polarizar la placa superior con signo positivo y alguno de los electrodos en la parte inferior con signo negativo, se forman líneas de campo eléctrico que trata de juntar ambas placas. Esto produce una torsión de la placa superior hacia el electrodo de polarización cargado con signo negativo. Si se elimina el campo eléctrico, la fuerza desaparece y la estructura vuelve a su posición original por la acción de la barra central. Si se alterna la polarización en los electrodos inferiores, entonces se puede lograr un movimiento de vaivén de la estructura. Este movimiento oscilatorio puede utilizarse para desviar una luz láser, y de esta manera realizar un barrido de una superficie de interés.

FABRICACIÓN DEL MICRO-ESPEJO

En la fabricación del micro-espejo se siguió la distribución mecánica mostrada en la Figura 2. Las dimensiones de este prototipo están en el rango de los micrómetros ($1\mu m = 1E^{-6}m$). Para su construcción se empleó un proceso comercial de micro-maquinado superficial de polisilicio

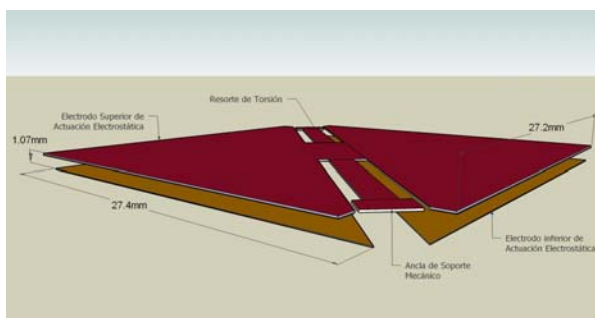


Figura 2. Diseño Mecánico Actuador

conocido como PolyMUMPs [9]. El proceso ofrece tres capas de silicio policristalino superficial las cuales pueden ser grabadas, mediante técnicas de fotolitografía, con patrones geométricos para formar estructuras microscópicas con capacidad de movimiento. En la Figura 3 se muestra un corte transversal del proceso PolyMUMPs. La fabricación inicia sobre un sustrato tipo-n. Sobre este sustrato se deposita una película de nitruro, una capa que sirve de material aislante tanto mecánico, como térmico y eléctrico. Sobre el nitruro se deposita la primera capa de polisilicio conocida como Poly0, con un espesor de 0.5um. Esta capa se encuentra anclada al nitruro, y generalmente se utiliza para diseñar conexiones eléctricas y electrodos de polarización. Sobre Poly0 se deposita una capa de oxido de sacrificio de 2.0um, conocida como Oxide1. Mediante un proceso similar, se deposita Poly1 y Oxide2, cuyos espesores son 2.0um y 0.75um, respectivamente. Sobre Oxide2 se deposita una capa de polisilicio de 1.5um de espesor conocida como Poly2. De esta manera, las capas Poly0 y Poly1 se encuentran separadas 2.0um, por Oxide1, mientras que Poly1 y Poly2 solo 750nm por Oxide2. Finalmente, se deposita una capa de oro de 0.5um de espesor conocida como Metal. Debido a sus propiedades eléctricas y ópticas, esta capa se utiliza para construir conexiones eléctricas y como superficie reflejante de la luz. Al final del proceso, las capas de sacrificio Oxide1 y Oxide2 son removidas mediante la acción de un ácido para liberar mecánicamente a las capas Poly1 y Poly2. Los patrones geométricos empleados en la construcción de un micro-sistema, como vigas voladas, placas paralelas, engranajes, etc., se trazan

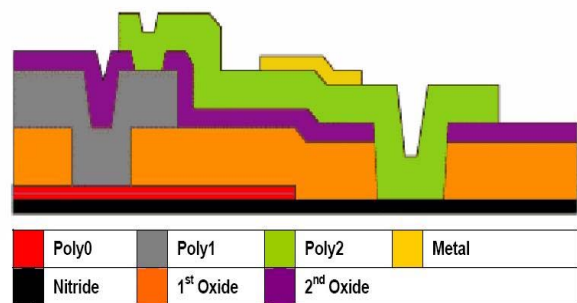


Figura 3. Proceso de Fabricación PolyMUMPs

utilizando una herramienta CAD muy similar a la empleada en el diseño de semiconductores. El resultado, es un conjunto de celdas en formato CIF que son utilizadas para generar las máscaras de litografía. Para el diseño del micro-espejo se utilizó el software L-Edit, de Tanner Inc. [10]. El diseño final del micro-espejo se muestra en la Figura 4, el cual incluye áreas de contacto eléctrico a los lados del diseño, para realizar las pruebas. Se utilizaron las capas Poly0 y Poly1 para crear los electrodos de polarización y la placa suspendida de actuación, respectivamente. Finalmente, se utilizó la capa Metal, oro de 750nm de espesor, sobre una capa de Poly2 para formar el rectángulo central que hace las veces de superficie reflejante para la luz. Además, el diseño incluye otro mecanismo de actuación basado en expansión térmica, el cual está realizado en las capas Poly1 y Poly2, unidas mediante un ancla rectangular grabada en la capa Poly1-Poly2-Via. Este mecanismo de actuación está dispuesto en forma ortogonal respecto al primero, lo que permite dos grados de libertad en el espejo. Sin embargo, dicho mecanismo queda fuera del alcance de este trabajo.

El electrodo de polarización en Poly0 es un triángulo isósceles con lados 80 μ m y 113 μ m, respectivamente. El electrodo de actuación en Poly1 está formado por dos triángulos isósceles contrapuestos, con esquinas truncadas y lados de 100 μ m y 140 μ m respectivamente. Los triángulos se encuentran unidos por su parte central mediante

un rectángulo de 30 μ m y 15 μ m por lado. De la parte central de este rectángulo se extiende una viga que sirve de soporte para toda la estructura y que hace las veces de resorte de torsión. A los extremos de la placa en Poly1 se agregaron topes mecánicos, mediante la capa Dimple, para evitar que la placa superior colapse sobre el electrodo de polarización en Poly0 por la fuerza electrostática.

FABRICACIÓN DEL MACRO-ESPEJO

El macro-espejo fue fabricado siguiendo la misma geometría mostrada en la Figura 2 pero con dimensiones en el rango de los milímetros. El factor de escalamiento entre los espejos es de 1:250. De esta manera, el espesor de la placa en Poly1 de 2 μ m en el micro-espejo corresponde a una placa de espesor de 0.5mm en el macro-espejo. Este es un grosor típico de la hoja de aluminio que puede encontrarse con facilidad, por esta razón se decidió establecer dicho factor de escalamiento. Las dimensiones del triángulo isósceles que forman el electrodo de polarización del macro-espejo son 27.4mm por lado. Se colocaron dos de ellos de manera simétrica sobre una superficie plástica. Las dimensiones del electrodo de actuación son de 27.2mm por lado en cada uno de los dos triángulos truncados que lo conforman. La separación entre el electrodo de actuación y el de polarización es de 1.07mm. Dichas dimensiones se muestran en la Figura 2, mientras que una imagen del prototipo fabricado se presenta en la Figura 5. El material empleado en su construcción, tanto de

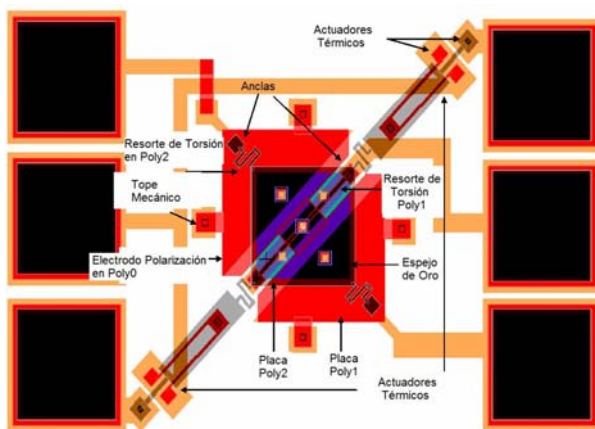


Figura 4. Diseño de Máscaras de Fabricación

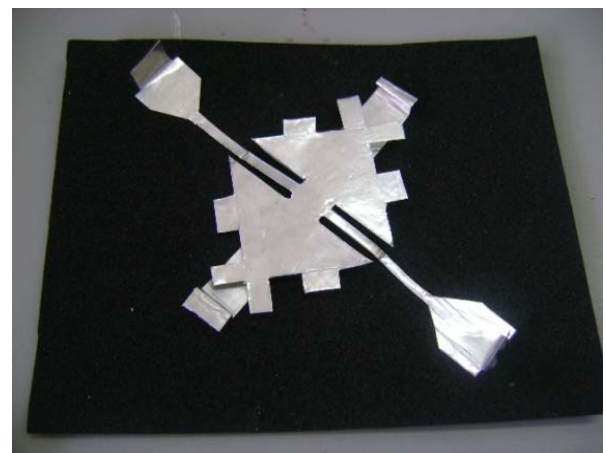


Figura 5. Diseño del Macro-Espejo

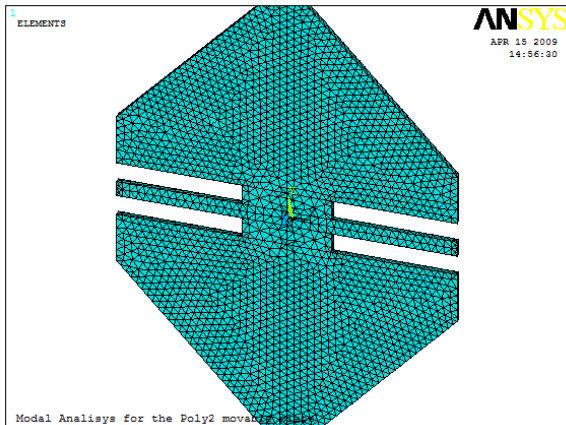


Figura 6. Detalle del mallado Análisis Modal

los electrodos de polarización como del electrodo de actuación, es hoja de aluminio de 0.5mm de espesor. Se emplearon tijeras normales para cortar las piezas, las cuales fueron colocadas sobre una superficie plástica con una resistividad eléctrica de aproximadamente $1E^{13}\Omega m$, lo que proporciona un excelente aislamiento eléctrico. Este aislamiento eléctrico se requiere debido al alto voltaje empleado para polarizar las estructuras, que sobrepasa los 1500V.

IV. ESTIMACIÓN DE PARÁMETROS

Como se mencionó en la sección anterior, el diseño propuesto en este trabajo consta de un par de electrodos de polarización y una placa suspendida que forma el electrodo de actuación. Los electrodos de polarización se encuentran fijos a una estructura aislante, mientras que el electrodo de actuación se encuentra suspendido mediante una barra delgada que funciona como un resorte de torsión. Consecuentemente, los parámetros mecánicos y eléctricos de las estructuras resultan de especial interés para este trabajo. Además, la respuesta en frecuencia de la estructura es muy importante si se pretende utilizar el diseño propuesto como un escáner. Sin embargo, debido a la complejidad de este tipo de análisis, en este trabajo se estudian exclusivamente los parámetros mecánicos y eléctricos de las estructuras propuestas y se deja el análisis en frecuencia como trabajo futuro.

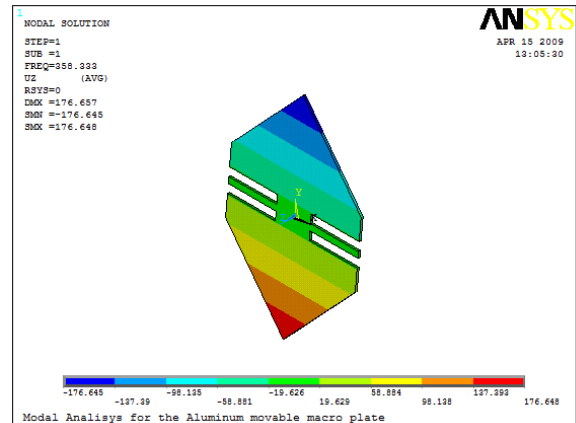


Figura 7. Primer Modo de Resonancia

El electrodo de actuación es la parte móvil en el diseño, ya que puede inclinarse hacia un lado u otro debido a la acción del campo eléctrico. Por lo tanto, el cálculo de la constante del resorte del sistema se centra exclusivamente en esta parte de la estructura.

CONSTANTE DEL RESORTE

La constante del resorte está relacionada con la resistencia al movimiento de un sistema mecánico. En el diseño propuesto, el resorte es una barra central que mantiene suspendido al electrodo de actuación. No existen formulas analíticas simples para el cálculo de la constante del resorte de una viga de torsión, sobre todo si se considera que dicha constante está distribuida en toda la estructura que mantiene a la placa suspendida. En este trabajo se utilizó la frecuencia natural del electrodo de actuación para encontrar su constante del resorte de manera indirecta. La técnica consistió en estimar la rigidez de la placa suspendida mediante un análisis modal. Se sabe que la frecuencia de resonancia de un sistema masa-resorte, sin considerar el factor de amortiguamiento, está determinada por la raíz cuadrada del cociente entre la constante del resorte y su masa, es decir:

$$f_i = \frac{1}{2\pi} \sqrt{\frac{k_i}{m}} \quad (4)$$

donde k_i es la rigidez de la estructura cuando oscila con una frecuencia f_i en el i -ésimo modo de

resonancia y m es la masa efectiva en movimiento de la estructura.

De esta manera, si en (4) se conoce f_i y m , la constante k_i se determina como:

$$k_i = \omega^2 m \quad (5)$$

donde $\omega = 2\pi f_i$ es la frecuencia natural, en radianes, de la estructura.

A fin de determinar k_i a partir de (5), es necesario determinar la masa de la estructura y la frecuencia natural en la dirección en la que se desea estimar la rigidez del sistema. Para encontrar la masa de las estructuras, se utilizó la relación volumen/densidad de los materiales empleados en su fabricación:

$$m = \rho V \quad (6)$$

donde ρ es la densidad del material en Kg/m^3 y V el volumen en m^3 .

Para el caso del electrodo de actuación en el micro-espejo, cuyo material de fabricación es polisilicio con una densidad de $\rho_1 = 2330 Kg/m^3$ y un volumen $V_1 = 3.1274822E^{-16}m^3$, de (6) se encuentra que la masa es $m_1 = 7.2870336E^{-11}Kg$. Por otro lado, para el macro-espejo, cuyo material de fabricación es aluminio con densidad de $\rho_2 = 2700 Kg/m^3$ y un volumen $V_2 = 1.5259259E^{-11}m^3$, la masa resulta ser $m_2 = 4.12E^{-5}Kg$.

A fin de encontrar la frecuencia de resonancia, se utilizó Simulación de Elemento Finito mediante la técnica de Análisis Modal. Se generaron dos modelos sólidos con una relación de escalamiento de 1:250. Las condiciones de frontera para este experimento consistieron en restringir el movimiento en los extremos del resorte de torsión en sus caras laterales. En las simulaciones para ambos modelos, se utilizó un mallado de tamaño 5 empleando un elemento finito sólido 3D con 20-nodos y 3 grados de libertad por nodo, que representan la traslación en los ejes X, Y, y Z. La Figura 6 muestra un detalle del mallado empleado. Se utilizó el método Block-Lanczos para calcular los primeros 5 modos de resonancia de la estructura. Este método construye el espacio nulo de la matriz de rigidez definida por los elementos finitos del mallado y expande los resultados a toda la estructura. Los resultados para ambos modelos se muestran en la Figura 7. El primer modo de resonancia corresponde al movimiento de torsión

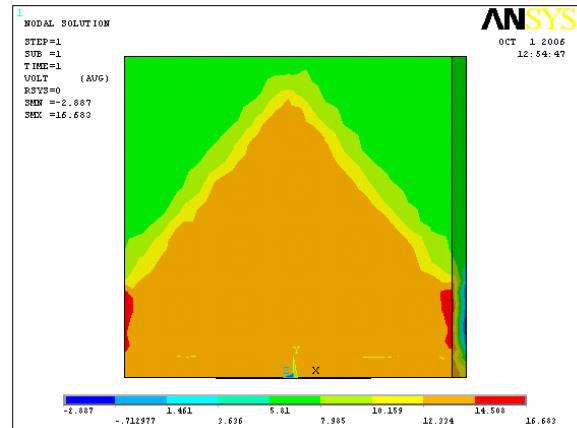


Figura 8. Distribución del Campo Eléctrico.

en la dirección en la que se desea estimar la rigidez de la estructura. Con las frecuencias y los valores de las masas, mediante (5), se encuentra que la constante del resorte para la micro-estructura es $k_1^1 = 0.9394$ y para la macro-estructura $k_1^2 = 0.292$. La forma en la que oscila la estructura en el primer modo de resonancia se muestra en la Figura 7. La parte superior se mueve hacia dentro de la figura mientras que la parte inferior hacia afuera. El resto de los modos de resonancia en la Tabla 1 pueden utilizarse para estimar la rigidez de las estructuras para diferentes direcciones de movimiento, con lo que se puede obtener una idea general del comportamiento dinámico de la estructura.

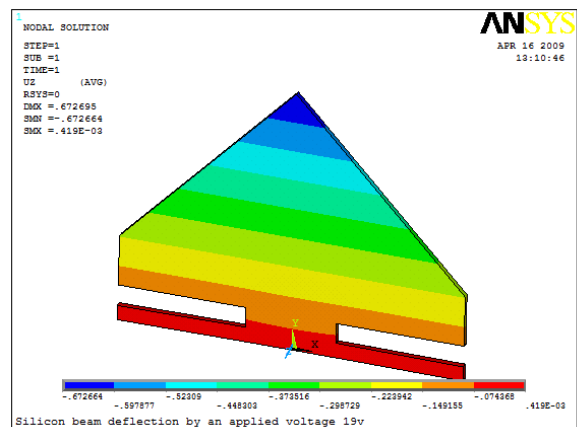


Figura 9. Deformación Fuerza Electroestática.

CAMPO ELÉCTRICO

Como se describe en (3), la fuerza electrostática generada por un par de placas paralelas está relacionada con la densidad de campo eléctrico en el capacitor de núcleo de aire formado por los electrodos de polarización y actuación. Esta densidad de campo eléctrico está determinada por la constante dieléctrica del aire, el área efectiva de polarización, el voltaje aplicado y la separación entre las placas. Sin embargo, como es de esperarse, la acción de la fuerza electrostática provocará que el electrodo de actuación se doble hacia el electrodo de polarización, con lo que aun con un voltaje constante, la fuerza electrostática tenderá a ser una fracción mayor a medida que pasa el tiempo, debido a que la separación entre placas disminuirá. Más aun, la distancia entre placas se volverá irregular, siendo menor en los extremos del electrodo de actuación y mayor en su centro. Por esta razón, se dice que el sistema esta retroalimentado, donde la variable que se retroalimenta es la separación entre placas. Este fenómeno transitorio ocurre durante unos milisegundos justo después de aplicar el voltaje de polarización. Si bien el análisis transitorio del fenómeno electrostático es necesario para una correcta comprensión sobre el comportamiento de la estructura, su estudio profundo se vuelve muy complicado, por lo que en general, se utilizan simuladores de estado estacionario.

Si se considera el efecto de amortiguamiento que produce el aire al ser comprimido debajo del electrodo de actuación, es necesario tomar en cuenta un factor proporcional a la primera derivada del desplazamiento, es decir la velocidad, a fin de aproximar la solución. Esto debido a que si el movimiento del electrodo de actuación es suficientemente lento, entonces el aire debajo de éste podrá escapar lentamente por los lados. Sin embargo, si el momento es muy rápido, el aire podría ser comprimido por unos instantes, lo que opondría mayor resistencia al movimiento del electrodo de actuación y provocaría movimientos erráticos de la estructura. Por esta razón, en esta etapa del trabajo, la investigación se centrará en estudiar el campo eléctrico en estado estacionario, dejando los efectos transitorios para futuras extensiones de este trabajo.

Tabla 1. Frecuencia de Resonancia

Modo	Macro-Modelo	Micro-Modelo
1	424.36Hz	180.71kHz
2	873.12Hz	357.12kHz
3	1327.20Hz	543.83kHz
4	1473.0Hz	600.85kHz
5	1886.20Hz	769.52kHz

A fin de estudiar los fenómenos eléctricos, se utilizó un modelo tridimensional para su simulación en computadora, mediante el método de elementos finitos. Debido a la simetría del fenómeno, se consideró la mitad del electrodo de actuación y uno de los electrodos de polarización dispuestos uno sobre otro rodeados por una masa de aire. Se aplicó un voltaje de 0V al electrodo de polarización y 16V al de actuación. El resultado de la distribución del campo eléctrico se muestra en la Figura 8. Como puede apreciarse, la mayor parte del campo eléctrico se encuentra concentrado entre los electrodos, lo cual resulta conveniente. También es posible apreciar que existen pequeñas fugas de campo eléctrico por los extremos de la estructura. Esto resulta inevitable en cualquier aplicación real.

Como resultado de la fuerza electrostática, el electrodo de actuación se doblará hacia el electrodo de polarización. Este es un fenómeno en el que ocurren dos dominios, el dominio eléctrico y el dominio mecánico. De hecho, como ya se ha mencionado, los cambios en un dominio afectan al otro y viceversa. Se dice entonces que los fenómenos se encuentran acoplados. En la técnica de elementos finitos, dichos fenómenos son

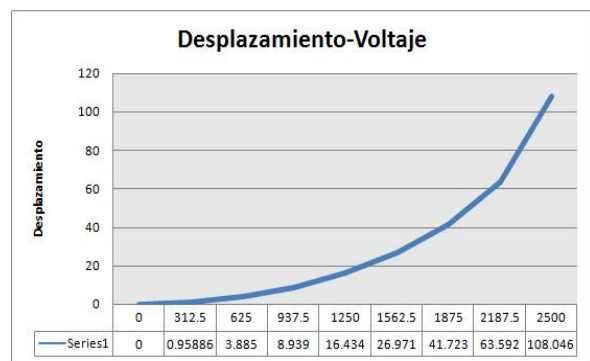


Figura 10. Curva Desplazamiento-Voltaje

resueltos mediante análisis multifísicos. En esta simulación se consideró la deformación mecánica en estado estacionario, que sufre el electrodo de actuación por la fuerza electrostática. Los resultados se muestran en la Figura 9. Como es de esperarse, la máxima deflexión de la estructura ocurre en la parte superior, debido al efecto de palanca presente en la estructura. Para este caso en particular, debido a la aplicación de una diferencia de potencial de 16V entre las placas, la máxima deflexión lograda es de -672nm. El signo negativo indica que la deflexión fue en el sentido negativo del eje Z, debido a la forma en que el modelo tridimensional fue capturado en la herramienta de simulación.

CURVA VOLTAJE-DESPLAZAMIENTO

A fin de predecir el comportamiento del electrodo de actuación del macro-espejo bajo la acción de diferentes voltajes, se realizaron 8 simulaciones igualmente espaciadas en el rango de voltaje de 312.5V a 2500V. El comportamiento en simulación del extremo superior del electrodo de actuación se muestra en la curva de la Figura 10. Como puede apreciarse, el comportamiento de la deformación mecánica por la fuerza electrostática describe una curva cuadrática, centrada en el origen y con asíntota cerca de los 2700V, donde se alcanza el voltaje de colapso de la estructura. Esto corresponde al punto de equilibrio inestable del sistema, ubicado a dos tercios de la separación original de los electrodos.

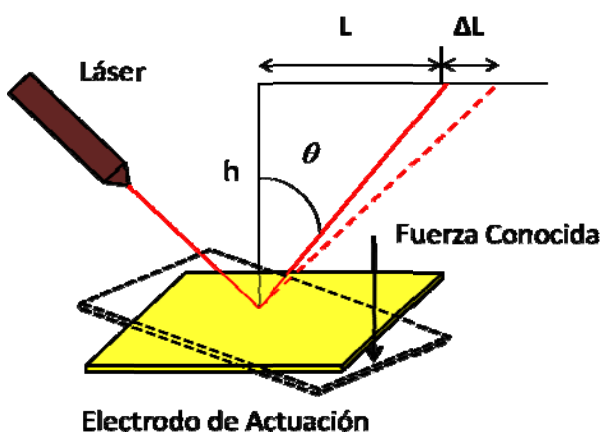


Figura 11. Verificación constante del resorte

V. EXPERIMENTACIÓN

Una vez fabricados los prototipos, se procedió a la realización de pruebas para verificar sus parámetros de funcionamiento. En esta etapa del trabajo, se realizó la experimentación necesaria para verificar la constante del resorte del sistema y la deformación máxima alcanzada por el electrodo de actuación bajo la acción de un voltaje conocido, todo en estado estacionario. Se reportan los resultados preliminares para el macro-espejo y se dejan los resultados del micro-espejo para una versión extendida de este trabajo.

VALIDACIÓN DE LA CONSTANTE DEL RESORTE

Para verificar la rigidez del sistema se realizó un experimento que consistió en aplicar una fuerza conocida a un extremo del electrodo de actuación a fin de medir el máximo desplazamiento experimentado por la estructura. Para esto, se utilizó un puente óptico NEWPORT, perfectamente alineado como soporte mecánico para el experimento. Sobre éste se colocaron las estructuras bajo estudio. Además, se empleó una luz láser montada mediante un sistema de posicionamiento sobre el mismo puente óptico e incidiendo sobre el electrodo de actuación. Como primer paso, se estableció un punto de referencia en ausencia de fuerza alguna sobre el electrodo de actuación. Después, se aplicaron varios pesos cuyo valor en gramos-fuerza se determinó con resolución de nano gramos. Enseguida, se midió el desplazamiento que sufrió la luz láser respecto del

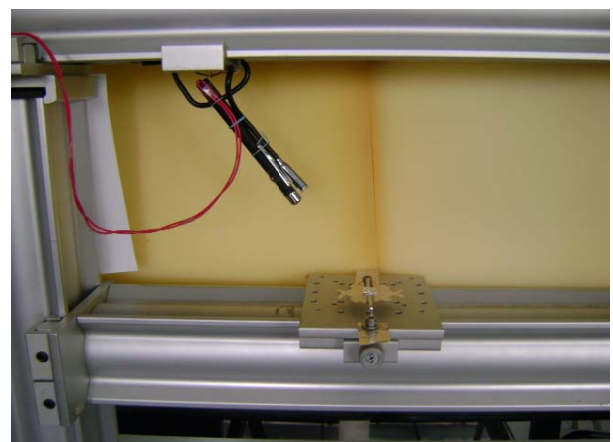


Figura 12. Puente óptico con láser

punto de referencia y mediante trigonometría se determinó el ángulo de torsión. El experimento se repitió con diferentes pesos en miligramos. Los pesos fueron determinados en nano gramos, con 4 dígitos decimales mediante una balanza digital METER AE200 perfectamente calibrada. Una vez medidos, los pesos fueron etiquetados para su fácil identificación.

En la Figura 11 se muestra un diagrama del experimento y en la Figura 12 se ilustra el puente óptico con la luz láser y el prototipo de experimentación. En la Figura 13 se proporciona un gráfico con los valores encontrados experimentalmente de la constante del resorte. El valor más alto encontrado en esta experimentación es de 0.160 y el menor 0.145. El valor promedio oscila entre 0.155 y 0.150. Las unidades de este gráfico son *N/grado* debido a que la constante calculada es de un resorte de torsión. Este valor contrasta con el obtenido en simulación, que es de 0.292. La diferencia podría deberse a defectos en la fabricación de la macro-estructura y/o errores en la realización de las medidas de la misma. Se planea ampliar la altura *h* en el esquema de experimentación de la Figura 11 a fin de tener mayor precisión en la realización de los experimentos. Debe notarse que al aumentar el valor de *h*, la distancia ΔL , que se genera por la aplicación de un peso conocido, se ve magnificada, lo que permite realizar las mediciones con mayor exactitud. En los experimentos reportados, se utilizó una distancia $h = 21,5\text{cm}$. Para futuros experimentos se planea utilizar una altura $h = 2.45\text{m}$.

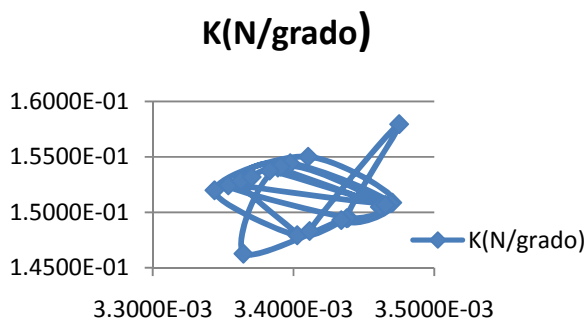


Figura 13. Constante del Resorte Experimental

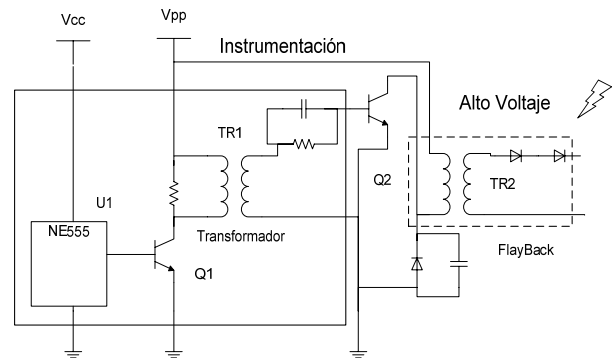


Figura 14. Generador de Alto Voltaje

VALIDACIÓN DE LA FUERZA ELECTROSTÁTICA

Al momento de la escritura de este trabajo, se realizan experimentos sobre el macro-espejo con el circuito de alto voltaje de la Figura 14. La instrumentación está formada por un oscilador que controla el primario del transformador TR1 de bajo voltaje. El secundario de TR1 a su vez controla el primario del flyback TR2. A la salida de TR2 el alto voltaje se rectifica, con lo que se obtiene una señal de CC. Las terminales de este circuito se conectan al electrodo de polarización y al electrodo de actuación. Hasta el momento, se ha logrado una torsión del electrodo de actuación menor a 5°, antes de que se produzcan arcos de voltaje entre las secciones puntiagudas del diseño. El voltaje aplicado fue de 1250V. Debido a esto, se han propuesto modificaciones al diseño de los actuadores para reducir las secciones punzantes, y evitar en lo posible la generación de arcos de voltaje en el macro-espejo. Sin embargo, se sabe que existe una barrera física para el diseño de macro-estructuras electrostáticas, que está determinada por la constante dieléctrica del aire.

VI. CONCLUSIONES

En este trabajo se ha presentado el diseño de espejos de actuación electrostática. Se han mostrado dos prototipos, uno cuyas dimensiones se encuentran en el rango de los micrómetros, fue fabricado mediante fotolitografía en un proceso comercial de micromaquinado superficial de polisilicio. El otro, cuyas dimensiones se encuentran en el rango de los milímetros, fue fabricado a mano utilizando hoja de aluminio de

0.5mm de espesor. Los prototipos guardan una escala de 1:250. La razón por la cual se realizaron dos prototipos con esta relación de tamaño es para investigar las implicaciones del escalamiento en las dimensiones con respecto a la fuerza electrostática y extrapolar los resultados del macro modelo al micro modelo.

VII. AGRADECIMIENTOS

Los autores agradecen a la Subsecretaría de Educación Superior e Investigación Científica por los apoyos de Fomento a la Generación y Aplicación Innovadora del Conocimiento y la Beca de Estudiante otorgados al proyecto “Diseño de MicroEspejos MEMS” mediante oficio PROMEP/103.5/08/2919.

VIII. REFERENCIAS

- [1] Marc J. Madou. Fundamentals of Microfabrication: The Science of Miniaturization. CRC Pub., 2nd ed., 2002. ISBN 0849308267.
- [2] Stephen D. Senturia. Microsystem Design, Springer, 2004. ISBN 0792372468.
- [3] Gregory T. Kovacs, Micromachined Transducers Sourcebook. McGraw-Hill, 1st edition, 1998. ISBN 0072907223.
- [4] Gabriel M. Rebeiz. RF MEMS: Theory, Design, and Tech. Wiley-Interscience; 1st edition, 2002. ISBN 0471201693.
- [5] John A. Pelesko, David H. Bernstein. Modeling MEMS and NEMS. CRC Pub., 1st edition, 2002. ISBN 1584883065.
- [6] Stephen A. Campbell. The Science and Engineering of Microelectronic Fabrication. Oxford University Press, USA; 2nd edition, 2001. ISBN 0195136055.
- [7] Richard C. Jaeger. Introduction to Microelectronic Fabrication. Prentice Hall; 2nd edition, 2001. ISBN 0201444941.
- [8] Matthew N. O. Sadiku. Elements of Electromagnetics. Oxford University Press, USA; 4th edition, 2006. ISBN 0195300483.
- [9] PolyMUMPs Fabrication Process, http://www.memscap.com/en_mumps.html
- [10] Tanner Layout Editor (L-Edit), Tanner Research Inc., <http://www.tanner.com/>.
- [11] R. Campos-Rodríguez, J. Mireles-García. Thermal and Electrostatic 2-DOF Micromirror, Vol. 1, pp 1-6, Proceeding of the TexMEMS-VIII International Workshop, The University of Texas at Dallas, Oct 9, 2006.

IX. AUTORES

Dr. Raúl Campos Rodríguez es Ingeniero en Computación, Centro Universitario de la Ciénega, UDG, 2000. Estudió la Maestría y Doctorado en Ciencias en Ingeniería Eléctrica en el CINVESTAV Guadalajara, 2002-2007. Trabaja activamente en el diseño de algoritmos para sistemas de eventos discretos y en el diseño de Microsistemas, o MEMS, específicamente en mecanismos basados en fuerza electrostática y expansión térmica.

Dra. Mildreth Alcaraz Mejía es Ingeniero en Sistemas Computacionales por el Instituto Tecnológico de Colima, 2001. Obtuvo los títulos de Maestro y Doctor en Ciencias con especialidad en Ingeniería Eléctrica en el CINVESTAV Unidad Guadalajara, 2004-2007, respectivamente. Actualmente, se dedica a la aplicación y desarrollo de técnicas de control embebidas con una orientación hacia el control de los Microsistemas.

Samuel Rosales Sánchez es estudiante de la carrera de licenciatura en informática, en el Centro Universitario de la Ciénega, de la UDG. Trabaja en este proyecto con una beca de PROMEP para estudiantes en proyectos de investigación.

Salvador Moisés González Santibáñez es estudiante de la carrera de ingeniería en computación en el Centro Universitario de la Ciénega, de la UDG. Trabaja en este proyecto con una beca de excelencia académica para motivación a la investigación.

Juan Ramón Camacho Espinoza es estudiante de la carrera de licenciatura en informática, en el Centro Universitario de la Ciénega, de la UDG. Trabaja en este proyecto como parte de su servicio social en actividades de investigación.

ARREGLO DE MICRO-ENGRANES PLANETARIOS

R. Campos-Rodríguez, M. Alcaraz-Mejía, M. Aguirre-Villa, A. Aceves-Paredes

Departamento de Ciencias Tecnológicas,
Centro Universitario de la Ciénega, Universidad de Guadalajara
Av. Universidad 1115, Col. Lindavista, Ocotlán, Jalisco, 47820, México
Tel. +52(392)9259400, ext. 8369, correo electrónico: {raul.campos,mildreth.alcaraz}@cuci.udg.mx

RESUMEN

Este trabajo presenta el diseño de un conjunto de engranes planetarios. Los dientes de los engranes individuales se diseñaron para tener un perfil de involuta a fin de minimizar la fricción producida por el contacto con los dientes de los demás engranes. El arreglo fue diseñado tomando en cuenta su fabricación mediante procesos de micro-maquinado superficial de polisilicio, cuya resolución de fabricación se encuentra en el rango de los micrómetros. Se presentan las condiciones de ensamble de los engranes, la relación del número de dientes entre los planetas, el sol y el aïillo exterior.

Palabras clave: Micro-Engranes, Engranes Planetarios, Microsistemas, MEMS.

I. INTRODUCCIÓN

Los engranes han sido un elemento clave en máquinas antiguas y en muchas otras que se emplean en la actualidad. Estos mecanismos permiten transmitir movimiento angular de manera infinita a otros elementos en un sistema de transmisión. Normalmente se encuentran incrustados en arreglos, cadenas, transmisiones y otros elementos. Los engranes tienen la habilidad de multiplicar el movimiento, dividiendo la fuerza. Inversamente, también pueden multiplicar la fuerza, dividiendo el movimiento. Idealmente, todos los engranes colocados en cualquier configuración conservan una relación constante, que es la velocidad angular. Sin embargo, en la práctica esta relación puede verse afectada por la fricción entre dientes, el desgaste, defectos en la alineación de los engranes y otros factores. A pesar

de esto, los arreglos de engranes de involuta sobrepasan con facilidad el 90% de eficiencia en la transmisión del movimiento bajo condiciones normales de operación. Otro factor importante de los engranes es que pueden diseñarse en múltiples formas y tamaños, como helicoidales, circulares, cónicos, de espuela y otros. Además, los engranes se pueden diseñar para tener dientes internos y externos y de diferentes tamaños. Todo esto ofrece enormes posibilidades en el diseño de sistemas de transmisión del movimiento basado en engranes.

Los engranes circulares son los más utilizados debido a su facilidad de fabricación y simplicidad en el ensamble. Generalmente, los engranes se encuentran acompañados de otros elementos, como cremalleras y cadenas, formando un sistema de transmisión completo. En todos estos arreglos y configuraciones, el perfil de involuta es uno de los preferidos para el diseño de los dientes de cada elemento que intervienen en el sistema de transmisión.

A pesar de la importancia de los engranes en las aplicaciones domésticas e industriales, el diseño de engranes miniaturizados se ha reportado poco en la literatura. En [5] se ofrece al público en general varias celdas de diseño a partir de las cuales los usuarios pueden definir instancias particulares mediante parámetros específicos. En [3] y [4], los autores proponen el diseño basado en parámetros de engranes con perfiles rectangulares. A pesar de que los algoritmos presentados permiten definir engranes de espuela a partir de un conjunto de parámetros, el perfil rectangular del diente reduce considerablemente la eficiencia de los engranes obtenidos. En [6], el autor dedica un capítulo al análisis de varios engranes ya fabricados, la

mayoría en SUMMIT-V de los Laboratorios Nacionales Sandia.

En este trabajo se presenta el diseño basado en parámetros de un conjunto de engranes planetarios. Se consideran aspectos del diseño de engranes de involuta individuales y las condiciones para su ensamble en torno al sol, así como aspectos generales de la generación de máscaras para su fabricación en procesos de micromaquinado superficial de polisilicio.

El resto de este documento se organiza de la siguiente manera. En la Sección II. , se presentan el perfil de involuta y sus principales ecuaciones. En la Sección III. , se abordan los aspectos del diseño de un arreglo de engranes planetarios. En la sección IV. , se revisa el proceso de micromaquinado de polisilicio SUMMIT-V para la fabricación de microistemas y finalmente en la Sección V. , se presentan las conclusiones de este trabajo.

II. PERFIL DE INVOLUTA

Una involuta es una curva trazada sobre otra curva previamente determinada [1]. El perfil de involuta se puede determinar matemáticamente sobre cualquier superficie siempre que se tenga una expresión analítica que la represente. Este perfil puede definirse tanto en el plano como en el espacio, aunque comúnmente cuando se habla de perfiles de involuta, se hace referencia a figuras en el plano. La forma más simple de concebir un perfil de involuta consiste en atar un extremo de una cuerda de una longitud fija, a la superficie sobre la que se desea obtener la involuta. El otro extremo debe estar libre pero enrollado completamente sobre la superficie en cuestión. Después se comienza a desenrollar la cuerda sobre la superficie mientras se mantiene lo más tensa posible. Los puntos que traza el extremo libre de la cuerda que se desenrolla forma el perfil de involuta de esta curva en particular.

Para el caso de un círculo, el perfil de involuta se ilustra en la Figura 1. A partir del centro del círculo base, con radio base r_b , se localizan el punto con coordenadas (X_c, Y_c) que se encuentra en un ángulo θ a partir de la vertical. El ángulo θ va tomando valores en el rango $[0, \alpha]$. El ángulo α

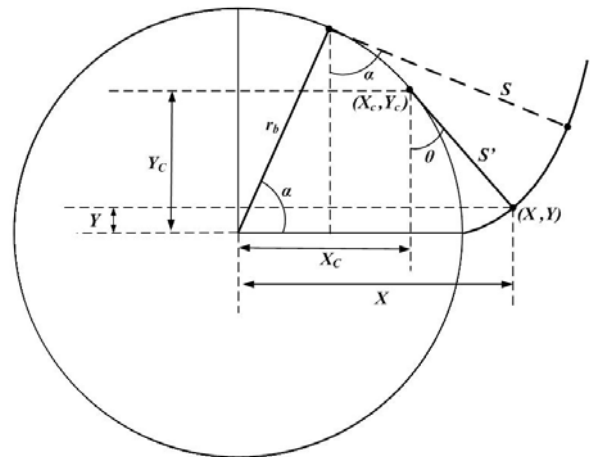


Figura 1. Definición del perfil de involuta

generalmente se proporciona en grados y se le conoce como el *ángulo de presión* del diente de involuta. Este es un parámetro importante en el diseño de los engranes y que está relacionado con la presión que se aplicará sobre cada diente de involuta [2]. La región del círculo que comprende α es la porción de la involuta que define el ancho del diente, y representa a la parte de la *cuerda* sobre la circunferencia que se está *desenrollando*. A partir de la figura, se deduce que el punto (X_c, Y_c) se calcula como:

$$\begin{aligned} X_c &= r_b \cos \theta \\ Y_c &= r_b \sin \theta \end{aligned} \quad (1)$$

donde θ satisface $0 \leq \theta \leq \alpha$.

Una vez determinado el punto (X_c, Y_c) , las coordenadas cartesianas (X, Y) que forman el perfil de involuta son:

$$\begin{aligned} X &= X_c + S' \sin \theta \\ Y &= Y_c - S' \cos \theta \end{aligned} \quad (2)$$

donde S' es la sección de la superficie S que se va desenrollando hasta una abertura igual al ángulo θ . Para cada valor del ángulo θ , se evalúa (2) para encontrar las coordenadas (X, Y) correspondientes en el perfil de la involuta. En teoría, θ puede tomar cualquier valor en el rango $[0, \alpha]$. Sin embargo, en la creación de un perfil de involuta en un sistema computacional, es necesario dividir el intervalo $[0, \alpha]$ en un número finito de valores los cuales pueda tomar el ángulo θ . Esto se logra introduciendo una variable de control que permite definir el número de elementos en que se dividirá el intervalo $[0, \alpha]$. Es decir, si t es una variable

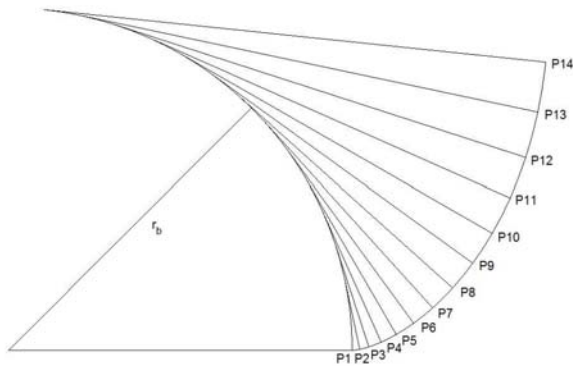


Figura 2. Detalle del Perfil de Involuta.

discreta la cual satisface que $0 \leq t \leq 1$, entonces a partir de la Figura 1, se deduce que la ecuación para el segmento S' es:

$$S' = \frac{2\pi r_b t \alpha}{360} \quad (3)$$

Mediante la adecuada manipulación de t en (3), es posible definir el grado de rugosidad, o inversamente de suavidad, del perfil de involuta. Este es un aspecto técnico que se debe tomar en cuenta al momento de la fabricación del engrane, ya que un perfil muy rugoso puede reducir la eficiencia del sistema de transmisión, mientras que un perfil muy fino, que es computacionalmente difícil de calcular, es innecesario debido a la resolución que el sistema de fabricación puede alcanzar. En la Figura 2 se muestra un perfil de involuta donde la sección del ángulo de presión se ha dividido en 14 segmentos.

Como dato histórico, si al ángulo α en (3), se permite tomar un valor arbitrario, mas allá de los 360° , entonces el perfil que se obtiene es la conocida *Espiral de Arquímedes*, nombrada de esta manera en honor al matemático griego Arquímedes [7]. A esta figura también se le conoce como espiral aritmética, que se define como el lugar geométrico que describe un punto moviéndose a velocidad constante, sobre una línea recta que a su vez gira a velocidad angular constante sobre un punto fijo llamado el origen de la espiral aritmética. En la Figura 3 se muestra dicho perfil cuando $\alpha = 720^\circ$. El ángulo α que define la Espiral de Arquímedes y el perfil de un diente de involuta se discute a continuación.

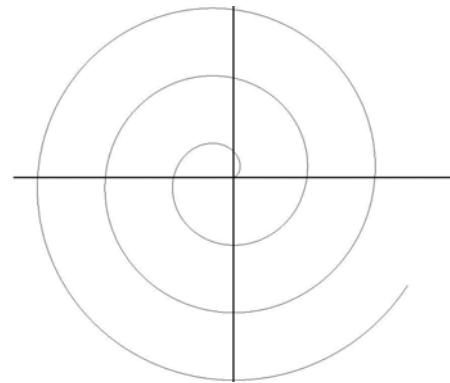


Figura 3. Espiral de Arquímedes.

ÁNGULO DE PRESIÓN

El ángulo de presión de un diente de involuta es importante siempre que los dientes de dos engranes entran en contacto. En la Figura 4 se muestra un par de circunferencias y se ilustra el concepto de ángulo de presión. Este ángulo se forma mediante la línea de acción de los dientes y la tangente de dos circunferencias imaginarias denominadas circunferencias primitivas con un radio igual al radio del pitch. Las circunferencias primitivas son un par de circunferencias que se tocan en un solo punto y que permiten definir la interacción de un par de engranes con dientes de involuta.

Los ángulos de presión típicos en la industria son 14.5° , 20° y 25° , aunque también es posible obtener perfiles de involuta con otros ángulos de presión

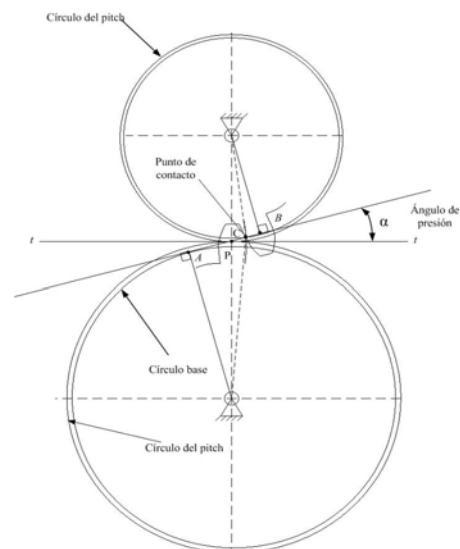


Figura 4. Ángulo de Presión.

[1], [2]. Para el caso de la fabricación de engranes con técnicas de micromaquinado superficial, el ángulo de presión solo se restringe a los segmentos mínimos que pueden ser transferidos al sustrato mediante el proceso de fotolitografía. Para el proceso de fabricación SUMMIT-V, el patrón mínimo es de 1 μ m [8]. Sin embargo, es recomendable apearse a los valores estándar tanto como sea posible.

PROPORCIONES DE LOS DIENTES

A fin de que un diseño de engranes sea eficiente, se deben tomar en consideración varios elementos como el número de dientes a colocar sobre un engrane. Es natural pensar que sobre una circunferencia determinada, el número de dientes de involuta que se pueden colocar no es arbitrario, sino que se limita a una cantidad máxima que lo hace eficiente para realizar el trabajo para el cual es diseñado. En la Figura 5 se muestra una sección de un engrane de involuta. Existe un círculo imaginario denominado el círculo del pitch que determina la región de ensamble de los dientes de dos engranes. El diámetro del pitch es el diámetro de este círculo. A partir de éste, se derivan varios parámetros de diseño del engrane que se deben considerar. A la relación del diámetro del pitch entre el número de dientes se le conoce como el módulo del engrane:

$$m = \frac{d_p}{N} \quad (4)$$

donde d_p es el diámetro del pitch y N el número de dientes en el engrane.

En (4), el valor de m generalmente está dado en milímetros/diente, y representa la parte de la circunferencia que le corresponde a cada diente. Entre mayor sea el número de dientes, menor será el valor numérico de m , lo que implica que los dientes estarán más juntos. Esto se deduce del hecho de que si d_p es el diámetro del círculo del pitch, entonces πd_p es la superficie de dicho círculo. Por otro lado, si N es el número de dientes y p_c es la región de la superficie del círculo del pitch que le corresponde a cada diente, conocida como paso circular, entonces $p_c N = \pi d_p$, de donde se deduce $d_p/N = p_c/\pi$. La interpretación de p_c se ilustra en la Figura 5. El inverso de m

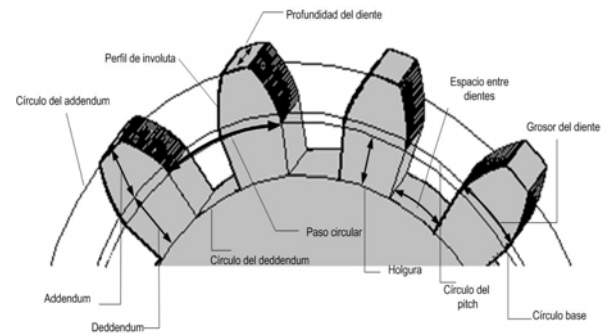


Figura 5. Detalle de engrane de dientes de involuta

representa el número de dientes por centímetro en el engrane, que informalmente proporciona una idea general de la densidad de dientes por centímetro lineal. Esto es un parámetro importante en la selección de engranes.

Existen otras dos distancias importantes llamadas *adendum* y *dedendum*, las cuales se definen a partir del círculo del pitch. El adendum es una distancia que sumada al radio del círculo del pitch define la distancia donde el perfil de involuta se trunca. Esto permite definir la agudeza de la parte final de diente. Un adendum muy grande provocará que el diente sea muy puntiagudo. En la práctica, el adendum generalmente es igual al valor numérico del módulo m . El dedendum es una distancia que se resta al radio del pitch y que permite definir la profundidad donde se insertan los dientes de otro engrane. En la práctica, el dedendum es 1.25 del valor del módulo m . El hecho de que el dedendum sea numéricamente 0.25 m mayor que el adendum permite que los dientes ensamble sin la posibilidad de que se atasquen al girar los engranes.

Existen otros parámetros importantes que por cuestiones de espacio no se detalla, como la profundidad y el grosor del diente, así como la holgura y el espacio entre dientes. Un lector interesado puede referirse a [1] y [2] para una descripción a detalle de estos parámetros de diseño de engranes de involuta.

III. ARREGLOS PLANETARIOS

Los engranes planetarios fueron objeto de un intenso estudio debido a su eficiencia, distribución de la carga y baja vibración, entre otros atributos. Una buena introducción sobre engranes planetarios

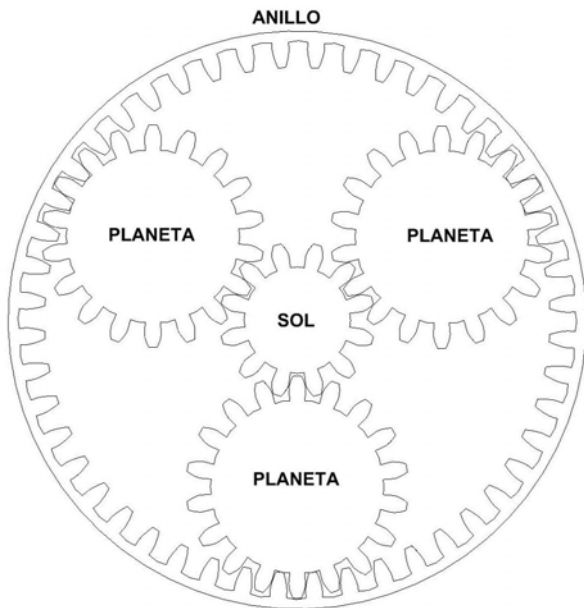


Figura 6. Arreglo de Engranes Planetarios

se encuentra en [2]. En un conjunto de engranes planetarios, al engrane central se le denomina como el sol, mientras que a los engranes en torno a éste se les denomina planetas. Los planetas tienen su eje móvil y giran alrededor del sol. Además existe una estructura que mantiene unido al conjunto de engranes. En algunas configuraciones existe un anillo externo, mientras que en otras se utiliza una estructura que sujeta a los ejes de los planetas y los mantiene unidos al sol. En la Figura 6 se muestra un conjunto de engranes planetarios, con un sol, tres planetas y el anillo exterior. En el diseño de este tipo de arreglos, la distancia entre ejes del sol a los planetas es un factor clave que permite el adecuado ensamble de los engranes. En la Figura 7 se muestra el concepto de distancia entre centros. Se adopta la convención de que el engrane más pequeño se llama piñón y el más grande se le denomina el engrane. Bajo dicha convención, d_g representa el diámetro del pitch del engrane mientras que d_{pi} representa el diámetro del pitch del piñón. Una condición necesaria para que los engranes embonen adecuadamente, es que sus módulos sean iguales, es decir:

$$m_g = \frac{d_g}{N_g} = \frac{d_{pi}}{N_{pi}} = m_{pi} \quad (5)$$

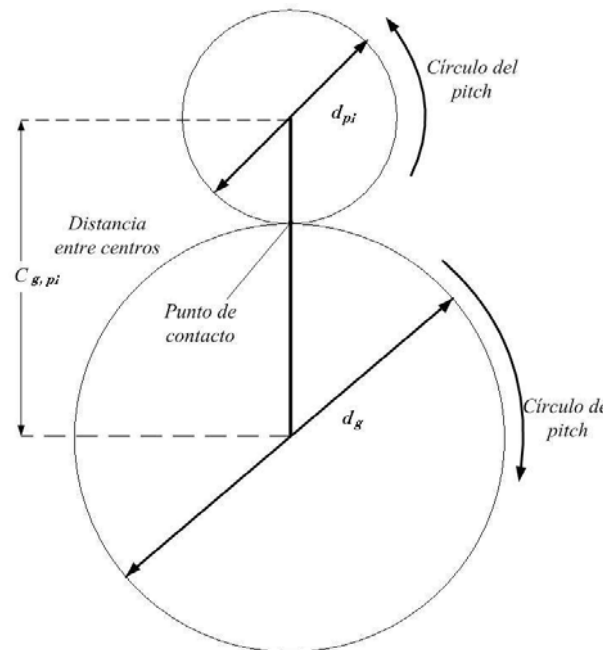


Figura 7. Distancia entre Centros

donde m_g, d_g, N_g representan, respectivamente, el módulo, diámetro y número de dientes del engrane, y m_{pi}, d_{pi}, N_{pi} representan el módulo, diámetro y número de dientes del piñón, respectivamente. A fin de que los engranes tengan un ensamble perfecto, la distancia entre sus centros, denominada $C_{g,pi}$ debe satisfacer

$$C_{g,pi} = \frac{d_g + d_{pi}}{2} \quad (6)$$

Utilizando (6), es posible determinar la distancia a la cual los centros del sol y los planetas se deben colocar. Sin embargo, a fin de lograr el adecuado ensamble de los planetas en torno al sol, es necesario conocer además el ángulo de colocación de cada planeta, el cual se determina como:

$$\alpha_p = \frac{360}{n_p} \quad (7)$$

donde α_p representa el ángulo de colocación para cada planeta en torno al sol y n_p el número de planetas.

Una vez establecida el ángulo de colocación de un planeta, es posible determinar la ubicación de los demás planetas sumando el valor del ángulo α_p en (7) para cada planeta. Mediante una conversión simple de coordenadas polares a cartesianas, se

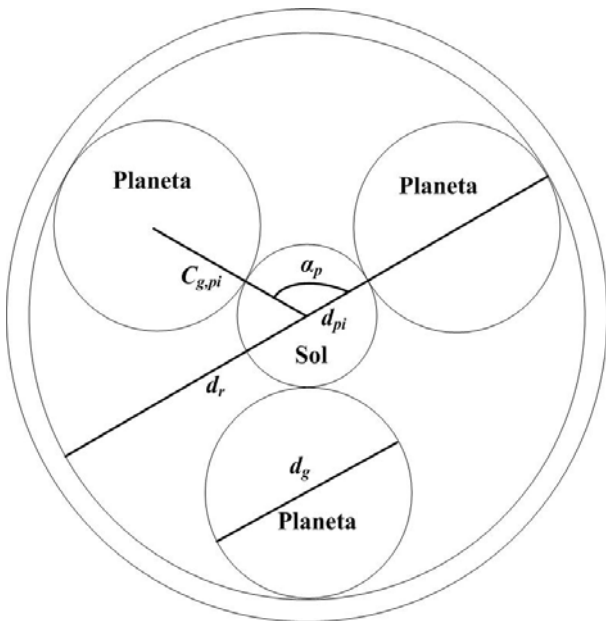


Figura 8. Conceptos Básicos Arreglos Planetarios

puede encontrar los valores (x_{p_i}, y_{p_i}) de los centros de cada i -ésimo planeta para su adecuado ensamble. Esto se resume en la siguiente ecuación

$$\begin{aligned} x_{p_i} &= C_{g,pi} \cos i\alpha_p \\ y_{p_i} &= C_{g,pi} \sin i\alpha_p \end{aligned} \quad (8)$$

donde i representa el i -ésimo planeta a ensamblar. Las coordenadas calculadas mediante (8) resultan muy adecuadas para su programación mediante una implementación en computadora.

Una vez definidos y alineados el sol y los planetas, se diseña el mecanismo que los mantendrá unidos. En este caso se considera un anillo exterior a los planetas y con dientes de involuta en su parte interior. Este mecanismo mantiene a los planetas unidos en torno al sol. El diámetro del anillo se determina como

$$d_r = d_s + d_p \quad (9)$$

donde d_r denota el diámetro del anillo exterior, d_s el diámetro del sol y d_p el diámetro de los planetas.

El anillo de soporte no solo es una estructura circular, sino que debido a los dientes en la parte interior, gira en el mismo sentido que los planetas. A partir del diámetro del anillo calculado con (9), se definen los perfiles de involuta para los engranes internos del anillo de soporte. Los conceptos más relevantes en la alineación de los

planetas y en la definición del anillo exterior se ilustran en la Figura 8.

IV. TECNOLOGÍA DE FABRICACIÓN

Hasta este momento, el trabajo se ha enfocado en obtener un procedimiento automático que permita definir el radio del sol, el número de dientes y la cantidad de planetas en el arreglo y que de manera automática genere el conjunto de arreglos. Además, es posible generar un conjunto de engranes planetarios nuevo con solo variar algunos parámetros. Ahora se tomará en cuenta la tecnología de fabricación con la que se pretende obtener el arreglo de manera física. Para este diseño se utilizará el proceso SUMMIT-V (Sandia Ultra-planar Multi-level MEMS Technology V, en inglés) patentado por los Laboratorios Sandia, en Albuquerque, Nuevo México [8]. SUMMIT-V es un proceso de fabricación de 5 capas de polisilicio, de las cuales 4 son mecánicamente móviles y 1 anclada al sustrato. En la Figura 9 se muestran las capas de SUMMIT-V. El proceso inicia sobre un sustrato de 6 pulgadas de tipo-n orientado en el plano $\langle 1,0,0 \rangle$. El sustrato está altamente dopado y sirve para realizar interconexiones eléctricas y como plano de tierra. Sobre el sustrato se agrega una capa de óxido de 0.63um. Esta capa funciona de aislante térmico y eléctrico, y es posible grabar patrones sobre ella. Sobre la capa de óxido se deposita una capa de nitruro de silicio, de 0.80um de espesor. Sobre la capa de nitruro se deposita la primera capa de polisilicio de 0.30um de espesor. A esta capa se le conoce como MMPOLY0. Sobre MMPOLY0 se deposita una capa de óxido de

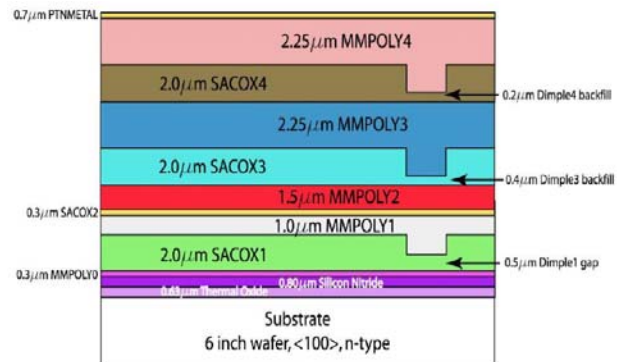


Figura 9. El proceso de fabricación SUMMIT-V

sacrificio de 2.0um de espesor conocida como SACOX1. Sobre la capa SACOX1 se pueden grabar patrones llamados DIMMPES, que son una especie de patillas para reducir la superficie de contacto y por consecuencia la fricción. El procedimiento de colocar polisilicio y óxido de sacrificio, así como la creación de DIMMPLES, se repite hasta llegar a MMPOLY4, con espesores determinados, según se ilustra en la Figura 9. El acabado final es una capa de metal de 0.70um de espesor conocida como PTNMETAL.

DISEÑO DE MÁSCARAS

De las capas de polisilicio que ofrece el proceso de fabricación SUMMIT-V, cuatro son móviles, estas son MMPOLY1, MMPOLY2, MMPOLY3 y MMPOLY4, y son candidatas para grabar los patrones geométricos que formarán el arreglo planetario. Según la Figura 6, se deduce que los dientes del sol, los planetas y el anillo deben tener contacto, por lo que dichos patrones deberán estar dibujados en la misma capa de polisilicio. Además, la o las capas donde se diseñen los patrones geométricos que producirán el arreglo planetario deben tener movimiento. Algo que no se menciona en el texto, pero que es igualmente importante, es el hecho de que las estructuras deben estar de alguna manera sujetas al sustrato, a fin de que no salgan de sus ejes cuando estén en movimiento. Por esta razón, se eligió la capa MMPOLY4 para la creación de *remaches* para los engranes, y la capa MMPOLY1 para ser la base del arreglo planetario y reducir la fricción. De esta manera, quedan las capas MMPOLY2 y MMPOLY3 para el diseño de los patrones de los engranes planetarios. Mientras se diseñan las máscaras, es posible remover la capa SACOX3 a fin de lograr una capa de polisilicio móvil que sume los espesores de las capas MMPOLY2 y MMPOLY3, esto es 3.75um de espesor, con lo que se logra una mayor superficie de contacto entre los dientes de los engranes, y a su vez, mejora la eficiencia del diseño. Además, se aprovecha la capa DIMMPLE para reducir la fricción de los engranes.

Se realizó un conjunto de scripts para AutoCAD para la generación automática de engranes de involuta, así como de arreglos planetarios de estos engranes. El script toma como datos de entrada el

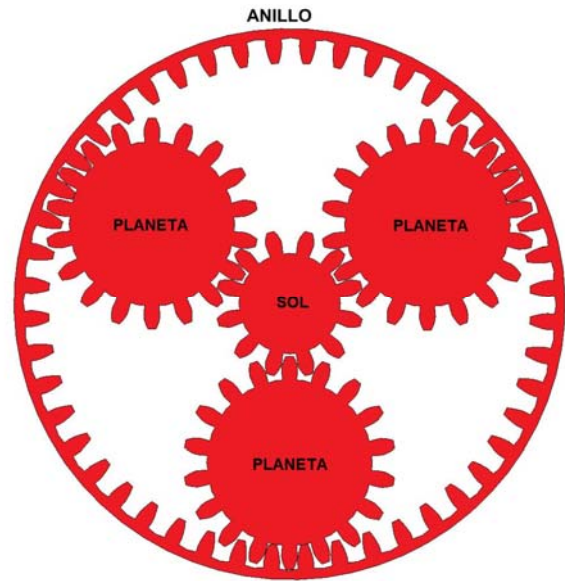


Figura 10. Máscara de Fabricación en MMPOLY2

ángulo de presión de los dientes de involuta, el número de dientes del sol y los planetas, y a partir de ellos, implementa las fórmulas para calcular todos los elementos necesarios en la creación del arreglo planetario de manera automática. Se creó una interfaz de usuario en VBA [9] que permite la fácil introducción de los datos y que proporciona acceso directo a los algoritmos implementados.

En la Figura 10 se muestra el diseño de un arreglo planetario en la capa MMPOLY2 con los siguientes datos:

- Número de soles: 1,
- Número de dientes del sol: 12
- Número de planetas: 3
- Número de dientes de los planetas: 18
- Número de dientes del anillo: 48
- Ángulo de presión de los dientes: 14.5°
- Ángulo de orientación de los planetas: 120°

V. CONCLUSIONES

Este trabajo abordó el diseño de engranes planetarios para su construcción mediante procesos de fabricación de microistemas. Se presentó el perfil de involuta y las ecuaciones que la definen. Se derivó una representación cartesiana de las ecuaciones de la involuta para su implementación mediante scripts en una herramienta de diseño. Se discutieron los aspectos más relevantes en el

diseño de engranes con dientes de involuta, como el ángulo de presión y la línea de ensamble de los engranes. Se discutieron las condiciones para el alineamiento de los planetas en torno al sol para un ensamble adecuado y se propuso el diseño del anillo exterior para mantener el arreglo unido. Se presento una breve descripción del proceso de fabricación SUMMIT-V y se presento una máscara de fabricación de un arreglo planetario con un sol con 12 dientes de involuta, tres planetas con 18 dientes de cada uno y un anillo exterior 48 dientes, todos con perfil de involuta de 14.5° de ángulo de presión.

VI. AGRADECIMIENTOS

Los autores agradecen a la Subsecretaría de Educación Superior e Investigación Científica por los apoyos de Fomento a la Generación y Aplicación Innovadora del Conocimiento y la Beca de Estudiante otorgados al proyecto “Diseño de MicroEspejos MEMS” mediante oficio PROMEP/103.5/08/2919.

VII. REFERENCIAS

- [1] Darle W. Dudley, Handbook of Practical Gear Design, Mechanical Engineering CRC Press, 1st edition, 1994. ISBN 1566762189.
- [2] Faydor L. Litvin and Alfonso Fuentes, Gear Geometry and Applied Theory. Cambridge University Press; 2nd edition, 2004. ISBN 0521815177.
- [3] E.S. Kolesar and C.A. Edwards, “Parameterized Approach to MEMS Mechanical Gear Design,” International Journal of Advanced Manufacturing Systems, Vol. 10, No. 2, pp. 45-50, 2007.
- [4] E.S. Kolesar and C.A. Edwards, “MEMS Gear Design with a Parameterized Approach,” Proceedings of the TexMEMS VII Conference, The Univ. of Texas at El Paso, T. Rivera Conference Center, El Paso, TX, September 22, 2005.
- [5] Automated Design for Micromachining, <http://www.sfu.ca/adm/>
- [6] Mahalik N. P., MEMS. McGraw-Hill, 2008. ISBN 0070634459.
- [7] Thomas Little Heath, A History of Greek Mathematics, Vol. 2, pp 377. Dover Publications, 1981. ISBN 0486240746.
- [8] Sandia Ultra-planar, Multi-level MEMS Technology-5 (SUMMIT-V™) <http://www.mems.sandia.gov/tech-info/summit-v.html>
- [9] Marion Cottingham, Mastering AutoCAD VBA. Sybex Editor; 1st edition, 2001. ISBN 0782128718.

VIII. AUTORES

Dr. Raúl Campos Rodríguez es Ingeniero en Computación por el Centro Universitario de la Ciénega, UDG, 2000. Estudio la Maestría y Doctorado en Ciencias en Ingeniería Eléctrica en el CINVESTAV Guadalajara, de 2002 a 2007. Trabaja activamente en el diseño de algoritmos para sistemas de eventos discretos y ha dedicado últimamente parte de su tiempo al diseño de Microsistemas, o MEMS, específicamente en mecanismos basados en fuerza electrostática y la expansión térmica.

Dra. Mildreth Alcaraz Mejía es Ingeniero en Sistemas Computacionales por el Instituto Tecnológico de Colima, 2001. Obtuvo los títulos de Maestro y Doctor en Ciencias con especialidad en Ingeniería Eléctrica en el CINVESTAV Unidad Guadalajara, en el 2004 y 2007, respectivamente. Actualmente, se dedica a la aplicación y desarrollo de técnicas de control embebidas con una orientación hacia el control de los Microsistemas.

Martín Gerardo Aguirre Villa es estudiante de la carrera de ingeniería en computación del 8° semestre en el Centro Universitario de la Ciénega, de la Universidad de Guadalajara. Trabaja parcialmente en este proyecto como parte de su servicio social en la opción de Actividades de Investigación.

Alejandra Aceves Paredes es estudiante de la carrera de ingeniería en computación del 8° semestre en el Centro Universitario de la Ciénega, de la Universidad de Guadalajara. Trabaja parcialmente en este proyecto como parte de su servicio social en la opción de Actividades de Investigación.

CORRECCIÓN CROMÁTICA DE IMÁGENES NATURALES

¹P. A. Patlán-Rosales, ²J. G. Aviña-Cervantes, ³A. F. Maldonado-Mendoza, ⁴E. D. Martínez-Rodríguez.

Universidad de Guanajuato - División de Ingenierías Campus Irapuato-Salamanca
Carretera Salamanca-Valle de Santiago km. 3.5+1.8 km

e-mail: ¹palfonso.01@gmail.com, ²avina@salamanca.ugto.mx, ³anglmlndnd@gmail.com,
⁴ing.edgardmr@gmail.com

RESUMEN

El ser humano percibe el mundo que le rodea a través de sus sentidos, con ellos es capaz de adquirir e interpretar sensaciones y estímulos que le permiten interactuar con su entorno. La visión es de gran importancia para identificar objetos y permite tener una descripción gráfica de la información de estos.

Un sistema de visión artificial ayuda a controlar sistemas que necesiten caracterizar su entorno. Este sistema es útil para navegación y manipulación de objetos a través de la detección.

La *constancia del color* es un fenómeno de la percepción del color, por la que una superficie parece mantener su *apariencia cromática* bajo diferentes condiciones de iluminación, esto es posible gracias a la compensación que hace el cerebro a la información que recibe del sistema de visión.

En este trabajo se implementaron algoritmos para la corrección del color como: *mundo blanco*, *mundo gris* y *Retinex*.

Palabras clave: *Constancia de color*, *mundo blanco*, *el mundo gris* y *Retinex*.

I. INTRODUCCIÓN

Color

El color no es una característica de una imagen u objeto, sino que es más bien una apreciación subjetiva humana. Por tanto, podemos definirlo

como una sensación que se produce en el ojo a causa de una estimulación en sus mecanismos nerviosos, debido a la energía luminosa de ciertas longitudes de onda [2].

En 1665 Isaac Newton descubrió que la luz del sol, al pasar a través de un prisma se dividía en varios colores, conformando un espectro, con lo cual consiguió la descomposición de la luz en los colores del espectro visible, los cuales son: violeta, azul, verde, amarillo, naranja y rojo (ver Figura 1) [1].



Figura 1. Rangos del espectro visible por el hombre.

En base a la teoría ondulatoria, se sabe que cada color es en realidad una onda electromagnética de frecuencia o longitud de onda determinada, que al momento que incide sobre un prisma óptico como luz blanca, se desvía en una dirección diferente, debido a que el índice de refracción de ese material varía con la frecuencia de onda penetrante, haciendo posible que el ojo humano pueda percibir los colores del espectro visible en forma de radiaciones monocromáticas.

La Comisión Internacional de Iluminación (ICI en inglés, CIE en francés), es un organismo internacional dedicado al estudio de la luz, iluminación, color y los espacios de color. Esta comisión fue creada para describir racionalmente

los colores de las luces, tal y como las interpreta el cerebro humano, y en 1970 estableció la definición más aceptada para color, la cual dice lo siguiente [3]:

El color es el aspecto de la percepción visual mediante el cual un observador puede distinguir diferencias entre dos campos del mismo tamaño, forma y textura, tales que puedan ser debidas a diferencias en la composición espectral de las radiaciones relacionadas con la observación.

Una fuente de luz puede radiar dos tipos de luz: acromática o cromática. La primera tiene como único atributo la intensidad de la luz; la segunda, emite una luz que posee radiancia (Watts), luminancia (lúmenes) y brillo, además, tiene una componente de color dominante.

Atributos del color

El color juega un papel fundamental en la percepción que se tiene de un objeto o de una escena. Sin embargo, para definir un color se deben identificar las características que lo describen. Las características fundamentales son: matiz, saturación y brillo [4]. Las cuales también son conocidas como dimensiones tonales.

Matiz es el atributo de la sensación visual, según el cual, un área parece similar a uno de los colores percibidos: rojo, amarillo y azul o la combinación de dos de ellos. Está relacionada con la longitud de onda dominante del espectro. Se mide de acuerdo con la proximidad que tiene un color con relación a otro (Figura 2).



Figura 2. Matiz de un color

Saturación es el inverso de la cantidad de gris que contiene un color. Cuanto más alto sea el contenido de gris, menor será la saturación (Figura 3),

esta característica hace referencia a la pureza de un color.



Figura 3. Saturación

Brillo es el atributo de la sensación visual según el cual un área o superficie, parece emitir más o menos luz (Figura 4).



Figura 4. Brillo del color Rojo

Sistema visual humano

La información procesada por el ojo humano entra en forma de luz a través de la lente y sobre la retina, que está situado en la parte posterior del globo ocular [2]. En la retina hay 3 tipos de conos: *L*, *M* y *S*, cada uno de ellos es sensible a ciertas longitudes de onda. Los tipo *L* (Long) representan una sensibilidad máxima para las longitudes de onda más largas y perciben los tonos rojos (500 a 600 nm). Los tipo *M* (Middle) tienen mayor sensibilidad a las longitudes de onda medias y perciben los tonos verdes (430 a 620 nm). Los tipos *S* (Short) son más sensibles a las longitudes de onda corta y perciben los tonos azules (370 a 530 nm) (Figura 5) [5].

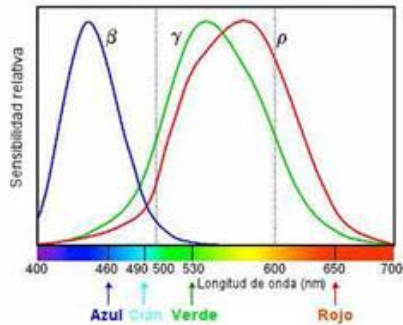


Figura 5. Sensibilidad Espectral Humana

Adaptación cromática

En el procesamiento digital de imágenes el color es una característica de gran importancia cuando se requiere segmentar una imagen, extraer características, detectar sombras, identificar objetos dentro de una imagen, etc. Esto debido a que una imagen a color aporta tres veces más información que una imagen monocromática.

Si el color es una característica útil para identificar un objeto, entonces este se debe mantener lo más constante posible cuando el objeto es visto en diferentes contextos. Las personas mantienen la percepción del color de un objeto a pesar de las variaciones de distancia y el ambiente luminoso en que se encuentran.

Cuando un objeto iluminado por dos fuentes luminosas distintas, se sigue percibiendo del mismo color se dice que hay *constancia del color* [2] y por tanto se está llevando a cabo un correcto proceso de adaptación cromática. McAdam [6] da una definición más formal al concepto de constancia del color y dice:

La apariencia del color de los objetos en una escena compleja es independiente de la cantidad de fotones que llega a los receptores de la retina, así, el cambio en el contenido espectral del iluminante, no afecta el color de los objetos, de ahí el nombre de constancia del color.

II. OBJETIVO

El objetivo de este trabajo es aplicar, combinar y desarrollar algoritmos que sean capaces de corregir el color en imágenes y que esto nos ayude a visualizar de una mejor manera las características en las imágenes y así implementarlos en procesamientos como la detección de objetos, reconocimiento de patrones y navegación visual para robots móviles.

III. DESCRIPCIÓN DEL PROBLEMA

Debido al rango de despliegue y procesamiento en los dispositivos electrónicos comerciales (ej. Cámara Digital), es común que las imágenes obtenidas no muestren los colores reales de la escena original.

Un problema que afecta a los dispositivos de captura es la luz natural, ya que esta cambia según el momento del día y las condiciones atmosféricas. En las horas de la mañana tiene un tono rosa, por la tarde tiene un tono amarillo y finalmente una tendencia azul por la noche. Todos estos cambios hacen que la misma escena cambie su apariencia cromática.

Los colores que registra una cámara digital dependen de la iluminación de la escena, la luz que entra por el diafragma y es registrada, no siempre es igual. Cierta luz puede ser natural o artificial, de manera que las características de ésta pueden tener grandes diferencias. Unas de estas características es la temperatura del color (Figura 6), es decir la componente del color dominante en la fuente luminosa.



Figura 6. Temperatura del color.

IV. METODOLOGÍA

Modelos de Adaptación Cromática

Los métodos modernos de adaptación cromática están basados, conceptual y matemáticamente, en el modelo de Johannes Von Kries. Dicho modelo dice que:

Los componentes individuales presentes en el órgano de visión son completamente independientes entre sí, y cada uno es fatigado y adaptado exclusivamente de acuerdo a su propia función.

Von Kries establece que cualquier cambio en las condiciones de adaptación se traduce en un simple rescaldado de las curvas de sensibilidad espectral de los mecanismos fundamentales [8].

La implementación moderna de la hipótesis de Von Kries puede expresarse en la siguiente ecuación:

$$\begin{pmatrix} L_2 \\ M_2 \\ S_2 \end{pmatrix} = \begin{pmatrix} K_L & 0 & 0 \\ 0 & K_M & 0 \\ 0 & 0 & K_S \end{pmatrix} \begin{pmatrix} L_1 \\ M_1 \\ S_1 \end{pmatrix} \quad (4.1)$$

Donde L, M y S representan las respuestas iniciales de los conos; K_L , K_M y K_S , son los coeficientes usados para escalar las señales iniciales de los conos; L_2 , M_2 y S_2 son las señales de los conos después de la adaptación. Con esto se obtiene una ganancia controlada en la cual, cada uno de los tres tipos de conos tiene un coeficiente de ganancia separado. Los coeficientes K_L , K_M y K_S son calculados para ser la respuesta inversa de los conos L, M y S, tomando en cuenta una escena blanca o bien la parte con máxima iluminación en la imagen, esto puede expresarse en las siguientes ecuaciones:

$$K_L = \frac{1}{L_{\max}} \quad (4.2)$$

$$K_M = \frac{1}{M_{\max}} \quad (4.3)$$

$$K_S = \frac{1}{S_{\max}} \quad (4.4)$$

Las ecuaciones (4.2), (4.3) y (4.4) son la representación matemática del Modelo de Von Kries, donde se cumple que cada una es fatigada o adaptada exclusivamente de acuerdo a su propia función. El modelo de Von Kries puede ser utilizado para calcular los colores correspondientes a dos condiciones de iluminación diferentes, para calcular la post-adaptación de las señales de la primera condición igualándolas a la post-adaptación de la segunda iluminación, de esta forma, se puede tener información sobre el segundo iluminante y revertir su efecto cromático para realizar la corrección, las siguientes ecuaciones expresan los conos después de la adaptación cromática:

$$L_2 = \frac{L_1}{L_{\max 1}} L_{\max 2} \quad (4.5)$$

$$M_2 = \frac{M_1}{M_{\max 1}} M_{\max 2} \quad (4.6)$$

$$S_2 = \frac{S_1}{S_{\max 1}} S_{\max 2} \quad (4.7)$$

Sin embargo, en la mayoría de los casos es conveniente expresar el modelo de adaptación cromática en términos de una matriz de transformación. El modelo de Von Kries se expresa también en esta notación, como:

$$\begin{pmatrix} L_2 \\ M_2 \\ S_2 \end{pmatrix} = \begin{pmatrix} \frac{L_{\max 2}}{L_{\max 1}} & 0 & 0 \\ 0 & \frac{M_{\max 2}}{M_{\max 1}} & 0 \\ 0 & 0 & \frac{S_{\max 2}}{S_{\max 1}} \end{pmatrix} \begin{pmatrix} L_1 \\ M_1 \\ S_1 \end{pmatrix} \quad (4.8)$$

Mundo Gris

Si una cantidad suficiente de colores está presente en una imagen, la medida de color es cercana al gris. Dependiendo del valor de esa medida sería posible estimar la ganancia para la corrección cromática de nuestras imágenes. Nos referimos al mecanismo basado en la hipótesis del mundo gris y su aplicación significa elegir una zona de la imagen para ver la estadística media de color: rojo R_s , verde G_s y azul B_s . Las ganancias de cada plan se estiman en dos formas diferentes:

1. La primera consiste en buscar cual de los tres colores es tomado como referencia. En la práctica se considera la decisión al color verde. El modelo Von Kries es aplicable si se considera una relación lineal entre los conos y las respuestas a imágenes en color:

$$K_R = \frac{G_s}{R_s}, K_G = 1.0, K_B = \frac{G_s}{B_s}$$

2. La segunda considera el nivel de intensidad de una zona seleccionando un parámetro de referencia. Las ganancias son:

$$K_R = \frac{I_s}{R_s}, K_G = \frac{I_s}{G_s}, K_B = \frac{I_s}{B_s}$$

Donde $I_s = \frac{R_s + G_s + B_s}{3}$ en estas condiciones

la medida de ganancias es igual a la unidad, esto se recomienda cuando se utiliza un método de adaptación cromática más general.

Mundo Blanco

En algunos casos el sistema visual humano normaliza las respuestas de sus fotorreceptores en los máximos valores de una zona hipotética blanca como referencia, y así lograr la constancia del color, nos referimos al mecanismo del mundo blanco. En este proceso el problema es de la detección de la zona blanca de referencia. El método permite promediar los componentes

cromáticos, esto reduciendo el área de referencia al valor del blanco (R_w , G_w y B_w).

En la fotografía, es común colocar en la escena un blanco para retocar imágenes después de ser adquiridas. La corrección de color se obtiene mediante la colocación a nivel de los tres colores R , G y B mediante las siguientes ecuaciones:

$$K_R = \frac{R_w}{R_s}, K_G = \frac{G_w}{G_s}, K_B = \frac{B_w}{B_s}$$

La metodología de este modelo de corrección cromática se describe mediante el siguiente algoritmo:

1. *Buscar el valor más alto para las componentes del espacio de color RGB:*

$$R_w = \max(R)$$

$$G_w = \max(G)$$

$$B_w = \max(B)$$

2. *Obtener valores para ajustar la imagen de acuerdo al punto más blanco:*

$$K_{RW} = 255/R_w$$

$$K_{GW} = 255/G_w$$

$$K_{BW} = 255/B_w$$

3. *Aplicar a todas las capas de la imagen este factor de corrección, esto se puede expresar mediante la siguiente ecuación:*

$$\begin{pmatrix} R_2 \\ G_2 \\ B_2 \end{pmatrix} = \begin{pmatrix} K_{RW} & 0 & 0 \\ 0 & K_{GW} & 0 \\ 0 & 0 & K_{BW} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

Retinex

El algoritmo Retinex fue propuesto por Edwin Land y John McCann [7], este algoritmo fue el primer intento para desarrollar un modelo

computacional que emule la constancia del color del sistema de visión humano.

Esta teoría es considerada como una mejora a la hipótesis de Von Kries. Su característica principal es que trata la distribución espacial de colores en una escena, con esto se trata de explicar la independencia de la apariencia del color en la distribución espectral de la luz reflejada.

El algoritmo Retinex [9] está basado en la respuesta logarítmica que posee el sistema visual humano.

Realce

El realce es un método que se utilizó para realzar cada color, permitiendo ver los colores de mejor manera. El método consiste en encontrar un factor a partir de las tres bandas, dividiendo cada píxel entre la luminosidad de la imagen buscando así un máximo valor de entre los tres colores. Este máximo valor de cada píxel sirve como referencia para que cada píxel, antes dividido entre la luminosidad, sea dividido entre este valor, dando así como resultado un factor por el cual se multiplica cada imagen de salida.

V. RESULTADOS

Los modelos antes mencionados fueron desarrollados bajo el sistema operativo Linux y en el lenguaje C++, además, a cada una de las técnicas fue aplicado un realce, obteniendo los siguientes resultados:

Imágenes originales



Figura 7: Imágenes Originales

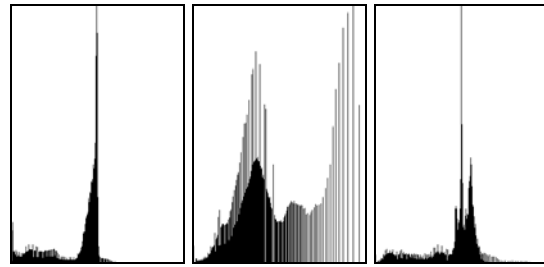


Figura 8: Histogramas de Luminancia para cada imagen Original

Mundo Gris con Realce



Figura 9: Imágenes corregidas con Mundo gris y aplicando un realce

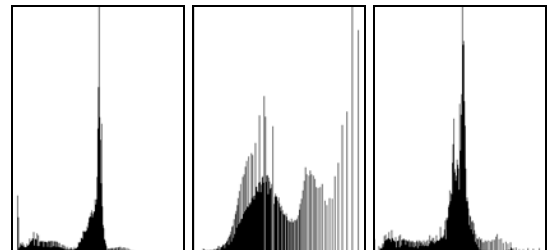


Figura 10: Histogramas de Luminancia para cada imagen corregida con el mundo gris.

En el mundo gris con realce se puede observar que la luminancia de la primera imagen de la Figura 8 se incrementa a un rango un poco más amplio, mientras que en la segunda imagen distorsiona el color y disminuye la luminancia, en la tercera imagen se obtuvieron muy buenos resultados debido a que el algoritmo encontró más tonos grises de referencia.

Mundo Blanco con Realce



Figura 11: Imágenes corregidas con Mundo Blanco y aplicando un realce

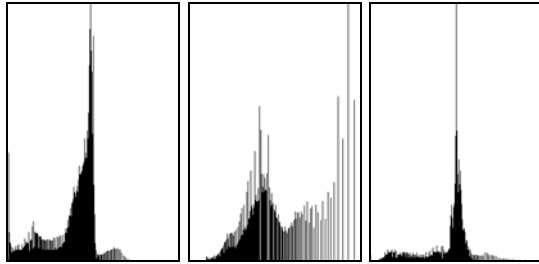


Figura 12: Histogramas de Luminancia para cada imagen corregida con el mundo blanco.

En el mundo blanco se puede observar mejoras con respecto a la iluminación de la primera y tercera imagen de la Figura 11, mientras que la segunda imagen distorsiona el color verde, debido a que el color de referencia no es muy blanco.

Retinex



Figura 13: Imágenes corregidas con Retinex

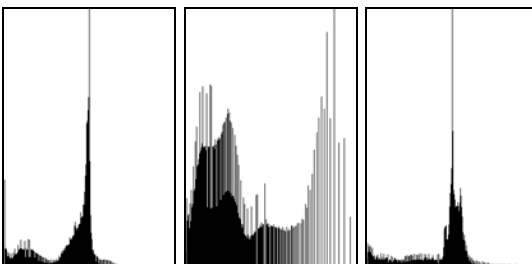


Figura 14: Histogramas de Luminancia para cada imagen corregida con Retinex.

El algoritmo de Retinex al contrario de los otros dos algoritmos de corrección obtuvo buenos resultados en las tres imágenes (Figura 13).

VI. CONCLUSIONES

Existen varios factores de los cuales depende la percepción del color, los cuales de una u otra forma no pueden ser controlados en el momento de la captura de la imagen y esto requiere que se realice una corrección posterior. Mediante las técnicas de adaptación cromática empleadas en esta investigación se obtuvieron resultados favorables en imágenes con iluminación no muy moderada como se muestra en la sección de resultados, mientras que en imágenes demasiado oscuras ningún método logra corregirlas de manera eficiente, debido a esto se necesita investigar y aplicar estos algoritmos en otros espacios de color como: HSV, CIELuv, HSL, entre otros.

VII. REFERENCIAS

- [1] Rafael C. Gonzalez and Richard Eugene Woods, *Digital image processing*, Prentice Hall, Third Edition, p. 395, ISBN-13: 978-0-13-168728-8, 2007.
- [2] Marc Ebner, *Color Constancy*, Jhon Wiley, ISBN: 987-0-470-05829-9, 2007.
- [3] Francesca Gasparini and Raimondo Schettini, *Color balancing of digital photos using simple image statistics*, Pattern Recognition, 37(6):1201–1217, 2004.
- [4] Richard J. Gerrig, Philip G. Zimbardo, *Psicología y vida*, Pearson Educación, ISBN: 9702605849, p. 110, 2005.
- [5] Mark D. Fairchild, *Color appearance models*, Jhon Wiley, ISBN: 978-0-470-01216-1, 2nd edition, p. 8-20, 2005.
- [6] Judd, D.B, McAdam, D.L, Wyszecki, G, "Spectral distribution of typical daylight as function of correlated color temperature", J. Opt. Soc. Am., pp.1031-41, 1964.
- [7] Land E. H. y McCann J. J., *Lightness and Retinex Theory*, 1971. Journal of the Optical Society of America, 1:1.
- [8] J. Von Kries, *Chromatic Adaptation*, In Festschrift der Albrecht -Ludwigs-Universität, 1902 [Translation: D. L. MacAdam, "Colorimetry- Fundamentals"],

SPIE, volume MS 77 of Milestone Series, 1993.

- [9] David H. Brainard and Brian A. Wandell, “*Analysis of the retinex theory of color vision*”, Journal Optical Society of America, vol. 3, no. 10, pp. 11, October 1986.

INSPECCIÓN DE TARJETAS ELECTRÓNICAS MEDIANTE TÉCNICA DE VISIÓN POR COMPUTADORA

Jorge Hernández-Constante¹, Marco Cedano-Olvera², Eduardo Velázquez-Mora³, Dora P. Hernández Roque⁴

^{1,3,4}Universidad de Guadalajara, División de Electrónica y Computación, Departamento de Electrónica, Av. Revolución 1500, Módulo "O", S.R. C.P. 44430, Guadalajara, Jal., México.
jorge.constante@red.cucei.udg.mx

²Universidad de Guadalajara, División de Electrónica y Computación, Departamento de Computación, Av. Revolución 1500, Módulo "O", S.R. C.P. 44430, Guadalajara, Jal., México.
marco.cedano@cucei.udg.mx

RESUMEN

Hay muchas ventajas al usar sistemas de visión artificial para llevar a cabo un sistema de control de calidad, sobre todo en procesos donde es imposible usar otros métodos. Sin embargo, los algoritmos de visión eficaces de tiempo real dependen de una relación apropiada de velocidad y costo de cómputo. Las aplicaciones de los filtros no lineales en esta área han producido resultados buenos e interesantes, debido a su eficacia para el reconocimiento de objetos en tiempo real. En este trabajo en desarrollo, se lleva a cabo una aplicación de reconocimiento de patrones que usa los filtros no lineales. Se presentan algunos resultados parciales durante una simulación de un proceso de inspección industrial en un caso de falla.

Palabras Clave: filtros no lineales, reconocimiento de patrones, inspección industrial.

I. INTRODUCCIÓN

Algunas industrias han estado buscando cómo automatizar sus procesos para mejorar la calidad de sus productos y evitar que los operadores humanos lleven a cabo una gran cantidad de labores consideradas como de mucha

fatiga o riesgosas [1]. No obstante, nos encontramos que una gran cantidad de tareas consideradas como livianas, tal como la inspección de dispositivos electrónicos, es aún una tarea realizada por operadores humanos. Debido a esta situación, estos procesos pueden tener deficiencias en la aplicación de los criterios de aceptabilidad basados en las normas de control de calidad, porque los inspectores humanos se fatigan en la última etapa de la jornada laboral y por trabajar horas extras. El proceso de inspección implica principalmente la adquisición de la información, el procesamiento de esta información y la toma de decisión por parte del operador. Esta es la función de un sistema automático de reconocimiento de patrones. En consecuencia, los sistemas automáticos con habilidades como la de detectar y reconocer características de objetos, sigue siendo un desafío.

En la industria de manufactura electrónica se emplean sistemas automáticos de inspección para analizar la calidad de los productos. Estos sistemas automáticos de inspección generalmente incluyen las siguientes etapas:

- Etapa 1: Se genera un banco de los patrones de los objetos por reconocer.
- Etapa 2: Se lleva a cabo la adquisición de la imagen bajo prueba y el procesamiento, se

aplican filtros con el fin de eliminar ruido, segmentar objetos, etc.

- Etapa 3: Se extraen características que ayuden a identificar, clasificar y localizar el objeto de interés en una escena dada. Las características a analizar por el sistema dependen de las características del objeto, el sistema de adquisición de imágenes y de los algoritmos que se estén empleando.
- Etapa 4: Se clasifican los objetos de acuerdo a las normas de calidad. Se analiza si cada uno de los objetos identificados cumple con los criterios de calidad establecidos por la empresa para un producto final dado.

Este trabajo presenta la aplicación de algoritmos para el reconocimiento de patrones, usando filtros no lineales; su propósito intencional es identificar componentes mal colocados en una tarjeta electrónica PCBA.

Hay diferentes alternativas propuestas: Usando reconocimiento de patrones [2], [3], [4], [5], el cálculo de la calidad de objetos, usando un sistema de visión y técnicas de lógica difusa para el extracto de las características de objetos [6], [7], el análisis de calidad de madera basado en la información sobre el color y la textura [8], y la inspección de la madera basado en el color y textura con agrupación supervisada [9], [10]. Originalmente estos procesos sólo podrían asegurarse por los seres humanos, debido a la subjetividad del acercamiento de la evaluación de la calidad.

Este trabajo es organizado como sigue: La sección 2 describe el sistema automático de inspección, también presenta una descripción breve de la teoría de filtros acoplados, así como los resultados que se obtuvieron. La sección 3 discute algunas conclusiones y comparaciones.

II. MÉTODO

El arreglo experimental está conformado por tres etapas: adquisición de imágenes, análisis automático de las imágenes adquiridas y la clasificación, como se muestra en la Fig. 1. En la etapa de adquisición, la imagen es capturada por una cámara CCD. La imagen digital es después utilizada en la segunda etapa, donde aplicamos técnicas de reconocimiento de objetos a fin de extraer las características que se usan para identificar y localizar los componentes dentro de la tarjeta electrónica. La clasificación se lleva a cabo en la tercera etapa conforme a un conjunto de parámetros definidos tal como la posición.

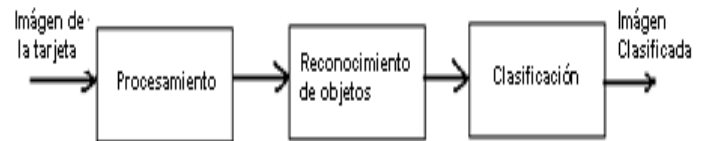


Fig.1 Sistema Automático de Inspección.

Los requisitos de aceptación para la manufactura de ensamblajes electrónicos están definidos por la norma IPC-A-610 emitido por “association connecting electronic industries, IPC” [11]). Este documento describe algunas características de las tarjetas de circuitos electrónicos relacionados a las condiciones deseadas, las cuales exceden las características mínimas aceptables, señaladas por la norma del producto final con la finalidad de ayudar a los evaluadores de procesos de la línea para determinar la necesidad de tomar acciones correctivas.

En los experimentos se usó una cámara, correspondiente a una cámara Sony Cybershot DSC-P93 CCD con sensor de 2592x1944 píxeles, 3V dc y lente con distancia focal de 38mm. Para el procesamiento de imágenes y la decisión se utilizó una PC Pentium IV, con 20Gb de espacio en disco duro, 128 Mb de RAM y monitor VGA con resolución de 1024 x 768 píxeles, vía una tarjeta de adquisición de imágenes. Se utilizaron dos decenas de imágenes de tarjetas. El sistema de iluminación consiste en una lámpara incandescente de 100

Watts, esta configuración produce una iluminación uniforme del objeto.

Existe gran variedad de algoritmos y arreglos ópticos que han sido propuestos con el fin de identificar patrones u objetos en imágenes. Entre los algoritmos propuestos destacan los filtros no lineales.

La salida del filtro no lineal está acoplada al objeto y su respuesta en frecuencia está dada por

$$NLF = |F(u,v)|^k e^{-j\phi(u,v)} \quad 0 < k < 1 \quad (1)$$

Donde:

$|F(u,v)|$ es el valor absoluto la Transformada de Fourier del objeto que deseamos reconocer, k es el valor que nos da la no linealidad en el filtro, $e^{j\phi}$ es la función exponencial j es el numero imaginario $\sqrt{-1}$, y ϕ es la fase de la transformada de Fourier

La métrica de desempeño de la *energía de pico-a-correlación* PCE se define como [12]

$$PCE = \frac{|E\{c(0,0)\}|^2}{E\{|c(x,y)|^2\}} \quad (2)$$

Donde el numerador es el valor esperado de la correlación del pico de intensidad y el denominador es el valor esperado de la energía media en el plano de correlación.

Las operaciones de correlación cruzada proporcionan características deseables en un sistema de reconocimiento de patrones, tales como, pico de correlación angosto, ruido de salida bajo, etc. La Fig.2 ilustra el esquema de operaciones computacionales para implementar el filtro no lineal.

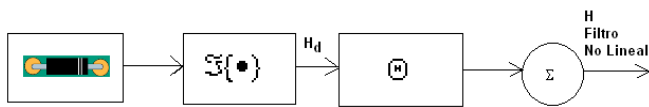


Fig.2 Diagrama a bloques para el filtro no lineal del objeto a identificar.

Un experimento ha sido diseñado para evaluar el desempeño del filtro no lineal considerando diferentes situaciones. En la sección 4.1 se diseñó un filtro acoplado capaz de detectar componentes diferentes y determinar cuál es su condición de aceptación o de defecto.

El caso que se analiza es determinar componentes de polaridad invertida. El conjunto de imágenes de entrenamiento se observan en la Fig. 3 y corresponden a un capacitor electrolítico, diodo y varios resistores del mismo tamaño. Para demostrar la factibilidad del filtro para identificar los componentes, lo aplicamos en las imágenes mostradas en la Fig.4. La Fig. 4a de estas imágenes corresponde a una sección de tarjeta PCBA que presenta a los componentes a identificar en condición de aceptable. La Fig. 4b corresponde a una sección de tarjeta PCBA que tiene un componente en la condición de defecto, es decir, un capacitor está mal orientado.

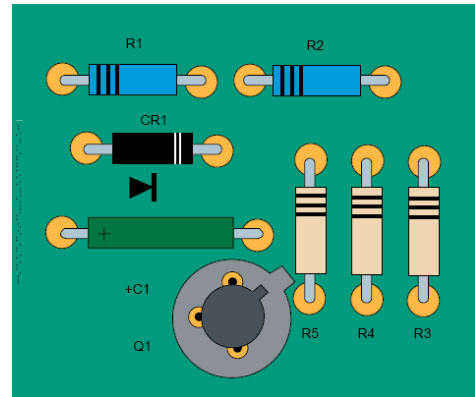


Fig. 3 Conjunto de componentes para diseñar el filtro no lineal, consistiendo en resistores, diodos, capacitores y multicontacto.

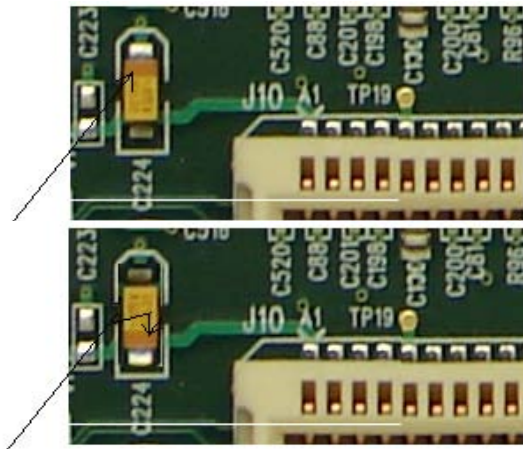


Fig. 4 Sección de PCBA: a) El componente está en condición de aceptación, b) El componente está en condición de defecto.

La correlación del filtro no lineal con cada una de estas imágenes se ilustra en la Fig. 5. Para el caso del PCBA mostrado en la Fig. 5a, observamos que la superficie de correlación presenta un máximo bien definido, cuyas coordenadas dentro del plano de correlación encontramos tres picos, el pico que corresponde al diodo en condición de aceptación tiene una mayor amplitud, de tal forma que podemos utilizar un valor umbral para discernir entre el pico que corresponde al elemento en condición de defecto y aquel en condición de aceptación. En la condición de defecto observamos varios picos casi de la misma amplitud, distinguiéndose así del caso de aceptación.

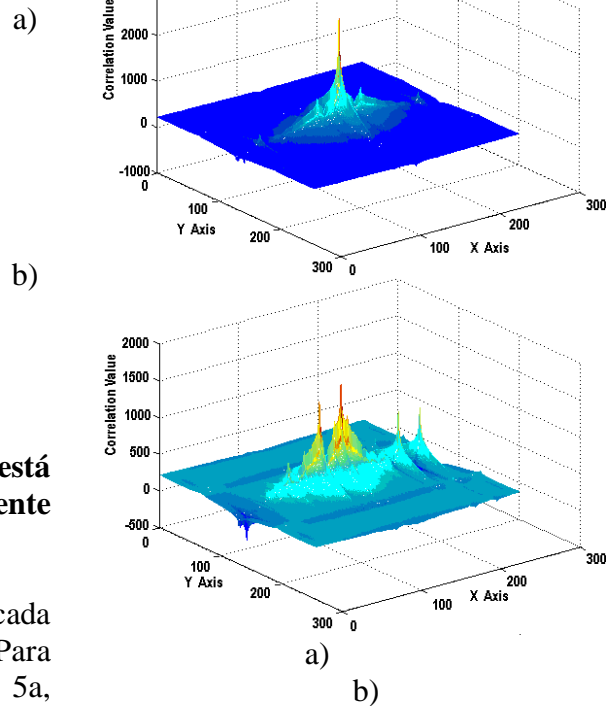


Fig.5 Salidas del filtro no lineal. A) Sección de PCBA en condición de aceptación, observamos el pico máximo que corresponde al componente aceptable, b) Sección de PCBA con el componente en condición de defecto, observamos varios picos casi de la misma amplitud.

Con este caso hemos mostrado la viabilidad del uso del filtro no lineal para identificar componentes en PCBA, encontrando que en el plano de correlación tenemos información suficiente para determinar la condición de aceptación o de defecto de los componentes identificados.

III. CONCLUSIONES

En este trabajo hemos revisado el concepto de filtros no lineales y los pasos para diseñar o entrenar este tipo de filtros, ya sea para reconocer objetos con características diferentes u objetos similares. El diseño de estos filtros se realiza a

partir de un conjunto de imágenes del objeto u objetos de interés.

La primera tarea consistió en encontrar un experimento estadístico numérico, el mejor valor del factor de intensidad de no linealidad k para el filtro no lineal. Los resultados se representan gráficamente, la energía pico a correlación (PCE) vs k , el resultado para el máximo valor de k es $k = 0.1$, mediante un experimento numérico. De la fig. 6, observamos que para los valores de $0.1 \leq k \leq 0.2$, el rendimiento del filtro no lineal es mejor para el filtro sólo de fase. Sin embargo, para valores de $0.3 \leq k \leq 1$, el rendimiento del filtro sólo de fase fue mejor que el para el filtro lineal y no lineal.

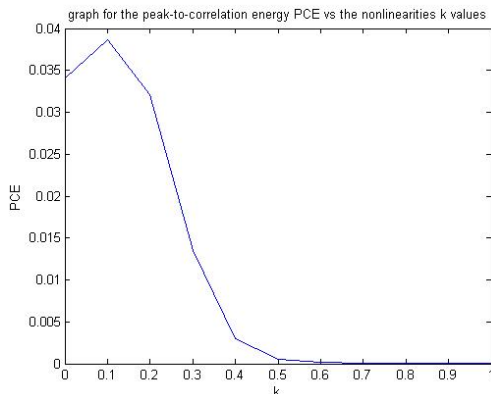


Fig.6 Gráfica para la energía de pico a correlación PCE vs los valores de k de no linealidades, desde $k=0$ hasta $k=1$ para la tarjeta electrónica de plantilla correlacionada con la tarjeta electrónica prueba.

En tarjetas de circuitos impresos, analizamos la robustez del filtro no lineal para discriminar entre componentes electrónicos en condición de aceptación y en condición de defecto. Demostramos que es factible identificar componentes electrónicos.

En resumen, concluimos que el filtro no lineal tiene capacidad para discriminar componentes en condición de aceptable de los que presentan condición de defecto. La optimización de la velocidad y costo de cómputo se logra al aplicar la Transformada Rápida de Fourier (FFT) conjuntamente con el filtro no-lineal. Una simple

resta de imágenes de tarjeta deseada menos tarjeta evaluada no sirve mucho para identificar alguna tarjeta porque la tarjeta a reconocer puede estar inmersa en ruido, puede tener diferente escala, puede estar rotada, puede estar en x posición, puede estar deformada, etc., y la resta no funcionaría. Hemos revisado bibliografía actualizada sobre inspección de tarjetas y circuitos electrónicos.

IV. REFERENCIAS

- [1] Josué Álvarez Borrego y María Cristina Chávez Sánchez. Introducción a la identificación automática de organismos y estructuras microscópicas y macroscópicas. Editorial Ediciones de la Noche, Universidad de Guadalajara, México. ISBN: 978-97027-1396-8. Hecho en México. pp 217-225.
- [2] González-F. Kober V. and Álvarez-B. Adaptive synthetic discriminant function filters for pattern recognition, *Optical Engineering* 45_5_, 057005 May 2006, pp 1-9.
- [3] Kober V., Mozerov M. and Álvarez-B. Pattern recognition based on rank correlations, *Proc. Spie's Optical Science and Technology 49th Annual Meeting 2-6 August in Denver Colorado USA vol. 5558*, 2006 pp 1-6.
- [4] Diaz-R, Kober V., and Álvarez-B. Pattern recognition with an adaptive joint transform correlator, *Vol. 45, No. 23 Applied Optics* 5929, 10 August 2006, pp 1-13.
- [5] Kober, V., Díaz-Ramírez V. H., González-Fraga, J. A. and Josué Álvarez-Borrego. Real time pattern recognition with adaptive correlation filters. In: G. Obinata and A. Dutta, "Vision systems, segmentation and pattern recognition", Chapter 27 Editorial Collegiums, Printed in Croatia., 2007, 515 p.
- [6] Kyllönen J., Pietikäinen M. Visual inspection of Parquet Slabs by Combining

color and Textures, FIN
<http://www2.automaatioseura.fi/jaostot/mv>
 n, 1995, pp 1-4.

- [7] Hirvonen J. Vision Pulp Analyzer, VTT Industrial Systems, Finland, <http://www.vtt.fi/tuo/indexe.html>, 2002, pp 1-4.
- [8] Gonzalez R. Digital Image Processing, Addison-Wesley, 1992, pp 628-631.
- [9] Skalak D. Prototype Selection for Composite Nearest Neighbor Classifiers, Thesis, University of Massachusetts, Amherst, MA 01003-46101997, 1997, pp 1-14.
- [10] Carew T. Automation of painted slate inspection, School of Electronic Engineering, Dublin City University, Ireland 2002, pp 9-29.
- [11] "IPC-A-610D Acceptability of Electronic Assemblies", Association Connecting Electronics Industries IPC(2005), Bannockburn, USA.
- [12] Coronel-Beltrán, A., Álvarez-Borrego, Josué, "Nonlinear filter for pattern recognition using the scale transform" in Applications of Digital Image Processing XXXI, edited by Andrew G. Tescher, Proceedings of SPIE Vol. 7073 (SPIE, Bellingham, WA 2008) 70732H.



MASI Marco Cedano-Olvera obtuvo su título de Maestría con especialidad en Sistemas Industriales en Febrero 15 de 2001 en el Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI) de la Universidad de Guadalajara. Actualmente es profesor e investigador en el Departamento de Computación de CUCEI.



MSC Eduardo Velázquez-Mora obtuvo su título de Maestría en Sistemas de Calidad con especialidad en Manufactura electrónica en Septiembre 07 de 2007 en el Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI) de la Universidad de Guadalajara. Actualmente es profesor e investigador en el Departamento de Electrónica de CUCEI.



Dora P. Hernández Roque es estudiante del Noveno Semestre de la Carrera de Ingeniero en Comunicaciones y Electrónica en el Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI) de la Universidad Guadalajara en el Departamento de Electrónica de CUCEI.

V. AUTORES



Dr. Jorge Hernández-Constante obtuvo su título de Doctor en Ingeniería y Tecnología con especialidad en Reconocimiento de Patrones en Noviembre de 2007 en el Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI) de la Universidad de Guadalajara. Actualmente es profesor e investigador en el Departamento de Electrónica de CUCEI.

IMPLEMENTACIÓN ÓPTIMA DE UN FILTRO ESPACIAL PARA MEJORAR EL DESEMPEÑO DE CLASIFICADORES DE SEÑALES CEREBRALES EN INTERFACES CEREBRO-COMPUTADORA

Escalona Diana, Gutiérrez David.

Centro de Investigación y de Estudios Avanzados del IPN-Unidad Monterrey
Parque de Investigación e Innovación Tecnológica (PIIT)
Autopista al Aeropuerto Km. 9.5, Lote 1, Manzana 29 Apodaca, N.L. Tel.: 52-81-1156-1740,
E-mail: {diana, dgtz}@hecate.mty.cinvestav.mx

RESUMEN.

En este trabajo presentamos una técnica de preprocesamiento para mejorar la clasificación de señales de Electroencefalografía (EEG) para aplicaciones de Interface-Cerebro Computadora (ICC) tomando en cuenta condiciones de operación reales, tales como baja relación señal a ruido (SNR, Signal to Noise Ratio), un número reducido de sensores de medición y número reducido de datos de entrenamiento en la etapa de clasificación. El método propuesto está basado en el diseño de un filtro espacial lineal de mínimo error cuadrático medio (LMMSE) que mejora la SNR de las señales de EEG antes de ser clasificadas. Los parámetros de dicho filtro son obtenidos a través de la versión optimizada del discriminante lineal de Fisher (DLF) tal que el área bajo la curva ROC (Receiver Operating Characteristics) sea máxima. Para comprobar la eficiencia del método se realizaron una serie de experimentos con datos simulados de EEG y se comparó el desempeño del clasificador que utiliza el filtro espacial contra el desempeño del clasificador de Mahalanobis bajo las distintas condiciones de operación planteadas. Los resultados numéricos muestran que el clasificador que utiliza el filtro espacial aumenta su desempeño para condiciones de SNR baja y el menor número posible de sensores de medición, mientras que el clasificador de Mahalanobis bajo las mismas condiciones de operación disminuye su desempeño. El método propuesto tiene el potencial de mejorar la eficiencia de los sistemas de ICC en condiciones reales con un costo computacional relativamente bajo.

Palabras clave: Electroencefalografía, interface cerebro-computadora, filtrado espacial,

discriminante lineal de Fisher, curvas ROC, clasificación de señales.

I. INTRODUCCIÓN

Una interface cerebro-computadora (ICC) puede definirse como un dispositivo de comunicación directa entre el cerebro y una computadora. Dicho dispositivo permite al usuario interactuar a través de sus pensamientos con un ambiente computarizado [1]. Una ICC tiene el potencial de ayudar a personas que sufrieron lesiones cervicales altas, parálisis cerebral, esclerosis múltiple o distrofia musculares, realizar alguna tarea a través de la comunicación de su cerebro y un dispositivo externo, por ejemplo, una silla de ruedas, prótesis o mover un cursor en una pantalla de computadora, etcétera.

En general, una ICC funciona de la siguiente manera: (i) Mediante técnicas como la electroencefalografía (EEG) o magnetoencefalografía (MEG), se captan señales eléctricas o magnéticas, respectivamente, asociadas a la actividad neuronal en la corteza cerebral; (ii) tales señales son interpretadas por una computadora previamente "entrenada" para mandar una orden al dispositivo externo.

En la etapa de procesamiento de las señales se contemplan métodos de filtrado, amplificación y compresión, con el fin de acondicionar las señales adquiridas a los niveles de amplitud y/o frecuencia deseados. Después, se extraen características específicas de las señales para identificarlas y relacionarlas de manera única. Algunas de las características más usadas en aplicaciones de ICC son parámetros espectrales, coeficientes de

modelos paramétricos, eigenvalores y eigenvectores, parámetros estadísticos. Todas estas características se relacionan con eventos fisiológicos específicos, tales como actividad sensorial o motriz, potenciales corticales, respuesta a actividades mentales o actividad de neuronas específicas. Sin embargo, la manera en que la computadora relaciona cualquier característica de la señal con algún evento fisiológico es por medio de un algoritmo de clasificación.

En los últimos años se han propuesto diversos clasificadores para su aplicación en ICC, muchos de los cuales reportan excelente desempeño en la clasificación de diferentes eventos motrices y cognitivos [2]. En la gran mayoría de estos estudios, el desempeño de la ICC se evalúa en términos de la velocidad y/o precisión en la clasificación de las señales cerebrales. Sin embargo, dicha evaluación del desempeño usualmente se lleva a cabo bajo condiciones de operación óptima y lejana a la realidad. Por ello, la evaluación de la ICC debe incluir pruebas bajo condiciones reales, tales como cuando se opera con un número limitado de mediciones, cuando se utiliza una base de datos reducida para el entrenamiento del clasificador o cuando se trabaja en presencia de ruido e interferencias.

En la práctica la cantidad de datos de entrenamiento y el número de sensores para las mediciones pueden ser ajustados con relativa facilidad, no así el ruido e interferencias debidas a distracción y/o estrés del paciente durante la medición de datos. Por ello, que se propone como alternativa un filtro espacial que ayude a disminuir el ruido e interferencias y por consiguiente que el desempeño del clasificador mejore.

II. MÉTODO

Se considera el caso en el que se tienen mediciones de actividad neuronal relacionada con algún evento motriz o cognitivo, tales mediciones provienen de un arreglo de m transductores de EEG, MEG o ambos. También se considera el caso en que las fuentes de actividad neuronal cambian su magnitud en el tiempo, pero mantienen su misma posición durante el periodo de las

mediciones. Esta suposición es válida en la práctica para experimentos de potenciales evocados, en donde una cierta actividad neuronal es provocada por estímulos visuales, audibles o sensitivos [3]. La estimulación se repite K veces con el fin de obtener K mediciones independientes de la actividad neuronal deseada.

Entonces, los datos de EEG/MEG son organizados en una matriz de m renglones para las muestras en el tiempo $t=1,2,\dots,N$. Así pues, la matriz con información espacio-temporal es de tamaño $m \times N$. Cuando las mediciones se repiten K veces, cada una de las mediciones independientes están definidas por el siguiente modelo:

$$Y_k = X + E_k, \text{ para } k = 1, 2, \dots, K, \quad (1)$$

donde Y_k es la matriz de mediciones para el experimento k , X es la matriz que contiene la señal deseada (evento neuronal) y E_k es la matriz de ruido con distribución probabilística Gaussiana con media cero y varianza σ_E^2 , no correlacionada en el tiempo ni en el espacio entre muestras.

Si además se considera el caso en el que se estudia un número I de eventos neuronales distintos, entonces las mediciones pueden ser organizadas en clases, tal que $Y_{k,i}$ corresponda a las mediciones obtenidas en el experimento k para el evento neuronal i , donde $i=1,2,\dots,I$ clases.

II.B. Clasificación.

El problema de clasificar datos de EEG/MEG consiste en que dada una medición Y , se debe identificar el evento neuronal i al que corresponde. Un ejemplo de clasificador es el de Mahalanobis, el cual está definido en términos de la distancia de Mahalanobis (d_M) para un grupo de datos Y de K mediciones independientes como:

$$d_M = |(Y - \bar{Y})^T R_{YY}^{-1} (Y - \bar{Y})|_F, \quad (2)$$

donde $\|\cdot\|_F$ es la norma de Frobenius, \bar{Y} es la media de los datos dado por:

$$\bar{Y} = \frac{1}{K} \sum_{k=1}^K Y_k, \quad (3)$$

y R_{YY} es estimada como:

$$\hat{R}_{YY} = \left(\frac{1}{K} \sum_{k=1}^K Y_k Y_k^T \right) - \bar{Y} \bar{Y}^T. \quad (4)$$

La clasificación de la matriz Y usando el clasificador de Mahalanobis sigue este procedimiento:

1. Para cada una de las i clases, le corresponden matrices $\{\bar{Y}, R_{YY}\}_i$ calculadas a partir de datos de entrenamiento;
2. Se calcula d_M usando (2) para cada clase $\{\bar{Y}, R_{YY}\}_i$ de la matriz Y ;
3. Y es asignado a la clase i para la cual d_M es mínima.

El clasificador de Mahalanobis ha sido utilizado en aplicaciones de ICC con técnicas de medición de EEG para diferentes procesos neuronales cognitivos o motores con una exactitud aceptable aún reduciendo el número de canales utilizados [4]. Sin embargo, esto es sólo viable para valores suficientemente grandes de K (es decir, cuando se estima una matriz de covarianza consistente) y alto SNR. Otro tipo de clasificadores no sólo requieren un valor grande de K , sino también un número grande de sensores m , resultando una gran carga computacional y un proceso de clasificación impráctico. Es evidente que un clasificador con un desempeño teórico del 100% no necesariamente trabajará de manera óptima a menos que durante su entrenamiento se hayan considerado factores reales que afecten su funcionamiento, tales como número reducido de transductores (m) para las mediciones, número reducido de datos de entrenamiento (K) y

baja SNR. Sin embargo, las variables m y K pueden ser ajustadas con relativa facilidad en aplicaciones reales, no así el SNR. Por ello, se propone un método basado en filtrado espacial para mejorar la SNR de la señal y, por ende, mejorar el desempeño del clasificador.

II.C. Filtrado espacial.

La matriz de covarianza de mediciones de EEG/MEG usualmente exhibe una estructura espacial considerable. Es entonces que, intentos recientes para mejorar el SNR de este tipo de mediciones se ha centrado en técnicas de filtrado espacial [5]. Un filtro espacial es aquel que permite amplificar señales que vienen de una dirección deseada, mientras suprime señales de otras direcciones. Este tipo de filtros han sido utilizados en diversas aplicaciones de arreglos de sensores, tales como radar y sismología [6]. El diseño del filtro propuesto está inspirado en el DLF para el caso especial de $I=2$ clases [7]. Esto es, considerando el modelo de datos dado por (1) y definiendo el error cuadrático medio (MSE, *mean square error*) entre la señal X y el filtro espacial F aplicado a los datos Y como:

$$e^2(F) = \text{tr}\{(X - FY)(X - FY)^T\}, \quad (5)$$

donde $\text{tr}\{\cdot\}$ denota la traza. El filtro de mínimo error cuadrático medio (*Linear Minimum Mean Square Error*, LMMSE) es [8]:

$$F_{LMMSE} = R_{XX} R_{YY}^{-1}, \quad (6)$$

donde R_{XX} y R_{YY} son la matriz de covarianza de la señal deseada y de los datos medidos, respectivamente. En la práctica R_{XX} no está disponible, por lo que se estimada a partir de los datos como:

$$\hat{R}_{XX} = \bar{Y} \bar{Y}^T. \quad (7)$$

donde \bar{Y} es la media dada por (3) y la matriz R_{YY} se estimada con (4).

Una implementación óptima del filtro espacial dado por (6) es sólo posible si se conoce la procedencia de la señal *a priori*, lo cual no es factible pues el filtro se aplica a las señales antes de identificar de donde provienen. Entonces, la mejor manera de conocer la procedencia de la señal es basarnos en probabilidades, es decir, posicionar nuestro filtro espacial hacia la dirección más probable donde encontraremos la señal deseada. Esta información podemos obtenerla del Discriminante Lineal de Fisher (DLF) que se basa en el cálculo del hiperplano óptimo cuya condición de optimización esta dada por el análisis de curvas ROC (*Receiver Operating Characteristics*).

Las curvas ROC son elementos que permiten observar, organizar y seleccionar discriminantes basados en su desempeño. Estas curvas han sido utilizadas en teoría de detección de señales para representar la proporción entre los aciertos y falsas alarmas de clasificadores. Su uso se ha extendido a la observación, análisis de comportamiento de sistemas de diagnóstico o la evaluación y comparación de algoritmos en máquinas de aprendizaje [9]. En este trabajo se utilizan para el análisis de discriminantes. La información que podemos asociar con el desempeño de discriminantes y las curvas ROC es el área bajo la curva, donde un área igual a 1 describe un discriminante perfecto y un área menor o igual a 0.5 es un discriminante inservible [10].

El hiperplano del DLF se calcula usando el vector normal definido, para el caso especial de $I = 2$ clases como:

$$V = (\gamma\{\hat{R}_{YY}\}_1 + (1-\gamma)\{\hat{R}_{YY}\}_2)^{-1} \cdot (\{\bar{Y}\}_2 - \{\bar{Y}\}_1), \quad (8)$$

donde $\{\hat{R}_{YY}\}_i$ y $\{\bar{Y}\}_i$ son, respectivamente, la matriz de covarianza y la media de la clase $i=1,2$, mientras γ es la constante de ajuste en el intervalo de $0 < \gamma < 1$. Cuando $\gamma = 0.5$ tenemos el DLF clásico.

Por otro lado, si variamos el valor de γ en la ecuación (8) en el intervalo $0 < \gamma < 1$ obtenemos distintos hiperplano, para lo que es necesario recurrir a algún criterio que nos proporcione la dirección más adecuada del filtro espacial. Se sabe que la mejor curva puede ser generada variando γ para cuando el área bajo la curva ROC sea máxima, y por tanto, el hiperplano sea óptimo [11]. Así, cada punto de la curva ROC estará asociado al óptimo valor de γ y a la dirección más probable que se puede obtener.

Es entonces que se propone aplicar una metodología similar para el caso de datos de EEG/MEG correspondientes a actividad contralateral (i.e., señales que provienen del hemisferio derecho o izquierdo del cerebro). Los parámetros obtenidos de la versión optimizada del DLF para calcular el filtro espacial es de la siguiente manera:

$$\hat{F} = \hat{C}_{XX} \hat{C}_{YY}^{-1}, \quad (9)$$

donde:

$$\hat{C}_{YY} = \gamma_{opt}\{\hat{R}_{YY}\}_1 + (1-\gamma_{opt})\{\hat{R}_{YY}\}_2, \quad (10)$$

y

$$\hat{C}_{XX} = (\{\bar{Y}\}_2 - \{\bar{Y}\}_1)(\{\bar{Y}\}_2 - \{\bar{Y}\}_1)^T. \quad (11)$$

El hiperplano óptimo para el caso $I = 2$ clases esta dado por:

$$V_{opt} = (\gamma_{opt}\{\hat{R}_{YY}\}_1 + (1-\gamma_{opt})\{\hat{R}_{YY}\}_2)^{-1} \cdot (\{\bar{Y}\}_2 - \{\bar{Y}\}_1), \quad (12)$$

Entonces el proceso de optimización tendrá los siguientes pasos:

1. Realizar la vectorización del modelo de mediciones dado por (1) para obtener un arreglo de $m \times N \times K$, con columnas correspondientes a mediciones independientes de cada evento conteniendo la información espacio-temporal de la actividad neuronal.

2. Calcular la matriz de covarianza y la medias a partir de los datos de entrenamiento vectorizados para cada clase $\{\bar{Y}, \hat{R}_{YY}\}_i$ utilizando las ecuaciones (4) y (7).
3. Calcular los parámetros óptimos del filtro :
 - 3.1. Calcular el hiperplano V dado por (8) utilizando las matrices de covarianza $\{\bar{Y}, \hat{R}_{YY}\}_i$ tomando un valor de γ en el intervalo $0 < \gamma < 1$;
 - 3.2. Usando V y γ , calcular:

$$v_0 = -\frac{\gamma\sigma_1^2 V^T \{\bar{Y}\}_2 + (1-\gamma)\sigma_2^2 V^T \{\bar{Y}\}_1}{\gamma\sigma_1^2 + (1-\gamma)\sigma_2^2}, \quad (13)$$
 usando

$$\sigma_i^2 = V^T \{\hat{R}_{YY}\}_i V, \quad (14)$$
 - 3.3. Calcular la función discriminante utilizando datos de prueba Y como:

$$D(Y) = V^T Y + v_0, \quad (15)$$
 - 3.4. Calcular la regla de discriminación (RD) de la siguiente forma:

$$RD = \begin{cases} Y \in P & \text{si } D(Y) > 0; \\ Y \in N & \text{si } D(Y) \leq 0. \end{cases} \quad (16)$$
 - 3.5. Calcular la curva ROC utilizando (16) y (15) para obtener los verdaderos positivos $VP = P(Y|p)$, los falsos positivos $FP = P(Y|n)$ y el área bajo la curva.
 - 3.6. Cambiar γ en todo el intervalo $0 < \gamma < 1$ repitiendo los pasos 3.1 a 3.5 hasta obtener los valores de γ_{opt} tal que el área bajo la curva ROC sea máxima para cada dato de prueba;
 - 3.7. Utilizar γ_{opt} para calcular \hat{C}_{YY} y \hat{C}_{XX} usando (10) y (11) respectivamente y sustituir en (9) para calcular el filtro óptimo.
4. Aplicar el filtro \hat{F} al elemento a clasificar, es decir, $\hat{Y} = \hat{F}Y$;

5. Alimentar \hat{Y} al clasificador y evaluar su desempeño con las curvas ROC.

II.D. Ejemplos Numéricos.

Para generar mediciones simuladas de EEG se utiliza el método de solución directa de un modelo esférico de cuatro capas descrito en [12] donde la actividad cortical es modelada como un dipolo de corriente con una magnitud que cambia en el tiempo, aproximándose a una típica señal de respuesta evocada. Este modelo es similar al usado previamente en [13]. La posición del dipolo se asume fijo sobre el área de la corteza motora en el hemisferio derecho o izquierdo del cerebro durante las mediciones. Entonces, las matrices de mediciones de EEG se realizan para $m=6$ y $m=16$ sensores distribuidos sobre el área de activación para una esfera de radio 10.5 cm como se muestra en la Figura 1. A los datos generados se les añade ruido aleatorio con distribución $N(0, \sigma^2)$ no correlacionado en tiempo y espacio, con diferentes valores de σ hasta obtener valores de SNR de -3, -6,-8 dB (i.e., condiciones de ruido bajo, medio y alto). Algunos ejemplos de la distribución del evento de la señal cerebral se muestran en la Figura 2.

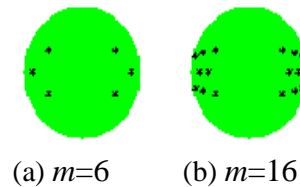


Figura 1: Distribución de sensores de medición.

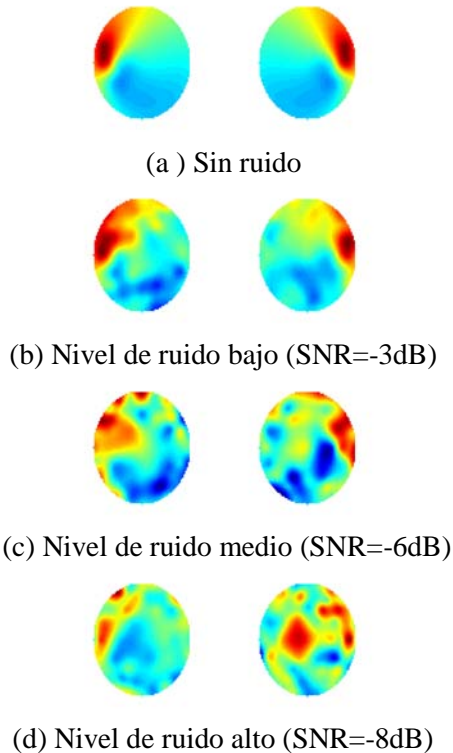


Figura 2: Datos simulados de EEG de actividad contra-lateral con diferentes niveles de ruido.

Los datos de EEG fueron alimentados al filtro espacial donde \tilde{Y} corresponde a la redistribución espacial de los datos originales medidos Y . Sin embargo, ésta redistribución espacial no resulta un efecto indeseado, sino resulta ventajosa cuando se tiene niveles de ruido alto y en consecuencia la discriminación de \tilde{Y} pueda ser fácilmente realizada. Un ejemplo de la redistribución de los datos se observa en la Figura 3, donde un sólo dato de prueba es aplicado al filtro espacial.

Así mismo se realizaron pruebas para evaluar el desempeño del clasificador alimentado por el filtro espacial, en comparación con el desempeño del clasificador de Mahalanobis para diferentes valores de m , K y SNR. De las 1300 mediciones simuladas de EEG se utilizaron 100 eventos destinados para la etapa de clasificación (datos de prueba) y K eventos usados como datos de entrenamiento, con $K = 15-600$.

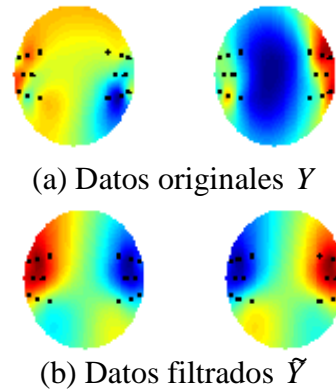


Figura 3: Redistribución espacial para datos simulados de EEG.

Los datos crudos fueron alimentados al clasificador de Mahalanobis y se calcularon las curvas ROC para evaluar su desempeño. Los datos filtrados fueron alimentados a un clasificador basado en la detección del pico positivo sobre la superficie de redistribución del potencial e igualmente se calcularon las curvas ROC para evaluar el desempeño y compararlo con el clasificador de Mahalanobis. Los datos de prueba, en ambos casos, fueron aleatoriamente ordenados. Así, todas las posibles combinaciones de m , K , SNR, previamente definidas, fueron analizadas con el clasificador de Mahalanobis y el clasificador alimentado con el filtro espacial. En la Figura 4 se muestran las curvas ROC para el caso especial donde $m = 16$, $K = 250$ y tres diferentes valores de SNR. Tales valores de m y K son un ejemplo correspondiente a valores típicos observados en condiciones de operación reales y nos ayuda a observar en el procedimiento de clasificación el efecto producido por los niveles de ruido. Claramente, la clasificación tras utilizar el filtro espacial aumenta el desempeño en comparación con el clasificador de Mahalanobis. En efecto, el clasificador de Mahalanobis tiene un desempeño bajo para el número de datos de entrenamiento usados en este ejemplo. Es bien sabido que el clasificador de Mahalanobis funciona de manera óptima cuando K es lo suficientemente grande. Sin embargo, esto no es factible en condiciones de operación real, ya que se requiere que el paciente esté sujeto a sesiones largas de adquisición de los

datos lo cual tiende a aumentar los niveles de ruido durante las mediciones debidas a estrés o distracción.

Por otro lado, el proceso de clasificación es un proceso aleatorio con una distribución de probabilidad de valor extremo [14] que puede ser caracterizada a partir de su media y desviación estándar, para lo cual es necesario repetir el experimento múltiples veces. Así, se calcularon las áreas bajo la curva ROC para 100 veces el proceso de clasificación cambiando aleatoriamente los datos, para valores de m , K y SNR definidos previamente. Los valores de media y desviación estándar son mostrados en las Figuras 5, 6 y 7 (las barras de error muestran la desviación estándar a partir de la media), donde para un nivel de ruido bajo (Figura 5) el clasificador basado en filtrado espacial está por arriba de la media del desempeño del clasificador de Mahalanobis, el cual para el valor de $K = 250$ funciona en un 100% al igual que el clasificador basado en filtrado espacial, con la diferencia de que este último tiene un desempeño entre el 70% (Figura 5b) y 80% (Figura 5a) para $K = 150$. Si aumentamos el ruido a un nivel medio (Figura 6), se observa que el desempeño del clasificador utilizando el filtro está de nuevo por arriba de la media del clasificador de Mahalanobis el cual para $K = 200$ baja su desempeño no así el clasificador basado en filtrado espacial que tiene un desempeño del 80% (Figura 6a) para $K = 150$. Si aumentamos aún más el nivel de ruido observamos la misma tendencia (Figura 7), el clasificador de Mahalanobis tiene un desempeño bajo y el clasificador basado en filtrado por arriba del 80% (Figura 7a) para $K = 200$. Sin embargo, es posible observar que el clasificador basado en filtro espacial también decae su desempeño para ciertos valores de K (Figuras 5b, 6b, 7b), esto es debido a que la matriz \hat{Y} tiene más variables y menos información de la etapa de entrenamiento y es de esperarse que el desempeño del clasificador decaiga cuando tenemos valores de $m = 16$, aún así bajo estas mismas condiciones observamos que el clasificador de Mahalanobis está por debajo de la media del clasificador basado en el filtro espacial.

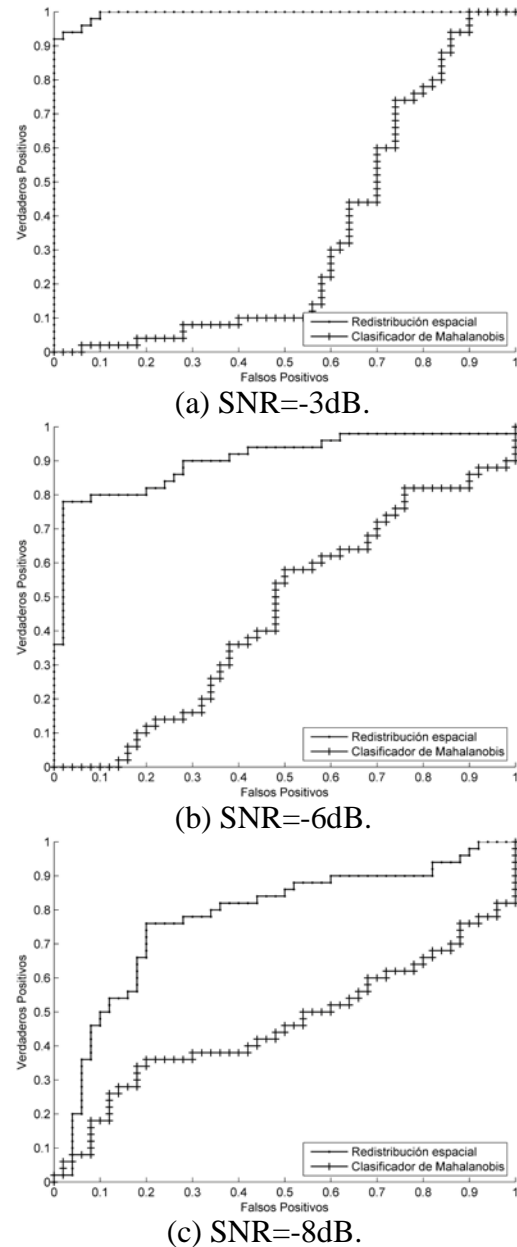


Figura 4: Curvas ROC para evaluar el desempeño del clasificador basado en filtro espacial en comparación con el de Mahalanobis.

Entonces, nuestros resultados muestran que para el caso de los datos filtrados, el desempeño de la media del clasificador está arriba del 80% de correctas clasificaciones cuando se acerca a $K = 250$ sólo usando 6 sensores durante la medición y con un nivel de ruido alto. Por otro lado, la variabilidad del proceso de clasificación es baja. Bajo estas mismas condiciones el clasificador de Mahalanobis tiene resultados desfavorables.

Los experimentos muestran un buen desempeño en el proceso de clasificación usando el menor número posible de sensores y datos de entrenamiento tras utilizar el filtro espacial. Esto comprueba la aplicabilidad de dicho método bajo condiciones de operación real. Por otro lado, con anterioridad se ha demostrado que métodos neuronales no han mostrado grandes ventajas en comparación a métodos lineales y cuadráticos, alcanzando sólo mejoras de alrededor del 6% en el número de clasificaciones correctas, pero incrementando considerablemente la complejidad computacional de la implementación [15], por lo que el método aquí propuesto promete poca complejidad computacional y un buen desempeño del clasificador bajo condiciones de operación real que podría llevar a la práctica a los sistemas ICC.

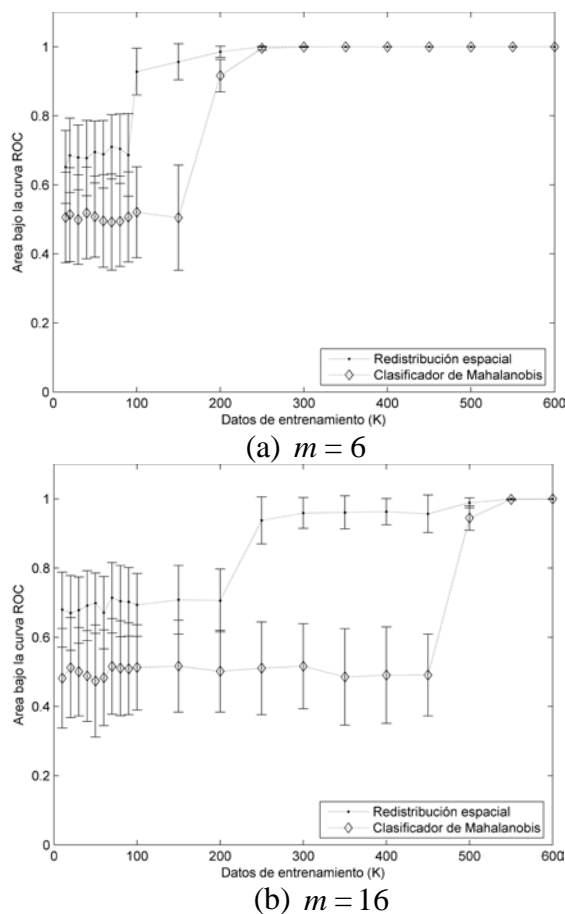


Figura 5: Análisis de medias y desviación estándar del clasificador basado en filtrado espacial y de Mahalanobis para ruido bajo.

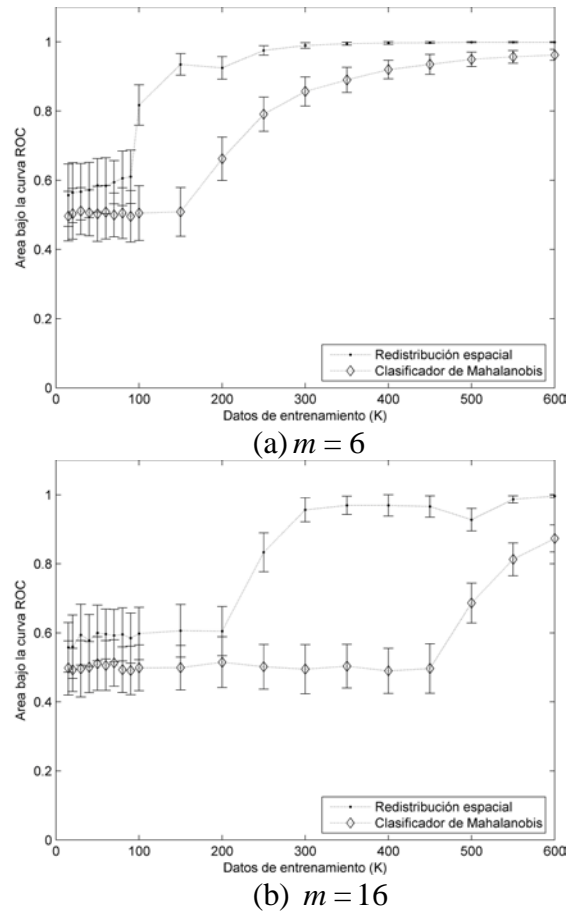


Figura 6: Análisis de medias y desviación estándar del clasificador basado en filtrado espacial y de Mahalanobis para nivel de ruido medio.

III. CONCLUSIONES

El método propuesto basado en filtrado espacial para datos de EEG de actividad contralateral con parámetros obtenidos por la optimización del DLF por análisis de curvas ROC incrementa la SNR. El filtro espacial aplicado a datos simulados de EEG favorece la etapa de clasificación. Por otro lado, el método propuesto tiene una potencial aplicación para mejorar el desempeño de clasificadores en condiciones de operación real de los sistemas ICC, es decir, consideraciones como bajo SNR, número reducido de datos de entrenamiento y el menor número posible de sensores de medición. Los trabajos a futuro estarán encaminados en términos de la implementación computacional para seleccionar el óptimo valor de γ y en generalizar el método para la clasificación de eventos multiclase o señales

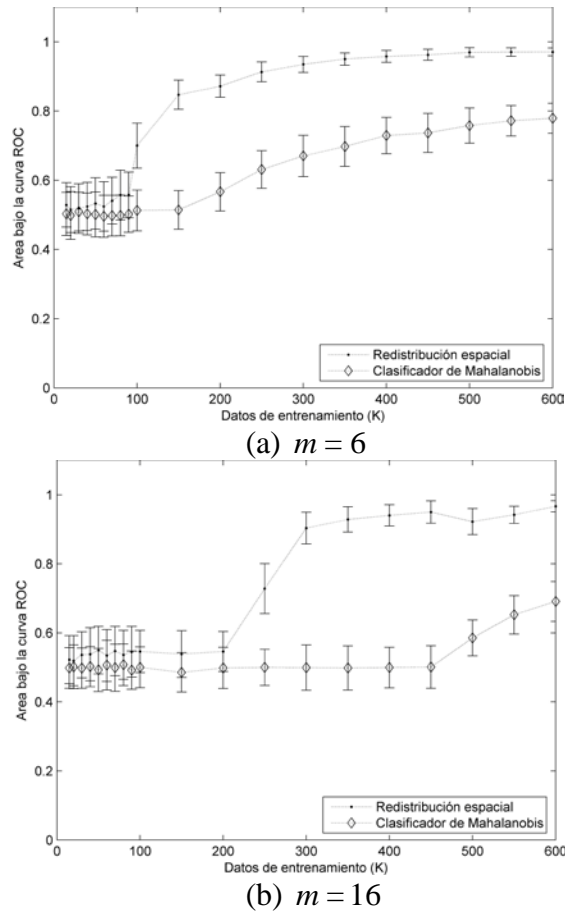


Figura 7: Análisis de medias y desviación estándar del clasificador basado en filtrado espacial y de Mahalanobis para nivel de ruido alto .

más complejas, así como aplicar el mismo método utilizando datos de magnetoencefalografía (MEG) para extraer mejores características de la señal y utilizarlas en la etapa de entrenamiento de los sistemas ICC.

IV. REFERENCIAS

[1] J.R Wolpaw, N. Birbaumer, D, J. Mcfarland, G. Pfurtscheller y T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767-791, 2002.

[2] A. Bashashati, M. Fatourehchi, R. K. Ward y G.E. Birch, "A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals," *Journal of*

Neural Engineering, vol.4, pp.R32-R57, 2007.

[3] J. W Rohrbaugh, R. Parasuraman y R Johnson, *Event-related brain potentials: basic issues and applications*, Oxford University Press, Nueva York, 1990.

[4] F. Babiloni, L. Bianchi, F. Semeraro, J. R. Millán, J. Mouriño, A. Cattini, S. Salinari, M. G. Marciani y F. Cincotti, "Mahalanobis distance-based classifiers are able to recognize EEG patterns by using few EEG electrodes," en *Proceedings of the 23th Annual Conference of the IEEE Engineering in Medicine and Biology Society*, Estambul, Turquía, 2001, pp.651-654.

[5] K. B. Yu y C.D. Mc Gillem, "Optimum filters for estimating evoked potential waveforms," *IEEE Transactions on Biomedical Engineering*, vol. 30, No. 11, pp.730-737, 1983.

[6] B. D van Veen y K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Signal Processing Magazine*, vol.5, pp. 4-24, 1988.

[7] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol.7, Parte II, pp. 179-188, 1936.

[8] S. J. Orfanidis, *Optimum signal processing*, an introduction. New York: MacMillan Publishers, 1985.

[9] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters Elsevier*, vol. 27, pp. 861-874, 2006.

[10] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition Society*, vol.30, No. 7, pp. 1145-1159, 1997.

[11] T. Cooke y M. Peake, "The optimal classification using a linear discriminant for two point classes having known mean and covariance," *Journal of Multivariate Analysis*, vol. 82, pp. 379-394, 2002.

[12] J. C. Mosher, R. Leahy, y P. S. Lewis, "EEG and MEG: forward solutions for inverse methods," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 3, pp. 245-259, 1999.

- [13] D. Gutiérrez, A. Nehorai, y C. Muravchik, "Estimating brain conductivities and dipoles source signals with EEG arrays," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 12, pp. 2113-2122, 2004.
- [14] D. Gutiérrez, "Designing a spatial filter to improve SNR in two-class discrimination problems for BCI applications," en *Proceedings of the 42nd Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, 2008, pp. 372-377.
- [15] D. Garrett, D. A. Peterson, C. W. Anderson y M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 141-144, Junio 2003.

becario del Programa Fulbright-García Robles durante sus estudios en UIC y desde 2005 es miembro del Sistema Nacional de Investigadores de CONACYT.

V. AUTORES

Diana Irazú Escalona Vargas obtuvo los títulos de Ingeniero en Biónica por la Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas (UPIITA) del Instituto Politécnico Nacional y de Maestría en Ingeniería y Física Biomédicas por el Centro de Investigaciones y Estudios Avanzados del IPN (Cinvestav) Unidad Monterrey en 2006 y 2009, respectivamente. Sus intereses están encamidos en el procesamiento de señales biomédicas en el área de las neurociencias y en el procesamiento digital de imágenes.

David Gutiérrez recibió el título en Ingeniería Mecánica Eléctrica (con mención honorífica) por la Universidad Nacional Autónoma de México (UNAM) en 1997, y los títulos de Maestría en Ingeniería Electrónica y Ciencias de la Computación y de Doctorado en Bioingeniería en 2000 y 2005, respectivamente, ambos por la Universidad de Illinois en Chicago (UIC). Entre 2005 y 2006 realizó una estancia posdoctoral en el Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS), UNAM. Desde Junio de 2006 es Investigador Adjunto en el Centro de Investigación y de Estudios Avanzados (Cinvestav), Unidad Monterrey. Sus líneas de investigación son en el área de procesamiento estadístico de señales médicas. También está interesado en imagenología médica, neurociencias y algoritmos de procesamiento en tiempo real. El Dr. Gutiérrez fue

APLICACIÓN DE DIFERENTES TÉCNICAS PARA EL RECONOCIMIENTO DE PATRONES DE TRAYECTORIAS EN UN ESPACIO EUCLIDIANO CON UN *WiiMote*

Lee Cosío Blanca Miriam¹, Delgado Mata Carlos²

Escuela de Ingeniería

Universidad Panamericana, Campus Bonaterra

Josemaría Escrivá de Balaguer No 101 Aguascalientes, Aguascalientes 20290, México

Tel. 910 6216, correo electrónico: ¹miriam.lee@up.edu.mx, ²cdelgado@up.edu.mx

RESUMEN

Este artículo presenta el trabajo en desarrollo para la aplicación de Modelos Ocultos de Markov (HMM¹) y Redes Neuronales Artificiales (ANN²) para el reconocimiento de patrones de una trayectoria en un espacio euclidiano a partir de una muestra de datos filtrada y parametrizada por la Transformada Rápida de Fourier (FFT³). La información proviene del movimiento de un dispositivo basado en acelerómetros utilizado por la consola Wii de Nintendo© (también conocido como 'WiiMote').

Los conceptos, metodología, evaluación y conclusiones se presentan y discuten en este artículo.

Palabras clave: Reconocimiento de patrones, Modelos Ocultos de Markov (HMM), Redes Neuronales Artificiales (ANN), Transformada Rápida de Fourier (FFT), Wii.

I. INTRODUCCIÓN

El ámbito del reconocimiento de patrones ha sido ampliamente estudiado con diferentes técnicas y aplicaciones.

Desde la creación de los dispositivos de interfaz con la computadora (HID⁴), se ha enfocado la investigación en el área de reconocimiento de patrones y se han abierto nuevos campos de aplicación. Es así que en la actualidad existen

diversos dispositivos que son capaces de proporcionar información relevante respecto a diferentes fenómenos de interés, sin embargo cada uno entrega la información codificada y por tanto se requiere de un algoritmo de procesamiento para cada uno.

Las aplicaciones más comunes de reconocimiento de patrones han sido para el procesamiento de voz y escritura. Investigaciones previas de *Bell Laboratories* [1] e *IBM Research* [2] han aplicado estos métodos para el reconocimiento de texto escrito a mano. En éstos, los métodos probabilísticos y estadísticos fueron los más utilizados por sus buenos resultados.

Trabajos similares se han desarrollado para la evaluación y caracterización de trayectorias y movimientos a partir de la información obtenida por acelerómetros. Proyectos como *WiiGee* [3], *GesApp*, parte del *Google Summer of Code 2008* [4] y el de *PREKOPCSÁK* [5] han investigado el reconocimiento de patrones de este tipo.

En particular, técnicas de filtrado con la Transformada Rápida de Fourier (FFT) y algoritmos de K-Medios [3] se han utilizado para la evaluación de información. Paralelamente, Modelos Ocultos de Markov (HMM) [4] [6] y las Redes Neuronales Artificiales (ANN) [6] [7] [8], y algoritmos de clasificación con redes bayesianas [4] se han utilizado para el entrenamiento y caracterización de modelos.

En este artículo, se analizarán diferentes técnicas de reconocimiento de patrones para la

¹ Hidden Markov Models, por sus siglas en inglés

² Artificial Neural Networks, por sus siglas en inglés

³ Fast Fourier Transform, por sus siglas en inglés

⁴ Human Interface Device, por sus siglas en inglés

clasificación y caracterización de patrones de movimiento y trayectorias.

II. CONCEPTOS

A la información obtenida a partir del movimiento de un HID se le conoce en la Inteligencia Artificial (AI⁵) como gesto. En particular, un gesto derivado de un dispositivo *WiiMote* está caracterizado por un vector de información que contiene los valores de la aceleración del movimiento o trayectoria realizados [3].

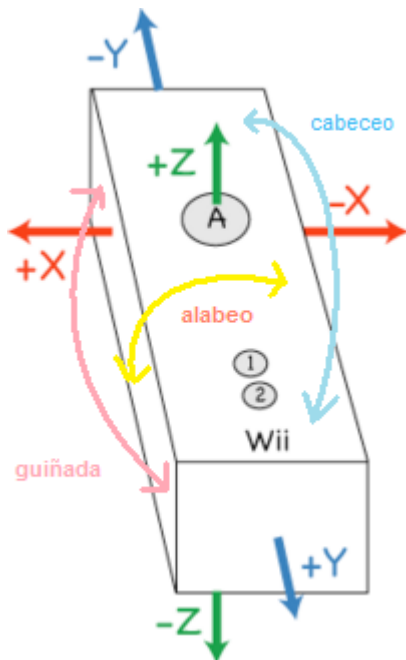


Figura 1 – WiiMote y sus grados de libertad [9]

El *WiiMote* es un HID diseñado para la medición de aceleraciones en un espacio euclidiano. Tiene seis grados de libertad; es decir, se puede mover a lo largo de los ejes X, Y, Z. Igualmente es capaz de rotar sobre estos mismos ejes, lo cual también se conoce como cabeceo (*pitch*), alabeo (*roll*), y guiñada (*yaw*) (Figura 1).

Para su comunicación con otros dispositivos, el *WiiMote* cuenta con un Bluetooth integrado Broadcom BCM2042 y utiliza el protocolo estándar Bluetooth HID. [10]

El *WiiMote* contiene un ADXL330, que es un acelerómetro lineal de tres ejes y fabricado por *Analog Devices*®. Tiene la capacidad de medir aceleraciones en el rango de +/- 3g con 10% de resolución. Esta característica del control es la que nos permitirá obtener la información de los gestos realizados.

El *WiiMote* tiene una alta frecuencia de muestreo, pues puede entregar hasta 100 valores por segundo para cada una de la aceleraciones en X, Y, Z. [4]

Si el *WiiMote* estuviera en caída libre con el eje Z perpendicular al piso, devolvería valores muy cercanos a 0 para la aceleración en Z. Por el contrario, si se mantuviera el *WiiMote* en reposo con la misma orientación, entregaría en este eje un valor positivo equivalente a la fuerza de gravedad g (aproximadamente 9.8 m/s²). En la dirección contraía este valor sería negativo. A partir de esta relación se puede conocer la orientación del *WiiMote*, siempre y cuando esté razonablemente quieto [9].

III. METODOLOGÍA

La metodología a seguir para el reconocimiento de gestos en el *WiiMote* es el que se muestra en la Figura 2.

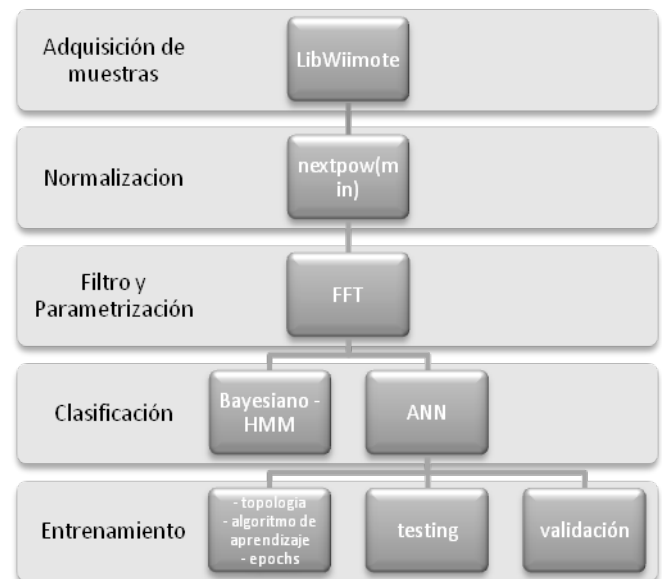


Figura 2 – Proceso para reconocimiento de gestos.

⁵ Artificial Intelligence, por sus siglas en inglés

Adquisición de muestras

Para la adquisición de muestras se utilizó la librería *LibWiiMote* [11], desarrollada en el lenguaje de programación C, que provee una interfaz para la comunicación con el *WiiMote* en un sistema operativo Linux.

En condiciones normales de operación, el *WiiMote* entrega valores de aceleración en X, Y, Z cada vez que le son solicitados (Tabla 1):

0.222	-0.154	0.880	0
0.259	-0.077	1.000	10
0.333	-0.154	1.040	20
0.370	-0.231	1.040	30
0.370	-0.308	1.080	40

Tabla 1 – Valores de aceleración X, Y, Z, ms

En este caso las primeras tres columnas corresponden a los valores de las aceleraciones en X, Y, Z y la última columna, es el tiempo en milisegundos (ms) en el cual se realizó la captura.

Normalización

Este proceso estandariza las muestras a una cantidad de valores predefinido para cada uno de los movimientos. En este caso, se utilizaron técnicas de normalización logarítmicas, polinomiales y lineales, siendo ésta última la seleccionada por su sencillez y velocidad de procesamiento.

Así se obtuvo una base de datos estable, que sin importar el número de entradas, siempre tuviera el mismo número de salidas.

Filtrado y parametrización – Transformada de Fourier

Debido a que el acelerómetro devuelve múltiples valores en un periodo de tiempo muy corto, y debido a que no todas las muestras tendrán la misma duración, y por tanto, la misma cantidad de datos, se requiere de un filtro y un parametrizador. Estas salidas son la entrada del modelo Markoviano y de la Red Neuronal, que previamente se ha estructurado para el número

fijo de entradas, que se obtienen de la salida del parametrizador.

Una forma muy sencilla de definir un movimiento (determinado por valores en X, Y, Z) sería obtener el máximo, mínimo, y media de cada eje, obteniendo lo siguiente [12]:

$$X_{\min}, X_{\max}, X_{\text{med}}, Y_{\min}, Y_{\max}, Y_{\text{med}}, Z_{\min}, Z_{\max}, Z_{\text{med}}$$

Sin embargo este filtro es insuficiente pues tiene un rango de error muy alto al momento de hacer la clasificación.

Mientras que en [3] se utiliza un algoritmo de K-medios, aquí se optó por utilizar la Transformada de Fourier para evaluar la función resultante de cada movimiento, y obtener el espectro de Fourier característico.

La transformada de Fourier es una señal continua de entrada $x(t)$ definida por (Ecuación 1):

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt, \quad \omega \in (-\infty, \infty)$$

Ecuación 1

Por su parte, la Transformada Discreta de Fourier (TDF) reemplaza la integral infinita por una sumatoria (Ecuación 2):

$$X(\omega_k) \triangleq \sum_{n=0}^{N-1} x(t_n)e^{-j\omega_k t_n}, \quad k = 0, 1, 2, \dots, N-1$$

Ecuación 2

En otras palabras, significa que la TDF transforma, como su nombre lo dice, cualquier señal de entrada en una sumatoria infinita de senos y cosenos a diferentes frecuencias.

Lo que se obtiene de la transformada son las diversas frecuencias de la función original respecto a la amplitud. Las primeras armónicas del espectro de Fourier son las predominantes de la función y determinan en gran medida la esencia de la función.

Clasificación

Para la clasificación de las muestras se aplicarán los siguientes métodos:

a) Modelos Ocultos de Markov

Los Sistemas actuales de reconocimiento de voz y movimientos, entre otros, utilizan Modelos Ocultos de Markov (HMM). En algunos sistemas de reconocimiento de voz, los estados en un HMM representan las unidades fonéticas. Un estado de transición define la probabilidad de que ocurra el siguiente estado (Figura 3).

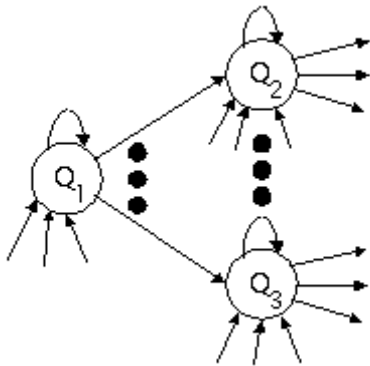


Figura 3 - Una representación simple de un HMM [8].

El término oculto o discreto se refiere al tipo de Modelo Markoviano en el cual las observaciones son una distribución de probabilidad del estado actual. El reto es determinar los parámetros ocultos de los parámetros observables. Un HMM se puede considerar como la red bayesiana dinámica más simple. Una especificación completa de un HMM requiere de la siguiente información[13]:

La evaluación de la probabilidad (o verosimilitud) de una secuencia de observaciones

La determinación de la mejor secuencia de estados del modelo

El ajuste de los parámetros del modelo que mejor se ajusten a los valores observados

En el caso del reconocimiento de voz, por ejemplo, un HMM se crea para cada palabra en un vocabulario. Una de las tareas es medir la secuencia de unidades fonéticas y determinar cuál HMM es capaz de generar la misma secuencia. [14] [15].

En el trabajo presentado en este artículo, un HMM se crea por gesto del *WiiMote* a evaluar.

b) Redes Neuronales Artificiales

Otro método que se está explorando es de las Redes Neuronales Artificiales (ANN). Para entender la implementación de la red neuronal que se está haciendo, a continuación se explica brevemente cómo funcionan las ANN.

Como su nombre lo dice, las ANN que simulan las redes neuronales que existen en el cerebro humano. A pesar de que las ANN no son inteligentes (en un sentido estricto), si pueden reconocer patrones y crear reglas simples para resolver problemas complejos. Sus capacidades de entrenamiento y aprendizaje también son muy amplias, por lo cual son utilizadas mucho en investigación de inteligencia artificial, como lo presentado en este artículo. Estas reglas sencillas le permiten a las ANN generalizar y por lo tanto clasificar después de un entrenamiento con varias muestras similares. [16]

La definición matemática de una ANN en general es la siguiente (Ecuación 3):

$$y(x) = g \left(\sum_{i=0}^n w_i x_i \right)$$

Ecuación 3

Donde x es la neurona con n entradas ($x_0...x_n$) y una neurona de salida $y(x)$, y donde ($w_0...w_n$) son los pesos. La función de activación, g , es la función que mide que tan alta debe ser la salida de la neurona, basada en la sumatoria de la entradas por sus pesos. La función de activación puede ser una simple función escalón, que devuelve 1 o 0.

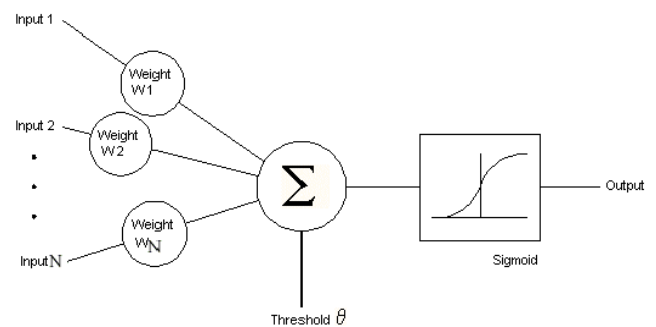


Figura 4 – Perceptrón de una capa [17]

La Figura 4 muestra un perceptrón, es decir, una ANN de una sola capa. Generalmente las ANN se distribuyen en capa de entrada, capa oculta, y capa de salida. Sin embargo, puede haber más de una capa oculta.

Las capas de entrada y de salida significan los valores de entrada y salida que queremos aproximar. Los pesos entre las capas ocultas determinan lo bien o mal que se desempeña la ANN. Estos pesos son los que se van ajustando en cada iteración del entrenamiento para satisfacer los valores de entrada.

Con esta base de información se procede a estructurar el algoritmo que manipula la red neuronal. Para esto, se hace uso de la librería para redes neuronales llamada FANN [18].

Esta librería nos permite configurar la red neuronal desde su creación, de tal manera que podamos ir adaptando sus características como topología (capas, entradas, y neuronas en las capas ocultas), algoritmo de entrenamiento, función de activación y su pendiente en las capas ocultas y la de salida, interconectividad, aprendizaje, etc.

Por el tipo y cantidad de información que se obtiene de las muestras, se optó por una topología de tres capas, con seis neuronas en la capa oculta.

También se debe considerar el generar varios conjuntos de muestras, unos de entrenamiento, otros de validación y otros de prueba. De esta forma la red tendrá suficiente información sobre la cual trabajar. De la misma forma que se trabajó con la información de inicio, se manipulará esta información para tener como resultado las entradas a la red neuronal.

Entrenamiento

Para el mejor desempeño durante el entrenamiento, se pueden ir alterando los pesos y el nivel de aprendizaje en cada iteración manualmente. Para esto se debe hacer un ciclo (epoch) y en cada epoch se pregunta el error de aprendizaje, es decir, la diferencia entre la salida real y la salida deseada, hasta llegar ya sea al error mínimo deseado, o al número de epochs

máximo. Como parte del entrenamiento podemos corroborar que la red neuronal elegida sea la ideal haciendo pruebas con información diferente a la información de entrenamiento, es decir, la información de prueba que capturamos y procesamos anteriormente. De esta forma también la ANN se da una idea de que tan específica o generalizadamente se han alterado sus pesos.

Como salida de este entrenamiento, se obtienen múltiples archivos .net que corresponden a la topología de la red neuronal en cada epoch. Una vez obtenidos los pesos ideales de la ANN, se continúa con

IV. PRUEBAS Y RESULTADOS

Una trayectoria de un movimiento circular se muestra en figura 5.

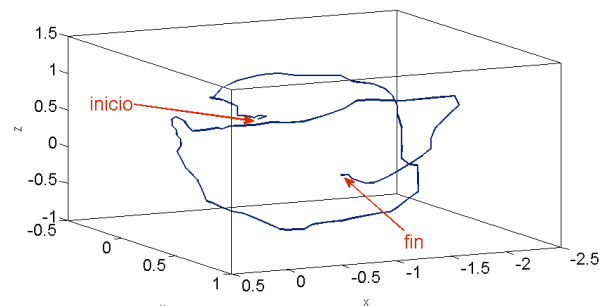


Figura 5 – Trayectoria de movimiento circular tomada con WiiMote

Se puede apreciar el punto del inicio y el final del movimiento, al igual que su dirección y trayectoria. Sin embargo se observa que la imagen a primera vista es poco comprensible, pues incluye cambios bruscos de concavidad, ruido y no linealidades.

Es por esto que la FFT ayuda a extraer las características esenciales de la función. La parametrización con Fourier se hace por separado para cada uno de las aceleraciones de los ejes X, Y, Z, como se muestra en la figura 6.

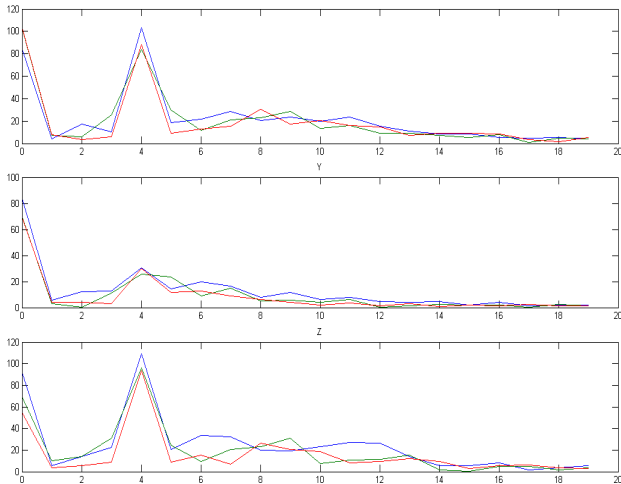


Figura 6 - Espectro de Fourier de una muestra de una trayectoria tomada

Nótese que cada movimiento tiene en el espectro de Fourier un valor medio en la intersección de la curva con el eje Y, valles y armónicas sucesivas, lo cual sirve para caracterizar el tipo de movimiento.

En particular se puede caracterizar el movimiento de un círculo por la posición y la amplitud de las armónicas dominantes para cada una de las aceleraciones tomadas en los ejes X, Y, Z.

Esta información es tomada para el procesamiento y entrenamiento con una ANN y un HMM.

V. CONCLUSIONES

En los últimos años se ha investigado extensamente el reconocimiento de patrones y de gestos en particular, sin embargo, aún no se establecen reglas y métodos óptimos para cada una de las posibles aplicaciones.

Es por eso que en este trabajo se han estudiado y analizados diversos métodos para la adquisición, filtrado, parametrización, caracterización, clasificación y reconocimiento de patrones aplicado a gestos con un control *WiiMote*.

En un trabajo futuro se presentará la comparación de los métodos ANN y HMM y se mostrará para que casos el desempeño es mejor.

REFERENCIAS

- [1] Lecun, Y., et al. “*Comparison of Learning algorithms for Handwritten Digit Recognition*”. International Conference on Artificial Neural Networks (ICANN).Paris, France, 1995. pp. 53-60.
- [2] Nathan, Krishna S., et al. “*Real-Time On-Line Unconstrained Handwriting Recognition Using Statistical Methods*”. International Conference on Acoustics, Speech, and Signal Processing (ICASSP).Detroit, MI, USA, 1995. Vol. 4, pp. 2619-2622. 10.1109/ICASSP.1995.480098.
- [3] Poppinga, Benjamin, et al. “*Gesture Recognition with a Wii Controller*”. 2nd International Conference on Tangible and Embedded Interaction (TEI).Bonn, Germany, 2008. 978-1-60558-004.
- [4] Borza, Paul-Valentin. “*Motion-based Gesture Recognition*”. Faculty of Mathematics and Computer Science, Babeş-Bolyai University of Cluj-Napoca, Romania. s.l. : Google Summer of Code, 2008. Bachelor’s Thesis. <http://code.google.com/p/accelges/>.
- [5] Prekopcsák, Zoltán. “*Accelerometer Based Real-Time Gesture Recognition*”. Proceedings of the 12th International Student Conference on Electrical Engineering (POSTER).Prague, Czech Republic, 2008.
- [6] Mlích, Jozef. “*WiiMote Gesture Recognition*”. Proceedings of the 15th Conference and Competition (Student EEICT).Brno, CZ, 2009. Vol. 4, pp. 344-349. 978-80-214-3870-5.
- [7] Murakami, Kouichi y Taguchi, Hitomi. “*Gesture Recognition using Recurrent Neural Networks*”. Proceedings of the Conference on Human Factors in Computing Systems (SIGCHI): Reaching through Technology.New Orleans, Louisiana, United States, 1991. págs. 237 - 242. 0-89791-383-3.

- [8] Chandra, Rohitash and Omlin, Christian W. “*Hybrid Recurrent Neural Networks: An Application to Phoneme Classification*”. [ed.] Hamid R. Arabnia, Jack Y. Yang and Mary Qu Yang. Proceedings of the 2007 International Conference on Genetic (GEM).Las Vegas, Nevada, USA, 2007. pp. 57-62. 1-60132-038-8.
- [9] Motion analysis. “*WiiLi.org Wii Linux*”. [En línea] http://www.wiili.org/index.php/Motion_analysis.
- [10] Wiibrew. “*Wiimote*”. [En línea] <http://wiibrew.org/wiki/Wiimote>.
- [11] Andersson, Joel. “*Libwiimote*”. [C Library] Simple Nintendo Wii Remote Framework for Linux. <http://libwiimote.sourceforge.net/>.
- [12] Wassner, Hubert. Sujet du TP “*neural computing*”. [En línea] 1 de Febrero de 2008. [Citado el: 16 de Diciembre de 2008.] <http://professeurs.esiea.fr/wassner/?Ico4168-neural-computing>.
- [13] Ferguson, Jack. “*Lectures on HMM Theory*”, Institute of Defense Analysis.
- [14] Ljolje, Andrej and Levinson, Stephen E. “*Development of an acoustic-phonetic Hidden Markov Model for continuous speech recognition*”. Transactions on Signal Processing.1991. Vol. 39, pp. 29-39.
- [15] Rabiner, L. R. “*A tutorial on Hidden Markov Models and selected applications in speech recognition*”. Proceedings of the IEEE.1989. Vol. 77.
- [16] Nissen, Steffen. “*Implementation of a Fast Artificial Neural Network Library (FANN)*”. Department of Computer Science, University of Copenhagen (DIKU). Copenhagen : s.n., 2003. Thesis.
- [17] Nikolaev, Nikolay. Single-Layer Perceptrons. “*311: Neural Networks*”. [En línea] http://homepages.gold.ac.uk/nikolaev/311_perc.htm.
- [18] Nissen, Steffen. “*Fast Artificial Neural Network Library (FANN)*”. [Free open source neural network library]
- [19] Delgado Mata, Carlos, et al. “*Low cost video game technology to measure and improve motor skills in children*”. AFRICON.Nairobi, Kenia, 2009.

AUTORES

Ing. Miriam Lee obtuvo su título de Ingeniería en Sistemas Computacionales con Especialidad en Negocios Electrónicos por la Universidad Panamericana.

Dr. Carlos Delgado es Ingeniero en Ciencias Computacionales por la Universidad Panamericana, Maestría en Ingeniería en Computación con especialidad en Realidad Virtual por la UNAM. Obtuvo el Doctorado en Agentes Inteligentes en Ambientes Virtuales por la University of Salford.

ARQUITECTURA GENÉRICA PARA EL PROCESAMIENTO DE MÚLTIPLES IMÁGENES EN TIEMPO REAL BASADA EN FPGA

Juan Manuel Xicoténcatl Pérez

Laboratorio de Investigación en Robótica y Electrónica Avanzada, Ingeniería en Mecatrónica
Universidad Politécnica de Pachuca
Carretera Pachuca-Cd. Sahagún km. 20 Rancho Luna ExHacienda de Sta. Bárbara C.P. 43830,
Zempoala, Hidalgo. Correo electrónico: jmxico@upp.edu.mx

RESUMEN

Se presenta una arquitectura genérica para el procesamiento de múltiples imágenes en tiempo real, la arquitectura se encuentra basada en arreglos sistólicos colocados en cascada y cuya conexión es variable. El procesamiento presenta importantes ventajas sobre diseños anteriores en cuanto a tiempo de procesamiento y prestaciones, siendo la principal característica su capacidad de filtrado espacial y en general, de procesamiento con máscaras. Los resultados muestran una velocidad de procesamiento de 60 fps a una velocidad de reloj de 20MHz.

Palabras Claves: FPGA, Sistemas sistólicos, procesamiento de imágenes

I. INTRODUCCIÓN

Diseñar un procesador hardware para el procesamiento de datos masivos, es un paso imprescindible para los sistemas de visión actuales. Usar arquitecturas para el procesamiento por ventanas es debido al hecho de que algoritmos complejos de procesamiento de imágenes utilizan primitivas para sus propósitos [1]. Para propósitos actuales, muchos procesadores secuenciales ya poseen la estructura y la capacidad de cómputo para el procesamiento en paralelo; sin embargo, sigue existiendo la limitante del procesamiento en tiempo real [2, 3].

Los avances en la tecnología de computadoras y los requerimientos de muchas de las aplicaciones actuales, han resultado en el uso de la visión por computadora en diferentes

campos tales como la robótica, sistemas multimedia, realidad virtual, inspección industrial, ingeniería médica y sistemas de navegación. Las limitaciones de estos sistemas, se encuentran en el volumen de los equipos que realizan las tareas de visión y la energía para alimentarlos [4].

Existen varios sistemas en la literatura que trabajan sobre ventanas en el procesamiento de imágenes [2, 3, 4]. Pero no realizan más que una tarea específica y eso sin considerar problemas de consumo de potencia o su habilidad para incorporarse a sistemas móviles.

En éste trabajo, se propone una arquitectura con la capacidad de procesar una imagen con diferentes más caras o ventanas, al mismo tiempo. Adicionalmente a éste esquema de trabajo, la arquitectura puede incrementar el tamaño de la ventana *en tiempo de ejecución*, lo que amplía el rango de aplicaciones. El resto de este trabajo está ordenado como sigue. La sección 2, muestra las bases del trabajo y presenta la arquitectura en sus componentes básicos. La sección 3 presenta los resultados obtenidos hasta el momento y la sección 4, da las conclusiones y trabajos a futuro.

II. DESARROLLO

Existen al menos dos criterios para diseñar una arquitectura para el procesamiento de imágenes y en particular, para la convolución espacial por medio de máscaras [6]:

2. Optimizar en base a la máscara o filtro espacial (por ejemplo, la máscara de promedio).
3. Optimizar en base al número de operaciones en la memoria de almacenamiento de la imagen

En el primer caso, la optimización se realiza en función del tipo de filtro a aplicar, por ejemplo, en el filtro de promedio, los valores de los píxeles dentro de una ventana, son sumados y promediados entre el número de píxeles de la ventana. El resultado entre ventanas adyacentes, difiere en apenas una columna de la ventana. Lo cual simplifica los recursos y la estructura a implementar. Un diseño de este tipo de estructuras se puede ver en la figura 1.

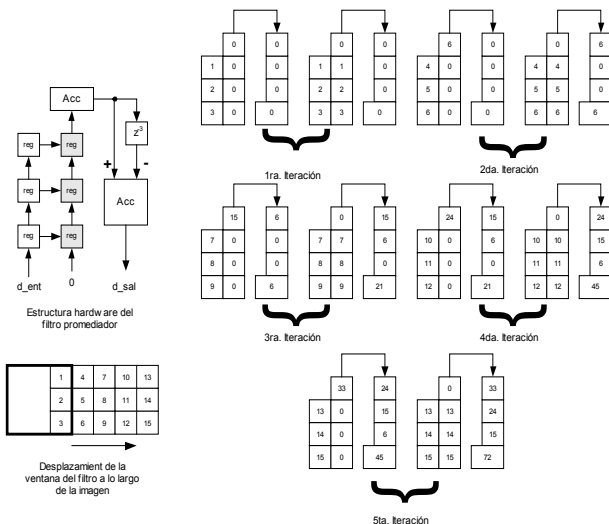


Figura 1. Se muestra la estructura hardware de un filtro de promedio, la división se realiza utilizando un multiplicador de punto fijo y no se muestra por simplicidad. Se muestran cinco iteraciones para ejemplificar el diseño, la latencia del proceso es de $2(n+1)$.

En la figura 1, se presenta la estructura hardware para un filtro de 3×3 . El filtro optimiza los recursos, al requerir, en este caso, 9 registros y un par de acumuladores. El elemento de retardo, permite restar la última columna del filtro del resultado. Ésta técnica, ahorra un 66.6% del tiempo de cómputo requerido por otras técnicas y representa un punto de partida para arquitecturas más complejas. El anterior diseño se puede aplicar a

filtros tales como el filtro gaussiano, utilizado para “suavizar” una imagen en tareas tales como detectores de borde o procesamiento piramidal, el cual se puede separar para procesar tanto por columnas como por renglones de forma separada.

La arquitectura anterior tiene la desventaja de implementar un filtro con características especiales. Sin embargo, si se desea abordar el diseño de una estructura genérica, en donde se desconoce la naturaleza del filtro, se debe optar por un diseño que tome en cuenta, no la optimización del filtro en sí, sino del número de operaciones de lectura/escritura de la memoria en donde se almacena a la imagen. Se propone la siguiente arquitectura para el procesamiento de ventanas en tiempo real.

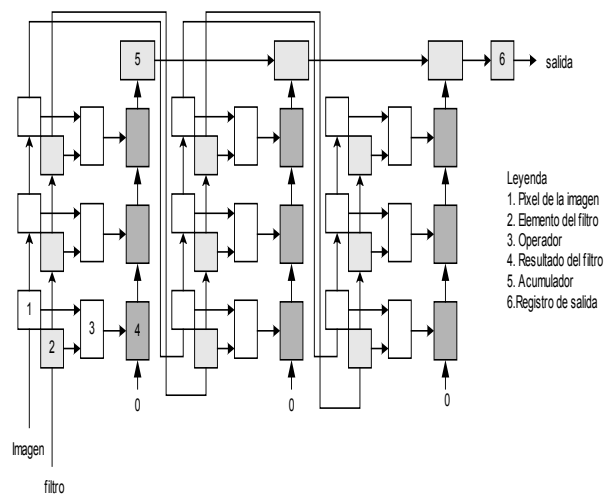


Figura 2. Diagrama a bloques del sistema para el filtrado de imágenes en tiempo real. Los datos del filtro y de la imagen son almacenados por semicolumnas. El procesamiento por semicolumnas, facilita el diseño de sistemas para el procesamiento de imágenes.

En la figura 2, se presenta en diagrama a bloques del sistema de filtrado. Como se menciona anteriormente, el cómputo es por semicolumnas, esto es, el proceso se realiza tomando una parte de una columna de la imagen. El proceso por columnas elimina la necesidad de buffers, elementos necesarios para procesar por renglones.

El funcionamiento del sistema de la figura 2, es como sigue: se cargan los datos del filtro en primer lugar, lo cual puede hacerse como si fuera un cuadro de imagen más, excepto que se habilitan los elementos de almacenamiento marcados como 2. A continuación, se introducen los elementos de la imagen por semicolumnas, una a la vez. Los módulos señalados con el número 3 en la figura 2, son bloques combinaciones y carecen de señales de control. Dichos módulos, representan el operador a realizar, ya sea una multiplicación, substracción, corrimiento a la derecha o izquierda (multiplicación y división por dos) o incluso comparaciones. Los módulos marcados como 4, reciben la información de los operadores y la envían mediante corrimientos a un acumulador de columna (módulo 5). Los acumuladores de columna transmiten dicha información a su vez, al acumulador general, cuya salida es el resultado del filtro. En la siguiente figura se ejemplifica a bloques el procesamiento realizado por esta arquitectura.

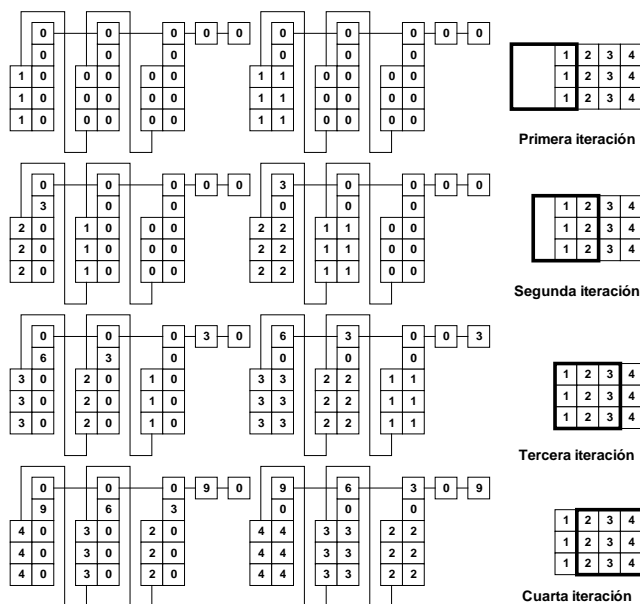


Figura 3. Se muestran cuatro iteraciones de la arquitectura. A la derecha se muestra el desplazamiento de la ventana de procesamiento sobre una imagen, se asume un filtro con coeficientes igual a. La latencia del procesamiento es de $3(n+1)$.

La figura dos muestra detalles de la estructura hardware, la latencia del

procesamiento es de $(3n+1)$, donde $n \times n$ es el tamaño del filtro, en éste caso es de 3×3 , o $n=3$. El procesamiento se realiza en semicolumnas a lo largo de una línea de barrido de la imagen (líneas horizontales). La estructura del sistema, no toma en cuenta la redundancia existente entre píxeles pertenecientes a líneas de barrido adyacentes. El número de ciclos de reloj utilizados para el procesamiento es, por lo tanto, para una imagen con dimensiones de $M \times N$, donde M corresponde a las columnas y N a los renglones, de $(M+3)N(n+1)$. Si la frecuencia de trabajo del sistema se define como f , el tiempo de procesamiento es de $(M+3)N(n+1)/f$ segundos.

En muchas ocasiones, el sistema de control provoca una caída en la frecuencia máxima de trabajo. En éste trabajo, el control ha sido diseñado de tal forma que cada operación de escritura/lectura y control de la arquitectura, toma exactamente un ciclo de reloj. El tiempo de procesamiento, en conjunto, sigue siendo de $(M+3)N(n+1)/f$.

Se ha mencionado que la arquitectura de la figura 2, no toma en cuenta la redundancia en la lectura de datos existente entre píxeles adyacentes en las líneas de barrido, lo que aumenta su tiempo de procesamiento. Para subsanar la falta de aprovechamiento de la redundancia, se propone la arquitectura siguiente.

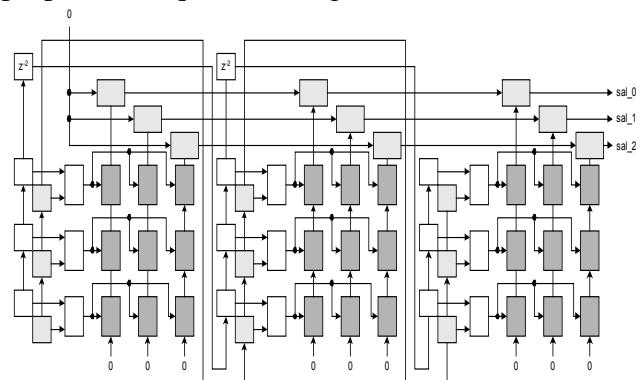


Figura 4. La arquitectura presentada, toma en cuenta la redundancia de datos entre líneas adyacentes, es posible notar más unidades de almacenamiento para los operadores, así como elementos de acumulación y retardo.

La arquitectura es semejante a aquella presentada en la figura 2, excepto porque existen más registros para los operadores, cuyo número no varía. Los elementos de retardo permiten almacenar los pixeles de la columna pertenecientes a la imagen a filtrar que no están siendo utilizados.

El funcionamiento es semejante al discutido para la figura 2, pero en ésta nueva arquitectura se siguen leyendo datos de la semicolumna en cuestión hasta r veces más datos, r es un número igual a $n-1$ y, es usado, para fijar el tamaño del elemento de retardo. Las características son:

1. Se usan $n \times n$ elementos para almacenar el segmento de la imagen a filtrar y el mismo número para los coeficientes del filtro.
2. Se usan $n \times n \times n$ unidades para implementar los operadores.
3. Se utilizan $r=n-1$ elementos de retardo con longitud r .
4. Las señales de control para los elementos de almacenamiento de los operadores se encuentran retrasadas un ciclo de reloj entre ellas. Por lo que se utiliza el mismo control tanto para la arquitectura de la figura 2 como de la arquitectura de la figura 3.

De las especificaciones anteriores es posible ver el potencial de la arquitectura de la figura 4, en términos del número de ciclos de reloj. El número de ciclos de reloj necesarios para que funcione la arquitectura es de $(M+3)N(n+1)/n$. Nuevamente, si f es la frecuencia de trabajo del sistema, se requieren de $(M+3)N(n+1)/(nxf)$ segundos para el procesamiento de toda la imagen.

Finalmente, se describe la arquitectura para el procesamiento múltiple de ventanas en tiempo real en la figura 5. Los registros de almacenamiento de coeficientes de filtro, almacenamiento de píxel de imagen, operador y registro del operador, son considerados un sólo módulo (EP, *Elemento Procesador*) para simplificar la figura.

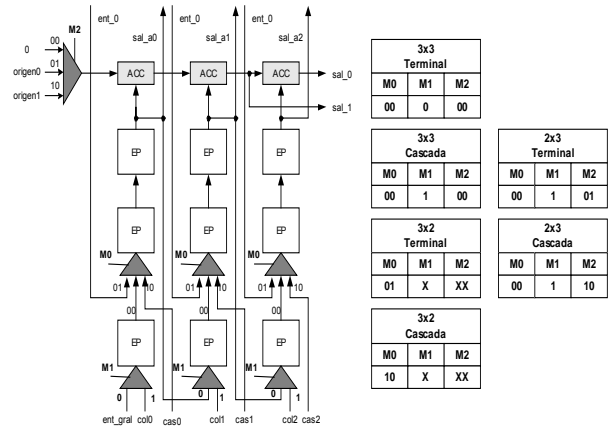


Figura 5. El módulo de procesamiento es definido por celdas de 3x3 y presenta cinco modos de funcionamiento, a partir de los datos que acepta de sus vecinos.

El sistema mostrado en la figura 5, es denominado *módulo básico*. Los cinco modos de funcionamiento permiten conectar dichos módulos en un arreglo capaz de permitir de manera simultánea el filtrado de imágenes.

Sin embargo, el procesamiento múltiple de ventanas de 3x3 no parece atractivo para propósitos prácticos, por ello, el arreglo de módulos básicos permite interconectarse de tal forma, que se puede incrementar el tamaño de la ventana de procesamiento en tiempo de ejecución. El control de la arquitectura es el mismo, excepto que se tiene varias señales que aunque se generan desde el principio, estas son llamadas conforme aumenta o disminuye el tamaño de la ventana de procesamiento. Se presenta a continuación lo que puede llamarse un arreglo típico y un par de configuraciones que permiten a la arquitectura trabajar con diferentes tamaños de imagen.

En la figura 6, se tiene un ejemplo de interconexión entre los módulos. Los diferentes módulos y su configuración se muestran, un procesador interno y dedicado es el encargado de dar la conexión y la configuración tanto de la memoria como de la salida de los datos.

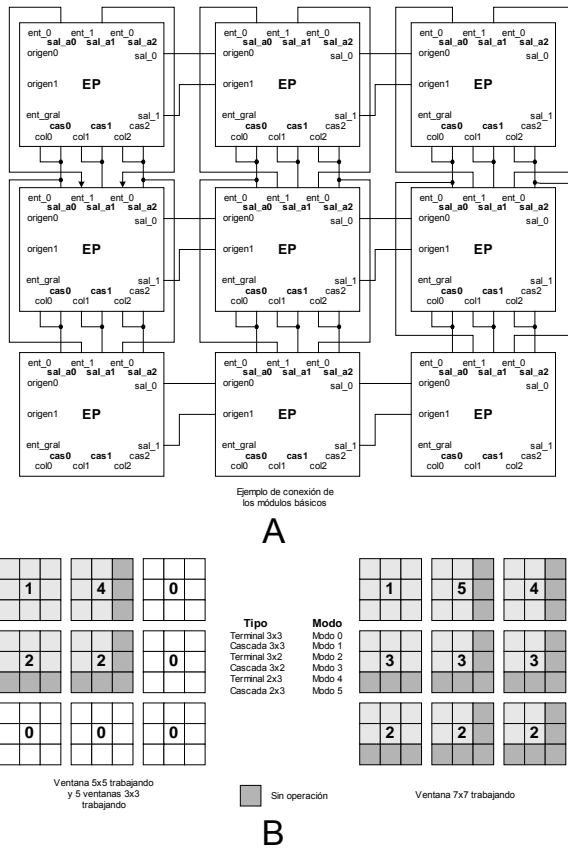


Figura 6. A) La conexión de los diferentes módulos de procesamiento en un arreglo de 3x3; B) Dos ejemplos de configuración, la configuración de todos los módulos n en modo es por default.

III. RESULTADOS

La arquitectura fue implementada utilizando código VHDL para su descripción y una tarjeta de desarrollo DIGILENT INC. Con un SPARTAN 3E de 0.5MGates. El reloj es de 20MHz y se ocupo el 92% de recursos del FPGA disponibles.

A la arquitectura descrita, se le añaden tres módulos complementarios: el módulo de lectura y configuración del arreglo óptico, (KAC2835 HeadBoard, 60 fps VGA y configurable a través de I²C), el módulo de manejo de la memoria DRAM, y el controlador del monitor VGA o memoria de video.

La versión fue sintetizada utilizando el XST del ISE 10.2i de Xilinx para el XC3S500Efg320-4. El modelo funcional fue probado con el ISE Simulator y la versión final

fue descargada con el iMPACT. Los detalles de la arquitectura son: Número de slices 4234 de 4656, numero de LUTS 8468 de 9312.

Se muestran a continuación algunas imágenes obtenidas con la arquitectura. No existe técnica de relleno alrededor de los bordes, por lo que se asume que los valores son cero.

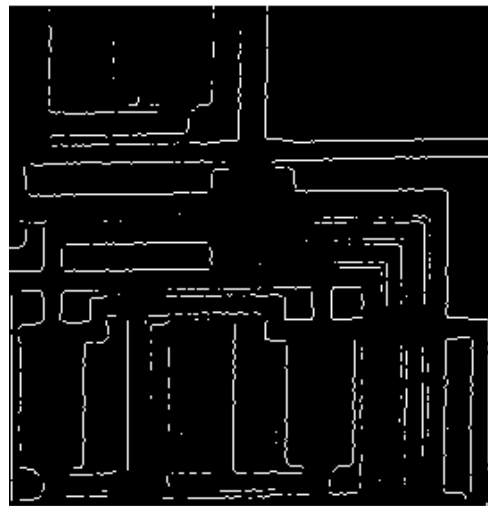
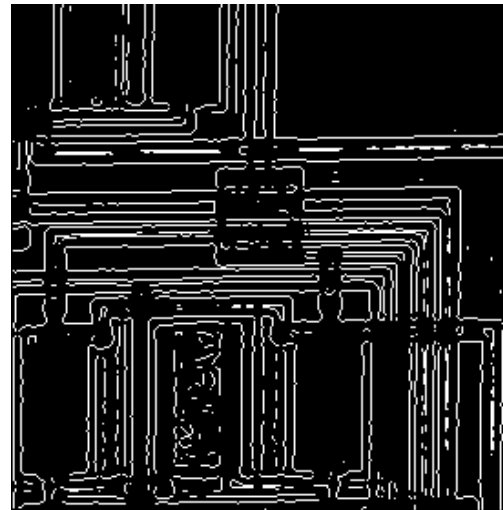


Figura 7. Se presentan dos ejemplos del procesamiento. a) Laplaciano del Gaussiano y b) Prewitt.

Se muestran dos ejemplos de filtrado, para establecer un comparativo, existen errores en los bordes debido al *padding*, pero ésta área se suele omitir en procesos posteriores.

IV. CONCLUSIONES

Se ha presentado en éste trabajo una arquitectura para el procesamiento de imágenes en tiempo real con múltiples máscaras. El sistema es ideal para procesar una imagen con múltiples filtros, por ejemplo, se puede hacer una estimación del gradiente en x y y con el operador sobel *al mismo tiempo* e inclusive, complementar los resultados con la máscara LOG (laplaciano del gaussiano).

V. TRABAJOS A FUTURO

La estructura de la arquitectura presenta características que la hacen inmejorable para aplicaciones de filtrado espacial; sin embargo, la estructura sistólica permite implementar también neuronas artificiales, algoritmos de correlación de flujo óptico y sistemas de seguimiento de objetivos con blancos móviles, o una mezcla de éstos *en tiempo real* y de forma *concurrente*.

AGRADECIMIENTOS

Este proyecto se encuentra respaldado por PROMEP a través del apoyo UPPACH-PTC-058.

VI. REFERENCIAS

- [1] R. Andraka. A "Survey of CORDIC Algorithms for FPGA based Computers" – Proc of the 1998, CM/SIGDA Sixth International Symposium on FPGAs, Monterey, CA, pp. 191-200. February 22-28, 1998.
- [2] *Spartan 3 FPGA Family: Complete Data Sheet*, Xilinx Inc., April 26, 2006.
- [3] S. O. Memik, A K Katsaggelos, and M. Sarrafzadeh. "Analysis and FPGA implementation of image restoration under resource constraints", IEEE Transactions on Computers, vol. 52, no.3. Mar. 2003.

- [4] J. E. Volder, "The CORDIC Trigonometric Computing Technique" IRE Transactions on Electronic Computers, vol. EC-8, no. 3, pp. 330–334, 1959.

VII. AUTOR

Juan Manuel Xicoténcatl Pérez, recibió el grado de doctor por el Instituto Nacional de Astrofísica Óptica y Electrónica (INAOE) en 2005. El grado de maestría en ciencias en 2000 y el de licenciatura en 1998. Actualmente se encuentra en la Universidad Politécnica de Pachuca como profesor investigador de tiempo completo. Sus intereses son las arquitecturas de visión por computadora, procesamiento de señales y microprocesadores.

IMPLEMENTACIÓN DE ALGORITMOS COMPUTACIONALES PARA LA IDENTIFICACIÓN DE PERSONAS MEDIANTE LA LECTURA DEL IRIS

Ricardo Fuentes Covarrubias¹, Andrés Gerardo Fuentes Covarrubias¹, Enrique Cabello Pardos².

¹ Facultad de Ingeniería Mecánica y Eléctrica, Universidad de Colima
Av. Universidad 333, Col. Las Víboras, Colima Col. 280400, México

Tel. (312)3161165, correo electrónico: fuentesr@uclm.mx, fuentesg@uclm.mx

² Universidad Rey Juan Carlos, Madrid España. Correo electrónico: enrique.cabello@urjc.es

RESUMEN

Se presentan los resultados de un proyecto de investigación en el área de la biometría basado en visión artificial orientado a la identificación de personas mediante la lectura del iris. La orientación del proyecto se centra en la localización, extracción y reconocimiento de los rasgos característicos del ojo humano y finalmente la localización del iris, sus contornos y las áreas que definen a la persona identificada.

Palabras clave: Biometría, visión artificial, reconocimiento automático.

I. INTRODUCCIÓN

En este trabajo se muestran los resultados obtenidos mediante el desarrollo de un sistema de identificación de personas basado en la biometría. Con la necesidad del desarrollo de sistemas de alta seguridad se ha optado por realizar investigaciones de métodos no tan fáciles de evitar como por ejemplo los NIP's, los códigos de barras y los algoritmos para el reconocimiento de voz, por mencionar algunos; se ha optado por tomar el reconocimiento de personas mediante la lectura del iris o el escaneo de la retina ya que estas partes del cuerpo humano no son fáciles de duplicar como lo es la huella digital o los ya antes mencionados por lo tanto esta investigación se enfoca a lo que es la Biometría del ojo humano que en conjunción con un sistema de computo nos permite obtener un buen sistema de seguridad y obtención de datos difíciles de burlar como por ejemplo el registro de empleados de una empresa basado en reloj chocador, en el cual no existe garantía de que sea el mismo trabajador el que registre su asistencia.

II. ANTECEDENTES

La Biometría se define como el análisis estadístico de observaciones biológicas [2]. Es también una ciencia emergente que compara electrónicamente características biológicas de un individuo (rostro, huella dactilar, voz, etc.) contra una población de una o más de tales características. La Identificación y Autenticación biométrica explota el hecho de que ciertas características biológicas son singulares e inalterables y son además, imposibles de perder, transferir u olvidar. Esto las hace más confiables, amigables y seguras que las contraseñas.

Tipos de Biometrías:

- Biometría Estática: Mide la anatomía del usuario.
 - Huellas digitales
 - Geometría de la mano
 - Termografía
 - Análisis del iris
 - Análisis de la retina
 - Venas del dorso de la mano
 - Reconocimiento facial
- Biometría Dinámica: Mide el comportamiento del usuario.
 - Patrón de voz
 - Firma manuscrita
 - Dinámica de teclado
 - Análisis gestual

II.1 EL ANÁLISIS Y RECONOCIMIENTO DE LA IMAGEN DEL IRIS

La estructura global del sistema propuesto se desglosa a continuación: una imagen de la región

del ojo humana circundante se obtiene a una distancia de 20 cm. de una cámara. En la fase del preprocesamiento, se realizan los pasos siguientes: Primero, localizar el iris, la porción de la imagen es procesada. Segundo, el sistema de coordenadas cartesianas de la imagen es convertida en un sistema de coordenadas polares para facilitar el proceso de extracto de rasgos. En la fase de extracto de rasgos, 2-D Wavelet la transformación se usa para extraer un vector de los rasgos de la imagen del iris. En la fase final, la identificación y fase de la comprobación; se explotan dos métodos de aprendizaje competitivos revisados para LVQ: clasificación de los vectores del rasgo y reconocimiento de la identidad de persona. Para mejorar la eficacia del sistema, algunos métodos se aplican a la fase de extracción del rasgo y a la fase de identificación.

Para localizar el iris, se debe encontrar el centro de la pupila primero, y enseguida determinar los límites internos y exteriores. Aplicando el método de la bisección a cada dos puntos en el mismo borde, se puede conseguir sólo un punto en el caso ideal que cruza cada línea perpendicular encima de la línea que conecta a dos puntos, pero se selecciona el centro que apunte como el punto frecuentemente cruzado. Después de determinar el punto del centro, se encuentra el límite interior y el límite exterior extendiendo el radio de la pupila y contando el número de los puntos del borde en el círculo virtual trazado. Dos círculos virtuales con el número máximo de puntos del borde dentro de cada rango correspondiente determinado por algún anterior son seleccionados como los dos límites que se desea encontrar. Consulte la secuencia de figuras 1, 2 y 3 para ilustrar lo anterior.

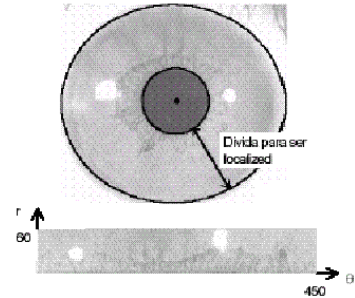


Figura 1. Resultados de la fase de preprocesamiento.

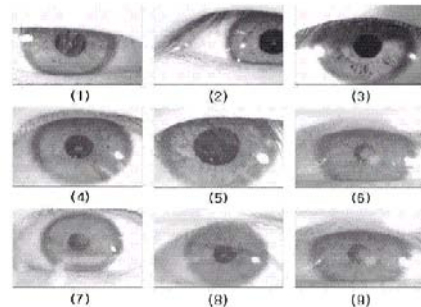


Figura 2. Tipos de fallos en la fase de preprocesamiento

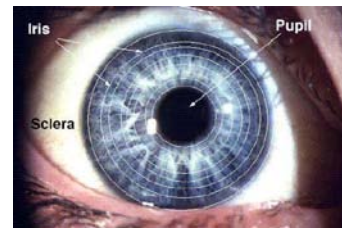


Figura 3. Fase de reconocimiento.

Para la identificación mediante el reconocimiento del iris se toman muestras en diferentes capas del iris en forma de aros, posteriormente se utilizan algoritmos para reconocimiento de imágenes los cuales permiten obtener un identificador diferente para cada persona

III. DESARROLLO DEL PROYECTO

III.1 INFRAESTRUCTURA NECESARIA.

Para el sistema desarrollado se utilizaron las siguientes herramientas computacionales: PC a 1.6 GHZ y 2GB de memoria RAM con Sistema operativo Windows XP Profesional. Lenguaje Visual Basic versión 6.0.

Cámara Panasonic modelo Autenticam de uso específico.

III.2 LOCALIZACIÓN DEL ÁREA DEL IRIS

En este punto se utiliza la transformada de Hough [3] para rectas para determinar la circunferencia del iris, Esta transformada busca formas geométricas en toda la imagen encontrando los parámetros de aquellas que contengan a más puntos en ella.

Para evitar la singularidad que pueden dar las rectas de pendiente infinitas, la ecuación de una línea recta puede escribirse como:

$$x \cos \theta + y \sin \theta = \rho$$

Quedando por tanto cada recta definida por su par de parámetros θ y ρ (Figura 4). Todos los puntos que pertenezcan a la misma recta coincidirán en las componentes θ y ρ .

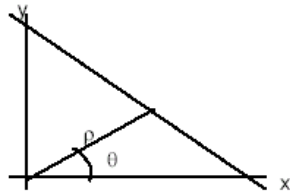


Figura 4. Gráfico de la transformada de Hough.

Al igual que en una recta puede definirse paramétricamente, se tiene que una circunferencia cumple la curva:

$$(x - x_0)^2 + (y - y_0)^2 = R^2$$

A diferencia del caso anterior ahora se tiene tres parámetros (x_0, y_0, R).

Para un punto perteneciente al borde (X_1, Y_1), el espacio de posible circunferencia vendrá dado por un cono (Figura 5) definido por la siguiente ecuación:

$$(x_1 - x_0)^2 + (y_1 - y_0)^2 - R^2 = 0$$

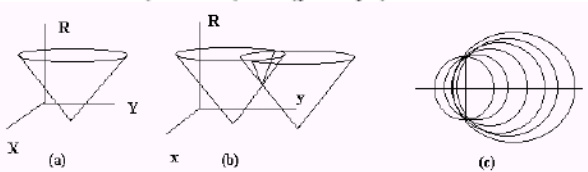


Figura 5. Transformada de Hough para circunferencia.

III.3 TRANSFORMADA DE SOBEL PARA LA DETECCIÓN DE BORDES

El operador de Sobel [3] se distingue del operador de Prewitt en que se duplican los valores de los píxeles al norte, sur, este y oeste, es decir, $K=2$.

De esta forma, cada píxel de entorno de vecindad es ponderado de acuerdo a las distancias de este al central (píxel evaluado). Como consecuencia de ello, Sobel es mas sensible que Prewitt a los bordes diagonales, mientras que este lo es para los horizontales y verticales. En la Práctica, sin embargo, no se aprecia gran diferencia entre ellos. Por otra parte Freí y Chen propusieron un valor de $k = \sqrt{2}$ con el fin de que el gradiente sea el mismo para bordes horizontales y verticales y diagonales. Estos operadores 3×3 tienen unas prestaciones bastantes similares.

Particularmente, los dos primeros (Prewitt y Sobel) son muy populares, debido a que los valores de los coeficientes hacen que las operaciones aritméticas se puedan programar de una manera eficiente o implementar fácilmente el hardware.

III.4 ELIMINACIÓN DE RUIDO

Para la eliminación de ruido se utiliza un método de eliminación de píxeles que pertenecen a la imagen pero que solamente afectan al procesamiento de la misma, para eso se emplea un contador de píxeles blancos en una matriz de 5×5 donde si el numero de píxeles es menor que 16 entonces realiza una conversión de los píxeles blancos en negros.

B	B	N	N	N
N	B	B	N	N
N	N	B	B	N
N	N	N	B	B
N	N	N	N	B

N	N	N	N	N
N	N	N	N	N
N	N	N	N	N
N	N	N	N	N
N	N	N	N	N

$B < 16$

Cambia la matriz en píxeles negros

N	N	B	B	B
B	N	N	B	B
B	B	N	N	B
B	B	B	N	N
B	B	B	B	N

N	N	B	B	B
B	N	N	B	B
B	B	N	N	B
B	B	B	N	N
B	B	B	B	N

$B > 16$

Deja la matriz en su forma original

El proceso de conteo de píxeles se realiza en toda la imagen, donde el proceso se realiza de 5 en 5 para X, y para Y hasta el fin de la misma. Consulte las figuras 6a y 6b.

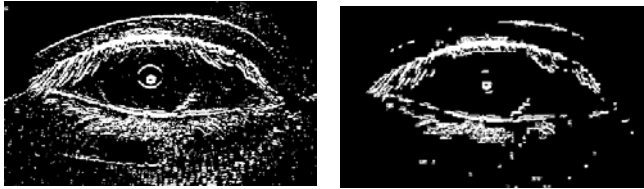


Figura 6a. imagen con ruido. 6 b. Imagen tratada sin ruido.

III.6 NORMALIZACIÓN DE PUNTOS

Se utilizó un método matemático, con el cual se normalizan tres puntos con respecto al lagrimal y al centro de la pupila donde el punto X_0, Y_0 es las coordenadas del lagrimal y el punto X_1, Y_1 es el centro de la pupila.

Para el reconocimiento del iris, la imagen obtenida puede tener un ángulo de inclinación, por lo tanto la normalización de estos puntos es indispensable para tomar un marco de referencia.

El lagrimal y la pupila van a estar en línea sin importar la posición de la cabeza de la persona. El punto a, b, y c, son los puntos que vamos a girar con respecto a α donde alfa es el ángulo formado entre el lagrimal y la pupila. Los nuevos puntos a y b, quedan sobre la recta lagrimal – pupila y el punto c queda perpendicular a esta.

En la figura 7 se puede observar como los puntos se normalizan con respecto al lagrimal donde los puntos originales están sobre las líneas claras y los puntos normalizados están sobre las líneas oscuras. También se puede observar que los puntos se interceptan en las líneas imaginarias trazadas en el centro de la pupila que en este caso es nuestro principal marco de referencia.

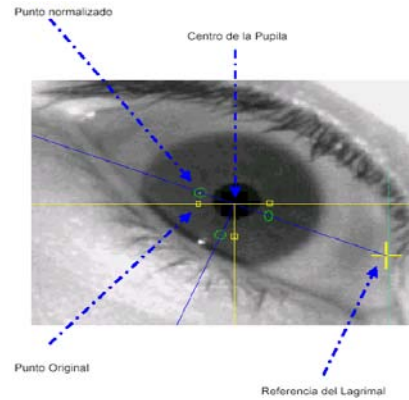


Figura 7. Imagen normalizada.

IV.6 TRANSFORMADA DE WAVELETS Ó FILTRO DE GABOR

Para la obtención de datos, de los puntos normalizados con respecto a la pupila y al lagrimal se aplica el método de Wavelets [1] donde a cada valor del píxel se le aplica una operación con las matrices mostradas en la figura 8:

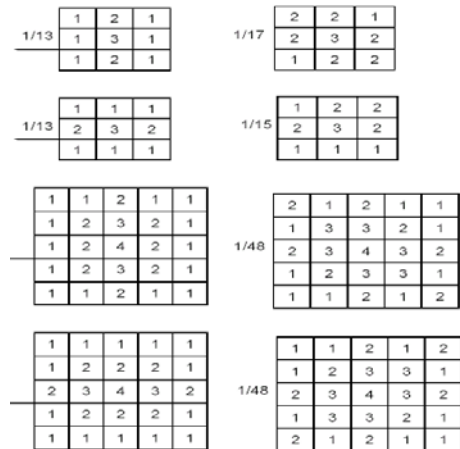


Figura 8. Transformada de Wavelets

Lo anterior permite obtener 8 números para cada punto. Estos números representan las variaciones de los niveles de los colores.

II.7 RECONOCIMIENTO

Con los 8 números de cada punto se calcula la distancia euclídea con el vector que se va a reconocer (P_i) con cada uno de los vectores (C_i) almacenados en la base de conocimientos del sistema. Donde posteriormente se eligen los dos

vecinos mas cercanos al resultado de la distancia euclídea

$$D \text{ Euclídea} = \sqrt{\sum_{i=0}^7 (P_i - C_i)^2}$$

Para la localización de Iris se toma como marco de referencia el punto blanco que se encuentra dentro de la pupila el cual es el reflejo del infrarrojo de la cámara, a partir de ahí se realiza un barrido de pixeles en todas direcciones para obtener el limite de la pupila con el iris donde ya obtenido el centro de este se procede a calcular el radio así como su posición en la imagen para de ahí aplicar 2D Wavelets.

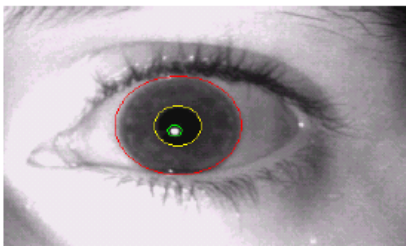


Figura 9. Localización del iris.

La figura 9 muestra como se realiza la búsqueda de objetos, donde primero se busca la zona que esta rodeada de color verde que es el reflejo de foco infrarrojo, y de ahí busca el limite de la pupila con el iris (circulo amarillo) y de ahí proceder a encontrar el iris y el lagrimal, es de hacer notar que existen causales de fallos ocasionados entre otros motivos debido a exceso de brillo, un mal enfoque del ojo o en su defecto a una insuficiente apertura del mismo.

III.8 INTERFACES DE USUARIO

El software desarrollado permite la identificación de una persona a través de varias etapas:

- a) Captura de la imagen base.
- b) Almacenamiento de la imagen con formato de imagen de Mapa de Bits (BMP).
- c) Posterior a este proceso se diseñó un Manejador de Bases de Datos en el cual se almacenaron los nombres de las imágenes de las personas que colaboraron en el proyecto.

- d) Obtención de las características de los mismos. En el sistema se realizó una rutina en la cual se obtienen tres muestras del ojo derecho y tres muestras del ojo izquierdo, donde las dos primeras muestras de cada iris se van a utilizar para obtener todas las características posibles. La tercer muestra se va a utilizar para identificar a la persona ya que en la Base de Datos Se Almacena El Nombre completo de cada Individuo. Consulte la imagen 10.

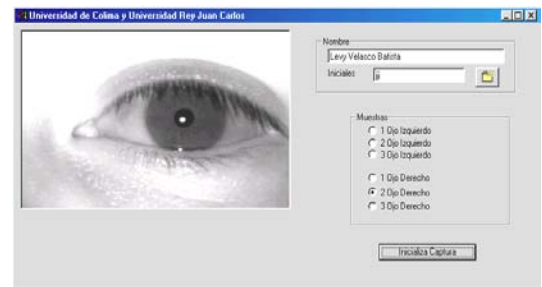


Figura 10. Captura de datos de individuos.

Con este sistema se obtuvo un total de 24 capturas habiéndose almacenado un total de 12 muestras del ojo Izquierdo y 12 muestras del ojo derecho. Solamente una de ellas se usará para el procesamiento.

La firma de cada imagen es esta:

C:\IrisRec\Imagenes\AGM\AGMIZQ104.bmp
 75.6153 75.7058 75.6153 76.7333 77.675 77.5625
 77.5 77.8541 73.6153 73.1764 73.0769 73.9333
 73.3 73.2083 73.15 73.229161.6923 60.7647
 60.5384 61.4666 60.725 60.25 60.15 60.4375

Para la ubicación de la pupila se hace un barrido de la imagen de izquierda a derecha para localizar un punto blanco que esté rodeado de puntos negros y almacenar la posición de estas coordenadas en variables globales, las cuales serán utilizadas posteriormente para localizar el lagrimal del ojo, con este método se hace un barrido hacia la izquierda o derecha según la imagen del ojo si es izquierdo se hace hacia la derecha hasta que encuentra un punto blanco y de ahí parte hacia abajo mientras siga encontrando un puntos negros y si encuentra un punto blanco ahí termina

diciendo que es el párpado inferior, la sumatoria de los puntos verticales divididos entre dos es la posición del lagrimal con respecto al mismo valor de X donde se encontró el punto blanco del párpado superior. A continuación se muestra en la figura 11 la interfaz de procesamiento de imágenes del sistema:

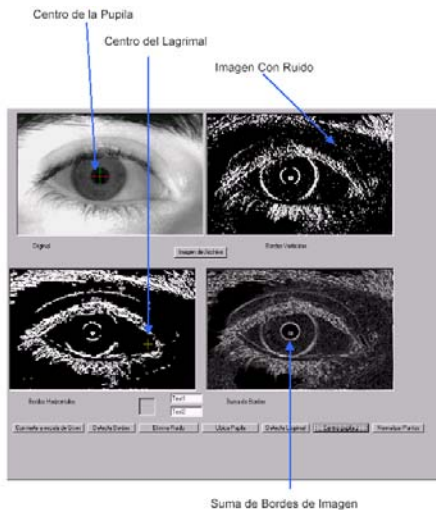


Figura 11. Interfaz del sistema.

La cual consta de los siguientes bloques de trabajo:

1. Convierte la imagen original a escala de grises.
2. Detecta los bordes de la imagen, así como verticales horizontales y realiza la suma de ellos
3. Elimina el ruido de la imagen que tiene los bordes horizontales.
4. Ubica la pupila utilizando como punto de referencia el reflejo del infrarrojo de la cámara.
5. Realiza una rutina para detectar el lagrimal con respecto a las coordenadas retornadas por la ubicación de la pupila
6. Ubica el centro de la pupila y elimina el reflejo del infrarrojo de la cámara

IV. CONCLUSIONES

El presente trabajo ha servido como base para el desarrollo de aplicaciones orientadas al control de acceso a recintos y áreas restringidas. Asimismo es base para el desarrollo de aplicaciones de control de dispositivos mediante algoritmos de visión artificial. Respecto a esta última área de trabajo, actualmente se desarrolla en la Universidad de Colima un prototipo de control para una silla de

ruedas mediante los movimientos de la cabeza, orientándolo a personas con cuadruplégia.

V. REFERENCIAS

- [1] Lawrence O'Gorman, Lawrence O'Gorman, (2008). *Practical Algorithms for Image Analysis*, Cambridge University Press.
- [2] Wayman, J. and Jain, A. and Maltoni, D. and Maio, D., (2005) *Biometric Systems: Technology, Design and Performance Evaluation*. Springer.
- [3] Michael Seul, Lawrence O'Gorman, Michael J, Sammon, (2001). *Practical Algorithms for Image Analysis, Description, Examples, and Code*. Cambridge University Press.
- [4] Daugman John. *How Iris Recognition Works*, *CirSysVideo(14)*, No. 1, January 2004, pp. 21-30. IEEE Abstract. 0402
- [5] Daugman, J.G.[John G.], *New Methods in Iris Recognition*, *SMC-B(37)*, No. 5, October 2007, pp. 1167-1175. IEEE DOI Link 0711BibRef.
- [6] Daugman, J., *Recognizing Persons by Their Iris Patterns*, In: *Biometrics: Personal Identification in Networked Society, 1998*, Amsterdam: Kluwer.
- [7] Bowyer, K.W.[Kevin W.], Hollingsworth, K.P.[Karen P.], Flynn, P.J.[Patrick J.], *Image Understanding for Iris Biometrics: A survey*, No. 2, May 2008.

VI. AUTORES

Ricardo Fuentes Covarrubias. Maestro en ciencias computacionales por la Universidad de Colima, Master en tecnologías de información y sistemas informáticos por la Universidad Rey Juan Carlos de España. Profesor e investigador en la Facultad de Ingeniería Mecánica y Eléctrica de la Universidad de Colima.

Andrés Gerardo Fuentes Covarrubias. Maestro en ciencias computacionales por la Universidad de Colima, Master en Tecnologías de Información y Sistemas Informáticos por la Universidad Rey Juan

Carlos de España. Profesor e investigador en la Facultad de Ingeniería Mecánica y Eléctrica de la Universidad de Colima.

Enrique Cabello Pardos. Doctor en informática por la Universidad Politécnica de Madrid, España. Licenciado en Ciencias Físicas por la Universidad de Salamanca, España. Profesor titular de la Escuela superior de Ingeniería Informática de la Universidad Rey Juan Carlos de España.

EVALUACIÓN ELECTROACÚSTICA DE IPODS

Servín Rivas Ilhuicamina Trinidad

Academia de Acústica, Departamento de Ingeniería en Comunicaciones y Electrónica
E.S.I.M.E. Zacatenco, I.P.N.
Unidad Profesional Lic. Adolfo López Mateos, Col. Lindavista, México, D.F. 07738, México
Tel. 57296000 ext. 54616, correo electrónico: iservinr@ipn.mx

RESUMEN

La exposición a niveles de presión acústica elevados durante largo tiempo conlleva el riesgo de pérdida de la audición o sordera, la cual se produce debido a la destrucción de las células ciliadas, encargadas de transmitir al cerebro los impulsos eléctricos asociados a la audición. El uso cada vez más frecuente de los aparatos denominados iPods y otros reproductores de audio que permiten almacenar gran cantidad de música, puede resultar perjudicial para los oídos, debido a la potencial exposición a niveles altos de audio durante periodos prolongados. Considerando que esta situación representa una problemática latente de salud, especialmente en la población juvenil de nuestro país, en este trabajo se propone un procedimiento de evaluación electroacústica empleando un sonómetro convencional, a fin de determinar los niveles sonoros producidos por los equipos de audio portátiles empleados usualmente por los alumnos de educación superior, y se describen los resultados preliminares obtenidos, observando que estos valores representan un riesgo para la audición.

Palabras clave: caracterización de iPods, pérdida auditiva debida al uso de iPods

I. INTRODUCCIÓN

Escuchar música de rock a volúmenes altos ha contribuido a la pérdida de la audición en algunos músicos emblemáticos de los años sesentas. Es conocido el caso de Pete Townshend, guitarrista del grupo inglés The Who, quien sufre de sordera y tinitus, molesto zumbido que se presenta en los

oídos, a causa de la exposición a los altos niveles de presión acústica de sus conciertos, y según admite él mismo, al elevado volumen con el que escucha la música a través de sus audífonos.

En parecida situación se encuentran entre otros, Brian Epstein, productor de The Beatles, el músico Sting, y es de esperarse que algo similar ocurra con una buena cantidad de músicos en nuestro país.

En este contexto destaca la creación de la organización norteamericana H.E.A.R. [1] (abreviatura de los términos en inglés *Hearing Education and Awareness for Rockers*) fundada por la cantante y bajista Kathy Peck y el médico Flash Gordon, con el fin de orientar a los músicos y al público en general en el cuidado del instrumento musical más valioso con el que podemos contar: la audición.

Desafortunadamente el riesgo de pérdida de la audición también está presente en otras personas, como es el caso de los asiduos usuarios de aparatos de audio que emplean audífonos, en los que predominan los jóvenes.

En los equipos de audio portátiles, como el *walkman* o el reproductor de discos compactos, la música se escucha usualmente mediante audífonos que se insertan directamente en los canales auditivos. El uso de estos equipos ha permitido escuchar la música de manera personal y privada, pero también ha traído consigo la potencial exposición a elevados niveles de presión acústica que pueden resultar perjudiciales para los oídos. Con la aparición de los aparatos denominados

iPods y otros reproductores de audio que utilizan formatos como el MP3, con los cuales es posible tener una gran cantidad de música almacenada, es previsible que aumente el riesgo de pérdida de la audición, debido a que la exposición a niveles altos de audio puede ser por periodos más prolongados.

El exponerse a niveles de presión acústica elevados durante largos intervalos de tiempo puede ocasionar pérdida de la audición o sordera, la cual se produce debido a la destrucción de las células ciliadas que se encuentran dentro de la cóclea, delicado mecanismo del oído interno encargado de convertir la energía sonora en señales eléctricas que se transmiten al cerebro, dando lugar a la audición. [2]

Inicialmente, la pérdida de la audición puede ser sólo temporal, volviendo a la normalidad después de algunas horas o días. Sin embargo, si la exposición ocurre de manera repetida, el mecanismo de la audición perderá su habilidad para recuperarse, resultando en una degeneración auditiva permanente, siendo el daño irreversible. En ocasiones también pueden presentarse molestias como el tinnitus.

En diferentes referencias [3,4,5] se advierte, principalmente por parte de audiólogos y médicos, acerca del riesgo que representa escuchar música a volúmenes elevados, a través de los audífonos que emplean los equipos de audio portátiles, ya que se puede provocar una forma de pérdida de la audición inducida por ruido.

Especialmente los jóvenes usuarios de estos equipos que tienden a aumentar el volumen de sus equipos, considerando quizá que ciertos géneros de música sólo se pueden disfrutar escuchándolos a niveles sonoros altos, no se percatan que se están causando un daño, pues la pérdida de la audición inducida por ruido se desarrolla de manera lenta e insidiosa, manifestándose cuando es demasiado tarde.

Cuando una persona está expuesta a sonidos fuertes durante periodos largos de tiempo, los síntomas de pérdida de la audición inducida por

ruido aumentan gradualmente. Con el tiempo, los sonidos que una persona escucha pueden percibirse distorsionados o atenuados, y puede ser difícil para la persona entender las conversaciones. Una persona con pérdida de la audición inducida por ruido puede no ser consciente de la pérdida, pero ésta puede ser detectada mediante una audiometría, estudio que permite evaluar la agudeza auditiva a diferentes frecuencias.

Estudios recientes muestran también que la exposición a niveles de ruido perjudiciales provoca la formación de radicales libres dentro del oído, que dañan las células ciliadas, produciendo pérdida de la audición inducida por ruido y tinnitus. [2]

A fin de prevenir la pérdida de la audición inducida por ruido, es conveniente evitar la exposición a niveles sonoros que resultan peligrosos o reducir el tiempo de exposición. Algunos organismos relacionados con el cuidado de la audición [2] sostienen que los niveles sonoros que exceden 85 dBA resultan riesgosos. Es posible también tomar como referencia los niveles máximos permisibles de exposición a ruido en ambientes laborales que se establecen en la norma oficial mexicana NOM-011-STPS-2001 “Condiciones de seguridad e higiene en los centros de trabajo donde se genere ruido” [6]. Esta norma estipula un nivel de exposición a ruido (NER) de 90 dBA para un tiempo máximo permisible de exposición (TMPE) de 8 horas, reduciendo el tiempo de exposición a la mitad cada vez que se tiene un incremento de 3 dBA, incremento que equivale a duplicar la energía, por lo que para un nivel de 93 dBA se permite un tiempo de 4 horas y para un nivel de 105 dBA sólo es aceptable un tiempo de 15 minutos.

II. METODOLOGÍA

Con el fin de analizar el riesgo de pérdida de la audición debida al uso de iPods o equipos similares, es necesario determinar la cantidad de sonido que impacta el mecanismo de la audición de las personas cuando usan iPods.

Para complementar este análisis, se ha realizado una encuesta a un grupo de estudiantes de la especialidad de Acústica de la carrera de Ingeniería en Comunicaciones y Electrónica de la E.S.I.M.E. Zacatenco, en la que se han considerado las siguientes preguntas: 1) ¿tienes iPod o algún reproductor de audio similar?, 2) ¿desde cuando haces uso de este tipo de reproductor?, 3) indica dos géneros de música que escuches usualmente, 4) ¿qué tan frecuente usas tu iPod por día?, 5) considerando una escala del 0 al 100, ¿con que valor de volumen acostumbras escuchar? 6) ¿cuánto tiempo escuchas tu iPod con este volumen?, 7) ¿recuerdas si el porcentaje de volumen ha aumentado últimamente? 8) ¿a que consideras que se deba el aumento de volumen?, 9) ¿has notado alguna disminución en tu capacidad auditiva últimamente?, 10) ¿consideras que el uso de equipos de audio con audífonos es dañino para el oído?

Para realizar la evaluación electroacústica de equipos de audio portátiles que utilizan audífonos, es conveniente conocer las magnitudes acústicas que pueden usarse como descriptores para este propósito.

El parámetro más común para medir la amplitud de un sonido es el nivel de presión acústica instantáneo, definido por:

$$L_p = 20 \log_{10} \left(\frac{p}{p_0} \right) \quad (1)$$

donde,

L_p = nivel de presión acústica en un instante de medición, en decibeles (dB)

p = valor raíz cuadrático medio de la presión acústica, en Pa

p_0 = presión acústica de referencia = 2×10^{-5} Pa

Cabe mencionar que el término empleado por la normativa nacional [6] es nivel de presión acústica, en vez de nivel de presión sonora, como suele emplearse de manera incorrecta comúnmente, ya que de acuerdo con la normativa vigente, un nivel

sonoro implica un nivel medido con una red de ponderación en frecuencia normalizada, como se verá a continuación.

Para evaluar los efectos del ruido en el ser humano, se ha convenido internacionalmente en utilizar el nivel de presión acústica con ponderación A o nivel sonoro A, expresado en dBA. La ponderación A busca incorporar la respuesta que tiene el oído humano en función de la frecuencia. El nivel sonoro A puede expresarse como:

$$L_{pA} = 20 \log_{10} \left(\frac{p_A}{p_0} \right) \quad (2)$$

donde,

L_{pA} = nivel de presión acústica con ponderación A, o nivel sonoro A, en decibeles A (dBA)

p_A = valor raíz cuadrático medio de la presión acústica con ponderación A, en Pa

p_0 = presión acústica de referencia = 2×10^{-5} Pa

Este parámetro puede medirse con un sonómetro, instrumento que permite medir el nivel de presión acústica en decibeles, incorporando redes con las que se obtienen las curvas de ponderación en frecuencia, cuyo propósito es simular la respuesta del oído. En la Fig. 1 se muestran las gráficas de respuesta a la frecuencia de las curvas de ponderación A, B y C. La red de ponderación A proporciona una respuesta que se aproxima a la curva invertida de un contorno de igual sonoridad para niveles de presión acústica bajos, la red de ponderación B corresponde a un contorno invertido para niveles de presión acústica medios y la red C a un contorno invertido para niveles de presión acústica altos; actualmente la red de ponderación A es la más ampliamente usada, siendo seguida por la C, que se aproxima a una respuesta plana, mientras que la red B ha caído en desuso.

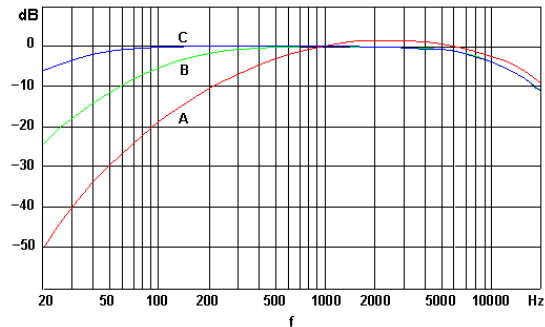


Fig. 1 Curvas de ponderación en frecuencia

Debido a que el ruido se caracteriza por variar continuamente en función del tiempo, se suele utilizar también para su cuantificación el nivel sonoro continuo equivalente L_{Aeq} , que corresponde al nivel constante en dBA que produciría la misma cantidad de energía acústica con ponderación A que un sonido variante en el tiempo, en un periodo establecido de tiempo. Se define mediante la siguiente expresión:

$$L_{Aeq} = 10 \log_{10} \left[\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \frac{p_A^2}{p_0^2} dt \right] \quad (3)$$

donde,

L_{Aeq} = nivel sonoro continuo equivalente, en dBA

p_A = valor raíz cuadrática medio de la presión acústica con ponderación A

p_0 = presión acústica de referencia = 2×10^{-5} Pa

t_1 = tiempo inicial de la medición

t_2 = tiempo final de la medición

Este parámetro se suele medir con un sonómetro denominado integrador, y es ampliamente utilizado para medir ruido, ya que da cuenta de la energía acústica correspondiente a un determinado lapso de tiempo.

También puede emplearse el nivel máximo de presión acústica con ponderación A, $L_{pAmáx}$, que corresponde al nivel sonoro más alto que se registra durante un intervalo de medición. Este parámetro puede medirse con algunos sonómetros convencionales e integradores.

Llama la atención que al hacer referencia en varios artículos [3,4,5,7,8,9] a los niveles producidos por los iPods, se habla de manera genérica de valores en decibeles, sin hacer distinción si se refiere a decibeles con ponderación A o decibeles sin ponderación, aunque por el contexto se infiere que se trata de niveles en dBA. Esta aclaración puede parecer superficial, pero los valores en dBA son usualmente menores que los expresados en dB, debido a la atenuación que sufren los sonidos de bajas frecuencias, como puede observarse en la Fig. 1.

El procedimiento que se propone para la evaluación electroacústica de iPods consiste en medir con un sonómetro convencional, con clase de exactitud 2, el nivel máximo de presión acústica con ponderación A, $L_{pAmáx}$, seleccionando la característica de ponderación temporal rápida, que tiene un constante de tiempo de 125 ms, para dar un mejor seguimiento a la fluctuante dinámica de amplitud de la señal de audio. Es importante que el sonómetro se calibre previamente con un calibrador acústico.

Para simular el conducto auditivo se sugiere utilizar un ducto de plástico suave, colocando en uno de sus extremos el micrófono del sonómetro y en el extremo opuesto el audífono del iPod a evaluar, como se muestra en la Fig. 2.



Fig. 2 Arreglo de medición para la evaluación electroacústica de iPods

A fin de ilustrar los resultados de las mediciones obtenidas con este procedimiento, se utilizaron cuatro iPods de diferentes marcas, seleccionando dos condiciones de volumen del equipo, una

considerando subjetivamente un volumen típico y otra con el volumen del equipo al máximo. Asimismo se tomó como referencia un tema musical de rock, que es uno de los géneros que predominan en el gusto de los jóvenes, según las encuestas. Se consideró conveniente hacer las mediciones al menos dos veces, para comprobar la repetibilidad de los valores obtenidos.

En las Tablas 1 y 2 se muestran los niveles sonoros máximos obtenidos en dBA. Los iPods se identificaron con números.

iPod	$L_{pAmáx}$, dBA
1	100.5
2	88.5
3	101.8
4	92.2

Tabla 1. Resultados. Condición: volumen típico

iPod	$L_{pAmáx}$, dBA
1	111.3
2	101.4
3	112.6
4	105.1

Tabla 2. Resultados. Condición: volumen al máximo.

Con respecto a los resultados de las encuestas, se pueden mencionar los siguientes aspectos relevantes: 1) Destaca el que de 32 estudiantes sólo 2 de ellos no cuentan con reproductores de audio portátiles, lo cual puede dar una primera idea de la población de usuarios. 2) En cuanto al género de los propietarios de estos equipos, 73 % corresponde a hombres y 27 % a mujeres. 3) Tocante a la antigüedad como usuarios, es observable un 37 % para los que usan equipos desde hace más de 3 años, variando en forma descendente hasta un 10 % para aquellos que tienen menos de un año de usarlos. 4) Los géneros musicales que prefieren los estudiantes son predominantemente rock, metal y electrónico. 5) La frecuencia de uso se refleja en 47 % para

quienes emplean los equipos una hora, 27 % para los que lo emplean dos horas, 20 % para los que lo utilizan tres horas, y 7 % para quienes lo usan cinco horas. 6) Con referencia al volumen con el que se escuchan los iPods, el 23 % lo utiliza al 70 % del volumen, en tanto que un 20 % lo usa al 40 % del volumen, presentándose porcentajes del 10 % para quienes utilizan el 90, 80 y 50 % del volumen, y porcentajes menores para otras condiciones de uso. 7) Respecto al tiempo con el que usan dicho volumen, predominan 43 % con una hora, 30 % con dos horas, y 13 % con tres horas. 8) La mayoría, representada por el 80 %, no recuerda haber aumentado el volumen con el que escucha su equipo. 9) Sólo el 23 % se ha percatado si ha aumentado el volumen para escuchar la música. 10) En cuanto a si se considera que el uso de iPods es dañino, un 70 % considera que sí, mientras que el 30 % restante considera que no.

III. CONCLUSIONES Y RECOMENDACIONES

Los resultados de las mediciones de los niveles sonoros máximos en dBA producidos por los iPods evaluados muestran que se encuentran en el intervalo de valores que resultan nocivos para la audición.

Es deseable ampliar el número de mediciones, a fin de documentar mejor cual es el escenario existente en torno al uso de iPods y el riesgo de pérdida de la audición en nuestros jóvenes estudiantes, pues al parecer la población de usuarios es considerable.

Entre tanto, podemos observar algunas recomendaciones. Una medida general para preservar la audición es evitar ruidos que son demasiado fuertes, que están demasiado próximos a nuestros oídos o que duran demasiado.

Asimismo es pertinente reproducir las sugerencias que se dan en una de las referencias [9], curiosamente una publicación especializada en el género de rock, para proteger nuestros oídos:

- 1) Usar protectores auditivos en los conciertos, si los niveles sonoros son muy elevados,

- 2) bajar el volumen del iPod, particularmente si se usa durante un tiempo prolongado,
- 3) conseguir mejores audífonos, ya que algunos utilizan sistemas de control activo de ruido, que permiten reducir el ruido de fondo del entorno,
- 4) dejar descansar la audición después de una exposición a niveles sonoros altos, ya que la audición requiere aproximadamente 18 horas para retornar a la normalidad,
- 5) dejar de fumar, ya que con ello se tendría una condición cardiovascular menos restrictiva, que favorecería el suministro de sangre a las células ciliadas.

A fin de cobrar conciencia acerca de la tragedia que significa quedarse sordo, puede resultar aleccionador un fragmento de una carta que escribiera en Heiligenstadt Ludwig van Beethoven en el que da cuenta de su sordera: "...Y no era posible que yo dijera a los hombres: '¡Habládmeme más alto, gritádmeme, que soy sordo!'. No me hubiera sido posible descubrir entonces la carencia de un sentido que debiera ser en mí más perfecto que en nadie, y que yo he poseído en otro tiempo, en la mayor plenitud, con una perfección que seguramente no tuvieron jamás los mejores de mi oficio. ¡Oh, esto no es posible tolerarlo! ..."

IV. REFERENCIAS

- [1] <http://www.hearnet.com>, Sitio consultado en mayo de 2009.
- [2] Berger H. Elliot et al, Noise and Hearing Conservation Manual, 4th edition, American Industrial Hygiene Association, 1986.
- [3] <http://media.www.dailyvidette.com/media/storage/paper420/news/2005/12/02/OtherCampus/Research.Indicates.Hearing.Loss.May.Be.Caused.By.Ipods-1118902.shtml>, Sitio consultado en abril de 2009.
- [4] National Institutes of Health, Noise and Hearing Loss – Consensus Conference, Journal of the American Medical Association, 263 (23), 1990.
- [5] <http://media.www.unews.com/media/storage/paper274/news2005/10/10/Forum/The-Ipod.Generation.Faces.Hearing.Loss-1017039.shtml>, Sitio consultado en mayo de 2009.
- [6] Norma oficial mexicana NOM-011-STPS-2001, disponible en: <http://www.stps.gob.mx/DGSST/normatividad/noms/Nom-011.pdf>
- [7] <http://www.sciencedaily.com/releases/2005/12/051216191834.htm>, Sitio consultado en mayo de 2009
- [8] <http://www.sciencedaily.com/releases/2008/10/081014093323.htm>, Sitio consultado en mayo de 2009
- [9] Ringen Jonathan, Music Making Fans Deaf, Revista Rolling Stone, 18 November 2005

V. AUTOR

Iluicamina Trinidad Servín Rivas, Ingeniero en Comunicaciones y Electrónica, egresado de la E.S.I.M.E. Zacatenco; se ha desempeñado como docente y ha ejercido la profesión en diversos campos de la Acústica.

PRIMERA ETAPA DE UN SISTEMA CAD PARA EL ANÁLISIS DE MAMOGRAFÍAS

Arturo Zúñiga-López, Joel C. Castán-Herrera, Andrés Ferreyra-Ramírez, Carlos Avilés-Cruz.

Departamento de Electrónica, Área de Instrumentación
Universidad Autónoma Metropolitana--Azcapotzalco
Av. San Pablo No. 180, Col. Reynosa Tamaulipas, México, D.F. 02200, México
Tel. 53189000, correo electrónico: azl@correo.acz.uam.mx

RESUMEN

En el presente trabajo se enfoca en el desarrollo de un sistema CAD en su primera etapa, para el análisis de mamografías digitales, para posteriormente realizar un diagnóstico objetivo; brindando así, la posibilidad para mejorar el diagnóstico y disminuir el número de falsos negativos.

El cáncer en general se presenta actualmente como la tercera causa de muerte en nuestro país y aproximadamente el 15% de las defunciones por cáncer corresponden al mamario.

El sistema actualmente utiliza OpenCV como librerías para el procesamiento de imágenes y NeuronDotNet para implementar redes neuronales (en etapa de desarrollo).

Palabras clave: Mamografía digital, cáncer de mama, sistema de detección de cáncer.

I. INTRODUCCIÓN

El cáncer de mama es actualmente, el más frecuente y el de mayor mortalidad entre las mujeres del mundo [1].

Entre los procedimientos de detección, que incluyen también la autoexploración y el examen clínico, la mamografía es la única técnica que puede ofrecer una detección suficientemente oportuna. En México, donde la mayoría de los tumores malignos de mama son diagnosticados en etapa avanzada [2], las mamografías se han utilizado casi totalmente con fines de diagnóstico.

Para diagnosticar un cáncer se realiza, entre otras pruebas, algún test de imagen, como puede ser una radiografía, ecografía o resonancia magnética. Mediante estos test pueden detectarse zonas con alta sospecha tumoral, cuyo diagnóstico debe confirmarse finalmente mediante la realización de una biopsia. Este tipo de imágenes, sin embargo, no son fáciles de interpretar, lo que provoca que el profesional encargado de analizarlas, a pesar de su experiencia, no sea capaz de detectar en ellas un porcentaje importante de tumores (falsos negativos).

Una posibilidad para mejorar el diagnóstico y disminuir el número de falsos negativos consiste en utilizar sistemas de diagnóstico asistido por computadora o computer-aided diagnosis (CAD). Un sistema CAD analiza la imagen médica y trata de detectar zonas sospechosas de contener alguna anomalía. Estas zonas son marcadas sobre la propia imagen con un doble objetivo: llamar la atención del profesional encargado de analizarla hacia la zona sospechosa y aportar una segunda opinión respecto al diagnóstico.

El cáncer en general se presenta actualmente como la tercera causa de muerte en nuestro país y aproximadamente el 15% de las defunciones por cáncer corresponden al mamario, ocupando el tercer lugar dentro de las causas de muerte por cáncer, pero, el primer lugar de muerte por cáncer en mujeres [3].

II. MAMOGRAFÍA

La mamografía (también llamada mastografía) es una imagen plana de la glándula mamaria obtenida con rayos X. La imagen se forma debido a la diferente atenuación sufrida por los rayos al atravesar los medios que constituyen la mama. La mama representa un verdadero desafío para la obtención de su imagen radiológica debido a que está constituida por tejidos muy similares entre sí y porque las lesiones buscadas por el radiólogo como indicadores de la posible presencia de un tumor son muy pequeñas o muy parecidas al tejido normal. Como estándar de mamografías se obtienen cuatro imágenes de rayos X, con dos vistas en cada una (dos vistas cráneo-caudal (CC) y dos medio-lateral-oblicuo (MLO)). Todas ellas son empleadas para detectar anomalías y valorar su malignidad o benignidad [4].

Los proyectos de diagnóstico asistido por computadora que actualmente se están desarrollando, tienen los siguientes objetivos:

- Mejorar el rendimiento del observador, conociendo previamente el tamaño o la localización hipotética del objeto. Se trata de métodos que dirigen la atención del radiólogo a regiones sospechosas, intentando resolver los problemas de omisión de hallazgos.
- Aportar una segunda opinión sobre el grado de correlación de las características de una imagen, con un diagnóstico determinado.
- Realizar comparaciones detalladas entre varias imágenes, con el objeto de aportar información en una misma exploración (por ejemplo, con y sin contraste) o en varias exploraciones deferidas de un mismo paciente.
- Proporcionar sistemas de autoevaluación y reciclaje, para radiólogos con diversos grados de experiencia y de aprendizaje, y para residentes en periodo de formación.

La clasificación de los distintos tipos de anomalías observables en una mamografía, que pueden ser indicativos de la presencia de un tumor maligno, se listan a continuación.

- Micro calcificaciones: Agrupación de pequeños puntos de gran intensidad (Fig 1).
- Masas circunscritas: Región más o menos circular con alto brillo (Fig 2.).
- Masas radiales: Región con alto brillo y forma estrellada (Fig. 3).
- Otras masas poco definidas: Región con alto brillo y bordes poco definidos (Fig. 4).
- Distorsiones: Distorsión de la estructura normal de la mama (Fig. 5).
- Asimetrías: Diferencias entre la apariencia de ambas mamas (Fig 6).



Fig 1. Micro calcificaciones.



Fig 2. Masas circunscritas.

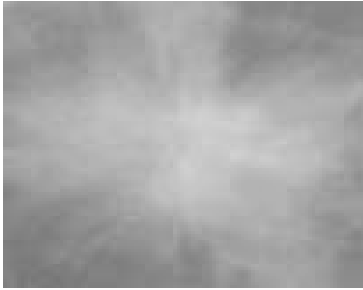


Fig 3. Masas radiales.



Fig 4. Masas poco definidas.



Fig 5. Distorsiones.



Fig 6. Asimetrías.

III. DESARROLLO

El trabajo se enfoca en la creación de un software capaz de realizar un diagnóstico sobre el análisis de una mamografía digital y presentar los resultados al médico para ayudarlo a dar un diagnóstico más objetivo.

Para el diseño del sistema se empleó el Lenguaje Unificado de Modelado (UML por sus siglas en inglés Unified Modeling Language), las etapas del diseño fueron:

- Análisis de cada módulo del sistema, especificando sus funcionalidades.
- Elaboración de un conjunto de diagramas de casos de uso.
- Descripción de cada caso de uso.
- Elaboración de los diagramas de secuencia correspondientes a cada caso de uso.

El software fue desarrollado bajo el lenguaje de programación de Microsoft Visual C#. Para el módulo de modificación de la mamografía, se empleó la librería OpenCV pero para poder ser usada en Visual C#, se necesitó de una API llamada Wrapper (DLL) for Intel's OpenCV Library, esta fue la interfaz entre Visual C# y la librería OpenCV que está desarrollada en C/C++, ya que sin esta API no se habría podido realizar el sistema bajo este lenguaje.

La librería OpenCV (Open source Computer Vision library) [5], es una librería abierta desarrollada por Intel. Esta librería proporciona un alto nivel de funciones para el procesamiento de imágenes. OpenCV implementa una gran variedad de herramientas para la interpretación de la imagen. Es compatible con Intel Image Processing Library (IPL) que implementa algunas operaciones en imágenes digitales. El rasgo esencial de la librería junto con su funcionalidad y la calidad, es su desempeño. Los algoritmos están basados en estructuras de datos muy flexibles, acoplados con estructuras IPL; más de la mitad de las funciones han sido optimizadas aprovechándose de la arquitectura de Intel.

Los riesgos asociados con el desarrollo de sistemas largos y complejos son enormes. Una forma de reducir los riesgos es construir sólo una parte del sistema, reservando otros aspectos para niveles posteriores.

El software en esta primera etapa consta de 3 módulos que se especifican a continuación

Módulo de modificación de la mamografía. En esta parte se manipularán las mamografías digitales con el fin de hacerlas más nítidas, con mayor contraste, obtener contornos y su vector de características.

Módulo de red neuronal artificial (todavía en pruebas). Aquí se tiene implementada la red neuronal artificial con las opciones para ser entrenada o para realizar la clasificación, ésta recibirá como parámetro el vector de características.

Módulo de interfaz de usuario. Es donde estarán contenidos los módulos de modificación de imagen y la red neuronal.

Para hacer pruebas del software se tomaron mamografías de la base de datos pública MIAS (Mammographic Image Analysis Society). Muchos trabajos relacionados usaron esta base de datos [6][7][8]. La base de datos consta de 322 imágenes con una resolución de 1024x1024 píxeles en formato PGM, con diferentes tipos de cáncer (los mencionados ya en líneas anteriores) e imágenes sanas.

El sistema tiene implementado hasta ahora en el primer módulo lo siguiente:

Invertir una imagen: esta opción muestra el negativo de una imagen como se muestra en la Fig. 7.

Histograma: con esta opción podemos ver el histograma de la imagen cargada, esto solo representa el número de píxeles de un mismo color en la imagen (Fig 8).

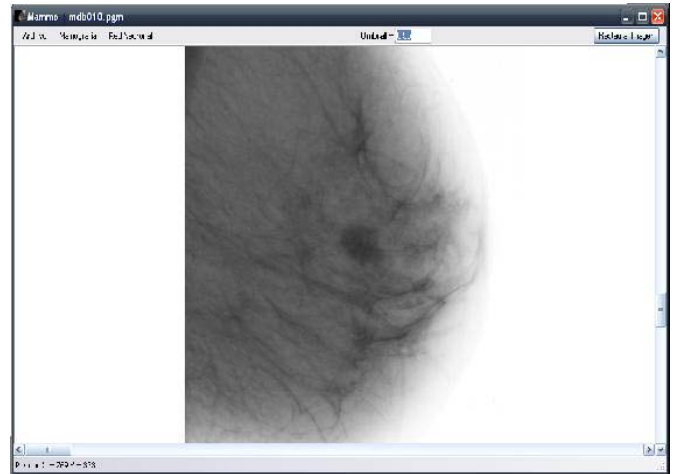


Fig. 7 Inversión de una mamografía.

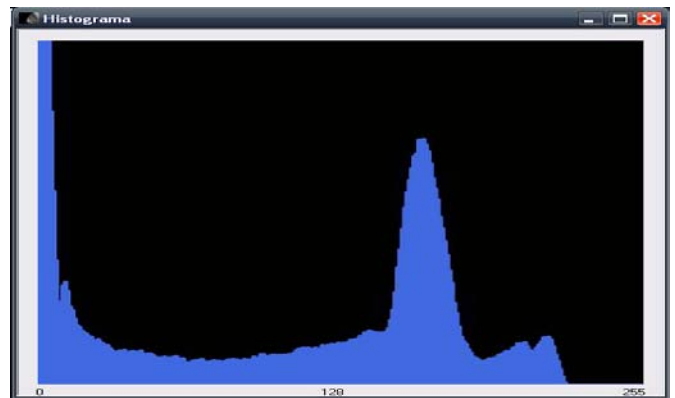


Fig. 8 Histograma de una imagen.

Normalizar: con este se obtiene la imagen normalizada (Fig. 9) de una mamografía (Fig. 10), es decir resalta los bordes, el contraste y aplica filtros para eliminar el ruido. Esta opción realiza los siguientes filtros [9] [10].

- PrewittX
- PrewittY
- Sharpen
- Pasa Bajos
- Pasa Media

El operador de Prewitt otorga el mismo peso a los píxeles contiguos en vertical y horizontal, que a los contiguos en diagonal.

Para realizar el realce propiamente dicho de la imagen empleamos los filtros Sharpen que

constituyen una variación de las máscaras de realce.

Los filtros de paso bajo deben tener todos sus coeficientes positivos y la suma de ellos deben ser igual a uno. Las máscaras que se utilizaron, son de 3x3.

El filtro espacial de media reemplaza el valor de un píxel por la media de los valores del punto y sus vecinos. Su efecto es el difuminado o suavizado de la imagen.



Fig. 9 Normalización



Fig. 10 Imagen Original

Para el módulo de red neuronal artificial, se está empleando la librería NeuronDotNet (esta librería esta diseñada para trabajar en Visual C#), se están haciendo todavía pruebas con una backpropagation

o propagación hacia atrás, esta es una de las redes más empleadas dentro del reconocimiento de patrones. En la Fig. 11 se puede observar su interfaz.

IV. CONCLUSIONES

El programa y la librerías de OpenCv funcionan correctamente y tienen un buen desempeño, sin embargo se pueden tener mejoras en cuanto a la modificación de la imagen ya que los filtros empleados, tal vez no son lo suficientes para obtener una imagen de alta calidad, con bordes bien definidos, con las regiones de interés resaltadas y alto contraste en las partes sanas de la mama.

Aunque en esta primera etapa, del trabajo de investigación ya se tiene un módulo programado (sin todo el potencial de la librería NeuronDotNet) para identificar patrones (utilizando una red neuronal backpropagation), apenas se han obtenido algunos resultados preliminares, el siguiente paso es probar la red neuronal más a fondo y definir bien el número de imágenes para entrenarla, y dado que existe un gran número de redes neuronales artificiales, se puede investigar más para plantear una mejor red o una que mejor se adapte al tipo de procesamiento que se está haciendo.

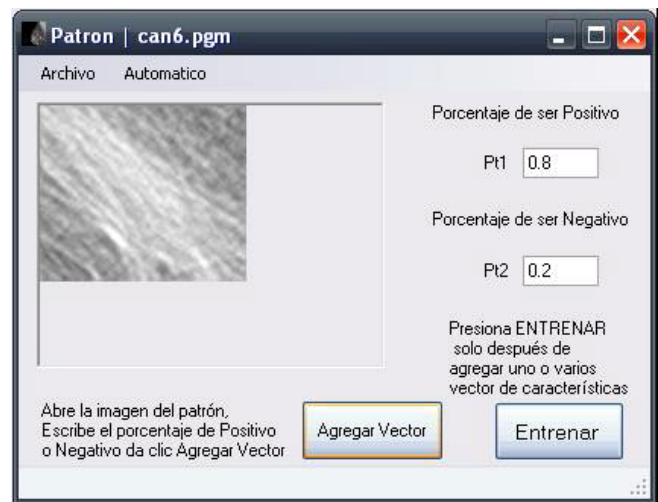


Fig. 11 Módulo de la Red Neuronal

V. REFERENCIAS

- [1] *Cancer Mondial*, International Agency for Research on Cancer, IARC, <http://www.dep.iarc.fr>
- [2] López-Carrillo L, Torres-Sánchez L, López-Cervantes M y Rueda-Neria C, “*Identificación de lesiones mamarias malignas en México*”, *Salud Pub Mex* 2001, 43:199-202.
- [3] Instituto Nacional de Estadística, Geografía e Informática (INEGI), <http://www.inegi.org.mx/inegi/contenidos/espanol/prensa/Contenidos/estadisticas/2008/cancer08.doc>
- [4] Brandan y Villaseñor, “*Detección del cáncer de mama: estado de la mamografía en México*”. *Cancerología* 1 (2006): 147-162.
- [5] Gary R. Bradski y Adrian Kaehler, *Learning OpenCV: Computer Vision with the OpenCv Library*, Edit. O'Reilly, 2008.
- [6] Antonie, M., Zaiane, O, Coman A. “*Application of data mining techniques for medical image classification*”, In In Proc. Of Second Intl. Workshop on Multimedia Data Minino (MDM/KDD'2001) in conjunction with Seventh ACM SIGKDD, 94–101, San Francisco, USA, 2001.
- [7] Karssemeijer N, Brake G. “*Detection of Stellate Distortions in Mammograms*”, *IEEE Transactions on Medical Imaging*. Vol 15, No. 5, October 1996, 611-614.
- [8] Rangayyan R.M, El-Faramawy N.M, Leo Desautels J.E, Alim O.A. “*Measures of Acutance and Shape for Classification of Breast Tumors*”, *IEEE Transactions on Medical Imaging*. vol. 16, no. 6, December 1997. 799-803.
- [9] R. C. González, R. E. Woods, *Digital image processing*, Edit. Addison-Wesley, 2007.
- [10] N. Efford, *Digital image processing: A practical introduction using JAVA*, Edit. Addison-Wesley, 2000.

TEORÍA DE LA RESONANCIA ADAPTIVA PARA EL ANÁLISIS DEL APRENDIZAJE, UTILIZANDO REDES NEURONALES ARTIFICIALES ART1.

Alejandro Cruz Sandoval, Juan Carlos Olgún Rojas, Andrés Ferreyra Ramírez.

Área de Instrumentación, Departamento de Electrónica, División de CBI
 Universidad Autónoma Metropolitana-Azcapotzalco
 Av. San Pablo No. 180, Col. Reynosa Tamaulipas, México, D.F. 02200, México
 Tel. 53189550 ext.1019, correo electrónico: acs@correo.azc.uam.mx

RESUMEN

Una de las características de la mente humana consiste en su habilidad para aprender nuevos conceptos sin necesitar para ello olvidar otros aprendidos en el pasado. Es probable que esta misma capacidad se pueda conseguir en las redes neuronales artificiales. Sin embargo, muchas de estas redes tienden a olvidar informaciones pasadas al tratar de enseñarles otras nuevas. Ya se han abordado el estudio de este tipo de situaciones neuronales con el concepto de perturbaciones singulares en el análisis de la estabilidad-plasticidad, obteniéndose buenos resultados, separando las redes en dinámicas lentas y dinámicas rápidas. El objetivo fundamental en este artículo es utilizar la teoría de resonancia adaptiva para abordar lo que S. Grossberg denomina el dilema de la estabilidad-plasticidad del aprendizaje [9]. Básicamente se diseñó un sistema neuronal que reacciona de forma *plástica* a nueva información recibida, pero que se mantiene *estable* a aquella que es irrelevante. Se utilizó la herramienta de Matlab, como medio de programación y simulación de la red neuronal diseñada, para validar su funcionamiento.

Palabras clave: Elasticidad, Memoria a corto plazo (STM), Memoria a largo plazo (LTM), Plasticidad, Parámetro de vigilancia (ρ), Resonancia, Redes Neuronales ART1.

I. INTRODUCCIÓN

La idea de plasticidad-estabilidad indica la capacidad de los seres vivos a que se pueden adaptar y aprender con extrema facilidad nuevas

situaciones, aunque éstas sean completamente diferentes y cambien drásticamente, al mismo tiempo que son capaces de pasar por alto estímulos irrelevantes. A veces hay ciertas situaciones que es conveniente recordar o almacenar, debido a su importancia para el ser vivo que las padece, aunque no se repitan muy a menudo, o aunque no hayan ocurrido recientemente situaciones de peligro, situaciones de supervivencia, etc. Por ejemplo un elefante es capaz de conducir a la manada a una laguna que nunca se seca, en períodos de gran sequía, aunque haya estado en ella una sola vez, muchos años atrás, y tenga que recorrer grandes distancias.

Las redes neuronales artificiales o un sistema neuronal artificial lo mismo que la redes neuronales biológicas, pueden incluir dos tipos de memoria: la memoria a corto plazo (STM) y que en términos de redes neuronales artificiales se identifica con los valores de activación de las neuronas; y la memoria a largo plazo (LTM) y que se refiere a los pesos o intensidad de las conexiones con las neuronas. El comportamiento de la red es entonces caracterizado por un tipo de red con conexiones hacia delante y hacia atrás (*feedforward/feedback*).

II. MÉTODO.

Descripción Teórica

La descripción matemática básica y un esquema hipotético de este sistema o campo puede ser modelado por las siguientes ecuaciones y la fig.1.

$$v_{ij} = 1 \quad (1)$$

$$w = \frac{1}{1+N} \quad (2)$$

donde v_{ij} es el peso de la conexión *feedback* en la neurona, w_{ij} es el peso de la conexión *feedforward*. La red neuronal ART es una clasificadora de vectores. Un vector de entrada se clasifica dentro de una de un número de categorías dependiendo de la similitud con los patrones previos. Si el criterio de similitud no se satisface se crea una nueva categoría. Así, la red ART resuelve el dilema entre plasticidad y estabilidad al habilitar que el aprendizaje ocurra en un estado de resonancia, es decir, hacer resonar la información de entrada con los representantes o prototipos de las categorías que reconoce la red. Si entra en resonancia con alguno, es suficientemente similar, la red considera que pertenece a dicha categoría y únicamente realiza una pequeña adaptación del prototipo almacenado representante de la categoría para que incorpore algunas características del dato presentado. Manteniendo con esto el equilibrio entre plasticidad y estabilidad en el aprendizaje neuronal.

valores de entrada en las neuronas que conforman las capas de entrada y salida. Otra parte del subsistema de atención, es el denominado control de ganancia, que actúa sobre las neuronas de entrada (STM) para estabilizar el funcionamiento de la red, aumentando la sensibilidad o grado de atención de estas neuronas, evitando una posible anticipación respecto a la llegada de la información del siguiente patrón a través de las conexiones *feedback* (v_{ij}). El subsistema de orientación se utiliza para detectar si la información de entrada pertenece o no a una determinada categoría conocida por la red, para lo cual se calcula el porcentaje de semejanza entre dicha información y el representante (prototipo) de cada categoría almacenada en los pesos v_{ij} (LTM), que deberá ser superior al valor de un parámetro ρ , denominado parámetro de vigilancia. De no ser así, este subsistema reinicia el valor de la neurona vencedora de la capa de salida y prueba otra categoría. Si ninguna es suficientemente parecida, se crea una nueva clase, de la que será prototipo la información de entrada. Es así que, el subsistema de atención realiza la tarea de reconocimiento (búsqueda de la categoría a la que pertenece la entrada), mientras que el subsistema de orientación actúa sobre el primero, permitiéndole responder a entradas no pertenecientes a ninguna categoría, creando para ellas nuevas clases. Por lo tanto, la representación funcional de la figura 1 podría ser representada por medio de la red neuronal de la figura 2, en donde ya se le añadieron dos neuronas que permiten implementar las funciones de control de ganancia y del subsistema de atención. El nodo de reset (R) tiene un funcionamiento especial y consiste en recibir las salidas de las neuronas de entrada a través de conexiones con peso -1 junto con los valores del patrón de entrada a la red a través de conexiones con pesos iguales al parámetro de vigilancia de la red, para compararlo, y si no son suficientemente parecidos, activar la salida correspondiente para, a través de una conexión de inhibición con peso negativo de gran valor, conseguir anular el efecto de la neurona que se ha comprobado que no representa la categoría adecuada para el patrón de entrada. Por otro lado la neurona (G) controla la ganancia de atención.

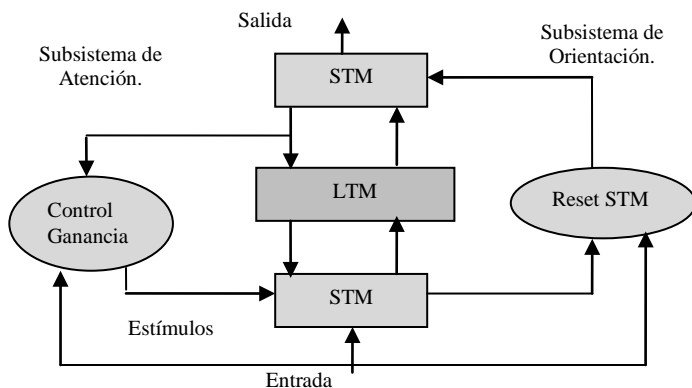


Fig.1 Representación funcional de la red ART

En el subsistema de atención fig. 1, la función básica es el reconocimiento y clasificación de las informaciones aprendidas. En este subsistema se localiza la memoria a largo plazo (LTM), compuesta por los pesos w_{ij} y v_{ij} de las conexiones entre las neuronas de las capas de entrada y salida, que confieren a la red su característica de estabilidad, en el sentido de retener la información aprendida. En la entrada y en la salida de la red, se encuentra el complemento total de la memoria, es decir, la memoria a corto plazo (STM), y que se refiere a la capacidad de recordar aspectos más inmediatos mediante la retención temporal de los

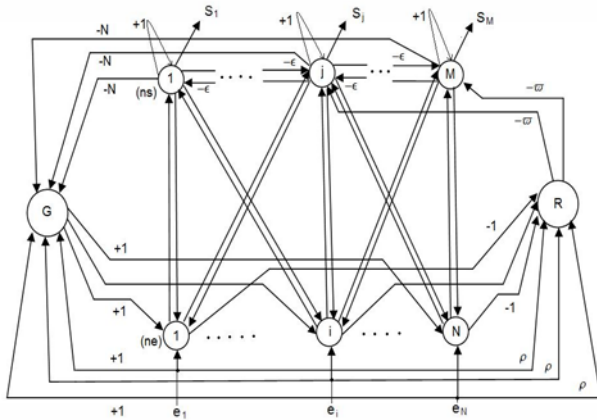


Fig.2 Arquitectura completa de la red ART1.

Para ello recibe como entrada los valores de las neuronas de la capa de salida a través de conexiones con pesos $-N$ y los del patrón de entrada mediante conexiones con peso igual a la unidad, generando, mediante una función de activación escalón, una salida que se aplicará a las neuronas de la capa de entrada de la red para ayudar a la activación de las neuronas de esta capa en función de la cantidad de neuronas de salida activadas. De esta forma, se puede aplicar la regla de activación denominada *Regla $\frac{2}{3}$* , que indica que una neurona de la capa de entrada se activará (salida 1) cuando dos de las tres entradas que recibe patrones de entrada, neurona G y conexiones (v_{ij}) esté activa teniendo valor 1. La entrada total de cada neurona de la capa de entrada puede ser modelada por la ecuación 3.

$$N_{e_i} = e_i^k + \sum_{j=1}^M v_{ij} S_{s_j} + S_G \quad (3)$$

Por lo que se refiere al nodo de reset (R), éste va a tener M salidas, una por neurona de la capa de salida de la red, de las cuales sólo se podrá activar una cada vez, la correspondiente a la neurona que hay que reiniciar.

Es importante mencionar que la función de activación para este nodo está representado por la ecuación 4, donde se involucra el parámetro de

vigilancia ρ y que es importante al momento de entrenar la red.

$$S_R = \begin{cases} \text{Reset de } ns_j & \sum_{i=1}^N \rho e_i^k - \sum_{i=1}^M S_{ns_i} > 0 \\ 0 & \sum_{i=1}^N \rho e_i^k - \sum_{i=1}^M S_{ns_i} \leq 0 \end{cases} \quad (4)$$

De esta forma, se produce el reset de la neurona vencedora cuando se cumple la relación de la ecuación 5.

$$\frac{\sum_{i=1}^M S_{ns_i}}{\sum_{i=1}^N e_i^k} < \rho \quad (5)$$

Funcionamiento y Aplicación de la Red Diseñada

- a.- Se presenta una información (vector) de entrada $E_k = (e_1^k \dots e_N^k)$.
- b.- Cada neurona (ne_i) de la capa de entrada recibe el valor del componente correspondiente (e_i^k) del vector y lo envía a todas las neuronas de la capa de salida a través de las conexiones correspondientes (w_{ji}) .
- c.- Cada neurona (ns_j) de la capa de salida compete con las demás de esta capa hasta que sólo una permanece activa. Las conexiones laterales son las que permiten realizar esta competición, y tienen un peso con un valor fijo $(-\epsilon)$ que debe ser menor que $\frac{1}{M}$, siendo M el número de neuronas de la capa de salida, para que la competición funcione correctamente. Después de cierto número de iteraciones, se llega a una situación de estabilidad en la que una de las neuronas resulta vencedora, generando una salida de valor 1, mientras que en las demás neuronas la salida es 0.
- d.- La neurona vencedora (ns_j) envía su salida (1) a través de las conexiones hacia atrás (v_{ijs}) .
- e.- Se compara la información de entrada $E_k = (e_1^k \dots e_N^k)$ con la información $X = (x_1 \dots x_n) = (v_{ijs} \dots v_{Nis}) = V_{js}$ recibida en la capa de entrada desde la salida a través de las conexiones hacia atrás. Se observa que el vector X y por lo tanto V_{js} es en realidad el representante o prototipo de la clase o categoría j^* -ésima si la

neurona que se ha activado, la vencedora, ha sido la neurona de salida (ns_j). Esta comparación se hace valorando la siguiente relación de semejanza:

$$Rel. Sem. = \frac{\|E_k \cdot X\|}{\|E_k\|} = \frac{\|E_k \cdot V_{j^*}\|}{\|E_k\|}$$

f.- Se compara la relación de semejanza entre ambas informaciones con un parámetro de vigilancia ρ , cuyo valor será fijado por el usuario en un rango de 0 a 1 y que influirá en el número de clases que establecerá la red, ya que cuanto mayor sea su valor, se está pidiendo al sistema que discrimine con mayor precisión, del tal forma que si $\rho = 1$ se estará indicando que una clase sólo puede pertenecer a patrones que sean exactamente iguales.

g.- Si la semejanza evaluada es igual o mayor que el parámetro de vigilancia, entonces se asume que la neurona que se ha activado a la salida es la que representa la categoría más apropiada para la información de entrada E_k , procediéndose a ajustar los pesos (v_{ij}) y (w_{ji}) de la red (memoria a corto plazo) para que el nuevo prototipo almacenado para representar dicha categoría incorpore algunas características del dato de entrada. Como la red ART1 trabaja con valores binarios, esto básicamente consiste en una operación lógica AND entre los bits del prototipo (v_{ij}) y los del patrón de entrada (e_i^k). Por este motivo, inicialmente los pesos (v_{ij}) se inicializan a 1, para que el prototipo coincida exactamente con el patrón de entrada.

La red ART1, por lo tanto, permite establecer clases de datos. Recibiendo datos en la entrada y activando la salida que indique la categoría o clase a la que pertenece. Si en la capa de salida hay N neuronas, el número de categorías que se establecen es N. Los prototipos o representantes de las clases son los pesos (v_{ij}) de las conexiones. El aprendizaje en este tipo de topología neuronal es de tipo *on line*, por lo que no se distingue entre etapa de entrenamiento y de funcionamiento. Los pesos varían durante el funcionamiento de la red cuando se aplica una información de entrada a la misma. Estas también utilizan un tipo de

aprendizaje no supervisado de tipo competitivo, además poseen los dos tipos de aprendizaje denominados aprendizaje lento (STM) y aprendizaje rápido (LTM).

Para comprender y analizar bien el funcionamiento de la red ART1, se procedió a diseñar una red neuronal de éste tipo para clasificar las 26 letras del alfabeto inglés. Las cuales son representadas mediante 26 píxeles, por lo tanto la red está integrada por 26 neuronas de entrada a través de las cuales se recibe la letra que debe clasificar. La red clasifica indicando a qué clase, de las establecidas hasta el momento, pertenece la letra. Si este no fuera el caso, deberá crear una nueva. Se fija un parámetro de vigilancia $\rho = 0.95$, es decir, se exige una semejanza mayor del 90% para considerar que dos letras son de la misma clase. La primera letra que se aplica a la red es la letra ñ que no es del alfabeto inglés y se encuentra la letra más parecida. Como se trata de la primera información, se realiza el aprendizaje completo de la misma, actualizándose los pesos de las conexiones para almacenar este primer ejemplar, que será el representante o prototipo de la primera categoría establecida por la red. La figura 3 es la letra reconocida por la red, mientras que la figura 4 muestra la misma letra pero con ruido. También se muestra en la tabla 1 las categorías encontradas sin variar el parámetro de vigilancia.

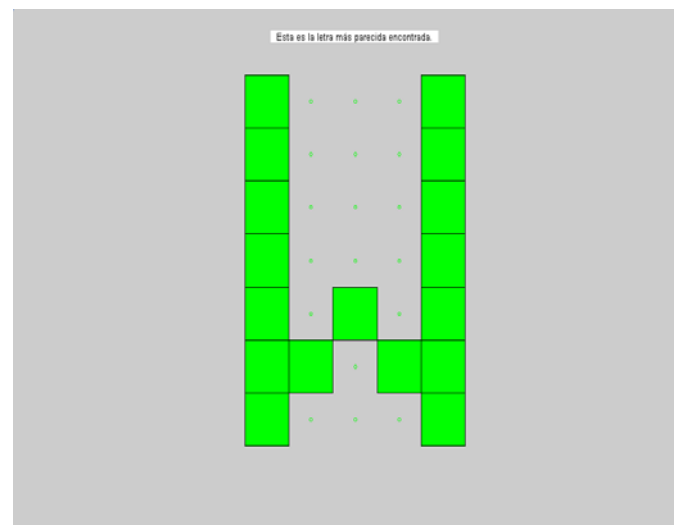


Fig. 3 Letra más parecida.

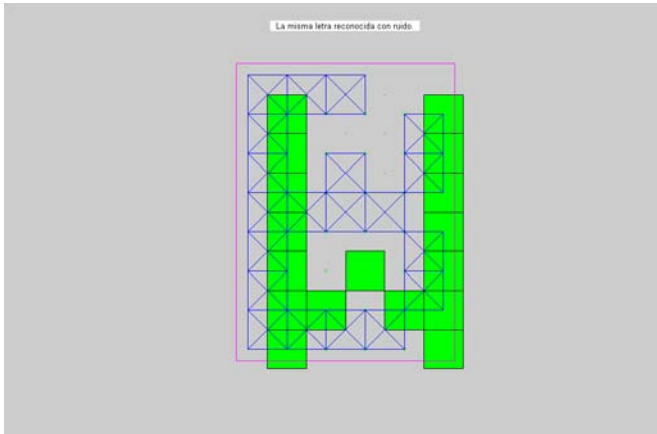


Fig. 4 Letra reconocida con ruido.

Al variar drásticamente el parámetro de vigilancia ρ de 0.95 a 0.75 se obtienen: la letra reconocida figura 5, letra con ruido figura 6 y categorías encontradas tabla 2.

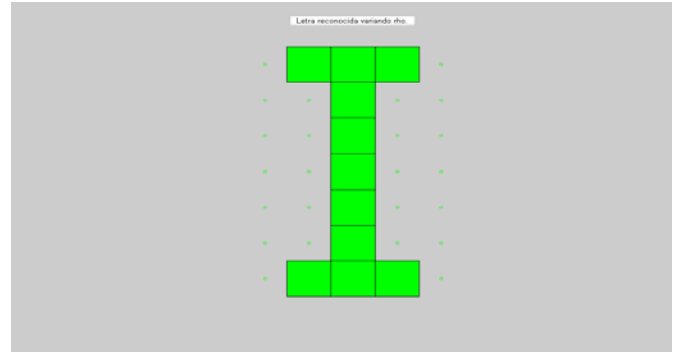


Fig. 5 Letra reconocida con $\rho = 0.75$

Clase	Letras
C ₁	A
C ₂	B
C ₃	C
C ₄	D
C ₅	E
C ₆	F
C ₇	G
C ₈	H
C ₉	I
C ₁₀	J
C ₁₁	K
C ₁₂	L
C ₁₃	M
C ₁₄	N
C ₁₅	O
C ₁₆	P
C ₁₇	Q
C ₁₈	R
C ₁₉	S
C ₂₀	T
C ₂₁	U
C ₂₂	V
C ₂₃	W
C ₂₄	X
C ₂₅	Y
C ₂₆	Z

Tabla 1 Primer categoría obtenida.

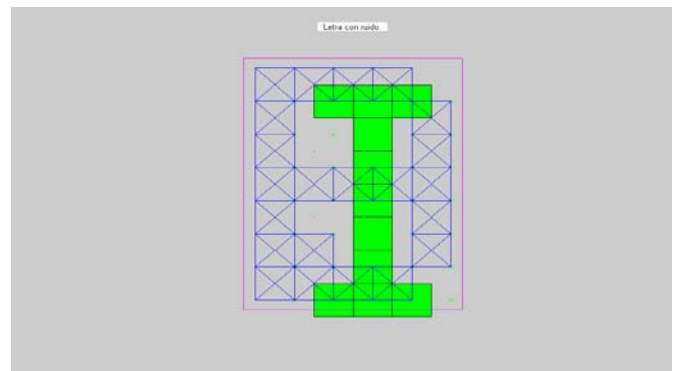


Fig. 6 Letra reconocida con ruido con $\rho = 0.75$

Clase	Letras
C ₁	A
C ₂	B,C,G,L
C ₃	D
C ₄	E,F,P
C ₅	H,M,U,V
C ₆	I,T
C ₇	J
C ₈	K
C ₉	N,W
C ₁₀	O,Q
C ₁₁	R
C ₁₂	S
C ₁₃	X
C ₁₄	Y
C ₁₅	Z

Tabla 2 Categorías con variación de parámetro.

A continuación se utilizó la letra B con dos píxeles invertidos (5.7% de ruido) lográndose la clasificación siguiente: Letra reconocida figura 7, las categorías encontradas son las mismas indicadas en la tabla 1, y en la figura 8 se muestra la letra con ruido al invertir dos píxeles.

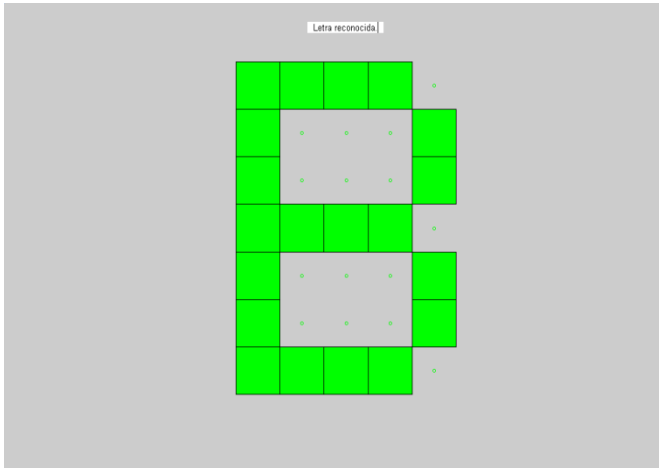


Fig. 7 Letra reconocida.

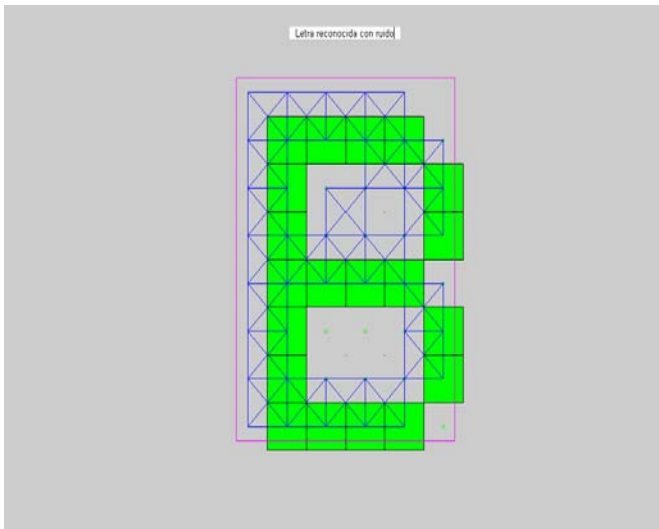


Fig. 8 Letra con ruido.

Finalmente en el experimento se hace una variación más drástica al variar ρ de 0.95 a 0.55, obteniéndose la letra reconocida en la figura 9, y la misma con ruido en la figura 10.

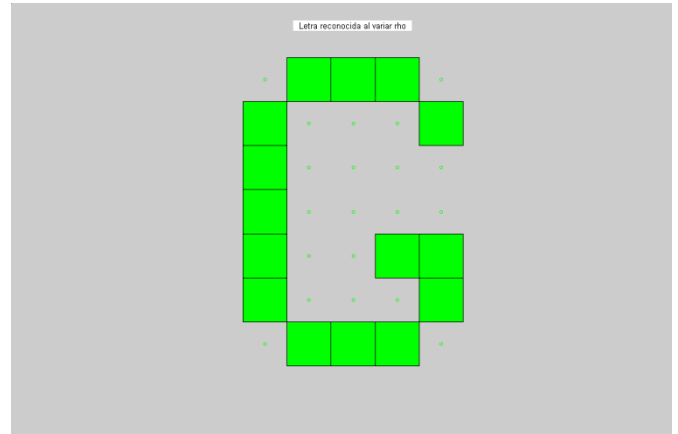


Fig. 9 Letra reconocida con $\rho=0.55$

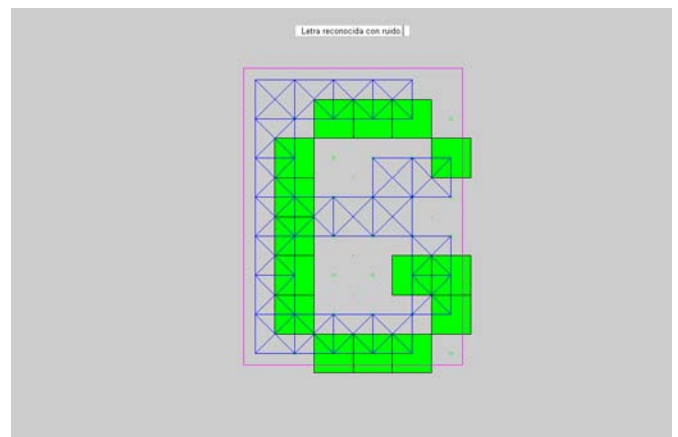


Fig. 10 Letra con ruido.

Finalmente como era de esperarse, se obtienen menos categorías las que se muestran en la tabla 3.

Clase	Letras
C ₁	A,M
C ₂	B,C,E,F
C ₃	D,G,O,S
C ₄	E,F,P
C ₅	H,K,L
C ₆	J
C ₇	N,P,R
C ₈	Q,U,V
C ₉	W,X
C ₁₀	Y
C ₁₁	Z

Tabla 3 Categorías obtenidas con drástica variación del parámetro ρ .

III. CONCLUSIONES

Se ha presentado el concepto de plasticidad-estabilidad de la memoria humana, en sus dos modalidades, es decir, la memoria a largo plazo (LTM) y la memoria a corto plazo (STM), utilizando la arquitectura de red neuronal ART1, por tener la cualidad de separarse en dos dinámicas o subsistemas y que aplican el principio de resonancia entre sus dispositivos neuronales. Se observa perfectamente que un factor importante para lograr un informe más veraz de los que aprende la red, es el parámetro de vigilancia, ya que este tiene un impacto importante en la forma de interpretar el medio circundante. Por otra parte, pese a sus limitaciones y por ser una red con gran sensibilidad ante ruido o distorsión y ante la variación de los datos de entrada, es importante el logro que pueden tener para la clasificación o categorización de datos en el reconocimiento de caracteres o imágenes. También es evidente que el tratar con datos discretos o binarios la hacen más limitada, sin embargo ya se tiene una extensión de esta red, para solventar estos problemas, la red ART2, que es capaz de manejar datos continuos o analógicos, y que en un posterior artículo se abordará plenamente.

IV. REFERENCIAS

- [1] Sánchez Camperos Edgar Nelson y Alanís Gracia Alma Yolanda, *“Redes Neuronales: Conceptos fundamentales y aplicaciones a control automático”* Pearson Prentice-Hall 2006.
- [2] Pedro Isasi Viñuela/Inés M. Galván León *Redes de Neuronas Artificiales Un enfoque práctico.* Pearson-Prentice Hall 2003.
- [3] Madan M. Gupta, Liang Jin, and Noriyasu Homma. *“Static and Dynamic Neural Networks.”* IEEE Press Wiley-Interscience. A John Wiley & Sons, Inc., Publication. 2003.
- [4] José R. Hilera/Victor J. Martínez *Redes Neuronales Artificiales Fundamentos,*

modelos y aplicaciones. Alfaomega-RAMA 2000.

- [5] Simon Haykin *“Neural Networks.”* Prentice Hall, 1999.
- [6] Carpenter, G. A. and S. Grossberg, *“The ART of Adaptive Pattern Recognition by a Self-Organization Neural Networks.”* IEEE Computer, pp. 77-88, Marzo, 1988.
- [7] Carpenter, G. A. and S. Grossberg, *“ART2: self-organization of stable category recognition codes for analog input patterns.”* Applied Optics, vol.26 No.23, Dec. 1987, pp. 4919-4930.
- [8] Carpenter, G. A. and S. Grossberg, *“Absolutely stable learning of recognition codes by a self-organizing neural network.”* American Institute of Physics (AIP) Conference Proceedings 151: Neural Networks for Computing, pp. 77-85, 1986.
- [9] Grossberg S. *“How does a brain build a cognitive code?”* Psychological Review, 87, 1980.

V. AUTORES

Dr. Alejandro Cruz Sandoval obtuvo su título de Doctor en Ciencias con especialidad en Control Automático en el CINVESTAV-IPN.

Ing. Juan Carlos Olguín Rojas es Ingeniero en Electrónica por la U.A.M.I, y actualmente cuenta con el 100% de créditos de maestría en control por la U.N.A.M.

Dr. Andrés Ferreyra Ramírez obtuvo su título de Doctor en Ciencias con especialidad en Control Automático en el CINVESTAV-IPN.

ESQUEMA DE GESTIÓN BASADO EN CALIDAD DE SERVICIO PARA DISTRIBUCIÓN DE CONTENIDO MEDIA STREAMING SOBRE REDES P2P

Guerrero Ibáñez Juan Antonio¹, Ramírez Alcaraz Juan Manuel², Buenrostro Mariscal Raymundo³, Acosta Díaz Ricardo⁴, Contreras Castillo Juan José⁵.

Facultad de Telemática

Universidad de Colima

Av. Universidad No. 333, Col. Las Viboras Colima, Colima, C.P. 28040, México

Tel. 312 3161075, correo electrónico: antonio_guerrero@ucol.mx¹, jmramir@ucol.mx², raymundo@ucol.mx³, acosta@ucol.mx⁴, juancont@ucol.mx⁵.

RESUMEN

En los últimos años, las características de las redes de pares o Peer-to-peer (P2P) las han hecho una alternativa viable para la distribución de contenido multimedia bajo demanda sobre redes IP. Sin embargo este tipo de contenido demanda ciertos requerimientos para dentro del entorno para proporcionar a los usuarios un servicio con calidad. En este artículo se propone un esquema para la gestión de red que entregue servicios de calidad dentro de un entorno de redes P2P. El esquema está formado por un mecanismo de selección de nodos y un modelo de tarificación que proporciona incentivos y beneficios a los nodos colaboradores dentro de este entorno de comunicaciones.

Palabras clave: redes punto a punto, tarificación, incentivos, media streaming.

I. INTRODUCCIÓN

La red de pares o Peer-to-Peer (P2P) es una arquitectura de red lógica de alto nivel construida sobre un conjunto de nodos interconectados a través de una estructura de red física. Estas redes han creado un nuevo paradigma de distribución de contenidos. A diferencia de los sistemas distribuidos tradicionales, las redes P2P son redes auto-organizables formadas por una gran cantidad de computadoras heterogéneas denominadas nodos o peers que colaboran entre sí para la distribución de contenidos a través de la red [1].

En los últimos años, las redes P2P han sido visualizadas como una alternativa viable para

distribuir contenidos multimedia bajo demanda (media streaming) sobre IP. Las aplicaciones media streaming se caracterizan por conservar la relación (variación) de tiempo entre las entidades de información (paquetes) dentro de un flujo de datos, aunque no tiene ningún requerimiento en demandar retardos de transferencia bajos.

La mayor diferencia entre un sistema P2P general y un sistema media streaming P2P radica en el modo de compartición de datos entre nodos. El primer sistema utiliza un modo del tipo “*abrir-después de descargar*” donde el contenido multimedia se ejecuta después de la descarga completa. El segundo sistema utiliza el modo “*reproducir-mientras se descarga*”, en el cual reproduce el contenido (streaming) mientras se descarga.

Los sistemas media streaming P2P son redes auto-expandibles, donde los nodos son heterogéneos, actúan como cliente/servidor, y el servicio dentro de las redes es del tipo multipunto-punto, es decir, varios nodos le proporcionan el servicio simultáneamente a un solo nodo [2].

Sin embargo, este tipo de contenido demanda ciertos requerimientos (retardos inferiores a 200ms., variaciones de retardo menores a 150 ms, tasas de transferencia mínima, tasa de error de bit máxima, etc.) dentro de la red para proporcionarle un buen servicio al usuario, el cual lo podemos denominar calidad de servicio (QoS por sus siglas en inglés Quality of Service).

Existen varios problemas que tienen que resolverse en este entorno para prestar servicios con calidad,

por ejemplo, el procedimiento de selección de nodos activos y la falta de mecanismos de tarificación adecuados.

La selección se convierte en un procedimiento importante debido a la dificultad de seleccionar y coordinar un gran número de nodos activos que proporcionen la calidad de servicio deseada por este tipo de aplicaciones. Por lo anterior podemos decir que la calidad de servicio es dependiente totalmente de la inteligencia de selección de nodos y la monitorización de los enlaces para la detección de degradación de QoS.

De igual forma, existe un problema en relación a la falta de mecanismos dinámicos que tarifiquen el uso de recursos dentro de este tipo de sistemas. El modelo de tarificación más utilizado es la tarifa plana, donde se paga un precio fijo por el acceso al servicio. Los nuevos servicios demandan modelos de tarificación adaptables, dinámicos, que valoren la calidad de servicio ofrecida.

Para diseñar un sistema de tarificación se deben considerar algunos factores importantes. El primero está relacionado con los beneficios que obtienen los receptores con el modelo de tarificación implementado. Es decir, los efectos que tiene la calidad de servicio en el nivel de satisfacción del usuario. El segundo factor radica en la motivación de los nodos para cooperar dentro de este entorno. Este es un problema serio, conocido como “free-riding” [3-4], donde los nodos solamente hacen uso del servicio, pero no colaboran en la distribución de contenidos. Por esta razón surge la necesidad de mecanismos de incentivos que motive a los nodos a participar en la distribución de contenido dentro de este entorno.

En este artículo atacamos estos dos problemas con el objetivo de proporcionar servicios con calidad dentro de un entorno de red P2P. Por un lado proponemos un mecanismo de selección de nodos que permite elegir los nodos adecuados para proporcionar un servicio con calidad. Por otro lado definimos un mecanismo de tarificación, denominado TCdS (Tarificación con Calidad de Servicio), que permite a los nodos transmisores la obtención de beneficios por sus contribuciones y a los nodos receptores beneficios por servicios de calidad y descuentos en situaciones de incremento en el tráfico en la red.

El resto del artículo está organizado de la siguiente forma. La sección II describe en forma breve el trabajo relacionado con este tipo de sistemas. La sección III comienza con la descripción de nuestra propuesta, explicando la arquitectura global de la red media streaming P2P utilizada en el trabajo. La sección IV describe el esquema de gestión de servicios media streaming propuesto. La sección V presenta los resultados obtenidos a través de simulaciones de la evaluación de desempeño del esquema propuesto. Finalizamos el artículo con las conclusiones del trabajo.

II. TRABAJO RELACIONADO

En los últimos años, la aplicación de redes P2P como una alternativa para la distribución de contenido media streaming ha sido objeto de investigación. Producto de esa investigación podemos mencionar algunos trabajos como el presentado en [5] donde los autores proponen el uso de CollectCast para la difusión de media streaming sobre sistemas P2P. Por otro lado, en la referencia [6] se propone un mecanismo de control de congestión usando un modelo de estimación de ancho de banda para la aplicación de media streaming dentro de este entorno. Finalmente, un algoritmo de asignación de tasa TCP-Friendly para múltiples codificaciones de flujos con diversidad de rutas es propuesto en [7].

Con respecto a la tarificación, algunas soluciones han sido propuestas dentro de la literatura. La referencia [8] propone una solución basada en un mecanismo seguro de prepago dentro de entornos P2P. Sin embargo el trabajo se centra en la seguridad más que en especificar el mecanismo de tarificación a detalle. Otros mecanismos como el uso de tarificación basado en fichas han sido propuestos en la literatura [9].

Sin embargo, este trabajo va más allá de esas propuestas, proponiendo un esquema de selección de nodos, y un mecanismo de incentivos que proporcione ganancias a los nodos participantes en la provisión de servicio dentro de un entorno de modelo competitivo.

III. ARQUITECTURA GLOBAL

Con el objetivo de familiarizar al lector con la arquitectura general de un sistema media streaming P2P a continuación se muestra un escenario de aplicación. La Fig. 1 muestra una arquitectura global de sistemas P2P planteando un esquema multipunto-punto, donde existen múltiples nodos transmisores y un solo nodo receptor dentro de un proceso de provisión de servicio.

Se pueden distinguir tres grupos dentro de la arquitectura. El primer grupo está relacionado con los nodos que participan actualmente dentro del sistema media streaming P2P. El segundo grupo está compuesto por los nodos que tienen el contenido media streaming requerido por el nodo receptor y que están a la espera de ser seleccionados para la distribución del contenido. Este grupo lo denominamos *nodos candidatos*. El tercer grupo, denominado *nodos activos*, está formado por los nodos que son seleccionados por el algoritmo propuesto para obtener el contenido requerido.

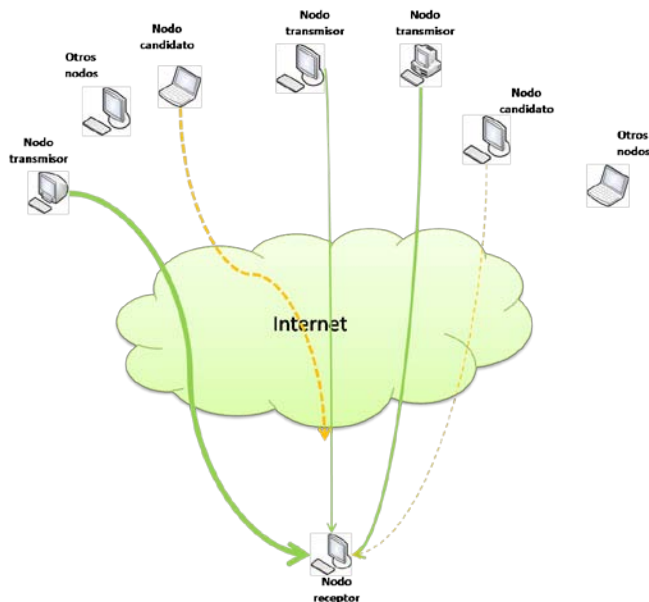


Fig. 1.- Arquitectura global de una red media streaming P2P.

Otro de los puntos importantes es la definición de la topología de red utilizada. La topología de una red P2P se clasifica en: centralizadas como por ejemplo la usada por Napster [10], sin estructura fija y descentralizadas como la usada por Gnutella [11], y finalmente la topología estructurada

descentralizada como la usada en CAN [12]. Para nuestro trabajo, definimos una topología híbrida compuesta por redes centralizadas incrustadas dentro de una red descentralizada, como se muestra en la Fig. 2.

La topología está formada por grupos de nodos (denominados *clusters*) formados por un conjunto de nodos gestionados por un nodo central que denominamos *super-nodo* (SN). Estos SN forman una red en forma de malla, permitiendo un trabajo descentralizado. Los SN controlan a cada uno de los nodos que forman parte del *cluster* a través del almacenamiento de información relacionada con la característica de QoS de cada nodo. Esta información relevante forma lo que denominamos *perfil del nodo*. La información del perfil de nodo es utilizada para el mecanismo de selección, el cual es explicado en secciones posteriores.

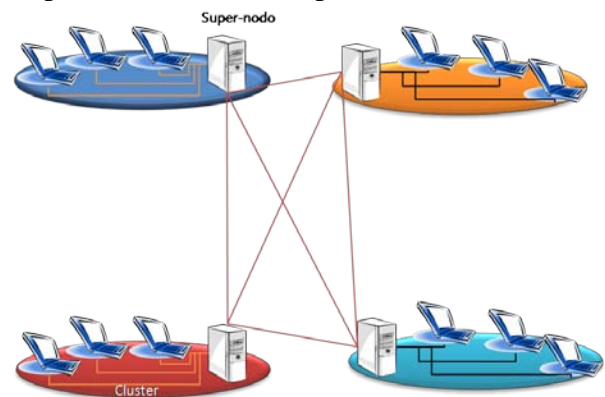


Fig. 2.- Topología media streaming P2P utilizada dentro del presente trabajo.

Dentro de este entorno, existirá una competición entre nodos, haciendo necesario la definición de un mecanismo de incentivos con el objetivo de generar una cooperación a través de la obtención de beneficios por parte de los nodos. De esta forma, en el esquema de incentivos propuesto, se define un *token* como el incentivo monetario para motivar a los nodos para que puedan definir un precio por la contribución de sus recursos. Este mecanismo es explicado a detalle en una sección posterior.

IV. ESQUEMA DE GESTIÓN PROPUESTO

El esquema de gestión propuesto está formado por un mecanismo de selección de nodos y un modelo de tarificación. El mecanismo de selección elige

los nodos adecuados para proporcionar un servicio con calidad. El modelo de tarificación permite a los nodos la obtención de beneficios por la cooperación en el proceso de prestación de servicios. A continuación explicamos a detalle cada uno de los mecanismos que componen el esquema de gestión propuesto.

A. MECANISMO DE SELECCIÓN

Como se describió anteriormente, este trabajo está basado en el modelo muchos-a-uno, en el cual un solo receptor recibe contenido media streaming desde múltiples transmisores en un ambiente P2P. Uno de los principales problemas es el proceso de selección de los nodos transmisores. Algunos métodos utilizados para la selección de nodos son el SBF (Smaller available outbound bandwidth first) que selecciona los nodos con las tasas de transferencia de salida más bajas, y el LBF (Larger available outbound Bandwidth First) que selecciona el nodo con la mayor capacidad de tasa de transferencia de salida a fin de minimizar el número de nodos seleccionados [13].

Recordemos que este es un entorno dinámico, donde el número de nodos puede variar con el paso del tiempo, pudiendo afectar la calidad del servicio que recibe el nodo que demanda el servicio. Algunos factores que pueden afectar el QoS son:

- 1.- Un transmisor detiene su contribución en la provisión de un servicio.
- 2.- Variación en el ancho de banda compartido.
- 3.- La inclusión de nuevos nodos dentro del sistema con mejores capacidades de QoS.
- 4.- Tráfico pesado en la red puede ocasionar una mayor pérdida de paquetes, o variaciones de retardo excesivos, repercutiendo en la calidad de servicio.

Por lo anterior, podemos considerar que la calidad del servicio es dependiente de la selección inteligente de nodos y la monitorización de la misma. En este sentido, proponemos un mecanismo de selección de nodos basado en las preferencias del nodo y el modelo competitivo entre los nodos.

Nuestro mecanismo está compuesto de varios procedimientos los cuales son explicados a continuación.

A.1. Procedimiento de búsqueda

El mecanismo de búsqueda de contenido define el envío, por parte del nodo receptor hacia su supernodo, de una petición indicando el contenido deseado. El supernodo contacta a los otros supernodos para recibir información sobre los nodos que tienen el contenido y el nivel de QoS que pueden ofrecer. De esta forma se crea el grupo de nodos candidatos. Los nodos son filtrados mediante la definición de políticas de calidad de servicio, evitando evaluar nodos que no cumplen con una tasa de transferencia y una variación de retardo mínimos.

A.2. Función Utilidad

A fin de seleccionar los nodos, definimos una función utilidad (1) que evalúa un conjunto de parámetros relacionados con la calidad del servicio y las preferencias del nodo n para la selección de los nodos adecuados para el servicio s [14]. Como se mencionó previamente, las aplicaciones media streaming se caracterizan por una preservación de la relación de tiempo entre los datos enviados en el flujo [15]. Sin embargo, otros parámetros deben ser analizados para determinar posibles problemas como puede ser congestión de la red. Dentro de nuestra propuesta, los parámetros considerados dentro de la calidad de servicio son la tasa de transferencia (b), el precio (p), el retardo (r), y la variación del retardo (v). Por otro lado las preferencias del nodo representan la calidad de servicio y al precio, es decir, para un nodo puede ser más relevante seleccionar los nodos más económicos, sin importar que proporcionen el nivel de calidad óptimo, pudiendo proporcionar un 80%, 20% para el parámetro precio (w_p) y calidad de servicio (w_{QoS}) respectivamente. Sin embargo, otro nodo puede asignar valores diferentes, por ejemplo, 30%, 70% para los mismos parámetros.

$$U_n(S) = w_p \times f_{n,s}(p) + w_{QoS} [f_{n,s}(b) \times f_{n,s}(r) \times f_{n,s}(v)] \quad (1)$$

Donde $f_{n,s}(x)$ para $x=p,b,r,v$ representa la función evaluación (2) que definimos para la evaluación individual de cada uno de los parámetros considerados en nuestra propuesta que proporciona el nodo n para el servicio s . Como cada parámetro tiene requerimientos diferentes, usamos la función

(2a) para la evaluación de la tasa de transferencia debido a que es mejor tener altas tasas de transferencia. Por otro lado, para el caso del precio y variación del retardo necesitamos que los nodos proporcionen niveles de variación de retardo bajos, para lo cual utilizamos la ecuación (2b).

$$f(Q_x, Q_y) = \exp\left[-(Q_x - Q_y)/2Q_y^2\right] \quad (2a)$$

$$f(Q_x, Q_y) = \exp\left[-(Q_y - Q_x)/2Q_x^2\right] \quad (2b)$$

Dentro de estas ecuaciones Q_x y Q_y representan el valor para el parámetro específico proporcionado por el nodo evaluado y el valor deseado respectivamente para el servicio.

La función utilidad generará valores entre 0 y 1. Los nodos serán seleccionados en base al nivel obtenido, y la limitante será el crédito disponible por el nodo, de esta forma se tratará de utilizar el menor número de nodos posibles que cumplan con los requerimientos de servicio y que no sobrepase el precio fijado para el servicio seleccionado.

A.3. Monitorización de QoS

El proceso de monitorización de la calidad del servicio se lleva a cabo tanto en el nodo receptor como en el supernodo del cluster al que pertenece. En caso de detectar alguna variación de la calidad del servicio, ya sea por parte del nodo receptor (el cual es controlado por la aplicación destino) o por el supernodo, en el caso de detección de pérdida de conexión con algún nodo transmisor, se ejecuta el proceso de selección para elegir nodos complementarios a fin de que se pueda recuperar lo más pronto posible la calidad de servicio proporcionada.

B. MODELO DE TARIFICACIÓN

El modelo de tarificación según calidad de servicio (TCdS) está formado por dos pilares fundamentales: el mecanismo de incentivos para nodos transmisores y un mecanismo de control de crédito de los nodos receptores. Dentro de esta sección se explica más a detalle cada uno de los componentes del modelo propuesto.

B.1. Mecanismo de incentivos

Las redes P2P pueden ser visualizadas como un tipo de red dentro de la cual los nodos ofrecen/consumen servicios. Partiendo del concepto anterior, la idea esencial del mecanismo de incentivos es la obtención de ganancias por parte de los nodos participantes a través de la competición entre nodos.

Dentro del mecanismo de incentivos se define el concepto de *token* como el incentivo monetario para motivar a los nodos a participar en la provisión del servicio. El control de crédito es llevado a cabo a través de una autoridad de crédito (AC).

Cada nodo define un precio por la distribución de un contenido media streaming, con lo cual se abre un amplio rango de oferta a los nodos receptores. El precio que define cada nodo refleja su nivel de competición dentro del modelo competitivo. De esta forma, el nivel de competición definido por cada nodo, se verá reflejado en los beneficios obtenidos. Los nodos tienen la libertad de adoptar su propia estrategia de precio.

Como parte del trabajo desarrollado, se definió un modelo de precio basado en demanda como el mecanismo para fijar el precio por el servicio. De esta forma, ésta estrategia calcula el precio en forma dinámica basado en la situación actual del nodo proveedor del servicio. Mediante el uso de ésta estrategia, el precio puede ser incrementado en horas pico con el objetivo de evitar la congestión de la red. El control del precio se realiza mediante la definición de políticas. Los nodos definen una serie de políticas para configurar dinámicamente el precio del servicio en diferentes situaciones. Considerando que los nodos deben proporcionar un servicio de calidad y un precio razonable, las políticas estarán enfocadas a aplicar restricciones. Cada política refleja un índice de precio que permite una competición entre nodos. Dentro del modelo propuesto definimos el índice que refleja el precio real por unidad de QoS, (en nuestro caso es tasa de transferencia, variación del retardo y retardo, de acuerdo a los parámetros relevantes definidos por el 3GPP (3rd. Generation Partnership Project [15]). Este índice se calcula mediante la ecuación (3).

$$I_{precio} = p \times \frac{1}{b} \times r \times v \quad (3)$$

Donde P es el precio designado por el nodo, B es la tasa de transferencia y r es el retardo ofrecido y v la variación de retardo ofrecido. Un alto nivel de QoS (alta tasa de transferencia y una variación del retardo baja), y un bajo precio permitirá al nodo tener un alto nivel de competición. La tabla 1 muestra un ejemplo de relación de perfiles de nodos que es mantenida dentro del supernodo.

ID_nodo	Parámetros del perfil			I_{precio}
	B (Kbps)	r (ms)	P (\$)	
Id-1	150	100	0.6	0.4
Id-2	120	130	0.4	0.43
Id-3	150	90	0.9	0.54

Tabla 1.- Registro de perfiles de nodos conectados al supernodo.

El valor del I_{precio} es utilizado como uno de los factores que se evalúan en el mecanismo de selección de nodos.

El precio se define mediante un perfil de nivel de servicio definido en cada nodo. Todos los nodos parten de una tarifa nominal definida para el servicio a partir de cierta calidad de servicio dependiendo el tipo de servicio (video, audio, etc.). El perfil está basado en categorías subjetivas que representan niveles de degradación de la calidad del servicio. Cada nodo define un número de categorías a fin de generar un nivel de competición contra el resto de nodos cuando la probabilidad de congestión se incrementa. La tabla 2 presenta un ejemplo de perfil de nivel de servicio definido por un nodo.

Categoría	Parámetros QoS			P (\$)
	B (Kbps)	r (ms)	P (\$)	
Excelente	150	100	Tarifa nominal	
Buena	120	130	-20%	
Regular	90	180	-40%	

Tabla 2.- Perfil de nivel de servicio definido por un nodo.

B.2. Mecanismo de Control de Crédito

Como se mencionó anteriormente, el mecanismo de pago está relacionado con el nivel de calidad de

servicio recibido por el nodo receptor. La gestión de cobro la realiza una autoridad de crédito (AC). El modelo que se sigue es un pago fuera de línea a fin de no impactar en la latencia de la red, el costo computacional y el costo de comunicación. Cada nodo tiene un crédito registrado en la AC. De esta manera los nodos podrán gastar su crédito cuando hagan uso de un servicio, e incrementar su crédito cuando participan en la provisión del servicio.

Para el control de crédito, cada nodo tiene un identificador único y una cuenta personal ($pa_n=[Id_n, C_n]$), donde Id_n es el identificador de nodo, y C_n es el crédito actual del nodo. Esta cuenta es usada por el AC para el control del servicio. El proceso de tarificación y pago se muestra en la Fig. 3.

El pago de crédito se realiza inmediatamente después de que el servicio finalice. A través de registros que contienen información sobre la contribución de cada nodo transmisor (duración, costo, nivel QoS) se calcula el cobro que se hará al nodo receptor y el pago por la contribución a cada nodo receptor. A fin de evitar el repudio, el supernodo lleva un registro similar de cada nodo transmisor dentro de su cluster.

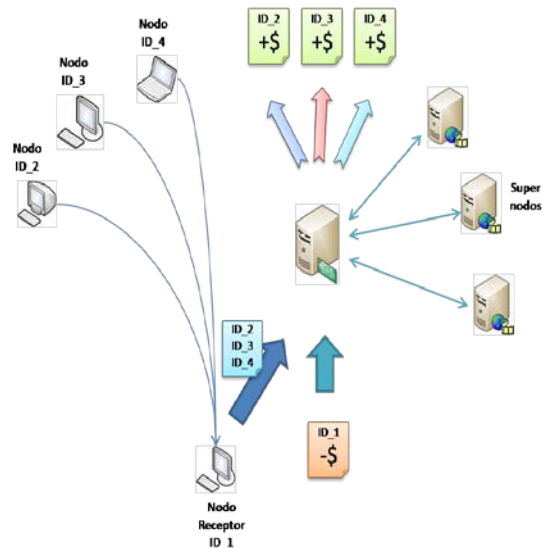


Fig. 3.- Proceso de tarificación de servicios.

El nodo receptor envía al AC el registro final del servicio con el registro de contribución de cada nodo. De esta forma, como medida de seguridad el AC verifica la información del registro mediante un comparativo del registro almacenado en cada supernodo. Como cada nodo transmisor define sus

propias tarifas, el pago que se realiza a cada nodo transmisor se calcula en base a la ecuación (2).

$$Pn_i = \sum I_{precio} \times N_{QoS} \times t \quad (2)$$

Donde I_{precio} representa el índice de precio del nodo, N_{QoS} representa el nivel QoS proporcionado y t representa el tiempo en que proporcionó ese nivel de servicio.

Finalmente, el cargo que se hará al nodo receptor será calculado por la ecuación (3).

$$D = \sum_{i=1}^n Pn_i \quad (3)$$

V. EVALUACIÓN Y RESULTADOS

En esta sección evaluamos el modelo propuesto. Se llevaron a cabo simulaciones intensivas a través de un simulador de eventos discretos desarrollado en C/C++ para validar los resultados [16]. A continuación se describe la topología utilizada en las simulaciones y se presenta un análisis de los resultados obtenidos.

A. DESCRIPCIÓN DEL SISTEMA SIMULADO

En el presente trabajo definimos una topología de simulación compuesta por 10 clusters con 20 y 30 nodos cada uno. 15 tipos diferentes de archivos media streaming se distribuyeron aleatoriamente entre los nodos de los diferentes clusters. Los enlaces de cada uno fueron definidos aleatoriamente por conexiones de 1 y 2 Mbps. Un enlace entre supernodos de 2Mbps. Definimos un precio nominal de \$20 por el servicio.

Cada nodo define un perfil de nivel de servicio que especifica el precio dependiendo el nivel de servicio. El perfil de servicio es generado en cada nodo en forma aleatoria pudiendo generar de 2 a 4 categorías asignando diferentes capacidades y precios obtenidos mediante una distribución logarítmica.

Los nodos receptores son generados en forma aleatoria mediante una distribución de Poisson con media de 5 a 30 peticiones por minuto. Finalmente analizamos el impacto de la estrategia de precio en las ganancias totales de un nodo, el gasto promedio por nodo, el impacto del tráfico en la calidad de servicio, y el número de nodos seleccionados a

través de nuestro mecanismo de selección en comparación con otros métodos utilizados.

B. RESULTADOS OBTENIDOS

A continuación se presentan los resultados obtenidos en la evaluación del sistema de tarificación propuesto.

La figura 4 muestra los resultados obtenidos de las ganancias promedio obtenidas por los nodos transmisores aplicando el modelo de incentivos propuestos y el comparativo con el modelo de tarifa plana.

Como se puede observar, el modelo de incentivos propuesto obtiene mejores resultados que la tarifa plana. Sin embargo, cuando el número de categorías es mínimo, su comportamiento es muy similar a la tarifa plana, con un aumento promedio del 5% de las ganancias. Por otro lado, los resultados muestran que conforme el número de categorías se incrementa, se obtiene un aumento en las ganancias de hasta el 21% en situaciones de poco tráfico y hasta un 28% en situaciones de alta demanda en la red.

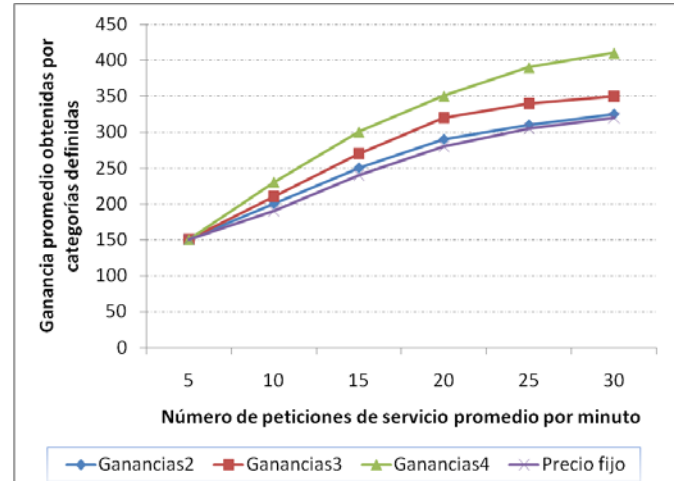


Fig. 4.- Comparativo de ganancias promedio entre el modelo de incentivos propuesto y la tarifa plana.

Por otro lado, la figura 5 presenta el gasto promedio de un nodo durante uso del servicio a través de la aplicaciones de los dos modelos de tarificación analizados. Los resultados muestran que el modelo de tarificación TCdS y el modelo de tarifa plana tienen un comportamiento similar en situaciones de poco tráfico. No obstante, cuando el

número de peticiones en la red se incrementa, y la calidad de servicio se degrada, el modelo TCdS obtiene un decremento en el gasto promedio del nodo en comparación al modelo de tarifa plana, con beneficios de hasta un 22% de ahorro para el nodo.

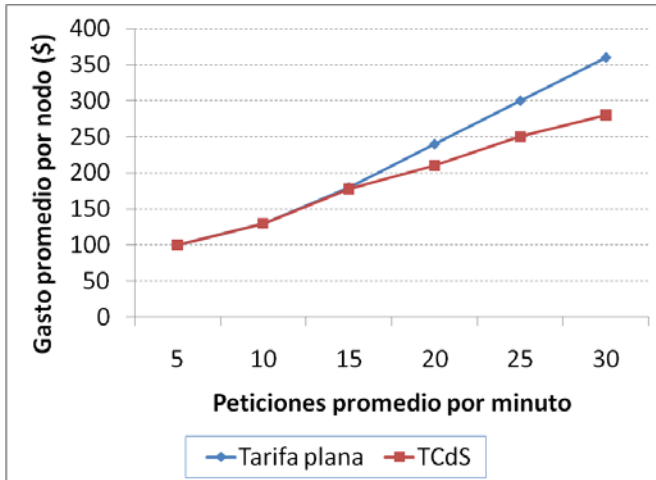


Fig. 5.- Comparativo de gasto promedio por nodo.

Finalmente, la figura 6 presenta el impacto que tiene el tráfico en el proceso de selección de nodos transmisores aplicando los diferentes algoritmos mencionados.

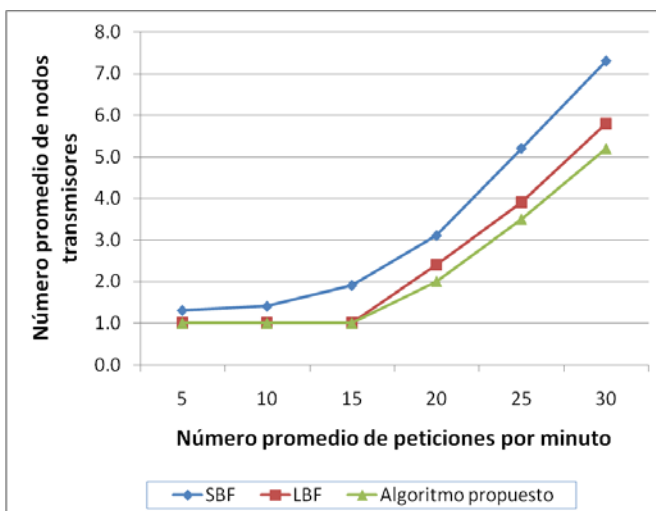


Fig. 6.- Impacto de las preferencias de nodo en el número promedio de nodos seleccionados.

Se puede observar que en situaciones de solicitudes de servicio bajas, la relación de nodos seleccionados no varía mucho entre los métodos LBF, SBF y el mecanismo propuesto. Sin embargo, conforme la cantidad de peticiones

aumenta, se obtiene un mejor desempeño del mecanismo propuesto en comparación a LBF y SBF, disminuyendo en un 29% y 10% respectivamente el número promedio de nodos seleccionados. Esto se debe a que se evalúan mayor número de parámetros en comparación de los otros mecanismos, seleccionando los nodos esenciales en base a las preferencias de los nodos.

VI. CONCLUSIONES

En este trabajo se propuso un modelo de gestión basado en calidad de servicio para provisión de contenidos media streaming a través de redes punto a punto. Los perfiles definidos en el modelo propuesto permiten a los nodos configurar su nivel de competición mediante la especificación de precios por el servicio basado en categorías subjetivas de degradación de la calidad de servicio. Este modelo genera un entorno de competición que permite a los nodos obtener beneficios con una estimación de precio apropiado.

Por otro lado, el mecanismo de selección de nodos permite la provisión de servicios de calidad basados en las preferencias de los nodos a través de la selección adecuada de los nodos que pueden prestar el mejor nivel posible, evitando una degradación de la calidad de servicio percibida por el usuario.

Finalmente se demostró que un modelo de tarificación basado en la calidad de servicio es una alternativa viable para entornos P2P ya que obtiene una serie de beneficios tanto para los nodos receptores primeramente mediante la percepción de servicios de calidad y en segundo lugar un ahorro económico en situaciones de gran demanda. Por otro lado, los nodos transmisores obtienen beneficios a través del mecanismo de incentivos por su participación en la prestación de un servicio, y esto motiva la participación de los nodos, lo que podría ayudar en la disminución del fenómeno free-riding que se presenta en este tipo de entornos.

VII. REFERENCIAS

- [1] Y.Liu, Z. Zhuang, X. Li, N. Lionel, "A distributed approach to solving overlay mismatch problem", In proceedings of the

- 24th IEEE International Conference on Distributed Computing Systems (ICDCS2004), pp. 132-139, 2004.
- [2] D. Xu, M. Hefeedas, S. Hambrusch, B. Bhagava, "On Peer-to-Peer media streaming", Purdue Computer Science Technical Report, 2002.
- [3] E. Adar, B. Huberman, "Free riding on Gnutella", Technical report, Xerox PARC, 10 August 2000.
- [4] S. Saroiu, P.K. Gummadi, S.D. Gribble, "A measurement study of peer-to-peer file sharing systems", In Proceeding of Multimedia computing and networking (MMCSN'02), 2002.
- [5] M. Hefeeda, H. Habib, B. Botev, D. Xu, B. Bhagava, "Promise: Peer-to-peer media streaming using Collectcast", ACM MM'03, Berkeley, CA, November 2003.
- [6] N. Aboobaker, D. Chanady, M. Gerla, M.Y. Sanadidi, "Streaming media congestion control using bandwidth estimation", In proceedings of 5th. IFIP/IEEE International Conference on Management Multimedia Networks and Services: Management of Multimedia on the Internet, pp. 89-100, vol. 2496, 2002.
- [7] J. Kim, R.M. Mersereau, Y. Altunbasak, "Network Adaptive Video Streaming using multiple description coding and path diversity", In proceeding of 2003 International Conference on Multimedia and Expo, vol. 2, pp. 653-656, July 2003.
- [8] B. Emmert, O. Jorns, "P²P – Prepaid Peer-to-peer services", Sixth International Conference on Peer-to-peer computing P2P'06, pp. 223-224, September 2006.
- [9] D. Hausheer, N. Liebau, A. Mauthe, R. Steinmetz, B. Stiller, "Token-based accounting and distributed pricing to introduce market mechanisms in a peer-to-peer file sharing scenario", In proceedings of the Third International Conference on Peer-to-peer computing (P2P'03), 2003.
- [10] Napster, <http://www.napster.com>
- [11] Gnutella, <http://www.gnutella.com>
- [12] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shen, "A scalable content-addressable network", Proceeding of the 2001 Conference on Applications, technologies, architectures, and protocols for computer communications, 2001.
- [13] H. Kim, S. Kang, H. Yeom, "Node selection for a fault-tolerant streaming service on a peer-to-peer network". In proceedings of International Conference on Multimedia and Expo 2003, 6-9 July 2003.
- [14] J.A. Guerrero Ibáñez, "Contribución a la gestión de la tarificación basada en políticas según calidad de servicio: aplicación a redes de comunicaciones móviles avanzadas heterogéneas", tesis doctoral, Barcelona, España, Septiembre 2008.
- [15] 3rd. Generation Partnership Project (3GPP), "Quality of Service (QoS) concept and architecture (Release 8)" Technical Specification 23.107 v.8.0.0, December 2008.
- [16] Omnet++, <http://www.omnetpp.org>.

VIII. AUTORES

D. en C. Juan A. Guerrero Ibáñez recibió el título de Ingeniero en sistemas computacionales en 1996 y el grado de maestro en ciencias en el área de Telemática en 1999, ambos por la Universidad de Colima. Obtuvo el grado de Doctor Ingeniero en Telemática en la Universidad Politécnica de Cataluña en Barcelona, España. Desde 1997 es profesor-investigador de tiempo completo de la Facultad de Telemática de la Universidad de Colima. El doctor Guerrero ha participado como ponente en varios congresos a nivel nacional e internacional, y tiene publicados varios artículos de carácter nacional e internacional. Sus principales áreas de interés son control y gestión de red y servicios, la gestión basada en políticas, y las redes de comunicaciones móviles avanzadas.

M. C. Juan Manuel Ramírez Alcaraz recibió su título de Ingeniero en Sistemas Computacionales en 1998 y su grado de Maestro en Ciencias área Telemática en 2001, ambos por parte de la Universidad de Colima, en la ciudad de Colima, Colima, México. Ha sido profesor durante 20 años en la Universidad de Colima y desde 2004 se desempeña como Profesor-Investigador de Tiempo Completo en la Facultad de Telemática de la misma Universidad. Cuenta con la certificación CCNA de Cisco y actualmente es alumno del doctorado en ciencias de la computación en el CICESE, Ensenada, México. El maestro en Ciencias Ramírez ha participado como ponente en varios congresos a nivel nacional e internacional, y tiene publicados varios artículos de carácter nacional e internacional. Sus principales áreas de interés son la gestión de redes y el cómputo de alto rendimiento (Clusters y Grids).

M. C. Raymundo Buenrostro Mariscal obtuvo el grado de ingeniero en comunicaciones y electrónica en 1992 en la Universidad de Colima. Obtuvo el grado de maestría en ciencias área telecomunicaciones en 1996 en el Centro de Investigación Científica y de Educación Superior de Ensenada. Ha ejercido profesionalmente como catedrático de la Facultad de Telemática de la Universidad de Colima desde 1996 como profesor-investigador de tiempo completo. Ha sido Director General de Servicios Telemáticos de la misma universidad desde 1998 hasta mayo de 2008. El maestro en ciencias Buenrostro ha participado como ponente en varios congresos a nivel nacional e internacional, tiene publicados una serie de artículos de carácter nacional e internacional, prototipos tecnológicos y capítulos de libros. Sus principales áreas de interés son la seguridad y gestión de redes, así como servicios de comunicación de voz, datos y video por Internet.

D. en C. Juan José Contreras Castillo obtuvo su licenciatura en sistemas computacionales por la Facultad de Ingeniería Mecánica y Eléctrica de la Universidad de Colima en 1996 y la maestría en Ciencias de la Computación por el Centro de Investigación Científica y Educación Superior de Ensenada (CICESE) en 1998. Es Doctor en Electrónica y Telecomunicaciones por el CICESE. El doctor Contreras ha participado como ponente en varios congresos a nivel nacional e internacional, y tiene publicados varios artículos en revistas arbitradas, es autor de varios capítulos de libros, así como de un libro sobre el uso de tecnologías para el apoyo a los migrantes Colimenses en Estados Unidos. Sus líneas de investigación son: la educación a distancia y la computación móvil para apoyar los procesos de enseñanza – aprendizaje.

M en C. Ricardo Acosta Díaz es Ingeniero en Sistemas Computacionales por el Instituto Tecnológico de Colima. Obtuvo la Maestría en Ciencias de la Computación en el Centro de Investigación Científica y Educación Superior de Ensenada en el año 2000, actualmente cursa el tercer año de Doctorado en Ingeniería Telemática en la Universidad de Vigo, España. El Maestro Acosta ha participado como ponente en varios congresos a nivel nacional e internacional, y tiene publicados varios artículos en revistas arbitradas, es autor de varios capítulos de libros, así como de un libro sobre el uso de tecnologías para el apoyo a los migrantes colimenses. Sus líneas de investigación son: el cómputo ubicuo, la informática médica y las bibliotecas digitales. Actualmente es Profesor-Investigador en la Universidad de Colima.

HERRAMIENTA VISUAL PARA ANALIZAR SECUENCIAS GENÓMICAS

Verduzco Reyes Gustavo, Bautista Thompson Ernesto, Tass Herrera Benjamín

Cuerpo Académico Ciencias de la Computación, Dependencia Área Ciencias de la Información
Universidad Autónoma del Carmen (UNACAR)
Av. 56 Núm. 4, CP. 24180, Cd. Del Carmen-Campeche
Tel. (938)3811018 Ext. 1506 y 1507, correo electrónico: {gverduzco,ebautista,btass}@pampano.unacar.mx

RESUMEN

En este artículo, se aborda la construcción de una herramienta computacional para analizar de forma visual el comportamiento de secuencias de ADN usando herramientas de bases de datos, descomposición y graficación de datos. El documento primero abordará la estructura interna de la herramienta, luego la interfaz gráfica del usuario y finalmente las pruebas y conclusiones.

Palabras clave: ADN, Base de Datos.

I. INTRODUCCIÓN

En la actualidad se ha intensificado el análisis de secuencias de ADN (Fig. 1) debido a que este contiene la información necesaria para el desarrollo y funcionamiento de un organismo. En su origen, el ser humano, al igual que las plantas y los animales pluricelulares, es una sola célula. Cuando esta alcanza cierto tamaño, se divide en dos, luego en cuatro, y así sigue el fraccionamiento, durante el cual van diferenciándose las células en musculares, nerviosas, cutáneas, etc. Más tarde se agrupan para formar tejidos (por ejemplo, un conjunto de células musculares produce tejido muscular). A su vez, los diversos tejidos forman órganos, como el corazón, los pulmones y los ojos. Bajo la delgada cubierta de cada célula hay un fluido gelatinoso llamado citoplasma. Luego, separado de este por una fina membrana, viene el núcleo, que se considera el centro de control, pues dirige casi todas las actividades de la célula. En él se encuentra el programa genético de la célula, “escrito” con ADN (ácido desoxirribonucleico). Las moléculas de

ADN se hallan enroscadas estrechamente en los cromosomas de las células. Los genes, que son segmentos de las moléculas de ADN, poseen toda la información necesaria para generar un organismo completo. “El programa genético que porta el ADN diferencia a cada ser vivo de los demás. Este programa logra que el perro sea distinto del pez; la cebra, de la rosa, y el sauce, de la avispa. Hace que uno sea diferente de toda otra persona que vive en la Tierra” [1].

Las “letras” (unidades fundamentales) del idioma del ADN son los componentes químicos llamados nucleótidos. Hay cuatro tipos de nucleótidos: timina, adenina, guanina y citosina, que suelen abreviarse T, A, G y C. “Estos cuatro nucleótidos deben verse como las unidades de un alfabeto de cuatro letras. Al igual que combinamos las letras del abecedario para formar palabras con sentido, también combinamos las A, T, G y C que componen nuestros genes para constituir ‘palabras’ de tres letras que resulten comprensibles a la maquinaria celular” [2]. Además, con las “palabras” genéticas se construyen “oraciones” que le indican a la célula cómo elaborar una determinada proteína. El orden en que estén dispuestas las letras del ADN determina si la proteína funcionará como una enzima que nos ayude a hacer la digestión, como un anticuerpo que combata una infección o como cualquier otra de los miles de proteínas del organismo.

Las investigaciones recientes permiten determinar que pequeños segmentos de ADN llamados **nucleosomas** son en parte los responsables de determinar el comportamiento de toda una

secuencia de ADN (regulación de expresiones de los genes y empaquetamiento de la cadena de ADN), de ahí entonces la importancia de su estudio [3, 4, 5, 6, 7]. Si se pudiera determinar cómo se activa un gen entonces se podrían generar medicamentos más efectivos para combatir enfermedades como la diabetes o la hipertensión, las cuales ocurren como resultado de la alteración de un conjunto de “instrucciones” o nucleótidos de letras en el ADN.

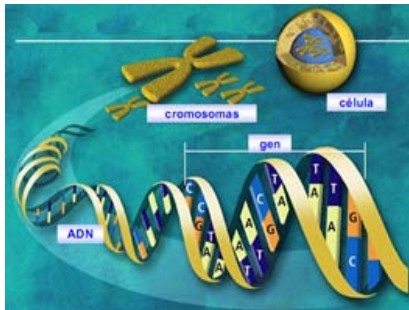


Figura 1. El ADN en la Célula

II. PROBLEMA

Para un Biólogo Molecular analizar secuencias de ADN para encontrar **zonas nucleosomales** es una tarea monótona y cansada en vista de que un organismo puede contener millones de nucleótidos que conforman el ADN. El estado del arte actual nos indica que hay varias herramientas para analizar cadenas de ADN tales como *Knob Linux al alcance de todos*, *Genome Projector*, *E-Cell*, *AJAX Generic Genome Browser*, *Galaxy*, *Endeavour*, *Dasty2* [8] y *BLAST* [9]; sin embargo no están enfocadas a la búsqueda específica de zonas nucleosomales sino a la visualización genérica de las distintas relaciones que guardan los nucleótidos que componen al ADN. La herramienta computacional que se presenta a continuación permite al investigador en genómica analizar patrones específicos de forma gráfica en secuencias del ADN de algún organismo [6]. Esto se logra mediante la separación de las secuencias de ADN en dímeros (pares de nucleótidos).

III. ARQUITECTURA DE LA HERRAMIENTA

La herramienta fue programada usando como lenguaje de programación Java 1.6, también se empleó un IDE para el lenguaje llamado Net Beans Ver. 5.5, con licencia “dual license consisting of the Common Development and Distribution License (CDDL) v1.0 and the GNU General Public License (GPL) v2”. Se utilizó para el manejo de base de datos FireBirds 2.0 y como Administrador de la Base de datos FlameRobin 0.8.0 Unicode bajo licencia “InterBase Public License” [10].

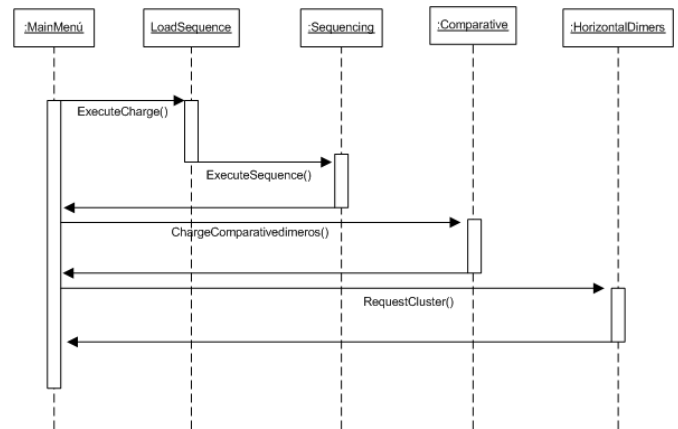


Figura 2. Arquitectura de la Herramienta

En la Fig. 2 se observan las distintas clases que componen la herramienta. Una de las ventajas de esta herramienta es que explota el manejador de BD para extraer información cuando normalmente se usa un lenguaje de alto nivel, manejando cadenas para extraer información relevante de secuencias de ADN.

Los módulos se describen con detalle a continuación:

Clase LoadSequence.- Esta clase se encarga de leer un archivo plano que corresponde con secuencias de ADN y las vacía en la base de datos. El proceso consiste en primero abrir el archivo, tal como se muestra en el primer recuadro y luego según el segundo recuadro, se divide la secuencia en 2 tokens o elementos, el primer token es la clave de la secuencia y el segundo token corresponde a la secuencia en sí mismo.

```
try {
```

```
FileReader fr = new  
FileReader(nombreArchivo.getText());  
BufferedReader br = new BufferedReader(fr);  
String s,codigo,sec;  
codigo="";sec="";  
this.GenerarTabla();  
Conexion conectar = new Conexion();  
Connection cone10 = conectar.AbrirConexion(); //  
Se abre conexion a la BD.  
// Se realiza la inserción de datos  
Statement consulta = cone10.createStatement();
```

```
while((s = br.readLine()) != null) {  
    StringTokenizer st = new  
StringTokenizer(s);  
    codigo = st.nextToken();  
    sec = st.nextToken();  
    areaSecuencia.append(s+"\n");  
    consulta.executeUpdate("INSERT into  
"+nombreTabla.getText()+"(ID_SEC,SECUE  
NCIA) VALUES('"+codigo+"','"+sec+"')");  
}
```

```
fr.close();  
JOptionPane.showMessageDialog(this,"El  
proceso se ha terminado");  
Explorar.setEnabled(false);  
Cargar.setEnabled(false);  
// consulta.close();  
} catch(Exception e) {  
System.out.println("Error: "+e); };
```

Los resultados de la ejecución del código se observan en las Figs. 3 y 4.

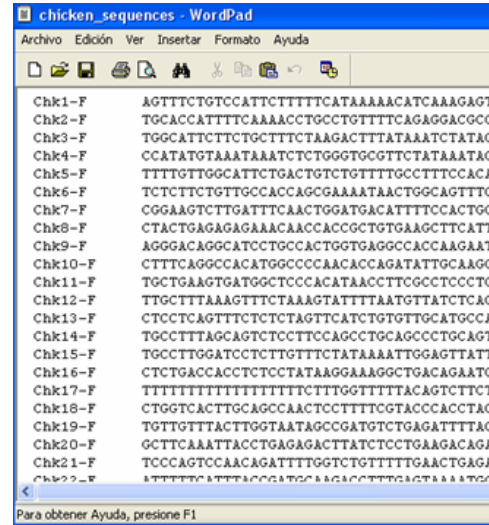


Figura 3. Archivo de texto

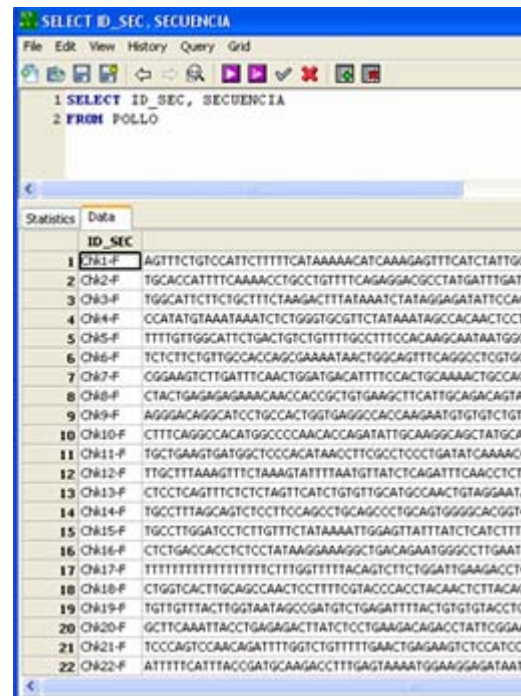


Figura 4. Vaciado del archivo plano en la base de datos

Clase Comparative: Es la responsable de realizar el filtrado de datos almacenado en una tabla de la base de datos. Se realiza una consulta SQL con agrupamiento de datos, de tal forma que por cada columna se obtiene un orden de frecuencia de aparición de dímeros. El primer recuadro del código fuente que aparece abajo muestra la consulta SQL la cual se realiza de forma dinámica, cambiando entre columna y columna, luego el segundo recuadro muestra como estos valores se

agregan a una tabla para ser visualizadas en pantalla, ver Fig. 5

```
for(int j=1;j<130;j++) {
try {
Conexion conectado3 = new Conexion();
Connection conectado3 =
conectado3.AbrirConexion();
// Se realiza la consulta de la clave
Statement buscando3 =
conectado3.createStatement();
String cambiapar =
“par”+String.valueOf(contador);
```

```
ResultSet buscado3 =
buscando3.executeQuery(“SELECT
“+cambiapar+”, count(*) FROM “+eligeTabla+”
GROUP by “+cambiapar+” ORDER by
count(*)”);
while (buscado3.next()) {
String base = buscado3.getString(1);
int frecuencia = buscado3.getInt(2);
cad1 = cad1+“\n”+base+ “=”+frecuencia;
// -- Se escribe las frecuencias en el archivo bajo
el formato <CA 5 par1>
cad2= base+” “+frecuencia+” par”+j+”\n”;
byte buf[] = cad2.getBytes();
```

```
//fo.write(buf);
cad2=””;
// Cambiando el valor de una celda
fila = 49-frecuencia;
```

```
if(miTabla.getModel().getValueAt(fila,columna)!=
null) {
String aux =
miTabla.getModel().getValueAt(fila,columna).toSt
ring();
aux = aux+”,”+base;
miTabla.getModel().setValueAt(aux,fila,columna);
} else
{miTabla.getModel().setValueAt(base,fila,column
a); }
```

```
...
} /// Aqui termina el For.
```

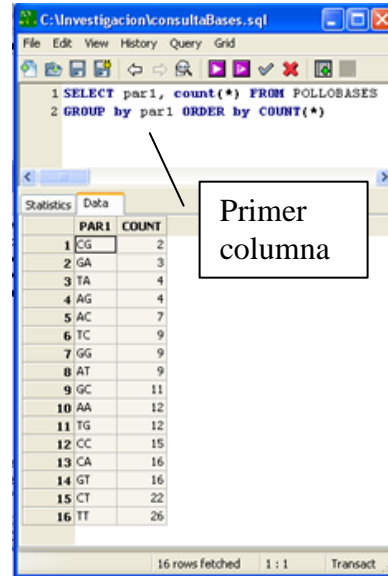


Figura 5. Análisis de resultados por columna.

Clase HorizontalDimers: Esta es la responsable de que los datos se puedan graficar de forma discretizada. En el código fuente del primer recuadro que aparece más abajo crea una referencia a la tabla que será coloreada de forma visual. El segundo recuadro es el que asigna un color distinto a los dímeros que están en un rango de 10 en 10.

```
class CustomTableCellRenderer2 extends
DefaultTableCellRenderer{
public Component
getTableCellRendererComponent (JTable table,
Object valor, boolean isSelected, boolean
hasFocus, int row, int column) {
Component cell2 =
super.getTableCellRendererComponent (table,
valor, isSelected, hasFocus, row, column);
if (isSelected) {
cell2.setBackground(Color.LIGHT_GRAY);
} else {
if (valor!=null) {
boolean bandera1 = false;
```

```
String cadena1 = valor.toString();
//System.out.println(mode.get(j).toString());
setToolTipText(cadena1);
int numero1 = 0;
// Se hace una relación entre color y Valor de
celda
try {
```

```

numero1 = Integer.parseInt(cadena1);
if (numero1>0 && numero1<=10)
cell2.setBackground(Color.yellow); else
if (numero1>10 && numero1<=20)
cell2.setBackground(Color.green); else
if (numero1>20 && numero1<=30)
cell2.setBackground(Color.red); else
if (numero1>30 && numero1<=40)
cell2.setBackground(Color.blue); else
if (numero1>40 && numero1<=50)
cell2.setBackground(Color.orange); else
if (numero1>50 && numero1<=60)
cell2.setBackground(Color.cyan);
bandera1=true;
    
```

IV. INTERFAZ GRÁFICA DE USUARIO (GUÍ)

Se ha hecho un esfuerzo para que la Interfaz gráfica de la herramienta sea agradable y de fácil comprensión al usuario, Ver Fig. 6



Figura 6. Interfaz principal

La herramienta tiene 4 secciones, estos se describen a continuación:

Subir secuencias: Originalmente una secuencia de ADN se puede obtener mediante secuenciadores automáticos, estos generan archivos que son de acceso público en repositorios de la WEB, un archivo de secuencias es plano (txt o dat). Estas secuencias son tomadas y descargadas en una tabla en la base de datos (BD) que se generó para este fin. Véase la Fig. 7



Figura 7. Cargar secuencia a BD.

Ordenar secuencias: En vista de que la secuencias pueden tener distintos tamaños en nucleótidos (número de letras de ADN) se requiere normalizar o igualar dichos tamaños, una vez igualados se requiere dividir la secuencia en pares de nucleótidos, esto para poder extraer información estadística de aparición de las nucleótidos. Véase la Fig. 8



Figura 8. Ordenamiento de secuencias

Graficar por dímero: Este módulo se va a encargar de graficar la frecuencia de aparición de dímeros como AA/TT/TA. Se puede graficar un dímero, dos o hasta 16 dímeros, esto depende de las necesidades del usuario, así mismo en todo momento se puede observar el código de colores asignado a cada dímero para un análisis más fácil del comportamiento de dichos dímeros. Véase la Fig. 9

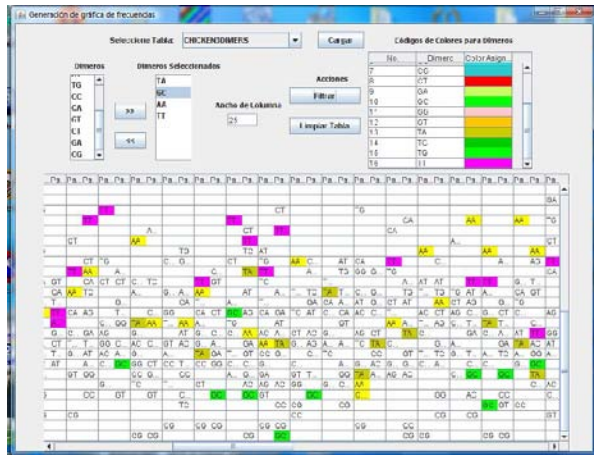


Figura 9. Graficado por dímero

Discretizar dímeros: Este módulo se encarga de mostrar una vista diferente de la frecuencia de dímeros, en ella los dímeros son agrupados acorde con una escala de aparición por ejemplo los dímeros que tienen frecuencia de 1-5 se le asigna un color, luego del 6-10, y así sucesivamente, permitiendo obtener una vista de agrupada de los dímeros. Verse Fig. 10

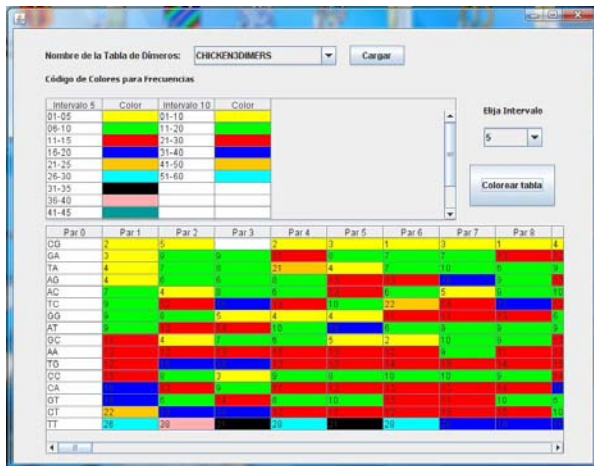


Figura 10. Discretización de dímeros

V. PRUEBAS A LA HERRAMIENTA COMPUTACIONAL

Las pruebas que se hicieron a la herramienta consistieron en observar si al tomar algunas muestras de secuencias de ADN era posible apreciar algún tipo de comportamiento. De manera inicial se tomaron dos dímeros de prueba bajo el uso del módulo 3, TT y GC. Véase la Fig. 11

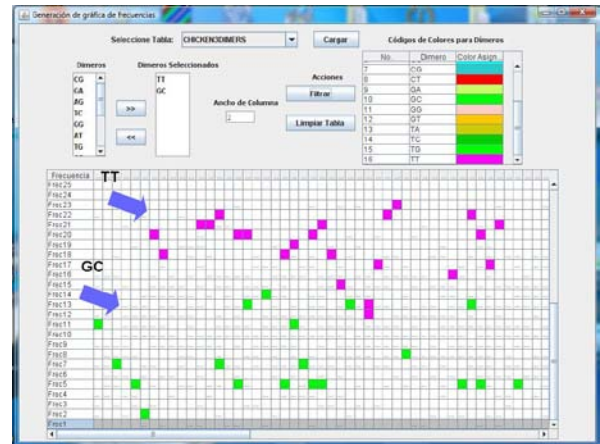


Figura 11. Comportamiento de TT y GC

En la Fig. 11 se puede observar que el comportamiento de TT es prácticamente opuesto al de GC mientras que el primero tiene un comportamiento de una frecuencia alta en el segundo es baja.

Otra prueba consistió en ver el comportamiento de 3 dímeros, específicamente TA, AA y TT. Véase la Fig. 12

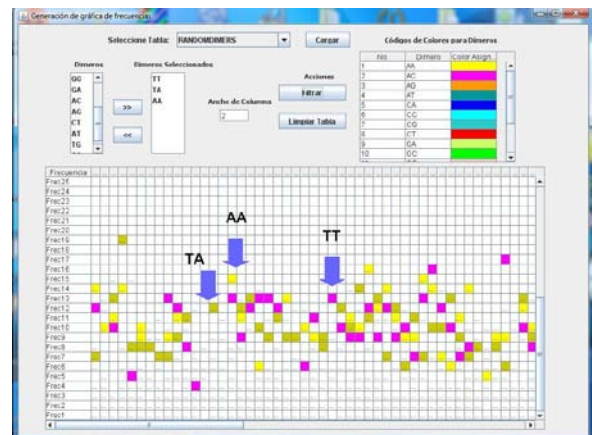


Figura 12. Dímeros TA, AA y TT

La Fig. 12 nos permite apreciar que el comportamiento de los dímeros en cuestión es muy similar, sus órdenes de frecuencia nos ayudan a identificar que las diferencias de aparición solo son del orden de 2 o 3 valores de frecuencia.

VI. CONCLUSIONES

El presente trabajo ha mostrado el funcionamiento completo de la herramienta de visualización de

secuencias de ADN, lo que incluye su estructura interna y la interfaz gráfica de usuario. La herramienta como tal permite a un biólogo tener una visión global del comportamiento de patrones gráficos en secuencias de ADN y le ahorra tiempo en su estudio. Por otro lado se ha planeado que esta herramienta tenga algunas mejoras como el hecho de que pueda ser funcional para el Internet y pueda ser empleada por diversos usuarios a la vez.

VII. REFERENCIAS

- [1] *WorldBook, ADN y su constitución*, <http://www.worldbookonline.com/>, [Consultado: 15 de Noviembre de 2008]
- [2] *Nacional Geographics, El Cuerpo Humano*, <https://genographic.nationalgeographic.com/genographic/overview.html>, [Consultado: 10 de Octubre de 2008]
- [3] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thastrom, Y. Field, I.K. Moore, J.P.Z. Wang, and J. Widom, "A Genomic Code for Nucleosome Positioning", *Nature*, Vol. 442, 2006, pp. 772-778.
- [4] Orrego-Cardozo, M., *Estudios sobre la Interacción de Subtipos de la Histona H1 con el DNA y la Cromatina*, Ph. D. Dissertation, Universidad Autónoma de Barcelona, Barcelona, January 2004.
- [5] A. Thastrom, P. T. Lowary, H. R. Widlund, H. Cao, M. Kubista, and J. Widom, "Natural and Non-natural Nucleosome Positioning DNA Sequences", *J. Mo. Biol.*, Vol. 288, 1999, pp. 213-229.
- [6] P. T. Lowary, and J. Widom, "Nucleosome Packaging and Nucleosome Positioning of Genomic DNA", *Proc. Natl. Acad. Sci. USA*, Vol. 94, February 1997, pp. 1183-1188.
- [7] Lacroix, Z., and T. Critchlow, *Bioinformatics Managing Scientific Data*, Morgan Kauffmann Publishers, 2003.
- [8] *Inmegen, Herramientas para análisis de ADN*, http://www.inmegen.gob.mx/index.php?option=com_content&task=view&id=525&Itemid=292, [Consultado el 05 de Noviembre de 2008]

- [9] NCBI, Uso del Blast para analisis de secuencias de ADN, <http://blast.ncbi.nlm.nih.gov/Blast.cgi> [Consultado: 05 de Noviembre de 2008]
- [10] Firebird Database Engine, http://www.firebirdsql.org/index.php?option=com_content&task=view&id=525&Itemid=292 [Consultado 15 de Octubre de 2008]

VIII. AUTORES

M. en C. Gustavo Verduzco Reyes obtuvo su título de Maestría en Ciencias con especialidad en Sistemas Distribuidos en CENIDET, Cuernavaca Mor-México, actualmente es profesor de tiempo completo de la UNACAR.

Dr. Ernesto Bautista Thompson es Dr. en Ciencias de la Computación por parte del CIC-IPN, México D.F., actualmente es profesor de tiempo completo de la UNACAR.

C.M.C. Benjamín TASS Herrera es candidato a Maestro en Ciencias por la Universidad de Morelos, Nuevo León, actualmente es profesor de tiempo completo de la UNACAR.

GENERACIÓN DE AMBIENTES VIRTUALES MEDIANTE MODELADO DECLARATIVO BASADO EN CONOCIMIENTO

Jaime Zaragoza, Félix Ramos, Véronique Gaildrat*

Grupo de Sistemas Distribuidos
Centro de Investigación y de Estudios Avanzados del Politécnico Nacional, Unidad Guadalajara
Av. Científica 1145, colonia el Bajío, Zapopan, 45015, Jalisco, México.

Tel. 3337773600, correo electrónico: {jzaragoz, [frames](mailto:frames@gdl.cinvestav.mx)}@gdl.cinvestav.mx

*VORTEX Groupe

Institut de Recherche in Informatique de Toulouse
Université Paul Sabatier, 118 Route de Narbonne, F-31062, Toulouse CEDEX 9, Francia
Tel. 33561556765, correo electrónico: veronique.gaildrat@irit.fr

RESUMEN

Crear mundos virtuales usando herramientas dedicadas es una tarea que requiere de tiempo y recursos en cantidades variadas. Aún los usuarios experimentados pueden encontrar difícil el crear representaciones virtuales para diversas necesidades, por ejemplo, simulaciones del mundo real, recreación de mundos fantásticos o antiguos, educación, entre otros. A través del método del modelado declarativo un usuario puede crear un escenario virtual al expresar algunas propiedades deseadas para el ambiente y las entidades en él. Este artículo presenta una aproximación novedosa para el modelado declarativo. La principal contribución es el uso de conocimiento previo sobre el ambiente y las entidades para asistir el proceso de la creación y validación del mundo virtual creado de ésta forma. Éste conocimiento incluye la información necesaria para permitir la evolución del mundo.

Palabras clave: Modelado declarativo, realidad virtual, explotación del conocimiento.

I. INTRODUCCIÓN

La Realidad Virtual se ha utilizado como método para simular entidades en el mundo real, de diferentes campos de la experiencia humana, como medicina, construcción, entretenimiento, y muchos otros. Ésta tecnología permite recrear casi cualquier tipo de escenario o mundo, y puede

representar casi cualquier tipo de escena o situación. Se ha propuesto diferentes métodos para diseñar y animar esos mundos virtuales.

Sin embargo, la mayoría de las veces esos mundos son creados un equipo multidisciplinario completo, compuesto por muchos artistas, modeladores e ingenieros, siendo necesarias desde algunas semanas hasta años para completar una visualización exitosa del ambiente deseado. Esto incluye el uso de herramientas especializadas, las cuales necesitan entretenimiento especial y muchas horas de práctica para obtener un resultado de apariencia profesional. La problemática de crear mundos virtuales se puede descomponer en dos partes: primero es la creación del ambiente, el segundo, describir la escena, es decir, como los personajes deben de evolucionar en el ambiente. En este trabajo tratamos con el primero de esos subproblemas, al tiempo que se ha conducido trabajo en el segundo subproblema [1].

Un método que permita la creación de escenarios virtuales tanto simples como complejos es un tema actual de investigación. El Modelado Declarativo es una aproximación a éste objetivo.

En éste artículo entendemos Modelado Declarativo como una metodología que permite crear un escenario virtual por medio de una descripción, escrita en un lenguaje lo más cercano al natural, donde se indica el tipo de ambiente, las entidades que lo poblarán, y como estarán distribuidas esas entidades. El Modelador Declarativo debe tomar

como entrada la descripción, y encontrar al menos una solución que satisfaga los deseos del usuario.

La principal deferencia de nuestro trabajo con otros es el uso de conocimiento sobre las propiedades de los objetos y otras entidades para auxiliar el proceso de validación de las posibles soluciones, y que el modelo resultante puede ser utilizado para manejar la evolución del escenario. Este modelo puede ser enviado a cualquier arquitectura subyacente para su visualización y animación, ya que toda información necesaria es provista con el modelo. La arquitectura también puede conducir la simulación del mundo virtual, pues también se incluye en el modelo las reglas para el ambiente y el comportamiento de las entidades. Este trabajo es parte del proyecto GeDA-3D [2], una arquitectura multi-agente distribuida en 3D. El objetivo de GeDA-3D es ofrecer una herramienta completa a cualquier usuario con la necesidad de simular un comportamiento dado.

II. TRABAJO RELACIONADO

WordsEyes [3], desarrollado en los laboratorios AT&T por Bob Coyne y Richard Asproad, es el trabajo más cercano al nuestro. WordsEyes es un convertidor automático texto-a-escena, el cual utiliza marcadores de parte del discurso y analizadores estadísticos para obtener una descomposición jerárquica de las oraciones provistas por el usuario. El sistema utiliza diferentes herramientas para asignar *descriptores*, representaciones gráficas de bajo nivel que corresponden con cada concepto en la oración, y que proveen las características físicas a los elementos en el escenario. Se utiliza cinemática inversa para conseguir las poses de las entidades en la escena, y se emplean algunos métodos para resolver ciertas abstracciones, como textualización, caracterización o personificación. El sistema está basada en el Internet, de manera que el usuario solo tiene que entrar al sitio Web del proyecto, escribir una descripción corta, y enviarla para obtener una vista estática del escenario creado.

El proyecto DEM²ONS [4] es un sistema multimodal compuesto por una interfaz modal y un modelador en 3D. La interfaz modal usa diferentes

métodos de entrada, desde ratón hasta dispositivos hápticos, permitiendo al usuario crear el escenario de una manera intuitiva. Los módulos en cargo del análisis sintáctico y las interfaces dedicadas manejan las entradas, las cuales se traducen en eventos normalizados. El modelo resultante es validado a través de ORANOS [5], un resolvidor de restricciones diseñado para permitir expansibilidad en aplicaciones de modelado declarativo. El sistema se basa en una Interfaz Gráfica de Usuario o GUI, escrita en el Open Inventor Toolkit [6], y permite interacciones simples, como modificar los elementos no dinámicos presentados en el escenario.

CAPS [7] es un sistema basado en restricciones, orientado a resolver la disposición de objetos dentro de un escenario. El sistema permite el modelado de escenarios grandes y complejos, usando una entrada consistente de restricciones intuitivas de posicionamiento que pueden provenir de distintos métodos de entrada, como texto escrito o dispositivos hápticos. Puede manejar varios objetos a la vez, usando pseudo-físicas para asegurar la estabilidad de la posición, y emplea conceptos tales como fragilidad o interacción para asegurar la validez de la misma. Permite interacciones y está orientado hacia la manipulación de escenarios, pero no provee método alguno para la auto-evolución de la escena.

III. MODELADO DECLARATIVO

El modelado declarativo define un proceso recursivo necesario para crear una solución a las propiedades declaradas por el usuario. Aplicado a la creación de un escenario, el modelado declarativo es un proceso continuo en el cual retroalimentación proveniente de soluciones previamente obtenidas es usada recursivamente, hasta que el usuario está satisfecho con la solución obtenida. El método está compuesto por tres fases: Descripción, Generación y Vista [8].

Reconocimientos:

Esta investigación es parcialmente apoyada por el proyecto CoECyT-Jal No. 2008-05-97094, mientras el autor J.A. Zaragoza agradece el apoyo de la beca CONACYT no. 190965.

En la fase de generación, el sistema recibe la descripción del usuario, escrita en un lenguaje de definición diseñado para la aplicación de modelado. Este lenguaje puede venir en una variedad de formas, ya sea simples entradas de texto a señales provenientes de dispositivos especializados o procesador del habla. El sistema valida la sintaxis de la entrada, la analiza, y crea una salida en un formato adecuado para el sistema, y entonces transfiere la información obtenida al siguiente paso. A continuación se procede a crear el modelo, normalmente iniciando con valores por defecto para las entidades en el mismo, y procediendo a asignar los valores para las propiedades usando distintos métodos para asegurar la consistencia de la semántica, la lógica y la posición de las entidades. Uno de los métodos más comunes es resolver un Problema de Satisfacción de Restricciones o CSP.

Podemos definir un CSP como un conjunto de variables $X = \{X_1, X_2, \dots, X_n\}$, un dominio D el cual indica los valores posibles para variable en X , y un conjunto de restricciones $C = \{C_1, C_2, \dots, C_n\}$ el cual especifica un subconjunto de variables y los valores posibles para cada una de ellas.

Una vez que cada variable ha sido asignada con un valor de su dominio, de tal forma que no se viola ninguna restricción, se ha alcanzado una solución. Para un CSP dado puede existir más de una solución, de tal manera que se puede diseñar un algoritmo CSP de tal manera que pueda localizar varias soluciones [9].

El paso de modelado puede crear varios modelos y entonces proceder a validarlos, o puede iniciar con uno y entonces resolver cualquier conflicto hasta que se alcance una solución. Este proceso involucra métodos de rastreo y avance, permitiendo al sistema descubrir si ha alcanzado un mínimo o máximo local. En el primer caso, pueden existir algunas restricciones que en apariencia no pueden ser resueltas, pero al cambiar algunas de las propiedades de las entidades se puede llegar a un nuevo arreglo que las satisfaga todas. En el segundo caso, todas las restricciones se cumplen, pero la solución es rígida y no permite ninguna modificación sin romper alguna restricción

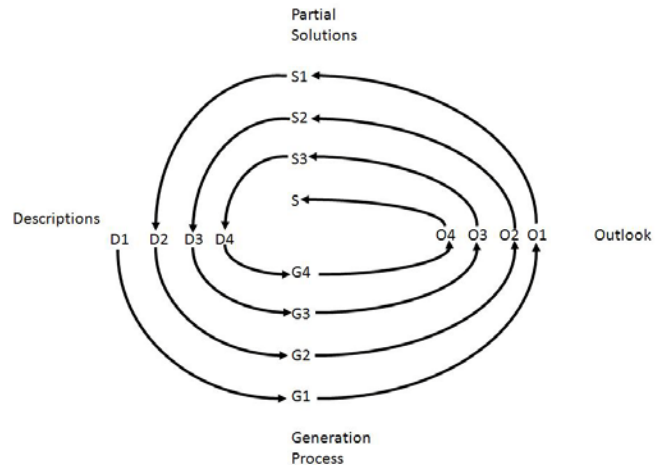


Figura 1. Proceso del Modelado Declarativo.

Nuevamente, algunos cambios pueden llevar a un nuevo conjunto de soluciones con valores más flexibles. Este paso puede incluir la participación del usuario, presentado las soluciones obtenidas y permitiendo al usuario decidir si alguna cumple con la imagen mental empleada para crear la descripción.

La fase de vista consiste en presentar al usuario las soluciones obtenidas. El modelo puede entonces ser modificado para acercarlo a la idea del usuario al iniciar el proceso, pero siempre dentro de las restricciones establecidas en el CSP. El usuario puede rechazar algunas o todas las soluciones encontradas, indicando así al sistema cual espacio de soluciones debe ser explorado para localizar futuras soluciones.

IV. UN MODELADOR DE AMBIENTES VIRTUALES BASADO EN CONOCIMIENTO

Nuestra propuesta tiene una característica principal, la cual no está completamente considerada en otros trabajos. Ya que el usuario expresa que es lo que debe ser posicionado en el escenario virtual, el sistema posee las indicaciones para las características y posición de cada elemento, las cuales pueden aplicadas dentro de ciertos rangos difusos. Por ejemplo, aún cuando la posición exacta de los elementos puede variar en la perspectiva de cada usuario, existe la certeza si es una referencia absoluta o relativa entre los elementos: “derecha” será siempre un área específica definida en relación a una entidad en

referencia, sin importar su orientación, tamaño o entidad en referencia, y puede ser definida dentro de una estructura que provee los parámetros para obtener esa área. Esa estructura de datos puede presentarse en la forma de una base de datos. Sin embargo, la información necesaria debe ser provista en combinación de las relaciones entre los conceptos en evaluación, por lo cual una solución mejor adaptada son las bases de conocimiento. Ya que una base de conocimientos no solo contiene la información de los elementos en un dominio dado, sino también las relaciones entre conceptos, permite derivar nuevo conocimiento de la información ya presente, expandiendo el alcance de la base de conocimientos.

La acción de extrapolar nueva información de conocimiento actual es llamada *inferencia*, y es una característica útil para validar conceptos y propiedades, ya que el usuario puede comenzar expresando características simples, las cuales pueden ser combinadas para inferir conocimiento complejo.

Con esta idea, diseñamos un modelador declarativo que permite la creación de mundos virtuales a partir de una simple descripción en texto, usando la explotación de bases de conocimiento como base para el proceso de modelado. El modelador está formado por cinco módulos, como se muestra en la figura 2.

V. ANÁLISIS LÉXICO SEMÁNTICO

El Analizador Léxico-Sintáctico recibe la descripción escrita en un lenguaje definido a la medida, llamado VEDEL o Lenguaje para Descripción de Ambientes Virtuales [10]. VEDEL es parecido al lenguaje natural, completamente orientado a conducir el proceso de descripción. Provee una ayuda natural a los usuarios con su estructura para organizar las ideas al describir mundos virtuales. También permite crear escenarios de manera incremental, al añadir o modificar elementos específicos para extender la descripción. Una descripción en VEDEL se compone de tres secciones, Ambiente, Actores y Objetos, delimitados por etiquetas de sección. Cada una de esas secciones está formada por

oraciones compuestas por declaraciones separadas por comas. Cada declaración corresponde a una propiedad para la entidad que está siendo descrita, y puede ser declaradas en cualquier orden, con una única restricción, la primera declaración debe corresponder al tipo de entidad, incluyendo un identificador opcional. Cada oración debe terminar con un punto (“.”).

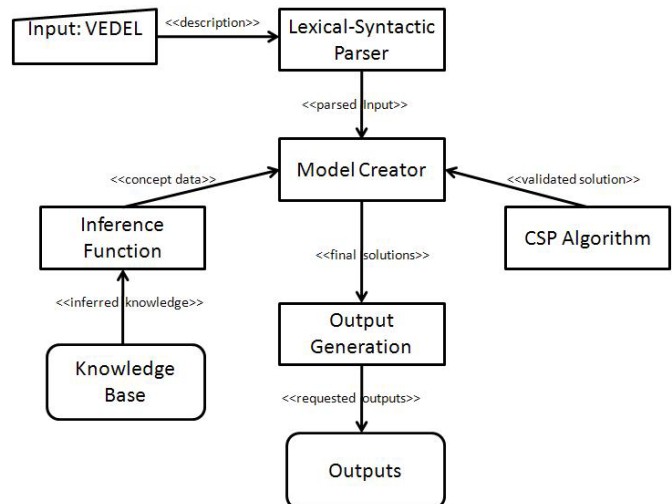


Figura 2. Arquitectura del Modelador

Las razones para definir un lenguaje de interacción como un subconjunto del lenguaje natural son: proveer un método estructurado simple y amigable para escribir y formatear la descripción, así como evitar la necesidad de métodos lógicos difusos y analizadores estadísticos requeridos en el análisis y validación del lenguaje humano diario. La estructura del lenguaje permite mantener un balance en el número de restricciones, evitando descripciones sobre o bajo-restringidas.

El analizador Léxico-Sintáctico es básicamente una máquina de estados que extrae elementos de la descripción, y los utiliza para formar una estructura de datos jerárquica que representa la información expresada en la descripción. La jerarquía está organizada con los tipos de entidad como las ramas superiores y sus propiedades como hojas, permitiendo una extracción de los datos por el módulo de análisis Léxico-Sintáctico, el cual solo verifica la composición correcta de la descripción y busca caracteres no permitidos, dejando la validación semántica para el módulo de Creación del Modelo.

[ENV]
house, night.
[/ENV]

[ACTOR]
man Antony, big, old, bald, sit BigCouch.
woman Brittany, young, near CenterTable.
dog Casper, sleep, front FirePlace.
[/ACTOR]

[OBJECT]
Couch BigCouch, against westWall, big,
facing east.
CoffeTable CenterTable, center
LivingRoom.
[/OBJECT]

Figura 3. Ejemplo de VEDEL.

VI. CREACIÓN DEL MODELO

El Creador del Modelo recibe la entrada analizada por el Analizador Léxico-Semántico, y procede a generar el modelo.

El modelador trabaja en pasos incrementales, iniciando con un modelo básico instanciado con valores por defecto para los elementos del ambiente virtual, asignando posteriormente los valores solicitados por el usuario.

La creación del modelo termina cuando todas las propiedades han sido instanciadas y éstas no violan ninguna restricción.

Primero, el modelador obtiene toda la información sobre el ambiente. Cualquier petición por información es manejada a través del componente Función de Inferencia, el cual accede directamente a la base de conocimientos. La función de inferencia es un singleton, y es usado a través de todo el proceso de obtención de información, formateando la información para la tarea de modelado y validando las peticiones hechas en la descripción. La información reunida desde la base de conocimientos sobre el ambiente es usada para construir las reglas del mundo virtual y asignar valores a sus propiedades. Si el ambiente posee alguna construcción o elemento especial, el modelador construye las entradas necesarias en el modelo para satisfacer esas indicaciones. Zonas específicas como paredes, puertas o áreas, o

elementos propios del ambiente (amueblado, vegetación) son generados de ésta manera. Las zonas específicas corresponden a información implícita sobre el ambiente y no tienen una representación visual explícita, mientras que los elementos propios del ambiente deben ser instanciados, colocados y representados como elementos individuales en el ambiente.

Una vez que el ambiente ha sido creado, el modelador continúa con las entidades que lo poblarán. Cada una de esas entidades es creada usando representaciones básicas instanciadas con valores por defecto almacenados en la base de conocimientos. Tales presentaciones son llamadas Avatares. Cuando se le solicita que cree una entidad en particular, primero verifica si el avatar correspondiente existe. En caso contrario, se hace la solicitud a la base de conocimientos para obtener la información de la entidad, y el avatar se añade al modelo. El avatar es entonces instanciado con los valores de la entidad en proceso. Cada petición es validada primero a través del avatar, para verificar si la propiedad es válida para la entidad. Posteriormente, se verifica a través de la función de inferencia si los valores asignados son correctos o corresponden a la propiedad en proceso. Cualquier conversión es conducida a través de la base de conocimientos. Por ejemplo, si el usuario solicita una silla coloreada en rojo, el modelador primero verifica que “silla” posea la propiedad “color”. Si la propiedad es validada, se procede a verificar que si uno de los posibles valores para la propiedad “color” de “silla” es “rojo”. Si todo resulta positivo, la función de inferencia retorna la información correspondiente al color “rojo” en el formato RGB.

POSICIONADO LAS ENTIDADES

Una vez que todas las entidades han sido creadas y sus propiedades instanciadas con los valores especificados en la descripción, el modelador procede a posicionar cada entidad en la posición correcta, de acuerdo a las declaraciones hechas en la descripción. Las entidades son colocadas primero en una posición por defecto, almacenada en la base de conocimientos. El proceso para obtener los parámetros para posicionar la entidad

es similar al usado para obtener los valores de las propiedades. La declaración es validada directamente a través de la función de inferencia, la cual recibe la información correspondiente a la posición actual de la entidad, su tamaño y su orientación, y, en el caso de posicionamiento relativo, también se envía los valores de la entidad en referencia. Los parámetros correspondientes al concepto son obtenidos de la base de conocimientos y posteriormente instanciados con los valores de la entidad o entidades. Esos nuevos parámetros son enviados al modelador, el cual los asigna a las propiedades de la entidad. La entidad en referencia puede ser cualquier entidad en el ambiente, ya sea el ambiente mismo o una zona específica. Los parámetros del concepto pueden ser definidos con parámetros difusos, de manera que el proceso de modelado presente cierta aleatoriedad, para permitir la generación de diferentes modelos durante el proceso.

Ya que cada entidad ha sido posicionada, el modelador envía el estado actual del modelo al componente CSP. Este componente verifica la posición de cada entidad, esto es, que no existan colisiones entre las entidades y que la posición de cada una de ellas corresponda a las declaraciones hechas en la descripción. Para cumplir esas tareas, cada entidad posee una serie de marcas de colisión, así como marcas de puntos característicos. Esas marcas especializadas están almacenadas en la base de conocimientos, y son instanciadas cuando la entidad es creada. Cuando el modelador cambia la posición de la entidad, esas marcas también son actualizadas.

Las marcas de colisión están definidas como una serie de esferas que envuelven toda la geometría de la entidad (figura 4), y son usadas como la restricción principal a ser satisfecha por el algoritmo CSP. Para ejecutar la verificación, el sistema primero verifica si la distancia euclidiana entre cualquier par de entidades es menor que la diagonal mayor correspondiente a una caja formada por las medidas de la entidad. Si esto es verdadero, se evalúa la ecuación de distancia entre dos esferas para cada par posible de marcas de colisión entre las dos entidades. Esta ecuación debe ser satisfecha para cualquier par de marcas de colisión entre entidades en conflicto, $C_1 = (x_1, y_1,$

$z_1, r_1)$ y $C_2 = (x_2, y_2, z_2, r_2)$, de manera que la relación siguiente se mantenga:

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 - (r_1 - r_2)^2 \geq t \quad (1)$$

Si cualquier par falla ésta evaluación, esto es, el resultado de la ecuación (1) es menor que un rango ($t \in \mathbb{R}, 1 > t > 0$) establecido por el administrador del sistema, una colisión ha sido detectada y el modelo debe ser modificado. Para realizar los cambios en la posición de las entidades, primero se realizan algunas verificaciones: si una de las entidades es un “*pivote*” (esto es, está asignada a una posición absoluta, como *norte*, *sur*, o *este*), el tamaño de las entidades, y la relación entre ellas.

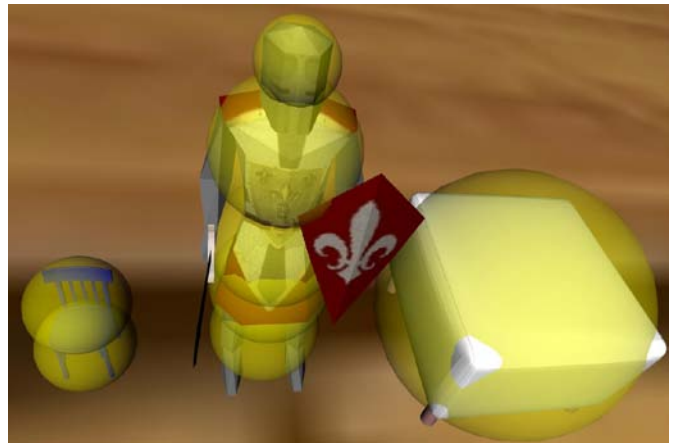


Figura 4. Marcadores de colisión.

Las entidades pivote son modificadas al final, y sólo si el modelo requiere de tales cambios para obtener una solución.

Enseguida, las relaciones entre elementos son exploradas, ya que algunos de los conceptos de posicionamiento requieren que dos entidades entren en contacto, tales como “*contra*”, “*dentro*” o “*sobre*”, de manera que en esos casos la colisión será desechada y otras validaciones tendrán lugar. En otros casos, las entidades que hagan referencia a otra serán movidas primero. Si la configuración del modelo no puede ser validada, la entidad en referencia será entonces movida. En otros casos, ambas entidades hacen referencia a una tercera entidad, en cuyo caso el algoritmo CSP hace uso de una técnica similar a la usada en los Sistemas-L [11] para posicionar ambas entidades en una nueva

locación que cumpla con las declaraciones hechas en la descripción mientras se resuelven el conflicto. Finalmente, el tamaño es considerado para decir cual entidad debe ser movida primero. Las entidades más pequeñas tienen mejor probabilidad de ser colocadas en una posición válida que los elementos grandes, por lo cual el sistema moverá primero las entidades más pequeñas en conflicto.

La verificación de colisiones es ejecutada hasta que no existen más conflictos. Si el modelo actual cae en un mínimo local, o no se puede encontrar una solución, el sistema restaura una solución parcial previa que contenga menos conflictos y procede a generar nuevas soluciones. Si la solución restaurada es el modelo inicial enviado por el creador de modelo, el sistema verifica si todas las entidades posibles han sido reasignadas, y reinicia el proceso. Si ya no existen más entidades a mover, el sistema informa que no se puede encontrar una solución.

Para posicionar exitosamente una entidad, la siguiente validación debe ser satisfecha: al menos uno de sus puntos característicos debe estar dentro del volumen de un elipsoide o un paraboloides instanciadas con los parámetros correspondientes a la información de la entidad en referencia y a los valores del concepto asociado. Esto es, para cada entidad en el ambiente, al menos uno de sus puntos característicos $C = (x, y, z)$ debe estar en la superficie o dentro de un volumen de validación $E = (dx, dy, dz)$, para elipsoides (3), o $P = (p, q)$, para una paraboloides (2), instanciadas con los valores de la entidad en referencia. Podemos representar esto como:

$$\left(\frac{x^2}{p}\right) + \left(\frac{y^2}{q}\right) - 2z \leq 0 \quad (2)$$

$$\left(\frac{x^2}{dx}\right) + \left(\frac{y^2}{dy}\right) + \left(\frac{z^2}{dz}\right) - 1 \leq 0 \quad (3)$$

Los parámetros para ambas ecuaciones están almacenados en la base de conocimientos, como parte del concepto de posicionamiento o como parte de la entidad en referencia, en cuyo caso también se pueden incluir la cantidad de puntos característicos que necesitan estar dentro del volumen para validar la posición, o incluso la lista

de puntos que deben ser evaluados. Las posiciones relativas son validadas siempre a través de volúmenes de paraboloides, para los cuales los parámetros p y q corresponden al tamaño de la entidad en referencia, y z es un valor definido por el administrador del sistema.

Relaciones espaciales tales como “contra” o “sobre” son verificados usando volúmenes elipsoidales, para los cuales uno de sus parámetros es menor a 1, y localizados sobre las superficies de las entidades. Para posiciones que requieren un volumen dentro de una entidad, o una posición en un área dada tal como zonas específicas, el volumen elipsoidal es instanciado con los valores de la entidad o la zona.

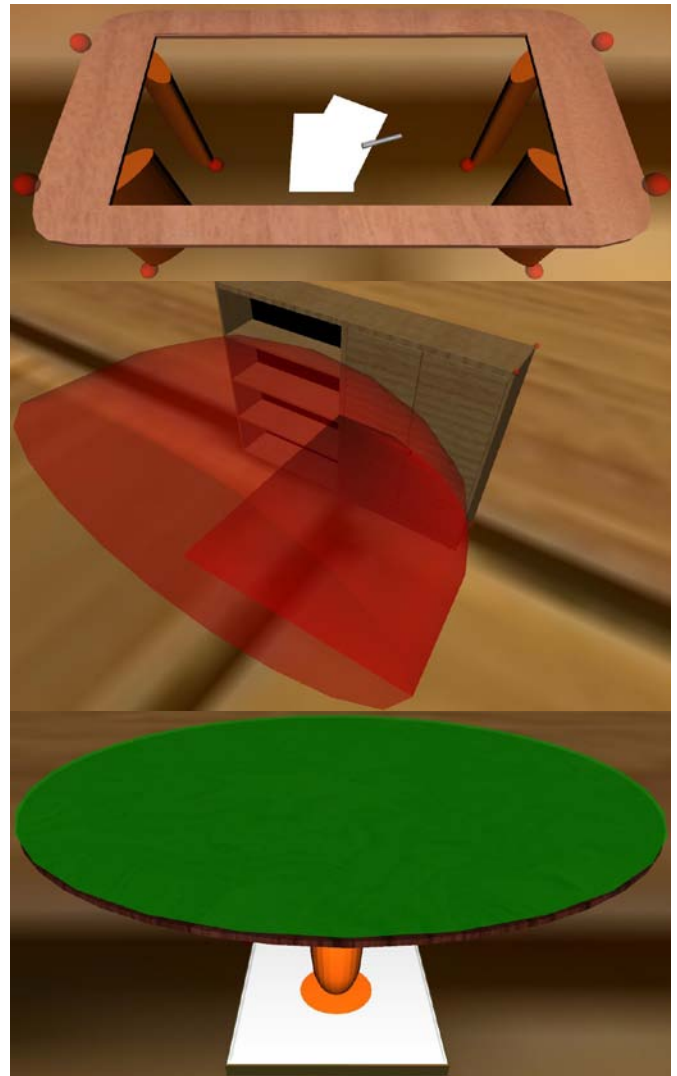


Figura 5. Puntos característicos y volúmenes de posicionamiento

Si la validación de la posición de una entidad falla, se calcula una nueva posición. Antes de asignar ésta nueva posición, se revisa una lista de posiciones previas para la entidad en evaluación. Si el vecindario de esas previas posiciones forma un grupo compacto, entonces una nueva posición fuera de ese grupo es asignada, y la validación es conducida nuevamente. Si la nueva posición no es válida, o no puede ser asignada fuera del grupo, el sistema no puede encontrar una nueva posición para la entidad y regresa a un estado del modelo previo. Al igual que con la verificación de colisiones, si el modelo previo es el estado inicial, la descripción solicitada no puede ser resuelta.

Cada entidad en el modelo almacena la lista de elementos con los cuales comparte alguna relación. De ésta manera, cada entidad puede seguir los movimientos hechos al modelo, y modificar sus propios parámetros en consecuencia, por ejemplo, si una entidad debe orientar en relación a otra, y la entidad en referencia es movida, la primera entidad puede ajustar sus parámetros de orientación para ajustarse a la nueva posición establecida para la entidad en referencia. Lo mismo aplica para posiciones relativas. Ya que cada entidad ha pasado ambas validaciones, el modelo es enviado al componente final, el Generador de Salida.

GENERANDO LAS SALIDAS

El componente de Generación de Salida recibe el modelo final. Este componente utiliza la metodología Controlador Vista-Modelo para generar las estructuras de datos de salida y los archivos necesarios para que la arquitectura subyacente pueda representar las acciones necesarias para crear la visualización del ambiente virtual, y conducir el proceso que activará la simulación. Estas salidas utilizan plantillas que reciben la información del modelo, y a través del MCV son instanciadas con los valores obtenidos a través del proceso de modelado. Este método permite la fácil modificación de los parámetros de salida, de manera que los usuarios pueden agregar o modificar la información enviada a la arquitectura subyacente, que puede ser también un visualizador 3D de cualquier tipo, en tanto las plantillas mantengan los estándares utilizados para

la visualización en 3D. La cantidad de salidas puede ser establecida en la base de conocimientos.

VII. TRABAJO PRESENTE

Actualmente contamos con un prototipo del modelador en el lenguaje Java, esto con la finalidad de proporcionarle características multi-plataforma. La base de conocimientos ha sido definida usando el sistema Protégé y fue creada utilizando el estándar OWL [12], seleccionada debido a su expansibilidad sobre la Internet. A continuación se muestran algunos ejemplos obtenidos usando la versión más reciente del modelador, para cuya visualización se utiliza el estándar X3D y un visualizador basado en X3D.

En el primer ejemplo, figura 6, "LivingRoom" es un concepto almacenado en la base de conocimientos que también establece el amueblado del escenario y la posición de cada elemento. El segundo ejemplo, figura 7, muestra diferentes modelos de casa obtenidos de una sola descripción. "House" establece el ambiente completo, incluyendo la mayoría del amueblado y las áreas en referencias para colocar a los actores. El último ejemplo, figura 8, muestra el mecanismo de resolución de conflictos para varios objetos colocados en la misma posición relativa.

VIII. CONCLUSIONES

En este artículo hemos descrito nuestra novedosa aproximación para el modelado declarativo utilizando bases de conocimiento para asistir en el proceso de generación del modelo y vista. Hemos probado nuestra propuesta por medio de un prototipo aún en evolución. Hasta ahora, hemos creado diferentes escenarios con elementos simples, generando incrementalmente entidades complejas. También hemos probado la facilidad para modificar los parámetros y conceptos, utilizando diferentes valores para el mismo modelo, alcanzando modelos significativamente diferentes a través de varias iteraciones del modelo. Otros trabajos tratan con la metodología del modelado declarativo, pero a partir de la

[ENV]
 LivingRoom.
 [/ENV]
 [ACTOR]
 ManSuit John.
 YoungWoman Sarah.
 [/ACTOR]
 [OBJECT]
 [/OBJECT]

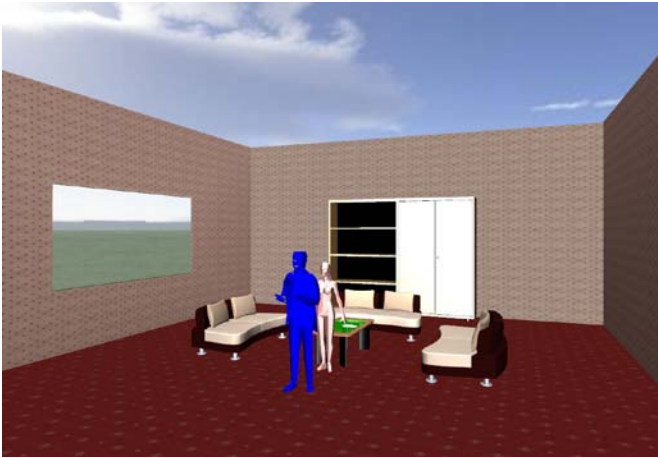


Figura 6. Ejemplo 1.

literatura disponible podemos afirmar que ninguno está orientado a crear mundos dinámicos y controlados por el usuario.

Aún cuando sólo tuvimos acceso a uno de ellos, WordsEye, es aparente que todos auxilian en el proceso de crear mundos complejos, pero la interacción con el usuario es limitada a la creación del escenario, y cambiar la vista con que se presentan los resultados. En el caso de WordsEye, su naturaleza basada en red no permite la modificación directa de los modelos usados para la representación en 3D, y la vista y calidad de la misma está limitada por la capacidad del servidor del sistema. En caso de DEM2ONS y CAPS, la literatura de ambos apunta que los métodos de entrada son muy especializados, y la salida está limitada a la visualización del modelo, sin interacción alguna. El modelador presentado en este artículo permite crear, a través de una descripción escrita en un lenguaje cercano al natural, un escenario en 3D, así como las estructuras de datos necesarias la evolución de una escena en 3D.

[ENV]
 house.
 [/ENV]
 [ACTOR]
 Knight, anywhere Kitchen.
 YoungWoman, anywhere Garden.
 woman, anywhere Livingroom
 [/ACTOR]
 [OBJECT]
 Chair, front woman0.
 Puff, front bed0.
 [/OBJECT]



Figura 7. Ejemplo 2.

La principal desventaja de esta investigación es que la base de conocimientos debe contener todos los conceptos así como la necesidad de una base de objetos 3D. Sin embargo, una vez que ambos aspectos han sido cumplidos, la creación de escenarios en 3D es posible para usuarios finales. Con esto se alcanza el objetivo principal de ésta investigación. El trabajo futuro incluye una

herramienta amigable para manejo del conocimiento y extensión de la base de objetos 3D, así como tratar con casos específicos que puedan requerir de métodos de modelado especiales.

```
[ENV]
void.
[/ENV]
[ACTOR]
Knight, center.
[/ACTOR]
[OBJECT]
CenterTable Table, front Knight0.
Chair, left Table.
Chair, left Table, color red
Chair, left Table, color blue
Chair, left Table, color gray
Chair, left Table, color green
facing Knight0.
[/OBJECT]
```



Figura 8. Ejemplo 3.

IX. REFERENCIAS

- [1] F. Ramos, F. Zúniga, and H. I. Piza, "A 3D-space platform for distributed applications management," International Symposium and School on Advanced Distributed Systems 2002. Guadalajara, Jal., México, Noviembre 2002.
- [2] P. Hugo, F. Zúniga, and F. Ramos, "A platform to design and run dynamic virtual environment," International Conference on Cyber Worlds, Tokio Japón, pp. 18–20, Noviembre 2004.
- [3] B. Coyne and R. Sproat, "Wordseye: An automatic text-to-scene conversion system," en SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques. AT&T Labs Research, 2001, pp. 487–496.
- [4] O. Le Roux, V. Gaildrat, and R. Caubet, "Design of a new constraints solvers for 3D declarative modeling" en International Conference on Computer Graphics and Artificial Intelligence (3IA), Limoges, 03/05/200-04/05/200, mayo 2000, pp. 75–87.
- [5] G. Kwaiter, V. Gaildrat, and R. Caubet, "DEM2 ONS: A high level declarative modeler for 3D graphics applications," en Proceedings of the International Conference on Imaging Science Systems and Technology, CISST'97, 1997, pp. 149–154.
- [6] D. Wang, I. Herman, and G. J. Reynolds, "The open inventor toolkit and the PREMO standard," Amsterdam, Holanda, Rep. Tec, 1996.
- [7] K. Xu, "Constraint-based automatic placement for scene composition," en In Graphics Interface, 2002, pp. 25–34.
- [8] V. Gaildrat, "Declarative modeling of virtual environment, overview of issues and applications," en International Conference on Computer Graphics and Artificial Intelligence (3IA), Atenas, Grecia, vol. 10. Laboratoire XLIM - Université de Limoges, mayo 2007, pp. 5–15.
- [9] S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach. Pearson Education, 2003.
- [10] J. A. Zaragoza Rios, "Representation and exploitation of knowledge for the description phase of declarative modeling", CINESTAV-PNI, Unidad Guadalajara, Guadalajara, México, Septiembre 2006.
- [11] P. Prusinkiewicz and A. Lindenmayer, The algorithmic beauty of plants. Springer-Verlag New York, Inc., 1990.
- [12] D. L. McGuinness and F. van Harmelen, "OWL Web Ontology Language overview," World Wide Web Consortium, W3C Recommendation, Febrero 2004, <http://www.w3.org/TR/2004/REC-owl-features-20040210/>.

EL SEGUIMIENTO DE LA TUTORÍA PERSONALIZADA EN LA UNIVERSIDAD DE COLIMA MEDIANTE TECNOLOGÍAS DE INFORMACIÓN

Ricardo Fuentes Covarrubias, Andrés Gerardo Fuentes Covarrubias, Alfredo Rodríguez Moreno, María Isabel Parada Sandoval

Universidad de Colima- Instituto de Innovación y Desarrollo Tecnológico- Facultad de Ingeniería Mecánica y Eléctrica – Av. Gonzalo de Sandoval 444, Col. Las Víboras Colima, México. C. P. 28045.
Tel. 01(312)3161150, correo electrónico: fuentesr@ucol.mx.

RESUMEN

Se presentan los resultados de un proyecto en el área de ingeniería de software para el seguimiento de la tutoría personalizada, para lo cual se diseñó un software mediante el cual se automatiza la tutoría personalizada y se da seguimiento al proceso desde la etapa de registro o alta, las entrevistas entre tutor y tutorado, el plan de trabajo, así como las distintas evaluaciones lo cual incidirá en un mejor seguimiento del proceso de la tutoría.

Palabras clave: tutoría académica, proceso enseñanza aprendizaje, tutorado, tecnologías de información.

I. INTRODUCCIÓN

La tutoría personalizada se define como la relación creativa entre un maestro y un alumno, o un grupo pequeño de estos, con la intención de tratar asuntos pertinentes al ámbito académico, de una manera mas personal, contribuyendo así a elevar la calidad del proceso enseñanza – aprendizaje. Siendo personal en el sentido de que en la relación maestro – alumno se establece un vínculo más directo y estrecho dándose con ello un acompañamiento a lo largo de la trayectoria de los estudiantes [7]. En la Universidad de Colima la atención personalizada, constituye una gran oportunidad para que la educación superior mexicana se transforme, teniendo como perspectiva una nueva visión y nuevo paradigma

para la formación de los estudiantes, concibiendo el aprendizaje a lo largo de toda la vida, el aprendizaje autodirigido y la formación integral con una visión humanista y responsable, ante las necesidades y oportunidades del desarrollo de nuestro país y de la región.

Este tipo de atención a los estudiantes a través de los tutores, coadyuva en abatir los índices de reprobación y rezago escolar, a disminuir los índices de deserción y abandono de los estudio y al mejoramiento de la eficiencia terminal.

II. ESTADO DEL ARTE

La tutoría académica personalizada en cada escuela y facultad de la Universidad de Colima, es brindada por los profesores y para su realización se apega a las siguientes directrices:

- a) Adquirir capacitación necesaria para la actividad tutorial.
- b) Conocer los aspectos institucionales y específicos del estudiante esenciales para la actividad tutorial.
- c) Construir una situación de interacción apropiada con el tutorado, estableciendo un clima de confianza que permita identificar estilos de aprendizaje, características de personalidad y posibles problemas que afecten su desarrollo como estudiante.
- d) Realizar diagnostico de necesidades de tutoría que permitan identificar problemas

específicos que afectan el rendimiento académico de los alumnos.

- e) Establecer un programa de atención, canalización y seguimiento de tutorado.

El objetivo que se persigue con la implementación de este software es dar un seguimiento ágil y expedito al proceso de la tutoría la cual se realizaba anteriormente en forma manual, utilizando formatos que se fueron estandarizando a través del intercambio de experiencias entre las escuelas y facultades así como por parte de la Dirección de Orientación Vocacional de la Universidad de Colima[9], pero la disponibilidad de la información no era la adecuada debido a las distancias entre los cinco campus universitarios en los cuales se encuentra organizada nuestra institución, lo cual no permitía que la información fluyera rápidamente pues los informes se entregaban al final del semestre y ello además impedía que se tomaran decisiones a tiempo cuando el caso requería de atención inmediata.

III. DESGLOSE DEL SISTEMA

III.1 ETAPA DE DISEÑO. CASOS DE USO

Como consecuencia de la problemática descrita en la etapa anterior se pensó en desarrollar una herramienta computacional que permitiera llevar el seguimiento de la tutoría en función a la organización existente en la Universidad de Colima.

En esta etapa se identificaron los principales actores así como sus funciones y actividades dentro del sistema:

- a) El alumno. Se da de alta en el sistema utilizando su número de cuenta y generando una contraseña personal, con lo anterior procederá a realizar el examen SITAE, le será asignado un tutor y evaluará la acción tutorial, consulte la figura 1.

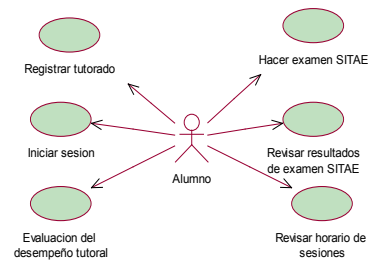


Figura 1. Casos de uso alumno.

- b) Profesor o Tutor. Se da de alta en el sistema utilizando su número de trabajador y genera una contraseña personal, analiza las respuestas dadas por sus tutorados en el examen SITAE, realiza el horario de sesiones así como la programación de las actividades tutoriales, contesta la encuesta de las valoraciones de las dificultades de la acción tutorial, hace anotaciones para el control de las entrevistas con sus tutorados las cuales son confidenciales debido a que cada tutor solo puede revisar los registros de sus tutorados. Consulte la figura 2.
- c) El coordinador de tutorías de cada escuela o facultad. Este es quien se encarga de realizar las asignaciones de tutores y tutorados, se le asigna una identificación como superusuario, puede generar reportes de tutores y tutorados y da seguimiento al proceso de la tutoría en cada plantel a través de la valoración del desempeño tutorial. Consulte la figura 3.

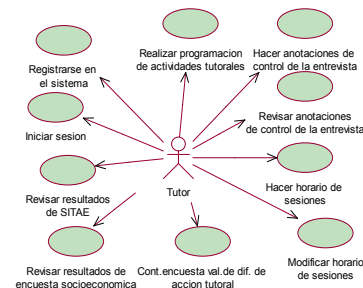


Figura 2. Casos de uso Profesor.

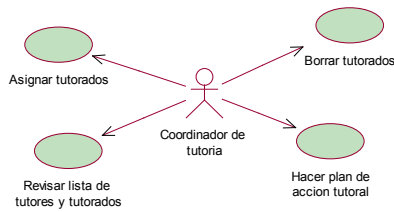


Figura 3. Casos de uso Coordinador de tutorías

III.2 INTERFASES DE USUARIO

Su diseño es amigable, muy fácil de utilizar y solo requiere una breve explicación de 10 minutos para que los usuarios puedan empezar a trabajar, en la pantalla inicial se hace una breve introducción al proceso de la tutoría como se muestra en la figura 4.



Figura 4. Vista de principal del sistema.

A continuación cada uno de los usuarios puede acceder al área que le corresponde en función de su clasificación: alumno, tutor o coordinador de tutoría para lo cual utilizará el numero de cuenta de alumno, el numero de trabajador de acuerdo a nomina o la clave de coordinador de tutorías la cual inicia con la letra “c” seguida del numero de trabajador es de hacer notar que el sistema solo reconoce a un coordinador de tutorías por escuela o facultad. Las figuras 5 y 6, muestran a continuación la vista del modulo Altas:

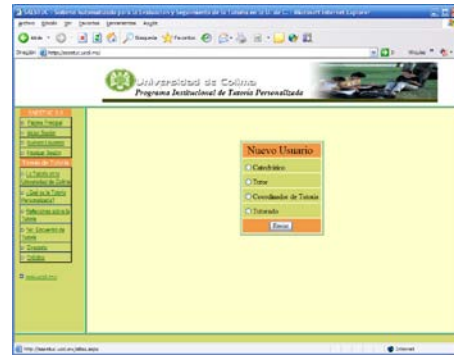


Figura 5. Altas al sistema

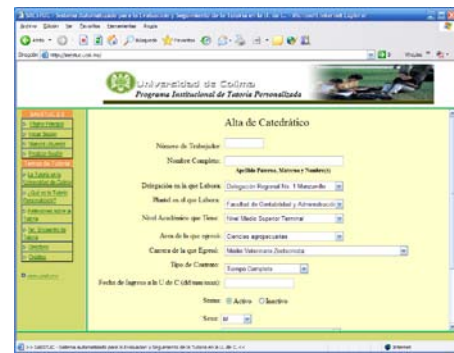


Figura 6. Complemento altas de profesores

Una vez que el profesor se da de alta en el sistema puede ingresar en el mediante su clave de trabajador y la clave de usuario que registro en el proceso de altas.

Al iniciar una sesión el profesor realiza entre otras cosas la programación de la acción tutorial como se muestra en la figura 7, el seguimiento de las entrevistas con sus tutorados y las demás actividades descritas en la sección anterior.

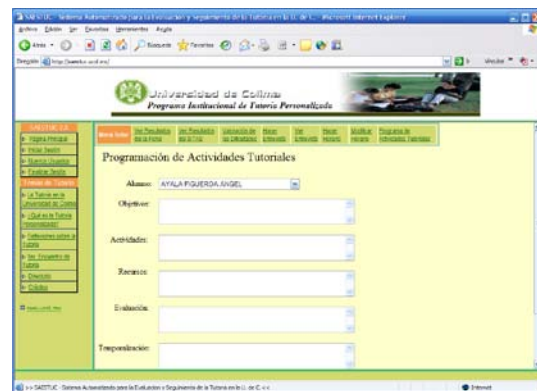


Figura 7. Sección para la programación de la acción tutorial por parte del tutor.

El proceso de la entrevista, le permite al tutor dar seguimiento a cada charla realizada con sus tutorados. El sistema cuenta con las siguientes opciones de trabajo para llevar el control de la entrevista.

1. Al elegir la opción Alumno, seleccionar el nombre del Tutorado con el que tuvo la entrevista.
2. Con la opción Motivo, indicar el motivo de la entrevista.
3. Al elegir de la lista desplegable Canalización si fue necesario canalizar con alguna otra persona o instancia al alumno, y si no es así elegir la opción Ninguno
4. En la caja de texto Observaciones y/o Compromisos generará alguna nota u observación acerca del tutorado o la entrevista Consulte la figura 8.

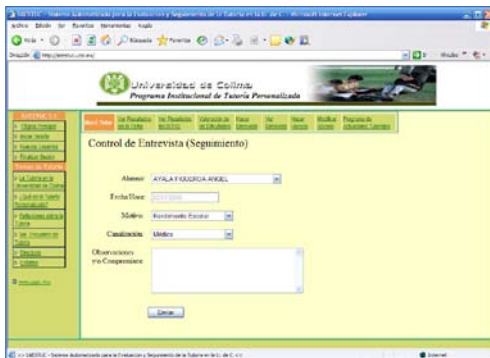


Figura 8. Vista de proceso de la entrevista.

En lo que respecta al coordinador de tutorías, éste puede realizar asignaciones de alumnos y tutores como se muestra en la Figura 9, sin embargo es de hacer notar que él solo puede observar en pantalla a los alumnos y profesores que corresponden a su escuela o facultad, otra área importante es el de reportes los cuales pueden ser por tutores y alumnos o solo por tutores.



Figura 9. Pantalla de asignación de tutores a tutorados.

Por ultimo, la Dirección de Orientación Vocacional como súper usuario tiene acceso al seguimiento de la tutoría en cada una de las escuelas y facultades de la Universidad de Colima y puede obtener reportes por escuela o facultad, por delegación o generar un reporte global.

III.3 ETAPA DE DESARROLLO. ASPECTOS TECNICOS.

Para la implementación del software se utilizó la suite Visual Basic .NET en una plataforma Windows 2000, ejecutándose en la Intranet de la Universidad de Colima mediante paginas dinámicas .ASP con una base de datos SQL a partir de la base de datos del sistema de control escolar de la Universidad de Colima (SICEUC).

IV. RESULTADOS

Actualmente se está desarrollando la versión 2.0 de esta herramienta la cual está enriquecida con el pilotaje realizado durante 4 años en los 5 campus de la Universidad de Colima en el nivel superior en el cual se involucraron aproximadamente 10,000 estudiantes. De igual manera en la nueva versión de SAESTUC se va a involucrar a los alumnos del nivel bachillerato.

Otro aspecto importante, es que las instancias de gobierno de la Universidad de Colima pueden conocer en cualquier momento cual es el estado de la tutoría pues el sistema genera reportes por

delegación, por escuela o facultad, por carrera y por tutor.

V. DESARROLLOS FUTUROS.

A cuatro años de la implementación de SAESTUC, se trabaja en nuevas alternativas que permitan un trabajo mas dinámico en el proceso de la tutoría, específicamente en lo referente a la entrevista, y se encuentra en proceso de experimentación, un módulo de computo móvil para la realización de las entrevistas de los tutores en un ambiente no controlado, es decir fuera de la oficina del tutor con el propósito de dar un seguimiento más oportuno al proceso de la tutoría. Consultar figuras 10 y 11. La última versión del software puede ser consultada en el link: <http://saestuc.ucol.mx>.

En esta etapa se utilizan PDA's Pocket PC marca Dell a 520 MHz, Batería extraíble y recargable, con tecnología inalámbrica 802.11b y Bluetooth™ integrada, Memoria SDRAM de 64 MB, Pantalla táctil TFT QVGA (240 x 320) transflexiva y a color de 3,5 pulg. y 16 bits, Procesador Intel® XScale™ PXA270 a 520 MHz., Ranuras para tarjetas de memoria integradas, Memoria Flash de 128 MB y Sistema Operativo Microsoft® Windows ® Mobile 2003.

Se utiliza como plataforma de desarrollo: Windows XP Professional (Service Pack 2), Microsoft Visual Studio 2003.NET (Visual Basic), Microsoft ActiveSync, Microsoft SQL Server 2000, Microsoft SQL Server CE, lo cual permite su total transparencia y compatibilidad con SAESTUC.

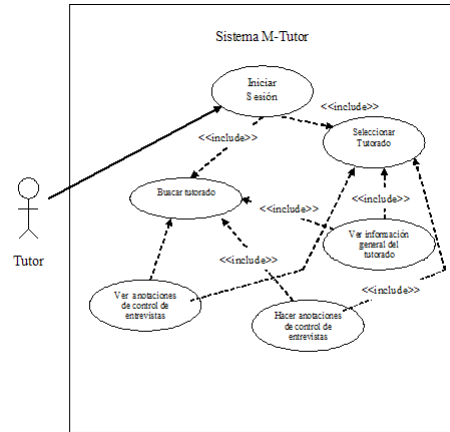


Figura 10. Modelo de casos de uso tutoría móvil.

Al ingresar al sistema, este cargará los tutorados asignados, mostrándolos en una lista para que puedan ser seleccionados por el tutor. Como lo muestra la figura 12.

Al tener un alumno seleccionado el tutor puede proceder a realizar la entrevista, de la misma forma, al tener seleccionado el alumno, se puede consultar el historial de las entrevistas realizadas a la fecha o iniciar una nueva entrevista.

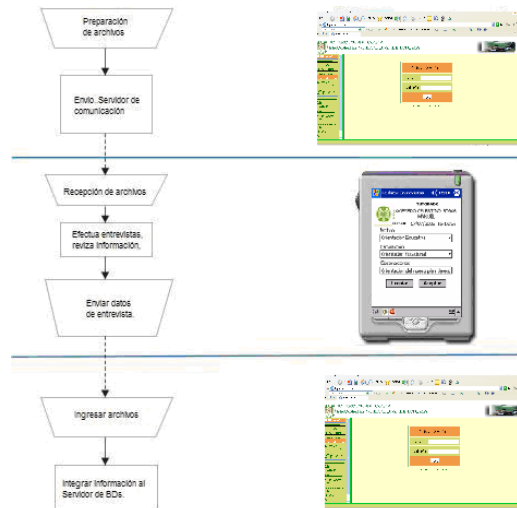


Figura 11. Interacción del sistema móvil y fijo.

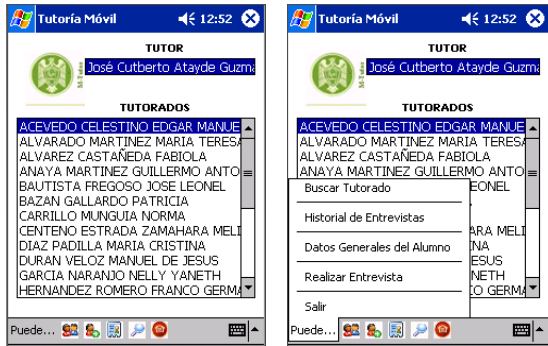


Figura 12. Interfaz para seleccionar alumno y opción a realizar.

VI. CONCLUSIONES

Esta herramienta permite simplificar el proceso de la tutoría pero de ninguna manera sustituye el trato personal entre tutor y tutorado, más bien auxilia en el seguimiento de la tutoría, desburocratizándola al sustituir los formatos manuales existentes y los cuales debían ser llenados y enviados por los tutores a las instancias correspondientes al final del semestre. En este sentido el tutor podrá dedicar su atención a aspectos más puntuales de la tutoría relacionados con la identificación de:

- Intereses y aspiraciones académicas
- Rendimiento escolar pues el sistema interactúa con el sistema de control escolar de la Universidad de Colima (SICEUC).
- Actitud ante el trabajo escolar
- Nivel de integración en la escuela
- Detectar necesidades que permitan mejorar el rendimiento escolar
- Conocer las expectativas profesionales de sus tutorados
- Ayudar al alumno en la toma de decisiones.

El resultado se reflejará en un proceso de la tutoría más dinámica y con una mejor interacción tutor – tutorado.

VII. REFERENCIAS

- [1] Álvarez, P. y González, M. (2008). Análisis y valoración conceptual sobre las modalidades de tutoría universitaria en el espacio europeo de educación superior. *Revista interuniversitaria de formación de profesores*.
- [2] González R. (2008): *Herramientas básicas para el acompañamiento tutorial*, revista mexicana de orientación educativa. 2ª. Época, vol. VI, número 14, 12-18. Mexico D. F.
- [3] Rincón Igeta, B. del (2000): *Tutorías personalizadas en la universidad*. Cuenca. Ediciones Castilla-La Mancha.
- [4] Michael Halvorson, (2008): *Aprenda ya Visual Basic 2005*. Edit. Mc Graw – Hill.
- [5] Jimmy Nilsson, (2001): *.NET Enterprise Design with Visual Basic .NET and SQL Server 2000*, First Edition, Sams Publishing. December 12.
- [6] Mridula Parihar, (2002): *ASP.NET Bible*, First Edition, Hungry Minds. October 10.
- [7] Fuentes Covarrubias Ricardo, Fuentes Covarrubias Andrés Gerardo, Villa Chávez Alma Ruth (2004): *Diseño de un software computacional para el seguimiento de la tutoría personalizada en la Universidad de Colima*. *Gaceta Universidad de Colima*.
- [8] Pablo Fernández Juárez (2004). *La Tutoría Académica Personalizada y su Importancia en la Eficiencia Escolar*, *Revista Mexicana de Orientación Educativa*. N° 2, Marzo 2003-Junio 2004. <http://www.remo.ws/revista/n2/n2-fernandez.htm>.
- [9] Castellanos Castellanos Ana Rosa, Venegas González Francisco Javier, Ramírez Domínguez José Luís (2003). *Sistemas Tutoriales en el centro occidente*

de México, Biblioteca de la Educación superior.

VIII. AUTORES

Ricardo Fuentes Covarrubias. Maestro en ciencias computacionales por la Universidad de Colima, Master en tecnologías de información y sistemas informáticos por la Universidad Rey Juan Carlos de España. Profesor e investigador en la Facultad de Ingeniería Mecánica y Eléctrica y el Instituto de Innovación y Desarrollo Tecnológico de la Universidad de Colima.

Andrés Gerardo Fuentes Covarrubias. Maestro en ciencias computacionales por la Universidad de Colima, Master en Tecnologías de Información y Sistemas Informáticos por la Universidad Rey Juan Carlos de España. Profesor e investigador en la Facultad de Ingeniería Mecánica y Eléctrica y el Instituto de Innovación y Desarrollo Tecnológico de la Universidad de Colima.

Alfredo Rodríguez Moreno. Egresado de la Maestría en computación e Ingeniero en sistemas computacionales Universidad de Colima.

Maria Isabel Parada Sandoval. Egresada de la Maestría en computación por la Universidad de Colima y Licenciada en informática por el Instituto Tecnológico Regional de Colima.

FORMACIÓN DE AGENTES MÓVILES BASADA EN ESTRATEGIAS DE AUTO-ORGANIZACIÓN

Miguel A. Sánchez-Acevedo, Félix Ramos-Corchado, Ernesto López-Mellado.

CINVESTAV Unidad Guadalajara, Av. Científica 1145, Col. El Bajío, 45010 Zapopan, Jal. México
Email: {sancheza, framos, elopez}@gdl.cinvestav.mx

RESUMEN

En este artículo se presenta una metodología para establecer formaciones estructuradas de agentes móviles en dos y tres dimensiones. Se utilizan estrategias de auto-organización para establecer la formación. La posición de los agentes en la formación es determinada al comunicarse con agentes vecinos ya formados. Los agentes son integrados progresivamente a la formación. El controlador del agente se obtiene a partir de un modelo de red de Petri; este modelo representa el estado del agente y la comunicación con sus vecinos. Se incluyen simulaciones para validar el comportamiento de los agentes.

Palabras clave: Formación de agentes móviles, auto-organización, interacciones locales, red de Petri.

I. INTRODUCCIÓN

Debido a la gran diversidad de aplicaciones que pueden desarrollarse con un grupo de agentes móviles con capacidades mínimas de razonamiento, muchos investigadores han definido estrategias para el control de formaciones [3], [4], [7], [9], [16], [17]; sin embargo, obtener un grupo robusto, escalable y adaptable es un gran reto para agentes autónomos.

Recientemente, los investigadores han introducido mecanismos de auto-organización observados en la naturaleza para resolver problemas complejos [2], [8], [10], [13], [14]. Siguiendo la idea de obtener un comportamiento global basado en interacciones locales, se han presentado diferentes trabajos para resolver el problema de formación de agentes móviles.

La propuesta presentada en [1] muestra un análisis de la adaptabilidad a las condiciones del ambiente de agentes en diferentes formaciones; para mantener la posición de cada agente en la formación, cada agente toma como referencia a su vecino. Una técnica para volar en una trayectoria preestablecida utilizando un grupo de vehículos aéreos autónomos se presenta en [5]. Cada vehículo genera las señales de control basado en el objetivo asignado, la posición a mantener en la formación y las restricciones presentadas en el ambiente. Para mantener la posición, se calculan dos fuerzas virtuales considerando la posición de los otros vehículos; esta posición es transmitida continuamente a todos los vehículos.

[15] presenta una técnica de control basada en la teoría de estabilidad en malla [12], donde dos helicópteros reales y seis helicópteros simulados son controlados mientras vuelan manteniendo la formación. En otra propuesta se presenta una arquitectura de control [18] para la formación de múltiples UAVs (vehículos aéreos no tripulados, por sus siglas en inglés); en este trabajo se utilizan dos modos de operación: el primer modo (modo seguro) es usado cuando no se detectan obstáculos en el ambiente y se utiliza un control predictivo para mantener la posición en la formación. El segundo modo (modo peligro) se utiliza cuando existen obstáculos en el ambiente; para calcular la trayectoria que evite los obstáculos se utiliza una red neuronal Grossberg.

Una propuesta centralizada para adaptar la formación con el fin de evitar amenazas se presenta en [6]. Una vez que la amenaza es detectada, un miembro de la formación (el mejor equipado) se encarga de calcular las posiciones de

los agentes en la formación con el fin de evitar la amenaza. Las nuevas posiciones son enviadas a cada elemento del grupo.

Para resolver el problema de formaciones en entornos tridimensionales, se introduce una propuesta basada en líderes virtuales combinada con campos potenciales [11]. Primero se generan los campos potenciales entre todos los elementos y después se calcula la trayectoria a seguir, generando de esta forma las señales de control para cada elemento de la formación.

El objetivo de este trabajo es establecer una formación estructurada de agentes móviles en un ambiente de dos y tres dimensiones utilizando estrategias de auto-organización. En esta propuesta la posición de los agentes no es predefinida, puesto que los agentes se van integrando progresivamente a la formación. Las estrategias de auto-organización producen un comportamiento complejo a partir de simples comportamientos individuales. Estos comportamientos son definidos mediante un conjunto de reglas if-then, las cuales son derivadas de un modelo de red de Petri que representa el estado del agente y la comunicación con sus vecinos.

La metodología propuesta se presenta en la sección II. La sección III presenta el controlador del agente. La sección IV presenta los resultados obtenidos en simulación. Finalmente las conclusiones obtenidas se presentan en la sección V.

II. ESTRATEGIA DE FORMACIÓN

II.1 Declaración del Problema.

En este trabajo se considera que los agentes pueden comunicarse solo con sus vecinos, y que las posiciones de los agentes en la formación no están pre-establecidas. Un agente puede tomar uno de cuatro roles posibles: *Líder*, este rol lo toma el agente localizado a la mitad de la formación; *Seguidor Izquierdo*, este rol lo toman los agentes ubicados a la izquierda de la formación con respecto al líder; *Seguidor Derecho*, este rol lo

toman los agentes ubicados a la derecha de la formación con respecto al líder; *Libre*, este rol lo toman los agentes que aún no se han integrado a la formación.

II.2 Estrategia de Formación

El comportamiento de los agentes se determina a partir de simples reglas generadas de acuerdo al estado del agente y de la información recibida de los vecinos. El líder mantiene la trayectoria a seguir y la formación a ser establecida. Esta información se comparte con los vecinos; de esta manera, cuando un nuevo agente es integrado a la formación, esta información es transmitida a él. Para mantener la formación, cada agente mantiene los mismos ángulos (yaw, pitch, roll). El tipo de formación es determinado por el ángulo α , mientras que la distribución de la formación en el espacio tridimensional es determinada por los ángulo β , roll, pitch, y yaw (Figura 1).

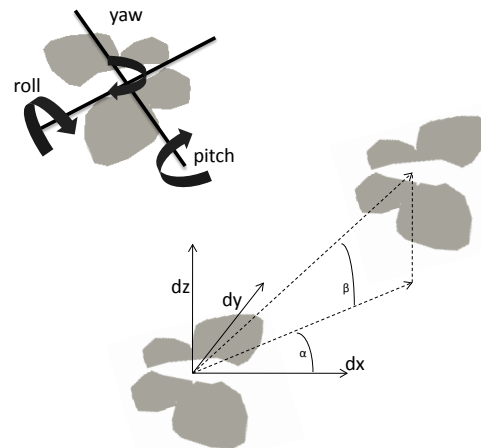


Figura 1. Ubicación de un agente con respecto a su vecino.

Cada agente en la formación emite una señal que puede ser detectada por agentes libres. Cuando un agente libre detecta esta señal, este agente tiende a moverse hacia donde la intensidad de la señal es más fuerte. Para evitar posibles deformaciones de la estructura debido a la integración de varios agentes al mismo tiempo, una vez que un agente libre inicia la interacción con un agente en la

formación, el agente en la formación transmite un mensaje a través de sus vecinos, el cual es retransmitido hasta que el líder lo recibe. El líder permite solo a un agente continuar con el proceso de integración de acuerdo al ID de los participantes para evitar posibles conflictos. Los agentes que pierden la elección son alejados para evitar que interfieran con el proceso de reconfiguración. Mientras se lleva a cabo el proceso de reconfiguración, los agentes en la formación no emiten señales, de tal forma que no se puedan integrar nuevos agentes hasta que el proceso haya concluido.

El agente libre inicia un intercambio de mensajes para definir su posición en la formación. La posición se determina de acuerdo al estado del agente vecino. Las posibles situaciones que un agente puede manejar se presentan a continuación. Por simplicidad, las figuras para explicar el comportamiento de los agentes están en dos dimensiones, no se considera el ángulo β , el cual se puede mantener mientras se lleva a cabo el proceso de reconfiguración.

- a) El agente es un líder sin vecinos (Figura 2).

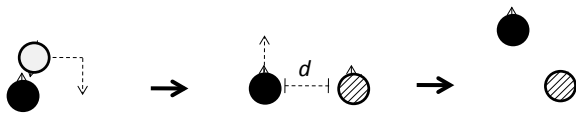


Figura 2. Un agente libre obteniendo su posición con respecto al líder.

- b) El agente es un seguidor derecho (Figura 3). Los agentes ubicados a la derecha del agente nuevo son recorridos a la derecha. Los agentes ubicados a la izquierda del agente nuevo avanzan una distancia d .

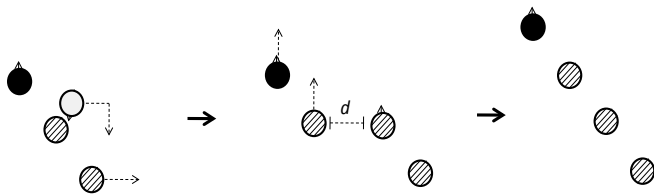


Figura 3. Un agente libre obteniendo su posición con respecto a un seguidor derecho.

- c) El agente es un seguidor izquierdo (Figura 4). Los agentes ubicados a la izquierda del agente nuevo son recorridos a la izquierda. Los agentes ubicados a la derecha del agente nuevo avanzan una distancia d .

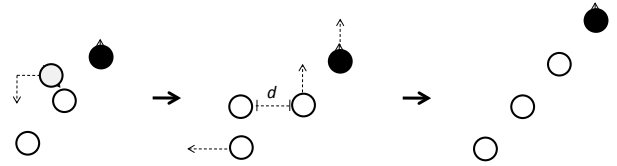


Figura 4. Un agente libre obteniendo su posición con respecto a un seguidor izquierdo.

Puede observarse como los agentes nuevos son integrados a la formación en el lugar donde encuentran a un agente ya formado, dando como resultado una deformación en la estructura. Para detectar esta deformación, el líder mantiene dos contadores. El primero mantiene el número de agentes integrados recientemente en el lado derecho de la formación; el segundo almacena el número de agentes integrados recientemente en el lado izquierdo de la formación.

II.3 Estrategia de Balance

Cada vez que un nuevo agente es integrado en la formación, se envía un mensaje, el cual es retransmitido hasta que el líder lo recibe. Una vez que el líder se da cuenta de que un nuevo agente ha sido integrado, actualiza su contador respectivo. Si ambos contadores son iguales, estos son reiniciados a cero; cuando la diferencia entre los contadores es igual a dos, se inicia el proceso de actualización de líder. Si el contador de vecinos izquierdos es mayor que el contador de vecinos derechos, entonces el líder cambia su rol a *seguidor derecho* y su vecino izquierdo cambia su rol a *líder* (Figura 5).

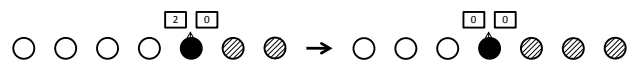


Figura 5. Actualización del rol de líder a seguidor derecho.

Si el contador de vecinos derechos es mayor que el contador de vecinos izquierdos, entonces el líder cambia su rol a *seguido izquierdo* y su vecino derecho cambia su rol a *líder* (Figura 6).

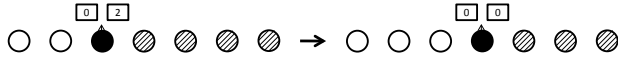


Figura 6. Actualización del rol de líder a seguidor izquierdo.

Para balancear la formación se lleva a cabo el mismo proceso con algunos incrementos. En lugar de cambiar solamente el rol de los agentes, se lleva a cabo un desplazamiento para reconfigurar la formación. Si el contador de vecinos izquierdos es mayor que el contador de vecinos derechos, todos los agentes que tienen el rol de *seguidores izquierdos* avanzan una distancia $2d$ (Figura 7).

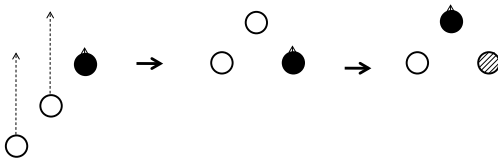


Figura 7. Balanceo de una formación desbalanceada a la izquierda.

En caso contrario, si el contador de vecinos derechos es mayor que el contador de vecinos izquierdos, entonces todos los agentes que tienen el rol de *seguidor derecho* avanzarán una distancia $2d$ (Figura 8).

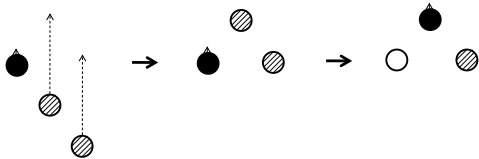


Figura 8. Balanceo de una formación desbalanceada a la derecha.

III. CONTROLADOR DEL AGENTE

Cada agente tiene el mismo controlador, el cual le permite reaccionar a eventos externos de acuerdo a su estado interno, el cual está representado por seis variables de estado representadas mediante una red

de Petri. Las variables se describen a continuación (Figura 9).

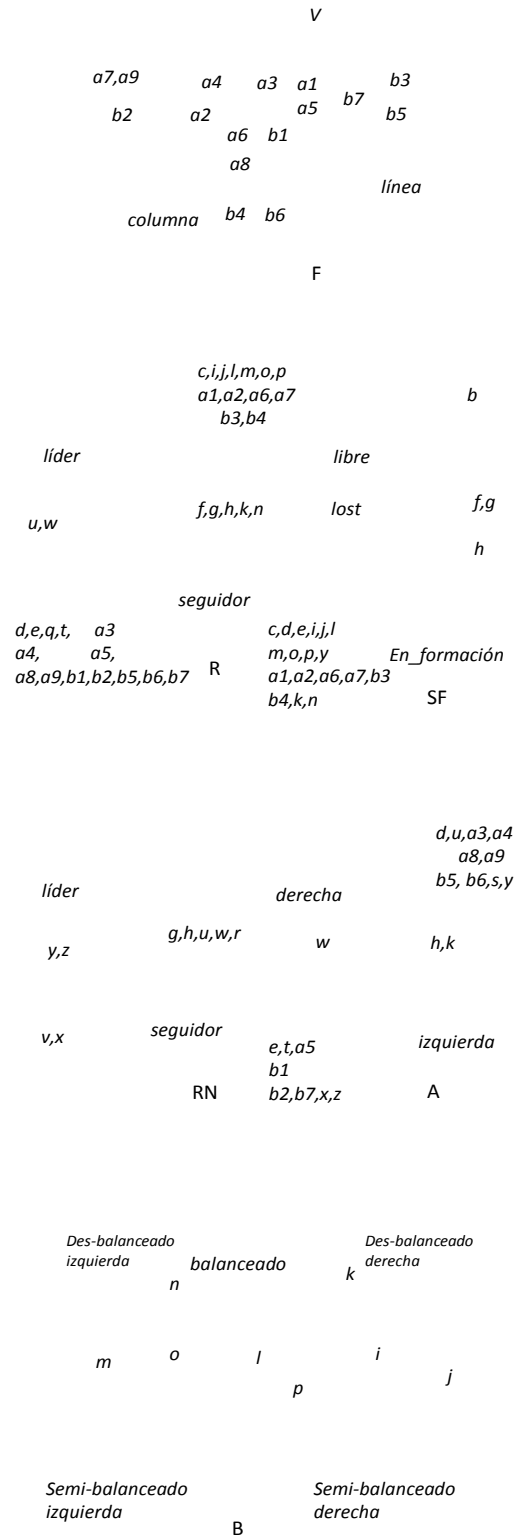


Figura 9. Red de Petri del controlador.

Formación (F): representa la formación que se esta realizando. *Role (R)*: mantiene el rol actual del agente. *Estado de Formación (SF)*: indica si el agente ha sido integrado a la formación o se mantiene libre. *Alineación(A)*: identifica si el agente se encuentra a la derecha o a la izquierda. *Rol del Vecino (RN)*: representa el rol del vecino localizado al frente; por ejemplo, para un agente con alineación izquierda, esta información corresponde al vecino derecho. *Balance (B)*: almacena la información necesaria para identificar cuando la formación esta des-balanceada y que lado es el que tiene más elementos. Todas estas variables se representan en la figura 9.

A partir del modelo de red de Petri, las etiquetas de sincronizado y las señales de entrada, se pueden derivar las reglas de comportamiento. Estas reglas definen las acciones que cada agente tiene que realizar.

IV. RESULTADOS DE SIMULACIÓN

El objetivo de esta simulación es validar e ilustrar como un grupo de agentes móviles distribuidos aleatoriamente en un ambiente de tres dimensiones, es capaz de establecer una formación. Se considera una formación en V con nueve agentes. Las simulaciones fueron realizadas en NetLogo [19].

El rol de líder se asigna a uno de los agentes, mientras que el resto son distribuidos aleatoriamente en el ambiente (Figura 10).

Cada agente libre se mueve hacia los agentes en formación, los cuales emiten una señal que es detectada por el resto de agentes y son integrados en la formación siguiendo la trayectoria del líder (Figura 11).

Conforme nuevos agentes son integrados al grupo, es posible que la formación se deforme debido a que los agentes no son integrados en posiciones predefinidas. El proceso de balanceo se lleva a cabo cada vez que una deformación es detectada (Figura 12).

Siguiendo el algoritmo propuesto, los agentes pueden establecer una formación a través de interacciones locales entre vecinos. A través de varias simulaciones se pudo demostrar como comportamientos complejos se pueden obtener a partir de reglas simples e interacciones locales.

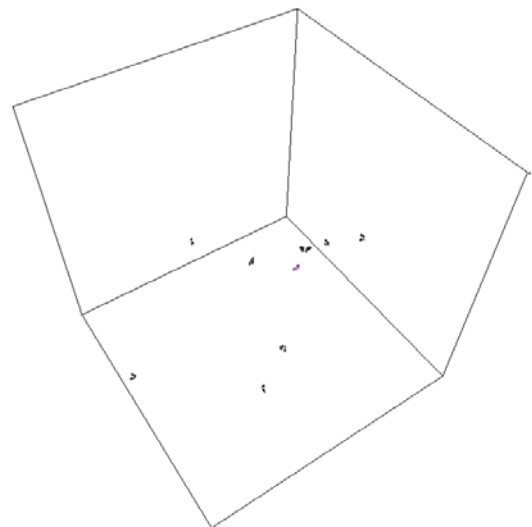


Figura 10. Grupo de agentes móviles distribuidos aleatoriamente en el ambiente.

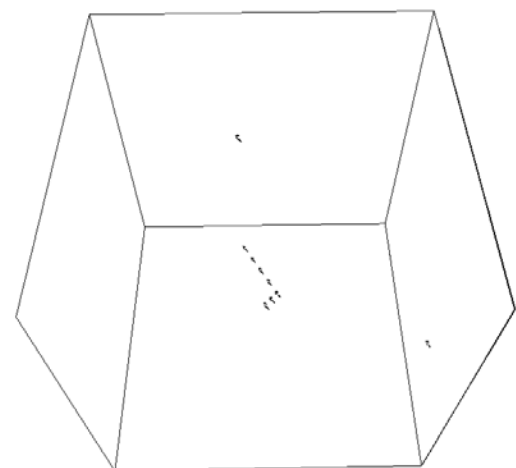


Figura 11. Agentes en una formación V.

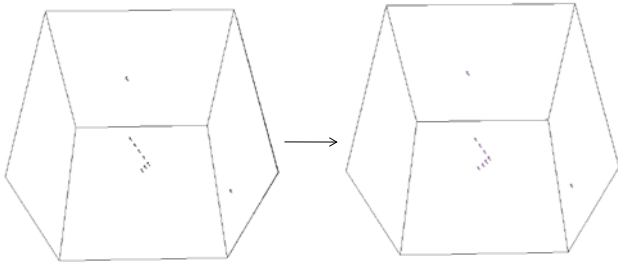


Figura 12. Balance de la formación.

V. CONCLUSIONES

Se ha presentado una metodología para construir algoritmos eficientes para el establecimiento de formaciones de agentes móviles. El comportamiento global de la formación emerge a partir de simples comportamientos individuales generados por el procedimiento de control; con esta estrategia los agentes son capaces de establecer la formación independientemente del número de agentes, así como intercambiar roles cuando se requiere.

Se ha demostrado a través de simulaciones como la auto-organización provee un comportamiento robusto, manteniendo las propiedades de escalabilidad y adaptabilidad en el grupo de agentes.

Gracias a la estrategia basada en interacciones locales y la complejidad del algoritmo, el procedimiento de control puede reducir el consumo de energía de los dispositivos móviles que implementen el agente.

VI. REFERENCIAS

- [1] F. Amigoni, M. S. Giani and S. Napolitano, "A simulation analysis of formations for flying multirobot systems", Proc. Eight Int. Conf. On Intelligent Autonomous Systems, Amsterdam, The Netherlands, pp. 149-156, March 2004.
- [2] D. Capera, J-P. Georgé, M-P. Gleizes and P. Glize, "The AMAS theory for complex problem solving based on self-organizing cooperative agents", Proc. 12th IEEE International Workshops on Enabling Technologies Infrastructure for Collaborative Enterprises, Linz, Austria, pp. 383-388, June 2003.
- [3] S. Chao-xia, H. Bing-rong and W. Yan-qing, "Cooperative exploration by multi-robots without global localization", Advanced Robotic Systems, Vol. 5, No. 2, pp. 129-138, June 2008.
- [4] M. Ferry, Z. Tu, L. Stephens and G. Prickett, "Towards true UAV autonomy", Proc. Information, Decision and Control, Adelaide, Australia, pp. 170-175, February 2007.
- [5] G. Hattenberger, R. Alami and S. Lacroix, "Planning and control for unmanned air vehicle formation flight", Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, pp. 5931-5936, October 2006.
- [6] G. Hattenberger, S. Lacroix and R. Alami, "Formation flight: Evaluation of autonomous configuration control algorithms", Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, California, pp. 2628-2633, November 2007.
- [7] H. Hexmoor, B. McLaughlan and M. Baker, "Swarm control in unmanned aerial vehicles", Proc. International Conference on Artificial Intelligence, Las Vegas, Nevada, Vol. 2, pp. 911-917, June 2005.
- [8] F. Heylighen, "Self-organization, emergence and the architecture of complexity", Proc. of the 1st European Conference on System Science, Paris, France, pp. 23-32, 1989.
- [9] O. Holland, J. Woods, R. D. Nardi and A. Clark, "Beyond swarm intelligence: the ultraswarm", Proc. IEEE Swarm Intelligence Symposium, Pasadena, California, pp. 217-224, June 2005.

- [10] R. Nagpal, “A catalog of biologically-inspired primitives for engineering self-organization”, Engineering Self-organising Systems, LNCS, Springer, pp. 53-62, 2004.
- [11] T. Paul, T. R. Krogstad and J. T. Gravdahl, “Modelling of UAV formation flight using 3D potential field”, Simulation Modelling Practice and Theory, Vol. 16, No. 9, pp. 1453-1462, October 2008.
- [12] P. Seiler, A. Pant and J. K. Hedrick, “Preliminary investigation of mesh stability for linear systems”, Proc. Asme Dynamic Systems and Control Division, Nashville, Tennessee, pp. 359-364, November 1999.
- [13] G. D. M. Serugendo, N. Foukia, S. Hassas, A. Karageorgos, S. K. Mostéfaoui, O. F. Rana, M. Ulieru, P. Valckenaers and C. V. Aart, “Self-organization: paradigms and applications”, Engineering Self-organising Systems, LNCS, Springer, pp. 1.19, 2004.
- [14] G. D. M. Serugendo, M-P. Gleizes and A. Karageorgos, “Self-organization and emergence in MAS: An overview”, Informatica, Vol. 30, No. 1, pp. 45-54, January 2006.
- [15] E. Shaw, H. Chung, J. K. Hedrick and S. Sastry, “Unmanned helicopter formation flight experiment for the study of mesh stability”, Cooperative Systems Control and Optimization, LNCS, Springer, pp. 37-56, March 2007.
- [16] G. S. Sukhatme, J.F. Montgomery and R.T. Vaughan, “Experiments with cooperative aerial-ground robots”, Robot teams: from diversity to polymorphism, T.Balch and L.E. Parker, A K Peters, Ltd., Ma, 2002.
- [17] P. Theodorakopoulos and S. Lacroix, “A strategy for tracking a ground target with a UAV”, Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, Nice, France, pp. 1254-1259, September 2008.
- [18] X. Wang, V. Yadav and S. N. Balakrishnan, “Cooperative UAV formation flying with obstacle/collision avoidance”, IEEE Transactions on Control Systems Technology, Vol. 15, No. 4, pp. 672-679, July 2007.
- [19] U. Wilensky, “Netlogo” Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. (1999). <http://ccl.northwestern.edu/netlogo/>.

VII. AUTORES

M.C. Miguel Angel Sánchez Acevedo obtuvo su título de Maestría en Ciencias con Especialidad en Ingeniería Eléctrica en el CINVESTAV Guadalajara. Actualmente estudia el Doctorado en Ciencias en la misma institución.

Dr. Félix Ramos Corchado es Doctor en Computación por la Universidad de Tecnología de Compiègne, Francia. Es miembro del Sistema Nacional de Investigadores. Actualmente, es profesor investigador en el CINVESTAV Guadalajara.

Dr. Ernesto López Mellado es Doctor Ingeniero con Especialidad en Robótica por la Universidad Paul Sabatier, Toulouse, Francia. Es miembro del Sistema Nacional de Investigadores. Actualmente, es profesor investigador en el CINVESTAV Guadalajara.

Reconocimientos:

Esta investigación es parcialmente apoyada por el proyecto CoECyT-Jal No. 2008-05-97094, mientras el autor M.A. Sánchez agradece el apoyo de la beca CONACYT no. 203124.

IMPLEMENTACIÓN DE UN MAPA AUTO-ORGANIZATIVO DE KOHONEN EN FPGA

Garduza-González, Sergio.

Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas
 Instituto Politécnico Nacional, CINVESTAV-IPN
 Av. IPN No. 2580, Col. Barrio La Laguna Ticomán, México, D.F. 07340, México
 Correo electrónico: sgarduza@ipn.mx

RESUMEN

Las Redes Neuronales Artificiales (RNA) no supervisadas, han sido implementadas en su mayoría, a nivel software, por lo tanto su ejecución ha sido en arquitecturas de propósito general. Particularmente, el Mapa Auto-Organizativo de Kohonen (MAO) es un algoritmo que por sus extensas aplicaciones, es deseable implementar a nivel hardware, para así acelerar el procesamiento de los datos. En este artículo se propone una arquitectura de hardware dedicado, que ejecuta el algoritmo del MAO. Para llevar a cabo la implementación se realizó una simplificación del algoritmo original y se seleccionó una arquitectura de Única Instrucción para Múltiples Datos (IUMD). Con el hardware diseñado, se entrenó un prototipo elemental de MAO de 5×5 neuronas de salida, en un espacio bidimensional. Los resultados mostraron que el tiempo de ejecución en el hardware propuesto fue menor frente a la tiempo de ejecución en una arquitectura de microprocesador de propósito general; también ofrece la ventaja de que se puede incrementar el número de neuronas de salida sin modificar el tiempo de ejecución. Con esto se verifica que esta arquitectura llega a ser una forma más rápida y eficiente para la ejecución de un MAO.

Palabras clave: Arquitectura IUMD, FPGA, Mapa Auto-Organizativo de Kohonen, Mapas sensoriales, Red neuronal artificial no supervisada.

I. INTRODUCCIÓN

Los modelos de RNA surgen como un intento por describir los fenómenos físicos, la tolerancia a fallas, y la capacidad de aprendizaje, presentes en

los sistemas neuronales biológicos. Los modelos neuronales biológicos y artificiales reducen el procesamiento a un conjunto elemental de unidades de información llamadas neuronas. En consecuencia, los modelos de RNA se han inspirado directamente de los modelos neuronales biológicos. Particularmente el MAO, fue inspirado a partir de las investigaciones realizadas en la corteza cerebral. A escala macroscópica la corteza cerebral se compone de diversas regiones espaciales constituidas por un elevado número de neuronas similares, las cuales colaboran para realizar las tareas específicas para las cuales se han especializado; por ejemplo, el análisis de señales sensoriales [1]. A escala microscópica, se ha observado que las señales aparecen en el mismo orden topográfico, tanto en la corteza cerebral como en el órgano sensorial que las recibió. A estas estructuras se les denominan Mapas Sensoriales. Inspirándose en esta proyección topográfica de la información, Teuvo Kohonen, propuso por primera vez un algoritmo de RNA no supervisada, denominado Mapa Auto-Organizativo de Kohonen [2].

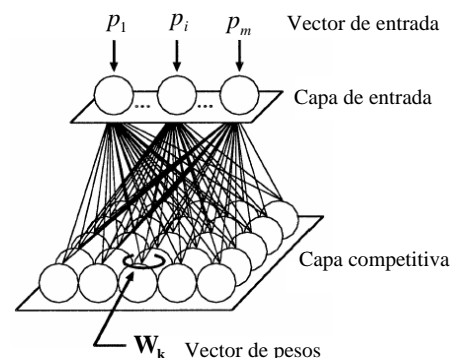


Fig. 1. Arquitectura del MAO de Kohonen con capa de salida bidimensional.

El tipo de redes de aprendizaje no supervisadas son aquellas a las que se proporciona una secuencia de vectores de entrada, pero no se proporciona un vector objetivo. Durante el proceso de aprendizaje la red debe descubrir por sí misma rasgos comunes, regularidades, correlaciones o categorías en los datos de entrada, e incorporarlas a su estructura interna de conexiones (pesos). De esta forma se dice que las neuronas deben auto-organizarse en función de los estímulos (señales o datos) provenientes del exterior. Para obtener resultados aceptables, la red requiere cierto grado de redundancia procedente del espacio sensorial, es decir, un número suficientemente grande de patrones de aprendizaje.

El MAO de Kohonen hace uso de un algoritmo para Auto-Organizar los datos y así poderlos visualizar. Este algoritmo fue desarrollado, en principio, para la visualización de relaciones no lineales de datos multidimensionales. Sin embargo, actualmente se han extendido sus aplicaciones. Una revisión amplia de aplicaciones para este tipo de red se presenta en [2, 3].

Debido a que este algoritmo debe procesar gran cantidad de información, una implementación a nivel software que se ejecuta en una arquitectura del tipo secuencial, resulta en largos tiempos de entrenamiento, por ello últimamente se han desarrollado implementaciones a nivel hardware en modalidad ASIC. Un ejemplo de esto se reporta en [4], el cual consiste en un ASIC procesador KNN (*Kohonen Neural Network*), asistido por un FPGA (*Field Programmable Gate Array*); éste KNN cuenta con 4 elementos de procesamiento y es utilizado para reconocimiento de patrones. La principal desventaja radica en que, siendo un ASIC, no es posible adaptarlo a una nueva aplicación y el costo de fabricación es elevado.

Para nuestra implementación hemos optado por la tecnología FPGA, ya que, además del bajo costo, como se ha mostrado en diversas aplicaciones, la principal ventaja es la capacidad de reconfiguración. De esta manera para el caso de un MAO se puede incrementar el número de redes de salida, haciendo uso de las compuertas

reconfigurables contenida en los bloques de un FPGA.

El MAO de Kohonen es una transformación matemática que toma los datos de entrada como vectores y los transforma hacia vectores de salida. En los siguientes apartados se presentan los elementos esenciales para llevar a cabo esta transformación y la prueba implementada.

II. ARQUITECTURA DE LA RED NEURONAL

La Fig. 1, muestra la arquitectura neuronal del MAO. En esta arquitectura las neuronas están organizadas en dos capas. La primera es la capa de entrada o sensorial, ésta consiste en m neuronas que se comportan como buffers, distribuyendo la información del exterior, a las neuronas de la segunda capa. El procesamiento se realiza en la segunda capa que forma el mapa de rasgos y consiste habitualmente en una estructura rectangular de $n_x \times n_y$ neuronas, que operan en paralelo. Cada neurona de entrada p_i está conectada a todas las neuronas del mapa, mediante un peso sináptico w_{xy} . De esta forma, el mapa se describe como una matriz de procesadores elementales localizados en un espacio bidimensional.

III. ALGORITMO DE ENTRENAMIENTO

Para que la red de salida aprenda un patrón de entrada en particular, se calcula primero la distancia entre cada vector de peso y el patrón de entrada. El nodo que tiene la distancia más corta es declarado nodo ganador y los pesos de este nodo son actualizados. Los pesos de los nodos dentro de la vecindad del nodo ganador también son actualizados.

El algoritmo procede primero por inicializar los siguientes parámetros:

- Pesos sinápticos (p_m).
- Parámetros de vecindad topológica.
- Razón de aprendizaje (α).

Una vez inicializada, hay tres fases esenciales en la formación del MAO.

COMPETENCIA

Se determina cuál es la neurona ganadora. El parámetro de coincidencia entre el vector de entrada y el vector de peso de cada neurona se estima por medio de la distancia Euclidiana entre el vector de entrada \mathbf{p}_j y el vector de peso \mathbf{w}_k :

$$\|\mathbf{p}_j - \mathbf{w}_k\| = \sqrt{\sum_{m=1}^n (p_{jm} - w_{km})^2} \quad (1)$$

La distancia mínima para un vector de entrada \mathbf{p}_j se expresa como $c(\mathbf{p}_j)$:

$$c(\mathbf{p}_j) = \arg \min_k \|\mathbf{p}_j - \mathbf{w}_k\|, \quad k = 1, 2, \dots, l \quad (2)$$

COOPERACIÓN

La neurona ganadora establece la localización espacial de una vecindad topológica de neuronas vecinas, que también serán actualizadas. Para esto hay que definir una función $h(\cdot)$, denominada función de vecindad. Una elección ampliamente utilizada es la función Gaussiana, definida en (3), puesto que permite una adaptación suave. Sin embargo, otras funciones más simples también son aceptables [5].

$$h_c = \exp\left(-\frac{d_{k,c}^2}{2\sigma^2}\right) \quad (3)$$

Donde d_{kc}^2 es la distancia lateral y σ la varianza, que es un parámetro importante para definir el área de cobertura de la función de vecindad.

La distancia lateral se encuentra como:

$$d_{k,c}^2 = \|\mathbf{r}_k - \mathbf{r}_c\|^2 \quad (4)$$

Donde \mathbf{r}_c , es el vector, cuyas coordenadas están definidas por la neurona ganadora.

La función (3), por un lado delimita una topología de vecindad simétrica (también llamada área de cobertura) y por otro lado, su amplitud (para esta función, σ) debe disminuir de forma monótona con el incremento de la distancia lateral.

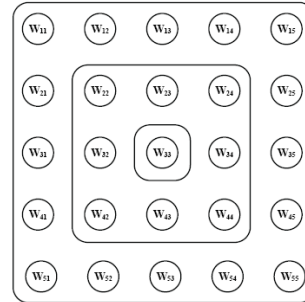


Fig. 2. Tres diferentes áreas para la función de vecindad rectangular, cuya neurona ganadora tiene como peso w_{33} .

Uno de los requisitos de convergencia es que el área de cobertura de la función de vecindad disminuya una vez concluida una época de entrenamiento, de tal modo que para la última época de entrenamiento sólo se actualizará la neurona ganadora.

ADAPTACIÓN SINÁPTICA

Esta última fase habilita las neuronas excitadas, es decir, ganadora y vecinas, para que incrementen o refuercen sus valores de pesos de acuerdo a la regla de Kohonen [2].

$$\mathbf{w}_k(n+1) = \mathbf{w}_k(n) + \alpha \cdot h_{k,c} \cdot (\mathbf{p}_j - \mathbf{w}_k(n)) \quad (5)$$

Donde:

α es la razón de aprendizaje (entre 0 y 1).

$w_k(n)$ es el peso actual de la neurona k para la época n .

$w_k(n+1)$ es el valor de peso actualizado.

IV. SIMPLIFICACIÓN DEL ALGORITMO PARA SU IMPLEMENTACIÓN

Algunas consideraciones sobre el algoritmo de aprendizaje son necesarias para optimizar recursos del FPGA. La primera de ellas es acerca del cálculo de la distancia Euclidiana. Ya que esto implica una raíz cuadrada, como lo indica (1), un cálculo simplificado es la distancia Manhattan, que se obtiene de:

$$\|\mathbf{p}_j - \mathbf{w}_k\| = \sum_{m=1}^l |p_{jm} - w_{km}| \quad (6)$$

En (6) sólo es necesario estimar una sumatoria de diferencias, ofreciendo una gran ventaja.

La razón de aprendizaje α se puede aproximar a una función del tipo:

$$\alpha \in \frac{1}{2^r}, \quad (r = 1, 2 \dots m) \quad (7)$$

Esta simplificación muestra que la razón de aprendizaje es inversamente proporcional a una potencia de base 2. En una cantidad binaria esto se obtiene al desplazar los dígitos una posición a la izquierda, para $r=1$, dos posiciones a la izquierda para $r=2$, y así sucesivamente.

Por último, la función de vecindad (Gaussiana), definida por (3), puede cambiarse por funciones regulares geométricas. A pesar de ello, se obtienen buenas aproximaciones [5, 6]. Por simplificación se ha elegido una forma topológica rectangular, como se muestra en la Fig. 2.

V. PROPUESTA DE ARQUITECTURA IUMD PARA EL MAO DE KOHONEN

Para la implementación del algoritmo del MAO Kohonen partimos de una arquitectura IUMD o “Instrucción Única para Múltiples Datos”. Esta arquitectura consiste de un número de Elementos de Procesamiento (EP), que ejecutan el mismo programa para diferentes datos, bajo la supervisión de una unidad de control; en otras palabras, esta unidad de control central despacha las instrucciones a diferentes unidades de procesamiento y por lo tanto cada procesador, aunque recibe la misma instrucción, es capaz de operar sobre diferentes conjuntos de datos. Las computadoras IUMD están especialmente diseñadas para realizar cómputos vectoriales sobre matrices o arreglos de datos.

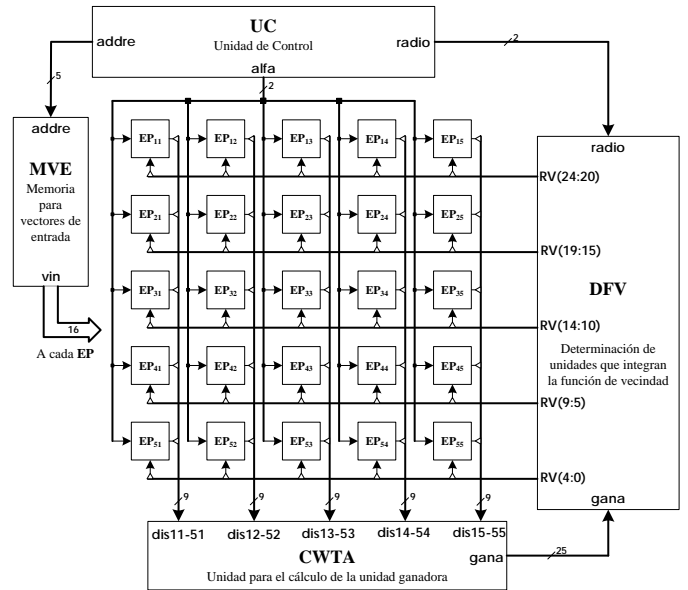


Fig. 3. Arquitectura IUMD con los bloques que implementa el MAO de Kohonen.

El algoritmo de aprendizaje simplificado se implementó en un FPGA Spartan-3 de XILINX®, programado en VHDL. La implementación consiste en un MAO bidimensional de 5x5 neuronas de salida; cada unidad de salida tiene segmento x y y , expresados con una precisión de 8 bits. Lo que en realidad entrega el FPGA en sus puertos son dos bytes que representan coordenadas espaciales. Estas coordenadas son utilizadas posteriormente para representar gráficamente la información.

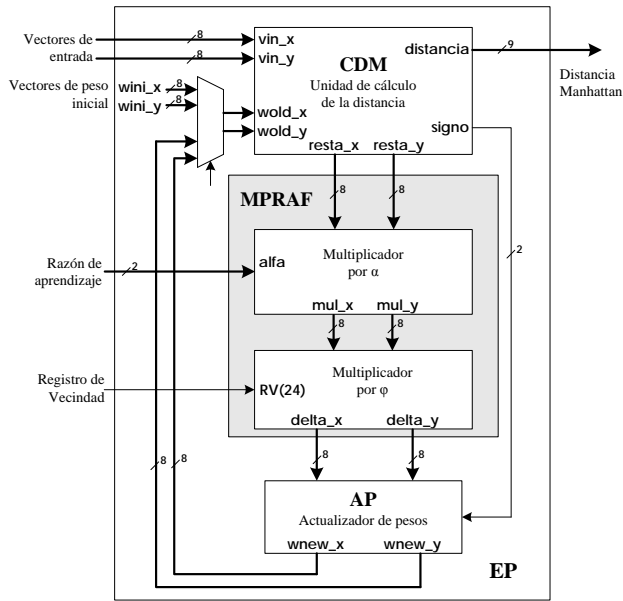


Fig. 4. Elementos que integran cada unidad EP (Elemento de Procesamiento).

La Fig. 3 muestra la arquitectura propuesta, donde se identifican los siguientes bloques:

- UC. Unidad de control.
- MVE. Memoria para los vectores de entrada.
- DFV. Determina las unidades que integrarán la función de vecindad.
- CWTA. Unit Winner-Take-All.
- EP. Elementos de Procesamiento, uno para cada neurona de salida.

En la unidad MVE se mantienen los vectores de prueba con los cuales se va a entrenar a la red. Cuando es iniciado el proceso de entrenamiento, la UC genera una dirección de 5 bits por el registro **addr**, el cual provoca que por **vin** se presente las componentes *x* y *y* del primer vector de prueba. Este vector llega a cada uno de los EP (Fig. 4). Estos a su vez toman el vector de entrada y con el valor de sus pesos iniciales (**wini_x**, **wini_y**) estiman la distancia Manhattan (en CDM) la cual es entregada por el registro **distancia** a la unidad DFV.

Hasta este punto, como vemos en la Fig. 3, todos los registros **distancia** llegan a la unidad CWTA, encargada de determinar cual unidad es la ganadora (la distancia mínima). El hardware

implementado en CWTA es implementado como en [7], y opera de forma serial. El número de ciclos que requiere es proporcional al número de bits contenidos en el registro **distancia** (11 ciclos); al término de la ejecución, la unidad CWTA, a través del registro **gana**, determina cuál de las neuronas es la ganadora.

Como se mencionó antes, la neurona ganadora determina la ubicación espacial de la función de vecindad, en consecuencia, la unidad DFV toma las coordenadas de la neurona ganadora e indica por el registro **RV** qué neuronas deberán actualizarse. La unidad DFV, en esencia, es una *Look-Up Table* para implementar la función rectangular como se muestra en la Fig. 2.

Un bit del registro **RV** es necesario para indicarle a cada unidad EP si deberá actualizar su peso o conservar su valor actual. Esto se determina en la unidad MPRAF. Si es necesario actualizar el peso, entonces, de acuerdo a (5), se le agrega:

$$\alpha \cdot h_{k,c} \cdot (p_j - w_k(n)) \quad (8)$$

Esta expresión es implementada por los bloques Mul- α y Mul- ϕ , en la Fig. 2. La unidad AP realiza la adición de la cantidad (8) o cero en caso de que no se tenga que actualizar la neurona.

Cada uno de los pasos descritos anteriormente es controlado por el módulo UC, el cual contiene una máquina de estado que se describe en la Fig. 5.

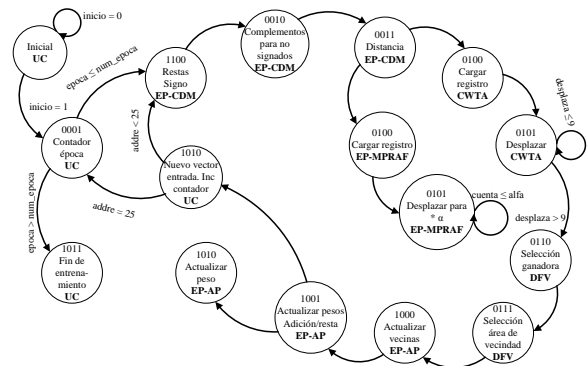


Fig. 5. Máquina de estado implementada en la Unidad de Control (UC). En cada estado se muestra la unidad de hardware que interviene.

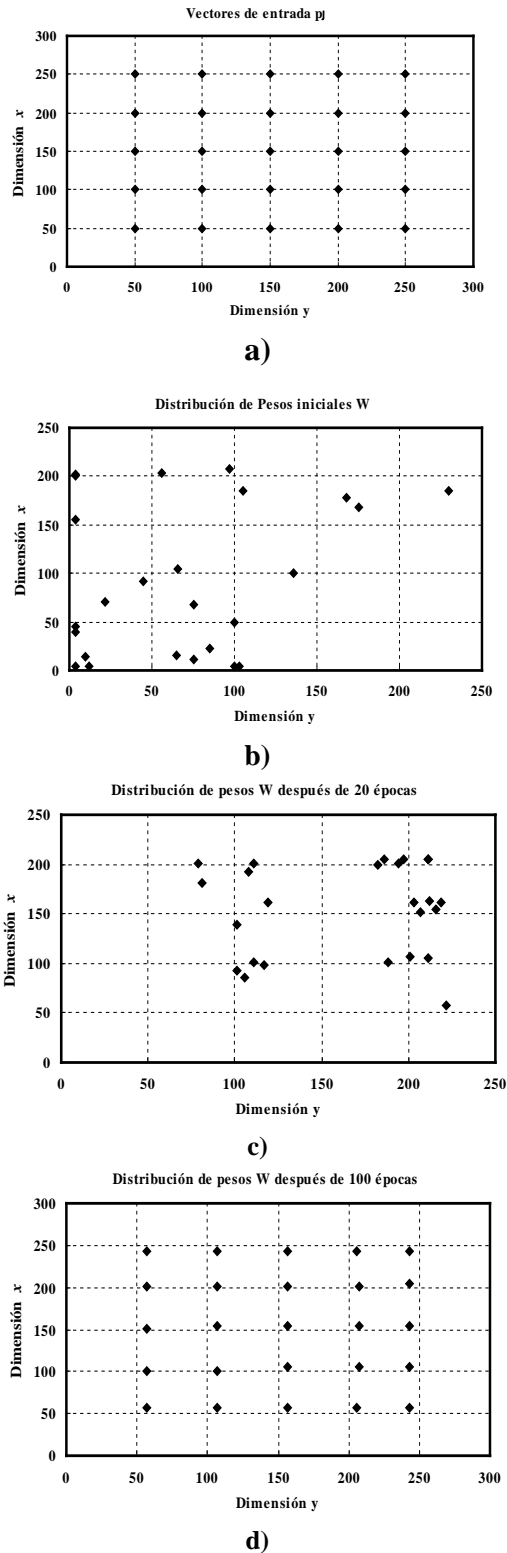


Fig. 6. Entrenamiento en el hardware propuesto para SOM de Kohonen de 25 neuronas de salida. a) Vectores de entrada, b) Vectores de pesos iniciales. En c) y d) Vectores de pesos

después de 20 y 100 épocas de entrenamiento, respectivamente.

Una vez procesado el primer vector, la UC genera la dirección de memoria siguiente para ejecutar el correspondiente vector de prueba, y se vuelve a ejecutar el ciclo, y así sucesivamente hasta concluir con todos los vectores de entrada. En ese momento habrá terminado una época de entrenamiento.

Aunque existen pruebas analíticas de convergencia, este algoritmo no lo implementa. Así que sólo queda observar la evolución gráfica de los pesos.

VI. RESULTADOS

Para verificar la arquitectura se implementó una prueba de $l = 25$ neuronas (bidimensional de 5×5), con un patrón de entrada como el mostrado en la Fig. 6a, los pesos se inicializaron con valores aleatorios (Fig. 6b). Después de 20 épocas de entrenamiento, se observa una convergencia de los valores (Fig. 6c), y finalmente en la Fig. 6d, para 100 épocas de entrenamiento se muestra una aproximación muy cercana al patrón de entrada.

De acuerdo al diagrama de estados (Fig. 5) en 21 ciclos de reloj se lleva a cabo la ejecución para una época de entrenamiento. El tiempo de ejecución total dependerá del tiempo del reloj (T_{ciclo}), el número de vectores de entrada (n_{vec}) y el número de épocas de entrenamiento (n_{ent}).

$$T_{ent} = n_{vec} n_{ent} T_{ciclo} \quad (8)$$

Así para este prototipo a una frecuencia de reloj de 25MHz se estimó un T_{ent} de 2.1ms. Para comparar su desempeño, hemos implementado el algoritmo simplificado del MAO en lenguaje C, y ejecutado en una PC con procesador Intel® Xeon® 5130@2GHz, resultando en un tiempo de ejecución de 85.95ms. Con esto podemos ver que un hardware dedicado acelera el tiempo de procesamiento, importante en aplicaciones con grandes cantidades de datos.

Los recursos de hardware utilizados por el FPGA se muestran en la tabla 1. De ahí se observa que el

Utilización de Lógica	Usado	Disponible	Utilización
Número de bloque conteniendo Flip-Flops	3,141	15,360	20%
Número bloques	3,189	7,680	41%
Número total de LUTs de 4 entradas	5,390	15,360	35%
Usados para lógica	5,362	15,360	34%
Número de bloques IOB	102	173	58%
Número de compuertas equivalentes	66,381		

Tab. 1. Resumen de recursos utilizados del FPGA, XILINX Spartan 3.

número de compuertas equivalentes usadas para este diseño es de 66,381 de 100,000, lo cual indica que estamos utilizando aproximadamente un 70% de los recursos de este dispositivo.

Es importante mencionar, que este diseño, al ser un prototipo, presenta desventajas en la cantidad de datos a procesar (25 neuronas), y la precisión de los datos, frente a la ejecución en software; sin embargo los planteamientos para la arquitectura del hardware mostrados en este artículo pueden ser fácilmente implementados en nuevas plataformas de FPGA.

VII. CONCLUSIONES

Se han descrito y analizado los elementos que contiene el hardware dedicado para la implementación del algoritmo de Kohonen simplificado. Estas simplificaciones no afectaron las características importantes para la formación del MAO. La arquitectura es una máquina IUMD, donde se construye un elemento de procesamiento para cada neurona, y los módulos de la implementación de la función de vecindad y el cálculo de la función ganadora son comunes a estos elementos. Partiendo de este diseño se abordó la evaluación del sistema, teniendo como principal resultado una disminución considerable en el tiempo de ejecución, respecto del algoritmo ejecutado en una computadora convencional.

VIII. ANEXO I. LISTA DE SÍMBOLOS Y ABREVIATURAS

Símbolos:

\mathbf{p}_j = Señales o vector de entrada a procesar.

$p_{j1}, p_{j2}, \dots, p_{jm} = m$ elementos del vector \mathbf{p}_j .

\mathbf{w}_k = Vector de peso a procesar.

$w_{k1}, w_{k2}, \dots, w_{km} = m$ elementos del vector \mathbf{w}_j .

$c(\mathbf{p}_j)$ = Distancia mínima entre \mathbf{p}_j y \mathbf{w}_k .

$h(\cdot)$ = Función de vecindad.

h_c = Función de vecindad tipo Gaussiana.

α = Razón de aprendizaje.

n = Épocas de entrenamiento.

l = Número total de neuronas.

Abreviaturas

MAO = Mapa Auto-organizativo de Kohonen.

CDM = Unidad de cálculo de distancia.

MRAF = Multiplicador por razón de aprendizaje y función de vecindad.

AP = Actualizador de pesos.

CWTA = Unidad para el cálculo de la unidad ganadora.

EP = Elemento de procesamiento.

UC = Unidad de control.

MVE = Memoria para los vectores de entrada \mathbf{p}_j .

DFV = Determinación de unidades que integran la función de vecindad.

Nomenclatura de registros

vin_x, vin_y = Contienen los elementos p_{jx} y p_{jy} de \mathbf{p}_j , 1 byte cada uno.

distancia = Contiene el valor de la distancia calculada. 9 bits.

signo = Signo que resulta de $p_{jm} - w_{km}$. 2 bits.

resta_x y resta_y = Contiene $p_{jx} - w_{kx}$ y $p_{jy} - w_{ky}$. 1 byte cada uno.

wini_x y wini_y = Elementos del vector de peso inicial.

wold_x y wold_y = pesos actuales. 1 byte cada uno.

alfa = selector de razón de aprendizaje. 2 bits.

mul_x y mul_y = Contiene el resultado de $\alpha \cdot (\mathbf{p}_j - \mathbf{w}_k(n))$. 1 byte cada uno.

delta_x y delta_y = contiene el resultado de $\alpha \cdot h_{i,c} \cdot (\mathbf{p}_j - \mathbf{w}_k(n))$. 1 byte cada uno.

wnex_x y wnew_y = Pesos actualizados: $\mathbf{w}_k(n) + \alpha \cdot h_{k,c} \cdot (\mathbf{p}_j - \mathbf{w}_k(n))$. 1 byte cada uno.

RV = Registro de vecindad de 24 bits.

adre = dirección de la MVE.

radio = radio de cobertura de $h(\cdot)$. 2 bits.

gana = indica cuál es la unidad ganadora. 25 bits, un bit para cada neurona.

IX. REFERENCIAS

- [1] H. Ritter, T. Martinetz, K. Schulten, *Neural computation and self-organizing maps: An introduction*, Reading, Mass.: Addison Wesley, 1992.
- [2] T. Kohonen, *Self Organizing Maps*, 3rd ed. Berlin, New York: Springer, 2001.
- [3] T. Kohonen, G. Barna, R. S. Chrisley, “*Statistical pattern, recognition with neural networks: benchmarking studies*”, Proc. IEEE Int. Conf. on Neural Networks, vol. I-61, 1998.
- [4] A. Rajah, M. Khalil Hani, “*ASIC design of a Kohonen neural network microchip*”, Proc. IEEE International Conference on Semiconductor Electronics, 2004.
- [5] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed., New Jersey: Prentice Hall, p. 38, 1999.
- [6] H. Tamukoh, K. Horio, T. Yamakawa, “*Fast Learning Algorithms for Self-Organizing Map Employing Rouge Comparison WTA and its Digital Hardware Implementation*”, IEICE Trans. Electron., vol. E87-C, no. 11, pp. 1787-1794, Nov. 2004.
- [7] T. Yamakawa, K. Horio, y T. Hiratsuka, “*Advanced self-organizing maps using binary weight vector and its digital hardware design*”, Proc. 9th Int. Conf. on Neural Information Processing, vol. 3, pp. 1330-1335, 2002.

X. AUTORES

M. en C. Sergio Garduza-González recibió el grado de Ingeniero en Electrónica y Comunicaciones en 2002 por la Universidad Veracruzana, Veracruz, México. En 2006 recibió el grado de M. en C. en Ingeniería Eléctrica por el CINVESTAV-IPN, D. F., México. Sus áreas de interés son: Hardware dedicado para RNA en FPGA, Herramientas EDA para el modelado de circuitos CMOS.

VISUALIZACIÓN DISTRIBUIDA EN TIEMPO REAL DE UN AMBIENTE VIRTUAL

Alma Martínez, Félix Ramos, Mario Siller.

Centro de Investigación y Estudios Avanzados del I.P.N (Cinvestav), Unidad Guadalajara
Av. Científica 1145, colonia el Bajío, Zapopan, 45015, Jalisco, México.
Tel: (33) 3777-3600, correo electrónico: {vmartine, framos, msiller}@gdl.cinvestav.mx

RESUMEN

Hoy en día, los ambientes virtuales distribuidos se utilizan en varias áreas de investigación, ya que estas permiten la simulación y el análisis de diversas situaciones del mundo real o del mundo sintético. En los casos de aplicaciones naturalmente distribuidas, estos ambientes se deben poder acceder simultáneamente por un gran número de usuarios situados en diferentes lugares geográficos. En este trabajo, presentamos una arquitectura distribuida para la visualización en tiempo real de Ambientes Virtuales en 3D, la cual utiliza las ventajas de la comunicación Peer-To-Peer. Garantizando una visualización coherencia y consistente entre los usuarios del Ambiente Virtual (AV).

Palabras clave: Sistemas Distribuidos, Peer-To-Peer, Áreas de interés, Render.

I. INTRODUCCIÓN

En los últimos años, la utilización de ambientes virtuales en diferentes áreas de investigación ha tomado un gran auge, esto debido a la facilidad que nos presenta para simular situaciones tanto de ambientes reales como ambientes sintéticos. Situaciones tales como es el comportamiento de multitudes en situaciones de emergencia [1], simuladores de vuelos para el entrenamiento de personal militar [2], así como en la educación [3].

El desarrollo de una ambiente virtual involucra diferentes aspectos, desde equipos encargados del diseño del AV, personal adecuado para el diseño del comportamiento de cada componente del AV, así como un grupo con habilidades especiales que

se encargue de integrar cada aspecto, lo cual puede llevar un tiempo considerable.

Para facilitar las tareas mencionadas, es útil considerar las características necesarias para la visualización de la evolución de un AV, esto tomando en cuenta la participación de varios usuarios. Para lo cual es necesario tener en cuenta la forma de distribución, en este tema últimamente se ha utilizado características de la comunicación Peer-To-Peer (P2P), la cual nos brinda la ventaja de disminuir la dependencia de servidores. En este trabajo, presentamos un nuevo enfoque para la visualización 3D en tiempo real de AV distribuidos basados en una comunicación P2P. Se diseño una distribución consistente y coherente para todos los usuarios implicados en una definición determinada del AV, utilizando dos características principales: la visualización local del AV y gestión de la comunicación entre ellos. Ambas características están pensadas para que la visualización se realice en tiempo real.

Este artículo está organizado de la siguiente manera. La siguiente sección describe el trabajo relacionado con nuestra área de investigación. En la sección III se describe la propuesta, la cual está dividida principalmente en dos sub-secciones, la visualización local del AV y la comunicación entre los usuarios del AV. La sección IV contiene las conclusiones obtenidas de este trabajo.

II. TRABAJO RELACIONADO

Para realizar una correcta distribución de la visualización del ambiente es necesario considerar factores tales como el ancho de banda, el retardo y pérdida de los mensajes, la topología de la red [4, 5, 6, 7]. Debido a que en el Internet, el retardo y el

ancho de banda es variable, así como la interacción de los usuarios es la pauta de la generación de mensajes en la red, resulta una tarea difícil mantener un ambiente virtual distribuido de forma coherente.

Debido a lo mencionado, existen trabajos enfocados en un área en especial. Tal es el caso de [8, 9, 10], el cual analiza el retardo existente, para la estimación del ancho de banda [11, 12], así como la carga de trabajo generada en el cliente y servidor [13, 14]. Pero ninguno de ellos toma en cuenta desde la generación de acciones por el usuario hasta la llegada hacia los demás usuarios en el ambiente, lo cual es importante debido a que se puede detectar cual es el cuello de botella y de esta forma administrar mejor los recursos para una mejor visualización.

Un campo de investigación es la comunicación entre los usuarios, para lo cual, existen diversos trabajos relacionados con el diseño de aplicaciones interactivas distribuidas, que analizan la comunicación Peer-To-Peer (P2P) [15, 16, 17]. Dichas investigaciones agrupan a los usuarios formando grupos reducidos que utilizan la arquitectura cliente-servidor para realizar su comunicación, provocando tener dependencia en los servidores de cada grupo, así como en la comunicación entre servidores. En este tipo de trabajos el servidor es quien realiza la actualización del ambiente, esto por medio de las solicitudes de los clientes. De tal forma, que el servidor está enviando constantemente el nuevo estado del ambiente, lo que puede provocar una saturación de mensajes en los canales de comunicación.

Es posible considerar a los videos juegos como un área de investigación relacionada con los ambientes virtuales [18], dado que este campo requiere el uso de la tecnología en los ambientes virtuales para facilitar el desarrollo de video juegos.

Un ambiente virtual puede ser dividido en 2 principales módulos [18, 19], sin embargo, la mayoría de los investigadores trabajan en la lógica del ambiente, y toman poco en consideración en representación gráfica.

III. VISUALIZACIÓN EN TIEMPO REAL DE UN AV

Nuestra propuesta distribuye la visualización de un ambiente virtual, permitiendo la interacción de diferentes usuarios localizados en diferentes áreas geográficas. El objetivo es brindar a los usuarios una visualización distribuida de los ambientes virtuales consistente y en tiempo real.

Para lograr este objetivo se han definido dos características principales de la distribución: el despliegue gráfico de forma local para cada usuario y la administración de la comunicación entre estos. Nosotros proponemos administrar ambas características de tal manera que se reduzca de manera considerable el retardo en los cambios de estado del ambiente, y de esta forma poder realizar una visualización en tiempo real, y además que el usuario perciba coherentemente la evolución del AV.

El usuario final puede solicitar diversas actualizaciones, las cuales son tomadas por el puerto visual, el cual consta de una interfaz en la cual se despliegan gráficamente dichas actualizaciones (esta interfaz esta creada sobre un motor 3D, comúnmente utilizados en los videojuegos, ya que estos permiten la manipulación en tiempo real de la entidades del ambiente), una segunda parte es la arquitectura encargada de procesar las ordenes y de realizar la comunicación con los demás usuarios (Fig. 1).



Fig. 1: Arquitectura para la visualización Local.

La comunicación entre los diversos usuarios está basada en una arquitectura Peer-to-Peer [20], esta decisión es debido a que se desea eliminar lo más posible la dependencia de servidores (Fig. 2). Si conservamos una arquitectura Peer-To-Peer cada usuario en el ambiente tiene la capacidad de comprender las acciones que se están tomando en diferentes partes del ambiente, de esta forma no se depende de un interprete como lo son los servidores.

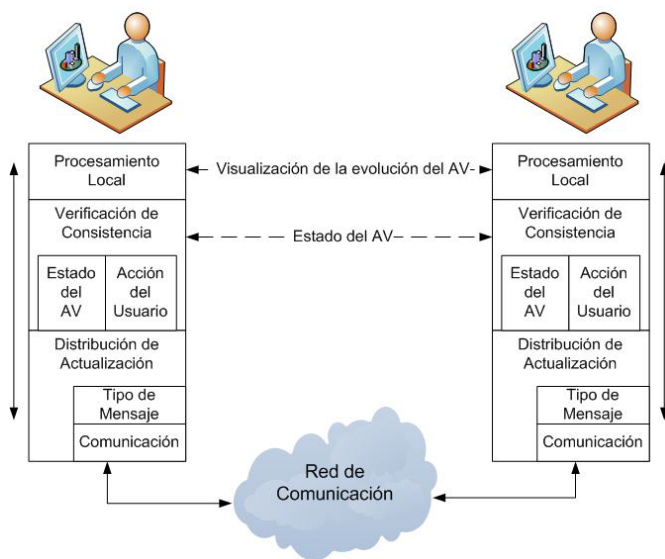


Fig. 2: Arquitectura Peer-To-Peer.

La arquitectura propuesta tiene una aproximación por capas, con el propósito de delimitar y disminuir la dependencia de las capas superiores e inferiores. El procesamiento local es el encargado tanto de actualizar la interfaz local, como de mantener una consistencia mínima local del ambiente, es decir, que las acciones que solicita el usuario tengan coherencia, por ejemplo: que las entidades que solicitan sean modificadas en realidad existan, que la modificación sea posible aplicarse, etc.

Una vez que se ha constatado que la actualización es posible, se realizan los cambios necesarios en el ambiente con los cuales se verifica la consistencia mínima local del ambiente, esto para saber el tipo de repercusión que debe visualizarse en el ambiente, de lo cual puede surgir alguna colisión entre las entidades o entre la entidad que realiza la

acción y el ambiente. El resultado de dicho cambio es notificado al usuario por medio de una cadena con formato XML, llamada evento. Un evento contiene la descripción del resultado de una solicitud. Un segundo paso es verificar la consistencia entre los usuarios del ambiente, para lo cual se generan mensajes de consistencia, estos contienen la descripción del área de interés al que pertenece el usuario en cuestión.

Los mensajes de consistencia son distribuidos por la capa de distribución de actualizaciones quien se encarga de seleccionar un medio fiable para su entrega hacia el coordinador del área de interés. Dicho coordinador tiene la tarea de realizar un acuerdo con los demás coordinadores el ambiente para establecer el estado correcto del ambiente.

A continuación se realiza una explicación más a detalle de las tareas realizadas para la visualización distribuida del ambiente en tiempo real, para lo cual se desglosan dos temas principales: la visualización en tiempo real, como se administran las actualizaciones de tal forma que el despliegue grafico no se vea pausado, la segunda, la distribución de mensajes, en lo cual se consideran la organización de los nodos dependiendo de su área de interés, así como los roles que toma cada usuario.

Visualización Local en Tiempo real

La primera tarea a realizar es analizar las acciones para constatar que son validas en el ambiente, es decir, que la entidad que se solicita existe y que la acción (animación) que se desea en realidad pertenece a dicha entidad.

Las entidades que podemos encontrar en un ambiente virtual (AV) se pueden clasificar en las siguientes:

$$AV = \{E, A, O\}$$

Escenario (E): es la descripción gráfica de la localidad donde se llevaran a cabo los cambios requeridos. Ejemplo: parque, casa, oficinas, escuela, ciudad, etc.

Avatares (A): son las entidades que representan a un entidad autónoma, por ejemplo: una persona, un animal, etc.

$$A = \{a_1, a_2, \dots, a_n\}$$

Un avatar (a) está formado por un conjunto de características, las cuales son: acciones (ac), personalidad (p) y un estado emocional (π), el cual contiene diferentes emociones (e).

$$a = \{p \times \pi \times ac\}$$

$$\pi = \{e_1, e_2, \dots, e_n\}$$

$$e = \{alegre, triste, enojado, etc.\}$$

Objetos3D (O): son entidades las cuales requieren de una estimulación externa para realizar alguna acción, por ejemplo: una puerta, sillas, cajas, etc.

$$O = \{o_1, o_2, \dots, o_n\}$$

Siendo los avatares los que provocan los cambios en el ambiente, estos constan de un esqueleto el cual es manipulado en tiempo real para realizar las animaciones solicitadas.

Un cambio en el ambiente es la repercusión de la ejecución de una solicitud del usuario. Dicha solicitud contiene tres elementos: el avatar que realizará el cambio, la acción y una marca de tiempo que nos indica en qué momento debe ser realizada la acción.

Una vez conocido el avatar se pasa a una cola de acciones del avatar, donde son ordenadas basadas en la marca de tiempo, dicho orden es tomado en cuenta por una ranura de tiempo dinámica (Fig. 3), con lo cual se indica cuales y cuantas acciones se ejecutarán.

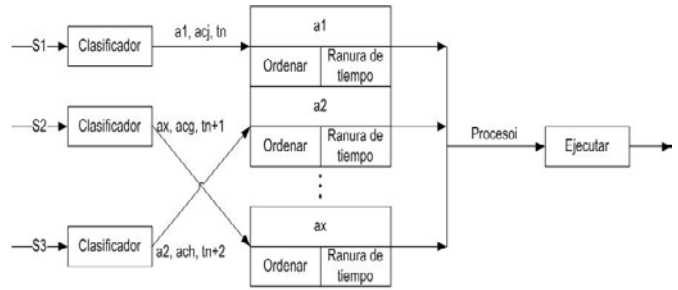


Fig. 3: Tratamiento de acciones.

Al ordenar las acciones de cada avatar se garantiza el obtener correctamente la secuencia de cambios en el ambiente, dando como resultado la evolución que se desea. La finalidad de manejar una ranura de tiempo es para administrar la carga de trabajo lo cual repercute en la visualización en tiempo real.

La ranura de tiempo está manejada de acuerdo a un incremento de tiempo (Fig. 4):

$$Ranura\ de\ Tiempo\ (RT) = marca\ de\ tiempo\ (mt) + incremento\ de\ tiempo\ (\Delta t)$$

Las acciones que estén dentro de la ranura de tiempo son agrupadas en un proceso. De esta forma el proceso contiene n acciones de diferentes avatares, la ejecución del proceso debe realizarse en tiempo real. Para lo cual se ha definido un tiempo máximo de 60 milisegundos, esto tomando como base el movimiento continuo definido en [21].

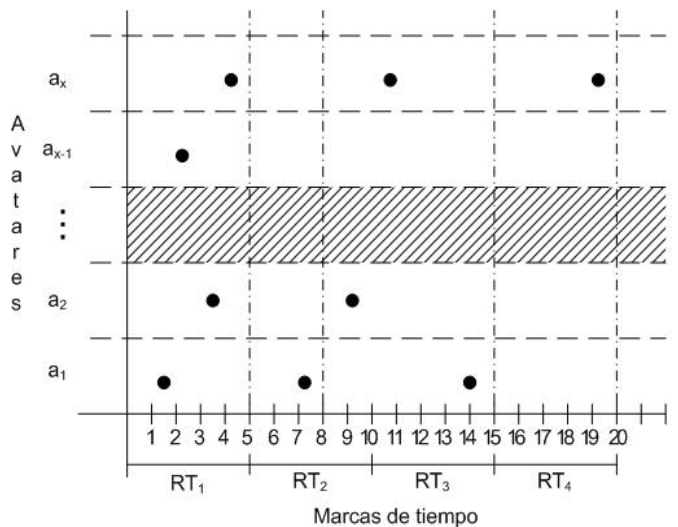


Fig. 4: Ranura de tiempo dinámica.

Una de las tareas de la ejecución de los procesos es el manejo de la concurrencia entre avatares o entre las partes de un avatar, para lo cual se generan bloqueos momentáneos del (los) avatar(es) o la(s) parte(s) que se requieren modificar.

Las solicitudes realizadas por un usuario deben ser notificadas a los demás integrantes del área de interés, para que estas sean tomadas en cuenta en tiempo y forma. Por lo cual se realizan dos tipos de mensajes: mensajes de actualización y mensajes de consistencia. El envío de los cuales se describe en la siguiente sección.

Distribución de mensajes

Para llevar a cabo la distribución de mensajes entre los nodos (usuarios) del ambiente, se deben considerar la organización de los nodos, el tipo de canal de comunicación, los protocolos involucrados, ya que estas características repercuten en el tiempo de entrega de paquetes, tiempo que se debe considerar para realizar una visualización en tiempo real.

Para lo cual se organizan los nodos en un grid heterogéneo (Fig. 5), donde podemos encontrar divisiones de acuerdo al área de interés de cada avatar.

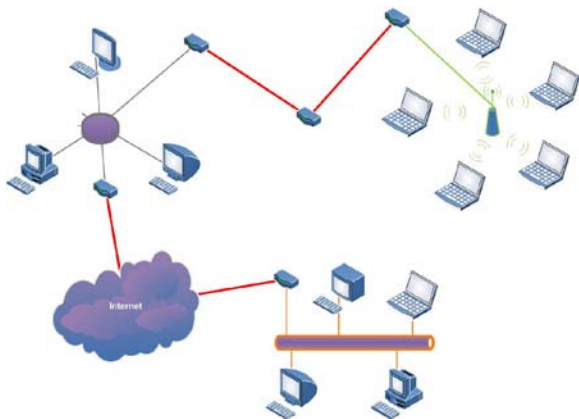


Fig. 5: Grid heterogéneo.

Para identificar a cual área de interés pertenece cada avatar se toman en cuenta dos aspectos:

Área de visión (Fig. 6): está dada por las características del avatar, el cual nos indica

el rango de visión para detectar los detalles que son visibles para cada avatar. De esta forma se evita el envío y/o el procesamiento de mensajes que no son perceptibles.



Fig. 6: Área de vision de un avatar.

Limitaciones dadas por el ambiente (Fig. 7): en un escenario es posible encontrar objetos que limitan el rango de visión de los avatares, por lo cual no es necesario entablar una comunicación entre ellos.

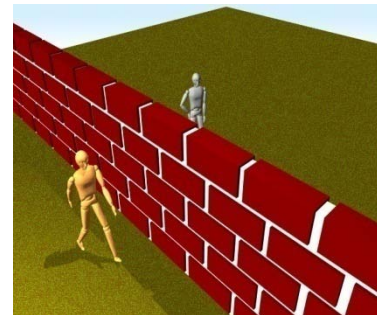


Fig. 7: Limitaciones del AV.

Con esto podemos identificar las áreas de interés involucradas en el ambiente (Fig. 8), las cuales pueden estar formadas por nodos que pertenecen a diferentes redes y que utilizan diferentes protocolos de comunicación.

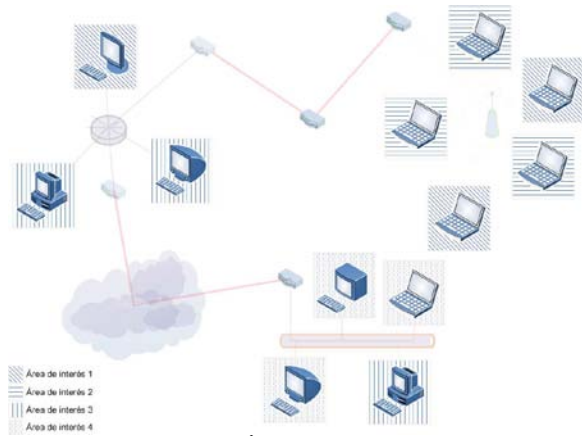


Fig. 8: Áreas de interés.

Una vez identificadas las áreas de interés es necesario asignar los roles a cada nodo, para cada área se elige un coordinador, para esto se toma en cuenta la distancia promedio entre los nodos y el ancho de banda. La primera característica que se analiza es la distancia promedio entre los nodos (Fig. 9), esto por medio del campo TTL de los mensajes, que indica el número de saltos en la red que realiza el mensaje. Esto permite dos características, el disminuir el retardo de entrega y la pérdida de mensajes.

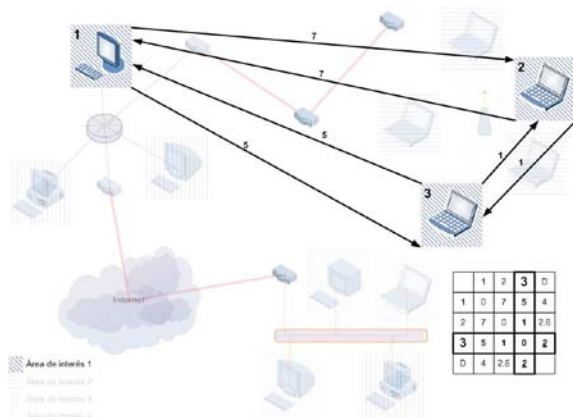


Fig. 9: Distancia promedio entre nodos.

En la figura anterior se muestran las distancias entre los nodos del área de interés 1, de los cuales se calcula la distancia promedio para cada nodo, siendo el nodo 3 el de menor distancia promedio del área, por lo cual se toma el nodo 3 como coordinador del área de interés 1.

La interacción entre los nodos se realiza entre usuarios y coordinado (Fig. 10), para llegar a un

acuerdo del estado actual del área de interés en cuestión, y entre coordinadores (Fig. 11) para validar el estado general del ambiente.

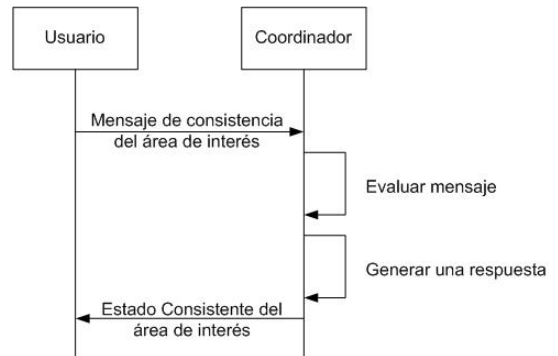


Fig. 10: Consistencia en un área de interés.

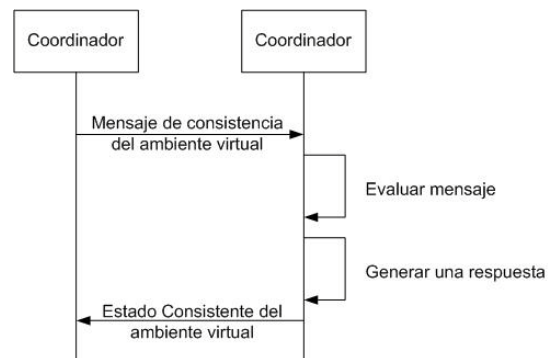


Fig. 11: Consistencia en el AV.

Es posible tener una sobre carga de trabajo en algún coordinador, esto dificultaría el trabajo necesario para la visualización en tiempo real. Para solventar este problema se elige otro coordinador del área. De esta forma puede existir más de un coordinador, los cuales se eligen de acuerdo a las distancias que tengan con los demás elementos, tratando de que cada uno tenga el mismo número de usuarios a su cargo. Cuando uno de los usuarios se desconecte o cambie de área de interés, se recalcula si el área de interés es soportada por un solo coordinador.

Otro de los aspectos que considera es cuando un usuario desea participar en la evolución del ambiente, para lo cual debe realizar una búsqueda de los ambientes virtuales disponibles, esto por medio de solicitar la agregación a una lista de multicast a la cual pertenecen los coordinadores de área. Una vez que los coordinadores contestan con

el(los) ambiente(s) al (los) que pertenece, el nuevo nodo selecciona a cual ambiente se conectará, y realizará una solicitud de agregación al ambiente.

Es posible que el nuevo usuario no conozca todos los elementos del ambiente, los cuales son solicitados al coordinador del área de interés al que pertenece.

IV. CONCLUSIONES

Con este trabajo se toman las ventajas de crecimiento de una “comunidad” de usuarios dada por los grid, aplicado a los ambientes virtuales en 3D. Al no depender de un servidor la evolución el ambiente puede continuar aun con las desconexiones de los nodos. El tráfico generado es administrado de tal forma que solo se envían los mensajes necesarios y por el canal más adecuado para ello.

Si cada usuario administra las acciones que debe realizar de acuerdo a la ranura de tiempo dinámica se garantiza una visualización en tiempo real. Las marcas de tiempo permiten ejecutar las acciones en la secuencia en que fueron creadas. Si no se realizaran las acciones acorde al tiempo que son solicitadas, es posible que no se mantenga una consistencia entre los usuarios del ambiente virtual. En consecuencia los coordinadores tendrían una sobre carga de trabajo para resolver el estado del ambiente virtual.

Las características analizadas nos permite la visualización de un ambiente virtual general, ya que no se basa en comportamientos prediseñados. Cuando son utilizadas animaciones prediseñadas la visualización se realiza de situaciones concretas. Por ejemplo: es posible visualizar tanto la evolución de un ambiente que representa a un parque como el que representa una oficina. Esto es porque las animaciones que se pueden realizar son micro-animaciones. Este tipo de animaciones permite manipular en tiempo real cada parte del avatar, dando más realismo a las animaciones.

V. AGRADECIMIENTOS

Esta investigación es parcialmente apoyada por CoECyTJal Proyecto No. 2008-05-97094, mientras que la autora Alma Martínez es apoyada por el CONACYT con la beca 47605.

VI. REFERENCIAS

- [1] *Emergency Evacuation Simulator* en <http://www.emergencysimulators.com/>
- [2] Dr. Venkat V.S.S. Sastry, Mr. Joseph Steel, Lt. Edward A. Trott. *A Virtual Environment for Naval Flight Deck Operations Training*. NATO Workshop “What is essential for Virtual Reality to meet military performance goals?”, Hague, The Netherlands, 2000.
- [3] Kevin L. Dean and Xylar S. Asay-davis and Evan M. Finn and Tim Foley and Jeremy A. Friesner and Yo Imai and Bret J. Naylor and Sarah R. Wustner and Scott S. Fisher and Kent R. Wilson. *Virtual Explorer: Interactive Virtual Environment for Education*. Massachusetts Institute of Technology. Vol. 9, No. 6, pp. 505–523, December 2000.
- [4] Funkhouser T., “*Network Topologies for Scalable Multi-User Virtual Environments*”, IEEE Virtual Reality Annual International Symposium, pp. 222–228, 1996.
- [5] Macedonia M. R, Zyda M. J., “*A Taxonomy for Networked Virtual Environments*”, IEEE Multimedia Magazine, pp. 48–56, 1997.
- [6] Singhal S., Zyda M. J.. “*Networked Virtual Environments*”, pp. 143, ACM Press, NY, NY, 1999.
- [7] Pullen M., “*Reliable Multicast Network Transport for Distributed Virtual Simulation*”, DIS-RT '99: IEEE Workshop on Distributed Interactive Simulations and Real-Time Applications, pp. 59–66, 1999.
- [8] Henderson, Tristan. “*Latency and User Behaviour on a Multiplayer Game Server*”, NGC '01: Proceedings of the Third International COST264 Workshop on

- Networked Group Communication. Springer-Verlag, 2001.
- [9] Majewski, Chris and Griwodz, Carsten and Halvorsen. "Translating Latency Requirements into Resource Requirements for Game Traffic", International Network Conference (INC 2006), pp 113-120, Steven Furnell and Paul Dowland, University of Plymouth, 2006.
- [10] Henderson, Tristan, Bhatti Saleem. "Modelling user behaviour in networked games", MULTIMEDIA '01: Proceedings of the ninth ACM international conference on Multimedia, pp 212-220, ACM, New York, NY, USA, 2001.
- [11] Ahmed Ait Ali, Fabien Michaut, Francis Lepage. *End-to-End Available Bandwidth Measurement Tools: A Comparative Evaluation of Performances*, CoRR, abs/0706.4004, 2007.
- [12] Ahmed Ait Ali and Francis Lepage. "IGMPS, a New Tool for Estimating End-to-End Available Bandwidth in IP Network Paths", Networking and Services, International conference on, pp 115, IEEE Computer Society, 2007.
- [13] Färber, J. "Traffic Modelling for Fast Action Network Games", Multimedia Tools Appl. 23, 1, 31-46. Kluwer Academic Publishers, Hingham, MA, USA, 2004.
- [14] Ajaltouni, Elie El, Boukerche, Azzedine and Zhang, Ming. "An Efficient Dynamic Load Balancing Scheme for Distributed Simulations on a Grid Infrastructure", DS-RT '08: Proceedings of the 2008 12th IEEE/ACM International Symposium on Distributed Simulation and Real-Time Applications, IEEE Computer Society, Washington, DC, USA, 2008.
- [15] Rueda Silvia, Morillo Pedro, Orduna Juan M. "A Peer-To-Peer Platform for Simulating Distributed Virtual Environments", ICPADS '07: Proceedings of the 13th International Conference on Parallel and Distributed Systems - Volume 2 (ICPADS'07), pages 1-8, Washington, DC, USA, 2007.
- [16] Silvia Rueda, Pedro Morillo, J. M. Orduna, Jo'e Duato. "On the Characterization of Peer-To-Peer Distributed Virtual Environments", William R. Sherman, Ming Lin, and Anthony Steed, editors, VR, pages 107-114. IEEE Computer Society, 2007.
- [17] Taylor Ian J. and Andrew Harrison. *From P2P and Grids to Services on the Web: Evolving Distributed Communities*. Springer, 2009.
- [18] Hsiao, Tsun-Yu and Yuan, Shyan-Ming. "Middleware for Massively Multiplayer Online Games", IEEE Internet Computing, pp 47-54, IEEE Educational Activities Department, Piscataway, NJ, US, 2005.
- [19] Grenville Armitage, Mark Claypool, Philip Branch, *Networking and Online Games: Understanding and Engineering Multiplayer Internet Games*. John Wiley & Sons, 2006.
- [20] Scott Douglas, Egemen Tanin, Aaron Harwood, Shanika Karunasekera. "Enabling Massively Multi-Player Online Gaming Applications on a {P2P} Architecture", Proceedings of the IEEE International Conference on Information and Automation, pp 7-12, Colombo, Sri Lanka, 2005.
- [21] Maiche Marini Alejandro. *Tiempo de reacción al inicio del movimiento: Un Estudio sobre la Percepción de Velocidad*. PhD perception, communication and time, Department of Educational Psychology, Universidad Autónoma de Barcelona, Barcelona, 2002.

VII. AUTORES

M. en C. Alma Martínez es ingeniera en Computación por la Universidad de Guadalajara, maestra en ciencias por el CINVESTAV-Guadalajara. Actualmente es estudiante de doctorado y realiza investigaciones sobre visualización y evolución cooperativa de ambientes virtuales distribuidos.

Dr. Félix Ramos es licenciado en electrónica de telecomunicaciones por la Universidad Autónoma Metropolitana y doctor en ciencias computacionales por la Universidad de Tecnología de Compiègne, Francia. Es miembro del Sistema Nacional de Investigadores. Actualmente, es profesor investigador en el Centro de Investigación y de Estudios Avanzados del IPN en Guadalajara.

Dr. Mario Siller es Ingeniero en Sistemas Electrónicos, Tecnológico de Monterrey y doctor en Ingeniería de Sistemas Electrónicos por la Universidad de Essex. Es miembro del Sistema Nacional de Investigadores. Actualmente, realiza investigaciones en el área de redes de comunicaciones en el Centro de Investigación y de Estudios Avanzados del IPN en Guadalajara.

APLICACIÓN DE LA TEORÍA DE GRAFOS PARA ENCONTRAR LAS DISTANCIAS MÁS CORTAS ENTRE LAS CIUDADES DE LA REPÚBLICA MEXICANA MEDIANTE LA IMPLEMENTACIÓN DEL ALGORITMO DE FLOYD

García García Oscar Galileo¹, Cacho Ruiz Mireya²

División de Desarrollo Biotecnológico, Departamento de Ciencias Básicas
Centro Universitario de la Ciénega

Av. Universidad 1115, Col. Linda Vista, Ocotlán Jalisco, México

Tel. 01(392) 92 59400 ext. ¹8312 y ²8332, e-mail: 1galileo@cuci.udg.mx , 2mireyacacho@cuci.udg.mx

RESUMEN

En el presente trabajo, se implementó un modelo matemático en forma de grafo, para representar un conjunto de n ciudades, *las cuales están simbolizadas en forma de vértice*, y éstos a su vez tienen aristas rotuladas con las distancias en kilómetros entre ellas. Dicho lo anterior, mediante la aplicación del algoritmo de Floyd Warshall se demuestra que es posible encontrar una solución óptima para determinar los caminos más cortos entre todos los pares de ciudades.

La idea principal, se plasmó en el desarrollo de una página web dinámica, que permite al usuario obtener la distancia más corta de un punto a otro de la República Mexicana; además muestra la ruta descriptiva en forma semántica y gráfica por medio del servicio de Google Maps y además, los usuarios registrados pueden colaborar agregando rutas adicionales entre ciudades y actualizando información de interés.

Palabras clave: modelo matemático, algoritmo de Floyd Warshall, página web dinámica, internet.

I. INTRODUCCIÓN

Para resolver el problema, un punto fundamental es entender con precisión el planteamiento y las bases de éste; una vez entendido el problema es más sencillo plantear un modelo matemático que lo represente. Alfred V. Aho [1] plantea que “Si es posible expresar ciertos aspectos de un problema

con un modelo formal, por lo general resulta beneficioso hacerlo, pues a su vez que el problema se formaliza, pueden buscarse soluciones en función de un modelo preciso y determinar si ya existe un programa que resuelva tal problema; aún cuando no sea tal el caso, será posible averiguar lo que se sabe acerca del modelo y usar sus propiedades como ayuda para elaborar una buena solución”

A lo largo del tiempo, se ha demostrado que es posible recurrir casi a cualquier rama de las matemáticas para obtener un modelo matemático, que ayude a representar de manera abstracta las características de un problema, y que a su vez, sirva como base para encontrar una solución computacional a éste.

Existen situaciones del mundo real que pueden ser representadas por medio de un Grafo, por ejemplo, en la Universidad de las Américas Puebla, desarrollaron un proyecto para determinar la ubicación y recolección de contenedores de basura en la colonia San Manuel de la Ciudad de Puebla [2], mediante la utilización de grafos y del algoritmo de Floyd Warshall, además de otras herramientas. Distintas representaciones en forma de grafo son útiles, como se puede apreciar en la **Figura 1**, para simbolizar la adyacencia de routers en redes de computadoras; determinar la estructura de servidores de correo en Internet; simbolizar la estructura organizacional de una

empresa; efectuar la asignación de frecuencias en radiocomunicaciones, entre otras.

Para el presente trabajo, se optó por aplicar un modelo matemático en forma de grafo, para representar un conjunto de n ciudades, *las cuales están simbolizadas en forma de vértice*; éstos a su vez, poseen aristas rotuladas con las distancias en kilómetros entre ellas, como se ilustra en la **Figura 2**. Dicho lo anterior, mediante la aplicación del algoritmo de Floyd Warshall, demostrar que es posible encontrar una solución óptima para determinar los caminos más cortos entre todos los pares de ciudades.



Figura 2 Mapa de la República Mexicana, con aristas entre capitales.

Nuestro objetivo es proporcionar información al instante a usuarios de todo el mundo, sobre rutas geográficas reales y al mismo tiempo ofrecerles la oportunidad de retroalimentar el sistema, por medio de una verdadera RIA que ofrezca todas las características de la Web 2.0 e información semántica accesible.

II. MÉTODO

Cairó y Guardatti [4] proponen la siguiente definición para grafo “Un Grafo es una estructura de datos no lineal donde cada nodo puede tener uno o más nodos adyacentes”. Un Grafo está formado por un conjunto de nodos o vértices conectados a través de arcos o aristas, los cuales pueden estar rotulados con información.

Dado un Grafo G formado por los vértices V , se puede encontrar solución a problemas de rutas o recorridos mediante la utilización de Teoría de Grafos, específicamente aplicando el algoritmo de Floyd, que dada una matriz de adyacencia del grafo G , calcula una matriz que contiene los caminos más cortos entre todos los pares de vértices de G .

Este algoritmo se puede resolver por programación dinámica, ya que es aplicable el principio de óptimo, que puede enunciarse para este problema de la siguiente forma: si en el camino que va de V_i a V_j , V_k es un vértice intermedio, los caminos de V_j a V_k y de V_k a V_j han de ser a su vez caminos

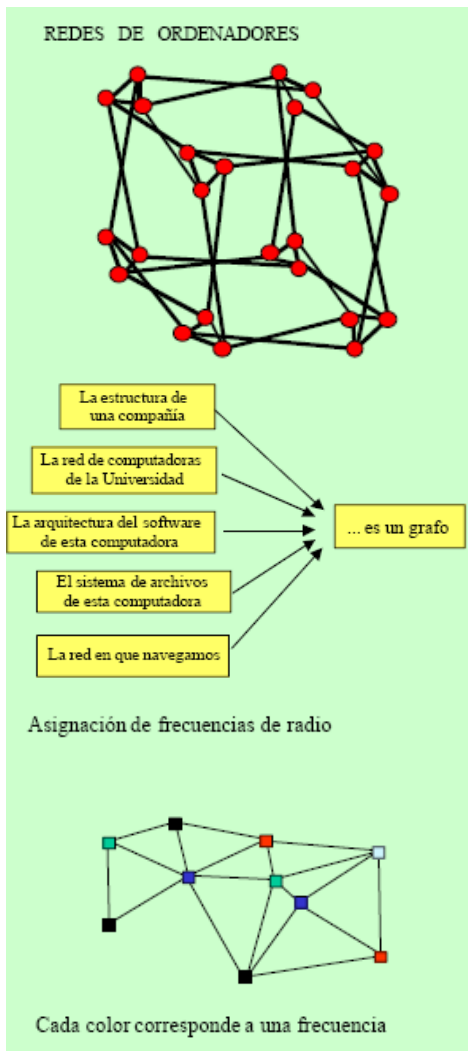


Figura 1 Representación en forma de grafo para diversas situaciones del mundo real. [3]

mínimos. Puede plantearse la relación que resuelve el problema como: $D_k(i,j)=\text{Min}\{D_{k-1}(i,k),D_{k-1}(k,j)\}$. Tal ecuación queda resuelta mediante el algoritmo de Floyd que, siguiendo el esquema de programación dinámica, utiliza una matriz para evitar la repetición de los cálculos. Con ello consigue que su complejidad temporal sea de orden $O(n^3)$ debido al triple bucle anidado en cuyo interior hay tan sólo operaciones constantes.

A continuación, se muestra el algoritmo base que se eligió para la resolución computacional del problema.

```

procedure Floyd(var A: array[1..n, 1..n] of real;C:
array[1..n, 1..n] of real);
{Floyd calcula la matriz A de caminos más cortos dada
la matriz de costos de arcos C}
  var
    i, j, k: integer;
  begin
    for i:=1 to n do
      for j:=1 to n do
        A[i, j]:=C[i, j];
    for i:=1 to n do
      A[i, j]:=0;
    for k:=1 to n do
      for i:=1 to n do
        for j:=1 to n do
          if (A[i, k]+A[k, j]<A[i,j]) then
            A[i, j]:= A[i, k]+A[k, j]
  end; {Floyd}
  
```

Para el caso que nos ocupa, se optó por trabajar con una base de datos, como se puede apreciar en la **Figura 3**, la cual está conformada básicamente por dos tablas, la primer tabla denominada **ciudad**, que dentro de la teoría de grafos representa las aristas o nodos, la cual será utilizada como entidad fuerte que almacenará las ciudades o puntos de interés; la segunda tabla, **distancia**, representa la relación entre dos ciudades o puntos, los vértices, donde se almacenarán los diferentes tipos de etiquetas o costos a procesar, como son distancia, tiempo y dinero, incluso información descriptiva sobre la ruta.

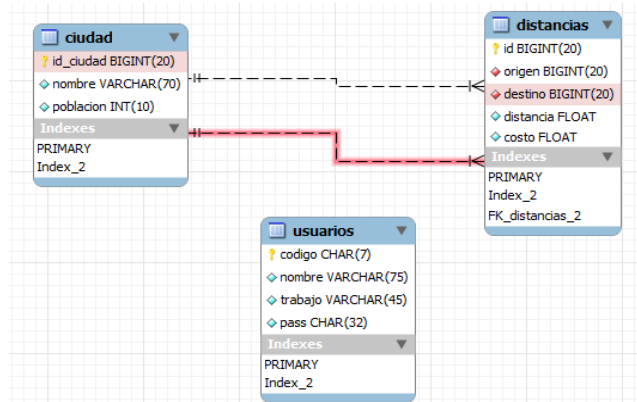


Figura 3 Modelo relacional de la base de datos.

Se optó por agregar una tabla adicional, *usuarios*, la cual atenderá las demandas de almacenamiento del módulo de usuarios registrados, brindando un sistema comunitario que podrá ser mejorado, quedando abiertos a la posibilidad de agregar rutas nuevas.

Se decidió utilizar el motor de almacenamiento InnoDB para MySQL, porque provee el soporte para preservar la coherencia y seguridad en la base de datos, restringiendo la eliminación de registros padre mientras estos tienen referencias que aún apuntan hacia él, así como la actualización en cascada en caso de cambios en los identificadores de ciudades.

Además se optó por utilizar la API pública que provee el servicio Google Maps, para brindar una interfaz gráfica y amigable al momento de buscar ciudades; el usuario podrá seleccionar directamente sobre el mapa la ciudad origen y destino, incluso almacenaremos las coordenadas geográficas de cada punto para que la ruta también quede expresada con una línea y mostrará las descripciones entre puntos intermedios.

Se eligió como lenguaje de programación el PHP, debido a que está basado en herramientas con licencia libre, soporta programación orientada a objetos, es multiplataforma y funciona del lado del servidor, es decir, puede realizar accesos a bases de datos, conexiones en red, y otras tareas para crear la página final que verá el cliente.

El algoritmo de Floyd Warshall utilizado en el presente proyecto, fue programado por Janne Mikkonen [5]; adicionalmente, se desarrolló una RIA con el patrón de diseño Modelo, Vista Controlador, utilizando la técnica de programación AJAX que caracteriza una aplicación Web 2.0 y la librería JQuery para estandarizar el Javascript entre navegadores.

INTEGRACIÓN DEL SISTEMA

El sistema divide sus funciones principales en dos procesos, **Ingreso** y **Consultas**, cuyo diagrama se puede apreciar en la **Figura 4**.

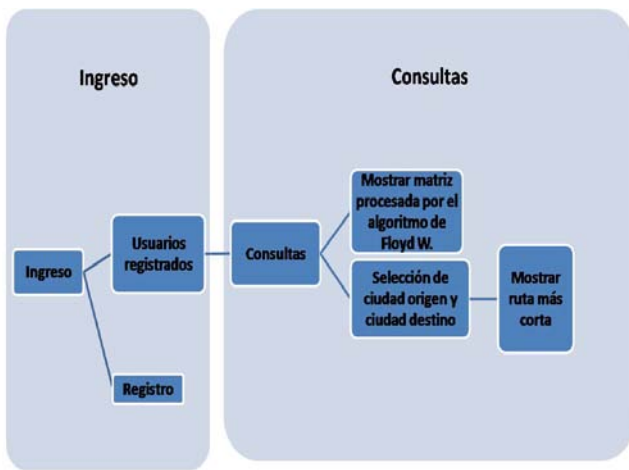


Figura 4 Diagrama de procesos que integran el sistema.

Al seleccionar **Ingreso** aparece el formulario que se aprecia en la **Figura 5**.

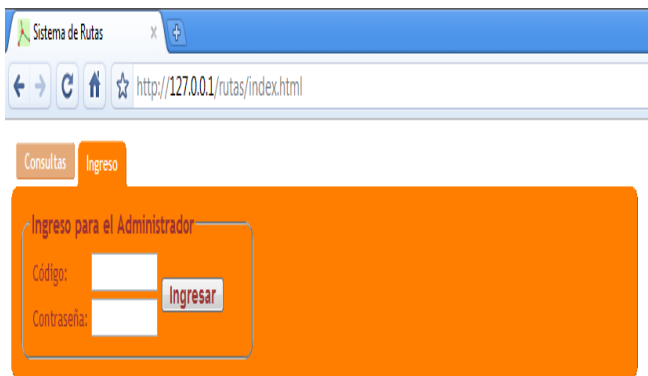


Figura 5 Formulario para ingresar al sistema.

Dicho formulario, permite el ingreso sólo a usuarios registrados, a través de un código de usuario y contraseña. La ventaja de contar con registro implica que los usuarios podrán ingresar información nueva al sistema.

En la opción **Consultas**, se muestra la matriz procesada por el algoritmo de Floyd Warshall (ver **Anexo 1**); mientras que en la parte inferior de la matriz se incluyó un menú dinámico, donde es posible seleccionar la ciudad de origen y el destino; como resultado se muestra las *ciudades que forman la ruta del origen al destino, acompañado de la distancia más corta en kilómetros*.

III. CONCLUSIONES

La aplicación de la teoría de grafos en la resolución computacional de un problema, simplifica enormemente el trabajo de encontrar soluciones óptimas a determinados situaciones del mundo real; aunado a esto, el uso de algoritmos eficientes, en nuestro caso el de Floyd Warshall, que auxiliien en la obtención del código fuente necesario para el desarrollo de la aplicación.

Hemos observado que el proyecto tiene un potencial enorme, debido a la incorporación de tecnologías estándares que garantizan la compatibilidad entre sistemas, es accesible y público. Actualmente ya se ofrecen en internet diferentes sistemas que encuentran la ruta de un punto a otro, pero hasta el momento ninguno provee la flexibilidad que sugiere este sistema, debido a que el cálculo de rutas que provee Google Maps aún no proporciona soporte completo entre ciudades de nuestro país y además no se actualiza frecuentemente; sin duda el presente proyecto cubre esta deficiencia en los servicios existentes y brinda información sobre los puntos de interés más consultados.

Algunas tablas auxiliares podrían agregarse en un futuro para almacenar información adicional que permitirá la administración del sistema.

Además se pretende almacenar la información que los usuarios consultan a menudo, como puede ser gasto de combustible, costo total de cacetos, costo viajando por autobús o por avión, etc. Para lograr modificaciones futuras, se trabajará mediante la utilización de información almacenada en tabla, ya que se pretende optimizar las veces que se ejecuta el algoritmo Floyd Warshall.

IV. REFERENCIAS

- [1] V. Alfred Aho, H. E. Hopcroft, U. D. Jeffrey, *Estructuras de datos y algoritmos*, Addison-Wesley Iberoamericana, 1998.
- [2] Resumen de la tesis digital: “Ubicación y recolección de contenedores de basura en la colonia San Manuel en la ciudad de Puebla”, presentada por Baiz Bonnin, Alejandro D., Romero Heredia, Yaire; Fecha última consulta: 07 septiembre 2009, disponible en el sitio:
http://catarina.udlap.mx/u_dl_a/tales/documentos/lii/baiz_b_ad/resumen.pdf
- [3] Imagen tomada de
<http://www.dma.fi.upm.es/grafos/Intro03.PDF>
- [4] O. Cairó, S. Guardatti, *Estructuras de datos*, México, McGraw Hill, 2006.
- [5] *Algoritmo de Floyd Warshall*, Fecha última consulta: 08 septiembre 2009, disponible en el sitio
<http://julmis.julmajanne.com/index.php/FloydWarshall>

V. AUTORES

M. en C. A. Mireya Cacho Ruiz en el año 2006 obtuvo el título de Maestría en Computación Aplicada con especialidad en bases de datos por parte de la Universidad de Guadalajara, es Profesora de Tiempo Completo del CUCiénega, con adscripción al Departamento de Ciencias Básicas. Actualmente está a cargo de la coordinación de la Maestría en Computación Aplicada. Tiene una antigüedad docente de 10 años, impartiendo materias en las carreras de Licenciatura en Informática e Ingeniería en Computación. Entre las publicaciones más recientes en su haber, se encuentra el artículo “Análisis y diseño de un sistema de administración para restaurantes” y las memorias en extenso “Impacto de las Tecnologías de la Información y Comunicación en el Personal Docente: Caso Centro Universitario de la Ciénega”.

Ing. Oscar Galileo García García, estudiante de la carrera de Ingeniería en Computación por la Universidad de Guadalajara, especializado en el desarrollo de RIA, fue becario de estímulos económicos para estudiantes sobresalientes durante tres semestres. Actualmente es auxiliar administrativo y es creador del sistema de reservación de aulas (<http://radio.cuci.udg.mx/horarioscta/inicio.php>), Sistema de Evaluación Docente (<http://148.202.149.3/sed>) y Centro de Información Estadística para la Planeación (<http://148.202.149.3/matrix>).

ANEXOS

Sistema de Rutas

http://127.0.0.1/rutas/index.html

Morelia	487	1596	1787	1458	809	camino	949	498	798	876	279	351	2136	3995	1760	1871	413	937	0	797	501
Oaxaca	992	1216	1407	1963	738	No hay camino	1454	1003	1587	1381	1068	856	2641	4784	1380	2122	664	1442	797	0	576
Pachuca	592	1375	1566	1563	588	No hay camino	1054	603	1187	981	668	456	2241	4384	1539	1546	88	1042	399	576	0
Puebla	629	1162	1353	1600	375	No hay camino	1091	640	1224	1018	705	493	2278	4421	1326	1759	301	1079	434	363	213
Queretaro	289	1502	1693	1260	715	No hay camino	751	300	884	678	365	153	1938	4081	1666	1673	215	739	198	703	303
Saltillo	502	1964	2155	903	1177	No hay camino	964	954	1042	722	692	684	1760	4408	2128	2285	869	85	795	1165	957
San Luis Potosí	168	1704	1895	1139	917	No hay camino	630	502	938	587	419	350	1817	4135	1868	1875	417	537	400	905	505
Tepic	478	2094	2285	1449	1080	No hay camino	940	892	504	897	227	660	2127	3489	2258	1823	807	1061	506	1295	895
Tlaxcala	659	1192	1383	1630	405	No hay camino	1121	670	1254	1048	735	523	2308	4451	1356	1789	331	1109	464	393	243
Toluca	519	2126	2317	1490	1426	No hay camino	981	1099	1108	938	589	400	2168	4305	2290	2290	832	1106	812	1496	920
Tuxtla Gutiérrez	1534	674	865	2505	1280	No hay camino	1996	1545	2129	1923	1610	1398	3183	5326	838	2664	1206	1984	1339	542	1118
Villahermosa	1410	381	572	2381	1156	No hay camino	1872	1421	2005	1799	1486	1274	3059	5202	545	2371	913	1860	1215	835	994
Xalapa	1456	864	1055	2427	1202	No hay camino	1918	1467	2051	1845	1532	1320	3105	5248	1028	2586	1128	1906	1261	464	1040
Zacatecas	129	1890	2081	1100	1036	No hay camino	591	688	829	290	319	311	1778	4035	2054	2061	603	716	422	1091	691

Origen: Destino:

**El camino más corto de Guadalajara a Aguascalientes es:
Guadalajara, Aguascalientes, 251 km**

Anexo 1 Demostración de lo que realiza el proceso *Consultas*.

MPISTE: “UN ENTORNO PARA CONTAR HISTORIAS BASADAS EN UBICACIÓN DE FORMA PERSONALIZADA EN DISPOSITIVOS MÓVILES”

Delgado Mata Carlos¹, Pooley Robert², Aylett Ruth², Robertson Judy².

¹Escuela de Ingeniería
Universidad Panamericana Campus Bonaterra
José Ma. Escrivá de Balaguer 101, Ags, Ags. 20290, México
Tel. 4499106216, correo electrónico: cdelgado@up.edu.mx

²Mathematics and Computer Science Department
Heriot Watt University
Riccarton Campus, Edinburgh, U.K.

RESUMEN

En este artículo se describe un entorno basado en ubicación física del usuario para contar historias de forma personalizada en dispositivos móviles. En primera instancia, se presenta la tecnología apropiada para ubicar físicamente al usuario. Posteriormente se pone el proyecto, llamado MPISTE, en contexto con proyectos relacionados. Se continúa con la descripción del proyecto y su diseño como parte de un proyecto para desarrollar un guía virtual afectivo. El trabajo descrito en este artículo consta de dos partes. La primera es el cliente, la cual fue desarrollado para teléfonos celulares inteligentes (*Smartphones*) y donde se utiliza conexión por medio de bluetooth a dispositivos de localización como RFID y GPS. La segunda es el servidor (*webserver*) el cual recibe peticiones de la primera y las atiende a través de servlets y una base de datos relacional. El protocolo de comunicación utilizado para comunicar el celular con el servidor es EDGE/3GSM.

Palabras clave: Dispositivos Móviles, RFID, Narrativa Interactiva.

I. INTRODUCCIÓN

Contar historias es inherente al ser humano. Con la ubicuidad de la disponibilidad de computadoras en hogares y en oficinas, se ha incrementado el interés en el área de narrativa interactiva en diversos

campos. Por mencionar algunos, están los juegos de video (como principal exponente de narrativa interactiva). El ejemplo predominante es el excelente juego publicado por Atari, *Fahrenheit*, el cual logra una experiencia única al ser una película interactiva, en la que las decisiones del jugador modifican sustancialmente la narrativa. Por otro lado se han desarrollado guías virtuales, los que se utilizan para contar historias basadas en contexto (ubicación física del usuario), las cuales se han aplicado con éxito en aplicaciones en interiores como museos, así como en exteriores, en particular en locaciones urbanas.

Métodos de ubicación.

Existen diversos métodos para obtener la ubicación del usuario, los cuales sirven para poner en contexto a la aplicación. A continuación se describen algunos métodos de ubicación que fueron investigados.

RFID. El identificador del radio frecuencia (por sus siglas en inglés) es un método de comunicación inalámbrica para identificar objetos y lugares que ha sido utilizado en diversas aplicaciones que permean nuestra vida diaria, como etiquetar libros, películas y CDs para evitar que se extraigan de una biblioteca o de una tienda de manera ilegal. Otra aplicación que ha adquirido notoriedad es el uso de la etiqueta (IAVE) en automóviles. Esta tarjeta

simplifica el cobro de casetas, al realizarse de forma automática cuando un automóvil pasa por casetas con lectores RFID en las autopistas mexicanas.

GPS. Es el sistema de posicionamiento global (por sus siglas en inglés) es utilizado de forma común para la ubicación de automóviles, tanto de pasajeros como de carga. Existen aplicaciones basadas en ésta tecnología, como asistencia en la navegación de automovilistas al combinar la posición del auto (a través de GPS) con la información del mapa correspondiente. El sistema GPS funciona al triangular la ubicación de al menos 3 satélites. Con el sistema GPS se puede obtener la ubicación del dispositivo con una exactitud en el rango de 5 metros. El sistema GPS es ideal para aplicaciones basadas en exteriores pero tienen severas limitaciones en aplicaciones ubicadas en interiores.

Bluetooth. Bluetooth es un sistema de comunicación inalámbrica que ha sido utilizada frecuentemente para crear redes personales. La mayoría de los dispositivos en la actualidad cuentan con bluetooth. Por ejemplo, los celulares, las laptops, las PDAS y los controles inalámbricos de consolas de videojuegos (Nintendo Wii) utilizan este sistema. Bluetooth puede utilizarse para aplicaciones basadas en ubicación al utilizarlos como marcadores inteligentes.

WLAN. Es una red local basada en comunicación inalámbrica (WiFi). Para desarrollar la red se colocan diversos puntos de acceso. Se puede obtener la ubicación física del usuario a través de la triangulación de información de éstos. Existe la restricción de que deben de estar colocados previamente estos puntos de acceso. Por lo anterior, en algunas aplicaciones en exteriores no es factible utilizar esta tecnología.

Infrarrojo. Está basado en radiación electromagnética. La amplitud de onda infrarroja es mayor que la de la luz visible. Esta tecnología es madura y utilizada comúnmente en dispositivos, como controles remoto para televisores y en

algunos PDAs y calculadoras. El problema es que la comunicación es uno a uno y no es recomendable para aplicación que trabajen con varios dispositivos móviles (clientes) a la vez.

Combinación de métodos. Se pueden combinar métodos para aprovechar las fortalezas de cada uno de éstos. Por ejemplo, se puede utilizar GPS para obtener la ubicación aproximada en un rango de 5 metros, y cuando se esté en cierto rango se puede utilizar tecnología RFID para ubicarse en rango de centímetros.

Historias interactivas.

Contar historias es cotidiano al ser humano desde hace miles de años. Ejemplo de lo anterior son las pinturas rupestres (arte prehistórico) descubiertas en diversas partes del mundo, y donde los más relevantes se encuentran en España y Francia.

Se han realizado estudios de la estructura de la historia entre la que destaca el trabajo seminal de Vladimir Propp [1], donde realizó un estudio sobre la estructura de cuentos populares rusos. Al separar una historia en fragmentos es posible contar historias en forma interactiva. Uno de los pioneros en generar historias de manera interactiva es Chris Crawford, [2] quien ha definido interactividad como *“un proceso cíclico entre dos o más agentes activos en donde cada agente de manera alternada escucha, piensa y habla”*. Generar historias interactivas, por medio de herramientas de desarrollo de videojuegos, también se ha utilizado para apoyar en el mejorar la literacidad en adolescentes [3].

Guías Virtuales. A partir de finales de la década de los noventa ha habido interés por desarrollar guías virtuales en dispositivos móviles. Uno de los primeros trabajos (Cyberguide) [4] fue desarrollado en el tecnológico de Georgia. Para este proyecto se utilizó el dispositivo móvil Apple Newton y se utilizó GPS para ubicación en exteriores e infrarrojos para interiores.

Se han utilizado guías virtuales en entornos virtuales. En [5] y [6] se muestra un trabajo interesante donde el guía cuenta historias desde su punto de vista. Por ejemplo, si en un entorno virtual de un lugar histórico, donde existieron historias de conquistadores y conquistados, el guía narrará la historia desde su punto de vista favoreciendo a los conquistadores o a los conquistados. En [7] se muestra una variación a este trabajo donde el guía cuenta historias dependiendo de su estado emocional y donde también se incluyó un componente para comunicar al usuario la ubicación física del usuario, por medio de GPS.

Aplicaciones basadas en ubicación.

Se han realizado aplicaciones donde la ubicación es importante para ponerlas en contexto. Se han utilizado algunas de las tecnologías descritas anteriormente como GPS, Infrarrojo, RFID para ubicar la posición física del usuario.

Este tipo de aplicaciones han tomado relevancia y se han desarrollado cubriendo una amplia gama de éstas, que van desde los juegos de video hasta aplicaciones críticas. Un ejemplo de los primeros, es el juego donde los usuarios se convierten en los personajes del videojuego Pacman [8]. Esto es, unos jugadores son los fantasmas y otro es Pacman. El juego se desarrolla en un entorno urbano. El método de localización que se utilizó fue RFID. Un proyecto similar es el proyecto colaborativo donde se simulaba una planicie africana [9] donde los usuarios tenían roles de depredador y presa y donde se motivaba al trabajo en equipo. El método de localización fue GPS.

La tecnología RFID también se ha utilizado en aplicaciones críticas, como lo es evacuación de un edificio [10] y para tener mayor seguridad en los procedimientos quirúrgicos en un hospital [11] al proveer de marcadores al cirujano, al paciente y a los medicamentos que se administrarán al éste último.

Un proyecto muy interesante es el juego “*Can You See Me*” [12], en donde conviven jugadores en línea que tratan de escapar de 4 jugadores profesionales, quienes toman el rol de perseguidores. Estos últimos se encuentran físicamente en una ciudad. La tecnología utilizada en el juego es GPS y WiFi.

Un proyecto donde se combinan adecuadamente la narrativa interactiva con la aplicación basada en ubicación es el proyecto descrito en [13], donde se tiene que encontrar al perpetrador de un crimen en un sector de una ciudad.

II. MÉTODO

El trabajo descrito en el presente artículo forma parte de un proyecto ambicioso para desarrollar un guía virtual afectivo. El proyecto completo se desarrolla en la Universidad de Heriot Watt en Edimburgo, Escocia. El subproyecto aquí descrito es un trabajo colaborativo entre la Universidad Panamericana y la Universidad de Heriot Watt.

El proyecto completo consta de los siguientes subproyectos.

1. Un guía virtual con actitud [7]. En donde un guía virtual cuenta historias dependiendo de su estado emocional y del perfil del usuario. El guía virtual está diseñado para exteriores y se ubica al usuario con GPS.
2. Un “framework” para la ubicación del usuario independiente del dispositivo de ubicación (GPS, RFID, Bluetooth) [14].
3. Personalización automatizada de los perfiles del usuario y del guía virtual.
4. Cadies: Aprendizaje al crear y compartir historias basadas en la ubicación del usuario.
5. MPISTE. El subproyecto, objeto del presente artículo, es un entorno para contar interactivamente historias personalizadas, basadas en ubicación.

El subproyecto MPISTE, en contexto del proyecto completo, cuenta con tres tipos de usuarios (tal como se muestra en la figura 1).

1. El usuario de la herramienta para generar contenido. Este contenido es utilizado por el usuario del guía virtual.
2. El usuario del guía virtual, quien es el que recibe las historias interactivas basadas en ubicación en su dispositivo móvil.
3. El moderador, quién modera el contenido que sube al servidor el usuarios de la herramienta de contenido.

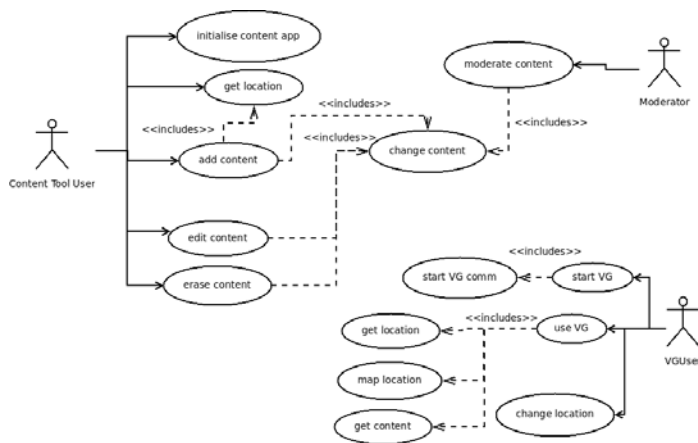


Fig. 1. Diagrama de caso de uso del subproyecto MPISTE en contexto.

Componentes del proyecto.

En la figura 2 se muestran los componentes del proyecto MPISTE. En la parte superior se muestra un dispositivo móvil el cual se conecta por medio de bluetooth a dispositivos de ubicación basados en tecnología RFID y GPS. En el celular (dispositivo móvil) se utiliza el “framework” abstracto para ubicación descrito en [14]. Por medio del protocolo de comunicación para celulares, EDGE/3GSM, el celular se conecta al servidor de contenidos, el cual cuenta con servidor web y con servidor de base de datos (MySQL).

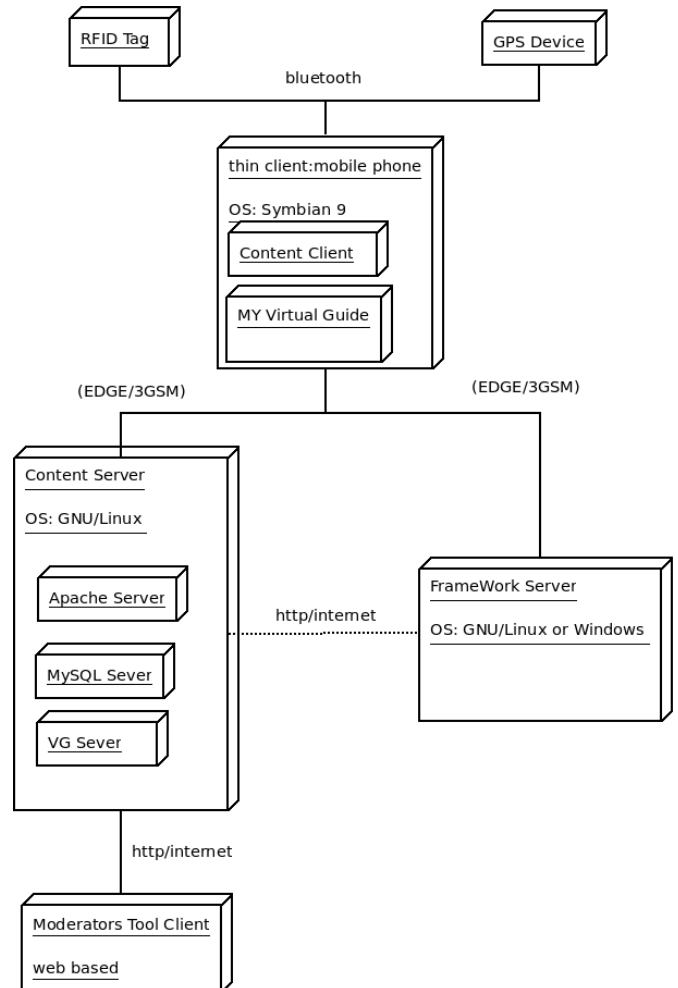


Fig. 2. Diagrama de bloques de los componentes del proyecto.

El código de la aplicación MPISTE esta conformada por dos partes. La primera, el cliente, fue desarrollada en J2ME para el dispositivo móvil (celular Nokia E62). Además de J2ME, el cual es una versión especial para dispositivos móviles, se utilizó el JSR 82 (Java Specification Request), que es el API para Bluetooth. En la parte del servidor se utiliza J2EE y se utilizan servlets que reciben peticiones a través de conexiones con protocolo EDGE/3GSM.

Para el desarrollo del cliente se utilizaron 3 capas., debido a que se han utilizado conceptos de ingeniería de software como patrones de diseño

para que la aplicación pueda ser extendida. En la figura 3 se muestran las tres capas.

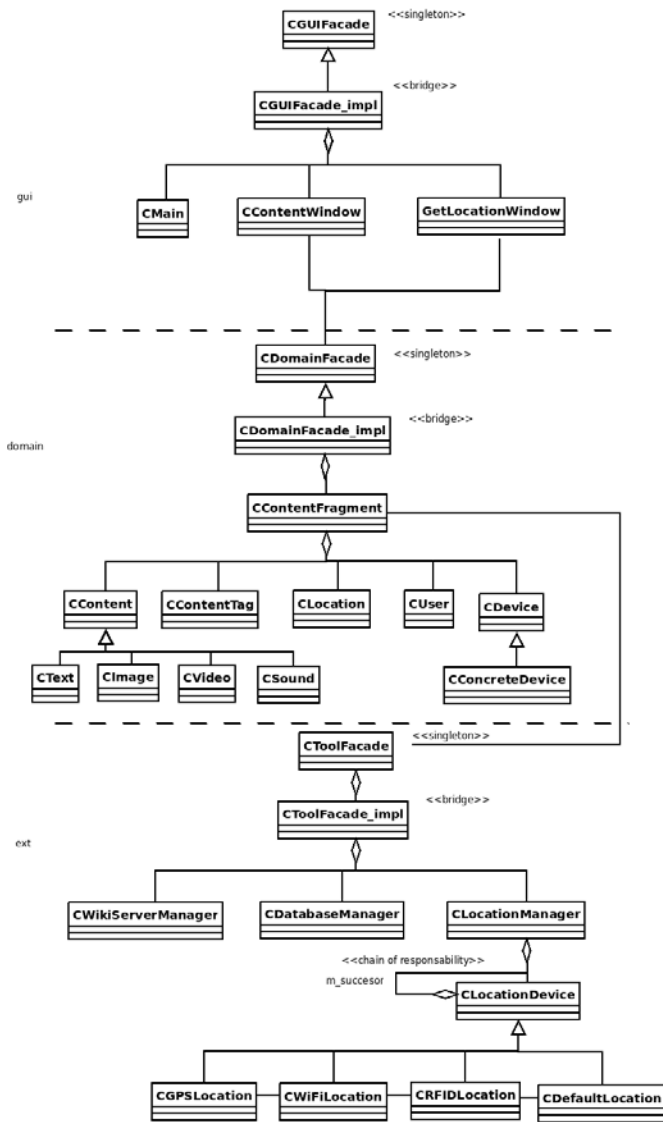


Fig. 3. Diagrama de clases para aplicación en el cliente.

1. La primera capa es conformada por las clases de interfaz del usuario como son las ventanas que se utilizan en el celular.
2. La segunda capa es conformada por las clases del dominio del problema, donde se encuentran clases relevantes a la aplicación como son el contenido del fragmento de información que se utiliza para generar interactivamente la historia.
3. La tercera capa es conformada por clases para realizar actividades auxiliares como

manejar la conexión con servidor por medio del protocolo EDGE/3GSM y para manejar la conexión (a través del “framework” abstracto para ubicación) con los dispositivos como GPS y RFID.

La aplicación para el servidor se desarrolló en 2 capas (como se muestra en la figura 4). La capa superior son servlets, o pequeños programas para aplicaciones web, los cuales reciben peticiones enviadas por los clientes (dispositivo móvil). Estas peticiones son atendidas y referidas a la clase correspondiente en la segunda capa. La segunda capa maneja la conexión con la base de datos y es donde se realizan consultas por medio de “queries”. El resultado de éstos es regresado al dispositivo móvil.

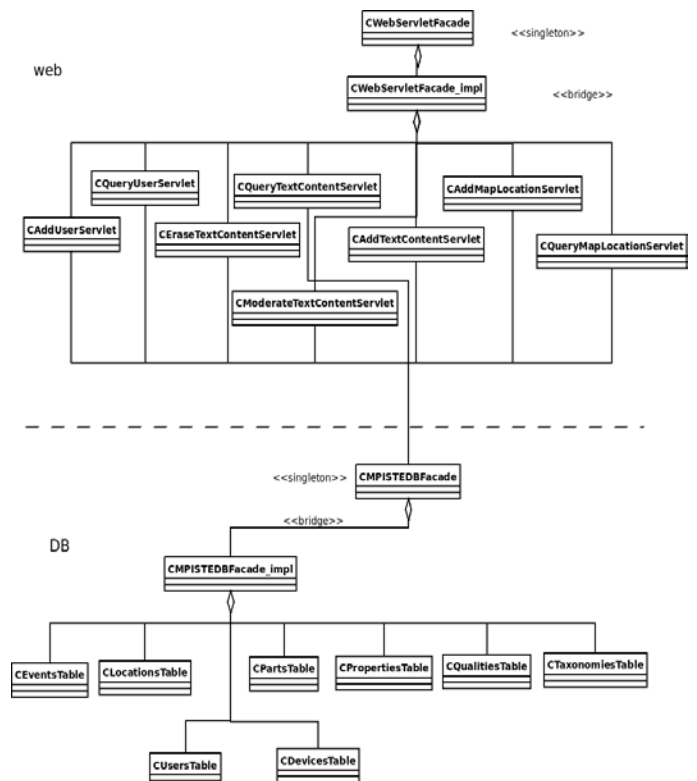


Fig. 4. Diagrama de clases para aplicación en el servidor.

MPISTE se probó con el “tree trail” que se ubica en el campus de la universidad de Heriot Watt. Este trail esta compuesto por un conjunto de 23 árboles que están etiquetados con números del 1 al 23. Para está prueba se colocaron “tags” RFID para

cada uno de ellos. En la base de datos se ligó la etiqueta RFID con la información correspondiente al árbol.

La aplicación se ha probado en un camino con árboles históricos. Esta aplicación es ideal puesto que la información, al ser sencilla se almacenó en una base de datos MySQL en muy pocas tablas. Además, es muy sencillo adecuar la aplicación para generar historias sobre sitios históricos en un entorno urbano, el cual es el siguiente paso en éste proyecto conjunto. La información se obtiene de la siguiente manera.

1. Se lee el tag RFID, por medio del lector en forma de pluma que el usuario sostiene con su mano derecha. Tal y como se muestra en la figura 5.
2. El número de identificación del tag es enviado por medio de bluetooth al dispositivo móvil, que el usuario sostiene con la mano izquierda en la figura 5.
3. El ID es enviado a través del protocolo EDGE/3GSM al *webserver*. La comunicación se realiza a través de servlets que aceptan las peticiones realizadas por medio del celular.
4. El servlet realiza una petición por medio de un query a la base de datos (MySQL).
5. La base de datos regresa información al servlet (*webserver*)
6. El servidor (*webserver*) envía los datos al celular.
7. En el celular se genera la historia al ligar los fragmentos de la historia de una manera coherente, de forma similar al trabajo descrito en [5].
8. El fragmento de historia se muestra en el dispositivo móvil como se aprecia en la figura 6.



Fig. 5. Lectura de tag RFID



Fig. 6. Historia generada a partir de la información recibida del servidor.

III. CONCLUSIONES

Se ha presentado un entorno basado en ubicación física del usuario para contar historias de forma personalizada en dispositivos móviles. Para ubicar al usuario se utilizó tecnología GPS y RFID. El trabajo descrito en este artículo (MPISTE) consta de dos partes. La primera es el cliente, el cual fue desarrollado para teléfonos celulares inteligentes (*Smartphones*) utilizando la tecnología Java Micro Edition, también conocida como J2ME. La segunda parte desarrollada fue el servidor (*webserver*) mediante el uso de la tecnología Java Enterprise Edition (J2EE). La aplicación servidor recibe peticiones del dispositivo móvil (*smartphone*) y las atiende por medio de servlets, los cuales a su vez realizan peticiones a una base de datos relacional, en este caso MySQL. El protocolo de comunicación utilizado para comunicar el celular con el servidor es EDGE/3GSM. El proyecto se extenderá para utilizarse en sitios con alto contenido de herencia cultural como es el “*Royal Mile*” en Edimburgo, Escocia, o el centro histórico de Aguascalientes, México.

AGRADECIMIENTOS

Los autores agradecen al *Leverhulme Trust* por el otorgamiento del *Leverhulme Fellowship* al primer autor, con lo que se cubrieron parte de los costos del proyecto descrito en el presente documento.

IV. REFERENCIAS

- [1] V. Propp, “Morphology of the Folk Tale”, University of Texas Press, 1968.
- [2] C. Crawford, “On Interactive Storytelling”, New Riders, USA, 2005.
- [3] J. Robertson, J. Good “*Story creation in virtual game worlds*”, Communications of the ACM Vol 48, no. 8, pp. 61-65, 2005.
- [4] G. D. Abowd, C. G. Atkeson, J. Hong, S. Long, R. Kooper, R., M. Pinkerton, “*Cyberguide: a mobile context-aware tour guide*”. Wireless Networks. 3, 5 (Oct. 1997), pp 421-433, 1997.
- [5] J. Ibañez, R.S. Aylett, and R. Ruiz-Rodarte. “*Storytelling in virtual environments from a virtual guide perspective*”, 3rd Virtual Reality Journal, Volume 7, Number 1 Springer London 2003, pp 30-42.
- [6] J. Ibañez, C. Delgado-Mata, R.S. Aylett, and R. Ruiz-Rodarte. “*Don't you escape I'll tell you my story, Intelligent Virtual Agents*”, 3rd Mexican International Conference on Artificial Intelligence, MICAI 2004, Mexico City, Mexico, 2004, Proceedings. Lecture Notes in Artificial Intelligence 2972 Springer 2004, pp 49-58.
- [7] M.Y. Lim, R.S. Aylett, “*Feel the difference: A Guide with Attitude!*”, Intelligent Virtual Agents, 7th International Conference, IVA 2007, Paris, France, September 17-19, 2007, Proceedings. Lecture Notes in Computer Science 4722 Springer 2007, pp 317-330.
- [8] O. Rashid, W. Bamford, P. Coulton, R. Edwards, J. Scheible, “*PAC-LAN: mixed-reality gaming with RFID-enabled mobile phones*”. Computers in Entertainment 4, 4 2006.
- [9] S. Benford, D. Rowland, M. Flintham, A. Drozd, R. Hull, J. Reid, J. Morrison, K. Facer, “*Life on the edge: supporting*

collaboration in location-based experiences". Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Portland, Oregon, USA, April 02 - 07, 2005). CHI '05. ACM, New York, NY, pp. 721-730, 2005

Applications", Internal Report, Heriot Watt University, 2009.

- [10] L. Chittaro, D. Nadalutti, "*Presenting evacuation instructions on mobile devices by means of location-aware 3D virtual environments*". In Proceedings of the 10th international Conference on Human Computer interaction with Mobile Devices and Services (Amsterdam, The Netherlands, September 02 - 05, 2008). MobileHCI '08. ACM, New York, NY, pp 395-398. 2008
- [11] J. E. Bardram, N. Nørskov, "*A context-aware patient safety system for the operating room*". In Proceedings of the 10th international Conference on Ubiquitous Computing (Seoul, Korea, September 21 - 24, 2008). UbiComp '08, pp 272-281, 2008.
- [12] S. Benford, A. Crabtree, M. Flintham, A. Drozd, R. Anastasi, M. Paxton, N. Tandavanitj, M. Adams, J. Row-Farr, "*Can you see me now?*". ACM Trans. Comput.-Hum. Interact. 13, 1 (Mar. 2006), pp. 100-133. 2006.
- [13] J. Paay, J. Kjeldskov, A. Christensen, A. Ibsen, D. Jensen, G. Nielsen, R. Vutborg "*Location-based storytelling in the urban environment*", OZCHI '08: Proceedings of the 20th Australasian Conference on Computer-Human Interaction, ACM, pp. 122-129, 2008.
- [14] S. Eldwaib, R. Pooley, "Reusable Framework for Location-Aware Mobile

CARNET MÉDICO DIGITAL PARA TELÉFONOS CELULARES

Mauricio de la Cruz¹, Félix Mata², Luis Bustos³

*Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas,
Instituto Politécnico Nacional.
Avenida Instituto Politécnico Nacional No. 2580, Colonia Barrio la Laguna Ticomán,
Delegación Gustavo A. Madero, CP. 07340 México D.F.
Correo electrónico: ^{1,2}mmatar@ipn.mx, ³almighty@ieee.org*

RESUMEN

Se presenta un sistema que provee una alternativa electrónica para la gestión del carnet médico de un paciente. El carnet se almacenará en el teléfono celular de cada paciente, y podrá ser accedido vía Bluetooth, mediante una aplicación que portarán los paramédicos en su celular. Con esto se pretende poner la información médica relevante de los pacientes al alcance de los paramédicos, cuando estos últimos requieran brindar atención médica de emergencia a los primeros; de esta forma el paramédico podrá visualizar en su celular la información concerniente a alergias y padecimientos que pudiese tener el paciente y así estará en posibilidad de darle un trato adecuado. Además, el sistema cuenta con una aplicación Web en una PC que permita, generar el carnet, enviarlo al celular del paciente así como registrar reportes que generen los paramédicos en su celular. La transmisión de la información entre la PC y los dispositivos se hará usando la tecnología Bluetooth.

Palabras clave: Sistema Web, Dispositivos móviles, Bluetooth.

I. INTRODUCCIÓN

Los dispositivos móviles hoy en día tienen muchas prestaciones y diversas formas de conectividad, en especial la que se da a través de la tecnología Bluetooth. Además, cuentan con diversas utilidades y funcionalidades, que nos permiten generar aplicaciones para estos. Los dispositivos móviles son usados tanto para comunicar a las personas como para el entretenimiento, pero existen muchas posibilidades para explotar el potencial de estos artículos, una de ellas es utilizarlos para almacenar información relevante de sus portadores.

El carnet médico es un documento personal muy importante que es indispensable a su titular, ya que contiene información relevante sobre el historial médico del individuo, además de que sirve para identificarse con el hospital al que están afiliados. Este documento generalmente es archivado por el propietario y es portado sólo cuando éste tiene contemplado hacer uso de los servicios médicos.

En la vida cotidiana las personas no están exentas de sufrir un accidente, que dependiendo de su gravedad podría requerir atención médica de emergencia. Sucede comúnmente en este tipo de situaciones, que el accidentado se encuentra imposibilitado para dar información al personal médico, ya sea por la naturaleza de las lesiones o por efectos psicológicos derivados del trauma.

En estos casos es común encontrar entre las pertenencias de las personas artículos como permisos de conducir, identificaciones, tarjetas de crédito, llaves y en muchas de las ocasiones dispositivos móviles (en la mayoría de estos casos celulares), pero es poco probable encontrar entre sus pertenencias información médica relevante para estas situaciones.

Según datos del INEGI, en marzo del 2008 los hogares con servicio de telefonía ascendían a 20 967 438 (75.5% del total de hogares), de los cuales 10 184 547 (36.7% del total de hogares) hogares contaban con telefonía fija y móvil y 6 760 936 (24.3% del total de hogares) contaban únicamente con telefonía móvil, de lo que podemos concluir que en el 61% de los hogares se cuenta con teléfonos celulares [1].

Se puede apreciar que un gran número de personas cuentan con un teléfono celular, y que con el tiempo éste ha dejado de verse como un artículo de lujo al que pocas personas tienen acceso.

Otro punto a considerar sobre los teléfonos celulares, es que son dispositivos cuya función principal implica que el usuario debe portarlos en todo momento, especialmente cuando está fuera de casa. Esto los hace idóneos para almacenar información a la que el usuario deba tener acceso en todo momento.

El sistema presentado utiliza las capacidades y tecnologías de los dispositivos móviles para crear un carnet electrónico, que se almacena en los dispositivos móviles, y que puede ser accedido vía Bluetooth por un paramédico, en caso de una emergencia.

II. PLANTEAMIENTO DEL PROBLEMA

En México existen diversas causas de mortalidad. De acuerdo con el INEGI, la diabetes, las enfermedades del corazón y los accidentes de transporte son las causas de más del 56% de las muertes en México [2].

Uno de los más grandes problemas que enfrentan los paramédicos es proporcionar una atención adecuada cuando se cuenta con poca o nula información sobre el paciente; ya que ésta es la primera atención que recibe un individuo por parte de un hospital cuando a éste le resulta imposible trasladarse.

El protocolo de atención precisa que antes de tratar al paciente, el paramédico debe hacerle una serie de preguntas que le permitan diagnosticar la condición del paciente así como saber de padecimientos previos y/o condiciones médicas particulares del paciente que pudieran requerir un trato especial. Sin embargo no se contemplan dos escenarios: el menos grave se presenta cuando el paciente no tiene conocimiento preciso de todos sus datos médicos, por lo que se puede obtener información errónea o incompleta; el segundo, es cuando debido a la naturaleza de un accidente, el afectado pierde la capacidad para comunicarse, en este caso,

los paramédicos se ven forzados a trabajar en la incertidumbre, arriesgando la vida del accidentado: cualquier decisión que tomen podría ser contraproducente para el paciente.

Es en casos como estos, se pretende hacer uso del sistema presentado, para proporcionar a los paramédicos la información necesaria para dar un tratamiento seguro y eficaz al paciente.

III. TRABAJOS RELACIONADOS

Existen diversas aplicaciones para dispositivos móviles enfocadas a la atención médica, por ejemplo, “*primary care plus ambulatory and Hospital Care Suite (PCP)*” el cual es una suite que permite consultar información médica en sistemas Palm OS, Windows Mobile, Mobile Wireless o ser consultado por Internet. En esta se detallan las monografías de prácticamente todas las enfermedades y condiciones que un médico puede encontrar en una oficina o un hospital. También está el caso de *Emedic* que es una enciclopedia médica. Y el proyecto de *medical records on your cell phone* [3] en el que investigadores en computación trabajan actualmente, con este proyecto las personas que cuenten con un dispositivo móvil podrán tener sus registros médicos (como radiografías, registros de presión, etc.), estos dispositivos deben tener una gran capacidad de procesamiento gráfico para poder utilizar esta aplicación.

Las dos primeras aplicaciones únicamente sirven de referencia para el diagnóstico, la tercera cuenta con la información del paciente, sin embargo, presenta el inconveniente de que se debe contar con dispositivos celulares con una capacidad que solo exhiben los dispositivos de nueva generación.

El carnet médico digital, puede ser utilizado en cualquier sistema que soporte la tecnología J2ME (*Java 2 Micro Edition*) y la especificación JSR82 lo que permite incluir a dispositivos no tan recientes.

IV. MARCO TEÓRICO

LENGUAJE DE PROGRAMACIÓN PARA DISPOSITIVOS MÓVILES

Uno de los lenguajes de programación más usados en el campo de los dispositivos móviles es J2ME, que está presente en diversas plataformas y sistemas operativos

MEDIO DE TRANSMISIÓN DE LOS DISPOSITIVOS MÓVILES CON LA TERMINAL DEL HOSPITAL.

Debido a que actualmente los dispositivos móviles, cuentan con diversos tipos de conectividad, entre ellos Bluetooth, se hará uso de esta tecnología de radiofrecuencia, cuya funcionalidad puede ser accedida por software mediante el uso del lenguaje de programación Java.

APIs PARA BRINDAR LA CAPACIDAD DE CONEXIÓN BLUETOOTH EN LOS MÓVILES (J2ME).

El medio de comunicación elegido es Bluetooth, para esto, el Java Community Process (JCP) define el JSR 82 (*Java Specification Requests*) que está implementado en J2ME. En este documento se especifican las clases, interfaces y métodos para brindar esta funcionalidad en dispositivos móvil como PDAs, celulares o Smart Phones.

Existen dos perfiles recomendados para la transferencia de archivos, uno de ellos es GOEP (Generic Object Exchange Profile) que establece una serie de protocolos y procedimientos para intercambiar objetos (por ejemplo transferir archivos). Emplea el estándar OBEX. Necesita el paquete Javax.Obex.

Por otro lado tenemos el perfil SPP (Serial Port Profile) que es usado para establecer una comunicación serial virtual entre 2 dispositivos vía Bluetooth. Necesita el paquete Javax.bluetooth.

Dados estos perfiles se tiene que la mayoría de los dispositivos móviles no soportan el estándar OBEX (o para su implementación, es necesario añadir paquetes especiales que correspondan a la versión y modelo de teléfono) y lo que se pretende es que las aplicaciones puedan ser utilizadas por

varios dispositivos móviles (debido a que las capacidades de éstos varían considerablemente), por lo tanto se opto por usar el segundo perfil que es soportado por la mayoría de los teléfonos con conectividad Bluetooth.

APIs PARA BRINDAR LA CAPACIDAD DE CONEXIÓN BLUETOOTH EN LA TERMINAL DEL HOSPITAL (J2SE)

Para hacer uso de la comunicación Bluetooth en la terminal del hospital, J2SE no implementa de manera nativa este tipo de comunicación. Estas APIs no son de uso libre y es necesario comprar alguna de las implementaciones comerciales. Algunos de los proveedores de la implementación del JSR 82 para J2SE son los siguientes:

Nombre de la compañía	Soporte JSR-82 JAVAX. BLUETOOTH	Soporte JSR-82 JAVAX. OBEX	Plataformas soportadas.	Sistemas Operativos soportados.
Atinav	Si	Si	J2ME, J2SE	Win-32, Linux, Pocket PC
BlueGiga	No	No	Waba JVM	uClinux
Ericsson	No	No	J2SE	Win-32, Linux
Esmertec	Si	NO	J2ME	Win-32, Palm OS, Pocket PC, entre otros
Harald	No	No	J2SE	Win-32, Linux
Possio	Si	Si	J2ME	Win-32, Linux
Rococo	Si	Si	J2ME, J2SE	
Smart Network Devices	Si	No	J2ME	HyNetOS
Super-Waba	No	No	Waba JVM	Palm OS
Avetana	Si	Si	J2ME, J2SE	Win-32, Linux, Palm OS, Pocket PC
BlueCo-ve	Si	Si	J2ME, J2SE	Win-32, Linux, Palm OS, Pocket PC
Zucotto	No	No	J2ME, J2SE	Win-32

Tabla 1. Bibliotecas para manejar Bluetooth

APLICACIÓN WEB

Para la aplicación Web se hizo uso de la tecnología JavaServer Pages (JSP). Ésta es una tecnología Java que permite generar contenido dinámico para

la Web, en forma de documentos HTML, XML o de otro tipo, además de contar con compatibilidad con las especificaciones para Bluetooth. Es una tecnología muy desarrollada y cuenta con mucha información.

V. SOLUCIÓN DESARROLLADA

El carnet médico digital es un sistema que proporciona 3 aplicaciones:

- **Aplicación para el teléfono del paciente.** Haciendo uso de la tecnología J2ME, se desarrolló una aplicación para el teléfono del paciente, la cual tiene la capacidad de almacenar la información del carnet medico así de como enviar y recibir la información utilizando la tecnología Bluetooth con la que actualmente cuentan la mayoría de los teléfonos celulares. Si el paciente quiere visualizar su información, deberá introducir un código (que si se introduce erróneamente 3 veces borrara la información contenida). El código le dará acceso a la información, con esto se busca cierto grado de confidencialidad que se requiere para esta información personal previniendo el caso de que por una situación ajena el celular cayera en manos equivocadas.
- **Aplicación para el teléfono del paramédico.** Esta aplicación pedirá una contraseña para acceder a sus funcionalidades. Ofrecerá las siguientes opciones: recibir carnets de pacientes (para lograr esto deberá introducir un número en el teléfono del paciente que permitirá la conexión entre ambos dispositivos), una vez recibida la información podrá desplegarla en el celular, debido a que podría ocurrir que recibiera información de personas con características físicas similares, recibirá adicionalmente la fotografía del paciente, esta información podrá ser almacenada y consultada posteriormente. También la aplicación tiene la capacidad de generar reportes de las personas que sean atendidas para posteriormente registrarlos en la base de datos del hospital.
- **Sistema Web.** Para el registro y transferencia de carnets y reportes. Radica en una PC con

conectividad Bluetooth, permite hacer los registros pertinentes de pacientes y paramédicos. Hace uso de la conectividad Bluetooth, para la comunicación entre dispositivos.

VI. ARQUITECTURA DEL SISTEMA

La arquitectura del sistema puede dividirse en dos grandes partes, una es la aplicación que radica en el hospital, que es el sitio donde se registran los administradores, paramédicos y pacientes, además, aquí se hace la transferencia de la tarjeta de identificación del paramédico y la del carnet del paciente, a sus respectivos dispositivos móviles.

La aplicación Web del hospital también cuenta con la capacidad de registrar la información recibida por la terminal del hospital vía Bluetooth que contiene los reportes generados en el móvil de los paramédicos (ver fig. 1).

La arquitectura está conformada por los siguientes módulos:

Administración de usuarios, en este módulo, el administrador es el encargado de gestionar la información de usuarios, clasificando a estos como: administradores, paramédicos y pacientes. En el escenario correspondiente al registro de paramédicos, se generará la tarjeta de identificación del paramédico (archivo XML) y será enviada a su celular. En el escenario correspondiente al registro de pacientes, el administrador podrá generar el carnet del paciente (archivo XML) y transferirlo al celular del paciente. Cuando se realicen modificaciones a los datos del carnet, el administrador podrá enviar dichas modificaciones al teléfono del paciente. En el caso de registro de administradores se darán altas a las sesiones de administrador. Los procesos de envío y recepción de datos, se harán vía Bluetooth.

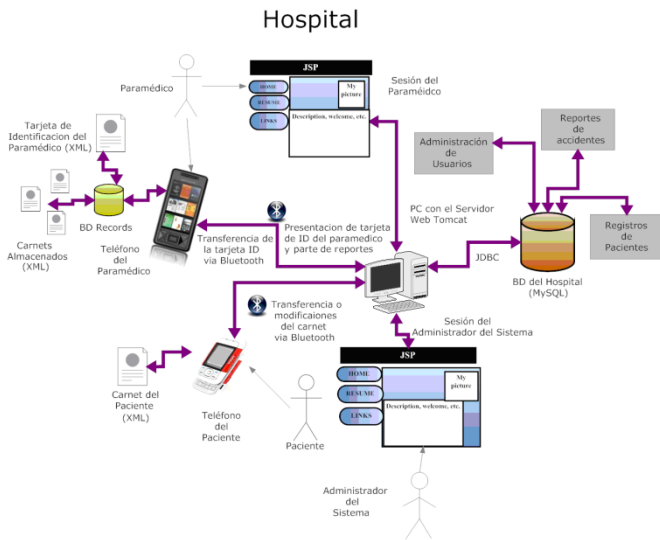


Fig. 1. Arquitectura del módulo residente en el hospital

Entrega de reportes (ver fig. 2), este proceso lo realizarán los paramédicos, para ello, deberán iniciar sesión presentando la tarjeta de identificación que les generó el administrador, una vez iniciada la sesión, podrán descargar los datos que fueron registrados en su celular a través de la aplicación, además podrán registrar los datos restantes del reporte y almacenar esos registros en la base de datos.

Atención de eventos, permite al paramédico que porte en su celular la aplicación, recibir el carnet electrónico correspondiente, del celular del paciente atendido.

La aplicación que porte el paramédico, será capaz de registrar los datos concernientes al evento atendido así como datos de la atención brindada al paciente.

Los escenarios de conflicto posibles, además de la solución que se presenta, se describen a continuación:

1. La existencia de diversos dispositivos con conectividad Bluetooth alrededor del paramédico puede dificultar la transferencia del carnet. Esto se resuelve de la siguiente forma:

Para el establecimiento de la comunicación Bluetooth, es necesario realizar primero una búsqueda

de dispositivos para ver con qué dispositivos se puede conectar y después realizar una búsqueda de servicios en cada dispositivo encontrado. Para evitar realizar búsquedas extra de dispositivos, el sistema los catalogará de acuerdo a su tipo de dispositivo para que la búsqueda de servicios solo se realice en celulares y no en otros dispositivos, logrando así reducir los tiempos de establecimiento de conexión. Además la búsqueda se realiza en el modo oculto que ofrece Bluetooth.

2. La existencia de varios dispositivos que cuenten con la misma aplicación alrededor del paramédico.

Para que el paramédico tenga el control sobre la recolección del carnet del paciente, por lo general, él mismo será encargado de activar la aplicación que está en el teléfono del paciente, para ello introducirá un código en su teléfono, mismo que posteriormente deberá introducir en el teléfono del paciente, con lo anterior se asegura que únicamente se obtendrán los datos del paciente.



Fig. 2. Arquitectura del módulo de atención de eventos

VII. RESULTADOS OBTENIDOS

Para esta etapa del desarrollo ya se tienen integrados los módulos que conforman al sistema, a continuación se da una breve explicación de los módulos y se muestran algunas de las pantallas correspondientes.

MÓDULOS DE REGISTRO

Tras haber iniciado sesión como administrador el módulo permite registrar un nuevo administrador o un paramédico, se piden los datos y se valida que

los campos no estén vacíos, posteriormente se realiza el registro y se muestra el resultado de la operación. La fig. 3 fue capturada al hacer el registro de un administrador. El registro procede de la misma forma cuando se trata de un paramédico.

Tras ser registrado correctamente un paramédico, se envía la tarjeta de ID del paramédico a su móvil, la recepción de la tarjeta de registro se muestra en la fig. 4.

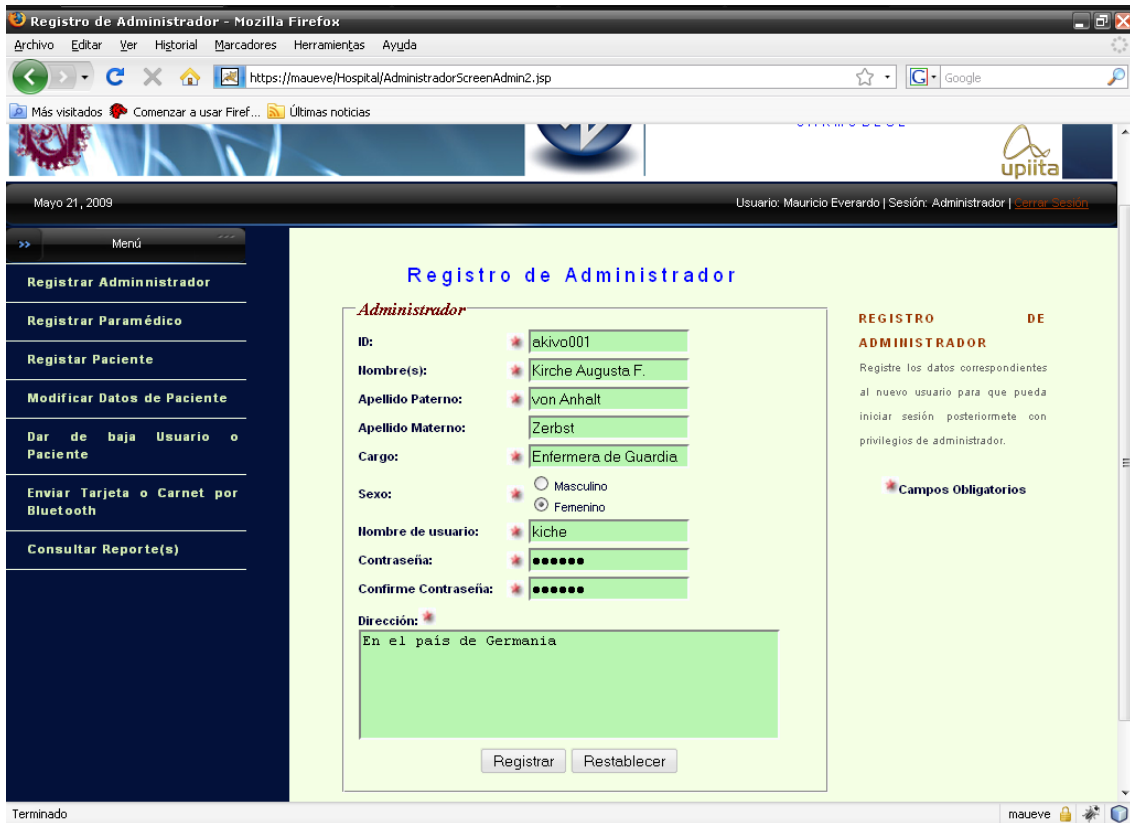


Fig. 3. Sistema Web, módulo de registro de administrador

REGISTRO DE PACIENTES

El registro de pacientes, difiere un poco con respecto al registro de administradores y paramédicos, ya que se debe almacenar más información, específicamente, la información médica requerida en el carnet. La fig. 5 muestra una captura de la pantalla de registro de pacientes, esta se puede acceder desde la sesión de administrador.



Fig. 4. Recepción de la tarjeta de registro

En la fig. 6 se muestra la segunda parte del registro de pacientes. En la primera parte se ingresan los datos personales, y en esta segunda parte se ingresan datos sobre los padecimientos, alergias y vacunas asociados al paciente. Como se puede observar, en el cuadro superior, se tienen dos columnas, en estas se pueden listar hasta 10 padecimientos, y 10 alergias que pudiera presentar el paciente. En el cuadro inferior se registran las vacunas con las que cuenta el paciente, así como las fechas en las que estas fueron administradas.

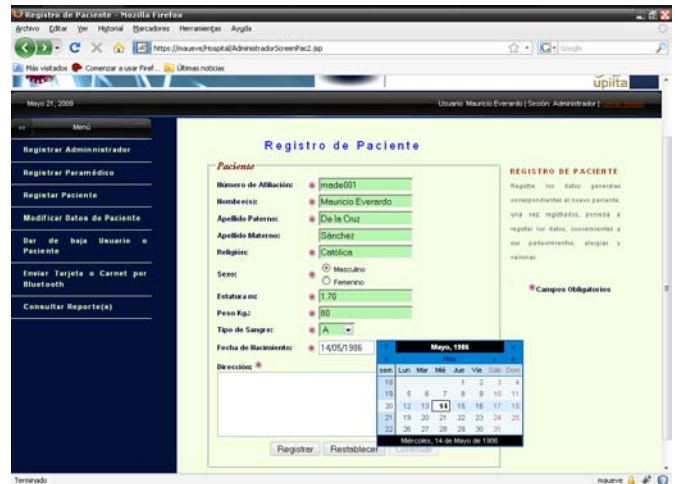


Fig. 5. Registro de Pacientes

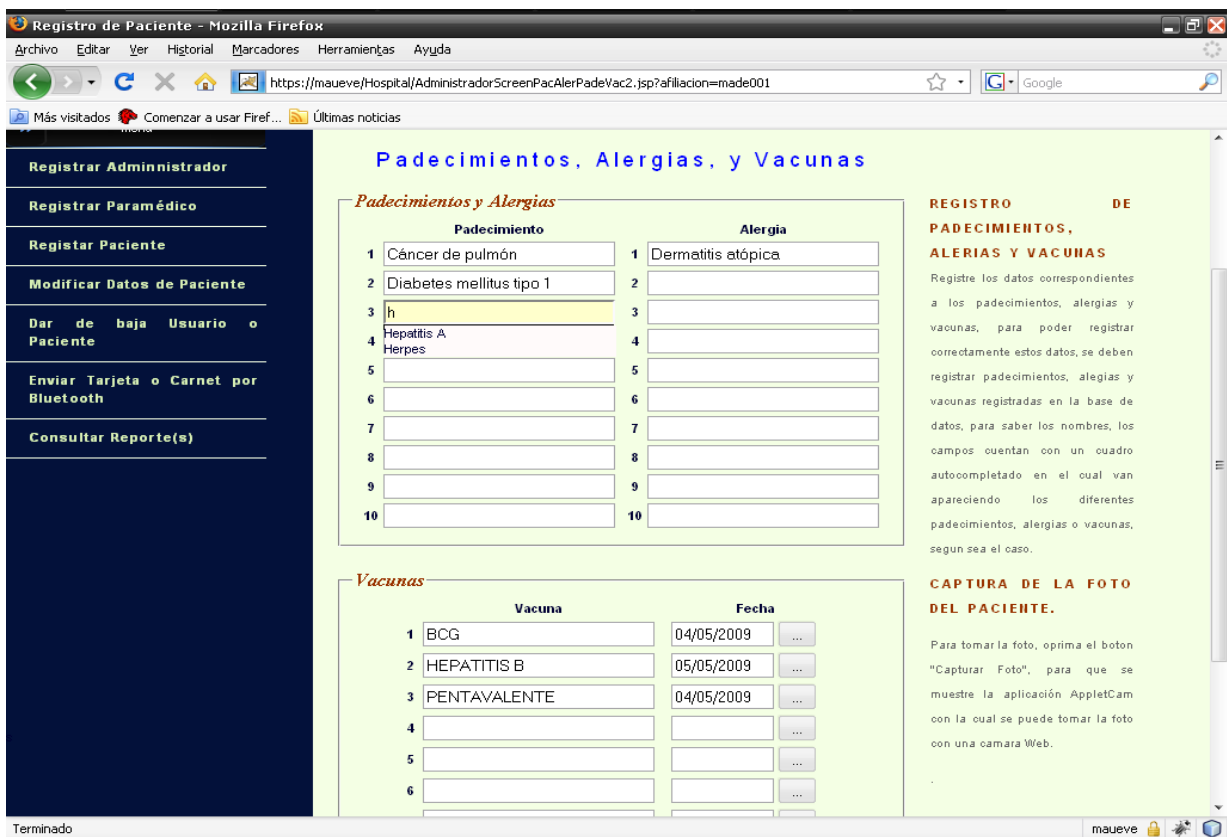


Fig. 6. Registro de pacientes (segunda parte)

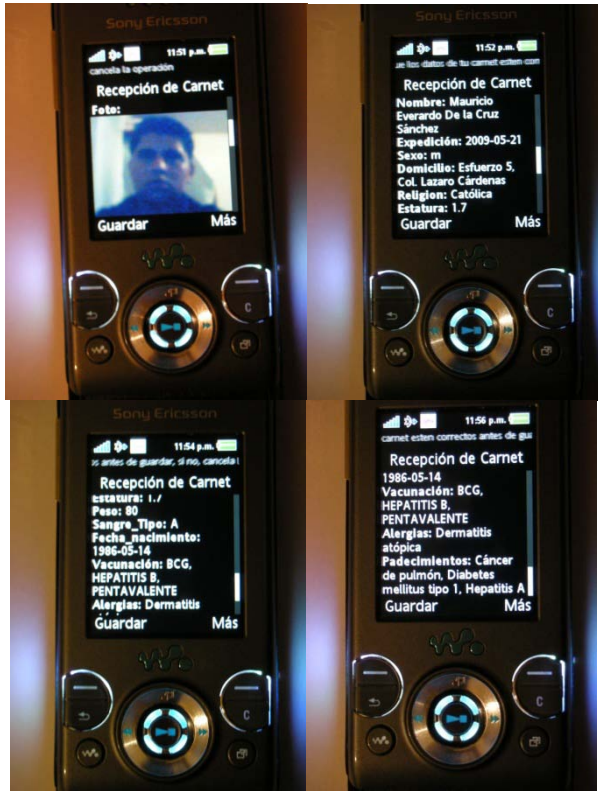


Fig. 7. Recepción del carnet

En la fig. 7 se puede observar el proceso de recepción de carnet en el celular del paciente. Este proceso es la última parte del registro del paciente, una vez que los datos se han introducido por medio del sistema Web, se hace la transferencia del carnet al teléfono.

APLICACIÓN DEL PARAMÉDICO

El paramédico tendrá en su teléfono una aplicación que le permita descargar el carnet del paciente. Posteriormente, el paramédico utilizará esta aplicación para registrar los datos del evento requeridos para la generación de un reporte, específicamente se registran los signos vitales del paciente, los tratamientos brindados, el material utilizado, y los datos del traslado.

En la fig. 8 se muestra un ejemplo de la aplicación al hacer una transferencia. En primer lugar, se ejecuta la aplicación del paramédico y se solicita preparar la recepción del carnet, con esto se genera un número pseudoaleatorio de 2 dígitos para la verificación de la transferencia. El segundo paso es utilizar el celular del paciente para realizar el envío,

mediante la opción *Enviar Carnet* del programa que se ejecuta en este dispositivo, la aplicación pedirá al usuario introducir el número de verificación generado en el paso anterior. Finalmente, el dispositivo hace la búsqueda Bluetooth del equipo del paramédico y se conecta a este para realizar la transferencia.

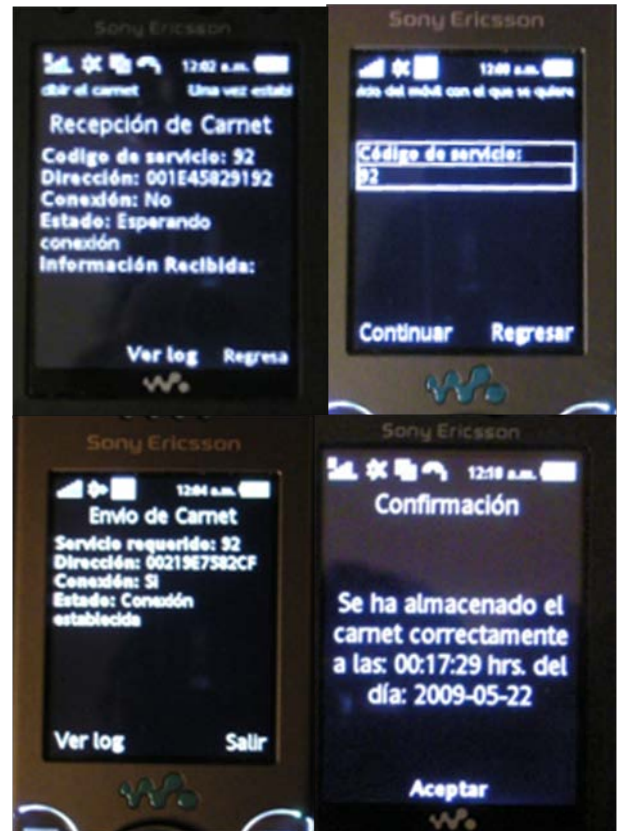


Fig. 8. Transferencia del carnet.

En la fig. 9 se muestra un ejemplo del uso de la aplicación del paramédico para el registro de reportes. Se inicia por registrar los signos vitales, como son temperatura y presión, esto se ve en la primera pantalla. Posteriormente, se hace el registro de los medicamentos administrados. Este registro se puede hacer varias veces si se administra más de un medicamento. De la misma forma, se registra el material utilizado en el evento. Para finalizar el reporte se requiere que el paramédico ingrese los datos de ubicación. Estos se piden al usuario en la pantalla final.

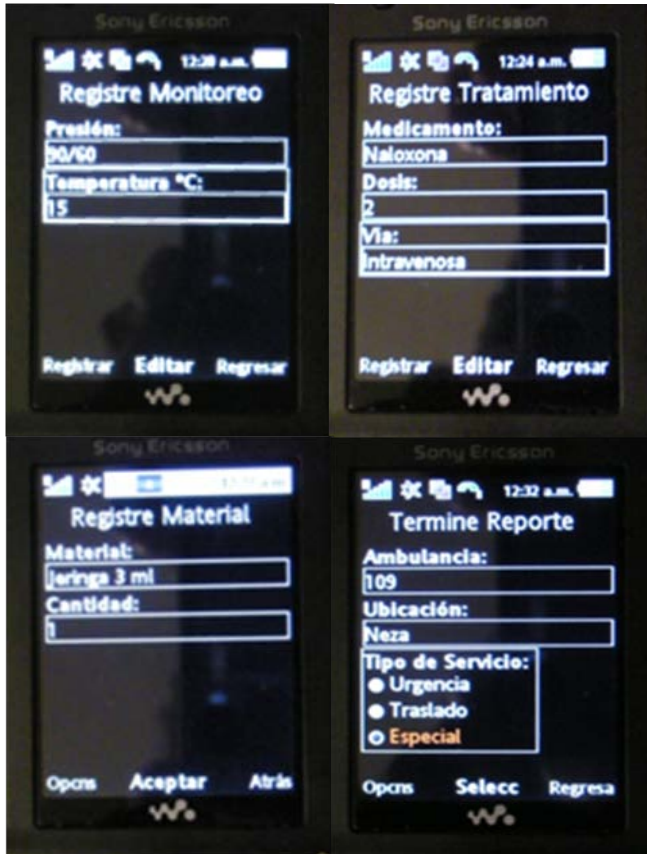


Fig. 9. Realización de reporte

SISTEMA WEB: SESIÓN DEL PARAMÉDICO

Esta es la parte del sistema Web para su uso por los paramédicos. A través de su sesión en el sistema, los paramédicos pueden ingresar sus reportes al sistema, así como realizar la consulta de reportes anteriores. Para tener acceso a su sesión, el paramédico deberá identificarse mediante una tarjeta digital de identificación, ésta la deberá presentar al sistema mediante una conexión a través de Bluetooth. Cabe destacar que se utiliza la capa SSL durante la sesión del paramédico para incrementar la seguridad del sistema ante intrusos.



Fig. 10. Transferencia de reportes

En la fig. 10, podemos observar la captura de una parte del la pantalla del sistema Web, durante la sesión de paramédico. El sistema utiliza un *applet* para administrar los reportes del paramédico, y para permitirle realizar la transferencia de estos desde su teléfono celular a la base de datos, a través de la conexión inalámbrica Bluetooth.

Además de ingresar nuevos reportes al sistema, el paramédico puede consultar antiguos reportes desde su sesión en el sistema. Las figs. 11 y 12, muestran la captura de un ejemplo de esta aplicación.



Fig. 11. Consulta de reportes.

Reporte

Datos Generales

Folio:	24
Afiliación:	made001
Fecha de Atención:	2009-05-22
Hora de Atención:	00:17:29
Ambulancia:	109
Servicio:	Especial
Ubicación:	Neza
ID de paramédico:	plule001
Número de paramédico:	666

Monitoreos

Húmero	Fecha	Hora	Presión	Temperatura °C
1	2009-05-22	00:20:12	90/60	15

Tratamientos

Húmero	Fecha	Hora	Medicamento	Dosis	Vía
1	2009-05-22	00:24:45	Naloxona	2	Intravenosa

Materiales

Húmero	Descripción	Cantidad
1	Jeringa 3 ml	1

REPORTE

El reporte consta de 4 partes que son: Datos Generales, Monitoreos, Tratamientos y Materiales.

Fig. 12. Ejemplo de un reporte.

VIII. CONCLUSIONES

Se presentó un sistema que asiste tecnológicamente a pacientes y personal médico, en las tareas que involucra un carnet medico, en este caso dicho carnet es un documento digital, cuya gestión se realiza a través de la comunicación Bluetooth y aplicaciones Java para móviles.

Se utilizó la pila *Bluetooth* con al API de WID-COMM (dado que otras APIS, como *Bluesock* y *BlueSoleil* el api de *bluecove*, presentó problemas en pruebas realizadas, por lo que se debe verificar la compatibilidad) e instalar JMF (Java Media Framework) para poder hacer uso de la cámara Web. El sistema fue probado en Windows XP además este sistema se probó en sistemas Linux con la implementación BlueZ móviles.

En la versión de escritorio, al hacer uso del api de *bluecove* se contó con las funcionalidades de la

especificación JSR-82 (Bluetooth) de manera completa, por lo que se pudo habilitar la funcionalidad de comunicaciones cifradas.

En los móviles utilizados fue posible habilitar la funcionalidad de cifrado que ofrece la especificación.

De acuerdo a los resultados obtenidos se considera llevar a cabo las siguientes tareas como trabajo a futuro:

- Extender el uso de la firma digital a los documentos utilizados entre aplicaciones, en conjunto con la verificación de la firma.
- Integración de la aplicación móvil del paramédico a otros sistemas médicos para adquisición de información vía bluetooth.

IX. REFERENCIAS

- [1] <http://www.inegi.org.mx/est/contenidos/esp/anol/rutinas/ept.asp?t=tnf220&c=9194>
- [2] <http://www.jornada.unam.mx/ultimas/2008/11/01/diabetes-males-del-corazon-y-accidentes-causan-56-de-muertes-en-mexico-inegi>
- [3] <http://www.avetana-gmbh.de/avetana-gmbh/produkte/jsr82.eng.xml>
- [4] M. Munoz, V. González, M. Rodríguez and J. Favela “*Supporting context-aware collaboration in a hospital: An ethnographic informed design*”, *lecture notes 2806*. Springer, New York, NY, 330-344
- [5] Stephanos Androutsellis-Theotokis , Diomidis Spinellis, “*A survey of peer-to-peer content distribution technologies*”, *ACM Computing Surveys (CSUR)*, v.36 n.4, p.335-371, December 2004
- [6] N. Bricon-Souf and C. R. Newman, “*Context awareness in healthcare: A review*”, *Intl. journal of Medical Informatics* 76, 2--12, 2007.
- [7] H. Bludau and A. Koop, “*2002. Lecture Notes in Informatics: Mobile Computing in Medicine*”, Kellen Verlag.
- [8] Paul J. Camp , James M. Hudson , Russell B. Keldorph , Scott Lewis , Elizabeth D. Mynatt, “*Supporting communication and collaboration practices in safety-critical situations*”, *CHI '00 extended abstracts on Human factors in computing systems*, April 01-06, 2000, The Hague, The Netherlands
- [9] Jonas Landgren and Urban Nulden, “*A study of emergency response work: patterns of mobile phone interaction*”, in *Proc. 2008 ACM Conference on Computer Supported Cooperative Work*
- [10] Kjeldskov and M. Skov., “*Supporting work activities in healthcare by mobile electronic patient records*”, in *Proc. 2004 6th Asia-Pacific Conference on Human-Computer Interaction*
- [11] http://www.sciencedaily.com/videos/2006/0306-medical_records_on_your_cell_phone.htm
- [12] <http://www.bluecove.org/>
- [13] Paul J. Camp, J. M. Hudson, R. B. Keldorph, S. Lewis, E. D. Mynatt, “*Supporting communications and collaboration practices in safety-critical situations*”, *CHI '00 extended abstracts on Human factors in computing systems*, April 01-06, 2000, The Hague, The Neatherlands

X. AUTORES

Ing. Mauricio de la Cruz es Ingeniero en Telemática por la Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas del Instituto Politécnico Nacional (UPIITA-IPN)

Dr. Miguel Félix Mata Rivera obtuvo el grado de doctor en el Centro de Investigación en Computación del Instituto Politécnico Nacional (CIC-IPN). Actualmente se desempeña como profesor investigador de la UPIITA en el área de telemática, dirigiendo proyectos relacionados con ontologías, telefonía celular y sistemas de información geográfica.

Luis Enrique Bustos Méndez es estudiante de noveno semestre de la carrera de Ingeniería Telemática en la Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas del Instituto Politécnico Nacional (UPIITA-IPN).

ALGORITMO DE PLANEACIÓN DE MOVIMIENTOS PARA UN BRAZO EN 3D

Cristian E. Boyain y Goytia Luna, Héctor Rafael Orozco Aguirre, Félix Ramos Corchado.

Centro de Investigación y de Estudios Avanzados del I. P. N.
CINVESTAV, Unidad Guadalajara
Av. Científica 1145, colonia el Bajío, Zapopan, 45015, Jal., México
Correo electrónico: {cboyain, horozco, framos}@gdl.cinvestav.mx

RESUMEN

En este artículo, se presenta un algoritmo de planeación dinámica para generar planes de movimiento enfocados a la animación de un brazo en un entorno 3D. Los planes de movimiento son generados usando un mecanismo aleatorio para buscar la siguiente acción a ejecutar sobre el brazo. Este mecanismo está basado en el algoritmo RRT (Rapidly-exploring Random Tree) y en el método de gradiente descendente como métrica del espacio de configuraciones. El algoritmo propuesto toma en cuenta la estructura física del brazo en base a estudios kinesiológicos para generar sólo movimientos físicamente posibles. Además se presenta un caso de estudio desarrollado en Java usando el motor jME (jMonkey Engine).

Palabras clave: Planeación dinámica de movimiento, Algoritmo RRT, Gradiente descendente, Kinesiología, jMonkey Engine.

I. INTRODUCCIÓN

La planeación es una de las actividades más cotidianas del hombre, tanto así, que no se llega a dar cuenta de lo frecuentemente que la realiza. Los planes que el hombre realiza pueden llegar a ser tan complicados como el plan de vida ó tan fáciles como el de tomar un objeto.

Aunque para el hombre el llevar a cabo actividades como la de desplazarse en una habitación o alcanzar algún objeto con la mano es fácil, para un robot o una creatura virtual autónoma (avatar) es una tarea mucho más complicada por lo que se requiere de un plan que cuente con toda la secuencia de estados necesarios para realizar las actividades mencionadas.

Para lograr obtener estos planes se utilizan diferentes enfoques, tales como el de algoritmos genéticos [1] para lograr mover un brazo en base a la planeación de caminos respectivamente [2]. En [3][4] se emplea la inteligencia de enjambre para mover un brazo robótico en 3D o para planear y optimizar caminos para robots móviles. Los enfoques probabilísticos [5][6][7][8][9] son a menudo utilizados para obtener rutas para robots móviles y para generar una rápida reacción ante cambios en el entorno durante la ejecución de un plan.

Por otra parte, en [10] se presenta una propuesta interesante de planeación para generar planes para lograr mover un brazo en 2D. En esta propuesta, se ataca el problema de la gran cantidad de configuraciones posibles cuando se tiene un brazo con muchos grados de libertad (DOF), debido a que al tomar en cuenta la topología del brazo y los obstáculos, es posible reducir el espacio de configuraciones al agrupar configuraciones homotópicas entre sí.

Otra propuesta es mostrada en [11], donde se reduce también el espacio de búsqueda, pero agrupando configuraciones en volúmenes esféricos que rodean cada articulación y donde se encuentran todos los caminos homotópicos para el movimiento de un brazo robótico. Además aquí, se utiliza un algoritmo de control reactivo para seleccionar entre estos caminos al presentarse una colisión con el ambiente durante la ejecución de un camino planeado.

La planeación de movimientos de un maniquí digital es presentada en [12], donde se establecen

tres pasos principales para lograrlo: primero se obtiene una trayectoria libre de colisiones, después la manipulación y la locomoción es sintetizada en la parte de animación y finalmente se verifica el plan para eliminar colisiones residuales que pudieran surgir de la animación generada.

Para manejar el problema de planear para un gran número de articulaciones, en [13] fue presentada una propuesta basada en el mapeo sobre una malla de posturas alcanzables. Además en este trabajo, se separa el movimiento de una creatura virtual en partes y se genera un plan individual de movimiento para cada parte. Para obtener el plan completo de movimiento, se unen todas las partes. Aunque también como se indica en [5], es posible utilizar un método aleatorio para decidir el siguiente movimiento utilizando sólo las articulaciones necesarias para cumplir el objetivo del plan, pudiendo agregar, si así se requiere, más articulaciones durante la ejecución del plan.

En este artículo, el objetivo principal es la obtención de planes para lograr la animación del brazo de un avatar de una manera mucho más natural y autónoma. Por lo que es presentado, un algoritmo de planeación dinámica de movimientos en el que se toma en cuenta la estructura fisiológica del avatar para generar movimientos físicamente posibles y mucho más naturales. La propuesta presentada en este trabajo utiliza el método de gradiente descendente para evaluar los posibles estados siguientes en un plan de animación.

Este trabajo está organizado de la siguiente manera: en la siguiente sección se describe de manera clara la funcionalidad del algoritmo RRT. En la tercera sección se describe la propuesta de este trabajo, donde se propone un algoritmo de planeación dinámica de movimientos para el brazo de un avatar. La cuarta sección muestra un caso de estudio donde se verifica la funcionalidad del algoritmo propuesto. Por último en la última sección se expresan las conclusiones obtenidas de este trabajo.

II. ALGORITMO RRT

El RRT es un algoritmo diseñado para realizar una búsqueda eficiente en un espacio muy grande de configuraciones. Este algoritmo fue desarrollado por LaValle y Kuffner [14]. A continuación se presenta la estructura básica del algoritmo:

```
Build_RRT ( $q_{init}$ ,  $K$ ,  $\Delta s$ )
G.init( $q_{init}$ )
for  $k \leftarrow 1$  to  $k \leftarrow K$  do
   $q_{rand} \leftarrow \text{Rand\_Conf}()$ ;
   $q_{near} \leftarrow \text{Nearest\_Vertex}(q_{rand}, G)$ 
   $q_{new} \leftarrow \text{New\_Conf}(q_{near}, \Delta s)$ 
  G.add_vertex( $q_{new}$ )
  G.add_edge( $q_{near}, q_{new}$ )
return (G)
```

El RRT inicia con una configuración q_{init} como raíz y posteriormente se construyen K vértices como se indica a continuación:

- Se elige una configuración aleatoria q_{rand} , donde, asumiendo una métrica p sobre el espacio de configuraciones, se evalúan todos los vértices obtenidos en G y se toma el más cercano a q_{rand} .
- Se genera una nueva configuración q_{new} , al moverse de q_{near} una distancia Δs en la dirección de q_{rand} . De esta forma se obtiene una secuencia de configuraciones en una estructura de árbol.

III. ALGORITMO DE PLANEACIÓN DINÁMICA DE MOVIMIENTOS PARA UN BRAZO EN 3D

En esta sección se describe un algoritmo para la animación de un brazo en 3D en base a la obtención de planes dinámicos de animación. Este algoritmo permite generar animaciones de una manera mucho más natural y eficiente, evitando así, invertir mucho tiempo en su elaboración. Con este algoritmo se logra una mejor naturalidad de animación del brazo, obteniendo distintas animaciones ante situaciones iguales, es decir, cuando se cuenta con el mismo estado inicial y se desea llegar al mismo estado final. En este trabajo

el objetivo primordial es que se puedan realizar movimientos semejantes a los de un brazo humano.

A continuación se definen formalmente los conceptos de estado, de acción, de meta y de espacio de configuraciones:

Un estado $e_k = (k, q_k, \vec{v}_k)$ representa la configuración actual para una k-ésima articulación del brazo, con una rotación representada por el cuaternión $q_k = (\theta_k, \hat{n}_k)$ (θ_k representa el ángulo de rotación y \hat{n}_k expresa un vector unitario que representa la dirección de la rotación) y un vector de traslación \vec{v}_k en R^3 que indica la posición actual de la articulación.

Una acción $a_k = (k, q)$ representa el movimiento a realizar sobre una k-ésima articulación, donde $q = (\theta, \hat{u})$ es el cuaternión que representa la rotación que se aplicará sobre la articulación (θ es el ángulo de rotación y \hat{u} representa el eje sobre el cual se rotará). Es posible definir una acción usando el formato PDDL (Planning Domain Definition Language) de la forma siguiente:

$a_k(k, q) =$
pre: estadoactual(e_k) \wedge *estadolibre*(e'_k)

pos: \neg estadoactual(e_k) \wedge *estadoactual*(e'_k) \wedge
 \neg *estadolibre*(e'_k)

El estado actual (*estadoactual*(e_k)) es un predicado para la configuración actual de una articulación y el estado libre (*estadolibre*(e'_k)) comprueba que no existan problemas con el estado alcanzable de la acción.

Una meta g para un plan generado por el algoritmo consiste en llegar a una posición final \vec{v}_{goal} , la cual representa un punto alcanzable por el brazo dentro de su espacio de configuraciones.

El espacio de configuraciones C representa todos los puntos en el espacio 3D que son alcanzables por cada configuración válida que puede adoptar el brazo, este espacio involucra al conjunto de articulaciones $J = \{j_1, j_2, \dots, j_i \mid j_i \text{ es una articulación del brazo}\}$ que componen el brazo. En este algoritmo el espacio de configuraciones C es

discretizado para establecer los grados de movimiento $D = \{n \in \mathbb{N} \mid n \bmod_{\Delta p} = 0 \wedge n \leq \theta_{max}\}$ que se pueden generar para una acción sobre una articulación k , donde θ_{max} indica el ángulo máximo de rotación para la articulación k , que para el caso de estudio analizado fue establecido con un incremento $\Delta p = 10$. Para dar una representación más entendible, \hat{u} se especificará dependiendo del movimiento que genere sobre alguno de los siguientes grados de libertad (DOF's): yaw, pitch y roll, formando así un conjunto de acciones para cada grado de libertad que posea cada articulación. Este conjunto de acciones estará definido por el grado de movimientos posibles que fueron adaptados de [15] a los 8 DOF's del avatar.

Articulación	Pitch	Roll	Yaw
Hombro	-90° a 90°	-50° a 180°	0° a 180°
Codo	-10 a 140°	---	---
Muñeca	-60° a 60°	-90° a 90°	-20° a 30°
Palma	0° a 90°	---	---

Tabla 1: Rango de movimiento para cada articulación del brazo del avatar

A continuación se explica el algoritmo de planeación propuesto. Este algoritmo es el siguiente:

```

motionPlanner( $\vec{v}_{goal}$ , J)
T.root( $e_{end}$ )
do
  for k  $\leftarrow$   $j_{ini}$  to k  $\leftarrow$   $j_{end}$  do ( $j_{end}$  es el efector final)
     $a_{rand} \leftarrow$  Rand_Action(k)
    AddAction (T,  $a_{rand}$ )
    ejecutarAccion (T)
while (gradientDescent( $e_{end}$ ) > threshold)
return (T)
    
```

Al aplicar el algoritmo se genera una estructura de árbol con la secuencia de movimientos que deben ejecutarse sobre el brazo para posicionar el efector en su objetivo final. Cabe mencionar que el algoritmo toma como entrada la posición final que debe tener el efector y el conjunto de articulaciones J del brazo, así empieza a trabajar a partir de la configuración actual que tenga el brazo.

Una vez que se establece la posición inicial del efector final como estado raíz del árbol, se generaran movimientos para cada una de las articulaciones involucradas y cada nueva acción generada será agregada a la estructura. Cuando ya se generaron nuevas acciones, estas se deberán ir ejecutando sobre el brazo, esto representa una diferencia con respecto al algoritmo original RRT, ya que aquí no se genera el plan de una forma fuera de línea, sino que este se va ejecutando conforme se van obteniendo las siguientes acciones.

Al ejecutar las acciones se determinará si es que el efector se encuentra lo suficientemente cerca de su objetivo, si no es así, se vuelven a generar nuevas acciones hasta que se logre obtener la posición deseada, dando como resultado un plan en una estructura de árbol que tendrá todas las acciones que se deben ejecutar.

A continuación se muestra como se obtiene la siguiente acción a ejecutar en base a la siguiente función:

```

Rand_Action(k)
If(distancia(k,k+1) ≥ distancia(k,  $\vec{v}_{goal}$ )
  Do
     $\hat{u} \leftarrow \text{randomDOF}(k)$ 
     $\theta \leftarrow \Delta p * \text{random}[1,4]$ 
     $q \leftarrow (\theta, \hat{u})$ 
  While (gradientDescent(q) >
gradientDescent(ek))
     $\theta_k \leftarrow \Delta p$ 
     $q_k \leftarrow (\theta_k, \hat{u})$ 
     $a_k \leftarrow (k, q_k)$ 
    return(ak)
else
  return (null)

```

La nueva acción a ejecutar sobre una articulación k se genera al azar y una vez verificado que se puede ejecutar esta se agrega al plan. Para hacer esto, primero se determina si es una acción válida para la articulación, verificando además, si esta no se encuentra cercana de la posición final, debido a que si está demasiado cerca, no podría llevarse el efector a su posición final. Para saber si la acción dará una mejor posición del efector final se utiliza

el método del gradiente descendente como métrica del espacio de configuraciones.

Para obtener el eje de rotación de la articulación k (DOF's), se emplea la siguiente función:

```

randomDOF(k)
switch(k)
  case codo, palma:
     $\hat{u} \leftarrow \text{random}(\text{pitch}, -\text{pitch})$ 
  case muñeca:
     $\hat{u} \leftarrow \text{random}(\text{roll}, -\text{roll}, \text{pitch}, -\text{pitch}, \text{yaw}, -\text{yaw})$ 
  default: //hombro
     $\hat{u} \leftarrow \text{random}(\text{roll}, -\text{roll}, \text{pitch}, -\text{pitch}, \text{yaw}, -\text{yaw})$ 
return ( $\hat{u}$ )

```

El eje de rotación se obtiene en base a los grados de libertad de la articulación k (yaw, pitch y roll). Cada eje de rotación es agrupado en alguno de estos grados de libertad. Para cada caso se escoge de forma aleatoria el eje de rotación de entre los grados de libertad validos para la articulación y este es regresado como resultado del algoritmo.

La Figura1 ilustra el funcionamiento de nuestro algoritmo.

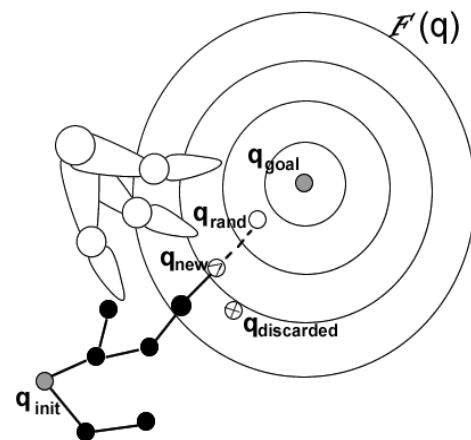


Figura1: Funcionamiento del algoritmo de planeación propuesto

IV. CASO DE ESTUDIO

Para la verificación de la funcionalidad del algoritmo propuesto en este trabajo, se utilizó el motor 3D jME [16] y el avatar que se muestra en la Figura 2. Sobre el brazo izquierdo de dicho avatar

se ejecutó el plan obtenido por el algoritmo, este brazo cuenta con ocho grados de libertad, tres para el hombro, uno para el codo, tres para la canilla o muñeca y uno para la palma (efector final).

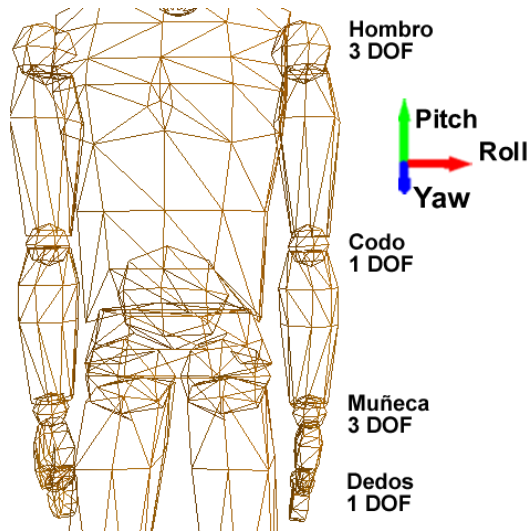


Figura2: DOF's de cada articulación del brazo del avatar

En la Figura 3 se muestra sólo una parte de la secuencia de configuraciones que se obtuvieron de la ejecución del algoritmo de planeación. Para una posición inicial del brazo del avatar (figura 3.A) se ejecutó una serie de movimientos (figura 3.B a figura 3.H) para lograr que la palma izquierda del avatar tocará la esfera que se encuentra en frente de él.

V. CONCLUSIONES

En este trabajo se presentó un algoritmo de planeación dinámica de movimiento para el brazo de un avatar. Este algoritmo hace uso de la información sobre las restricciones de movimiento de las diferentes articulaciones presentes en el brazo, además de que contempla el uso de cuaterniones para generar las rotaciones de las articulaciones y así obtener movimientos fluidos evitando problemas como la pérdida de grados de libertad (gimbal lock) que tienen otros métodos de rotación. El uso de la aleatoriedad en la determinación del siguiente estado evita la búsqueda exhaustiva sobre el espacio de configuración. Los resultados presentados

demuestran que el algoritmo funciona desde aplicaciones de tareas sencillas como la de alcanzar un objeto con la mano hasta tareas mucho más complejas como la de agarrado de objetos.

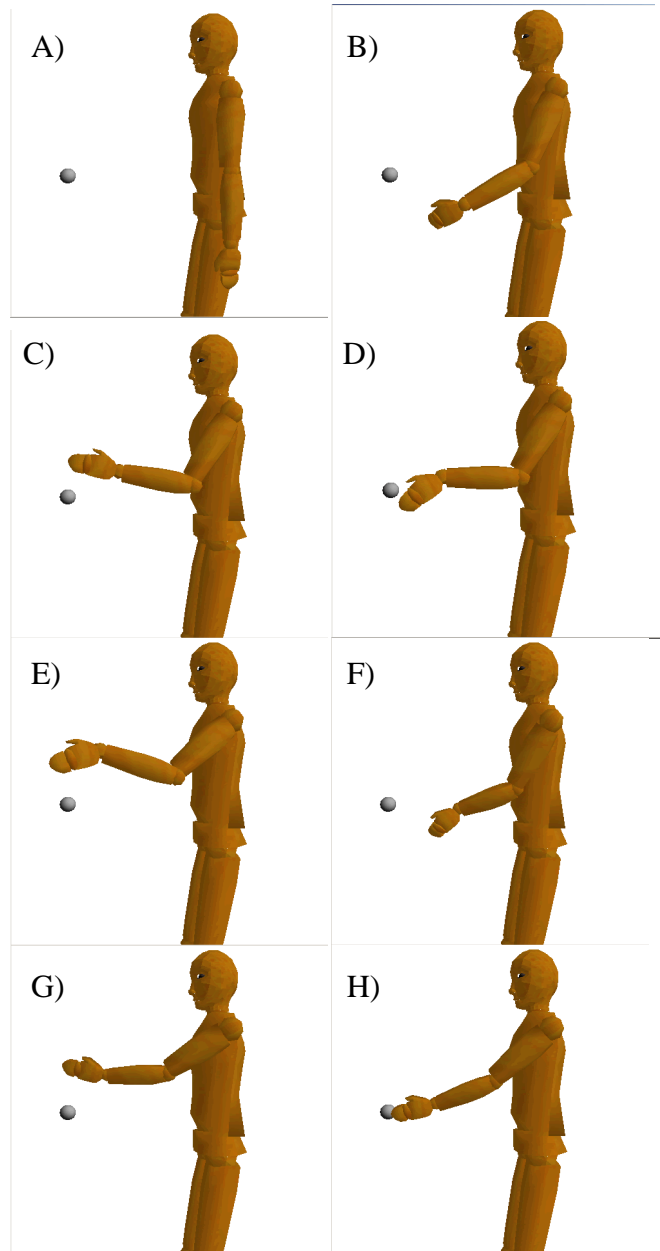


Figura3: Plan de movimiento obtenido

VI. AGRADECIMIENTOS

Este trabajo está parcialmente apoyado por el CoECyT-Jal proyecto No. 2008-05-97094, mientras que sus autores Cristian E. Boyain y Goytia Luna y Héctor Rafael Orozco Aguirre

cuentan con beca de CONACYT No. 212740 y 203060, respectivamente.

VII. REFERENCIAS

- [1] M. Uc, A. Rodríguez, F. Ramos, *Reinforcement Learning and Dynamic Planning Applied to Virtual Humans Animation*, International Conference on Electrical and Electronics Engineering September, 2007.
- [2] Michael Gerke, *GeneticPath Planning for Mobile Robot*, American Control Conference June, 1999.
- [3] Mohd Murtadha Mohamad, Nicholas K. Taylor and Matthew W. Dunningan, *Articulated Robot Motion Planning Using Ant Colony Optimisation*, 3rd International IEEE Conference Intelligent Systems September, 2006.
- [4] Beatriz A. Garro, Humberto Sossa and Roberto A. Vásquez, *Evolving ant colony system for optimizing path planning in mobile robots*, fourth Congress of Electronics, Robotics and Automotive Mechanics, September, 2007.
- [5] Eiichi Yoshida, *Humanoid Motion Planning using Multi-Level DOF Exploitation Based on Randomized Method*, Intelligent Robots and Systems, 2005.
- [6] Chenyu MA, Wei LI, Yang Yang, *Robot Motion Planning with Many Degrees of Freedom*, IEEE. 1995.
- [7] Dave Ferguson, Nidhi Kalra and Anthony Stentz, *Replanning with RRTs*, International Conference on Robotics and Automation, May, 2006.
- [8] Claudia Esteves, Gustavo Arechavaleta and Jean-Paul Laumond, *Motion Planning for Human-Robot Interaction in Manipulation Tasks*, International Conference on Mechatronics & Automation, July, 2005
- [9] Anthony Stentz and Dave Ferguson, *Anytime, Dynamic Planning in High-dimensional Search Spaces*, international conference on Robotics and Automation, April, 2007.
- [10] Jaesik Choi and Eyal Amir, *Factor-Guided Motion Planning for a Robot Arm*, IEEE/RSJ International Conference on intelligent Robots and Systems October, 2007.
- [11] Oliver Brock and Oussama Khatib, *Real-Time Replanning in High-Dimensional Configuration Space Using Sets on Homotopic Paths*, IEEE International Conference on Robotics & Automation, April, 2000.
- [12] Gustavo Arechavaleta, Claudia Esteves and Jean-Paul Laumond, *Planning Fine Motions for a Digital Factotum*, IEEE Intelligent Robots and Systems, September, 2004.
- [13] Hwan-Joo Kwak and Gwi-Tae Park, *Motion Separating Based Whole Body Motion Planning For Humanoid Robots Using a Gradient Descent Method*, IEEE/ASME International Conference on Advanced Intelligent Mechatronics, July, 2008.
- [14] S. LaValle and J. Kuffner, *Rapidly-Exploring Random Tree: Progress and Prospects*, Algorithmic and Computational Robotics: New Directions 293-308, A. K Peters, Wellesley, 2001
- [15] Kathryn Luttgens and Nancy Hamilton, *Kinesiology: Scientific basis of human motion*, Brown & Benchmark, 1997.
- [16] Mark Powell, *jMonkey Engine*, <http://www.jmonkeyengine.com>, 2003.

VIII. AUTORES

Ing. Cristian E. Boyain y Goytia Luna es Ingeniero en Computación por la Universidad Autónoma de Zacatecas, en el 2007 ingresó como alumno de la Maestría en Ciencias en Ingeniería Eléctrica en el CINVESTAV, unidad Guadalajara.

M. en C. Héctor Rafael Orozco Aguirre es Maestro en Ciencias en Ingeniería Eléctrica con especialidad en computación por el CINVESTAV, unidad Guadalajara, desde el 2005 es profesor de asignatura del Centro Universitario de Ciencias Exactas e Ingenierías de la Universidad de Guadalajara y en el 2006 ingresó a estudiar el Doctorado en Ciencias en Ingeniería Eléctrica en el CINVESTAV, unidad Guadalajara.

Dr. Félix Francisco Ramos Corchado obtuvo su título de Doctor en Computación de la Universidad de Tecnología de Compiègne Francia, tiene 15 años de experiencia en la investigación, la docencia y el desarrollo tecnológico.

SISTEMA DE SELECCIÓN DE HORARIOS PARA PROFESORES SiSHoP

Nolasco Salcedo Ma. Del Carmen, García García Oscar Galileo, Roberts Samuel Mclean

Área de Sistemas de Información, Departamento de Ciencias Tecnológicas, División de Desarrollo Biotecnológico
Universidad de Guadalajara, Centro Universitario de la Ciénege
Av. Universidad 1115, Col. Linda Vista, Ocotlán, Jalisco
Correo electrónico: mcns08@gmail.com , oscargalileo@gmail.com , panmansam@gmail.com

RESUMEN

El sistema SiSHoP propone una nueva metodología para la asignación de espacios y horarios de las clases que se imparten en el Centro Universitario de la Ciénege.

Este nuevo enfoque, propone una administración más eficiente por niveles, reflejando una parte de la estructura jerárquica del Centro, desde los Directores de División, seguido de los Jefes de Departamento, luego los Coordinadores de Carrera y finalmente los más beneficiados con el sistema: los profesores.

El objetivo principal es otorgarle a cada usuario las actividades que por reglamento les concierne, facilitando los procedimientos de captura y análisis de la información.

El principal fuerte del sistema, es la sencilla interfaz para la selección de los horarios por parte de los profesores, respetando así la decisión de aquellos docentes con mayor categoría y antigüedad.

Palabras Claves: Sistema de Información, Sistema de Administración, Sistema de Selección de Horarios.

I. INTRODUCCIÓN

Las Tecnologías de la Información y Comunicación permiten la utilización de masas de información pero no sólo basta la disponibilidad de la cantidad de información, se trata de contar con información de calidad [1].

Los sistemas de información administrativos no sólo son capaces de suministrar información de calidad y oportuna, sino que también pueden respaldar la toma de decisiones.

Las aplicaciones hipermedia han evolucionado en los últimos años y se han concentrado mayormente en la *web*. Son aplicaciones dinámicas, de constante actualización e incluso personalizables, capaces de adaptarse a los tipos de usuarios y en casos avanzados, a cada usuario en particular.

Estas características encuentran el medio ideal en la *web*, ya que de otra forma sería costoso su mantenimiento y evolución.



Figura 1. Módulos SiSHoP (Esquema General)

Las tecnologías de la información y de la comunicación (TIC) se han asentado plenamente en las universidades a nivel mundial a lo largo de la última década. Sin duda, esto está sufriendo un cambio significativo en la dinámica interna de las instituciones: En la administración académica y en los servicios a los estudiantes y profesores y, de forma relevante, en la actividad de investigación y difusión [3].

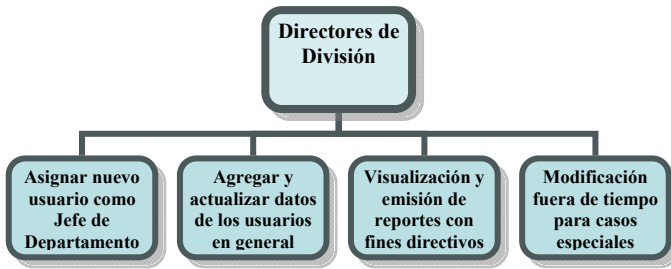


Figura 2. Funciones del Director de División

Es necesario reconocer que las tecnologías de la información y comunicación no son un fenómeno aislado, no se producen en el vacío sino que están insertas en una dinámica socio-cultural que es la sociedad global o la sociedad red.

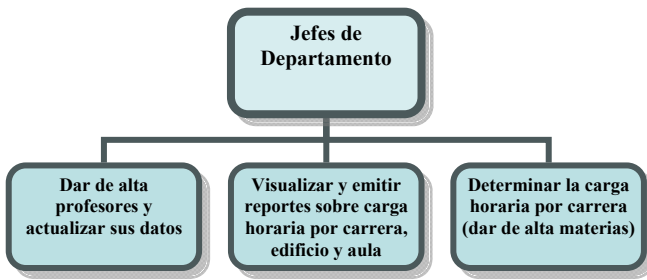


Figura 3. Funciones del Jefe de Departamento

Cada vez es más usual que algunos medios y recursos tecnológicos, sobre todo los informáticos vayan aportando sus posibilidades a la organización y administración de los centros universitarios. Caso típico es el de la informática, que cada vez es más usual que se utilice para diferentes tipos de gestiones: académica y administrativa de los centros en cuanto a las tareas de admisión de alumnos y control de las matrículas, académica de los estudiantes en cuanto al curso que se encuentran cursando y los exámenes que van superando, o simplemente económica.

La que hemos convenido en denominar sociedad de la información y el conocimiento ahonda sus cimientos tanto en las transformaciones tecnológicas a las que acabamos de hacer referencia como en la sistematización de los procesos de generación y transferencia [2]

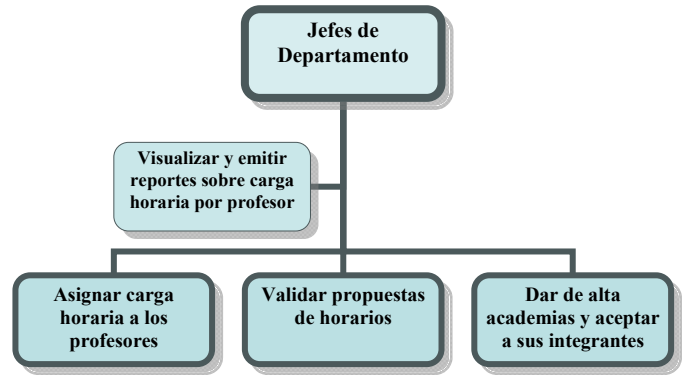


Figura 4. Funciones del Coordinador de Carrera

Los avances tecnológicos propician el desarrollo de nuevas pautas de estructuración y análisis de la información que facilitan su comprensión y potencial utilidad en la resolución de problemas y en la toma de decisiones [2].



Figura 5. Funciones del Profesor

En la actualidad las instituciones universitarias de punta se caracterizan por disponer y utilizar de manera eficiente distintas herramientas para gestionar en forma eficaz la información, como premisa para la adecuada gestión de los diversos procesos, de modo que les permita ir a la par con el avance tecnológico.

SiSHoP es un sistema cliente/servidor que permite una administración descentralizada del sistema en un ambiente de red, donde todos los usuarios participarán desde la comodidad de su computadora, retroalimentando con sus decisiones dentro del cargo, logrando así una gestión dinámica.

Actualmente, el proceso de asignación de espacios y horarios experimenta diversos problemas, disminuyendo su efectividad y fallando en casi todos los puntos que interviene. Cada semestre los espacios parecen insuficientes, los reglamentos casi nunca se respetan, los tiempos para correcciones a veces no existen y casi nunca se solicita la opinión de los profesores.

Es por eso que en el presente trabajo, describiremos el funcionamiento de SiSHoP, un sistema que pretende solventar los problemas actuales acerca de la asignación de espacios y horarios, descentralizando la tarea y ofreciendo una interfaz amigable e intuitiva, haciendo de estos procedimientos una tarea fácil y distribuida.

Además de que las ventajas que le aportará al Centro Universitario de la Ciénega el Sistema de Selección de Horarios para Profesores (SiSHoP) son:

1. Mejorar la calidad del servicio de atención a los profesores.
2. Mayor agilidad en la distribución/rapidez.
3. Ahorro de gestión/papeleo.

4. Mayor agilidad en la gestión.
5. Comodidad.
6. Mejor imagen de la institución.
7. Mayores posibilidades de innovación, de renovación de la gama de procesos.

II. DESARROLLO

SiSHoP es una RIA (Rich Internet Application: Aplicación de Internet Enriquecida) capaz de ofrecer una interfaz amigable e intuitiva, utiliza el servidor web Apache [4], PHP como lenguaje de servidor [5], la base de datos MySQL [6] y Javascript del lado del cliente [7], para realizar peticiones asíncronas [8] en todas sus funciones.

Existe un nivel de seguridad avanzado sin utilizar SSL, por medio de los algoritmos criptográficos MD5 y SHA1 programados en Javascript [9], la idea es desarrollar un sistema rápido, flexible, intuitivo y seguro.

Reduce el trabajo que hasta hoy realizan en conjunto los Jefes de Departamento, Coordinadores de Carrera y algunos Profesores.



[Ayuda](#)

Horas / Dias	Aula 0001	Aula 0002	Aula 0003	Aula 0004	Aula 0005	Aula 0006	Aula 0007	Aula 0008	Aula 0009	Aula 0010	Aula 0011	Aula 0012
07:00 a 08:00 hrs	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre
08:00 a 09:00 hrs	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre
09:00 a 10:00 hrs	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre
10:00 a 11:00 hrs	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input checked="" type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre
11:00 a 12:00 hrs	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input checked="" type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre
12:00 a 13:00 hrs	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre
13:00 a 14:00 hrs	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre
14:00 a 15:00 hrs	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre
15:00 a 16:00 hrs	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre
16:00 a 17:00 hrs	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre
17:00 a 18:00 hrs	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre
18:00 a 19:00 hrs	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre
19:00 a 20:00 hrs	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre
20:00 a 21:00 hrs	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre	<input type="checkbox"/> Hora Libre

Figura 6. Ejemplo del horario gráfico y dinámico

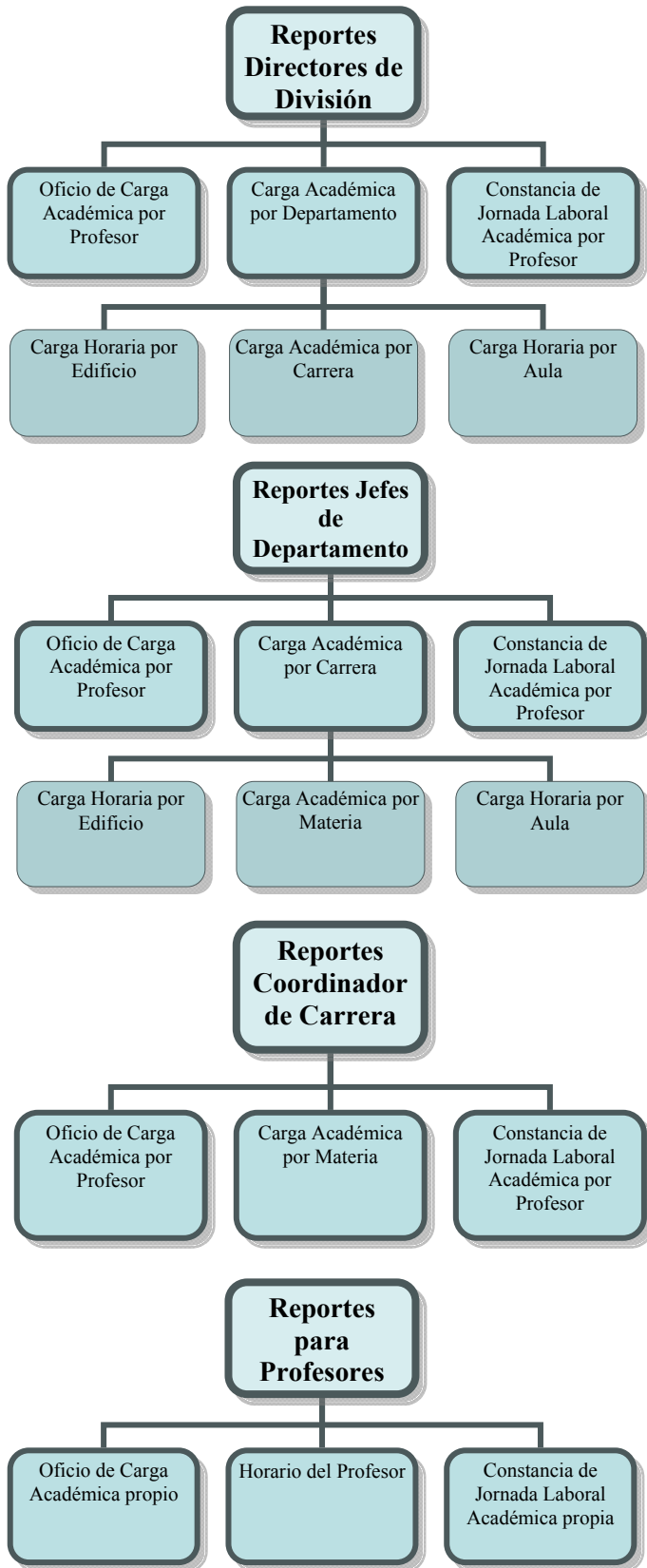


Figura 7. Reportes por Tipo de Usuario

La base de datos básicamente describe una parte de la estructura organizacional del Centro Universitario de la Ciénega (CUCI), se utilizaron únicamente estándares (XHTML, XML, JSON, CSS, Javascript, SQL) para garantizar la compatibilidad entre sistemas y accesibilidad, además aporta una serie de "servicios web" (SOAP, REST) que alimentan a otros proyectos, debido a esto puede ser considerado un sistema distribuido.

Los módulos que componen el Sistema de Selección de Horarios para Profesores son:

1. **Directores de División:** Este módulo será utilizado para la administración principal del sistema, da de alta a los jefes de departamento, visualiza reportes cuando se requiere y puede hacer modificaciones fuera de tiempos para casos especiales.
2. **Jefes de Departamento:** Por medio de éste módulo se darán de alta nuevos profesores y actualizarán datos de los existentes, se determina la carga horaria por carrera, entre otras funciones que se podrán agregar para extender las posibilidades del sistema.
3. **Coordinadores de Carrera.** Este módulo ha sido desarrollado para que el coordinador asigne materias a cada profesor, además validará la propuesta de horario de los profesores y también dará de alta a las academias, comités consultivos, de titulación, técnicos; y aceptará a sus integrantes (actividades colegiadas).
4. **Profesores.** Este módulo le brindará a los usuarios, en su carácter de profesor, la posibilidad de seleccionar él mismo, el aula y horario en que impartirá sus materias, donde la decisión final estará supeditada por el Coordinador de carrera, además de respetar las decisiones de aquellos docentes con mayor antigüedad y categoría en su nombramiento. Una función extra será la selección de actividades colegiadas, donde

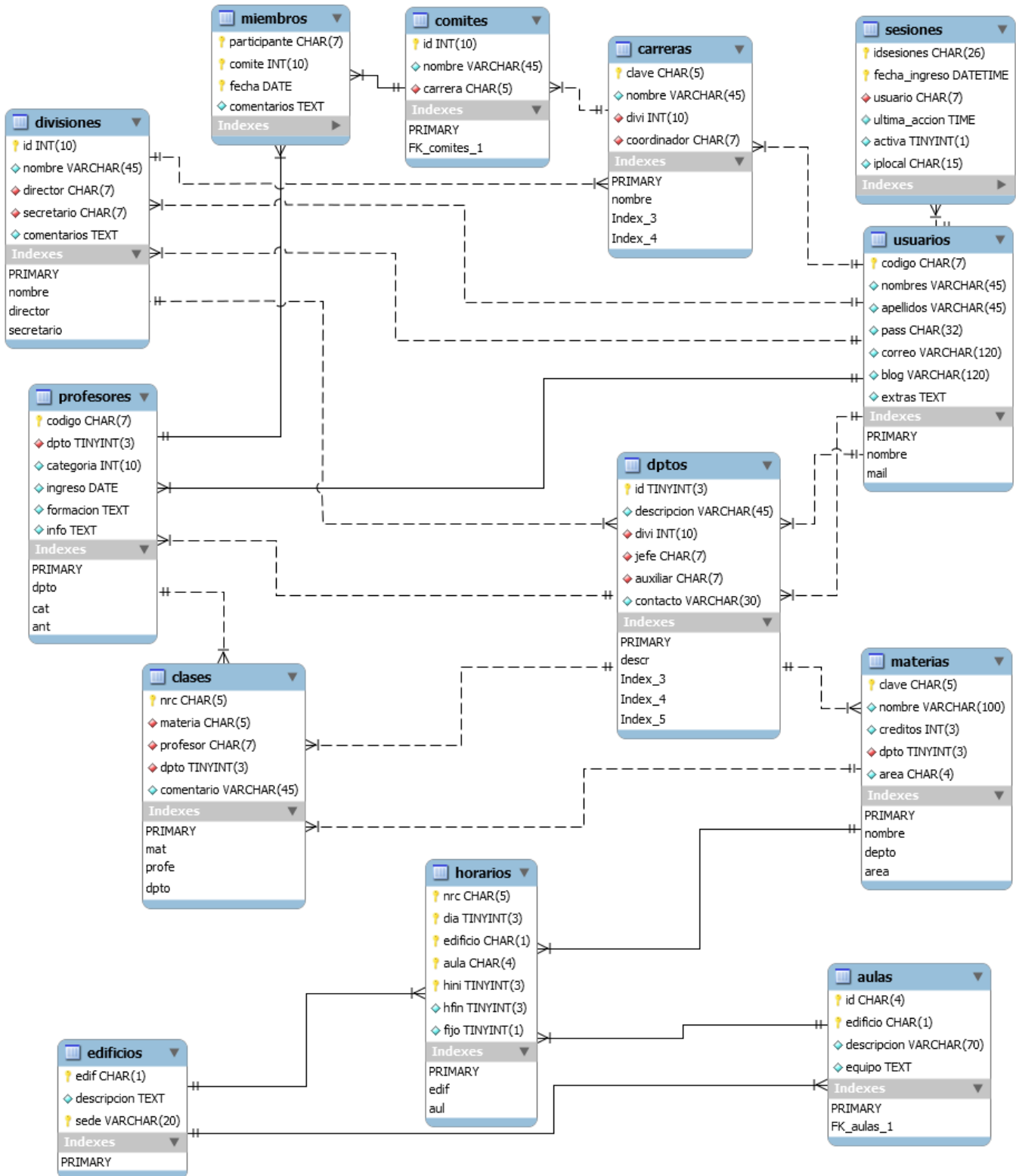


Figura 8. Diagrama UML de la Base de Datos

cada profesor podrá solicitar su ingreso a alguna de las academias, así mismo el

Coordinador de carrera validará a los miembros definitivos.

Cabe mencionar que la selección de horarios y aulas se realizará por medio de un horario gráfico y dinámico, donde el profesor visualizará los espacios disponibles y no disponibles, una vez que el profesor haya seleccionado su horario, éste será validado por el coordinador de carrera.



Figura 9. SiSHoP (página de inicio)

Finalmente, la última fase de SiSHoP es la impresión de diversos reportes, los cuales podrán imprimirse según el tipo de usuario que ingrese al sistema, en la figura 7 desglosaremos cada uno.

Los reportes más importantes son el Oficio de carga Académica que se le da a cada profesor al inicio del semestre, así como su horario y la Constancia de Jornada Laboral Académica.

En la figura 8 podemos apreciar el diagrama UML de la base de datos, que describe las entidades y relaciones programadas para este proyecto.

Para una mejor administración, se restringen el acceso a las diferentes áreas del sistema según el tipo de usuario; de acuerdo a la definición de las funciones descritas anteriormente.

Un ejemplo de la potencia de SiSHoP es la sencilla página de inicio (figura 9), donde los usuarios simplemente ingresan su código de trabajador y contraseña, el sistema automáticamente determinará los permisos correspondientes,

permitiendo que un usuario pueda tomar varios papeles (la mayoría de los directivos también son profesores), sin perjudicar la lógica de las funciones para cada rol y garantizando una gestión equilibrada y justa.

III. CONCLUSIONES

En la actualidad existen una serie de herramientas o sistemas de información que facilitan la toma de decisiones, el avance tecnológico nos ha conducido a desarrollar herramientas más inteligentes, precisas, eficaces y eficientes, que al formar parte del entorno contribuyen de manera satisfactoria a mejorar y hacer más productivos los diferentes procesos que se llevan a cabo en un determinado momento.

Es de resaltar la importancia de SiSHoP dentro del procedimiento de asignación de espacios, implicaría una mejor administración y distribución de las aulas y los horarios de clases, además proporcionaría un ambiente de colaboración más libre y equilibrada.

El desarrollo de aplicaciones *web* involucra decisiones no triviales de diseño e implementación que inevitablemente influyen en todo el proceso de desarrollo.

SiSHoP, aún no se encuentra en operación, sin embargo, puede favorecer a una atención oportuna para los profesores y de mejor calidad. Además de que el alcance de la aplicación y el tipo de usuarios a los que está dirigida son consideraciones tan importantes como las tecnologías elegidas para desarrollarlo.

Es por ello que el uso de una metodología de diseño y de tecnologías que se adapten naturalmente a ésta, son de vital importancia para el desarrollo de aplicaciones complejas.

SiSHoP marcará el inicio de la era 2.0 en Software para la Universidad de Guadalajara, siendo el primero de su tipo y seguido de muchos más, que sólo pretenden incrementar la calidad de los

servicios que ya ofrece nuestra máxima casa de estudios.

La Sociedad de la Información, la tecnología informática y la transdisciplina, en estos ámbitos, se han de complementar, para posibilitar una efectiva transformación de las instituciones hacia la calidad.

No se trata de cambiar por cambiar, sino por el hecho de hacer de la institución, un Centro universitario de mayor calidad caracterizado por la innovación y la renovación permanente de su administración.

En este sentido, este es el desafío: ser capaces de vislumbrar el horizonte antes que llegemos a él.

IV. REFERENCIAS

- [1] Sáez Baeza, Chiara (2009), *Invisibilización de la comunicación alternativa: propuestas de entrada y salida*. RLCS, Revista Latina de Comunicación Social, 64, páginas 416 a 423. La Laguna (Tenerife): Universidad de La Laguna.
- [2] Domingo Docampo (2003), *La Transformación de la Universidad*, Talleres de Formación de la AQU Catalunya, <http://www.gts.tsc.uvigo.es/~ddocampo/online/grial.pdf>
- [3] Duart, j. M.; Lupiáñez, F. (2005). *Las TIC en la universidad: estrategia y transformación institucional*. RUSC. Revista de Universidad y Sociedad del Conocimiento. issn: 1698-580X.
- [4] Servidor HTTP Apache <http://httpd.apache.org/>
- [5] Scripting del lado del servidor <http://www.php.net/>
- [6] Base de datos más popular del mundo <http://www.mysql.com/>
- [7] Potente librería Javascript <http://jquery.com/>
- [8] AJAX, una RIA y estándares usados <http://es.wikipedia.org/wiki/AJAX>
- [9] Seguridad por medio de Criptografía <http://pajhome.org.uk/crypt/index.html>

V. AUTORES

M.Sc. Ma. Del Carmen Nolasco Salcedo obtuvo su título de Maestría en Ciencias de la Computación con especialidad en Programación, Por la Universidad Martha Abreu de las Villas, Cuba, 2002

Ing. en Computación Oscar Galileo García García es egresado de la carrera de Ingeniería en Computación de la Universidad de Guadalajara, Centro Universitario de la Ciénega. Desarrollador del Sistema de Reservación de Aulas para el Centro de Tecnologías para el aprendizaje, Sistema de Evaluación Docente y Centro de Información Estadística para la Planeación, los cuales se utilizan en el Centro Universitario de la Ciénega.

Roberts Samuel Mclean es estudiante de séptimo semestre de la carrera de Lic. en Informática de la Universidad de Guadalajara, Centro Universitario de la Ciénega.

CONDUCTA INFLUENCIADA POR LA PERSONALIDAD EN AVATARES ANTE SITUACIONES DE DESASTRE: UNA PROPUESTA APOYADA EN LA PSICOLOGÍA

Víctor Fernández Cervantes, Luis Real Novo, Héctor Rafael Orozco Aguirre, Félix Ramos Corchado, Marco Ramos Corchado, Verónica Pérez.

Centro de Investigación y de Estudios Avanzados del I.P.N, Unidad Guadalajara
Av. Científica 1145, Col. El Bajío 45015 Zapopan, Jal., México

Correo electrónico: {vfernand, lreal, horozco, framos}@gdl.cinvestav.mx, mramos@univ-tlse1.fr

RESUMEN

En este trabajo se expone una metodología con fundamento en la psicología y en la teoría de patrones para la generación de conductas realistas de criaturas virtuales (avatares) en situaciones de desastre modeladas en ambientes virtuales 3D. Para poder generar estas conductas se obtiene un perfil de personalidad aplicando la evaluación psicológica del Inventario Multifásico de Personalidad de Minnesota (o MMPI, por sus siglas en inglés: Minnesota Multiphasic Personality Inventory). El resultado de la evaluación da una serie de valores para escalas de personalidad que son asignadas a los avatares. Para la generación de las conductas se aplica un modelo basado en la teoría de patrones, de esta manera se pueden generar comportamientos realistas influenciados por la personalidad.

Palabras clave: MMPI, Personalidad, Conducta influenciada, Avatares.

I. INTRODUCCIÓN

La personalidad es a menudo definida como el conjunto de rasgos psicológicos y mecanismos dentro del individuo que están organizados y son relativamente estables en el tiempo [1]. Estos rasgos influyen de manera directa la conducta del ser humano y sus interacciones y adaptaciones hacia su ambiente intrapsíquico,

social y físico. Los rasgos de personalidad son formas de patrones persistentes que están ligados a cómo el individuo percibe, piensa y se relaciona con el mundo en que vive y consigo mismo. Estas características son usadas para distinguir a una persona de otra y se ven por tanto reflejadas en su comportamiento en una amplia gama de contextos tales como el social y el personal.

De acuerdo a la investigación hecha por el médico canadiense J. S. Tyhurst [2], ante una situación de desastre aproximadamente del 12% al 25% de las víctimas se encuentran tensas y excitadas, aunque pueden actuar eficazmente, y a menudo pueden estar demasiado ocupadas como para preocuparse por actuar. Por otro lado, el 75% de ellas estarían pasmadas, asombradas y abrumadas. Las reacciones anteriores generan lo que se conoce como "síndrome de desastre" [3], el cual se caracteriza por la falta de emociones, inhibición de la actividad, docilidad, indecisión, falta de sensibilidad, conducta automática y manifestaciones fisiológicas de temor. El porcentaje restante presentaría un comportamiento totalmente inapropiado, donde se verían reacciones de ansiedad, reacciones histéricas y psicosis. Es importante destacar que estas últimas reacciones son raras, a pesar de que la creencia popular, y los medios de comunicación indican lo contrario [4]. Con la

investigación de Tyhurst se demuestra que las personas reaccionan de manera diferente ante una misma situación de desastre, lo cual pone de manifiesto que desde el punto de vista psicológico, los rasgos de personalidad son los que definen las tendencias a actuar y la conducta de cada persona.

Desde su aparición el Inventario Multifásico de Personalidad de Minnesota (MMPI) ha demostrado ser una buena herramienta para obtener el perfil de personalidad del ser humano. Este inventario de personalidad es el más empleado en la práctica psicométrica, para las áreas de la psicología clínica, educativa y laboral; por ser el que cuenta con mayor grado de confiabilidad. El MMPI evalúa por tanto los principales rasgos de personalidad y trastornos emocionales del individuo.

En éste trabajo nos apoyamos en base a los resultados de muchos años de investigación en México sobre las escalas de personalidad de los mexicanos y sus patrones de interrelación [5]. Con estos resultados se puede realizar una mejor interpretación de las evaluaciones dadas por el perfil de personalidad y generar conductas más realistas en los avatares. Mediante la interpretación de las escalas de personalidad obtenemos tendencias de comportamiento, las cuales utilizamos como reglas que definen la conducta de los avatares aplicando la teoría de patrones. De esta manera es posible hacer una simulación de la realidad ante situaciones de desastre y generar comportamientos creíbles para los avatares dentro de los ambientes virtuales 3D que son modelados.

Este trabajo está organizado de la siguiente manera: en la siguiente sección se da una revisión del trabajo relacionado sobre la simulación de comportamiento de criaturas

virtuales y como se ha aplicado la personalidad en los mismos. En la tercera sección, se da una breve descripción del MMPI, en la cuarta sección se presenta nuestra propuesta basada en psicología para la generación de conductas y comportamientos de avatares ante una situación de desastre aplicando la teoría de patrones [9], en esta sección se muestra una serie de animaciones obtenidas en la simulación virtual. Finalmente, en la última sección se dan las conclusiones obtenidas.

II. TRABAJO RELACIONADO

En los trabajos que se han revisado hasta el momento, no se ha encontrado indicio alguno de que se aplique la personalidad como tal desde el punto de vista psicológico, con lo cual el comportamiento simulado es altamente cuestionable, ya que no se apega a una simulación real de la conducta del ser humano ante una situación de desastre, debido principalmente a que en la mayoría de ellos se crean comportamientos predefinidos. Con estos comportamientos se establecen reacciones dependiendo de las metas individuales de la criatura virtual (avatar), sin embargo no hay algo que respalde esto desde el punto de vista psicológico, ya que el comportamiento es predefinido por la persona que realiza la simulación.

En [6], se presenta una simulación desastre para multitudes, donde algunos avatares huyen más rápidamente que otros, mientras otros son contagiados poco a poco al darse cuenta de lo que ocurre en el ambiente. Aunque la simulación expuesta en dicho trabajo parece ser realista, no lo puede ser en sí, ya que no sucede en cuanto a tiempo y forma, con lo cual no se apega a la realidad, principalmente porque el comportamiento expuesto es predefinido y no se

obtiene en base a la personalidad de los avatares.

Otra forma de exponer comportamientos fue presentada en [7]. En este trabajo se presenta un caso de estudio para una cafetería donde hay dos tipos de meseros y tres clases de clientes. Para los meseros existe uno que es sociable y conversador y otro que es apático y reservado, en el caso de los clientes cada uno de ellos de manera individual busca relacionarse con otros que sean de su misma clase o bien tratar de conversar con los meseros siempre y cuando estos estén disponibles. Aunque de cierta manera se utiliza un concepto de personalidad, esta no es modelada correctamente ya que no se aplica un estudio de los rasgos de conducta en las relaciones sociales, sino que se trata a la misma como un conjunto predefinido de preferencias hacía el tipo de entidad con la cual se desea interactuar.

Por otra parte, en los trabajos donde se exponen simulaciones de multitudes [8], es impresionante ver la cantidad de avatares que son capaces de manejar, sin embargo, no se aplica la personalidad en la toma de decisiones, ya por ejemplo, si se simulará una salida de emergencia para un edificio en llamas, sería imposible percibir actitudes reales cuando todos los avatares se comportan de manera correcta [6], esto debido a que la personalidad implica un factor muy importante cuando la conducta puede ser violenta y el instinto obliga a querer salir antes que otros.

Finalmente, una de las áreas donde actualmente se intenta modelar el comportamiento del ser humano de una manera rápida, vistosa y poco costosa, es el área de los videojuegos, donde

todo está centrado en involucrar a los usuarios en la historia del juego. Así los personajes buenos contienen grandes virtudes y los malos todo lo contrario, pero es poco lo que coincide la simulación con la realidad, porque la personalidad va mucho más allá de modelar buenas o malas conductas.

III. MMPI

El MMPI cuenta con tres adaptaciones, una de las cuales fue realizada en México por la UNAM [5]. Éste instrumento consta de un cuestionario de 567 preguntas a las cuales se debe responder de manera afirmativa o negativa. Algunas de estas preguntas fueron cambiadas para ser adaptadas a la cultura mexicana. Las diez escalas clínicas de personalidad que son evaluadas con el MMPI son las siguientes: hipocondriasis (1.-Hs), depresión (2.-D), histeria (3.-Hi), desviación psicopática (4.-Dp), masculinidad-feminidad (5.-Mf), paranoia (6.-Pa), psicastenia (7.-P), esquizofrenia (8.-E), hipomanía (9.-Ma) y finalmente extroversión social (0.-ES). Cada una de estas escalas tiene 5 niveles donde el tercer nivel es considerado como un valor normal en la escala, y un nivel inferior o superior a éste indica que existe un problema en la conducta de la persona. Por consiguiente, los extremos inferiores y superiores se consideran como una psicopatología (la persona requiere de tratamiento psicológico). Sin embargo, es de estos niveles de donde se obtiene mayor información acerca del posible comportamiento que dan estos rasgos de personalidad.

IV. PROPUESTA

Nuestra propuesta se compone de los siguientes seis pasos (ver figura 1). El primero de ellos es la aplicación del MMPI. El segundo es el

análisis e interpretación de los resultados para obtener los valores de las escalas clínicas, el tercero consta de la obtención de la personalidad en base a la escala clínica predominante, es decir la de mayor valor, de esta escala es de donde se obtiene la información de las posibles conductas que se pueden presentar ante determinadas situaciones, así es posible dotar al avatar de una personalidad real que fue obtenida de la evaluación psicológica de un individuo. En el cuarto paso se plantea una situación en un ambiente virtual en específico donde serán expuestas las conductas acorde a la interpretación de los rasgos de personalidad y de los eventos que el avatar perciba del ambiente, en el quinto paso se generan las reacciones que deben ser realizadas por el avatar, para ello se aplican patrones de comportamiento en base a acciones que serán ejecutadas en el último paso ante una situación de desastre.

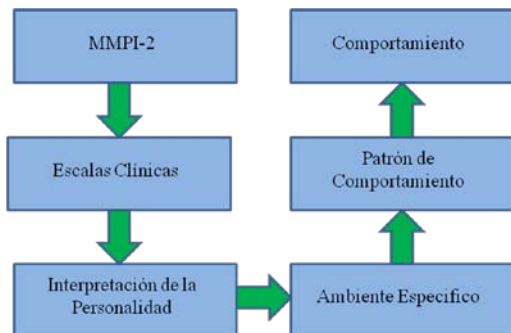


Figura 1. Pasos para la generación de conductas y comportamientos de avatares ante una situación de desastre aplicando la teoría de patrones

Interpretación de la personalidad

En este modulo después de calificar el test obtenemos un código para valor de cada una de las escalas de personalidad, éste código fue definido dependiendo el nivel donde caiga el valor de la escala, de la siguiente manera: * para

el nivel muy alto, + para el nivel alto, = para el nivel medio, - para el nivel bajo y / para el nivel muy bajo. Una vez hecho esto se hace un ordenamiento descendente de las escalas en base a su valor y al código asignado para obtener la escala más predominante. Por ejemplo, si se obtiene la siguiente cadena ==-+*=*/== y se usa el valor de la escala y el código para el ordenamiento se obtiene lo siguiente 7.-P, 5.-Mf, 1.-Hs, 2.-D, 4.-Dp, 6.-Pa, 9.-Ma, 0.-ES 3.-Hi, 8.-E. De manera queda ordenada la personalidad de forma descendente (de la escala más predominante a la menos predominante).

Ambiente específico.

En nuestra implementación, el ambiente del sistema representa todas las posibles acciones que pueden ser llevadas a cabo por los avatares. En nuestro caso la situación de desastre que se eligió modelar fue el caso de un temblor. En el ambiente generado las posibles acciones a ser llevadas a cabo por los avatares son (ver tabla 1): paralizarse si su hipocondriasis es muy alta, correr si su psicastenia es media, alta o muy alta, gritar si su histeria es alta, huir si su depresión es muy alta y ayudar si su extroversión social es media.

ε	Evento	Rasgos	Reacción
01	Temblor	Hs*	Paralizarse
02	Temblor	P=, P+, P*	Correr
03	Temblor	Hs+	Gritar
04	Temblor	D*	Huir
05	Temblor	ES=	Ayudar

Tabla 1. Acciones a ejecutar por los avatares ante una situación de sismo en base a su personalidad

Patrones de comportamiento

Para definir los patrones de comportamiento utilizaremos los generadores provistos por la teoría de patrones [9]. Aquí la unidad empleada es el generador de patrones (ver figura 2, parte izquierda) que se puede conectar a otros por medio de sus lazos para formar un grafo de comportamiento (ver figura 2, parte derecha). Esta conexión está dada por la función ρ , la cual indica si el valor β_i de un lazo puede ser conectado con el valor β_j de otro.

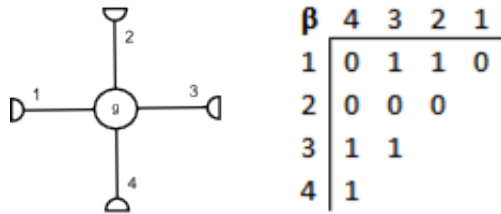


Figura 2. Generador de patrones (parte izquierda) y su conectividad (parte derecha)

De esta forma podemos utilizar un mismo generador tantas veces como sea necesario. El grafo obtenido de la conexión de los generadores de patrones, se llama grafo de configuración. Este grafo determina el comportamiento del avatar ante un evento determinado.

En este modelo las configuraciones están ligadas a los comportamientos, de manera que mediante el análisis del grafo resultante entre los generadores disponibles se puede conocer la acción que llevará a cabo el avatar.

El conjunto de generadores G que describe los comportamientos del avatar está compuesto por cuatro subconjuntos de cinco elementos (ver ecuaciones 1 a 5). Cada subconjunto describe una escala de personalidad, mientras que los

generadores indican el nivel de valor de esa escala.

$$G = \{G_1 \cup G_2 \cup G_3 \cup G_4\} \tag{1}$$

$$G_1 = \{Hs-1, Hs-2, Hs-3, Hs-4, Hs-5\} \tag{2}$$

$$G_2 = \{D-1, D-2, D-3, D-4, D-5\} \tag{3}$$

$$G_3 = \{P-1, P-2, P-3, P-4, P-5\} \tag{4}$$

$$G_4 = \{ES-1, ES-2, ES-3, ES-4, ES-5\} \tag{5}$$

En la tabla 2 podemos ver los subconjuntos y los escalas de personalidad a los que están ligadas. Así la personalidad está dada por un elemento de cada uno de estos subconjuntos (ver figura 3). Por lo que siempre tendremos cuatro generadores para formar el grafo de comportamiento que define la acción que realizará el avatar ante una situación en particular.

Rasgo	Subconjunto	Abreviatura
Hipocondriasis	1	Hs
Depresión	2	D
Psicastenia	3	P
Extroversión Social	4	ES

Tabla 2. Relación entre los subconjuntos de generadores y las escalas de personalidad

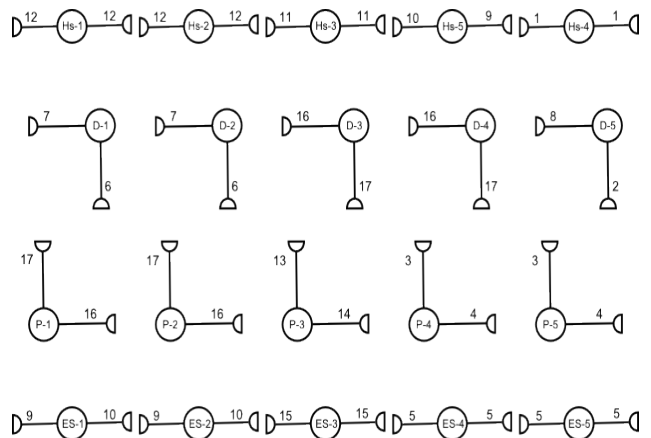


Figura 3. Conjunto de generadores para definir el patrón de comportamiento

Para formar el grafo σ de comportamiento se aplica una función de conexión ρ (ver figura 4), la cual permite conectar dos o más generadores. Así, si el valor de los lazos corresponde a 1, el valor 0 indica la ausencia de conexión.

β	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
5	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
10	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figura 4. Función de conectividad ρ

Generación de las acciones aplicando los patrones de comportamiento

Para este trabajo se consideraron cinco acciones posibles (ver tabla 1) para un avatar ante una situación de sismo. Estas acciones están definidas por configuraciones del modelo propuesto. Estas configuraciones tienen grafos bien definidos y la orientación que tiene cada uno de ellos es despreciable en el presente modelo, pero el lugar que toman los generadores es determinante en algunos casos, mientras que otros pueden tomar un lugar cualquiera sin afectar el resultado.

A continuación se explican las conexiones disponibles y los generadores determinantes en cada una de ellas:

- Inicialmente se tiene la acción paralizarse (ver figuras 5, 7 y 12), en este caso solo hay

un generador determinante, el asociado a la hipocondriasis de nivel muy alto. La posición del generador Hs siempre es la misma, en el extremo, este generador es quien inicia la configuración.

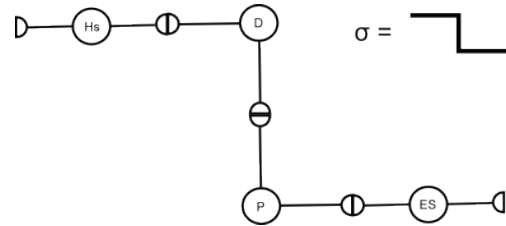


Figura 5. Configuración para la acción paralizarse

- La segunda acción es correr (ver figuras 6 y 7). Esta acción tiene dos generadores determinantes, depresión en niveles bajos combinada con niveles altos de psicastenia, en esta configuración la conexión de estos dos generadores dará siempre una 'S'.

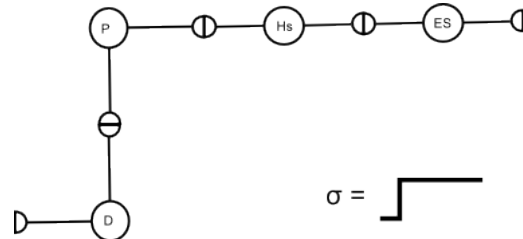


Figura 6. Configuración para la acción correr



Figura 7. Acciones de paralizarse (avatar izquierdo) y correr (avatar derecho) en la simulación generada

- En la figura 8 se muestra la acción gritar (ver figura 10), la cual también tiene dos generadores determinantes, pero en este caso no necesitan estar combinados. Lo anterior significa que la presencia de cualquiera lograría esta configuración. Los factores son: hipocondriasis de nivel 4 o niveles bajos de extroversión social. Este grafo siempre es una línea con vueltas de 90° en los extremos que pueden presentarse en el mismo sentido o en sentidos opuestos.

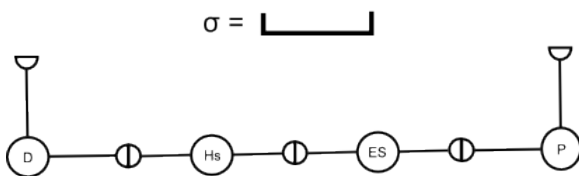


Figura 8. Configuración para la acción gritar

- La figura 9 muestra la acción huir (ver figuras 10 y 12). Esta acción depende de la aparición de cualquiera de dos generadores. Los factores son: nivel medio de hipocondriasis o nivel alto de depresión.

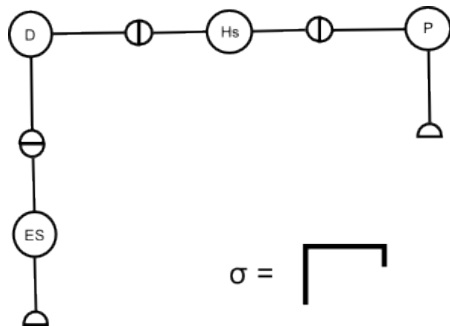


Figura 9. Configuración para la acción huir

- Por último, en la figura 11 se tiene la acción ayudar (ver figura 12). Esta acción al igual que la acción correr necesita la combinación de dos generadores para formar el patrón. Los factores deben ser niveles bajos de extroversión social y nivel medio de psicastenia.



Figura 10. Acciones de gritar (avatar izquierdo) y huir (avatares del lado derecho) en la simulación generada

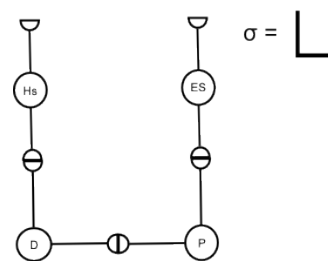


Figura 11. Configuración para la acción ayudar



Figura 12. Acciones de paralizarse (avatar izquierdo), huir (avatar del centro) y ayudar (avatar derecho) en la simulación generada

Algoritmo para la obtención de las acciones y el comportamiento de los avatares

Para determinar la conexión mencionada en la subsección previa, se desarrolló un algoritmo que recibe como entrada un vector de generadores, cada uno asociado a un rasgo de comportamiento de los antes mencionados.

Después se comprueban las conexiones que pueden ser establecidas entre los generadores y estas se ingresan en un nuevo vector como pares de generadores que indican el orden en el que se hicieron las conexiones. El algoritmo empleado fue el siguiente:

Algoritmo para determinar el comportamiento de un avatar
 entrada: vector de comportamiento $X = [x_1, x_2, \dots, x_n]$
 salida: configuración de conexión σ

1. Crear $\sigma = \emptyset$
2. Para $i = 1$ hasta $i = \|X\| - 1$
3. Para $j = i + 1$ hasta $j = \|X\|$
4. Si se puede establecer la conexión entre x_i y x_j
5. $\sigma = \sigma \cup (x_i, x_j)$
6. terminar bucle
7. Devolver σ

El algoritmo anterior devuelve el vector de configuración con las conexiones realizadas, lo cual indica el tipo de grafo obtenido en la configuración.

Análisis de la configuración

Una vez que se tiene la configuración del comportamiento que debe ser expuesto por los avatares, se puede determinar el mismo haciendo un análisis de la conexión, ya sea revisando las conexiones internas o externas, así como tomar en cuenta el orden en el que se unieron los generadores. Como ya se vio hay generadores clave en las configuraciones y una vez que se encuentre una combinación esperada de estos o un generador determinado, como en el caso de la acción paralizarse, el análisis solo requiere la confirmación de las conexiones subsecuentes.

La principal ventaja del modelo expuesto éste trabajo es que conforme aumente la cantidad de acciones y en base a las escalas de personalidad, se puede expandir para definir tantos patrones de comportamiento como sea necesario. Aunque para nuestro caso de estudio sólo se tomaron en

cuenta cuatro de las diez escalas de personalidad que define el MMPI, ya que sólo estas intervienen en las acciones que pueden ser tomadas por los avatares ante situaciones de sismo.

V. CONCLUSIONES

En éste trabajo se expuso una metodología para la simulación realista de comportamiento de avatares ante situaciones de desastre, como lo es un sismo. Mediante el análisis de la conducta expuesta por los avatares en las simulaciones, se puede hacer un mejor diseño de edificios y rutas más rápidas de evacuación, de esta manera se pueden tomar mejores decisiones al planear el diseño y hacer la construcción de lugares concurridos o de eventos masivos.

La personalidad asignada a los avatares fue generada apoyándonos en un base psicológica mediante el uso y la aplicación del MMPI, de esta forma fue posible la generación de conductas realistas por parte de los avatares, dichas conductas fueron coherentes con el evento que percibían estos y lo más destacado es que fueron obtenidas en base a los valores de sus escalas de personalidad.

VI. AGRADECIMIENTOS

Este trabajo está parcialmente apoyado por el CoECyT-Jal proyecto No. 2008-05-97094, mientras que sus autores Victor Fernández Cervantes, Luis Real Novo y Héctor Rafael Orozco Aguirre cuentan con beca de CONACYT No. 212753, 212758 y 203060, respectivamente.

VII. REFERENCIAS

- [1] I. López, J.J. Aliño, M. Valdés, *Manual Diagnóstico y Estadístico de los Trastornos Mentales*, Editorial Masson, Barcelona, España, 2002.

- [2] J. S. Tyhurst, “*Individual Reactions to Community Disaster. The Natural History of Psychiatric Phenomena*”, American Journal of Psychiatry, 1951.
- [3] R. Perry and M. Lindel. *The Psychological Consequences of Natural Disaster: A Review of Research on American Communities*, Mass Emergencies, 1978.
- [4] A. S. Demi and M. S. Miles, *Understanding Psychologic Reactions to Disaster*, Journal of Emergency Nursing, 1983.
- [5] G.-M. Emilia Lucio and G. M. Ivonne - *Maqueo, Dra. María Ivonne León, Uso e Interpretación del MMPI-2 en Español*, Manual Moderno, 2003.
- [6] A. Treuille, S. Cooper and Z. Popović. “*Continuum Crowds*”, ACM Transactions on Graphics, SIGGRAPH, 2006.
- [7] G. Francisco, L. Miguel, B. Fernando and V. Guillermo, “*Animating Groups of Socially Intelligent Agents*”, International Conference on Cyberworlds, University of Valencia, 2007.
- [8] M. Fiorenzo, Y. Barbara and M. Jonathan. “*Real-Time Scalable Motion Planning for Crowds*”, International Conference on Cyberworlds, Virtual Reality Laboratory, EPFL, Switzerland, 2007.
- [9] G. Ulf. *Elements of Pattern Theory*. Johns Hopkins University Press, 1996.

VIII. AUTORES

Ing. Víctor Fernández Cervantes es egresado de la carrera de Ingeniería en Computación del Centro Universitario de la Cienega de la Universidad de Guadalajara, desde el 2007 es alumno de la Maestría en Ciencias en Ingeniería Eléctrica con especialidad en computación en el CINVESTAV, unidad Guadalajara.

M. en C. Luis Eulalio Real Novo es egresado de la Maestría en Ciencias en Ingeniería Eléctrica con especialidad en computación por el CINVESTAV, unidad Guadalajara.

M. en C. Héctor Rafael Orozco Aguirre es Maestro en Ciencias en Ingeniería Eléctrica con especialidad en computación por el CINVESTAV, unidad Guadalajara, desde el 2005 es profesor de asignatura del Centro Universitario de Ciencias Exactas e Ingenierías de la Universidad de Guadalajara y en el 2006 ingresó a estudiar el Doctorado en Ciencias en Ingeniería Eléctrica en el CINVESTAV, unidad Guadalajara.

Dr. Félix Francisco Ramos Corchado obtuvo su título de Doctor en Computación de la Universidad de Tecnología de Compiègne Francia, tiene 15 años de experiencia en la investigación, la docencia y el desarrollo tecnológico.

Dr. Marco Antonio Ramos Corchado estudió su doctorado en la Universidad de Toulouse I, Francia, desde el 2007 es profesor de tiempo completo en el departamento de computación de la Universidad Autónoma del Estado de México.

Verónica Pérez brindó asesoría en la parte de psicología que fue necesaria para el desarrollo de éste trabajo.

HERRAMIENTA VISUAL PARA EL DESARROLLO DE SOFTWARE MEDIANTE ABSTRACCIONES GRÁFICAS

Alejandro Medina Santiago¹, Eva Valdez Alemán¹, Francisco Alberto Zavala Rosado²

¹Tecnológico de Estudios Superiores de Ecatepec
Ecatepec de Morelos, Estado de México. 55210, México
e-mail: alejandro_medina@tese.edu.mx, evaldez@tese.edu.mx

²Departamento de Tecnologías de la Información, Universidad Tecnológica de Campeche
Ciudad del Carmen, Campeche. 24381, México
e-mail: fzavala@utcam.edu.mx

RESUMEN

Actualmente hay programas que a partir de un conjunto de diagramas, son capaces de crear un esqueleto en código fuente, en diversos lenguajes de programación. Sin embargo, el usuario debe escribir todavía el código faltante y dominar un lenguaje de programación en que se codifica la funcionalidad de los elementos. Es posible generar un programa ejecutable sin escribir código fuente, mediante diagramas de flujo, estructuras de control y operaciones básicas. Por tanto es importante desarrollar un software capaz de diseñar y crear aplicaciones de escritorio, sin escribir código fuente, mediante el uso de elementos gráficos.

Palabras Claves: código fuente, diagrama, elementos gráficos, gramática, estructuras de control, entidades anidadas.

I. INTRODUCCIÓN

El componente central de un programa es sin duda el código fuente. Este código son las líneas de instrucciones que se escriben, siguiendo reglas de sintaxis, orden y semántica, mediante el cuál se le indica al compilador las acciones que debe realizar. Para los programadores es sin duda la parte más difícil de elaborar. Un lenguaje visual, es la herramienta que te permite añadir objetos visuales (botones de comando, Cuadros de Texto, Etiquetas, Imágenes, Cuadros de Diálogo, etc.), para proporcionarle a las

aplicaciones una mejor interfaz gráfica, que añada facilidad de uso al usuario final.

Existen actualmente diversos lenguajes visuales, capaces de crear robustas aplicaciones con interfaces gráficas de excelente nivel, tal es el caso de Microsoft Visual Studio .Net ó la última versión de Borland Delphi. También tenemos aplicaciones como Rational Rose (IBM) que mediante diagramas de clases, pueden generar un esquema o esqueleto vacío de código fuente.

Sin embargo, también se tienen que programar los métodos y eventos, mediante código fuente. Las estructuras de control se utilizan para controlar el flujo de ejecución de las instrucciones del código.

Las estructuras de control pueden ser representadas mediante elementos gráficos. Estos elementos gráficos están compuestos de líneas, flechas unidireccionales y multidireccionales, además de figuras geométricas en dos dimensiones tales como círculos, cuadros, rectángulos, rombos, etc.

A partir de estos elementos gráficos, se puede “interpretar” el equivalente código fuente que el usuario requiere, y poder generar un archivo ejecutable.

La sección 2 describe la forma de detectar los elementos gráficos para iniciar el proceso

adjunto a la separación de estos elementos a una gramática además de la validación del diagrama original y finalmente como generar el archivo ejecutable. La sección 3 muestra el desarrollo lógico y físico de la aplicación en un lenguaje de programación. En la sección 4 se muestran los resultados obtenidos trabajando directamente en la aplicación desarrollada. La sección 5 describe las conclusiones finales.

II. GENERAR ARCHIVO EJECUTABLE A PARTIR DE UN DIAGRAMA DE FLUJO

A.- DETERMINACIÓN DE UNIDADES GRÁFICO-LÉXICAS

Las estructuras de control más comunes se agrupan en: secuencia, las cuales indican el flujo y orden de los procesos a ejecutar, decisión, usadas cuando tenemos dos o más opciones a seguir lo cual depende del resultado de comparaciones, iteración, utilizadas en los casos de acciones repetitivas o cuantitativas, recursión, las cuales permiten a cierto grupo de instrucciones calcular un resultado que sirve como entrada de datos del mismo y recalculan el resultado final.

Por tanto, debemos implementar un módulo de “abstracción” de los elementos gráficos a estructuras de control. Este módulo está basado en comparaciones con los elementos ya establecidos. Las rutinas de secuencia son como los de la Figura 1, y son los bloques bases para todos los elementos.

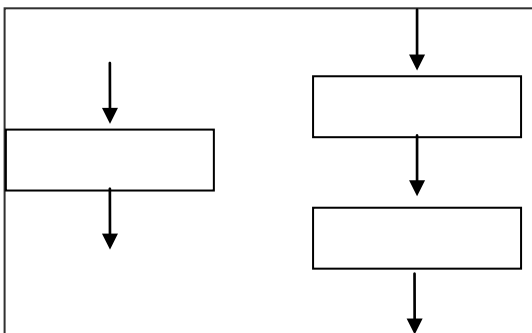


Figura 1. Estructura de secuencia

Los rectángulos representan una entidad, que puede contener una operación básica, otra estructura de control u otra entidad.

Las flechas se utilizan para tener unido todo el diagrama, ayudando a detectar errores desde el inicio, y sobre todo marcar el flujo y orden de todos los procesos.

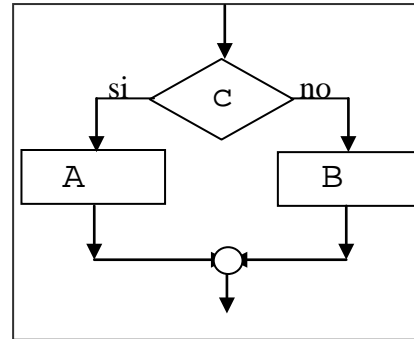


Figura 2. Estructura de decisión

Dentro de las estructuras de decisión como la de la Figura 2, tenemos principalmente al “if” y al “select case” ó también conocido como “switch”, que permite elegir entre múltiples opciones a la vez.

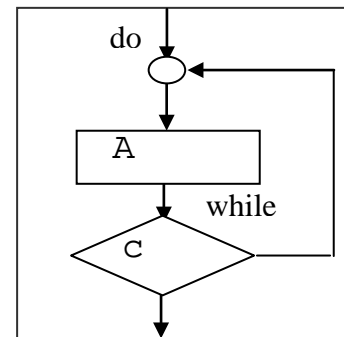


Figura 3. Estructura cíclica

Las estructuras de iteración principales son, For, While, Do – While, entre otras, conocidas generalmente como “ciclos” o “bucles”, ver Figura 3 y representan un camino secuencial que se repetirá un determinado número de veces.

En todas las estructuras, los identificadores A y B, son entidades que pueden contener una operación básica, otra estructura de control u otra entidad. El identificador C se refiere a una

comparación, que puede ser simple o compuesta.

De igual forma se deben definir todos los elementos gráficos que el sistema validará. Como podemos apreciar, todas las estructuras terminan en una punta de flecha, lo que facilita la conexión entre ellas. En este proceso, debemos convertir estos elementos en oraciones o estructuras canónicas.

B.- OBTENCIÓN DE LA GRAMÁTICA

El objetivo de este proceso es obtener la gramática de los elementos gráficos de ejemplo proporcionadas por el usuario.

Esta gramática servirá para interpretar y validar si todas las estructuras contenidas dentro de otras estructuras son válidas, ya que tenemos que dentro de una entidad, podemos tener una o más entidades, y así sucesivamente, es decir, tenemos estructuras “anidadas” [1, 3].

Partimos describiendo a una entidad con el símbolo, A es una asignación, E una estructura, O un conjunto de operaciones matemáticas, v el nombre de una variable, N un número y d un solo dígito numérico. Por tanto:

$S \rightarrow A | E | O$
 $A \rightarrow v = v | v = O | v = N$
 $O \rightarrow vsv | vsN | OsO | vsO | NsO | NsN$
 $s \rightarrow + | - | / | *$
 $N \rightarrow d | dN | d.N$
 $d \rightarrow 0 | 1 | 2 \dots | 9$
 $E \rightarrow IF | DO - WHILE | S$

Y así sucesivamente. Esta gramática se puede simplificar y reducir, para tenerla finalmente de forma optimizada. De esta gramática depende en gran parte el siguiente módulo de validación, por lo que se le debe asignar suficiente tiempo para diseñarla e implementarla. También se tienen que definir las equivalencias o “traducciones” de las oraciones canónicas a las líneas de código fuente del lenguaje de programación [4, 5].

C.- VALIDACIÓN DEL DIAGRAMA

En este proceso debemos validar las oraciones obtenidas del diagrama, con la gramática obtenida. Esto con la finalidad de detectar información faltante, estructuras completas y correcta formulación de estructuras anidadas. Esto significa, que el usuario puede colocar estructuras dentro de otras estructuras, sin tener límites de nivel, y la gramática es capaz de validar estas estructuras anidadas. Esto se logra usando procedimientos normales, que recorran el árbol de la gramática a través de comparaciones y sustituciones [1, 5].

Es importante resaltar, que aunque en teoría no hay límite de estructuras anidadas, en la realidad debemos tomar en cuenta aspectos como la cantidad de memoria asignada a la aplicación, el tamaño del archivo que contiene elementos gráficos, entre otros. Ya que aunque se tenga un diagrama válido, su gran tamaño podría llegar a indicar errores sin causa aparente.

D.- GENERAR EL ARCHIVO EJECUTABLE

Al tener un conjunto de oraciones en su forma canónica, totalmente válidas en sintaxis, podemos iniciar la traducción correspondiente a su equivalencia en el lenguaje de programación. La complejidad de este módulo depende directamente de la sintaxis y reglas del lenguaje del código fuente. Por ejemplo, si el lenguaje a utilizar es Visual Basic, será menos complejo, que los lenguajes como Java ó C ++. En este proceso se “traducen” todas las equivalencias a código fuente, para generar el archivo ejecutable.

Este proceso final, depende directamente del éxito de los anteriores, ya que si se inicia con oraciones erróneas, este proceso fácilmente se puede ciclar o bloquear, y en ciertos casos, se proporcionará un archivo ejecutable fallido, que podría contener errores, o simplemente funcionar diferente al diagrama inicial que el usuario planteó.

III. DISEÑO Y DESARROLLO DE LA APLICACIÓN

A.- DISEÑO

El diseño de la aplicación se basa literalmente en el procedimiento detallado en la sección 2. Consiste en proporcionar al usuario las herramientas visuales para poder crear un diagrama de flujo (inicio, fin, entradas, salidas, decisiones, ciclos, etc.), y permitirle unirlos mediante flechas de dirección. Conforme el usuario avanza en la formación del diagrama, se debe ir formulando el código equivalente en una matriz de arreglos [2]. Finalmente, podemos ejecutar ese código equivalente y brindar el escenario que le proporcione al usuario una visión de los resultados que arroja su diagrama de flujo.

B.- DESARROLLO

La aplicación se desarrolló en Microsoft Visual Basic 6, y consta de 19 formularios, 6 módulos de clase y 8 módulos básicos. El formulario principal, ver Figura 4, muestra del lado izquierdo la barra de herramientas que representan las estructuras del diagrama de flujo. Estas estructuras se pueden seleccionar para ir creando el diagrama deseado. El menú principal contiene las opciones de administración de archivos más comunes tales como, abrir, guardar, salir, etc. Todo el espacio disponible, se puede utilizar como área de trabajo para elaborar los diagramas de flujo.

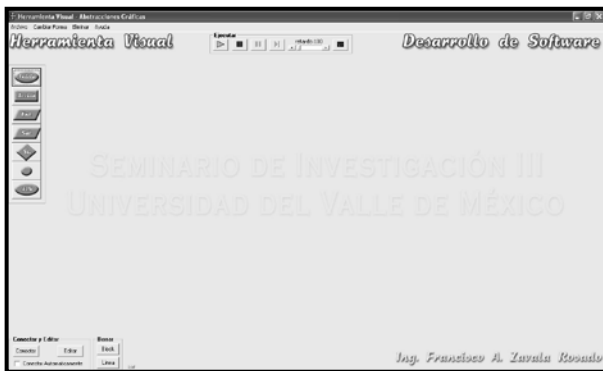


Figura 4. Pantalla principal

ESTRUCTURAS COMUNES

Tal y como lo muestra la Figura 5, las estructuras disponibles en el sistemas son, inicio, acción, entrada, salida, condición, conector y fin. Con estas estructuras, podemos crear diagramas de flujo de niveles básico e intermedio sin limitantes en el número de objetos, aunque en realidad, las limitantes las proporciona el equipo de cómputo donde se este ejecutando la aplicación.



Figura 5. Estructuras disponibles

BARRA DE EJECUCIÓN

En la parte media superior del formulario principal, encontramos la barra de ejecución, ver Figura 6, que contiene las opciones que controlan la ejecución y flujo del programa en cuestión, tales como, iniciar, detener, pausar, avanzar, duración del proceso y la pantalla de salida. Esta barra esta disponible en todo momento dentro del sistema.



Figura 6. Barra de ejecución

HERRAMIENTAS DE CONEXIÓN

En la parte izquierda inferior del formulario principal, se encuentra el grupo de opciones, ver Figura 7, que permiten la interconexión de las estructuras del diagrama de flujo. La interconexión se logra uniendo visualmente dos estructuras mediante una flecha direccional, que indica el sentido del flujo de datos y acciones dentro del diagrama creado. Estas flechas pueden ser de forma manual, donde el propio usuario debe seleccionar continuamente los dos objetos a conectar y presionar el botón de conexión, ó de forma automática, marcando esta opción, el sistema enlazará los dos últimos objetos colocados mediante una flecha con dirección hacia el último objeto creado. También se encuentran en este grupo las opciones para eliminar tanto flechas de conexión como objetos dibujados.

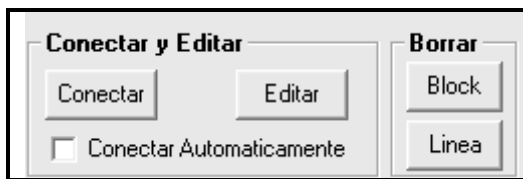


Figura 7. Grupo de Conexión

CREACIÓN DEL ARCHIVO EJECUTABLE

Para crear el archivo ejecutable, primero creamos un archivo de texto, que contendrá la interpretación de las estructuras dibujadas, esto es, propiedades, atributos y las conexiones entre ellas. Esta información se va formando y/o actualizando, conforme el usuario realiza cambios en el diagrama de flujo. Esta misma información es la usada para guardar el diagrama en archivo, para su uso posterior. Por último, simplemente tomamos una copia del archivo ejecutable de nuestro sistema, y lo guardamos con el nombre que el usuario proporcione, con la única diferencia, que el programa no mostrará la interfaz principal, sino simplemente, la ventaja de ejecución, con los resultados programados en el diagrama de flujo.

IV. RESULTADOS

La aplicación, instalada en un equipo de cómputo de medianas características, deberá ejecutarse sin ningún problema. Se presentan dos ejemplos característicos a las que fue sometida la aplicación. El primero es de nivel básico y el segundo de nivel intermedio.

PRIMER EJEMPLO

Se dibujó el diagrama de flujo de un programa que solicite un número entero e identifique si el número es positivo o negativo. Primero se dibujó el diagrama de flujo, ver Figura 8, mediante las herramientas que representan las estructuras. Cabe mencionar, que el diagrama no necesariamente debe tener esa forma ni contener el mismo número de estructuras para producir el mismo resultado.

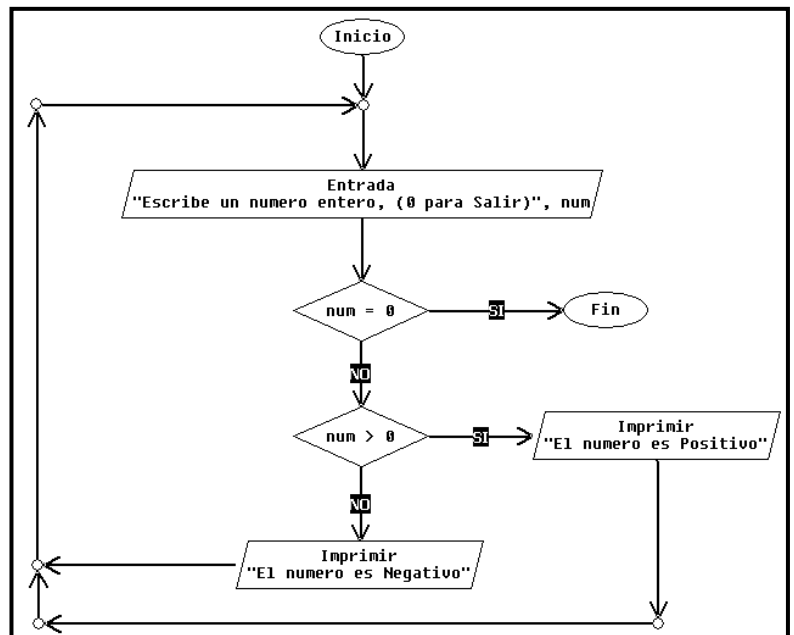


Figura 8. Diagrama de Flujo de nivel básico

Se guardó el archivo en el disco duro, y se ejecutó, obteniendo los resultados de salida mostrados en las Figuras 9 y 10.

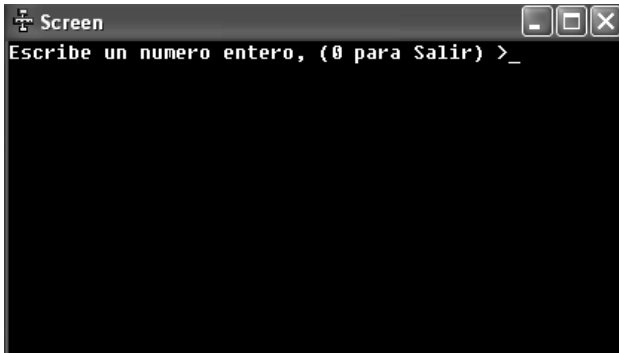


Figura 9. Inicio de la Ejecución del Diagrama de Flujo

Se muestra claramente que los resultados son los esperados, pues en la pantalla se imprimen resultados correctos acorde a lo que el usuario escriba.

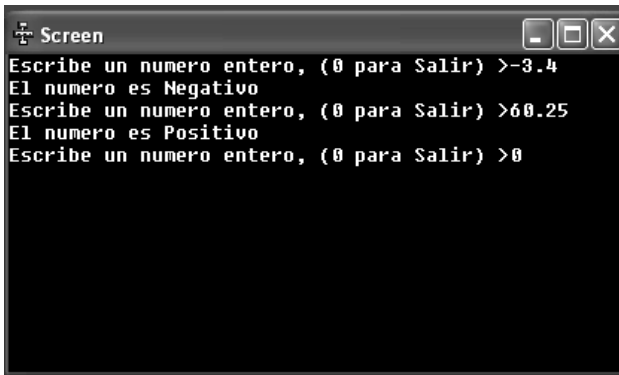


Figura 10. Ejecución del Diagrama de Flujo

Por último se procedió a crear el archivo ejecutable, por lo cual se obtuvo el resultado final esperado y el archivo de texto que contiene las instrucciones a seguir por parte del archivo ejecutable. El archivo de texto contiene los siguientes datos:

Entrada

Escribe un numero entero, (0 para Salir)

```
num
id3
92    <--TYPE
620   <--LEFT
376   <--TOP
112   <--WIDTH
50    <--HEIGHT
16777215 <--BACKCOLOR
0     <--BORDERCOLOR
```

```
0     <--BORDERCOLOR
IF_GREATER
num
0
id4
91    <--TYPE
817   <--LEFT
380   <--TOP
204   <--WIDTH
40    <--HEIGHT
16777215 <--BACKCOLOR
0     <--BORDERCOLOR
0     <--BORDERCOLOR
Imprimir
El número es Positivo
id5
91    <--TYPE
573   <--LEFT
486   <--TOP
204   <--WIDTH
40    <--HEIGHT
16777215 <--BACKCOLOR
0     <--BORDERCOLOR
0     <--BORDERCOLOR
Imprimir
El número es Negativo
.....
```

Esta es la información obtenida mediante la interpretación o abstracción de los elementos gráficos que componen el diagrama de flujo.

SEGUNDO EJEMPLO

Se dibujó el diagrama de flujo de un programa que solicite un número entero e identifique si al número se le puede calcular la factorial, de ser así, que lo calcule y lo imprima. Primero se dibujó el diagrama de flujo, ver Figura 11, mediante las herramientas que representan las estructuras. Cabe mencionar, que el diagrama no necesariamente debe tener esa forma ni contener el mismo número de estructuras para producir el mismo resultado.

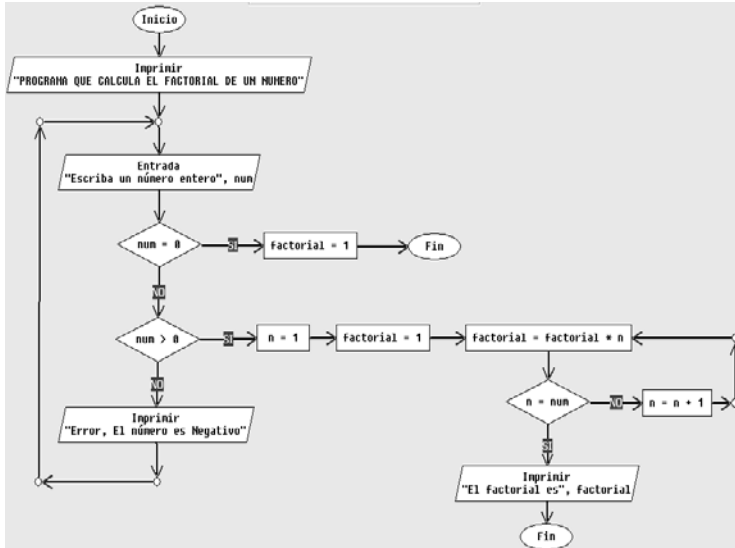


Figura 11. Diagrama de Flujo de nivel intermedio

Se guardó el archivo en el disco duro, y se ejecutó, obteniendo los resultados de salida mostrados en las Figuras 12 y 13.

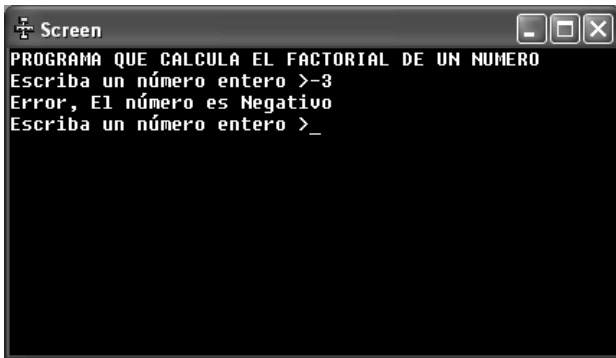


Figura 12. Inicio de la Ejecución del Diagrama de Flujo

Se muestra claramente que los resultados son los esperados, pues en la pantalla se imprimen resultados correctos acorde a lo que el usuario escriba.

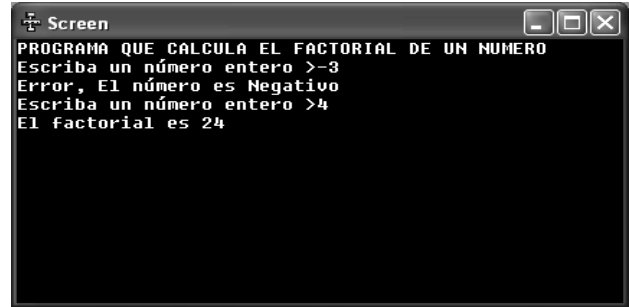


Figura 13. Ejecución Completa del Diagrama de Flujo

Por último se procedió a crear el archivo ejecutable con lo cual se logra el objetivo principal de la aplicación. Como lo muestran los dos ejemplos anteriores, en niveles básicos e intermedios, la aplicación funciona muy bien y no genera ningún error. En el segundo ejemplo, la extensión del diagrama de flujo ya es significativa, aun así, la aplicación sigue trabajando con el mismo rendimiento.

RENDIMIENTO REAL

Para justificar la factibilidad de la aplicación, se desarrollaron tres proyectos de nivel básico y tres de nivel intermedio. Esto se realizó de dos maneras, escribiendo el código fuente en un compilador genérico de C++, y la otra, dibujando el diagrama de flujo en la aplicación. Para no dar ventajas a ninguna parte, primero se diseñó el algoritmo de los 6 ejemplos, y a partir de ahí, se contabilizaron los tiempos de desarrollo. Los resultados se muestran en la Figura 14.

Se tienen que tomar en cuenta, que los tiempos mostrados en la gráfica de la Figura 14, fueron logrados por personas con conocimientos intermedios de programación. Es válido cuestionar estos resultados si tomamos en cuenta que existen variables que alterarían hasta en un 30 % los tiempos mostrados. Tales variables son: nivel de conocimiento de programación del usuario, el compilador y lenguaje de programación seleccionado, facilidad del usuario para escribir y/o dibujar, entre otros.

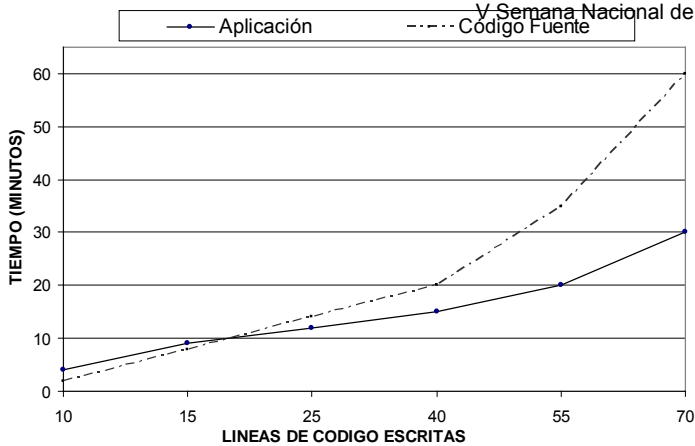


Figura 14. Gráfica comparativa de rendimientos

En los puntos graficados, la coordenada x corresponde al número de líneas de código fuente, y la coordenada y representa el tiempo. Por lo cual tenemos 6 puntos coordinados que corresponden a los 6 ejemplos de prueba. Utilizando el paquete matemático Matlab 6.5, podemos obtener un polinomio que aproxime la curva de la Figura 14. Haciendo uso de la instrucción polyfit y utilizando el método de aproximación de mínimos cuadrados obtenemos:

$$T(n) = 0.0003n^3 - 0.032n^2 + 1.2994n - 5.6398 \quad \text{----- (1)}$$

Donde n representa al número de líneas de código fuente y T representa el tiempo empleado en obtener el archivo ejecutable a partir del diagrama de flujo. También se puede apreciar que el polinomio es de tercer grado. Se puede obtener otra función de mayor grado, y por consiguiente mayor precisión, pero la dificultad de evaluación también crece, por lo que no se recomienda.

V. CONCLUSIONES

En este artículo se ha presentado un procedimiento para que a partir de un diagrama de elementos gráficos, se pueda obtener una aplicación o archivo ejecutable específico, mediante abstracciones gráficas de los elementos individuales. Esto permitirá a los programadores ahorrar tiempo y esfuerzo, al momento de escribir sus programas, sobre todo

a los principiantes. Hay que realizarle pruebas exhaustivas después del diseño al sistema, para observar sus resultados reales, y así poder liberar el sistema con resultados cualitativos.

El diseño completo del sistema es realizado en Microsoft Visual Basic 6.0, esto por la facilidad que presenta este lenguaje para formular las estructuras gráficas a utilizar. Los alcances del sistema dependen directamente de la cantidad de estructuras y operaciones que acepte la aplicación, además del periodo de prueba necesario para poder liberarlo.

Se mostraron dos ejemplos de diagrama de flujo que fueron convertidos exitosamente a su correspondiente archivo ejecutable. Sin embargo, la aplicación se sometió a más pruebas en busca de posibles fallas, las cuales fueron corregidas en su momento.

Esta aplicación está dirigida sobre todo a programadores de nivel básico e intermedio ya que en un nivel avanzado el rendimiento no sería significativo, debido a los alcances de la propia aplicación.

Este procedimiento puede ser solo un pequeño paso para integrar finalmente, un sistema donde se puedan realizar aplicaciones robustas y complejas, sin escribir código fuente, mediante la implementación de relaciones, entidades, clases, y objetos de tipo gráfico. Se recomienda usar un lenguaje de programación robusto y estructurado para su codificación, para ofrecer toda la infraestructura y seguridad a las aplicaciones a desarrollar mediante el sistema.

VI. REFERENCIAS

- [1] Elsa Barruecos Rodríguez, *Elaboración de Gramáticas en Sistemas Evolutivos Generador de Esquemas lógicos de Bases de Datos*, Seminario de Titulación, IPN- UPIICSA 1988.
- [2] Jesús Manuel Olivares C., *Programación Orientada a Objetos*

utilizando JAVA, Diapositivas informativas. IPN-UPIICSA, 2005.

- [3] James Noble and Robert Biddle, *Program Visualization for Visual Programs*, Computer Science, Victoria University of Wellington 1998.
- [4] Norman Murray, Norman Paton and Carole Goble, *KalNdoquery: A Visual Query Language for Object Databases*, Department of Computer Science, University of Manchester. Oxford Road, Manchester, M 13 9PL, UK.
- [5] Frederic Van de Veire, Przemyslaw Szmaj, Jaroslaw Francik, *Siamoa - A System For Visual Programming, Program Visualisation And Debugging*, University of Sciences and Technologies of Lille, France, Technical University of Silesia, Gliwice, Poland.

VII. AUTORES

D. en C. Alejandro Medina Santiago obtuvo el título de doctor en ciencias en la especialidad de ingeniería eléctrica en el diseño de circuitos integrados para redes neuronales y lógica difusa.

M. en C. Eva Valdez Alemán obtuvo su título de maestra en ciencias de la educación en la especialidad de diseño curricular.

Ing. Francisco Alberto Zavala Rosado obtuvo su título de ingeniero en sistemas computacionales con la especialidad de diseño de software.

GENERACIÓN DE PROTOCOLOS DE INTERACCIÓN EMPLEANDO PROGRAMACIÓN GENÉTICA

Gutiérrez-García José Octavio^{1,2} y González-Padilla Omar Alfredo¹

¹ Centro de Investigación y de Estudios Avanzados del IPN, Unidad Guadalajara,
Av. Científica 1145, Col. El Bajío, Zapopan 45015, Jalisco, México

²Grenoble Institute of Technology, LCIS Research Laboratory,
50, rue Barthélémy de Laffemas, BP 54, 26902 Valence Cedex 9, France
{jgutierrez, oglez}@gdl.cinvestav.mx

RESUMEN

En este artículo se propone un método para la generación de protocolos de interacción de agentes, utilizando programación genética. El método es orientado a metas. Una meta es definida como la obligación que tiene un agente de alcanzar un estado dado. Los mensajes de los protocolos son definidos por medio de precondiciones y efectos, ambos expresados en términos de obligaciones. De esta manera se logra expresar la interacción en obligaciones (metas), las cuales son utilizadas para guiar la generación automática del protocolo.

Palabras clave: Sistemas Multi-Agente, Normas sociales, obligaciones, protocolos de interacción, programación genética.

I. INTRODUCCIÓN

Los protocolos de interacción son requeridos para proveer los lineamientos de interacción que los agentes deben de seguir con el fin de alcanzar sus objetivos. La interacción entre agentes, por naturaleza, es dinámica y por tanto los protocolos de interacción deben de ser generados y adaptados de acuerdo a sucesos dinámicos, los cuales podrían alterar las metas de los agentes. En ambientes dinámicos es común que los agentes cambien de prioridades y por tanto sus metas sean otras, consecuentemente la creación de los protocolos de interacción debe estar basada en las metas de los agentes. Para lograr esto, los protocolos de interacción deben de ser explícitamente definidos

en términos de metas. Esto puede ser alcanzado por medio de la definición de protocolos de interacción usando normas sociales, particularmente obligaciones. Los agentes pueden considerar el cumplimiento de las obligaciones como las metas a perseguir, y estas en cambio, pueden guiar la interacción del agente.

La herramienta utilizada para generar los protocolos de interacción es la Programación Genética, primeramente definida por J. R. Koza en el trabajo presentado en [5]. La programación genética es normalmente aplicada a la generación automática de programas de computadoras; esta es presentada como una extensión a los algoritmos genéticos. Dentro del enfoque de la programación genética, una solución a un problema, es encontrada de la siguiente manera: primero se genera una población de individuos aleatoriamente, estos representan programas de computadora que pudiesen ser posibles soluciones para un problema dado, a continuación se utilizan técnicas evolutivas para desarrollar la población a subóptimas u óptimas soluciones, una vez que son alcanzados ciertos criterios, se detiene el proceso de evolución y se selecciona el mejor individuo, el cual representa la solución, en caso de existir una.

Las metodologías para guiar la interacción entre agentes por medio de normas sociales han sido ampliamente estudiadas, como se puede ver en las siguientes referencias [1], [2], [3], [4], [6], y [10]. En el trabajo presentado en [4], se presenta un formalismo para modelar la interacción basada en obligaciones. Las obligaciones son imposiciones a

uno mismo a alcanzar una situación determinada, demandada por una fuerza social.

Dado que las obligaciones son compromisos con uno mismo, permiten plantear la interacción entre agentes de una manera descentralizada, conforme a las obligaciones que los agentes van adquiriendo. La creación de la secuencia de mensajes es vista desde la perspectiva de un solo agente, pronosticando que los agentes participantes acepten la inducción de ciertas obligaciones. En caso de la no aceptación de alguna obligación se regenera la conversación tomando en cuenta la nueva información. Manejar la creación de las conversaciones de esta manera, elimina la existencia de un agente central, el cual crea las secuencias de mensajes válidos, que posteriormente pasa a los demás agentes. Esto centraliza al sistema multi-agente, yendo en contra de los principios de los sistemas distribuidos.

Entre las técnicas para generar protocolos de interacción, se encuentra el trabajo realizado por Quenum et al. [7], [8] y [9], donde los protocolos de interacción son derivados de protocolos genéricos, los cuales proveen información mínima de la conducta que el agente debe seguir. Otra investigación enfocada a la adaptación dinámica de protocolos se encuentra en [11], donde se propone un lenguaje de coordinación llamado *Reo* que permite la reconfiguración dinámica de los protocolos de interacción. Las investigaciones mencionadas anteriormente no son orientadas a metas, limitando su capacidad a situaciones semi-previstas.

En este artículo se propone un método para la generación de protocolos de interacción de agentes, utilizando programación genética. Este método consiste en 2 fases. La primera fase, involucra el modelado de mensajes como operadores que remueven y añaden obligaciones a los agentes. La segunda fase utiliza estos operadores además de las metas (obligaciones a cumplir), para guiar la generación automática del protocolo de interacción por medio de la programación genética.

El artículo está estructurado de la siguiente manera: la sección 2, explica como los protocolos de interacción pueden ser definidos en términos de obligaciones; la sección 3, presenta el uso de la programación genética para generar conversaciones entre agentes; la sección 4, contiene un ejemplo ilustrativo del método aquí propuesto; finalmente, la sección 5 presenta las conclusiones obtenidas.

II. OBLIGACIONES Y LOS PROTOCOLOS DE INTERACCIÓN

Una obligación es una imposición a sí mismo de alcanzar algún estado en particular, se representa simbólicamente de la siguiente manera: $O(a, f)$, lo cual significa que el agente a esta obligado a alcanzar el estado f .

El intercambio de mensajes entre agentes, crea y libera obligaciones, para ser preciso, las precondiciones para enviar un mensaje y sus efectos están definidas en términos de obligaciones. Como ejemplo de interacción basada en obligaciones, considere el siguiente ejemplo: cuando un saludo es recibido, una obligación de contestar el saludo es inducida por el emisor hacia el receptor, y si esta es aceptada, entonces la interacción toma lugar.

Las obligaciones pueden ser creadas y liberadas. A continuación una breve descripción de los operaciones para administrar obligaciones.

Crear obligaciones. La creación de una obligación para alcanzar un estado f por medio del envío de un mensaje $m()$, se representa de la siguiente manera:

$$\text{CrearO}(m(a), O(a,f), t)$$

Donde el predicado $m(a)$ indica que el mensaje fue enviado por el agente a , y t representa el instante cuando el mensaje fue enviado.

Liberar obligaciones. La liberación de una obligación toma lugar cuando el estado f ha sido

alcanzado por el agente a . Esto se enuncia de la siguiente manera:

$$\text{LiberarO}(m(a), O(a,f), t)$$

La definición de las operaciones sobre obligaciones esta formalizada con cálculo de eventos, el cual es un formalismo basado en lógica temporal para representar eventos y sus efectos. En [4] se presenta en detalle las definiciones.

A partir de las obligaciones se pueden modelar los roles que los agentes juegan dentro de un protocolo de interacción. Un rol tiene cinco elementos, los cuales son descritos a continuación:

1. **Contexto de Interacción (W).** Este contiene todas las obligaciones que los agentes pueden adquirir, así como aquellas que puede inducir a otros agentes. Se define de la siguiente manera:

$$W = \{f(\varepsilon)_1, f(\varepsilon)_2, \dots, f(\varepsilon)_n\}$$

2. **Dominio del Contexto (K).** Este provee los parámetros que indican las propiedades y/o nivel de las obligaciones a las cuales el agente se compromete.

$$K = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_3\}$$

3. **Obligaciones iniciales (I).** Cuando un agente instancia un rol, se le asigna automáticamente un conjunto de obligaciones, por el simple hecho de ejecutar el rol. Esto se define de la siguiente manera:

$$\forall f(\varepsilon)_i \in I \mid I = \{f(\varepsilon)_1, f(\varepsilon)_2, \dots, f(\varepsilon)_k\} \\ \text{tenemos } O(a, f(\varepsilon)_j)$$

Estas obligaciones se consideran como las restricciones iniciales para el método basado en programación genética.

4. **Mensajes (M).** Un mensaje es una tupla (P, M, E) , donde P significa precondiciones; M mensaje, y E efectos del mensaje.

Las precondiciones son expresadas por un conjunto de conjunciones y/o disyunciones de un predicado de cálculo de eventos $\text{HoldsAt}(h,t)$, el cual indica que el estado h está activo en el instante t . El valor de h puede ser un obligación a alcanzar un estado $O(a,f)$ o un simple estado f , lo cual significaría que la obligación ya fue alcanzada.

Los efectos de enviar un mensaje son representados por medio de la liberación y asignación de obligaciones.

Los mensajes están definidos como operadores, con el objetivo de aplicar a la programación genética, esto se podrá ver en las siguientes secciones.

5. **Estado de Interacción (S).** El estado actual de un agente está determinado por las obligaciones que este tiene asignadas en un momento dado, así como las obligaciones que ya han sido liberadas, todo esto agrupado por conversación, quedando definido de la siguiente manera:

$$\text{Estado de Interacción}_{Ci} = \\ \{\text{HoldsAt}(O(a,f), t) \cup \text{HoldsAt}(g,t)\} \mid \forall f, \forall g \in W$$

El estado de interacción es actualizado mediante la emisión y recepción de mensajes.

Los roles interrelacionados son agrupados en organizaciones. Una organización es definida como una tupla (W, K, G, P, R, S) , donde W denota contexto de interacción; K dominio del contexto; G grupo de agentes miembros de la organización; P propietario de la organización; R conjunto de roles; y S Estado de interacción. Tanto W, K , como S son resultado de la unión apropiada de los correspondientes elementos pertenecientes a los roles de la organización. Para más detalles acerca de las definiciones del rol y organización ver [4].

III. PROGRAMACIÓN GENÉTICA PARA LA CREACIÓN DE PROTOCOLOS DE INTERACCIÓN

El objetivo es obtener una secuencia de mensajes que pueda liberar un conjunto de obligaciones metas. Los dos elementos más importantes para trasladar este problema de manera que pueda ser abordado mediante la programación genética son las *Funciones* y los *Terminales*. Para esta aplicación, los terminales son representados por el conjunto de mensajes que un agente perteneciente a una organización puede enviar; las funciones, toman como parámetros los terminales y funcionan como conectores y/o modificadores de las secuencias de mensajes.

Hay tres fases principales de la programación genética, a continuación una breve descripción de cada una de ellas:

- 1) Crear la población inicial de individuos. Para el caso tratado, un individuo representa una secuencia de mensajes. La manera en que se obtiene esta población es la siguiente: se toman aleatoriamente elementos del conjunto de funciones y terminales, se agregan uno a uno y se colocan en posiciones aleatorias dentro de un árbol, el cual representa la estructura del individuo. Este proceso se realiza de manera incremental hasta llegar a un máximo establecido, en lo que respecta al tamaño del árbol.
- 2) Evaluar la aptitud de cada secuencia de mensajes, basándose en métricas de desempeño, las cuales son definidas en función al problema que se aborda.
- 3) Mantener un grupo selecto de individuos, normalmente los que presentaron mayor aptitud, después evolucionar la población, por medio de técnicas de recombinación y mutación. La recombinación se lleva a cabo tomando 2 individuos y seleccionando en cada uno de ellos un nodo corte, el cual será la raíz del nuevo subárbol, finalmente

se intercambian los subárboles entre los individuos. En el caso de la mutación, los cambios se realizan elemento por elemento. Nosotros consideramos que un elemento debe de mutar con una probabilidad del 10%; si el nodo es una función se selecciona aleatoriamente un elemento de conjunto de funciones y se intercambia con el elemento seleccionado para mutar, lo mismo aplica para los terminales.

Los pasos 2 y 3 son repetidos hasta alcanzar una solución. Un individuo representa una solución si cumple que todas las obligaciones asignadas como metas hayan sido alcanzadas.

Los parámetros a considerar en la programación genética son listados a continuación:

- **TERMINALES:** Estos están representados por el conjunto de mensajes que pueden ser enviados por los agentes
- **FUNCIONES:** Estas permiten la conformación de secuencias de mensajes válidas. Las funciones reciben como entrada mensajes o secuencias de los mismos. Las funciones utilizadas son las siguientes:

Entrada: BloqueMensajes

Función: EJECUTAR()

Salida: BloqueMensajes/nil

Entrada: BloqueMensajes1,
BloqueMensajes2

Función: UNIR():

Salida: BloqueMensajes3/nil

La función **EJECUTAR** toma como parámetro de entrada un conjunto de mensajes y los verifica, tomando en cuenta las restricciones requeridas por los mensajes (precondiciones) antes de ejecutar y cada vez que se ejecuta, actualiza el estado de interacción. En caso de que el bloque de mensajes pueda ser ejecutado,

este regresa el bloque de mensajes, en caso contrario regresa un valor nulo.

La función **UNIR** agrega un bloque de mensajes a otro; en este caso *BloqueMensajes2* es agregado al final de *BloqueMensajes1*, regresando una secuencia de mensajes llamada *BloqueMensajes3*. Esta función también verifica las restricciones de precondiciones y efectos de los mensajes, en caso de cumplirse, regresa la secuencia de mensajes, en caso contrario regresa un valor nulo.

- CASO DE EVALUACIÓN: Con el objetivo de evaluar cada individuo de la población es necesario definir un caso de evaluación. Este está definido por el conjunto de obligaciones iniciales, aquellas asignadas al agente por el hecho de ejecutar un rol determinado, así como por las obligaciones metas.
- MÉTRICAS DE APTITUD: Estas son necesarias para poder determinar la aptitud que presenta cada uno de los individuos, con respecto a la solución esperada. Entre estas métricas se encuentran:
 - Por cada dos bloques de mensajes alineados por medio de la función **UNIR**, se agregan 2 puntos. Dado que las secuencias de mensajes se esperan complejas y largas, en la mayoría de los casos, se supondría que a mayor número de mensajes enviados, mayor probabilidad de obtener una solución.
 - Por cada vez que una función **EJECUTAR** regrese un valor distinto a nulo, se agrega 1 punto.
 - Cuando la secuencia de mensajes libera todas las obligaciones meta, se agregan 10 puntos.
 - Mensajes que liberan más obligaciones que las que crean son preferidos, por tanto se agrega 1

punto por cada obligación liberada, dentro de la secuencia de mensajes.

La suma de todos los puntos da como resultado la aptitud del individuo, a mayor número de puntos, mayor aptitud.

- PREDICADO DE ÉXITO: Dado que múltiples secuencias de mensajes pueden liberar las obligaciones meta, aunque no todas sean óptimas, se considera que se ha alcanzado una solución óptima cuando después de 10 generaciones continuas no se ha obtenido una mejora en lo que respecta a la aptitud de los individuos.

IV. EJEMPLO ILUSTRATIVO

Como caso de estudio se utiliza la organización aerolínea, cuya especificación es presentada en la figura 1.

```

W={EnviarVuelos(Fecha, Origen, Destino)↑↓*, Pagar(Precio)↑↓*□,
  Reservar(Vuelo-ID, Clase, Fecha)↑↓*, EnviarBoletoE(Vuelo-ID)↑↓*};
R = {Cliente(C), Aerolínea(A)} P = {Aerolínea} I = { ∅ }
K={Fecha, Clase, Vuelo-ID, Tarjeta, Precio, Origen, Destino}
M={
P1: {∅}
M1: M(C, A, SolicitarVuelos(Fecha, Origen, Destino))
E1: {CreateO(M1, O(A, EnviarVuelos(Fecha, Origen, Destino)), t)}
P2: {∅}
M2: M(C, A, AceptarVuelo(Vuelo-ID, Clase, Fecha))
E2: {CrearO(M2, O(C, Pagar(Precio)), t),
      CrearO(M2, O(A, Reservar(Vuelo-ID, Clase, Fecha)), t)}
P3: {HoldsAt(O(C, Pagar(Precio)), t) ∧
      HoldsAt(O(A, Reservar(Vuelo-ID, Clase, Fecha)), t)}
M3: M(C, A, RealizarPago(Precio, Tarjeta))
E3: {LiberarO(M3, O(C, Pagar(Precio)), t),
      CrearO(M3, O(A, EnviarBoletoE(Vuelo-ID)), t)}
P4: {HoldsAt(O(A, EnviarVuelos(Fecha,Origen, Destino)), t)}
M4: M(A, C, VuelosDisponibles(Lista))
E4: {LiberarO(M4, O(A, EnviarVuelos(Fecha, Origen, Destino)), t)}
P5: {HoldsAt(O(C, Pagar(Precio)), t) ∧
      HoldsAt(O(A, Reservar(Vuelo-ID, Clase, Fecha)), t)}
M5: M(A, C, ConfirmarReservacion(Vuelo-ID))
E5: {LiberarO(M5, O(A, Reservar(Vuelo-ID, Clase, Fecha)),t)}
P6: {HoldsAt(O(A, EnviarBoletoE(Vuelo-ID)), t)}
M6: M(A, C, EntregarBoletosE(Vuelo-ID))
E6: {LiberarO(M6, O(A, EnviarBoletoE(Vuelo-ID)), t)}
    
```

Figura 1. Organización: Aerolínea

Esta organización controla la interacción de los agentes para comprar boletos de avión. La

organización tiene dos roles: cliente y aerolínea. Una secuencia de mensajes válida es la siguiente: Como primer paso, el cliente solicita a la aerolínea una lista de vuelos (M_1) que cumplan con ciertos parámetros, entonces, la aerolínea regresa una lista de vuelos disponibles (M_4), filtrada por las restricciones del cliente, después, el cliente selecciona un vuelo y solicita su reservación (M_2), la aerolínea confirma la reservación (M_5) y espera la información de la tarjeta de crédito del cliente (M_3), finalmente, la aerolínea envía el boleto electrónico al cliente (M_6).

Los terminales están representados por los mensajes:

M_1 : M(C, A, SolicitarVuelos(Fecha, Origen, Destino))
 M_2 : M(C, A, AceptarVuelo(Vuelo-ID, Clase, Fecha))
 M_3 : M(C, A, RealizarPago(Precio, Tarjeta))
 M_4 : M(A, C, VuelosDisponibles(Lista))
 M_5 : M(A, C, ConfirmarReservacion(Vuelo-ID))
 M_6 : M(A, C, EntregarBoletosE(Vuelo-ID))

Para actualizar el estado de la interacción y poder evaluar las secuencias de mensajes, los mensajes son definidos como operadores, ejemplo el mensaje M_3 :

Operador:

M(C, A, RealizarPago(Precio, Tarjeta))

Precondiciones:

O(C, Pagar(Precio))

O(A, Reservar(Vuelo-ID, Clase, Fecha))

Efectos:

Añadir O(A, EnviarBoletoE(Vuelo-ID))

Quitar O(C, Pagar(Precio))

Un *caso evaluación*, para evaluar una secuencia de mensajes dentro de la organización presentada, podría ser cualquier estado de interacción que contenga la obligación EnviarBoletoE(Vuelo-ID) activa. Dentro del escenario planteado, esto significa que la aerolínea ha recibido el pago correspondiente por el vuelo. La otra restricción que debe ser provista, es el conjunto de obligaciones iniciales de los agentes, con el fin de establecer el estado origen de la interacción. Para el caso particular de la organización aerolínea, el conjunto de obligaciones iniciales es vacío.

Un posible individuo perteneciente a la población, generada aleatoriamente podría ser:

UNIR (EJECUTAR(M_5),
 UNIR(EJECUTAR(M_2),
 EJECUTAR(M_1)))

La secuencia de mensajes de este individuo es (M_5 , M_2 , M_1). Está es una secuencia invalida, ya que no cumple con las restricciones planteadas.

El siguiente individuo es una secuencia de mensajes valida:

UNIR (EJECUTAR(M_1),
 UNIR (UNIR(EJECUTAR(M_4),
 UNIR(EJECUTAR(M_2),
 EJECUTAR(M_5))),
 UNIR(EJECUTAR(M_3),
 EJECUTAR(M_6))))))

Esta secuencia de mensajes (M_1 , M_4 , M_2 , M_5 , M_3 , M_6) cumple con las restricciones planteadas, además de pasar el caso de evaluación, alcanzando la obligación meta.

El resultado de aplicar las métricas de aptitud para el primer individuo es de 6 puntos, 4 de los cuales son debido a la presencia de dos funciones *UNIR* y los 2 restantes debido a que dos funciones *EJECUTAR* regresan un valor distinto a nulo. Para el segundo individuo se obtienen 30 puntos. La mayor puntuación se debe a que la secuencia de mensajes constituye una solución al problema planteado.

V. CONCLUSIONES Y TRABAJO FUTURO

Con el caso de estudio presentado en este artículo, se mostró como la programación genética puede ser usada para generar los protocolos de interacción, a partir de un conjunto de mensajes, cada uno de los cuales tiene precondiciones y efectos, expresados en obligaciones.

Un protocolo de interacción requiere de la intervención de varios agentes, por tanto,

normalmente la generación automática del protocolo es llevada a cabo como un proceso centralizado, el cual es realizado por un solo agente, que posteriormente provee las secuencias de mensajes válidas a los demás participantes, sin embargo, esto contradice el paradigma distribuido de los sistemas multi-agente, donde no existe un control central.

El método mostrado en este artículo, dota a los agentes de la capacidad para generar una posible secuencia de mensajes, tomando como punto de referencia sus propias metas; la generación de las secuencias de mensajes, para dos agentes que participan dentro de una organización, puede no ser la misma, ya que cada uno de ellos tiene diferentes metas, sin embargo comparten el dominio de interacción que provee la organización, esto permite que al encontrarse en situaciones donde el agente no reciba el mensaje esperado, replantee su secuencia de mensajes en base a la nueva información que provee el mensaje imprevisto. Inclusive, en algunos casos el estado de la interacción podría permitir todas las secuencias de mensajes promovidas por los participantes, sin importar que sean diferentes.

El uso de la programación genética para la obtención de los protocolos de interacción es sólo justificado en escenarios complejos, donde exista una gran cantidad de posibles secuencias, debido a que para secuencias de mensajes pequeñas, métodos tradicionales de planeación encadenada hacia atrás, podrían resultar más efectivos.

El trabajo futuro está dirigido a considerar restricciones de tiempo dentro de las precondiciones, además de la inclusión de una relación de preferencia sobre las obligaciones que permita evaluar con más detalle a los individuos de una población.

La implementación de la plataforma de sistemas multi-agente, dentro de la cual se puede llevar a cabo la interacción descrita en este artículo, así como la generación de los protocolos de interacción mediante el uso de la programación genética, se encuentra en fase de desarrollo.

VI. AGRADECIMIENTOS

Este trabajo está parcialmente apoyado por el CoECyT-Jal proyecto No. 2008-05-97094, mientras que sus autores Gutiérrez-García José Octavio y González-Padilla Omar Alfredo cuentan con beca de CONACYT No. 191493 y 203003, respectivamente.

VII. REFERENCIAS

- [1] Artikis, A., Sergot, M., y Pitt, J.: Specifying Norm-Governed Computational Societies, *ACM Transactions on Computational Logic*, 2007.
- [2] Carabelea, C., y Boissier, O.: Coordinating Agents in Organizations Using Social Commitments. *Electronic Notes in Theoretical Computer Science*, 150(3):73-91, 2006.
- [3] Fornara, N. y Colombetti, M. Specifying and enforcing norms in artificial institutions. In Boella, G., van der Torre, L., and Verhagen, H., editors, *Normative Multi-agent Systems*, number 07122 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany, 2007.
- [4] Gutiérrez-García, J.O., Koning, J.-L., y Ramos-Corchado, F.F. From obligations to organizational structures in multi-agent systems. In proceedings of PRIMA 2008, ed. Bui, The Duy et al., Springer LNAI no. 5357, pp. 206-213, Hanoi, Vietnam, 2008.
- [5] Koza, J. R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: The MIT Press. 1992.
- [6] Mallya, A. U., y Singh, M. P.: An algebra for commitment protocols. *Autonomous Agents and Multi-Agent Systems*, 14(2):143-163, 2007.

- [7] Quenum, J. G., Slodzian, A., y Aknine, S. Automatic derivation of agent interaction model from generic interaction protocols. In AOSE, 2003, pp. 138–152.
- [8] Quenum, J. G., Aknine, S., y Slodzian, A. An approach to configure interactions from generic protocols, October 2004.
- [9] Quenum, J. G., Aknine, S., Briot, J.-P., y Honiden, S. A modeling framework for generic agent interaction protocols. In Workshop on Declarative Agent Languages and Technologies, May 2006.
- [10] Winikoff, M. Implementing commitment-based interactions. In E. H. Durfee, M. Yokoo, M. N. Huhns, and O. Shehory, editors, 6th International Joint Conference on Autonomous Agents and Multiagent Systems, IFAAMAS, 2007.
- [11] Zlatev, Z., Diakov, N., y Pokraev, S. Construction of negotiation protocols for e-commerce applications. SIGecom Exch., 5(2):12–22, 2004.

VIII. AUTORES

José Octavio Gutiérrez-García, es maestro en ciencias por el CINVESTAV-IPN Guadalajara y estudiante de doctorado del CINVESTAV-IPN Guadalajara en co-tutela con el “Grenoble Institute of Technology”. Sus áreas de interés son protocolos de interacción, normas sociales y sistemas multi-agente.

Omar Alfredo González-Padilla, es maestro en ciencias por el CINVESTAV-IPN Guadalajara y estudiante de doctorado del CINVESTAV-IPN Guadalajara en co-tutela con la “Université de Technologie de Compiègne”. Sus áreas de interés son la capitalización del conocimiento y los sistemas multi-agente.

MODELADO Y SIMULACIÓN DE SISTEMAS MULTI-AGENTES ORIENTADOS A EVENTOS

López-Neri Emmanuel, López-Mellado Ernesto, Ramírez-Treviño Antonio.

CINVESTAV, Unidad Guadalajara
Av. Científica 1145, 45015 Zapopan, Jal. México
e-mail: {eneri, elopez & art}@gdl.cinvestav.mx

RESUMEN

En este artículo se presenta una metodología para el modelado y simulación de sistemas multi-agentes orientados a eventos (SMAOE). Ésta permite la integración de los dos enfoques más importantes de simulación: el de sistemas de eventos discretos (SED) y sistemas basados en multi-agentes (SBMA). La metodología permite obtener modelos jerárquicos, modulares y flexibles, características requeridas en el modelado y simulación de sistemas dinámicos complejos. Se utiliza el formalismo de redes de Petri n-LNS para describir el ambiente, la estructura del agente y sus actividades. Además, permite incorporar al modelo del agente un proceso de toma de decisiones y la correspondiente implementación de las redes n-LNS para su simulación.

Palabras Clave: Modelado, Orientado a Eventos, Multi Agente, Redes de Petri, n-LNS.

I. INTRODUCCIÓN

El modelado de sistemas con entidades que pueden percibir el ambiente que les rodea, y además actuar en él de acuerdo a cierto conocimiento o experiencia; es el dominio de estudio de los sistemas basados en multi-agentes (SBMA). Este enfoque permite el estudio del comportamiento dinámico de sistemas complejos a nivel microscópico (como aquellos en los que el ser humano se encuentra involucrado) y el estudio de comportamientos a nivel macroscópico que surgen de la interacción de múltiples agentes [1].

Una de las categorías más importantes de los sistemas basados en multi-agentes son los SBMA espaciales. Consisten en una cuadrícula que representa un ambiente geoméricamente distribuido, una colección de entidades móviles y un conjunto de reglas para actualizar ambos componentes. Algunos de los SBMA integran el enfoque de sistemas de eventos discretos (SED) [2], de manera que las acciones de los agentes son descritas como eventos. Ésta integración permite obtener en un mismo modelo las ventajas de ambos enfoques.

SWARM [3] fue el primer modelo desarrollado bajo este enfoque de integración y fue escrito en Objective-C. Actualmente existen diferentes versiones de SWARM, algunas escritas en Java [4] u otros lenguajes, que son utilizadas en el modelado de SBMA espaciales [5][6]. La herramienta de modelado REPAST [7] inició como una implementación Java de SWARM, sin embargo, se apartó progresivamente de la idea original empleando los autómatas celulares para la representación del ambiente de los agentes. Wagner observó que el trabajo de SWARM no permite simbolizar el modelo mental del agente (modelo de toma de decisiones), entonces utilizando el lenguaje UML propone el uso de un meta modelo para la simbolización del modelo mental del agente [8]. JAMES [9], por el contrario, utiliza un formalismo para la representación de la parte discreta empleando DEVS [10], el cual es un formalismo ampliamente manejado por la comunidad de simulación. El sistema SPADES [11], se enfoca en proveer un middleware diseñado bajo el enfoque de SED sobre el cual los agentes están montados, que permite la distribución de éstos. Así mismo,

MASON [12] es una librería con énfasis en la ejecución eficiente sobre la máquina virtual de Java. Theodoropulos y Logan [13] propusieron una arquitectura para acelerar la simulación de SBMA adoptando técnicas de la comunidad de simulación paralela y distribuida, tales como la sincronización optimista, afirmando que es la más adecuada para la simulación de SBMA, sin embargo la mayor parte de su trabajo se centra en cuestiones de representación espacial y manejo del espacio compartido.

La mayoría de las plataformas SBMA siguen un paradigma basado en el uso de bibliotecas, donde un lenguaje de programación huésped es extendido a través de procedimientos genéricos para facilitar el desarrollo de SBMA. En este enfoque la biblioteca incluye un marco de desarrollo (un conjunto de conceptos para el diseño de SBMA) junto con las herramientas para la simulación y visualización. Sin embargo al usar este paradigma, un concepto de diseño y metodologías apropiadas para la integración de los SBMA y SED están ausentes.

De los modelos estudiados y que han abordado la integración de los SBMA y SED, podemos establecer tres características requeridas del concepto de diseño para este tipo de modelos:

- Debe ser flexible, es decir, no estar confinado a estructuras exclusivas de SBMA o SED.
- Debe soportar modularidad y reusabilidad para poder beneficiar investigaciones futuras.
- No debe eliminar alguno de los beneficios, dados por SED (simplicidad y manejo de eventos eficientes) y SBMA (modelado de comportamiento complejos).

Además se requieren de herramientas de diseño simples y robustas que soporten ambos enfoques.

En el presente artículo se propone una metodología para el modelado y simulación de sistemas multi-agentes orientados a eventos, que cumple con las características necesarias en el concepto de diseño de un modelo de éste tipo. Se utiliza el formalismo n-LNS de redes de Petri para describir la estructura del sistema en tres niveles de abstracción. El primer nivel describe la red de tráfico; el segundo nivel modela la estructura de las entidades móviles o agentes, y el tercer nivel especifica las actividades del agente.

II. EL FORMALISMO N-LNS

El formalismo sigue el enfoque de redes anidadas introducido por R. Valk [14], en el cual es propuesto un esquema de dos niveles de redes llamado SOE (Sistema de Objetos Elementales). Una extensión a la propuesta de Valk, nombrada n-LNS, es realizada por Sánchez [15] donde muestra una definición mucho más detallada del formalismo.

A. DEFINICIÓN

Un modelo n-LNS consiste principalmente de un número arbitrario de redes organizadas en n niveles de acuerdo a una jerarquía; el valor de n depende del grado de abstracción deseado en el modelo. Una red puede tener como marcas símbolos y redes de niveles más profundos; las redes de nivel n permiten como marcas solo símbolos, similar a las RPC (redes de Petri coloreadas). Las interacciones entre redes son declaradas a través del etiquetado simbólico de las transiciones.

La Fig. 1 muestra un fragmento de los componentes de un modelo 4-LNS. El nivel 1 es representado por la red NET_1 , el nivel 2 por las redes $NET_{2,1}$ y $NET_{2,2}$, las redes $NET_{3,1}$, $NET_{3,2}$, $NET_{3,3}$, y $NET_{3,4}$ componen el nivel 3, y las redes $NET_{4,1}$, $NET_{4,2}$, $NET_{4,3}$ forman el nivel 4.

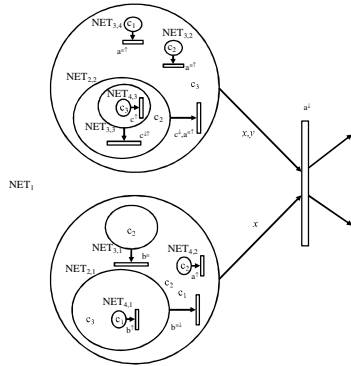


Fig. 1: Componentes de un modelo 4-LNS

Una red de nivel i es una tupla $NET_i = (typenet_i, \mu_i)$, la cual está compuesta por una estructura PN, el peso de los arcos ($\pi((p, t), lab)$ o $\pi((t, p), lab)$) expresado como multi-conjuntos de variables y símbolos, y una función de etiquetado de transiciones definiendo la interacción entre redes. μ_i es la función de marcado (ver Fig. 2).

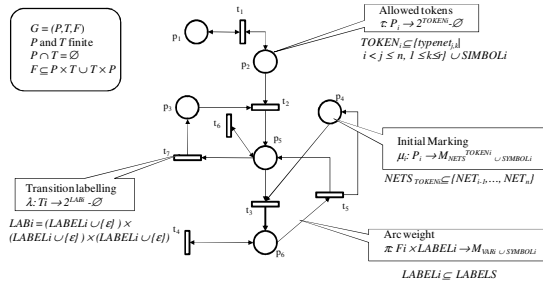


Fig. 2: Ejemplo de una red de nivel i , con el formalismo n-LNS definida como $Net_i = (typenet_i, \mu_i)$, donde $typenet_i = (G, TOKEN_i, LABEL_i, VAR_i, \tau, \lambda, \pi)$.

Un modelo n-LNS, llamado sistema de red, es una n -tupla $NS = (NET_1, NET_2, \dots, NET_n)$ donde NET_1 es la red de nivel más alto, y $NET_i = \{NET_{i,1}, NET_{i,2}, \dots, NET_{i,r}\}$ es un conjunto de r redes de nivel i .

Los componentes de un modelo pueden interactuar entre ellos a través de la sincronización de transiciones. El mecanismo de sincronización es

incluido en las reglas de habilitación y disparo de las transiciones; este establece que dos o más transiciones etiquetadas con el mismo símbolo deberán estar sincronizadas. Una etiqueta puede tener los atributos $\equiv, \downarrow, \uparrow$, los cuales expresan respectivamente sincronización local, interna y externa.

B. DISPARO Y HABILITACIÓN DE TRANSICIONES

Una transición t de una red de nivel i NET_i está habilitada con respecto a la etiqueta lab si: existe un ligado b_t que asocie el conjunto de variables que aparecen en todo $\pi((p, t), lab)$; debe de cumplirse que $\forall p \in \bullet t, \pi((p, t), lab)_{\langle b_t \rangle} \subseteq \mu_i(p)$. ($\langle b_t \rangle$ no es necesario cuando el nivel de la red es n); y las condiciones de uno de los siguientes casos se cumple:

Caso 1. Si no hay atributos entonces el disparo de t es ejecutado de manera autónoma.

Caso 2. Si lab tiene atributos se debe considerar la combinación de las siguientes situaciones:

- $\{\equiv\}$ Se requiere la habilitación simultánea de las transiciones etiquetadas con lab^\equiv que pertenecen a otras redes que se encuentran en el mismo lugar p del siguiente nivel de red superior. El disparo de estas transiciones es simultáneo y todas (localmente) las redes sincronizadas se mantienen en p .
- $\{\downarrow\}$ Se requiere la habilitación de las transiciones etiquetadas con lab^\downarrow pertenecientes a otras redes de nivel inferior dentro de $\bullet t$. Estas transiciones se disparan simultáneamente y las redes de menor nivel y los símbolos declarados por $\pi((p, t), lab)_{\langle b_t \rangle}$ son removidos.
- $\{\uparrow\}$ Se requiere la habilitación de al menos una de las $t \in p \bullet$, etiquetadas con lab^\uparrow , de la red superior donde NET_i está contenida. El disparo de t provoca la transferencia de la red NET_i y de los símbolos declarados en $\pi((p', t'), lab)_{\langle b_t \rangle}$.

El disparo de transiciones en todos los niveles de red modifica el marcado removiendo $\pi((p, t), lab)_{<bt>}$ en todos los lugares de entrada y agregando $((t, p), lab)_{<bt>}$ a todos los lugares de salida

En la Fig. 1, NET_1 está sincronizada a través de la transición etiquetada como a^\downarrow con $NET_{2,2}$, $NET_{3,2}$, $NET_{3,4}$ y $NET_{4,2}$ por medio de las transiciones (localmente sincronizadas) etiquetadas como a^\uparrow ; todas estas transiciones deberán estar habilitadas para poderse disparar. El disparo simultáneo de las transiciones remueve estas redes de los lugares de entrada.

Las redes $NET_{2,1}$, $NET_{3,1}$ y $NET_{4,1}$ están sincronizadas a través de las transiciones etiquetadas con b^\downarrow , b^\equiv , b^\uparrow respectivamente; el disparo de las transiciones cambia el marcado de $NET_{2,1}$ y $NET_{3,1}$. La red $NET_{4,1}$ es removida del lugar de $NET_{2,1}$. La red $NET_{3,3}$ es removida del lugar de entrada de $NET_{2,2}$ y $NET_{4,3}$ es removida de $NET_{3,3}$; esta interacción es establecida por c^\downarrow , c^\uparrow , c^\uparrow , respectivamente.

III. DEFINICIÓN DE SIMULACIÓN BASADA EN MULTI-AGENTE

Un SBMA es un sistema complejo donde varias entidades interactúan con objetivos individuales bajo ciertas reglas de evolución y siguiendo ciertas políticas de comportamiento del ambiente en el que se desenvuelven. El ambiente requiere además ser definido como abstracción de primera clase (no solo de manera explícita en el modelo) ya que proporciona a los agentes las condiciones de entorno necesarias para su existencia, siendo una parte esencial para la definición de SBMA espaciales [16]. Por consiguiente, podemos definir un SBMA como: $SBMA = \{Amb, Agt, Obj, RE, PC\}$. En la Tabla 1 se listan los componentes de un SBMA y sus respectivas definiciones.

Componente	Definición
<i>Amb</i>	El ambiente es el lugar donde los agentes residen. Este es un límite que permite definir qué es lo que está dentro y que es lo que está fuera del sistema a modelar. El ambiente es parte importante en la definición de un SBMA, ya que debe existir como clase de abstracción de primer orden, es decir no solo de manera implícita.
<i>Agt</i>	Los agentes son entidades autónomas, participando activamente en el sistema. Estos tienen comportamiento y toma de decisiones para lograr sus objetivos individuales. La acción de los agentes resultado de la toma de decisiones pueden incluir alguna de las siguientes actividades: movimiento, sentido y/o modificación del ambiente. Los agentes pueden diferir de otros en sus comportamientos, objetivos, la forma en la que observan el ambiente, diferentes opciones de comportamiento y formas de actuar en el ambiente.
<i>Obj</i>	Los objetos son las partes no autónomas del ambiente. Los objetos pueden ser desde recursos, herramientas, dispositivos, conceptos hasta inclusive categorías. Los objetos son utilizados por los agentes o modificados por los agentes.
<i>RE</i>	Reglas de Evolución. Estas son las reglas base que gobiernan el ambiente del SBMA. Estas reglas son las restricciones que los agentes deben obedecer y no se pueden romper.
<i>PC</i>	Políticas de Comportamiento. Las políticas establecen los mecanismos que rigen el comportamiento de los agentes en el ambiente. Estas pueden ser ignoradas por el agente de acuerdo a atributos o parámetros de éste.

Tabla 1: Definiciones y Componentes del SBMA.

IV. DEFINICIÓN DE SISTEMA DE EVENTOS DISCRETOS

En un SED los cambios en el estado del sistema suceden en intervalos de tiempo no constantes, llamados eventos. Law [17] define el estado de un sistema como una colección de variables necesarias para describir un sistema en un instante en particular. Podemos definir entonces a un SED como el modelado de los eventos que realizan un cambio en las variables de estado del sistema. La estructura de un modelo SED consiste de cuatro elementos como se muestra en la Tabla 2. Estos elementos forman el modelo básico de un SED [18].

Elementos del SED	Definición
Parámetros	VARIABLES de entrada en el sistema pero que no cambiarán en cada intervalo discreto.
VARIABLES de Estado	VARIABLES representando el estado del sistema. Estas pueden cambiar a cada intervalo discreto.
Eventos	Ocurrencias interesantes del sistema.
Relaciones de programación (calendarización)	Conexión de programación entre eventos en el sistema.

Tabla 2: Estructura de un modelo SED.

v. MODELADO Y SIMULACIÓN DE SISTEMAS MULTI-AGENTES ORIENTADOS A EVENTOS (SMAOE)

Un sistema bajo el enfoque de SMAOE utiliza el formalismo de redes de Petri n-LNS para definir la estructura del sistema en tres niveles de abstracción. El primer nivel detalla la estructura física del sistema o ambiente; el segundo nivel modela la estructura de las entidades móviles o agentes, y el tercer nivel especifica las actividades del agente. En la Fig. 3 se pueden apreciar los tres niveles de abstracción utilizados en el modelo.

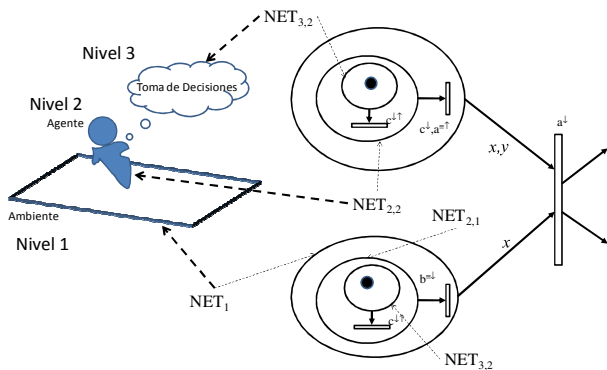


Fig. 3: Representación de un sistema usando 3-LNS

Cada lugar de la red de nivel superior tiene asociada además una descripción de segmento, el cual contiene los objetos del sistema requerido en la definición previa de un SBMA. El modelo de un SMAOE queda definido de la siguiente manera:

Definición 1. El modelo del SMAOE es una 3-tupla, definida por $SMAOE = (3-LNS, \{S_i\}, \{Agt_i\})$, donde:

- 3-LNS es la definición del ambiente y agentes descrita por el formalismo n-LNS de tres niveles (ver Fig. 3).
- $\{S_i\}$ representa el conjunto de descripciones de segmentos del sistema.
- $\{Agt_i\}$ representa el conjunto de descripciones de agentes del sistema.

El modelo deberá cumplir con las siguientes características:

- La movilidad de los agentes debe asegurarse sincronizando transiciones de las redes de tipo $ActivityNET_3$ con las redes del tipo $AgentNET_2$ y ésta a su vez con la red del tipo $EnvironmentNET_1$.
- No deberán existir segmentos aislados (la red de tipo $EnvironmentNET_1$ es conexa)

Utilizaremos el siguiente caso de estudio para presentar el modelo de un SMAOE. En este ejemplo, un agente realiza la búsqueda de un objeto en un mundo simulado a partir de la interacción con otro agente y objetos informativos [21]. Se modela a un individuo intentando obtener la ruta para llegar al objeto buscado. La Fig. 4 muestra las principales características del ejemplo. El agente entra al modelo de forma aleatoria, con la base de conocimientos, atributos y parámetros necesarios para encontrar las fuentes de información, interactuar con ellas y obtener así, la ubicación del objeto buscado. Una vez que el agente haya obtenido la información de un objeto fuente, navegará por el mundo simulado para encontrar el objeto. Hay dos métodos de transporte para poder llegar al objeto buscado; si el agente utiliza un objeto estacionario (mapa) se puede caminar hasta la ubicación del objeto buscado que le llevará aproximadamente unos 10 minutos, dependerá de la velocidad del agente para caminar. Si el agente obtiene la ruta a partir de otro agente, utilizará el tranvía que se encuentre

disponible. El tranvía llega a las estaciones cada 5 minutos y dura 1.5 minutos en trasportar un agente a la siguiente estación.

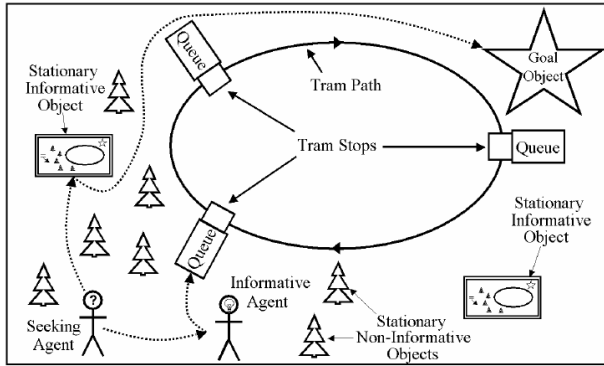


Fig. 4: Escenario utilizado como caso de estudio[21]

C. OBJETOS

Definición 2. La descripción de un objeto obj_i es un 3-tupla definida por $obj_i = (type_i, value_i, w_i)$ donde:

- $type_i$ es el tipo de objeto que se describe y que será identificado por el agente.
- $value_i$ es el valor del objeto i
- w_i es la posición relativa del objeto en el ambiente

Los objetos identificados del caso de estudio son los mapas indicando ubicaciones de las estaciones y puntos de interés en el mundo simulado, por ejemplo:

$$o_1 = ("map", "next_leon", 0.0)$$

$$o_2 = ("map", "next_estacion1", 0.0)$$

$$o_3 = ("map", "en_estacion1", 0.0)$$

D. MODELO DEL AGENTE

Las variables de estado (propias del SED) se capturan en el agente como atributos, y se representan por el conjunto $ATTR_i$. Por ejemplo,

los atributos del agente i están definidos por el conjunto $ATTR_i = \{velocidad, hambre, humor, agresividad, posición\}$. Los parámetros (propios del SED) son representados por el conjunto $PARAM_i$. Por ejemplo, los parámetros del agente i se definen por el conjunto $PARAM_i = \{tiempo_retardo, capacidad, tamaño, color, plan\}$.

En la descripción de un SBMA, el agente debe seguir ciertas políticas de comportamiento (PC), las cuales pueden ser ignoradas por éste de acuerdo a ciertos parámetros que involucren este tipo de decisión. Por ejemplo, cada que el agente observa un objeto informativo (mapa), deberá interactuar con él. En la Tabla 3 se muestra la implementación de esta PC.

ID Regla	Condición	Acción
R ₁	IF $spatialSituation == enEstación_1$ AND $Event == llegaTren$ AND $temporalRelation == después$	THEN $abordarTren(agt_i, arrivalTime, agt_i, currentPosition + agt_i, traveledDistance + agt_i, safeDistance, segment_i, agt_i),$ $bajarTren (Tren.evTime + tren_i, getDelayToStart, segment_i, agt_i)$

Tabla 3: Ejemplo de captura de una PC

Las reglas de evolución (RE) son restricciones físicas impuestas por características propias del sistema y por lo tanto, no pueden ser ignoradas por el agente. Por ejemplo, si en el camino del agente se atraviesa otro agente, el primero deberá re-evaluar su evento futuro previamente generado (ver Tabla 4).

Las PC y las RE son capturadas en una base de conocimiento y junto con los atributos y parámetros describen a un agente en particular utilizando la siguiente definición:

Definición 3. La descripción de un agente agt_i es un 3-tupla definida por $agt_i = (ATTR_i, PARAM_i, knowledgeBase_i)$ donde:

- $ATTR_i$, es un conjunto de atributos del agente.
- $PARAM_i$ es un conjunto de parámetros del agente
- $knowledgeBase_i$ es la base de conocimiento del agente

Rule ID	Condición	Acción
R _{1A}	IF $spatialSituation[actor_i]$ $== trafficLight$ AND $Event[actor_i]$ $== yellowToRed$ AND $temporalRelation[actor_i,$ $yellowToRed] == After$	THEN $stop(actor_i, arrivalTime,$ $actor_i, currentPosition +$ $actor_i, traveledDistance +$ $actor_i, safeDistance, segment_i, actor_i),$ $start(Event[actor_i], time +$ $actor_i, getDelayToStart, segment,$ $actor_i)$

Tabla 4: Ejemplo de captura de una RE

La red NET_{3,1} (ver Fig. 5) captura el comportamiento discreto del agente [19] y se le asocia la descripción del agente agt_i . La Tabla 5, muestra la descripción de la red NET_{3,1}. Se han agregado dos funciones al formalismo n-LNS para la descripción de la red tipo $ActivityNet_3$: σ y φ .

Definición 4. $\sigma: T \rightarrow LABEL_{i,k} \times ATTR_i$ es una función de asignación de atributos del agente a cada evento de la red NET_{3,1}, lo que permite saber que atributo es modificado cuando dicho evento es ejecutado, donde:

- $LABEL_{i,k}$ es un conjunto finito de etiquetas definidas para la red tipo k de nivel i.
- $ATTR_i$ es el conjunto de atributos del agente i.

Definición 5. $\varphi: T \rightarrow LABEL_{i,k}$ es una función de asignación de inicio de toma de decisión del agente a algunos eventos de la red NET_{3,1},

La toma de decisiones, es decir el mapeo de percepciones en acciones, involucra la descripción propia del agente agt_i y el estado de otros agentes en el ambiente. En algunos casos, los agentes pueden tener sistemas de percepción de eventos muy simples que disparen múltiples eventos de acción o puede involucrar procesos de toma de decisión sofisticados. El uso de este modelo no

limita la estructura del modelo de toma de decisiones que el agente use. En general, para que dicho modelo pueda ser utilizado, el sistema de inteligencia que sea modelado deberá involucrar una interfaz de entrada y una de salida que permita la interacción con el ambiente. Por ejemplo, pueden ser usados los sistemas expertos, redes neuronales, lógica difusa, etc.

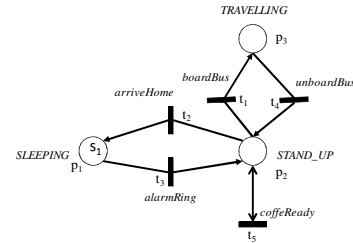
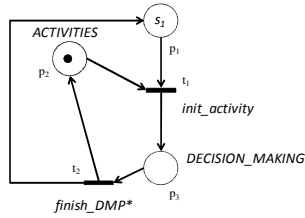


Fig. 5: Actividades del agente “Empleado” descritas por la red NET_{3,1}

NET _{3,1}
TOKEN _{3,1} = {s ₁ }; LABEL _{3,1} = {arriveHome, boardBus, unBoardBus, alarmRing, coffeReady}; VAR _{3,1} = {∅}
λ
$\lambda\{t_1\} = \{(boardBus, \uparrow)\}, \lambda\{t_2\} = \{(arriveHome, \uparrow)\}, \lambda\{t_3\} = \{(alarmRing, \uparrow)\},$ $\lambda\{t_4\} = \{(unBoardBus, \uparrow)\}, \lambda\{t_5\} = \{(coffeReady, \uparrow)\}$
τ
$\tau(p_1) = \tau(p_2) = \tau(p_3) = \{s_1\}$
π
$\pi((p_1, t_3), alarmRing^\uparrow) = \pi((t_3, p_2), alarmRing^\uparrow) = \pi((p_2, t_2), arriveHome^\uparrow) = \pi(($ $t_2, p_1), arriveHome^\uparrow) = \pi((p_2, t_5), coffeReady^\uparrow) = \pi((t_5, p_2), coffeReady^\uparrow) =$ $\pi((p_2, t_1), boardBus^\uparrow) = \pi((t_1, p_3), boardBus^\uparrow) = \pi((p_3, t_4), unBoardBus^\uparrow) = \pi((t_4,$ $p_2), unBoardBus^\uparrow) = s_1$
σ
$\sigma\{t_1\} = \{(boardBus, posición)\}, \sigma\{t_5\} = \{(coffeReady, humor)\}$
φ
$\varphi\{t_3\} = \{(alarmRing)\}, \varphi\{t_2\} = \{(arriveHome)\}, \varphi\{t_4\} = \{(unBoardBus)\}$

Tabla 5: Funciones λ, τ y π de la red NET_{3,1}.

En el nivel dos del modelo nombrado $AgentNET_2$ (ver Fig. 6), se provee de la estructura necesaria para que el agente pueda desplazarse sincronizándose con la red $EnvironmentNet_1$ y las redes del tipo $ActivityNET_3$.



*DMP: Decision Making Process

Fig. 6: Modelo del agente descrita por la red NET_{2,1}

NET _{2,1}
TOKEN _{2,1} = {s ₁ , Net _{3,1} }
LABEL _{2,1} = { arriveHome, boardBus, unBoardBus, alarmRing, coffeReady } ;VAR _{2,1} = {x : typeNet _{3,1} }
λ
$\lambda\{t_1\} = \{(arriveHome, \cdot), (alarmRing, \cdot), (unBoardBus, \cdot)\}$
τ
$\tau(p_2) = \tau(p_3) = \{ActivityNet_{3,1}\}, \tau(p_1) = \{s_1\}$
π
$\pi((p_2, t_1), arriveHome) = \pi((p_2, t_1), alarmRing) = \pi((p_2, t_1), unBoardBus) = \pi((t_1, p_3), arriveHome) = (t_1, p_3), alarmRing) = (t_1, p_3), unBoardBus) = \{x\}$

Tabla 6: Funciones λ, τ y π de la red NET_{2,1}

E. MODELO DEL AMBIENTE

La red del primer nivel del modelo es nombrada *EnvironmentNET₁*. Cada lugar de esta red representa una unidad de modelado básica llamada segmento. En contraste con la definición de espacio discreto utilizada por los autómatas celulares, la cual divide cada segmento en celdas y permite al agente moverse de celda en celda; el espacio de un segmento en el modelo es de tipo continuo, es decir, la posición de cada agente en el sistema es definida por un número de punto flotante.

El ambiente se define como un conjunto interconectado de segmentos. Su principal función es la de contener a los agentes y objetos.

Definición 6. La descripción de un segmento s_i es un 2-tupla definida por $s_i = (O_i, type_i)$ donde:

- O_i es el conjunto de objetos en el segmento i .

- $type_i$ representa el tipo de segmento, este será utilizado por el agente para descubrir los permisos que tiene sobre este segmento.

Los segmentos que están ubicados en los bordes del sistema son llamados segmentos fuente y pozo, respectivamente en estos se describe la entrada y salida de agentes del sistema.

Los siguientes pasos nos permiten establecer la conexión de la red NET₁ y la descripción s_i para cada lugar de la red; el marcado inicial es definido con los objetos y agentes que inician en la red.

1. Para cada segmento s_i , un lugar p_i es asignado.
2. Una transición t_{ij} es agregada por cada par de segmentos (s_i, s_j) físicamente relacionados (existe una conexión física que permita a los agentes moverse entre estos segmentos), junto con los arcos $(p_i, t_{ij}), (t_{ij}, p_j)$ y $(p_j, t_{ji}), (t_{ji}, p_i)$.
3. Agregar una transición t_i por cada segmento s_i fuente o pozo; arcos (t_i, p_i) o (p_i, t_i) son agregados respectivamente.
4. Las transiciones que provocan un cambio de segmento se les sincronizará con la red NET₂. Por ejemplo, la transición t_3 con el evento *displace* mueve al agente a otro segmento entonces la función $\pi((p_1, t_3), displace)$ será agregada (ver Tabla 7).

Las marcas contenidas en la red ambiente son redes de tipo *AgentNET₂*. En la Tabla 7 se resumen los componentes del modelo del ambiente.

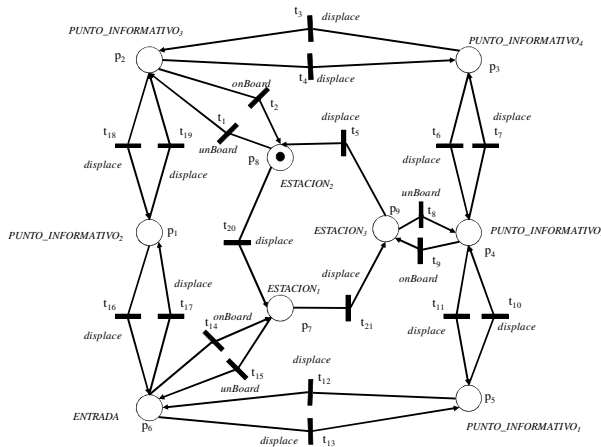


Fig. 7: Modelo de la red ambiente

NET ₁
TOKEN ₁ = { NET _{2,1} }
LABEL ₁ = { displace }
VAR _{1,1} = { x : NET _{2,1} }
λ
$\lambda\{t_1\}=\lambda\{t_2\}=\lambda\{t_3\}=\lambda\{t_4\}=\{(displace, 1)\}$
τ
$\tau(p_1) = \tau(p_2) = \tau(p_3) = \tau(p_4) = \{NET_{2,1}\}$
π
$\pi((p_1,t_3), displace^1) = \pi((p_4,t_4), displace^1) = \pi((p_3,t_2), displace^1) = \pi((p_2,t_1), displace^1) = \pi((t_1,p_1), displace^1) = \pi((t_2,p_2), displace^1) = \pi((t_3,p_4), displace^1) = \pi((t_4,p_3), displace^1) = \{x\}$

Tabla 7: Funciones λ, τ y π de la red ambiente NET₁

VI. CONCLUSIONES

Se presentó una Metodología de Modelado Orientada a Eventos Multi-Agente (MOEMA). El modelo base de entidades y estructuras que se obtiene de la metodología utiliza el formalismo de redes de Petri de tres niveles (3-LNS), para capturar el comportamiento de las entidades dinámicas, así como sus relaciones en un sistema, logrando simplicidad y modularidad en el diseño. Además, la descripción del modelo NET₃ permite seleccionar el nivel de detalle microscópico deseado del comportamiento del agente.

VII. REFERENCIAS

[1] E. Bonebeau, "Agent-based modeling: Methods and techniques to simulate human systems," PNAS, vol. 99, no. 3, pp. 7280-7287, 2002.

[2] B. ZEIGLER, T. Kim, and H. PRAEHOFER, *Theory of Modeling and Simulation*. Academic Press., 2000.

[3] N. Minar, R. Burkhart, C. Langton, and M. Askenazi, "The Swarm simulation system: a toolkit for building multi-agent simulations," Santa Fe Insitute Working Paper 96-06-042, 1996.

[4] S. D. Group. (2004, Aug.) Swarm 2.2. [Online]. <http://wiki.swarm.org>

[5] M. Batty, J. Desyllas, and E. Duxbury, "Safety in Numbers? Modelling crowds and designing control for the Notting Hill carnival," Urban Studies, vol. 40, no. 8, pp. 1573-1590, 2003.

[6] T. Schelhorn, D. O'Sullivan, M. Hakley, and M. Thurstain-Goodwin, "STREETS: An agent-based pedestrian model," Centre for Advanced Spatial Analysis (University College London) Working Paper 9, 1999.

[7] M. J. North, N. T. Collier, and J. R. Vos, "Experiences in creating three implementations of the Repast agent modeling toolkit," ACM Transactions on Modeling and Computer Simulations, vol. 16, no. 1, pp. 1-25, 2006.

[8] G. Wagner, "The agent-object-relationship metamodel: towards a unified view of the state and behavior," Information Systems, vol. 28, no. 5, 2003.

[9] B. Shattenberg and A. M. Uhrmacher, "Planning agents in JAMES," in Proceedings of the IEEE, vol. 89, 2001, pp. 158-173.

- [10] G. Wainer and N. Giambasi, "Specification, modeling and simulation of timed Cell-DEVS models," Comp. Dpto., FCEN/UBA, Technical Report, 1998.
- [11] P. Riley, "SPADES: System for parallel agent discrete event simulation," AI Magazine, vol. 24, no. 2, pp. 41-42, 2003.
- [12] S. Luke, C. Cioffi-Revilla, K. Sullivan, and G. Balan, "MASON: A multiagent simulation environment," Simulation, vol. 81, no. 7, pp. 517-527, 2005.
- [13] G. Theodoropoulos and B. Logan, "A framework for the distributed simulation of agent-based systems," in *Proceedings of the 13 th SCS European Simulation MultiConference*, 1999, pp. 58-65.
- [14] R. Valk, "Petri nets as token objects. An introduction to elementary object nets," in *International Conference on application and Theory of Petri nets*, 1998, pp. 1-25.
- [15] R. Sánchez-Herrera and E. López-Mellado, "Modular and Hierarchical Modeling of Interactive Mobile Agentes," in *IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, 2004, pp. 1740-1745.
- [16] A. Weyns, A. Omicini, and J. Odell, "Environment as a first class abstraction in multi-agent systems," vol. 14, no. 1, pp. 5-30, Feb. 2007.
- [17] A. M. Law, *Simulation modeling and analysis*, 4th ed. McGraw-Hill, 2007.
- [18] T. R. Boon-Leng, "A study to model human behavior in discrete event simulation (DES) using SIMKIT," Master's Thesis, Naval Postgraduate School, Monterey, California, 2007.
- [19] E. Lopez-Neri, E. Lopez-Mellado, and A. Ramirez-Treviño, "Microscopic Modeling Framework for Urban Traffic Systems Simulation," in the 7th International conference on system simulation and scientific computing, Beijing, China, 2008.
- [20] E. Lopez-Neri, E. Lopez-Mellado, and A. Ramirez-Treviño, "Hierarchical Modeling of Urban Traffic systems for Multi-Agent based simulation," in *International conference on Modeling, simulation and control 2008*, San Francisco, California, 2008a.
- [21] B. Dubiel and O. Tsimhoni, "Integrating agent based modeling into discrete event simulation," in *Proceedings of the 2005 Winter Simulation Conference*, 2005, pp. 1029-1037.

VIII. AUTORES

M. en C. Emmanuel López Neri, obtuvo su título de Maestría en Ciencias con la especialidad en Ingeniería Eléctrica. Actualmente es candidato a Doctor en Ciencias por el CINVESTAV Unidad Guadalajara.

Dr. Ernesto López Mellado, obtuvo su título de Doctor Ingeniero con Especialidad en Robótica por la Universidad Paul Sabatier, Toulouse, Francia. Actualmente es investigador en el CINVESTAV Unidad Guadalajara.

Dr. Antonio Ramírez Treviño, obtuvo su Doctorado en Ingeniería de Sistemas e Informática, por la Universidad de Zaragoza, España. Actualmente es investigador en el CINVESTAV Unidad Guadalajara.

TECNOLOGÍAS DE BAJO COSTO PARA MEDIR Y MEJORAR LA MOTRICIDAD EN LOS NIÑOS “ADVENTURES IN THE JUNGLE”

Quezada Patiño Oscar David, Ruvalcaba Manzano Ricardo, Gómez Pimentel Daniel y
Delgado Mata Carlos

Escuela de Ingeniería
Universidad Panamericana
Aguascalientes, Mexico, C.P. 20290
Tel. 4499106216, correo electrónico: cdelgado@up.edu.mx

RESUMEN

En este artículo se describe un video juego colaborativo para medir y mejorar la motricidad en los niños. El video juego utiliza tecnologías novedosas de bajo costo. La primer tecnología es el desarrollo de una pantalla multi-táctil, por medio de una camara Sony Eye para PS3 y un proyector, para el desarrollo y medición de habilidades motrices finas. La segunda tecnología es un tapete del juego Dance Dance Revolution para el desarrollo de habilidades motrices gruesas. También se describe el “framework” desarrollado para la comunicación de ambas tecnologías con el videojuego. El proyecto está desarrollado en niveles donde se enfoca a la medición y desarrollo de habilidades motrices fines y gruesas en los primeros niveles y el impulso al trabajo colaborativo en los últimos.

I. INTRODUCCIÓN

Actualmente existen diferentes formas de observar, medir e intervenir las habilidades que un niño desarrolla durante su niñez. Los métodos empleados para detectar dichas habilidades se enfocan en realizar ejercicios y juegos los cuales son supervisados por un experto. La forma tradicional de realizar los ejercicios es por medio de actividades pre-diseñadas las cuales se apoyan en métodos simples y en muchos casos con material didáctico pero sin contar con el apoyo de tecnología o con muy poca intervención de ella. Los métodos empleados actualmente realizan observaciones y mediciones de diversos factores

que determinan la eficacia de los movimientos, así como el intelecto que los niños pudieran aplicar para resolver problemáticas presentadas.

En el trascurso del artículo hablaremos del desarrollo de un entorno tecnológico compuesto por software y hardware, el cual ofrece un apoyo para complementar y simplificar el proceso de obtención de información y automatizar el desarrollo de las pruebas anteriormente mencionadas. Por lo que se buscará contestar las siguientes preguntas:

¿La tecnología puede ayudar a medir habilidades en el desarrollo de un niño?

¿El uso de ésta se puede desarrollar de una forma que sea divertida, didáctica y transparente para el niño?

¿El uso de la tecnología ayuda de una manera precisa y confiable a los métodos empleados para la medición de los movimientos?

Para contestar estas preguntas se trabaja sobre un entorno que se enfoca en la intervención de la psicomotricidad de los niños.

Psicomotricidad: se define como la interacción entre cuerpo, movimiento y mente para el desarrollo y expresión de la persona en la vida diaria. Esta se divide en dos grandes aéreas, fina y gruesa.

Psicomotricidad gruesa: Se refiere a los movimientos de los grandes músculos del cuerpo. Estas habilidades se desarrollan desde el

nacimiento. Comienzan a partir del control de la cabeza y el torso, y que continuar hasta que el bebé comienza a gatear, después estar de pie y caminar, correr, saltar y la gama de actividades que un adulto puede hacer. La Tabla I muestra el desarrollo normal de las habilidades psicomotrices gruesas en los infantes. (1)

Psicomotricidad fina: Es la habilidad desarrollada principalmente con las manos al usar herramientas, al usar los cubiertos o al usar una pluma para escribir. (2)

implementación de actividades físicas en el video juego, ya que por estadística se muestra un incremento del 77 % en obesidad infantil masculina y un 47 % en obesidad infantil femenina de acuerdo a estadísticas de 1999 al 2006 debido a falta de actividades físicas. (5)

Es importante mencionar que juego como definición, también es una herramienta que es usada con fines de transmitir valor y contenido educativo; siendo así, los videojuegos, un utensilio aprovechable a la educación. Dentro de las funciones educativas de los videojuegos encontramos:

Edad	Caminando y corriendo	Saltando	Brincando	Lanzando y atrapando	Pedalear en una dirección
2-3 Años	Caminar más rítmicamente; contrario al movimiento brazo-pierna. Caminar a prisa, incrementando hasta correr.	Saltar varios centímetros alrededor del piso con ambos pies, sin movimiento en los brazos.	Brincar 1 a 3 veces en un mismo pie manteniendo la parte superior del cuerpo en forma rígida.	Lanzar pelotas con sólo la extensión del antebrazo; los pies permanecen estacionarios.	Empuja un caballo de juguete con los pies; con poca dirección.
3-4 Años	Subir las escaleras alternando los pies. Caminar bajando las escaleras guiado con un pie. Camina en línea recta.	Saltar con un movimiento de brazos de forma coordinada. Grandes saltos con un solo pie.	Brincar de 4 a 6 veces en el mismo pie, flexionando la parte superior del cuerpo y con movimientos balanceados.	Lanzar la pelota con ligeras rotaciones del cuerpo, pero sin mover el peso hacia adelante. Flexiona los codos al momento de lanzar; Atrapar la pelota contra el pecho.	Pedalear y direccionar un triciclo.
4-5 Años	Caminar bajando las escaleras alternando los pies. Se corre con mayor fluidez. Galope y se salta con un pie.	Saltar hacia arriba y mejorando los saltos hacia adelante. Incrementar la distancia de los saltos.	Brincar 7 a 9 veces en el mismo pie. Mejorar la velocidad de salto.	Lanza la pelota con aumento de la rotación de cuerpo y algunos movimientos del cuerpo hacia adelante. Capturar la pelota con ambas manos, si fracasa, aún pueden atrapar la pelota contra el pecho.	Pasear en triciclo rápidamente, sin problemas de dirección.
5-6 Años	Aumento de la velocidad de la carrera. Galope más fluido. Cierzo parecido a saltar.	Salto con un solo pie. Grandes saltos de aprox. 1 metro de distancia.	Brincar 14 metros con el mismo pie en 10 segundos. Brincar con rítmica alternando 2 brinco con un pie y 2 con el otro.	Lanzar la pelota con todo el movimiento del brazo e incrementar el paso hacia adelante al momento de lanzar. El cuerpo se adapta al trayecto de la pelota y a su tamaño.	Paseos en bicicleta con ruedas de entrenamiento.

Tabla I – Habilidades motrices en niños, de (3)

Desde un principio el proyecto fue desarrollado en un ambiente que sea agradable para el niño y se sienta motivado. *“En el caso de los niños, un buen ambiente físico motiva su aprendizaje y el juego, influye en el concepto que el párvulo se forma de la gente a su alrededor. Fomenta los sentimientos positivos del pequeño sobre sí mismo y lo motiva a ser más activo en su proceso de aprendizaje, así como a interactuar dentro de este espacio libre.”* (4)

Un punto importante a considerar, es la

- Transmitir contenidos y valores.
- Adquirir habilidades relacionadas con el mundo digital: más destreza y seguridad.
- Desarrollar actividades como la superación y la autoestima.
- Aprender a compartir con los compañeros.
- Implicarse, tomar decisiones y ejecutar acciones.
- Potenciar habilidades psicomotrices.
- Adquirir nuevos conocimientos.
- Aprender a dominarse y a auto controlarse.
- Trabajar la fantasía.
- Favorecer la coordinación entre las manos y la

vista. (6)

II. MÉTODO

El entorno es un escenario que te envuelve en un ambiente de jungla divertida que nos ayuda a que los niños no descubran el objetivo del proyecto.

Por medio del mismo se miden y desarrollan habilidades motrices finas y gruesas con la ayuda de una pantalla “Multi- touch” y un dispositivo para video juegos llamado “Dance Dance Revolution Mat” (DDR Mat), además de transmitir contenido educativo arrojándonos los resultados pertinentes.

Componentes del entorno

El entorno con el que se desarrollarán las habilidades de los infantes está compuesto en tres grandes módulos; Juego, Pantalla Multi-Táctil y DDR Mat de Xbox 360, éstos dos últimos están encargados de mandar la información para ser procesada por el video juego como se muestra en la figura 1.

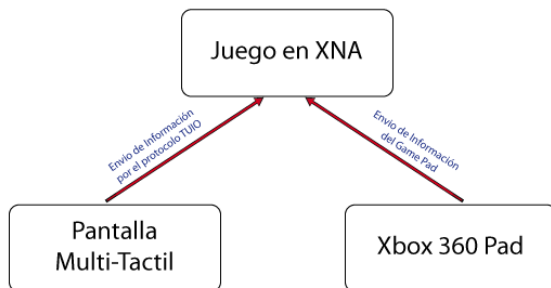


Fig. 1 Esquema del Entorno

En los siguientes puntos se describen detalladamente en que consiste cada uno de los componentes del entorno. Por restricciones de espacio no se detalla la configuración.

Como funciona el Entorno

A) Juego

El juego ha sido desarrollado sobre la

interface Microsoft® XNA™ (7) desarrollada pformas Xbox 360, Zune y Windows(8). Como anteriormente se mencionó, el juego está ambientado en la jungla con personajes y escenarios de acuerdo el tema, tal como se muestra en la figura 2.

Al comenzar con el juego se obtienen los datos de los niños a participar como se ve en la figura 3, para poder interpretar los resultados y estadísticas finales, apoyado de datos como edad y grado escolar para una mejor evaluación.



Fig. 2 Menú principal del juego.

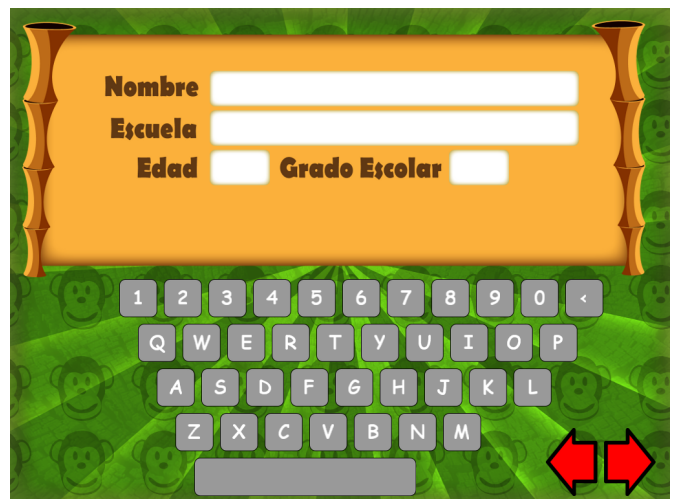


Fig. 3 Pantalla de captura de información general.

Nivel del juego con Motricidad Gruesa:

Nivel 1: Se basa en una persecución en la

cual uno de los personajes del juego mostrado en la figura 4 debe librarse de ser atrapado por un depredador. Los niños por medio de relevos deben simular estar corriendo en el DDR Mat para darle movimiento al personaje estando así motivados por apartarse de su depredador, apoyados de la colaboración de su compañero.

Este nivel nos permite desarrollar y medir la habilidad psicomotriz de carrera mostrándonos la rapidez con que mueven sus piernas. (9)

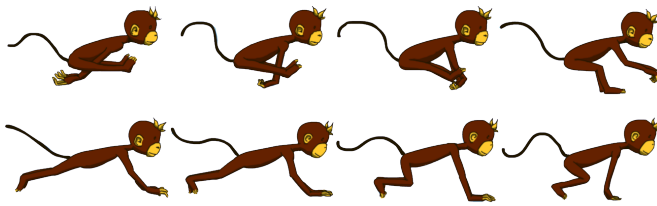


Fig. 4 Personaje del nivel 1.

Niveles Motricidad Fina:

Nivel 2: En pantalla caen cocos de manera continua con un número, el cual debe ser atrapado por medio de una cuerda que se forma al seleccionar con ambas manos los números situados a un costado de la pantalla, los cuales deben ser dos números que multiplicados, sumados o restados (dependiendo la dificultad por edad), den como resultado el número contenido dentro del coco, realizando dicha acción antes de que el coco llegue al piso. Entre mas rápido sea atrapado mayor es el puntaje. También se investiga la factibilidad de que la dificultad se auto modifique para evitar la frustración del niño y que éste continúe jugando. Aquí se mide y desarrolla la precisión y la capacidad de respuesta del niño.

Nivel 3: Este nivel es una adaptación del conocido juego “kullors” de Jon Bartram mostrado en la figura 5, el cual el objetivo es simple: Combinar dos personajes entre sí del mismo color para eliminarlos o juntar dos distintos colores para conseguir el color deseado apoyado de una barra de colores que nos indica el resultado.

Este nivel nos ayuda a desarrollar precisión al tener que seleccionar los personajes para combinarlos,

además desarrollar sus habilidades en el conocimiento de los colores y retención de memoria.

Una vez concluidos los niveles del video juego, el sistema reúne la información obtenida en cada nivel, es almacenada en una base de datos local. Para dicho propósito se cuenta con una base de datos desarrollada con SQLite el cual nos permite manejar la información de manera clara y precisa sin apenas consumir recursos. Por otro lado es necesario agregar una DLL para el manejo de la base de datos.



Fig. 5 Juego Kulors.

Ésta base de datos es alimentada de forma automática sin que el niño o el espectador se percate de ello, ya que se conocen los resultados obtenidos en cada uno de los indicadores analizados. Los datos una vez almacenados, son procesados y convertidos a información útil para el tutor o analista, obteniendo graficas de rendimiento así como graficas de comparación, dando como resultado una herramienta más para la medición del avance motriz e intelectual del niño.

Como un agregado más a dicha herramienta se puede considerar la posibilidad de agregar un modulo extra, el cual se encargue de publicar la información en una página web, la cual podrá ser consultada por los padres, analistas o el mismo niño con el fin de aumentar su competitividad.

B) Pantalla Multi-Táctil

La pantalla multi-táctil está compuesta por un vidrio opaco que se encuentra sostenido por un marco de madera en el cual con un video proyector se muestra el video juego. Durante el juego los niños colocarán sus dedos en la pantalla de vidrio como se muestra en la figura 6 para que posteriormente sean reconocidos como apuntadores. Para que esto se lleve a cabo es necesario colocar una cámara web debajo del vidrio (de ahora en adelante Display) ya que la cámara manda las imágenes capturadas hacia una librería (TUIO Tracker), (10). Para dicho propósito se utiliza una cámara QuickCam® Pro for Notebooks, ya que ésta cámara cumple con las especificaciones necesarias.

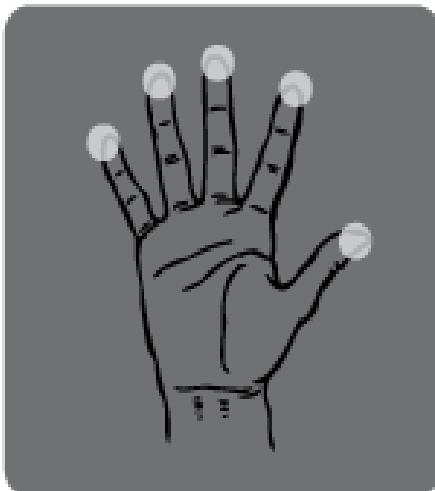


Fig. 6 Esquema de Pantalla Multi-Táctil

Como se mencionó en el párrafo anterior para realizar el proyecto es necesario contar con un "TUIO tracker" y se ha decidido utilizar la librería Touchlib desarrollada por NUI Group. Touchlib (11) está encargada de procesar las imágenes capturadas por la cámara web con las cuales por medio de algoritmos detecta patrones y son convertidos en apuntadores que son enviados al juego para su posterior procesamiento. Touchlib funciona como un servidor que envía paquetes de información al video juego. Los paquetes enviados

están estructurados de tal forma que cumplen con el protocolo TUIO (12), el cual es usado como estándar para la comunicación de pantallas interactivas multi-táctil. En la siguiente ilustración (Figura 7) se muestra un esquema de como está conformada la pantalla multi-táctil.

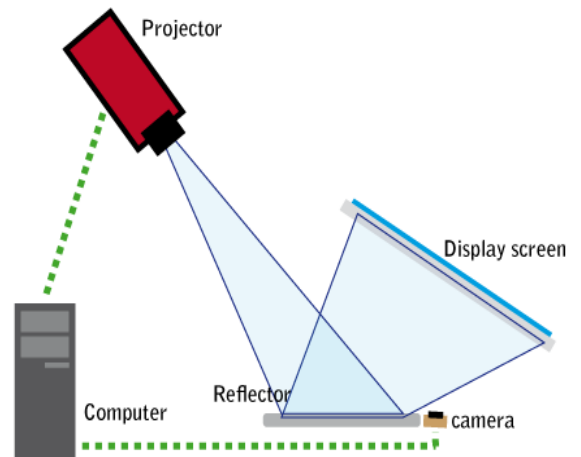


Fig. 7 Esquema de Pantalla Multi-Táctil

C) DDR Mat

El tercer componente es el tapete Dance Dance Revolution (DDR Mat) el cual está enfocado a promover el desarrollo motriz de los niños mediante actividades físicas que son empleadas durante el juego.

El game pad utilizado es el mismo que el empleado en Dance Dance Revolution, ya que el dispositivo está desarrollado para funcionar con el Xbox 360, por consiguiente la facilidad de poder adaptar el game pad a la API utilizada (XNA). Éste componente está formado por el DDR Mat y una funda, la cual recubre el dispositivo (como se muestra en la figura 8) dándole un aspecto diferente ya que el diseño de la funda fue realizado de acuerdo al esquema manejado, proporcionando una integración con el video juego.

Para que el componente pueda funcionar sólo es necesario conectarlo por medio del puerto USB y no es necesario instalar drivers ya que el XNA es el encargado de manejar el dispositivo.

Todos los eventos generados cuando se presione un objeto del DDR Mat son enviados por éste mismo, por lo que fungirá como servidor de eventos los cuales son capturados para su posterior procesamiento en el video juego.



Fig. 8 Dance Dance Revolution Mat y Funda

Requerimientos de software para video juego

El sistema tiene los siguientes requerimientos mínimos para poder ser ejecutado:

- Windows XP/Vista
- Runtime XNA 3,0
- FrameWork 3,5
- Driver para Cámara QuickCam® Pro for Notebooks
- Tener instalado “Adventures in the Jungle”

III. CONCLUSIONES

Con el proyecto queda demostrado que la tecnología de los video juegos ayuda a desarrollar las habilidades motrices e intelectuales de los infantes. Todo esto lográndose con actividades divertidas, ya que desde el inicio del juego los niños mostraron interés y entusiasmo por el hecho de encontrarse frente a un entorno interactivo que los hacia estar atentos en todo momento. Cabe

mencionar que dichos resultados se obtuvieron sin conocimiento del niño ya que para ellos es solo un juego. Los resultados fueron posibles gracias a la utilización de componentes de bajo costo.

IV. REFERENCIAS

- [1] Delgado Mata, Carlos et al., “*Low cost video game technology to measure and improve motor skills in children*”. AFRICON 2009, Nairobi Kenia, IEEE, 23-25 de Septiembre 2009, por publicarse.
- [2] Mesonero Valhondo, Antonio. “*Psicología de la Educación Psicomotriz*”. s.l.: Universidad de Oviedo. pág. 167. ISBN: 8474688051. ISBN-13: 9788474688054.
- [3] Berk, L. E. “*Infants, Children, and Adolescents.*” 5ta. 2004.
- [4] Gálvez Sosa, María Isabel y Rodríguez Ledezma, Nadia Carolina, “*La importancia del juego*”, Universidad de las Américas Puebla. [En línea] 14 de Mayo de 2005. http://catarina.udlap.mx/u_dl_a/tales/documentos/ldf/galvez_s_mi/capitulo3.pdf.
- [5] Islas Ortega, Laura y Peguero García, Mónica. Instituto Nacional de Salud Publica. [En línea] 2006. [Citado el: 10 de Junio de 2009.] http://www.insp.mx/Portal/Centros/ciss/nls/boletines/PME_14.pdf.
- [6] Bartrina, Elisabet. 3er Sector. [En línea] 08 de ENERO de 2008. [Citado el: 01 de JUNIO de 2009.] http://www.3ersector.org.mx/index2.php?option=com_content&do_pdf=1&id=4421.

- [7] Wikipedia. [En línea] [Citado el: 12 de Junio de 2009.] http://en.wikipedia.org/wiki/Microsoft_XNA.
- [8] XNA. [En línea] [Citado el: 12 de Junio de 2009.] <http://www.xna.com/>.
- [9] Fernández García, Emilia, Gardoqui Torralba, Ma. Luisa y Sánchez Bañuelos, Fernando. “*Evaluacion de las habilidades motrices básicas*”. s.l.: INDE Publicaciones, 2007. pág. 18. ISBN:978-84-9729-087-6.
- [10] Group, Natural User Interface. Natural User Interface. [En línea] 2006. [Citado el: 8 de Junio de 2009.] Natural User Interface Group.
- [11] Teiche, Alex, y otros. “*Multi-Touch Technologies*”. 1st edition. 2009.
- [12] Kaltenbrunner, Martin. “*TUIO: A Protocol for Table-Top Tangible User Interfaces*”. [En línea] <http://mtg.upf.edu/files/publications/07a830-GW2005-KaltenBoverBencinaConstanza.pdf>.

USO DE REALIDAD AUMENTADA PARA APOYAR EL APRENDIZAJE EN NIÑOS DE 2 A 5 AÑOS (JUEGO MEMORIA)

Luis Gustavo Quezada Patiño, Carlos Delgado Mata

Universidad Panamericana Campus Bonaterra
Av. Josemaría Escrivá de Balaguer, Aguascalientes No, 101 C.P. 20290, México
Tel. (449) 910 62 16, correo electrónico: cdelgado@ags.up.mx

RESÚMEN.

En este artículo se propone el uso de una herramienta para apoyar el aprendizaje en niños entre 2 y 5 años. El medio utilizado es el conocido juego de memoria el cual es llevado al siglo XXI con el uso de librerías Open Source y de realidad aumentada.

En el presente artículo se propone el uso de realidad aumentada o realidad mixta para apoyar el aprendizaje y con los resultados obtenidos proponer una estrategia que ayude a los niños a aprender o incluso agilizar su mente.

Se utilizan herramientas como Visual Studio 2008 y Microsoft Expression Blend 2, además de la herramienta de Diseño Blender para el Diseño de imágenes u objetos en 2D y 3D.

Los resultados de las pruebas con los niños y en base a teorías pedagógicas se analizarán y se propondrá un trabajo futuro que permita optimizar el aprendizaje e inteligencia de los niños

Palabras clave: Realidad Aumentada, tecnología, aprendizaje, memoria, Open Source, caricaturas, videojuegos, estimulación temprana, aprendizaje significativo.

I. INTRODUCCIÓN

La estimulación en niños de corta edad, en este caso, entre 2 y 5 años, es un punto muy importante

a tomar en cuenta. Está comprobado que durante esta etapa, incluso antes, la plasticidad del cerebro es mayor; lo cual quiere decir que es mucho más fácil enseñar y moldear las capacidades intelectuales de estos [1].

Una de las teorías en las que nos podemos basar para intentar comprender como funciona nuestro cerebro y la forma en la que aprendemos y relacionamos cosas; es con el estudio de la **Teoría del aprendizaje significativo** [2]. Según Ausubel, existen dos condiciones para que se realice el aprendizaje significativo.

El contenido del aprendizaje es significativo; cuando es posible relacionar los conocimientos previos con los nuevos. Esto significa que es necesario utilizar cosas sencillas de uso común, o conocidas para el niño, y así pueda aprovecharlas en su favor al recordar fácilmente los nuevos conocimientos.

Al utilizar esta teoría como base se intenta utilizar dibujos, caricaturas u objetos en 2D y 3D que ayuden a relacionar fácilmente lo que se quiere aprender.

Utilizando el sistema de los videojuegos se intenta lograr que el niño adquiera el gusto por aprender. Al darse esto, el niño podrá relacionar los objetos, e inconscientemente lograr una retención mayor de la información.

Memoria Sensorial	Sistemas de Memoria <i>Memoria Reciente</i>	Memoria Remota
Se refiere al <i>almacenamiento</i> de información sensorial (lo que se ve, oye, huele, gusta o toca).	Se refiere a la <i>retención</i> de información por cerca de 20 seg.	Se refiere al <i>almacenamiento</i> de información por largos períodos de tiempo
	Tiene una capacidad de almacenamiento limitada	Tiene una capacidad de almacenamiento ilimitada
	Requiere de la integridad del <i>hipocampo</i> (estructura cerebral)	Requiere de la integridad de las áreas de asociación de la corteza cerebral
La pérdida de información es rápida a menos que se transfiera a la memoria reciente	La pérdida de información es frecuente y rápida	La pérdida de información es menos probable
La capacidad de retener imágenes visuales se mantiene hasta bien entrada la tercera edad.	Las fallas de memoria reciente son frecuentes en la vejez	La capacidad de recordar hechos pasados permanece hasta etapas avanzadas de deterioro mental, en las que también se afecta
	La capacidad de <i>retención</i> depende del tipo de material a recordar. Así por ejemplo, los adultos mayores conservan la capacidad de retener dígitos hacia adelante, mientras que la retención de dígitos hacia atrás disminuye con el paso de los años	La capacidad de memoria de largo plazo para información <i>recién aprendida</i> desciende de manera significativa con el avance de la edad, mientras conserva la capacidad de evocar el material aprendido en el pasado más distante

Fig. 1 Memorias, de [3]

En la tabla de la figura 1 se muestra una clasificación de la memoria, la cual muestra tres tipos; la sensorial, reciente y remota. Estas describen el ciclo de vida que tiene el almacenamiento de información en nuestro cerebro.

Memoria Sensorial: Es la parte de almacenamiento temporal y captura todo lo que sentimos, este es el principio básico de lo que la herramienta quiere lograr. El punto es poder identificar que es lo que nos atrae y como, nos ayuda a relacionar las cosas.

Memoria Reciente: Es la parte más delicada de la memoria ya que la información es fácilmente desechada y depende mucho del material a recordar, nuevamente se trata de encontrar la mejor forma para recordar fácilmente las cosas.

Memoria Remota: Se refiere a la información que almacenamos y se queda ahí por mucho tiempo, la cual vamos perdiendo con el paso del tiempo. Aunque esto se puede compensar

con el ejercicio mental, los juegos como la memoria y sobre todo, a una corta edad nos ayudan a ejercitar esa parte del cerebro.

Juego de Memoria

El principal punto por el cual se utilizó este tipo de juego como referencia, es porque, desde que se conoce se ha sabido la importancia que tiene al ayudar a la concentración y al mejoramiento de las habilidades mentales, además de ser sencillo y fácilmente comprendido por las personas que lo juegan.

Video Juegos

El juego forma parte primordial en el desarrollo de los niños, ayuda en gran medida y junto a la intervención del ambiente, a forjar su intelecto y aumentar sus conocimientos y habilidades.

Utilizando este hecho y haciendo uso de la tecnología, se han creado video juegos, los cuales pueden ayudar a mejorar las capacidades de los

niños, siempre y cuando sean supervisadas por un adulto.

Algunos de las motivaciones que se persiguen con los videojuegos educativos son los siguientes:

- *Transmitir contenidos y valores.*
- *Adquirir habilidades relacionadas con el mundo digital: más destreza y seguridad.*
- *Desarrollar actividades como la superación y la autoestima.*
- *Aprender a compartir con los compañeros.*
- *Implicarse, tomar decisiones y ejecutar acciones.*
- *Potenciar habilidades psicomotrices.*
- *Adquirir nuevos conocimientos.*
- *Aprender a dominarse y a auto controlarse.*
- *Trabajar la fantasía.*
- *Favorecer la coordinación entre las manos y la vista.* [4]

Objetivos del proyecto.

Se intenta dar a los niños una herramienta que sea divertida, didáctica y transparente, que a su vez les permita desarrollarse con otros, jugar y aprender al mismo tiempo. No solo en forma intelectual sino física y motriz.

El segundo objetivo es visualizar aquello que les atrae más a los niños. El uso de caricaturas o elementos conocidos para ellos, incluso elementos abstractos en 2ª y 3ª dimensión y sobre todo registrar sus avances. Al procesar estos datos se obtendrá información que tratará de responder que tipos de objetos o figuras sirven en mayor medida como incentivo para los niños.

Los objetos a comparar son los siguientes:

- Objetos simples bidimensionales como: formas básicas geométricas (cuadrados, círculos, triángulos, etc.)
- Objetos simples tridimensionales como: cubos, pirámides, etc.
- Figuras complejas tridimensionales.

- Imágenes de objetos comunes.

Los resultados obtenidos se miden y se comparan contra el uso de una memoria tradicional.

El tercer objetivo es utilizar tecnología de código libre u Open Source, para poder hacer de esta herramienta, económica y que pueda ser utilizada por cualquier institución educativa.

Para poder lograr estos objetivos se utilizó tecnología de bajo costo para ayudar a mejorar las habilidades de los niños [5], ésta se describe a continuación.

Realidad Aumentada

La **Realidad Aumentada** (AR) es una nueva rama de interfaces hombre-máquina, donde los elementos reales conviven con los elementos virtuales. Estos sirven para aportar información adicional a los primeros [3]. En la figura 2 se muestra como se pueden combinar elementos del mundo real con el mundo virtual.



Fig. 2 Realidad Aumentada, de [3]

La realidad amentada es una tecnología que cuenta con varios años de desarrollo. Esta tecnología permite crear mundos y objetos virtuales que pueden ser utilizados para hacer más llamativa la experiencia en juegos que utilizan elementos del mundo real. Esta tecnología ha sido utilizada en el desarrollo de videojuegos tales como Eye of Judgment para Playstation 3 (mostrado en la figura 3).

Este video juego combina elementos físicos y virtuales, permitiéndole al jugador interactuar perfectamente entre la realidad y el mundo virtual.



Fig. 3 Eye of Judgment

En la figura 4 se muestran elementos típicos en proyectos de realidad aumentada, como son: Un tablero con patrones que pueden ser identificados mediante el uso de una cámara, para posteriormente ser mostrado por medio de lentes especiales y/o una pantalla.

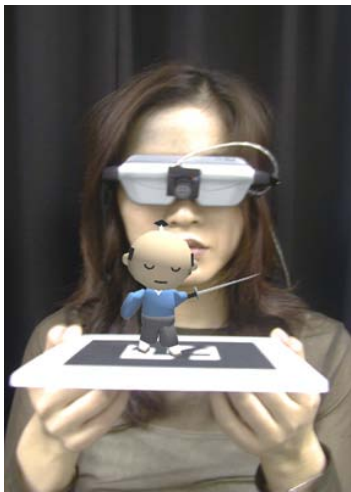


Fig. 4 Ejemplo de realidad aumentada, de [6]

La RA puede ser utilizada en proyectos educativos, de entretenimiento [7, 8] y de simulación en diversos campos como la medicina [9].

También se intenta aprovechar el hecho de que en la actualidad los niños conviven fácilmente con la tecnología, ya que ahora la disponibilidad de equipo computacional es muy amplia.

Patrones o marcadores

En la figura 5 se muestra un patrón (o marcador), que son los diseños en imágenes únicas, que pueden ser reconocidos por la librería de realidad aumentada.

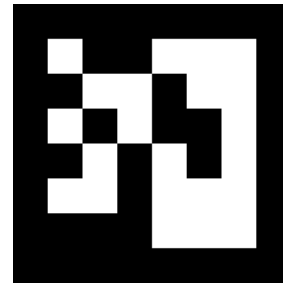


Fig. 5 Imagen o patrón utilizado en la herramienta

Estos patrones son utilizados en la interacción con el medio, utilizándose como tarjetas que serán reconocidas por la herramienta y mostrando su interpretación en la pantalla.

Al utilizar estos patrones se puede simular el juego de la memoria.

XAML

Es un lenguaje declarativo optimizado en forma de XML que define la forma de representar interfaces u objetos gráficos. Estas interfaces son interpretadas por el WPF y pueden ser utilizadas en Visual Studio 2008.

WPF (Windows Presentation Foundation)

Es una interfaz realizada para Windows, influenciada por tecnologías modernas como HTML y Flash. Permite interpretar XAML y realizar interfaces de usuario.

II. MÉTODO.

La estructura del proyecto está dividida en 4 partes:

- Editor de contenido.
- Convertidor de objetos 3D a XAML
- Herramienta de realidad aumentada.
- Interprete de datos.

El **editor de contenido** utilizado, es, el programa Open Source y multiplataforma llamado Blender [10] el cual permite generar imágenes u objetos 3D.

Los objetos que son realizados con esta herramienta, son convertidos y exportados a un formato entendible por la herramienta. Estos objetos son interpretados y se muestra en la pantalla.

El **convertidor Blender a XAML** es un script [11], que es instalado y configurado para funcionar con Blender, este permite exportar el objeto 3D creado a XAML [12]. Una vez que los objetos se encuentran en el formato correcto basta con ser declarados en la herramienta y copiar el archivo al repositorio, para que puedan ser utilizados.

La **herramienta de realidad aumentada** es la parte principal del proyecto. En ésta se ejecuta el programa que permite trabajar con la detección de patrones y así simular un juego de memoria. Este permite seleccionar el número de cartas con las que se va a jugar (la selección debe de ser N x N). Por otro lado, también es posible seleccionar el tipo de juego. La herramienta guarda los resultados en un archivo de estadísticas los cuales son interpretados posteriormente.

El **intérprete de datos** analiza y regresa información ya procesada permitiendo evaluar los resultados. Una vez analizados se llegará a conclusiones que permitirá hacer modificaciones o decidir la mejor forma de aprovechar la herramienta.

En la figura 6 se muestra el diagrama de funcionamiento del proyecto junto con los roles que desempeñan los usuarios.

El **diseñador** crea los objetos e imágenes en el editor de contenido que serán después convertidos a formato XAML, de ahí se toman y son utilizados por la herramienta, el **jugador** interactúa y los datos son tomados por el **interprete** y finalmente se les da un valor, al ser revisados por el **interprete de datos**.

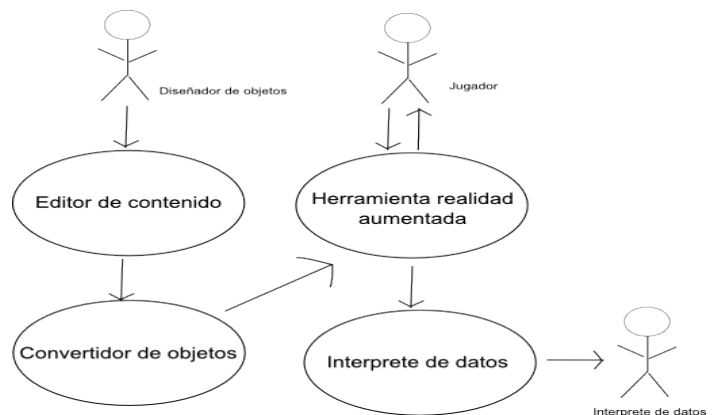


Fig. 6 Diagrama de funcionamiento

Para manipular y realizar esta versión del juego se utilizó la librería **ARToolKit Plus** [13] y el software de Microsoft Windows Presentation Foundation **WPF** [14], junto a Visual Studio 2008 con el Framework 3.0.

La dinámica del juego es sencilla, al iniciar se piden datos del niño como su nombre, edad, estatura, peso y escolaridad, y almacenados en un archivo XML [15]. También se lleva un registro de sus tiempos y número de aciertos.

Después de registrar la información se muestran las cartas para fijar la dimensión del juego N X N, y posteriormente comenzar a jugar.

El juego termina cuando todos los pares de objetos son identificados.

Los resultados son leídos e interpretados por el programa analizador, y muestra los resultados en forma de estadísticas. Para mostrar los resultados y darle significado se utiliza; análisis de datos estadísticos, sin llegar a ser tan complejos, puede arrojar resultados interesantes y así calcular el desempeño de los niños.

Algunos de los resultados arrojados son:

- Tiempos promedios según los tipos de objetos.
- Tiempos agrupados por edades.
- Tiempos agrupados por escolaridad.
- Tiempos agrupados por género.

Interfaz

La interfaz es muy sencilla y fue desarrollada con Visual Studio 2008. Funciona con una webcam que pueda procesar un mínimo de 28 FPS, en este caso se utilizo una **Logitech® QuickCam® Pro for Notebooks**, pero es recomendable utilizar cámaras con mayor resolución para mejorar la experiencia.

En la figura 7 se muestra la pantalla inicial en este momento no hay interacción, y no se muestran los objetos.



Fig. 7 Imagen sin interacción.

Una vez inicializada la captura se puede observar como las figuras que son idénticas cambian de color demostrando que son una pareja, tal y como se muestra en la figura 8.

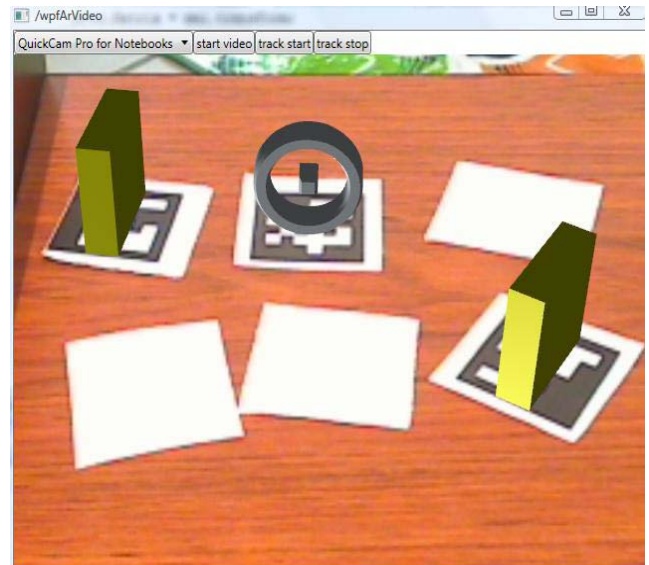


Fig. 8 Interacción

Los objetos idénticos interactúan y cambian de color, dando como resultado una retroalimentación positiva, que indica que la combinación es correcta.

Al terminar se felicita al jugador y se le invita a volver a jugar. En este momento la información recabada ya fue almacenada para su uso futuro.

Requerimientos

Los requerimientos del sistema son muy sencillos.

- Windows XP o superior, funciona con Windows Vista.
- Framework 3.0 o superior.
- Procesador Pentium 4 o superior.
- Instalar Memoria Aumentada.

III. CONCLUSIONES

En el presente artículo se ha presentado una herramienta que combina las bondades de un juego sencillo como lo es el Juego de Memoria y la tecnología de la realidad aumentada. Demostrando nuevamente que el uso de los videojuegos aplicados a una causa, ayuda al crecimiento, y motivación de los niños.

El desarrollo de este proyecto no requirió de una gran inversión para poder realizar una herramienta útil en el apoyo del crecimiento y mejoras de habilidades físicas e intelectuales de los niños.

Se explorará en un proyecto posterior el uso de realidad aumentada para la creación de una memoria combinada, como la mostrada en la figura 9.



Fig. 9 Ejemplo de memoria combinada, de [16]

En la figura anterior se muestra como se puede utilizar el principio del juego de memoria y ser utilizado para enseñar algo a los niños. Un ejemplo es relacionar números con los objetos que sean identificados adecuados para las condiciones del niño y por consiguiente utilizarlos para hacer operaciones matemáticas o que puedan relacionar y aprender mejor.

IV. REFERENCIAS

- [1] Arango de Narváez, María Teresa, Infante de Ospina, Eloísa. *Estimulación temprana*.
- [2] Ausubel, Paul David. *Teoría del aprendizaje significativo*. Nueva York.
- [3] Universidad de los Andes. “Memorias” <http://biblioteca.uandes.cl/documentos/BZegers3.jpg>
- [4] Bartrina, Elisabeth. “Los videojuegos también son educativos”. [En línea] 08 de ENERO de 2008. [Citado el: 01 de JUNIO de 2009.] http://www.3ersector.org.mx/index2.php?option=com_content&do_pdf=1&id=4421
- [5] Delgado Mata, Carlos et al., “*Low cost video game technology to measure and improve motor skills in children*”. AFRICON 2009, Nairobi Kenia, IEEE, 23-25 de Septiembre 2009, por publicarse.
- [6] Patrick Sinclair, “*ARToolkit*”, Universidad de Washington. <http://metade.org/research/phd/artoolkit/>
- [7] Fischer, J., Bartz, D., and Straßer, W. 2004. “*Occlusion handling for medical augmented reality using a volumetric phantom model*”. En Proceedings of the ACM Symposium on Virtual Reality Software and Technology (Hong Kong, November 10 - 12, 2004). VRST '04. ACM, New York, NY, 174-177
- [8] Cheok, A., Yang, X., Ying, Z., Billinghurst, M., and Kato, H. 2002. “*Touch-Space: Mixed Reality Game Space Based*” en Ubiquitous, Tangible, and Social Computing. Personal Ubiquitous Comput. 6, 5-6 (Jan. 2002), 430-442
- [9] Oda, O., Lister, L. J., White, S., and Feiner, S. 2007. “*Developing an augmented reality racing game*”. En Proceedings of the 2nd international Conference on intelligent Technologies For interactive Entertainment (Cancun, México, January 08 - 10, 2008), 1-8.
- [10] Roosendaal, Tom and Selleri, Stefano, “*The Official Blender 2.3 Guide: Free 3D Creation Suite for Modeling, Animation and Rendering*”, No Starch Press.
- [11] “*XAML exporter for Blender*” <http://www.codeplex.com/Wiki/View.aspx?ProjectName=xamlexporter>

- [12] Lori A. MacVittie, “*XAML in a Nutshell*”, O'Reilly, March 2006
- [13] *ARToolKitPlus* and *WPF*
<http://sites.google.com/site/augmentedrealitytestingsite/>
- [14] MacDonald, M, “*Pro WPF in C# 2008: Windows Presentation Foundation with .NET 3.5*”. Apress, 2007, pp. 1–17
- [15] Bray, Tim, Paoli Jean and Sperger-McQuen C.M. “*Extensible Markup Language (XML) 1.0*”, REC-xml-19990210
- [16] Santacruz, Ma. del Rocío, *Material didáctico propuesto para elaborarlo y trabajarlo con alumnos de problemas de aprendizaje*, Universidad Autónoma de Tlaxcala, Documento Técnico.

METODOLOGÍA PARA EL DISEÑO DE SISTEMAS EMBEBIDOS BASADA EN MODELOS DE REDES DE PETRI

M. Alcaraz-Mejía, R. Campos-Rodríguez, L. García-Contreras, J. Urquieta-Jiménez, M. Aguirre-Villa

Departamento de Ciencias Tecnológicas,
Centro Universitario de la Ciénega, Universidad de Guadalajara
Av. Universidad 1115, Col. Lindavista, Ocotlán, Jalisco, 47820, México
Tel. +52(392)9259400, ext. 8369, correo electrónico: {mildreth.alcaraz, raul.campos}@cuci.udg.mx

RESUMEN

En este trabajo se presenta una plataforma embebida para el control de sistemas de eventos discretos. El sistema real de propósito específico se modela como una Red de Petri con señales de entrada y salida. Esto permite adaptar un esquema de control que puede ejercer acción sobre el sistema mediante comandos de entrada y retroalimentar el estado mediante señales de salida. Se proporciona una metodología para traducir el modelo en Red de Petri a una implementación de firmware embebido.

Palabras clave: Sistema Embebido, Control, Sistemas de Eventos Discretos, Redes de Petri.

I. INTRODUCCIÓN

Los sistemas de eventos discretos se encuentran prácticamente en todos los aspectos de la vida cotidiana, entre los que se consideran los sistemas embebidos. Estos sistemas poseen varias características a las que se debe prestar especial atención. Por ejemplo, un sistema embebido se programa comúnmente una sola vez, es decir, no permite realizarle actualizaciones después que éste se encuentra trabajando. En los raros casos en los que esto se permite, resulta muy complicado y costoso. Otra característica crítica de los sistemas embebidos es que normalmente se utilizan para aplicaciones donde la confiabilidad es un factor primordial [1], [2].

Debido a las características que presenta un sistema embebido, las técnicas de programación de software convencionales no son aptas para su

utilización en el desarrollo de los mismos, ya que los sistemas embebidos requieren de pruebas y análisis más exhaustivos. Por ello, uno de los principales retos a los que se enfrentan los ingenieros que diseñan sistemas embebidos es que no existen metodologías formales relacionadas con los aspectos del diseño y la implementación de hardware, o la verificación del software necesario para realizar una adecuada implementación de estos sistemas [2].

A pesar de su importancia, la literatura relacionada con metodologías formales para el diseño de sistemas embebidos es relativamente nueva. En [3], el autor adopta un enfoque práctico sobre la implementación de sistemas embebidos basados en el lenguaje de programación C y C++. Ofrece una amplia gama de consejos y trucos prácticos basados en la experiencia del autor. En [4], los autores hacen una revisión de las metodologías formales para el diseño, la implementación y validación de sistemas embebidos. Sin embargo no ofrecen aspectos técnicos ni prácticos de implementación, ni técnicas para la traducción de dichas metodologías a código funcional. En [5], los autores se enfocan a la programación de sistemas embebidos de manera práctica. Abordan varios aspectos del diseño de sistemas embebidos, como la selección del procesador, tamaño de buses y periféricos de comunicación. Consideran el diseño del firmware basado en diagramas de flujo y pseudocódigo. En [12], los autores presentan un formalismo basado en Redes de Petri y Lógicas Temporales, para la verificación de sistemas embebidos. Sin embargo, la metodología para



Figura 1. Sistema Generalizado

pasar del modelo a la implementación no es directa. En [13], el autor se enfoca al diseño de controladores en hardware y software desde un punto de vista eminentemente práctico. Incluso considera el uso de un sistema operativo embebido para la realización de controladores complejos y difíciles de implementar en un entorno operativo diferente.

En este trabajo se propone una metodología basada en modelos en Redes de Petri para la implementación de sistemas embebidos. Se abordan aspectos del formalismo de las Redes de Petri y se consideran aspectos técnicos de su traducción a código ejecutable.

El resto de este trabajo se organiza de la siguiente manera. En la Sección II. se presentan los aspectos generales de los sistemas embebidos. En la Sección III. se introducen las nociones básicas de las Redes de Petri. En la Sección IV. se presenta la metodología de programación propuesta en este trabajo y se muestra un ejemplo de aplicación. Finalmente, en la Sección V. se ofrecen las conclusiones de este trabajo.

II. SISTEMAS EMBEBIDOS

Un Sistema Embebido es un sistema de propósito específico, que realiza sólo una o pocas funciones, generalmente corre en tiempo real, y se encuentra incrustado como parte de otro componente o sistema (no stand-alone) [1].

Los Sistemas Embebidos utilizan la potencia de un microcontrolador, como un MCU (microcontroller unit) o un DSC (digital signal controller) en su

implementación particular. Estos microcontroladores combinan una unidad de procesamiento (como el CPU de la PC) con circuitos adicionales llamados periféricos, más otros circuitos en el mismo chip para hacer un pequeño módulo de control que requiere solo unos pocos componentes externos para formar un sistema completo.

Debido a su propósito específico, estos sistemas utilizan pocas capacidades de hardware, lo que permite reducir su tamaño considerablemente y, por lo tanto, su costo y consumo de energía. Sin embargo, esta reducción de hardware exige a su vez consideraciones extras en el diseño del código de control para garantizar un alto rendimiento [2]. Este código, se conoce como firmware debido a que se almacena en un chip tipo ROM (Read Only Memory) o en una memoria Flash.

Como se mencionó al inicio de esta sección, otra característica importante de estos sistemas es que presentan restricciones de tiempo real e interactúan en ambientes críticos, por lo que se debe aumentar al máximo su confiabilidad [2].

Un sistema embebido puede tener alguno(s) de los siguientes periféricos: un pequeño teclado, una pequeña pantalla, leds, botones, displays de un dígito o pocos caracteres, sensores de temperatura, de luz, de movimiento, de humo, entre otros [5].

Los sistemas embebidos son utilizados en diversas áreas, como en electrónica, equipo médico, sistemas de transporte, telecomunicaciones, electrodomésticos, seguridad, automatización, entre otras. La idea general de un sistema se presenta en la Figura 1. Esta consiste de una caja negra con entradas y salidas, las cuales pueden ser analógicas o digitales. En el caso general de un sistema embebido, como se muestra en la Figura 2, la caja principal cuenta con varios componentes de hardware, como el procesador, la memoria, e interfaces de comunicación [3].

III. TÉCNICA DE MODELADO

Las Redes de Petri constituyen una herramienta ideal para el diseño y análisis de sistemas de eventos discretos, como en el caso de los sistemas embebidos. Este modelo permite aplicar análisis formal para la verificación de propiedades que aseguran el correcto diseño del sistema, tanto de hardware como de software, antes de pasar a la etapa de implementación. Al aplicar este tipo de análisis se ahorra mucho tiempo de depuración del código, y por consecuencia, reduce el costo final del sistema.

Para obtener el modelo matemático que represente al sistema, en este trabajo se utilizan las Redes de Petri Interpretadas (RPI) para la representación de sistemas con entradas y salidas [6], [7]. Las RPI son una extensión a las Redes de Petri (RP) [8], cuyo formalismo se presenta a continuación.

Definición 1. Una Estructura de Red de Petri (ERP) es un grafo bipartito $G = (P, T, I, O)$ con dos clases de nodos, los lugares $P = \{p_1, \dots, p_m\}$ y las transiciones $T = \{t_1, \dots, t_n\}$, y dos funciones de relación de flujo, la función de entrada $I: P \times T \rightarrow Z^+$ y la función de salida $O: P \times T \rightarrow Z^+$. La ERP se puede representar convenientemente como una matriz, lo cual simplifica la representación matemática, así como las demostraciones formales.

Definición 2. Las matrices de incidencia, pre $C_{m \times n}^-$ y post $C_{m \times n}^+$, de una ERP (P, T, I, O) , se definen como $C_{m \times n}^- = c_{ij}$, donde $c_{ij} = I(p_i, t_j)$, y

$$C_{m \times n}^+ = c_{ij} \quad \text{donde}$$

$$c_{ij} = O(p_i, t_j) \text{ para } 1 \leq i \leq m, 1 \leq j \leq n,$$

respectivamente. La matriz de incidencia $C_{m \times n}$ se define como $C_{m \times n} = C_{m \times n}^+ - C_{m \times n}^-$. Los índices m y n representan los números de lugares y transiciones en la ERP, respectivamente.

El comportamiento dinámico de una RP se realiza a través de un juego de marcas. Una función de marcado $M: P \rightarrow Z^+$ representa el número de marcas en cada lugar.

Definición 3. Un Sistema de Red de Petri o simplemente una Red de Petri (PN) es un par (G, M_0) donde G es una ERP (P, T, I, O) y M_0 es una distribución inicial de marcas, conocida como Marcado Inicial, la cual representa las condiciones iniciales de la RP.

Las marcas en los lugares de la RP, permiten definir el comportamiento dinámico a través del disparo de las transiciones habilitadas. Informalmente, una transición puede ser disparada si existe un número suficiente de marcas en cada lugar de entrada en un momento dado.

Definición 4. Dada una transición $t_j \in T$ de una RP (Q, M_0) , se dice que t_j está habilitada en el marcado M_k , denotado como $[M_k]t_j$, si se cumple que $\forall p_i \in P, M_k(p_i) \geq I(p_i, t_j)$. El conjunto de todas las transiciones habilitadas en el marcado M_k se denota como $[M_k]$.

El disparo de una transición habilitada en una RP permite alcanzar un nuevo marcado, el cuál puede ser calculado utilizando la ecuación de estados siguiente.

Definición 5. Una transición $t_j \in T$ habilitada en un marcado M_k , i.e., $[M_k]t_j$, puede ser disparada alcanzando un nuevo marcado M_{k+1} , denotado como $M_k \xrightarrow{t_j} M_{k+1}$. El nuevo marcado M_{k+1} se calcula como:

$$M_{k+1} = M_k + C \cdot \vec{t}_j \quad (1)$$

donde \vec{t}_j es un vector de n -entradas, conocido como el vector de Parikh de la transición t_j , definido como $\vec{t}_j(i) = 0, i \neq j, \vec{t}_j(j) = 1$.

Definición 6. Una Red de Petri Interpretada (RPI) es un par (Q, M_0) , donde

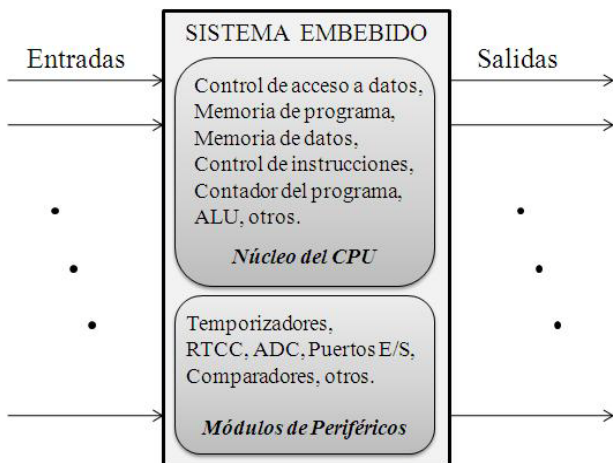


Figura 2. Sistema Embebido

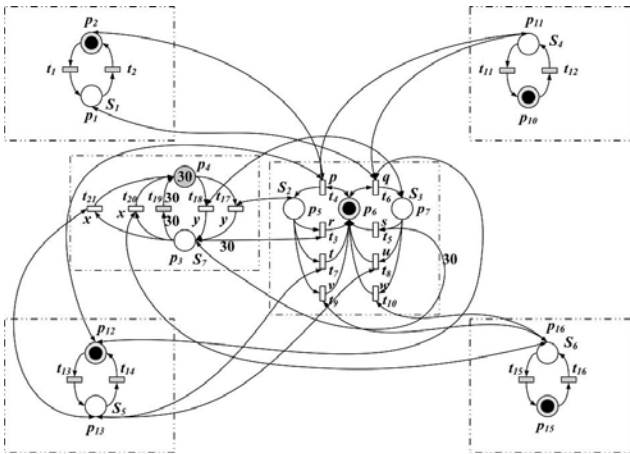


Figura 3. Red de Petri Interpretada

El marcado inicial de la RPI es

$$M_0 = [0\ 1\ 0\ 3\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0]^T$$

La función de salida es:

$$\varphi = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

El conjunto de símbolos de entrada es:

$$\Sigma = \{p, q, r, s, u, x, y\}$$

El etiquetado de las transiciones es el siguiente:

$$\lambda(t_4) = p, \lambda(t_6) = q, \lambda(t_2) = \lambda(t_5) = r,$$

$$\lambda(t_7) = \lambda(t_8) = s, \lambda(t_9) = \lambda(t_{10}) = u,$$

$$\lambda(t_{17}) = \lambda(t_{18}) = y, \lambda(t_{20}) = \lambda(t_{21}) = x,$$

$$\lambda(t_1) = \lambda(t_2) = \lambda(t_{11}) = \lambda(t_{12}) = \lambda(t_{13}) = \lambda(t_{14}) = \lambda(t_{15}) = \lambda(t_{16}) = \lambda(t_{19}) = \lambda(t_{22}) = x$$

IV. METODOLOGÍA DE PROGRAMACIÓN

En este trabajo se propone una metodología para traducir un modelo en RPI a su equivalente en programación embebida. Para ello debemos distinguir claramente las siguientes dos definiciones.

Definición 10. Una *Entrada* es aquella que se encuentra conectada al microcontrolador para que éste lea una señal o dato. Esta señal o dato es a su vez salida de otro sistema.

Definición 11. Una *Salida* es aquella que se encuentra conectada al microcontrolador para que

éste escriba una señal o dato. Esta señal o dato es a su vez entrada de otro sistema.

Los pasos generales para convertir una RPI que representa un sistema de eventos discreto, a un programa de control embebido, son los siguientes:

1. Identificar las Entradas $C_1^e, C_2^e, \dots, C_n^e$.
2. Identificar las Salidas $C_1^s, C_2^s, \dots, C_m^s$.
3. Para cada Entrada $C_j^e, 1 \leq j \leq n$, asignar los bits de los Registros de los Puertos de Entrada $\{Rx1, Rx2, \dots, \dots\}$ a los que está conectado físicamente con el microcontrolador, y en los cuales se leerá la señal de entrada.
4. Para cada Salida $C_j^s, 1 \leq j \leq m$, asignar los bits de los Registros de los Puertos de Salida $\{Rx1, Rx2, \dots, \dots\}$ a los que está conectado físicamente con el microcontrolador, y en los cuales se escribirá la señal de salida.
5. Implementar la RPI con el formato mostrado en la Figura 4.

El procedimiento detallado para la implementación de la RPI se describe a continuación:

- a. Definición de entradas y salidas. En esta parte es necesario definir con la etiqueta `#define` los bits del registro de control que se utiliza para cada entrada $C_j^e, 1 \leq j \leq n$. Opcionalmente, las entradas y salidas se pueden definir como constantes con `#define` para el registro x correspondiente.
- b. Definir el número de lugares y el número de transiciones de la RPI con `#define`, cuyos valores son constantes durante la ejecución del programa.
- c. Declarar las matrices de incidencia Pre y Post. Se recomienda declararlas para ser almacenadas en la memoria del código. Esto permite optimizar el uso de la memoria del programa. Para ello se utiliza el comando `const` previo al tipo de dato definido para las matrices. En este caso se utiliza el `unsigned char`, que permite guardar números en el

```

#include <picxxx.h>

// Definiciones de las entradas y salidas
#define Cel      Rxa
#define Csl      Rxb
.
.
// Definición de constantes de la RPI
#define NL        <Lugares>
#define NT        <Trans>

// Constantes en memoria de código
const unsigned char MatPre[NL][NT] = { { <I> } };
const unsigned char MatPost[NL][NT] = { { <O> } };
.
.
//Global Variable Declarations
unsigned char EstadoActual[NL]={ <Edo Inicial> };
unsigned char EstadoSiguiente[NL];
.
.
//Function Prototypes
void Inicializar();
void LeerEntradas();
unsigned char TransControlablesHabilitadas();
void EscribirSalidas();
void CalculoEstadoSiguiente();
unsigned char VerificacionErrores();
void CalcularEstadoActual();
void SeleccionarTransicion();
.
.
void main (void){
    Inicializar();
    while(1){
        CalcularEstadoActual();
        LeerEntradas();
        VerificacionErrores();
        if (TransControlablesHabilitadas()){
            SeleccionarTransicion();
            EscribirSalidas();
            CalcularEstadoSiguiente();
        } // fin if
    } //fin while
} //fin main

```

Figura 4. Código General de Implementación de una RPI

rango de [0..256]. Debe notarse que las matrices Pre y Post están formadas por números enteros positivos.

- d. Declarar las variables globales. En este caso se sugiere declarar como variables globales el estado actual y el estado siguiente para la RPI, las cuales tienen tamaño NL (Número de Lugares).

- e. Declarar los prototipos de las funciones a utilizar. Para este caso, se sugiere tener al menos las 8 funciones declaradas en el código que se muestra en la Figura 4.
- f. Escribir la función principal (main) para la implementación de la RPI, como se muestra en la Figura 4.

El programa o función principal (main) se describe a continuación:

1. En la función `Inicializar()` se limpian todas las banderas, se limpian todas las entradas y salidas, se definen pines de entrada y salida (en caso de no incluirlas en la cabecera) se configuran temporizadores, entre otras.
2. Enseguida se inicia un ciclo mientras infinito (`while(1)`). Esto es, mientras el microcontrolador esté energizado, se deberá realizar lo que se encuentra dentro del ciclo.
3. La función `CalcularEstadoActual()` permite actualizar el estado actual de la RPI (`EstadoActual`), con el “estado siguiente” calculado previamente en la RPI (`EstadoSiguiente`). Inicialmente, el estado actual es el marcado inicial.
4. La función `LeerEntradas()` realiza una lectura de los bits en los registros definidos como Entradas (`bits = 1`). Esto permite reconstruir el estado actual del Sistema Físico y compararlo contra el estado actual del modelo en RPI. En este caso, cada entrada leída es una salida en la RPI. Si el cambio se dio por una transición no-manipulable (evento externo), entonces se debe emplear algún mecanismo para reconstruir la trayectoria ejecutada por el sistema, en caso de que el esquema de control requiera retroalimentación de trayectorias. En esta etapa del trabajo, no se aborda el tema reconstrucción de estados y trayectorias.
5. La función `VerificacionErrores()` es recomendable en la implementación de una

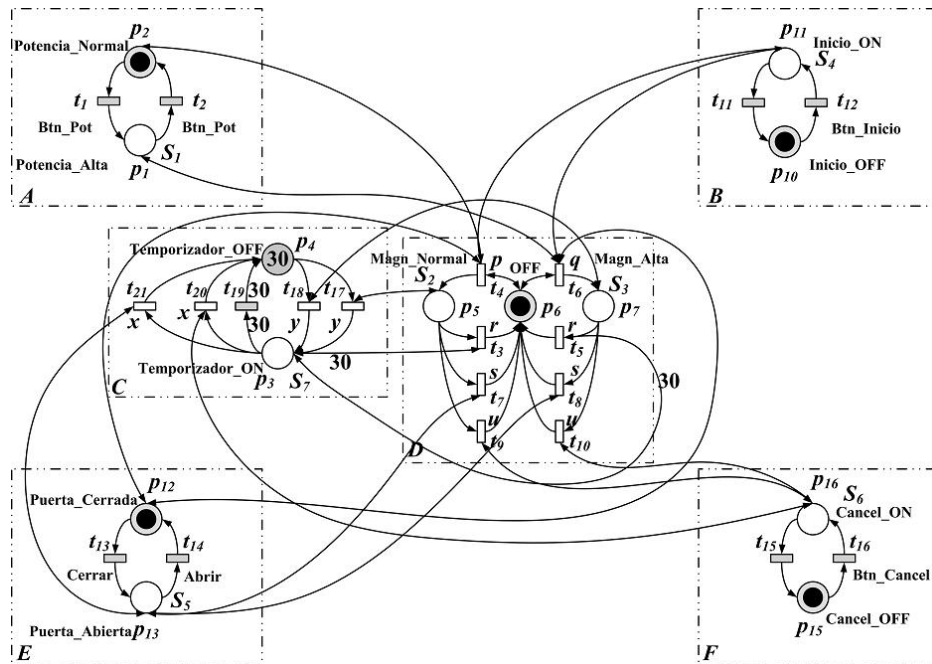


Figura 5. Modelo en RPI de la lógica de control de un horno de microondas simple

RPI en sistemas embebidos. Esta realiza una comparación entre el estado actual del Sistema y el estado actual de la RPI, mediante la ecuación de salidas en (2). Estos estados deben ser equivalentes. De lo contrario, se deberá realizar un diagnóstico y, en su caso emplear un mecanismo de recuperación de errores. En esta etapa del trabajo, no se aborda el tema de diagnóstico y recuperación.

6. `TransControlablesHabilitadas()` permite verificar la existencia de transiciones manipulables habilitadas. Para esto, se utiliza la ecuación de habilitación en Definición 4 y la de manipulable en Definición 7.
7. La función `SeleccionarTransicion()` permite seleccionar una transición del conjunto de transiciones manipulables habilitadas para su disparo. En esta función se puede implementar un algoritmo de control, ya sea control por regulación [11] o control supervisor [8], el cual permitirá decidir acertadamente el siguiente evento a realizar.

8. La función `EscribirSalidas()` envía los comandos de control para realizar las actividades correspondientes a la transición seleccionada. El conjunto de comandos de control disponibles es Σ de la RPI, asignados a cada transición por λ .
9. `CalcularEstadoSiguiente()` utiliza la ecuación de estados de la RPI en (2) para calcular el estado siguiente de la RPI, después de la ejecución de los eventos correspondientes a la transición disparada.

EJEMPLO DE APLICACIÓN

Para ilustrar el método de convertir la RPI en un programa de control embebido, utilizaremos como ejemplo un electrodoméstico de uso común en los hogares, el horno de microondas y un PIC de 16 bits. En la Figura 5 se encuentra el modelo en RPI de un horno de microondas. Cada recuadro A, B, C, D, E, F representa una variable de estado para los componentes Botón de Potencia, Botón de Inicio, Temporizador, Magnetron (generador de las microondas), Puerta del Horno, Botón de Cancelar, que intervienen en el sistema del horno

de microondas, respectivamente. Los recuadros A, B, E y F representan el sistema de interfaz con el usuario. El recuadro D representa el dispositivo a controlar, es decir, el Magnetron, y el recuadro C representa su temporizador. Cada sensor de salida asociado a cada componente del horno, son entradas necesarias para el control del magnetron (generador de las microondas), mismas que están conectadas físicamente al PIC. Los únicos eventos controlables internamente son los asociados al magnetron y el temporizador. Se debe observar que los eventos asociados a los dispositivos de la interfaz con el usuario, son controlados externamente por él. Es decir, son incontrolables para el PIC.

Con la finalidad de simplificar el ejemplo y hacerlo más didáctico, se consideran las siguientes hipótesis:

1. El horno (magnetron) sólo funciona por tiempos de 30 segundos, y
2. Si se cancela o se abre la puerta, el horno deja de funcionar y el tiempo se reinicializa.

Para convertir la RPI de la Figura 5 a código embebido se utilizan los pasos descritos anteriormente. Estos son los siguientes:

1. Entradas al PIC:
{ $S_1, S_2, S_3, S_4, S_5, S_6$ }
2. Salidas del PIC:
{PrenderMagn_N, PrenderMagn_A, Reset, Potenc₁}
3. { $s_1 \rightarrow RB1, s_2 \rightarrow RB2, s_3 \rightarrow RB3, s_4 \rightarrow RB4, s_5 \rightarrow RB5, s_6 \rightarrow RB6$ } Normal \rightarrow con configuración del $TRISB = 0x003F$
4. {PrenderMagn_N \rightarrow RC1, PrenderMagn_A \rightarrow RC2, ApagarMagn_N \rightarrow RC3, ApagarMagn_P \rightarrow RC4} con configuración del $TRISC = 0x0000$
5. El resultado de la conversión al código embebido se muestra en la Figura 6.

A continuación se presenta una muestra de la ejecución del programa presentado.

1. Limpiar banderas (INTCON = 0;); Limpiar registros (PORTB = 0; PORTC = 0;); Definir pines de entrada y salida (TRISB = 0x003F; TRISC = 0x0000;); otras.

2. Calcula el estado actual. EstadoActual = $[0\ 1\ 0\ 30\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0]^T$
3. Leer las entradas PORT_Read = PORTB, inicialmente, los valores de PORTB_Read = 0x0000. Aquí se están leyendo todas las entradas al mismo tiempo.
4. Comparar la reconstrucción del estado actual del sistema real con el estado actual de la RPI. En este caso, sería $[0\ 1\ 0\ 30\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0]^T = [0\ 1\ 0\ 30\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0]^T$. Por lo tanto, no tenemos desviaciones.
5. No hay transiciones habilitadas.
6. En un ciclo posterior, suponiendo que un usuario pulsó el botón de inicio, entonces al leer las entradas, se tiene que PORTB_Read = 0x0004. La reconstrucción del estado es: $[0\ 1\ 0\ 30\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0]^T \neq [0\ 1\ 0\ 30\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0]^T$. Entonces, se disparó la transición no-manipulable t_{12} . Se debe actualizar el estado actual.
7. Ahora sí existen transiciones habilitadas manipulables.
8. Las transiciones habilitadas manipulables son: $\{t_4\}$, por lo tanto se selecciona ésta para su disparo.
9. Ahora se debe realizar el evento correspondiente a la transición t_4 : Prender el Magnetron con Potenc₁ Normal \rightarrow PORTC=0x0001.
10. Ahora, el estado siguiente de la RPI es $[0\ 1\ 0\ 30\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 1\ 0]^T$.

Este último marcado, permite el inicio del temporizador. Si no ocurre un evento externo, como que el usuario abra la puerta o pulse el botón de cancelar, entonces, cuando termine el temporizador, se tendrá que apagar el horno con el evento r de la transición t_3 , es decir, PORTC=0x0003.

V. CONCLUSIONES

Se ha presentado una metodología para convertir una RPI a código de control para su

```
#include <pic24fj256gb106.h>

// Definiciones de las entradas y salidas
#define S1 RB1
#define S2 RB2
#define S3 RB3
#define S4 RB4
#define S5 RB5
#define S6 RB6
#define S7 RB7

#define PrenderMagn_N RC1 //evento p
#define PrenderMagn_A RC2 //evento q
#define ApagarMagn_T RC3 //evento r
#define ApagarMagn_P RC4 //evento s
#define ApagarMagn_C RC5 //evento u
#define Reset RC6 //evento x
#define Tic RC7 //evento y

// Definición de constantes de la RPI
#define NL 13
#define NT 21

// Constantes en memoria de código
const unsigned char MatPre[NL][NT] =
{{0,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,0,0,0,0}
{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0}
{0,0,1,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,0,0,0,1,0,0,1,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,0,0,1,0,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,0,0,0,0,0,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1}
{0,0,0,1,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0}
{0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,0,1,0,0,0,0,0,1,0,0,0,0,0,0}};

const unsigned char MatPost[NL][NT] =
{{1,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,0,0}
{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0}
{0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,0,1,0,1,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,0,0,1,0,1,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,0,0,0,0,0,1,1,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1}
{0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0}
{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}
{0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0}};

//Global Variable Declarations
unsigned char EstadoActual[NL] = {0,1,0,30,0,1,0,0,0,1,1,0,0,1,1,0,0,0,0};
unsigned char EstadoSiguiente[NL];
```

Figura 6. Parte del código de implementación de la RPI que representa un horno de microondas simple

implementación en un sistema embebido. Se introdujo detalladamente la noción de una RPI, se ejemplificó el diseño de modelos en esta técnica, y la forma en que estas pueden ser traducidas a código embebido. Se pudo observar que la traducción de RPI a código embebido es sencilla y directa. Se utilizó el PIC de 16 bits PIC24FJ256GB106 de Microchip® para la

programación de la aplicación. Como parte del trabajo futuro se pretende agregar algún esquema de control y de recuperación de errores a la metodología.

VI. AGRADECIMIENTOS

Los autores agradecen a la Subsecretaría de Educación Superior e Investigación Científica por los apoyos de Fomento a la Generación y Aplicación Innovadora del Conocimiento y la Beca de Estudiante otorgados al proyecto “Metodologías para el Diseño de Sistemas Embebidos” mediante oficio PROMEP/103.5/08/2919.

VII. REFERENCIAS

- [1] Ken Arnold, *Embedded Controller Hardware Design* (Embedded Technology Series), Publisher Newnes; 1st edition (January 15, 2001). ISBN-10: 1878707523.
- [2] Doug Abbott, *Linux for Embedded and Real-Time Applications* (Embedded Technology), Publisher: Newnes; September 2002, ISBN-10: 0750675462.
- [3] Michael Barr, *Programming Embedded Systems in C and C++*, O'Reilly Media, Inc.; 1999. ISBN 1565923545.
- [4] Stephen Edwards, Luciano Lavagno, Edward A. Lee, and Alberto Sangiovanni-Vincentelli, “*Design of Embedded Systems: Formal Models, Validation and Synthesis*”, pp: 86-107 in Readings in hardware/software co-design, by Giovanni De Micheli, Rolf Ernst, Wayne Hendrix Wolf. Morgan Kaufmann; 1st edition, 2001. ISBN 1558607021.
- [5] Stuart Ball, *Embedded Microprocessor Systems: Real World Design*, Third Edition, Publisher: Newnes; 3rd edition, November 18, 2002, ISBN-10: 0750675349.
- [6] Jorg Desel and Javier Esparza, *Free Choice Petri Nets* (Cambridge Tracts in Theoretical Computer Science), Publisher: Cambridge University Press; New Ed

- edition, September 8, 2005, ISBN-10: 0521019451.
- [7] Manuel Silva Suárez, *Las redes de Petri: En la automática y la informática*, Editorial AC, 1985, ISBN-10: 8472880451, ISBN-16: 9788472880450.
- [8] R. Campos-Rodríguez, M. Alcaraz-Mejia, J. Mireles-García, *Supervisory Control of Discrete Event Systems using Observers*, Proceedings of the 15th Mediterranean Conference on Control & Automation. ISBN: 978-1-4244-1282-2, Julio, 2006. Atenas, Grecia.
- [9] L. Recalde and M. Silva and J. Ezpeleta and E. Teruel, *Petri nets and manufacturing systems: An examples-driven tour*, in Lectures on Concurrency and Petri Nets. Advances in Petri Nets, Springer-Verlag, 2004.
- [10] Alcaraz-Mejia, M.; Lopez-Mellado, E.; Ramirez-Trevino, A.; Rivera-Rangel, *Petri Net Based Fault Diagnosis of Discrete Event Systems*, IEEE Conf. on Systems, Man and Cybernetics, 2003, Págs: 4730 – 4735, Vol. 5, ISBN: 1062-922X, October, 2003. Washington, DC, USA.
- [11] R. Campos-Rodríguez, A. Ramírez-Treviño, E. López-Mellado, “Regulation Control of Partially Observed Discrete Event Systems”, en Proceeding of the 2004 IEEE International Conference on Systems, Man and Cybernetics, Vol. 2, Págs1837-1842.
- [12] Luis Alejandro Cortes, Luis Alej , Ro Corts, Petru Eles , Zebo Peng, “Verification of Embedded Systems using a Petri Net based Representation”, in Proc. ISSS, 2000.
- [13] Arnold, K., *Embedded Controller Hardware Design*, (2001). Newnes Pub.; 1st edition 2001. ISBN 1878707523.

VIII. AUTORES

Dra. Mildreth Alcaraz Mejía es Ingeniero en Sistemas Computacionales por el Instituto Tecnológico de Colima, 2001. Obtuvo los títulos de Maestro y Doctor en Ciencias con especialidad en Ingeniería Eléctrica en el CINVESTAV Unidad Guadalajara, en el 2004 y 2007, respectivamente. Actualmente, se dedica a la aplicación y desarrollo de técnicas de control embebidas con una orientación hacia el control de los Microsistemas.

Dr. Raúl Campos Rodríguez es Ingeniero en Computación por el Centro Universitario de la Ciénega, UDG, 2000. Estudió Maestría y Doctorado en Ciencias en Ingeniería Eléctrica en el CINVESTAV Guadalajara, de 2002 a 2007. Trabaja activamente en el diseño de algoritmos para sistemas de eventos discretos y ha dedicado últimamente parte de su tiempo al diseño de Microsistemas, o MEMS, específicamente en mecanismos basados en fuerza electrostática y la expansión térmica.

Laura García Contreras es estudiante de la carrera de ingeniería en computación del 8° semestre en el Centro Universitario de la Ciénega, de la Universidad de Guadalajara. Trabaja en este proyecto como parte de su tesis de licenciatura.

Joel Urquieta Jiménez es estudiante de la carrera de ingeniería en computación del 8° semestre en el Centro Universitario de la Ciénega, de la Universidad de Guadalajara. Trabaja en este proyecto como parte de su tesis de licenciatura.

Martín Gerardo Aguirre Villa es estudiante de la carrera de ingeniería en computación del 8° semestre en el Centro Universitario de la Ciénega, de la Universidad de Guadalajara. Trabaja en este proyecto con una beca de PROMEP para estudiantes en proyectos de investigación.