

PISTAS

Educativas

NUEVA ÉPOCA • No. 128 • FEBRERO 2018 • ISSN: 2448-847X

Número Especial



SEP
SECRETARÍA DE
EDUCACIÓN PÚBLICA



TECNOLÓGICO NACIONAL DE MÉXICO



PISTAS EDUCATIVAS

Pistas Educativas, Año 2018, No. 128, publicación especial SENIE 2017, publicada y editada por el Tecnológico Nacional de México dependiente de la Secretaría de Educación Pública, a través del Instituto Tecnológico de Celaya, Arcos de Belén Núm. 79, piso 3, Colonia Centro, Delegación Cuauhtémoc, CP 06010, Ciudad de México, Tel. 5536011000 Ext. 65064,

d_vinculacion05@tecnm.mx, Editor Responsable Héctor Rojas Garduño. **Reserva de derechos al uso exclusivo No. 04-2016-120613261600-203, ISSN: 2448-847X**, ambos son otorgados por el Instituto Nacional del Derecho de Autor.

Responsable de la última actualización de este número Julián Ferrer Guerra, Subdirector de Planeación y Vinculación, Instituto Tecnológico de Celaya, Antonio García Cubas Pte #600 esquina Av. Tecnológico, Colonia Alfredo V. Bonfil, CP 38010, Celaya, Gto, Tel. 4616117575 Ext 5106, fecha de término de la impresión o modificación.

Pistas Educativas es un espacio de libertad intelectual con responsabilidad; más allá del compromiso adquirido de formar ingenieros y administradores competentes, está el mandato constitucional para toda institución educativa de promover el desarrollo armónico de todas las facultades del ser humano y de educar para la democracia, como un sistema.

Las publicaciones de los artículos son sometidas a revisión por un comité de arbitraje y el contenido es responsabilidad de los autores y no necesariamente reflejan la postura del editor de la publicación. Queda prohibida la reproducción parcial o total de los contenidos e imágenes de la publicación sin previa autorización del Instituto encargado o si lo permite poner las condiciones.

DIRECTORIO TecNM

Manuel Quintero Quintero

Director



DIRECTORIO TecNM en Celaya

Ignacio López Valdovinos

Director

José Antonio Vázquez López

Subdirector Académico

Martín Campos Moreno

Subdirector de Servicios Administrativos

Julián Ferrer Guerra

Subdirector de Planeación y Vinculación

Teresita de las Nieves Armengol Rico

Jefe Departamento de Desarrollo Académico



PISTAS EDUCATIVAS

pistaseducativas@itcelaya.edu.mx

PISTAS **Educativas**

COMITÉ EDITORIAL

Editor General

MC. Julián Ferrer Guerra
Subdirector de Planeación y Vinculación
Tecnológico Nacional de México en Celaya

Editores Ejecutivos

Dr. José Antonio Vázquez López
Subdirector Académico
Tecnológico Nacional de México en Celaya

MDPH. Teresita de las Nieves Armengol Rico
Jefe Departamento de Desarrollo Académico
Tecnológico Nacional de México en Celaya

Editor Responsable

Ing. Héctor Rojas Garduño
Coordinador de Métodos y Medios Educativos
Tecnológico Nacional de México en Celaya

PISTAS EDUCATIVAS No. 128 (SENIE 2017), febrero 2018

Contenido

EDITORIAL.....	1
MODELADO Y CONTROL DE UN CONVERTIDOR BOOST EN DCM EMPLEADO EN UN SISTEMA FOTOVOLTAICO PARA TRABAJAR EN MODO RED Y EN MODO ISLA <i>Miguel Ángel Abundis Fong, Óscar Carranza Castillo, Jaime José Rodríguez Rivas, Rubén Ortega González.....</i>	2-18
INTERCOMUNICADOR ENLAZADO A RED DE TELEFONÍA CELULAR <i>Joel Fernando Acevedo Ruiz, Aldrin Barreto Flores, Verónica Edith Bautista López, Salvador Eugenio Ayala Raggi.....</i>	19-35
DESARROLLO DE UN PROCESO DE AUTENTICACIÓN FACIAL EN UN SISTEMA ANDROID UTILIZANDO EL ALGORITMO LDA (ANÁLISIS DE DISCRIMINACIÓN LINEAL) <i>Francisco Emiliano Aguayo Serrano, Jesús Carlos Pedraza Ortega, Edgar Alejandro Rivas Araiza, José Erik Rivas Araiza.....</i>	36-51
COMPARACIÓN DE LAS TÉCNICAS DE DETECCIÓN DE CRUCE POR CERO Y LA TRANSFORMADA Z-CHIRP PARA MEDIR FRECUENCIAS EN EL RANGO ULTRASÓNICO <i>Guadalupe Aguilar Cerda, Luis Morales Velázquez.....</i>	52-66
REALIDAD AUMENTADA CON MARCADORES CUADRADOS Y NATURALES PARA NAVEGACIÓN QUIRÚRGICA <i>Eliana Margarita Aguilar Larrarte, Oscar Andrés Vivas Albán, José María Sabater Navarro.....</i>	67-85
DETECCIÓN DE NECESIDADES Y DEFINICIÓN DE CONTENIDOS PARA LA ENSEÑANZA DE LA METODOLOGÍA DEL PERIODISMO DE DATOS: EL CASO DE DATAÍSTA <i>Belén Alazañez Cortés, Zayra Monserrat Miranda Aguirre, Jocelyn Lizbeth Molina Barradas, Rocío Abascal Mena.....</i>	86-100
METODOLOGÍA PARA LA IMPLEMENTACIÓN DEL MPM EN VHDL Y LA EMULACIÓN DE AMPLIFICADORES DE POTENCIA EN UNA TARJETA FPGA <i>Edgar Allende Chávez, José Ricardo Cárdenas Valdez, José Alejandro, Galaviz Aguilar, Andrés Calvillo Téllez, José Cruz Núñez Pérez.....</i>	101-117
METODOLOGÍA PARA LA INTEGRACIÓN DE UN MANIPULADOR MÓVIL BAJO SOFTWARE LIBRE <i>Oswaldo Alquisiris Quecha, Francisco Aguilar Acevedo, Ignacio Algreto Badillo, J. Jesús Arellano Pimentel.....</i>	118-130
APROXIMACIÓN AL RECONOCIMIENTO DE EMOCIONES FACIALES BASADO EN POSICIÓN DE PUNTOS DE INTERÉS <i>Víctor Manuel Álvarez Pato, Ramiro Velázquez Guerrero.....</i>	131-141
DISEÑO Y ANÁLISIS DE LA DE UN MOTOR DE CC MEDIANTE LA SELECCIÓN ÓPTIMA DE PARÁMETROS <i>Jesús Alejandro Álvarez Tostado Uribe, Irma Martínez Carrillo, Carlos Juárez Toledo.....</i>	142-153
DESIGN, CONSTRUCTION AND CHARACTERIZATION OF A THREE-CHANNEL COSMIC RAY DETECTOR BASED ON ALUMINUM BLOCKS ELECTRONICS <i>Luis Arceo, Julián Félix.....</i>	154-168
DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA DE MEDICIÓN COMBINADA DE 4-PUNTAS <i>Francisco Javier Arizaga Ayala, Arturo III Espinoza Duarte, José Antonio Gallardo Cubedo, Armando Gregorio Rojas Hernández.....</i>	169-178
COMPARACIÓN DE TARJETAS ARDUINO UNO ORIGINALES Y CLONES COMO INSTRUMENTO DE MEDICIÓN <i>Miguel Ángel Bañuelos Saucedo.....</i>	179-190
ANÁLISIS DE RENDIMIENTO DE LA PC ODROID C2 PARA SU USO EN ESQUEMAS DE CIUDADES INTELIGENTES <i>Ernesto Bardales Hernández, Saúl Lazcano Salas.....</i>	191-206

SISTEMA DE ADQUISICIÓN DE DATOS DE BAJO COSTO PARA UN INVERNADERO BASADO EN TECNOLOGÍA DE ACCESO LIBRE <i>Felipe de Jesús Becerra Woo, Araceli Gárate García, Tania Aglaé Ramírez del Real, Ervin Jesús Álvarez Sánchez.....</i>	207-218
ONLINE PARAMETRIC IDENTIFICATION OF MASS-SPRING-DAMPER MECHANICAL SYSTEMS USING ACCELERATION MEASUREMENTS <i>Francisco Beltrán Carbajal, Gerardo Silva Navarro, Luis Gerardo Trujillo Franco.....</i>	219-229
ROBOT MÓVIL (3,0) UNA EVALUACIÓN DE RENDIMIENTO <i>Saúl Enrique Benítez García, Jorge Luis de la Cruz Osorio, Miguel Gabriel Villarreal Cervantes.....</i>	230-247
BLOOD PRESSURE MEASUREMENT SYSTEM BASED ON OSCILLOMETRIC METHOD <i>Jessica Bolaños Olvera, Roque A. Osornio Ríos, Rosalía Reynoso Camacho.....</i>	248-265
INGENIERÍA ONTOLÓGICA APLICADA EN EL DISEÑO DE UN SISTEMA DE ONTOLOGÍAS PARA LA GESTIÓN DE HORARIOS <i>Maricela Claudia Bravo Contreras, Francisco Pavón Gutiérrez, José Alejandro Reyes Ortiz, Roberto Alfonso Alcántara Ramírez.....</i>	266-285
SELF-ORGANIZING MOBILE ROBOTS BASED ON MULTI-AGENT COORDINATION TECHNIQUES IMPLEMENTED WITH AERIAL VISION AND COMMUNICATION GATEWAY BETWEEN WIFI AND RF <i>Cynthia Daniela Briones Valencia, Zandor Alfredo Machaen Terriquez, Luis Martin del Castillo, Gustavo Alejandro Torres Blanco.....</i>	286-299
COMPARATIVA KINECT VS MYO APLICANDO LA PRUEBA NASA-TLX EN UN ENTORNO DE RVI PARA INSPECCIÓN EN AEROGENERADORES <i>Daniel Cantón Enríquez, J. Jesús Arellano Pimentel, Miguel Ángel Hernández López, Francisco Aguilar Acevedo.....</i>	300-317
ANÁLISIS DE ATAQUES DE RED DEL TIPO DHCP SPOOFING, TCP SYN FLOOD Y PAQUETES MALFORMADOS <i>Josué Cirilo Cruz, Arturo Zúñiga López, Carlos Avilés Cruz, Juan Villegas Cortez.....</i>	318-334
DISEÑO DE UN DEMODULADOR DE FM MEDIANTE PLL PARA LA INTERROGACIÓN DE SENSORES INTERFEROMÉTRICOS DE FIBRA ÓPTICA <i>Jesús Lorenzo Cisneros Hernández, Alejandro Rodríguez Antonio, Abimael Jiménez Pérez, José Mireles Jr. García, Rafael E. González Landaeta, Ángel Saucedo Carvajal.....</i>	335-351
REVISIÓN DE MÉTODOS PARA LA ESTIMACIÓN DE LOS ESTADOS DE CARGA Y SALUD DE UNA BATERÍA <i>Alina Araceli Contreras Sillero, Nimrod Vázquez Nava, Claudia Verónica Hernández Gutiérrez, Jeziel Vázquez Nava, Joaquín Vaquero López.....</i>	352-368
APLICACIÓN MÓVIL PARA EL CÁLCULO DE RUTAS “LOBOBICI” EN CIUDAD UNIVERSITARIA BUAP BASADA EN BÚSQUEDAS <i>Eliúh Cuecuecha Hernández, José Javier Martínez Orozco, Daniel Méndez Lozada, Adán Zambrano Saucedo, Aldrin Barreto Flores, Verónica Edith Bautista López, Salvador Eugenio Ayala Raggi.....</i>	369-381
SISTEMA DE RECONOCIMIENTO DE VOCALES DE LA LENGUA DE SEÑAS MEXICANA <i>Eliúh Cuecuecha Hernández, José Javier Martínez Orozco, Daniel Méndez Lozada, Adán Zambrano Saucedo, Aldrin Barreto Flores, Verónica Edith Bautista López, Salvador Eugenio Ayala Raggi.....</i>	382-394
PROPUESTA DE UN ENTRENADOR MIOELÉCTRICO BASADO EN UNA APLICACIÓN MÓVIL <i>Humberto de la Cruz Regalado, Carlos Edgar López Barrera, Eduardo Emmanuel Rodríguez López, Luis Mariano Sandoval González, Alfredo Ramírez García.....</i>	395-411
FUSIÓN MORFOLÓGICA DE IMÁGENES IR Y VISUALES UTILIZANDO EL MODELO LIP <i>Oscar Ricardo Delfín Santiesteban, Iván Ramón Teról Villalobos.....</i>	412-431
MÓDULO DE CONTROL DE CARGA PARA EVALUAR CELDAS DE COMBUSTIBLE -HARDWARE- <i>Shirley Yahaira Echánove Gómez, Marco Antonio Rodríguez Blanco, Juan Manuel Tadeo Sierra Grajeda, Luis Enrique Vidal Burelo.....</i>	432-447

EFFECTO DE LA LONGITUD DEL DIÁMETRO EN LA ESTABILIDAD TÉRMICA DE LA CAPA LIBRE DE LAS MEMORIAS RAM MAGNÉTICAS

Marco A. Escobar, Rafael Guzmán-Cabrera, Miguel Torres-Cisneros, Jorge Ramón Parra-Michel, Rafael Martínez-Peláez..... 448-458

DISEÑO Y CONSTRUCCIÓN DE PROTOTIPO DE ENTRENAMIENTO PARA PRÁCTICAS EN INSTRUMENTACIÓN Y CONTROL

Eladio Flores Martínez, Pablo Reyna Guerra, José Luis Jiménez Reyes, Javier Garrido Meléndez, Quetzalcóatl Cruz Hernández Escobedo..... 459-472

SISTEMAS PARA LA EXTRACCIÓN DE FRASES CLAVE EN DOCUMENTOS CIENTÍFICOS

Gerardo Flores Petlascalco, Mireya Tovar Vidal, Hilda Castillo Zacatelco, José A. Reyes-Ortiz..... 473-486

SIMULACIÓN Y CONTROL DEL PROCESO DE MACERACIÓN DE UNA CERVECERÍA ARTESANAL

Jesús Antonio Flores Tovar, Miguel Magos Rivera, José Antonio Lara Chávez, José Manuel Domínguez Martínez, Juan Alberto Godínez Viveros..... 487-505

DISEÑO E IMPLEMENTACIÓN DE UN MULTÍMETRO DIGITAL CON FUNCIONES AMPLIADAS DE BAJO COSTO

Rafael García Arredondo, Juan Carlos Gómez Cortez, Diego de Jesús Padierna Arvizu, José Eleazar Peralta López, Francisco Javier Pérez Pinal, Luís Antonio Ramírez Arredondo, Julio Cesar Regalado Sánchez..... 506-522

SISTEMA DE MONITOREO DE TEMPERATURA EN RED

Ricardo Godínez Bravo, Miguel Magos Rivera..... 523-538

IMPLEMENTACIÓN DE EFECTOS DE SONIDO PARA GUITARRA ELÉCTRICA EN LA TARJETA C6713 DSK

Claudia Gómez Borrás, Javier Alducin Castillo..... 539-556

DISEÑO DE UN CIRCUITO DE CONTROL DE ILUMINACIÓN PARA UN SISTEMA FORMADOR DE IMÁGENES DE PURKINJE

Armando Gómez Vieyra, Ezequiel Martínez Solís, Juan Jesús Ocampo Hidalgo, Karla Beatriz Vergara Vázquez, Uriel Calderón Uribe, Geovanni Hernández Gómez..... 557-571

ANÁLISIS COMPARATIVO DE LOS TIEMPOS DE EJECUCIÓN SOBRE SBC PARA DOS SISTEMAS OPERATIVOS DE TIEMPO REAL

Diana Lizet González Baldovinos, José Luis Cano Rosas, Pedro Guevara López..... 572-585

GUÍAS DE DISEÑO WEB PARA FACILITAR EL ACCESO A LA INFORMACIÓN DESDE TELÉFONOS INTELIGENTES

Beatriz A. González Beltrán, Araceli Granados García..... 586-606

ESTUDIO DE UNA ANTENA DE MICROCINTA FRACTAL TIPO E PARA LA BANDA DE LOS 2.4 GHZ

Iván R. González Rangel, Javier Vargas Sánchez, Genaro Hernández Valdez, Mario Reyes Ayala, J. R. Miranda Tello, Edgar A. Andrade González..... 607-620

CONDITIONING AND SIGNAL AMPLIFICATION STAGES FOR A SMART GAS MICROSENSOR MEMS

J. L. González Vidal, M. A. Reyes Barranca, E. N. Vázquez Acosta..... 621-636

TÉCNICA DE CONMUTACIÓN SUAVE PARA UN CONVERTIDOR REDUCTOR-ELEVADOR DOBLE CON APLICACIONES EN ILUMINACIÓN

Pablo Israel Guzmán Tafoya, Nimrod Vázquez Nava, René Osorio Sánchez..... 637-651

DISEÑO Y SIMULACIÓN DE RODILLA MECÁNICA MONOCÉNTRICA

José Alexis Hernández Aguilar, Ervin Jesús Álvarez Sánchez, Andrés López Velázquez..... 652-668

DISPOSITIVO DE ILUMINACIÓN LED CON INCORPORACIÓN DE ELECTRÓNICA DIGITAL Y CONTROL DESDE ANDROID POR BLUETOOTH

Mario Alberto Hernández Alves, Leonardo Sánchez, José A. Reyes Ortiz..... 669-685

DESIGN AND FABRICATION OF A 64-QAM MODULATOR FOR ANALYSIS OF SIGNALS BETWEEN STAGES

Jorge Andrés Hernández Carrillo, José Ricardo Cárdenas Valdez, Virgilio Rosendo Pérez, Manuel de Jesús García Ortega, Andrés Calvillo Téllez..... 686-698

INTEGRACIÓN DE UN SISTEMA CEREBRO COMPUTADORA EMPLEANDO SOFTWARE LIBRE <i>Irving Ulises Hernández Miguel, Alejandro Jarillo Silva, Víctor Alberto Gómez Pérez.....</i>	699-715
ELEMENTOS DE LOS PARQUES EÓLICOS QUE DEBEN SER CONTROLADOS PARA SU INTERCONEXIÓN CON REDES ELÉCTRICAS <i>Jorge Eduardo Hernández Miranda, Irvin López García, Eduardo Campero Littlewood, Francisco Beltrán Carbajal, Víctor Manuel Jiménez Mondragón.....</i>	716-729
DETECCIÓN ACTIVA DE FALTAS EN SISTEMAS DE EVENTOS DISCRETOS <i>Karen Hernández Rueda, María E. Meda Campaña, Bernardo Haro Martínez.....</i>	730-748
CONTROL DEL FLUJO DE POTENCIA HACIA LA RED ELÉCTRICA DE UN SISTEMA DE GENERACIÓN EÓLICA EMPLEANDO UN GENERADOR DE INDUCCIÓN DE DOBLE ALIMENTACIÓN <i>Pedro Hernández Tenorio, Jaime José Rodríguez Rivas, Oscar Carranza Castillo, Rubén Ortega González.....</i>	749-766
OBTENCIÓN DEL MÁXIMO ANCHO DE BANDA PARA LA ADQUISICIÓN Y RECONSTRUCCIÓN DE SEÑALES ANALÓGICAS CON LA TARJETA SPARTAN 3E <i>Enrique Gerardo Hernández Vega, Jorge Alberto Ortiz Gallo, Daniel Eduardo Morales Fernández, Alejandro Verduzco Hernández.....</i>	767-780
MÉTODO DE INSTRUMENTACIÓN INDIRECTA BASADO EN ONDAS ACÚSTICAS DERIVADAS DE VIBRACIONES MECÁNICAS PARA LA ESTIMACIÓN DE VELOCIDAD ANGULAR EN MAQUINARIA ROTATIVA <i>Enrique Gerardo Hernández Vega, Sergio Iván Chavarría Estrada.....</i>	781-798
DESIGN, CONSTRUCTION AND SIMULATION OF A UNIFORM MAGNETIC FIELD GENERATOR WITH STEEL NUCLEUS TO DEFLECT COSMIC RAYS <i>Karla Natalia Herrera Guzmán, Raúl Alejandro Gutiérrez Sánchez, Jorge Luis Arceo Miquel, Julián Félix.....</i>	799-814
DISEÑO E IMPLEMENTACIÓN DE SENSORES DE GAS QCM DE ALTA SENSIBILIDAD PARA UNA NARIZ ELECTRÓNICA <i>Juan Jesús Jiménez Arellano, Severino Muñoz Aguirre, Juan Castillo Mixcoatl, Georgina Beltrán Pérez, José Lorenzo Muñoz Mata.....</i>	815-828
SIMULACIÓN DE ESTRATEGIAS DE BÚSQUEDA EN ANIMALES CON POSIBLES APLICACIONES EN COMPUTACIÓN Y ROBÓTICA <i>Joel Ricardo Jiménez Cruz.....</i>	829-847
EMULACIÓN EN FPGA DE TÉCNICA PARA CORRECCIÓN DEL DESEQUILIBRIO I/Q APLICADO EN UN MODULADOR DIGITAL 256-QAM <i>Sergio Alberto Juárez Cazares, Aldo Bonilla Rodríguez, José Cruz Núñez Pérez.....</i>	848-863
SINTONIZACIÓN DE UN CONTROLADOR PI APLICADO A UN HORNO EXPERIMENTAL A PARTIR DE LA IDENTIFICACIÓN DE MÚLTIPLES PUNTOS DE LA RESPUESTA EN FRECUENCIA UTILIZANDO UN ALGORITMO GENÉTICO <i>Cecilia de los A. Keb Chulin, César I. Coyoc y Coyoc, J. Rubén Lagunas Jiménez, Víctor M. Moo Yam.....</i>	864-877
A NEURO-FUZZY BASED CONTROL OF A SIMULATED SOFC IN A GRID CONNECTED ENVIRONMENT <i>Sohail Khan, Juan Carlos Olivares Galván, Rafael Escarela Pérez.....</i>	878-890
SISTEMA DE MONITOREO PARA UN EQUIPO DE ESTUDIOS DE TIEMPOS Y MOVIMIENTOS <i>José Antonio Lara Chávez, Miguel Magos Rivera, Miguel Ángel Figueroa Sánchez, Miguel Ángel López Ontiveros, Lisaura Walkiria Rodríguez Alvarado, Jesús Loyo Quijada.....</i>	891-908
MÉTODOS NUMÉRICOS EN INGENIERÍA UAM AZCAPOTZALCO: BAOC (BIG ACADEMIC OPEN COURSE) <i>Hugo Pablo Leyva, Rafaela Blanca Silva López, Rafael Morales Gamboa.....</i>	909-925
DISPOSITIVO TELEMÉTRICO PARA MONITOREO DE FRECUENCIA CARDIACA Y SATURACIÓN DE OXÍGENO <i>Lorena Lomelí Herrera, Federico Aguayo Ríos, Rafael Martínez Peláez.....</i>	926-943
IMPLEMENTACIÓN DE BLOQUES PARA CONTROLADORES DIFUSOS ANALÓGICOS CON CIRCUITOS CMOS Y OPAMPS	

Edgar López Delgadillo, Luis Alejandro Flores Oropeza, Luis Enrique Arámbula Miranda, Alfonso Vela Rivera, Martín Isaac Falcón Segovia..... 944-954

INSTRUMENTACIÓN VIRTUAL DE UN SISTEMA DE GENERACIÓN EOLOELÉCTRICO INTERCONECTADO A LA RED
Adolfo Rafael López Núñez, Jesús Darío Mina Antonio, Roberto Carlos Gómez Hernández, Gabriel Calderón Zavala, Oscar Hernández Martínez..... 955-971

PROTOTIPO DIDÁCTICO PARA NIÑOS DE PRIMARIA BASADO EN LA IDENTIFICACIÓN DE EMOCIONES Y SUS CONSECUENCIAS
Erick López Ornelas, Rocío Abascal Mena..... 972-984

CONSTRUCCIÓN DE MAPAS DE ISÓCRONAS PARA LA ZONA PONIENTE DE LA CIUDAD DE MÉXICO
Erick López Ornelas, Rocío Abascal Mena, Santiago Avilés Vázquez..... 985-998

METODOLOGÍA PARA LA CORRECCIÓN DE DISTORSIÓN GEOMÉTRICA Y RECONSTRUCCIÓN 3D DE UN OBJETO MEDIANTE PERFILOMETRÍA WAVELET 1D
Claudia Victoria López Torres, Gonzalo Elías Blanco Silva, Jesús Carlos Pedraza Ortega, Juan Manuel Ramos Arrequin, José Emilio Vargas Soto, Mayra Azucena Cíntora García..... 999-1013

SUPERVISIÓN DE TEMPERATURA Y HUMEDAD PARA EL CÁLCULO DE BALANCE ENERGÉTICO EN UN INVERNADERO CON TIEMPOS DE MUESTREO OBTENIDOS DE FORMA EXPERIMENTAL
Sergio Eduardo Luna Arauz, Andrés Alfonso Andrade Vallejo, Pedro Guevara López, Juan Enrique Rubiñoz Panta..... 1014-1027

SENSOR CAPACITIVO DE ALERTA PARA IDENTIFICAR IMPUREZAS EN ACEITE DE MOTORES DIESEL
Hiram U. Luna López, Antonio Ramos Carrasco, María Elena Anaya Pérez, Dainet Berman Mendoza..... 1028-1041

IMPLEMENTACIÓN EN “HARDWARE IN THE LOOP” DEL SISTEMA CARRO-PÉNDULO INVERTIDO CON BASE EN EL MICROCONTROLADOR HÉRCULES RM57L843 DE TEXAS INSTRUMENTS
Marcelino Martínez Aragón, Fermín Hugo Ramírez Leyva, José Anibal Arias Aguilar..... 1042-1058

DIGITAPRENDE: UNA APLICACIÓN PARA LA ALFABETIZACIÓN DIGITAL DE ADULTOS MAYORES
Daniel Martínez Espino, Alba Rocío Núñez Reyes, Dra. Rocío Abascal Mena..... 1059-1075

IMPLEMENTACIÓN DE UN SISTEMA DE ADMINISTRACIÓN Y CONTROL PARA UNA MICRO-RED DE CD UTILIZANDO PLATAFORMAS DE NATIONAL INSTRUMENTS
Juan José Martínez Nolasco, José Alfredo Padilla Medina, Elías José Juan Rodríguez Segura, Agustín Sancén Plaza..... 1076-1093

SIMULADOR TRIDIMENSIONAL DE LA CINEMÁTICA DEL ROTOR DE UN AEROGENERADOR TRIPALA CON BASE EN LA CONVENCION D-H
Ana Patricia Matus Vicente, Miguel Ángel Hernández López, Francisco Aguilar Acevedo, J. Jesús Arellano Pimentel..... 1094-1107

SISTEMA DE EVALUACIÓN POR COMPETENCIAS INTEGRADO A UNA PLATAFORMA EDUCATIVA INSTITUCIONAL
Víctor Hugo Medina Sandoval, Arturo García Nevares, Miguel Ángel Rodríguez Ortiz..... 1108-1121

INSTRUMENTACIÓN Y MONITOREO POR RED INALÁMBRICA DE SENSORES MEDIANTE XBEE PARA UN PROCESO DE POLIMERIZACIÓN
Cristian Mena Saucedo, Brian Manuel González Contreras, Miguel Ángel Munive Rojas, Fermín Martínez Solís..... 1122-1141

CASO APLICATIVO DEL SISTEMA DE GESTIÓN DIGITAL: GESTIÓN DE PROYECTOS DE INVESTIGACIÓN
Iris Iddaly Méndez Gurrola, César Augusto Briseño Moreno, Rafaela Blanca Silva López..... 1142-1157

IDENTIFICACIÓN DE LOS FACTORES ADVERSOS QUE INFLUYEN EN LOS JÓVENES EGRESADOS PARA INCORPORARSE AL CAMPO LABORAL
Luis Alberto Morales Rosales, Mariana Lobato Báez, Ignacio Algreto Badillo, Héctor Rodríguez Rangel... 1158-1173

CONVERTIDOR BIDIRECCIONAL MULTIFASE PARA APLICACIONES DE MICRO REDES DE CD
Jorge Rubén Morfín Orozco, Heriberto Rodríguez Estrada, Elías Rodríguez Segura..... 1174-1190

SEGMENTACIÓN DE IMÁGENES APLICANDO LA HERRAMIENTA COMPUTACIONAL P3S <i>Antonio Orantes Molina, Irwin Jovany Salinas Vargas, Raúl Cruz Barbosa, Rosebet Miranda Luna, Verónica Rodríguez López.....</i>	1191-1205
SIMULACIÓN BASADA EN AGENTES PARA EL CONTROL INTELIGENTE DE SEMÁFOROS MEDIANTE LÓGICA DIFUSA <i>Héctor Rafael Orozco Aguirre, Saul Lazcano Salas, Víctor Manuel Landassuri Moreno.....</i>	1206-1223
PREDICCIÓN DE POTENCIA FOTOVOLTAICA MEDIANTE REDES NEURONALES WAVELET <i>José Daniel Ortiz López, Luis J. Ricalde Castellanos, Braulio J. Cruz Jiménez, Ricardo J. Peón Escalante..</i>	1224-1236
SISTEMA AUTOMÁTICO DE INSPECCIÓN DE COMPONENTES MEDIANTE VISIÓN POR COMPUTADORA <i>Iván César Palacios Aguayo, Ramiro Velázquez Guerrero.....</i>	1237-1251
CONTROL DIFUSO PARA UN CONVERTIDOR CD-CD APLICADO A SISTEMAS FOTOVOLTAICOS EN LOS MODOS MPPT Y CV <i>Julio Cesar Peña Aguirre, Agustín Ramírez Agundis, Elías J. J. Rodríguez Segura, Juan José Martínez Nolasco.....</i>	1252-1269
DESARROLLO DE SOFTWARE SCADA PARA PLANTA PILOTO DE CONCRETO SECO CON PROTOCOLO ETHERNET/IP <i>Yolanda Pérez Pimentel, Ismael Osuna Galán.....</i>	1270-1285
REDUCCIÓN DE CROSS-TALKING POR MEDIO DEL USO DE FOCALIZADORES EN APLICACIONES DE ULTRASONIDO <i>Jovan Alejandro Ramírez Guzmán, Ignacio Herrera Aguilar, Gerardo Águila Rodríguez, Oscar Osvaldo Sandoval González.....</i>	1286-1296
ANÁLISIS DEL MÉTODO DE CORRIMIENTO DE FASE PARA ESCANEADO Y RECONSTRUCCIÓN 3D DE OBJETOS <i>Carlos Alberto Ramos Arrequin, Juan Carlos Moya Morales, Rodrigo Escobar Díaz Guerrero, Jesús Carlos Pedraza Ortega.....</i>	1297-1311
VOLTMETRO BLUETOOTH Y DESPLIEGUE EN SMARTPHONE <i>Fernando Reyes Avilés, Carlos Avilés Cruz.....</i>	1312-1330
SISTEMA PARA EL MONITOREO DE OPINIÓN CENTRADO EN ENTIDADES A PARTIR DE TWITTER <i>José Alejandro Reyes Ortiz, Ezra Saucedo Vargas, Ángeles Belém Priego Sánchez.....</i>	1331-1346
CLASIFICACIÓN DE REPORTES CLÍNICOS PARA APOYAR EL DIAGNÓSTICO DEL CÁNCER <i>José Alejandro Reyes Ortiz, Beatriz Adriana González Beltrán, Mireya Tovar Vidal.....</i>	1347-1361
MODELADO Y CONTROL DEL GIROSCOPIO ECP-750 <i>José Manuel Reyes Rodríguez, Fernando Reyes Cortés, J. Eligio Moisés Gutiérrez Arias.....</i>	1362-1375
SIMULACIÓN DEL CONTROL Y LA COORDINACIÓN DE UN ROBOT EXPLORADOR EN UN AMBIENTE AGRÍCOLA <i>Marisol Rodríguez Cuevas, Andrés Fernando Jiménez López, Pedro F. Cárdenas H.....</i>	1376-1391
DETECCIÓN DE FALLA DE RODAMIENTO EN UNA CADENA CINEMÁTICA VÍA EMISIÓN ACÚSTICA <i>Luis Alejandro Romero Ramírez, Luis Morales Velázquez, Roque A. Osornio Ríos, René de Jesús Romero Troncoso, Daniel Moríñigo Sotelo.....</i>	1392-1406
DISEÑO, CONSTRUCCIÓN Y PRUEBAS DE UN DETECTOR HÍBRIDO DE RAYOS CÓSMICOS DE 4 CANALES <i>Francisco Javier Rosas Torres, Julián Félix.....</i>	1407-1420
TRANSMISIÓN-RECEPCIÓN DE AUDIO VÍA LUZ VISIBLE <i>Sergio Sandoval Reyes, Arturo Hernández Balderas.....</i>	1421-1433
DISEÑO E INTEGRACIÓN DE UNA COMPUTADORA A BORDO PARA VUELOS ESTRATOSFÉRICOS <i>Lauro Santiago Cruz, Indira Citlalli Cortés Sánchez, Mario Alberto Mendoza Bárcenas.....</i>	1434-1447
CASO APLICATIVO DEL SISTEMA DE GESTIÓN DIGITAL: GESTIÓN DE ESPACIOS FÍSICOS	

Rafaela Blanca Silva López, César Arostégui Ramírez, Iris Iddaly Méndez Gurrola, Hugo Pablo Leyva.....	1448-1465
ANÁLISIS DE UN SISTEMA DE ENFRIAMIENTO DEL CPU DE UNA COMPUTADORA EMBEBIDA POR MEDIO DE UNA CELDA PELTIER	
Víctor Daniel Tejeda Mejía, Miguel Ángel Olivares Robles, Pedro Guevara López.....	1466-1478
SIMULACIÓN “HARDWARE IN THE LOOP” DE UN INVERSOR TRIFÁSICO CONECTADO A LA RED ELÉCTRICA	
Manuel Tlapa Juárez, Edgar Peralta Sánchez, Juan Marcos Ruiz Dávila, Sergio Alejandro Cardeña Moreno, Félix Quirino Morales.....	1479-1495
IMPLEMENTACIÓN DE UN DETECTOR DE CAÍDAS PARA SU APLICACIÓN EN PACIENTES HOSPITALIZADOS Y PERSONAS DE LA TERCERA EDAD	
José Luis Vázquez Ávila, Walter Ariel Silva Martínez, Rafael Sánchez Lara, Casandra Sánchez Galván.....	1496-1507
DISEÑO DE SISTEMAS HIPERCAÓTICOS PARA IMPLEMENTACIÓN EN DISPOSITIVOS LÓGICOS PROGRAMABLES ENFOCADO A APLICACIONES DE SEGURIDAD	
Jorge Gustavo Vázquez Duran, Ramón Ramírez Villalobos, Luis Néstor Coria de los Ríos, Manuel de Jesús García Ortega.....	1508-1517
SISTEMA DE CÁLCULO DEL CONSUMO ELÉCTRICO DE LA UAM AZCAPOTZALCO	
Rodrigo Vázquez López, Eduardo Campero Littlewood, Felipe González Montañez, Juan Carlos Olivares Galván, Raúl Arturo Ortiz Medina.....	1518-1530
IMPLANTACIÓN DE UNA LPWAN PARA MONITOREO DE TEMPERATURA Y HUMEDAD EN UN INVERNADERO	
José Ignacio Vega Luna, Mario Alberto Lagos Acosta, Gerardo Salgado Guzmán, Víctor Noé Tapia Vargas, Francisco Javier Sánchez Rangel, José Francisco Cosme Aceves.....	1531-1548
INVENTARIO DE MÁQUINAS EXPENDEDORAS USANDO UNA LPWAN	
José Ignacio Vega Luna, Mario Alberto Lagos Acosta, Gerardo Salgado Guzmán, Víctor Noé Tapia Vargas, Francisco Javier Sánchez Rangel, José Francisco Cosme Aceves.....	1549-1566
MÉTODO EXPERIMENTAL DE ESTIMACIÓN DE LA FUNCIÓN DE TRANSFERENCIA DE UN MOTOR DE CD UTILIZANDO ENCODER DE CUADRATURA	
Jorge Fernando Vera Centeno, Ignacio Herrera Aguilar, Gerardo Águila Rodríguez, Oscar Osvaldo Sandoval González, Blanca Estela González Sánchez.....	1567-1582
MODELOS DE TECNOLOGÍAS DEL BIG DATA ANALYTICS Y SU APLICACIÓN EN SALUD	
Gustavo Verduzco Reyes, Ernesto Bautista Thompson, Jorge A. Ruiz Vanoye, Alejandro Fuentes Penna... 1583-1599	
ROBOT CARTESIANO DE 3 GDL PARA INSPECCIÓN DE ESFUERZOS RESIDUALES MEDIANTE PRINCIPIO DE FOTOELASTICIDAD	
Ángel Vergara Betancourt, Fernando García Ortiz, José Guadalupe Gaona Reyes, Carlos Cortés Martínez. 1600-1615	
BRAZO ROBÓTICO CONTROLADO POR MEDIO DE VISIÓN COMPUTACIONAL UTILIZANDO UN KINECT	
Celina Villicaña González, María Teresa Orvañanos Guerrero, Eduardo Rodríguez Figueroa.....	1616-1632
DEVELOPMENT OF A WIRELESS SIGNAL ACQUISITION SYSTEM FROM SENSORS FOR COMFORT AND ENERGY QUALITY	
Israel Zamudio Ramírez, Arturo Yosimar Jaen Cuellar, Roque Alfredo Osornio Ríos.....	1633-1647
MAXIMUM PRINCIPLE FOR TIME MINIMIZATION OF CIRCUIT DESIGN PROCESS	
Alexander Zemliak.....	1648-1663
ANALYSIS OF DIFFERENT STRATEGIES FOR CIRCUIT OPTIMIZATION	
Alexander Zemliak, Fernando Reyes Cortés.....	1664-1677

EDITORIAL

Pistas Educativas en su número 128, tiene el agrado de presentar el número especial de la décimo tercera semana nacional de Ingeniería Electrónica –SENIE 2017-, que se llevó a efecto entre el 4 y el 6 de octubre del 2017 bajo la organización conjunta de la División de Ciencias Básicas e Ingeniería de la Universidad Autónoma Metropolitana Azcapotzalco, así como la Universidad de la Salle Bajío, fungiendo esta última institución como anfitriona del evento. Cada uno de los artículos que aquí se publican fue sometido a la consideración de dos investigadores expertos en el tema (doble ciego), teniendo el Comité Técnico de SENIE 2017 la responsabilidad en lo que se refiere a su calidad técnica. Por su parte, Pistas Educativas tuvo bajo su cuidado la edición de los textos de acuerdo, como siempre, con sus normas de publicación. A través de esta edición, el Tecnológico Nacional de México en Celaya, por medio de Pistas Educativas, avanza en su cometido de ser un vehículo para la difusión del conocimiento, albergando en esta ocasión las experiencias y logros de los que dan cuenta en sus artículos los estudiantes, profesores e investigadores de buena parte del sistema educativo nacional que se congregaron en SENIE 2017.

MODELADO Y CONTROL DE UN CONVERTIDOR BOOST EN DCM EMPLEADO EN UN SISTEMA FOTOVOLTAICO PARA TRABAJAR EN MODO RED Y EN MODO ISLA

Miguel Ángel Abundis Fong

Instituto Politécnico Nacional, SEPI-ESIME
abundisfong@hotmail.com

Óscar Carranza Castillo

Instituto Politécnico Nacional, Escuela Superior de Cómputo y SEPI-ESIME
ocarranzac@ipn.mx

Jaime José Rodríguez Rivas

Instituto Politécnico Nacional, SEPI-ESIME
jjrodriguezr@ipn.mx

Rubén Ortega González

Instituto Politécnico Nacional, Escuela Superior de Cómputo y SEPI-ESIME
rortegag@ipn.mx

Resumen

En este artículo se presenta el modelado de un convertidor CD-CD tipo Boost en el Modo de Conducción Discontinua (DCM); así como el diseño de sus lazos de control, siguiendo el esquema de control por Modo Corriente Media (ACC). El convertidor electrónico de potencia tipo Boost forma parte de un sistema que emula la generación fotovoltaica mediante paneles solares en el contexto de microrredes. El objetivo es utilizar el convertidor Boost para elevar y regular el voltaje generado por los paneles, de manera que este sirva como tensión de entrada a un inversor monofásico tipo puente completo, para su posterior conexión con la red eléctrica, o bien, para su operación en modo Isla alimentando una carga. Los controladores diseñados son validados mediante simulación, verificando que cumplan con los requisitos de sobretiro máximo y tiempo de

asentamiento. Los resultados obtenidos, comprueban que el controlador diseñado tiene una respuesta transitoria aceptable ante perturbaciones en la entrada del sistema.

Palabras Claves: Control modo, corriente media, convertidor Boost, microrredes, modo de conducción discontinua.

Abstract

In this paper, a Discontinuous Conduction Mode model and an Average Current Control scheme are developed for a DC-DC Boost type power converter. The Boost converter is part of a system that emulates photovoltaic generation using solar panels in the context of microgrids. The objective is to use the Boost converter to rise and regulate the voltage generated by the panels, so that it serves as the input voltage to a single-phase full-bridge inverter, for later connection to the power grid, or for its operation In Island mode by supplying power to a load. The designed controllers are validated by means of computer simulation, verifying that they satisfy the requirements of maximum overshoot and settling time. The results obtained show that the designed controller has an acceptable transient response to disturbances at the system input.

Keywords: Average current control, Boost converter, discontinuous conduction mode, microgrids.

1. Introducción

En los sistemas eléctricos de potencia actuales se procura tener una mayor participación de las fuentes de energía renovables. Este hecho plantea el desafío de contar con un mejor diseño y control de los convertidores electrónicos de potencia que son parte integral en los esquemas de Generación Distribuida de las Redes Inteligentes (*Smart Grids*) [Sharkawi, 2013].

La energía solar fotovoltaica es una de las fuentes renovables que más desarrollo ha tenido en los últimos años. Estos sistemas de generación presentan las desventajas de una baja eficiencia y baja potencia, además de que es intermitente a lo largo del día. Sin embargo, es uno de los recursos más abundantes y, en

ocasiones, es la opción más viable para proveer de electricidad a regiones sin acceso a la red eléctrica [Labouret, 2010]. Por tal motivo, es necesario contar con convertidores electrónicos que regulen con precisión los niveles de energía generada y que la distribuyan a sus cargas, cumpliendo con la amplitud y frecuencia establecidas, y con los niveles de distorsión armónica total permisibles. En [He, 2004] se presenta un modelo ACC híbrido para un convertidor Boost en conducción continua (CCM) utilizando un microcontrolador de 8 bits. Un modelo ACC robusto se describe en [Benavent et al., 2005] para disminuir las limitaciones derivadas del cero en el semiplano derecho de la función de transferencia de voltaje de salida respecto al control. Sin embargo, ambos trabajos contemplan sólo la operación en modo Isla. La transición entre los modos Isla y Red es estudiada en [Carrillo et al., 2016] para la operación en CCM, presentándose así el efecto de recuperación inversa en el Boost.

En el presente artículo se propone el control ACC como esquema de control de un convertidor Boost, el cual es parte de un sistema de generación fotovoltaica. El convertidor Boost opera en el Modo de Conducción Discontinua (DCM) a manera de evitar el efecto de recuperación inversa en el interruptor de potencia [Mohan, 2003]. La figura 1 muestra el esquema del sistema bajo estudio.

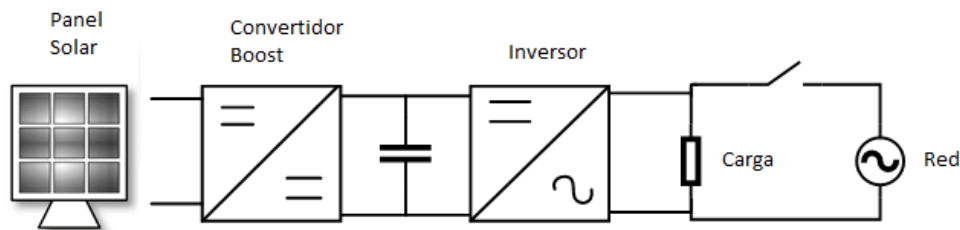


Figura 1 Esquema del sistema fotovoltaico.

El sistema está compuesto por una fuente programable de CD que representa la energía variable generada por un arreglo de paneles solares. A su salida se conecta el convertidor Boost, y a la salida de este, un inversor tipo puente completo. En la salida del inversor se conecta un filtro LCL para reducir el efecto de los armónicos y un transformador de relación 1:1 para proporcionar

aislamiento. El inversor y el filtro LCL están fuera del alcance de este artículo. El sistema está contemplado para operar en modo Isla, alimentando una carga; así como, para su operación en modo Red, inyectando la energía a la red de distribución eléctrica. La figura 2 muestra el circuito del sistema para el modo Isla, y la figura 3, para el modo Red.

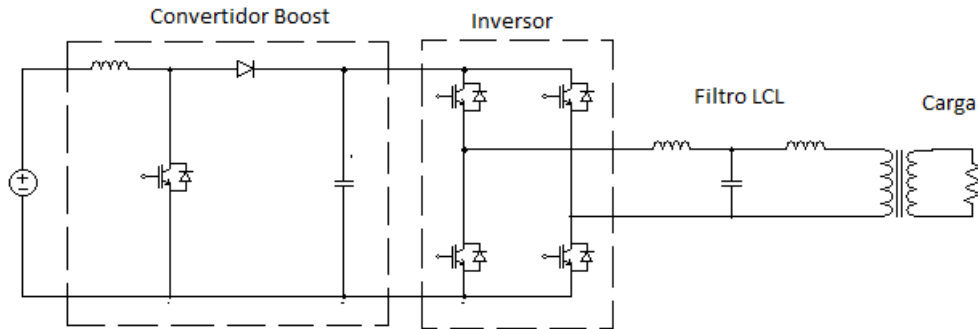


Figura 2 Sistema fotovoltaico en modo Isla.

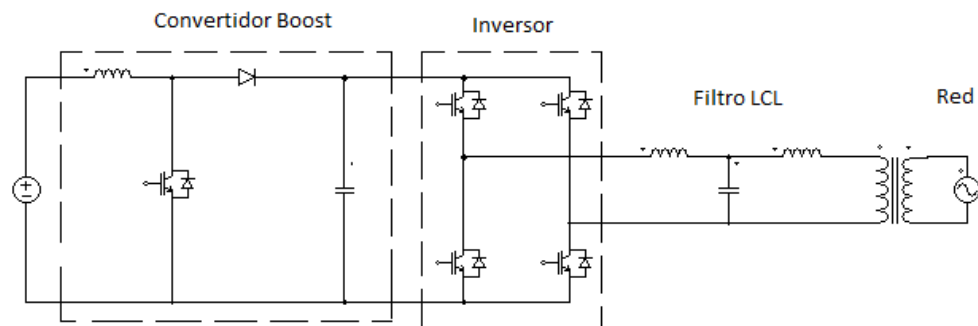


Figura 3 Sistema fotovoltaico en modo Red.

2. Métodos

Para realizar el análisis del comportamiento del convertidor Boost, debido a que este es un circuito no lineal, se debe obtener un circuito equivalente lineal alrededor de un punto de operación. Para ello, se utiliza el método del conmutador PWM [Garcerá, 1998]. Este método consiste en sustituir por un circuito equivalente la celda de conmutación, la cual, está presente en todos los convertidores. Dicha celda se conforma por el inductor, el interruptor de potencia, el capacitor y el diodo. Los nodos de los que forman parte estos elementos se designan como terminal activa (A), pasiva (P) y común (C), como se muestra en la

figura 4. Para el caso en que se trabaja en DCM, se tienen los circuitos equivalentes de gran señal y de pequeña señal de la figura 5 [Vorperian, 1990].

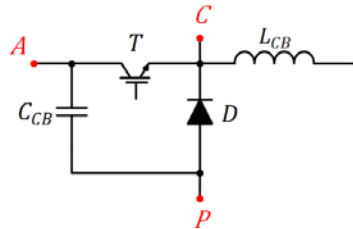


Figura 4 Terminales de la celda básica de conmutación.

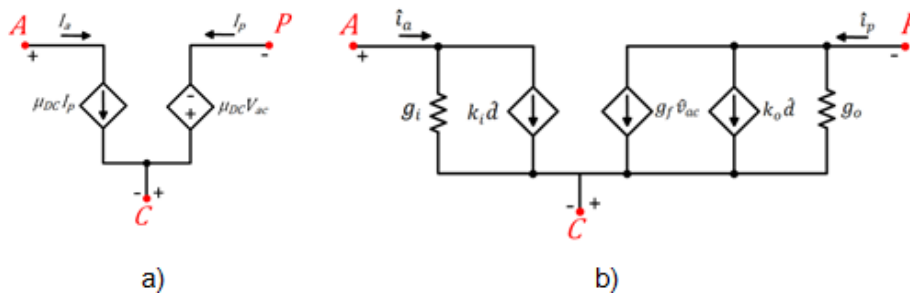


Figura 5 Circuitos equivalentes, celda básica conmutación modo conducción discontinua.

En los circuitos de la figura 5, I_a e I_p son las corrientes que circulan por las terminales A y P de la celda básica, respectivamente, y V_{ac} es la tensión entre las terminales A y C. Además, las variables \hat{i}_a , \hat{i}_p y \hat{v}_{ac} son: la corriente que circula por A, la corriente que circula por P y la tensión entre las terminales A y C en pequeña señal, respectivamente. El valor de μ_{DC} se expresa en ecuación 1, donde D es el ciclo de trabajo del interruptor, f_s es la frecuencia de conmutación, L es la inductancia del convertidor y V_{cp} es la tensión entre las terminales C y P.

$$\mu_{DC} = \frac{D^2 V_{cp}}{2L f_s I_a} = \frac{D^2 V_{ap}}{2L f_s I_p} \quad (1)$$

Del circuito se obtiene la relación de V_{CP} y de I_a que se expresan en ecuaciones 2 y 3.

$$V_{cp} = \mu_{DC} V_{ac} \quad (2)$$

$$I_a = \mu_{DC} I_p \quad (3)$$

La sustitución del circuito de la figura 5a en el convertidor Boost, da lugar a los circuitos equivalentes de gran señal para los modos Isla y Red que se muestran en la figura 6 y en la figura 7, respectivamente.

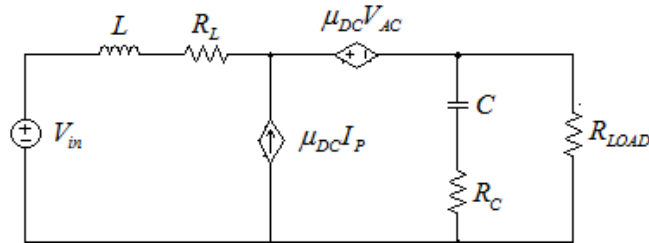


Figura 6 Modelo lineal a gran señal del convertidor Boost en modo Isla.

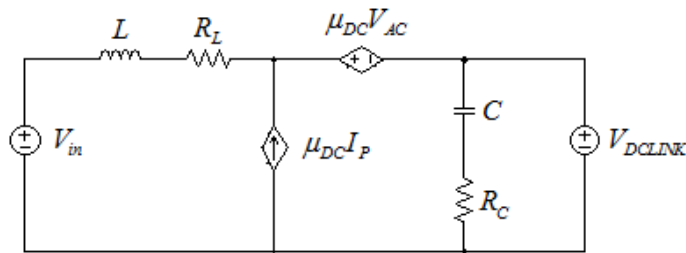


Figura 7 Modelo lineal a gran señal del convertidor Boost en modo Red.

Los valores de los parámetros k_i , k_o , g_f , g_i y g_o de la figura 5b son como se describe en [Vorperian, 1990]. La sustitución del circuito de la figura 5b en el convertidor Boost, da lugar a los circuitos equivalentes de pequeña señal para los modos Isla y Red que se muestran en la figura 8 y en la figura 9, respectivamente; y en donde R_L y R_C representan las resistencias internas del inductor y el capacitor.

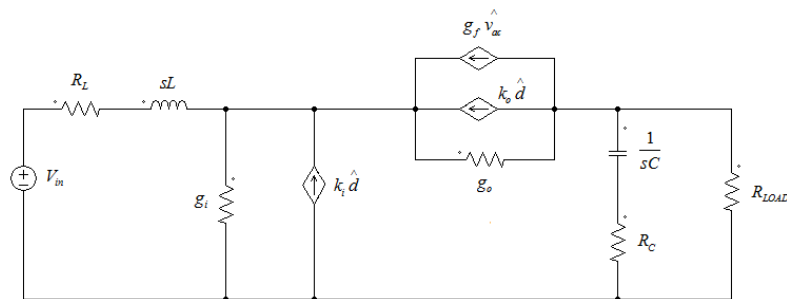


Figura 8 Modelo lineal a pequeña señal del convertidor Boost en Modo Isla.

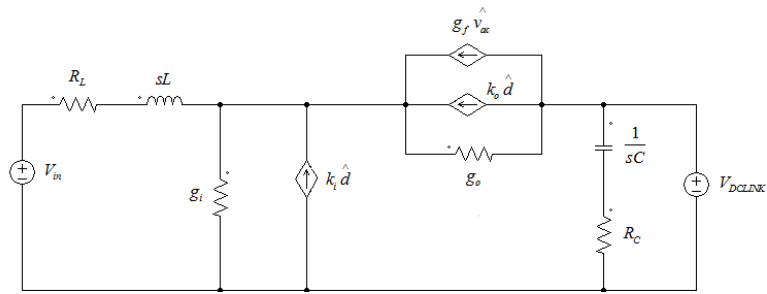


Figura 9 Modelo lineal a pequeña señal del convertidor Boost en Modo Red.

Para el circuito equivalente en modo Isla, la carga que alimenta el convertidor es modelada por una resistencia; y en el caso del circuito equivalente de modo Red, la salida del Boost tiene como carga una fuente de CD, debido a que, en este modo de operación, es el inversor quien controla la tensión en el Bus de CD. Para el sistema fotovoltaico se considera un arreglo de paneles solares que entrega una potencia máxima de 2 kW en ambos modos, con un voltaje de 96 V. Para el convertidor Boost se consideran los valores nominales mostrados en la tabla 1. La frecuencia de conmutación de los interruptores IGBT de potencia es de 10 kHz.

Tabla 1 Valores nominales del convertidor Boost.

ELEMENTO	VALOR
Potencia de salida, P_{out}	2 000 W
Tensión de salida, V_{out}	200 V
Corriente de salida, I_{out}	10 A
Resistencia de carga, R_{LOAD}	20 Ω
Resistencia del inductor, R_L	75 m Ω
Inductor, L	50 μ H
Resistencia del capacitor, R_C	50 m Ω
Capacitor, C	1.1 mF

Los controladores de todos los lazos se sintonizan siguiendo el criterio de estabilidad de Bode. Se consideran como valores mínimos un Margen de Fase (MF) de 50° y un Margen de Ganancia (MG) de 7 dB, para asegurar la estabilidad del sistema [Garcerá, 1998]. A continuación, se presenta el análisis del convertidor para ambos modos de operación y el diseño de los controladores propuestos.

En el modo Isla, el convertidor Boost es el encargado de regular la tensión del Bus de CD, para este propósito se emplea el esquema ACC, como se muestra en la

figura 10, donde R_{iB} es la ganancia del sensor de corriente, β_B es la ganancia del sensor de voltaje, F_{mB} es la ganancia del modulador PWM que está relacionada con el valor pico-pico de la señal comparadora diente de sierra y $DG(s)$ representa un retardo digital necesario para la implementación de los controladores mediante un Procesador Digital de Señales (DSP). En este esquema de control se realiza un lazo interno de corriente y uno externo de voltaje. Las funciones de transferencia se obtienen a partir del modelo lineal presentado en la figura 8. Para el sistema se consideran ecuaciones 4, 5 y 6.

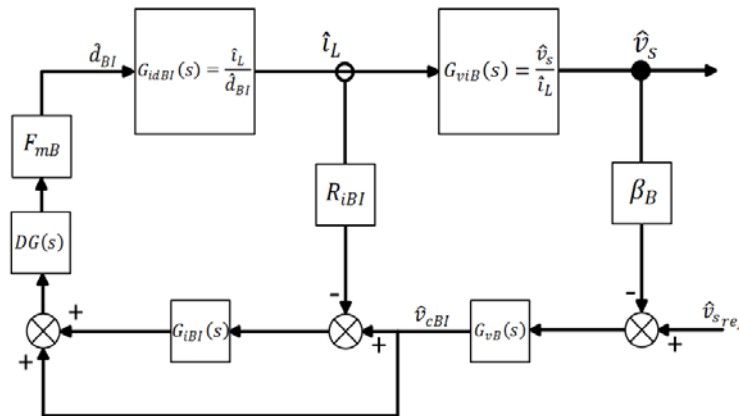


Figura 10 Diagrama de bloques en modo Isla.

$$F_{mB} = \frac{1}{V_{pp}} = \frac{1}{1} = 1 \quad (4)$$

$$R_{iB} = 0.042 \quad (5)$$

$$\beta_{iB} = 0.015 \quad (6)$$

La función de transferencia del retardo digital $DG(s)$, está dada por la aproximación de Padé [Carrillo, 2017] y se muestra en ecuación 7, donde T_{mu} es el período de muestreo.

$$DG(s) \approx \frac{1 - \left(\frac{sT_{mu}}{2}\right) + \left(\frac{(sT_{mu})^2}{12}\right)}{1 + \left(\frac{sT_{mu}}{2}\right) + \left(\frac{(sT_{mu})^2}{12}\right)} \quad (7)$$

La variable que controlar en el lazo interno es la corriente media que circula por el inductor, la cual corresponde a la corriente media de entrada al convertidor. La función de transferencia que se utiliza para este lazo de control es $G_{idB}(s)$ que relaciona la variación de la corriente en el inductor i_L respecto a la variación del ciclo de trabajo d del interruptor, la cual se muestra en ecuación 8, y donde Y_L es la admitancia del inductor con su resistencia interna en serie (ecuación 9) y Y_o es la admitancia de salida (ecuación 10).

$$G_{idB}(s) = \left. \frac{\hat{i}_L}{\hat{d}} \right|_{\hat{V}_{in}=0} = \frac{-Y_L [(Y_o + g_o)(k_i + k_o) - (k_o g_o)]}{(Y_o + g_o)(Y_L + g_i + g_o + g_f) - g_o(g_o + g_f)} \quad (8)$$

$$Y_L(s) = \frac{1}{Z_L(s)} = \frac{1}{R_L + sL} \quad (9)$$

$$Y_o(s) = \frac{1}{Z_o(s)} = \frac{sC(R_C + R_{LOAD}) + 1}{R_{LOAD}(sCR_C + 1)} \quad (10)$$

Para realizar el lazo de control, la magnitud del compensador de corriente G_{iBI} a la frecuencia de conmutación, debe cumplir las dos condiciones que se muestran en ecuaciones 11 y 12.

$$|G_{iBI}(s)|_1 < 20 \log_{10} \left(\frac{S_e L}{(V_{out} - V_{in}) R_{iB}} \right) \quad (11)$$

$$|G_{iBI}(s)|_2 < 20 \log_{10} \left(\frac{2S_e L}{V_{in} R_{iB}} \right) \quad (12)$$

Donde V_{in} es la tensión de entrada, V_{out} es la tensión media de salida del Boost y S_e es la rampa externa de la señal portadora del modulador. Estas condiciones se refieren a la inestabilidad en la amplitud del rizado de la señal de control. La condición $|G_{iBI}(s)|_1$ establece que la señal de control que se obtiene del compensador de corriente debe ser menor que la onda portadora, mientras que la condición $|G_{iBI}(s)|_2$ implica que el rizado de la señal de control debe ser menor que el doble de su valor medio, tal que no llegue a anularse su valor instantáneo [Carranza, 2012]. Para este caso ecuación 13.

$$S_e = 10000 \quad (13)$$

Los valores de las ecuaciones 11 y 12 dan como resultado ecuaciones 14 y 15.

$$|G_{iBI}(s)|_1 < -18.8263 \text{ dB} \quad (14)$$

$$|G_{iBI}(s)|_2 < -12.1104 \text{ dB} \quad (15)$$

Por lo que, el controlador de corriente sintonizado está por ecuación 16.

$$G_{iBI}(s) = 0.0026342 \frac{(s+19637)}{(s)} \quad (16)$$

Se verifica que el controlador sintonizado cumpla ecuaciones 14 y 15, mediante el análisis de su diagrama de Bode. Como se observa en la figura 11, el compensador sintonizado tiene una magnitud de -51.2 dB a la frecuencia de conmutación, cumpliendo así con las condiciones requeridas. La función de transferencia de lazo abierto está dada por ecuación 17.

$$T_i(s) = F_m R_i G_{iBI}(s) G_{idB}(s) DG(s) \quad (17)$$

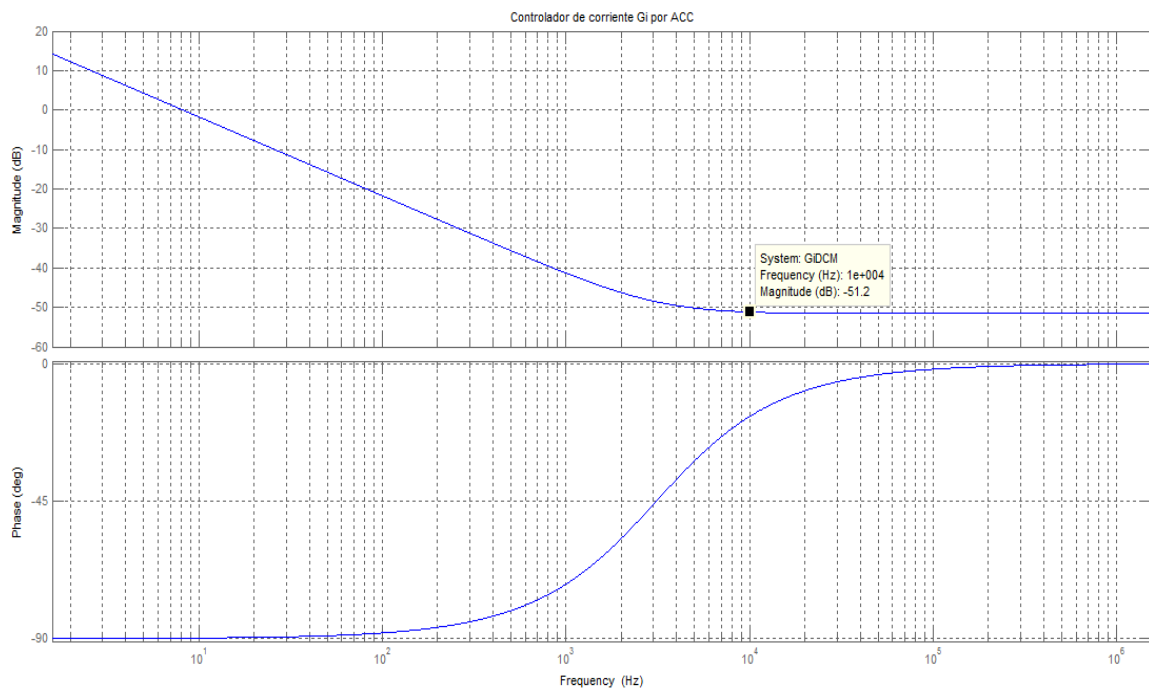


Figura 11 Diagrama de Bode de G_{idB} .

La función de transferencia de lazo cerrado que relaciona la corriente del inductor con la salida del compensador de tensión está expresada por ecuación 18.

$$G_{icB}(s) = \frac{\hat{i}_L}{\hat{v}_c} = \frac{(1 + G_{iBl}(s))(F_m DG(s)G_{idB}(s))}{(1 + T_i)} \quad (18)$$

Para el lazo externo de tensión, la función de transferencia de interés es la que relaciona la tensión de salida del convertidor con la señal de salida del compensador de tensión, mostrada en ecuación 19.

$$G_{vcB}(s) = \frac{\hat{v}_o}{\hat{v}_c} = \frac{[(g_f + g_o)(k_i + k_o) - (k_o)(Y_L + g_i + g_o + g_f)] F_m [(1 + G_{idB}) - (R_{iB} G_{iBl} D G G_{icB})]}{(Y_o + g_o)(Y_L + g_i + g_o + g_f) - g_o(g_o + g_f)} \quad (19)$$

El compensador sintonizado se muestra en ecuación 20.

$$G_{vBl}(s) = 0.00057981 \frac{(s + 72856)}{(s)} \quad (20)$$

Para la función de transferencia de lazo abierto no se toma en cuenta el retraso digital, ya que se encuentra incluido en lazo interno de corriente y su efecto en el lazo externo es despreciable. Por lo tanto, la función de lazo abierto se determina mediante ecuación 21.

$$T_v(s) = G_{vcB}(s) \beta_{iB} G_{vBl}(s) \quad (21)$$

El ancho de banda del lazo es de 14 Hz, el Margen de Ganancia es 97.6 dB, y el Margen de Fase es de 84.3°, como se muestra en la figura 12. La función de lazo cerrado está expresada por ecuación 22.

$$G_{vrB}(s) = \frac{G_{vB}(s)G_{vcB}(s)}{(1 + T_v(s))} \quad (22)$$

En el modo Red, la tensión en el Bus de CD es regulada por el inversor. Por lo tanto, el Boost se controla únicamente a través del lazo de corriente, el cual se encarga de mantener el valor requerido de corriente promedio del inductor. El diagrama de bloques del lazo de control se muestra en la figura 13. Del análisis del circuito equivalente de la figura 9, se obtiene la función de transferencia que

relaciona la corriente en el inductor con el ciclo de trabajo del interruptor y está dada por ecuación 23.

$$G_{idBG}(s) = \frac{\hat{i}_L}{\hat{d}} \bigg|_{\hat{V}_{in}=0} = \frac{-(g_i + g_o)(k_i + k_o)}{Z_L(1 + g_f) + (g_o + g_f)} \quad (23)$$

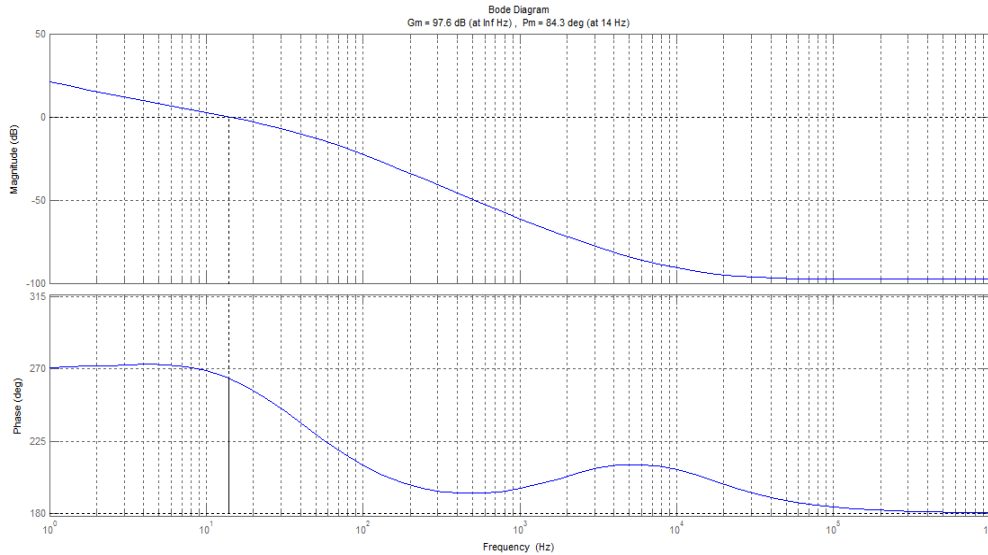


Figura 12 Ancho de banda, Margen de Ganancia y Margen de Fase para el modo Isla.

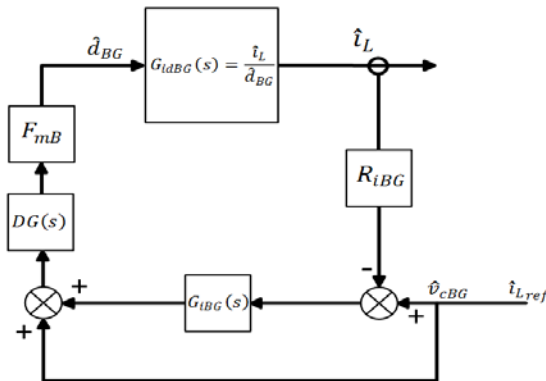


Figura 13 Diagrama de bloques en modo Red.

Para evitar cambios de compensadores cuando se trabaja en modo red y en modo isla, se utiliza el mismo compensador de corriente empleado en modo isla (ecuación 24), y se verifica que se cumplan las ecuaciones 11 y 12 para la operación en modo Red.

$$G_{iBG}(s) = 0.0026342 \frac{(s+19637)}{(s)} \quad (24)$$

La función de transferencia en lazo abierto está expresada por ecuación 25.

$$T_{iG}(s) = G_{idBG}(s)R_{iB}DG(s)F_mG_{iBG}(s) \quad (25)$$

La función de transferencia a lazo cerrado se obtiene por ecuación 26.

$$G_{icBG}(s) = \frac{G_{idBG}(s)DG(s)F_m(1+G_{iBG}(s))}{(1+T_{iG}(s))} \quad (26)$$

El ancho de banda del lazo es de 92 Hz, el Margen de Ganancia es 46.2 dB, y el Margen de Fase es de 69°, como se muestra en la figura 14.

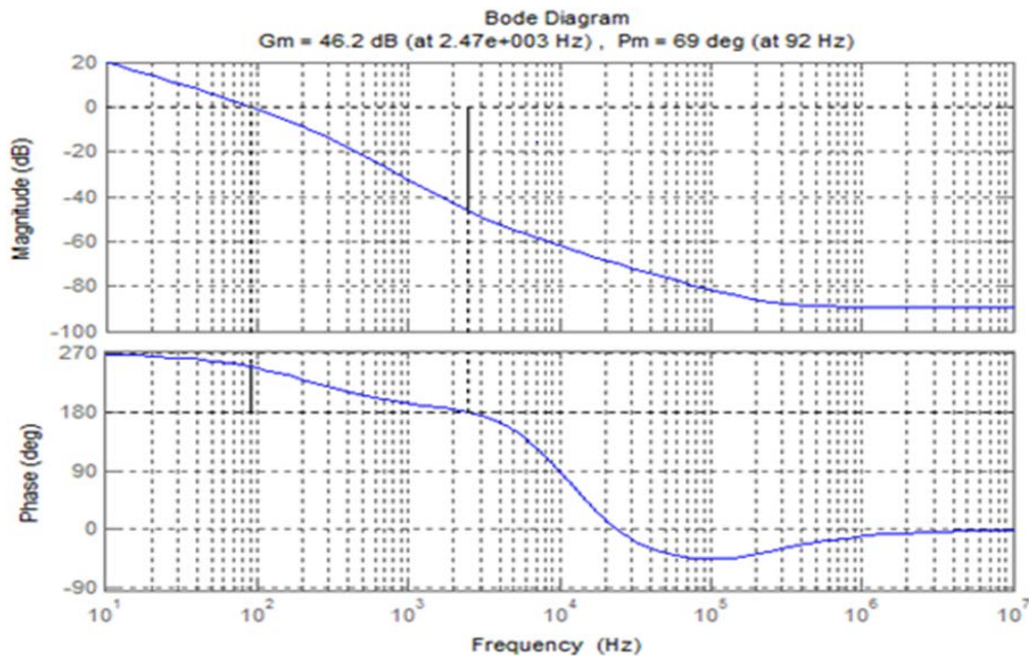


Figura 14 Ancho de banda, Margen de Ganancia y Margen de Fase para el modo Red.

3. Resultados

A continuación, se presentan los resultados en simulación del sistema bajo estudio y de los lazos de control propuestos. Para este propósito, se simula el sistema en PSIM, como se muestra en la figura 15, para el modo Isla.

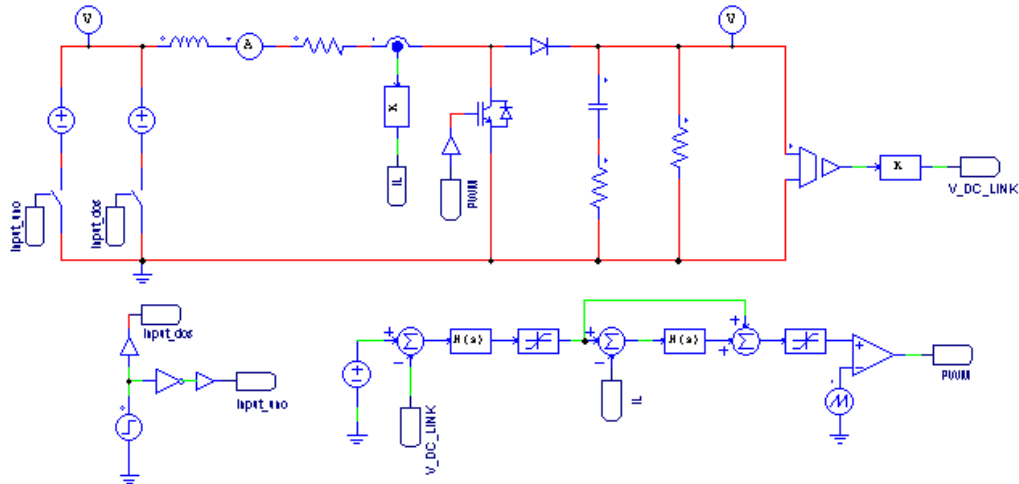


Figura 15 Convertidor Boost en modo Isla.

El sistema es simulado incluyendo perturbaciones en la tensión de entrada, lo que representa la variación en la salida del panel fotovoltaico. En la figura 16 se muestra que el Boost proporciona el valor de tensión media deseada aun cuando se tiene la perturbación en la tensión de entrada; se tiene un sobretiro máximo del 4.7% del valor de referencia, con un tiempo de asentamiento de 0.2 s.

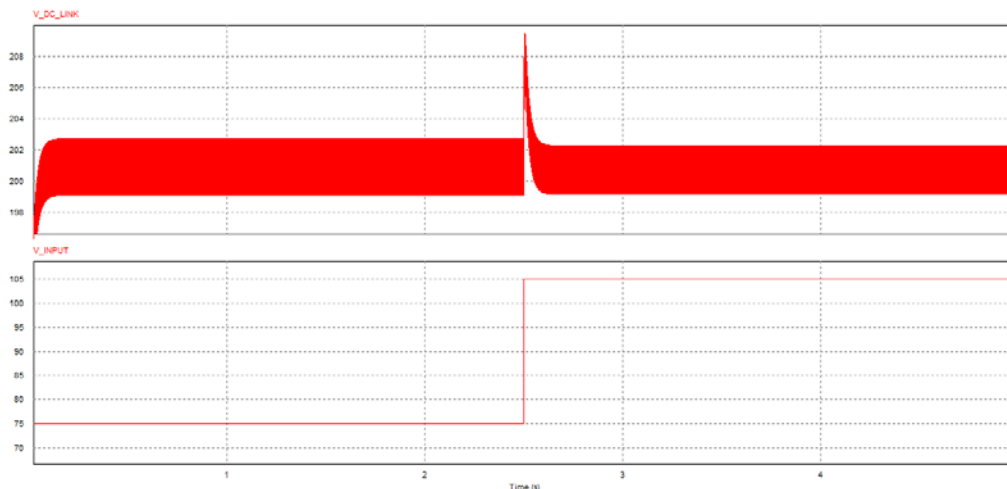


Figura 16 Tensión en el Bus de CD y tensión de entrada al convertidor en modo Isla.

Las perturbaciones consideradas están en el rango de 75 a 105 V. La figura 17 muestra que la corriente en el inductor permanece en DCM y que su valor promedio es el requerido de 22 A.

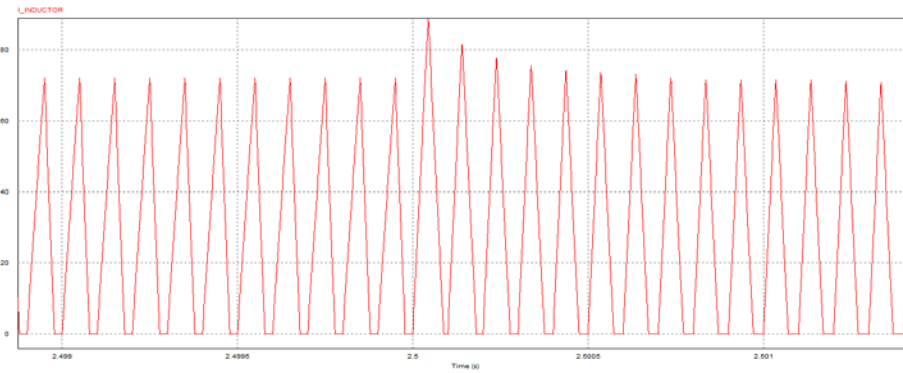


Figura 17 Respuesta transitoria de la corriente en el inductor en el modo Isla.

En la figura 18 se muestra el sistema para el modo Red. De igual manera, en la figura 19 se muestra la corriente en el inductor que permanece en DCM y con valor promedio de 22 A para el modo Red, cerca del momento de la perturbación, la cual es a 2.5 segundos para ambos modos de operación.

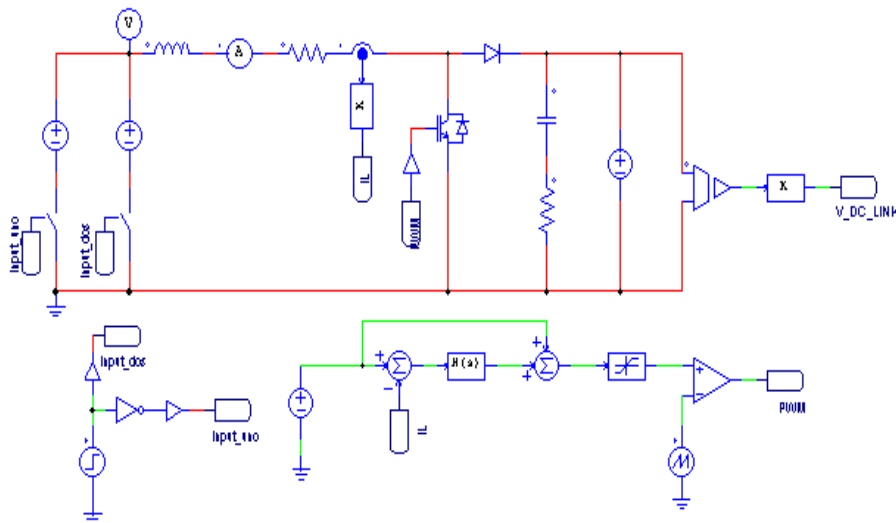


Figura 18 Sistema simulado en modo Red.

4. Discusión

Como se observa a partir de los resultados en simulación, los controladores de tensión y corriente cumplen su función de manera satisfactoria en ambos modos de operación. Para el caso de operación en modo Isla, el sobretiro en el nivel de tensión debido a la acción del controlador es tan sólo del 4.7%, lo cual representa una respuesta aceptable; además de que el rizo en estado estacionario de la

tensión es menor al 2%. En el caso del controlador de corriente, la corriente en el inductor tiene un sobretiro de aproximadamente 22% en su valor pico y un tiempo de asentamiento de 0.02 s, esto es debido a que la velocidad de respuesta del controlador de corriente es diez veces más rápida que la velocidad de respuesta del lazo de tensión. Para la operación en modo Red, la referencia de corriente se modela como una fuente de voltaje de CD con el valor de corriente promedio requerido; el sobretiro en el valor pico de la corriente del inductor es menor en comparación a la operación en modo Isla, siendo este del 19.5%.

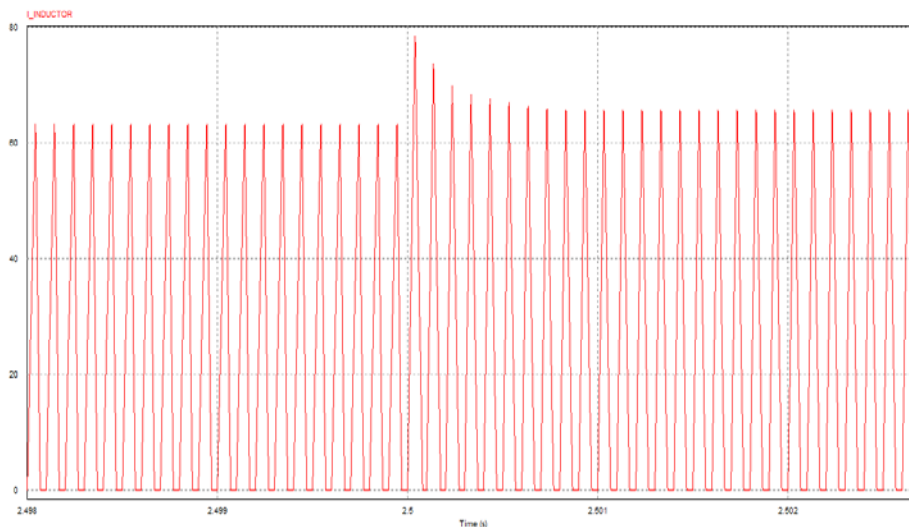


Figura 19 Respuesta transitoria de la corriente en el inductor en el modo Red.

5. Conclusiones

En este trabajo se presenta un esquema de control modo corriente media para un convertidor Boost operando en modo Isla y en modo Red. Los controladores diseñados de los lazos de tensión y corriente son del tipo PI. Los controladores son validados mediante la simulación del sistema en PSIM, donde se observa que se cumplen los requisitos de sobretiro máximo y de tiempo de asentamiento, durante la respuesta transitoria ante perturbaciones en la entrada del sistema.

Los esquemas de control serán implementados a través de un DSP y del prototipo del convertidor Boost que actualmente se está desarrollando, con la finalidad de validar experimentalmente los esquemas de control y el convertidor.

6. Bibliografía y Referencias.

- [1] Benavent, J. M., et al., Robust Model-Following Regulator for Average Current-Mode Control of Boost DC-DC Converters. 2005 IEEE International Symposium on Industrial Electronics. Dubrovnik, Croatia, 2005
- [2] Carranza, O., Estudio de técnicas de control de rectificadores Boost Trifásicos con filtro LCL para reducción de la distorsión armónica en corriente, aplicadas al procesado eficiente de energía en aerogeneradores síncronos de imanes permanente operando a velocidad variable. Universidad Politécnica de Valencia, España, 2012.
- [3] Carrillo, F., et al., Transition controllers between Island and Grid mode of a Photovoltaic system in a Smart Grid. VIII Congreso Internacional de Ingeniería Electromecánica y de Sistemas, Ciudad de México, México, 2016.
- [4] Carrillo F., Implementación de un sistema que emule la generación de energía fotovoltaica con operación en modo Isla y en modo Red. Instituto Politécnico Nacional, México, 2017.
- [5] Garcerá, G., Figueres, E., Abellán, A., Conversores Conmutados: Circuitos de potencia y control. Universidad Politécnica de Valencia, España. 1998.
- [6] He, D., Nelms, R. M., Average Current-Mode Control for a Boost converter using an 8-bit microcontroller. 2004 IEEE International Symposium on Industrial Electronics. Ajaccio, France, 2004
- [7] Labouret, A., Viloz, M., Solar Photovoltaic Energy. The Institution of Engineering and Technology, 2010.
- [8] Mohan, N. Power Electronics, Converters, Applications and Design, 3rd edition. Wiley, 2003.
- [9] Sharkawi, M. A., Electric Energy: An Introduction, 3rd edition. CRC Press, 2013.
- [10] Vorperian V., Simplified Analysis of PWM Converters Using Model of PWM Switch Part II: Discontinuous Conduction Mode. IEEE Transactions on Aerospace and Electronic Systems. Vol. 26, No. 3, Mayo 1990.

INTERCOMUNICADOR ENLAZADO A RED DE TELEFONÍA CELULAR

Joel Fernando Acevedo Ruiz

Benemérita Universidad Autónoma de Puebla

jo_fe_ar@hotmail.com

Aldrin Barreto Flores

Benemérita Universidad Autónoma de Puebla

aldrin.barreto@correo.buap.mx

Verónica Edith Bautista López

Benemérita Universidad Autónoma de Puebla

vbautista@cs.buap.mx

Salvador Eugenio Ayala Raggi

Benemérita Universidad Autónoma de Puebla

saraggi@ece.buap.mx

Resumen

La mayoría de timbres y sistemas de intercomunicación instalados en los hogares se vuelven inútiles cuando los habitantes se ausentan, poniendo en riesgo las pertenencias de las familias pues delincuentes consideran esto una oportunidad perfecta. Se desarrolla un dispositivo tipo timbre que no evidencie la ausencia de personas, pues se comunica mediante una llamada telefónica al celular del propietario manteniéndolo siempre en contacto y con conocimiento de quienes visiten su hogar. La instalación del dispositivo es sencilla y resalta su practicidad pues la comunicación siempre es al teléfono celular del dueño, evitando así utilizar un intercomunicador fijo al interior de la casa. El dispositivo consiste en un módulo de telefonía celular FONIA 3G de Adafruit, el cual está manejado por un microcontrolador PIC18F2550 mediante comunicación serial UART. Dispone en sus entradas de un sensor de movimiento PIR y un botón,

ambos manejados digitalmente, al igual que la pantalla tipo OLED de 128x64 pixeles controlada mediante I2C. El sistema está alimentado con una batería recargable de litio de 1300 mAh.

Palabras Claves: Celular, comunicación, microcontrolador, seguridad.

Abstract

The majority of doorbells and intercom systems installed in homes become useless when the inhabitants are absent, putting at risk the belongings of the families because criminals consider this a perfect opportunity. It develops a device type bell that does not evidence the absence of people, because it communicates by means of a telephone call to the cell phone of the owner keeping it always in contact and with knowledge of those who visit his home. The installation of the device is simple and highlights its practicality because the communication is always to the cell phone of the owner, thus avoiding the use of a fixed intercom in the house. The device consists of an Adafruit FONA 3G cellular phone module, which is managed by a PIC18F2550 microcontroller via UART serial communication. It has in its inputs a PIR motion sensor and a button, both digitally handled, as well as the 128x64 pixel OLED screen controlled by I2C. The system is powered by a rechargeable lithium battery of 1300mAh.

Keywords: Cellphone, communication, microcontroller, security.

1. Introducción

Inicialmente los timbres fueron utilizados de forma aislada como mecanismos de alerta o avisos. Consistían en alarmas sonoras producidas por el choque mecánico entre una campana y un martillo que se movía por efecto de un campo electromagnético. Con el desarrollo de la tecnología electrónica se crearon dispositivos que implementaron los timbres clásicos en tamaños reducidos logrando instalaciones prácticas y más económicas. A la vez, se añadió la función de comunicación mediante voz al crear los intercomunicadores que, instalados en las fachadas de los edificios y viviendas, reemplazaron al timbre común.

El visitante llega al edificio en cuestión y toca el timbre. Si el habitante se encuentra dentro, puede comunicarse mediante el intercomunicador instalado. Sin embargo, en caso contrario, el intercomunicador queda totalmente inútil y evidencia la situación ante el visitante.

La inutilidad del intercomunicador cuando el propietario está ausente representa grandes problemas sobre todo para casas con un solo habitante, como las siguientes:

- Evidencia que no hay nadie en la vivienda aumentando posibilidades de robo.
- Pérdida de paquetes, documentos y cartas que no pudieron ser entregados. (El usuario no pudo darle instrucciones a quien intentó entregarlos).
- Problemas con servicios a domicilio cuando éstos arriban en ausencia del habitante.
- Familiares y visitas con complicaciones cuando no reciben respuesta en el intercomunicador.
- Desatención de recibos, citaciones, multas y otros.

Ante tales problemáticas, se pensó en una solución que fuera capaz de comunicar al visitante con el propietario de la construcción en todo momento, independientemente del lugar en el que éste último se encontrara.

Para ello se tomó en cuenta el uso de las redes de telefonía celular ya que esto ahorraría el uso de un dispositivo extra y por el contrario aprovecharía el teléfono celular que la mayoría de gente ya posee [Jiménez, 2007].

El habitante en cuestión podrá atender sin excepción a cualquier persona que necesite dejar un aviso, una citación, paquete, carta, servicio, etc. Podrá dar y recibir indicaciones a quienes usen el intercomunicador. Dando un aporte de solución práctica a la necesidad de constante comunicación entre las personas de una calle, colonia, población [De la Mora, 2004].

Dispositivos relacionados:

- Ring DoorBell PRO – Timbre inteligente: es un dispositivo de la compañía Ring capaz de transmitir video en resolución 1080p y audio a un

Smartphone en cualquier lugar, utilizando el acceso a internet mediante red WiFi de 2.4 Ghz y 5 Ghz. El dispositivo incluye reconocimiento de imagen desde su aplicación para Smartphone, lo que permite definir áreas de vigilancia en el día o en la noche gracias a su sistema de visión nocturna. Su precio es de 249 USD [DoorBell, 2017].

- SmartBell – Portero automático: incluye una videocámara y una pantalla táctil para la interacción con el visitante. Se pueden grabar videos para ser reproducidos después, así como dejar mensajes a los visitantes por medio de su pantalla. También puede programar acciones para controlar puertas automáticas, cuenta con un sensor de movimiento y realiza transmisión de video al Smartphone del habitante mediante acceso a internet por WiFi. Permite el uso de múltiples usuarios para la comunicación con el Smartphone que el visitante elija. Se encuentra en etapa de financiamiento y su precio final se ha fijado en 279 USD [SmartBell, 2017].
- SkyBell: Es un intercomunicador con videocámara, micrófono, altavoz, sensor de movimiento y un led infrarrojo integrados. Puede realizar un enlace de audio y video a cualquier Smartphone por medio de internet desde una red WiFi. Su costo es de 199 USD [SkyBell WiFi, 2017].
- DoorBot: Es un timbre que permite ver quien está de visita y hablar con él desde un Smartphone, además de poder controlar la cerradura electrónica de la puerta de la casa. Está diseñado con una carcasa de aluminio y utiliza 4 baterías AA. Funciona por medio de una red WiFi con acceso a internet. Su precio es de 169 USD [Ring Products, 2017].

2. Métodos

Planeación de dispositivo

El dispositivo encargado de dar solución a este problema será un dispositivo inteligente. Se trata de un intercomunicador que funcionará en esencia de forma típica, reaccionará ante la presión de un botón o “timbre” con el que trabajará de forma complementaria junto a un sensor de presencia, cuando lo determine

pertinente realizará una llamada por medio de telefonía celular al número del propietario.

La función de lo anteriormente mencionado es evitar realizar llamadas innecesarias o bien continuar realizando una llamada en caso de que la persona que activó el intercomunicador se haya retirado. Se usaron las recomendaciones para establecer el diagrama de flujo que aparecen en [Joyanes, 1990] y el enfoque algorítmico presente en [García, 2005]. El proceso de funcionamiento puede observarse en el diagrama de flujo que aparece en la figura 1.

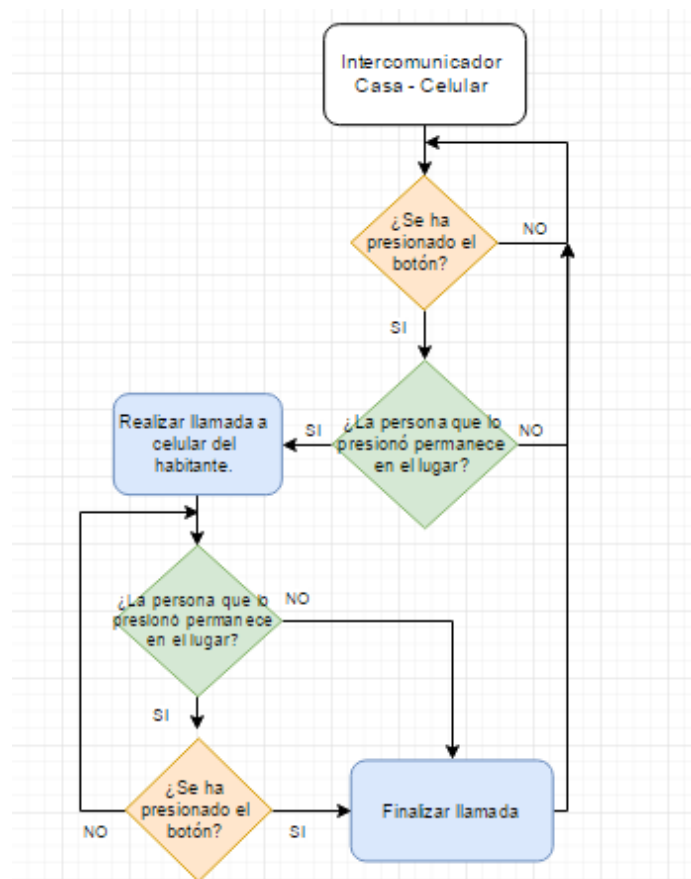


Figura 1 Diagrama de flujo del intercomunicador.

Se consideró adecuado que el intercomunicador celular cuente con un medio para indicar el estado del funcionamiento para una correcta interacción con el usuario, de ésta forma poder indicar cuando la llamada está siendo realizada, ha sido establecida y/o cuando ha sido finalizada.

El dispositivo, para cumplir su función, se programará en lenguaje C por sus características de alto nivel y sintaxis [Stroustrup, 1999], será instalado en el exterior de una vivienda, por lo que los sensores y componentes deberán funcionar adecuadamente ante las características ambientales que esto significa, por ejemplo, la humedad, la luz solar, el ruido auditivo, etc.

Selección de hardware:

Por los detalles mencionados en el apartado anterior, se eligieron los siguientes componentes y módulos para conformar el dispositivo objeto de este reporte.

Microcontrolador

El microcontrolador PIC18F2550 fue elegido para gobernar el dispositivo. Las principales características de un microcontrolador como se explica en [Angulo, 2003], memoria RAM, FLASH, ROM, sus múltiples puertos de entrada y salida, y en este caso puerto USB, comunicación UART, y altas frecuencias de trabajo, hicieron al PIC18F2550 el idóneo para esta aplicación con intención de mejoras continuas.

Comunicación con Red de Telefonía Celular

Para establecer las llamadas celulares se trabajará con el módulo FONA 3G de Adafruit [Adafruit Industries, 2017], el cual se aprecia en la figura 2 y es un módem celular compatible con señales GSM y 3G, capaz de hacer y recibir llamadas y SMS. Se controla mediante comandos AT a través de un puerto serial UART. El módulo integra un sistema completo para usar y cargar una batería LiPo de 3.7 V, haciéndolo ideal para este proyecto en el que la portabilidad es importante [Adafruit, 2017].



Figura 2 Módulo de módem celular Adafruit FONA 3G.

Sensor de Presencia

Para sensor la presencia y movimiento de las personas se seleccionó un módulo comercial estándar de Sensor PIR, el cual nos proporciona una salida digital con base en la detección realizada y puede controlarse fácilmente en su sensibilidad y tiempo de muestreo mediante dos POT en el PCB del módulo.

Indicadores de Estado-Componentes de Salida

Como componentes de salida o medios de interacción con el usuario del intercomunicador se escogieron una pantalla y un LED con las siguientes características.

Display gráfico OLED 0.96", figura 3.

- Pantalla OLED controlada por medio del protocolo I2C.
- Admite alimentación en DC de 3.3 a 5 V.
- Consumo eléctrico de aproximadamente 18 mA.
- Ángulo de visión de 160°.

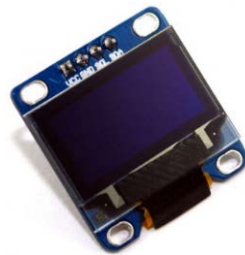


Figura 3 Display OLED 0.96" 128x64 pixeles.

LED ultrabrillante color Blanco.

- Diámetro de 5 mm.
- Alimentación de 3.5 V.
- Consumo eléctrico de 20 mA.
- Ángulo de iluminación de 30°.

Esquemáticos de Conexión

Teniendo seleccionados los elementos de hardware a utilizar, queda el esquema de conexión por bloques definido como en la figura 4. Como entradas

del sistema tendríamos al sensor de movimiento PIR y el botón del timbre. Como elemento de interacción entrada-salida la tarjeta Adafruit FONA 3G, y como piezas de salida la pantalla OLED y el LED blanco.

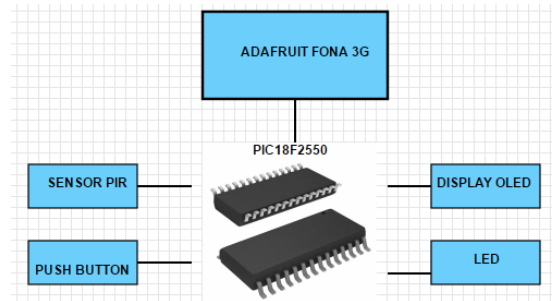


Figura 4 Diagrama de bloques del sistema.

Se utiliza los pines del puerto UART del microcontrolador para la comunicación serial con el módulo Adafruit Fona 3G, la cual fue la forma de comunicación y control elegida con bases en las ventajas que se encuentran en [Matpic, 2017], [Torres, 1999] y [Usategui, 1997].

El Display OLED está conectado a los pines RB0 y RB1 del microcontrolador por ser estos los correspondientes al protocolo I2C. El botón del timbre está conectado al pin RB2, en esa misma conexión se encuentra una resistencia de 10k en configuración pull-up. El LED indicador funcionará por medio del pin RB3 mientras que la señal del sensor de movimiento PIR es recibida por el pin RB5. Los puertos y pines utilizados se aprecian en el esquemático de la figura 5.

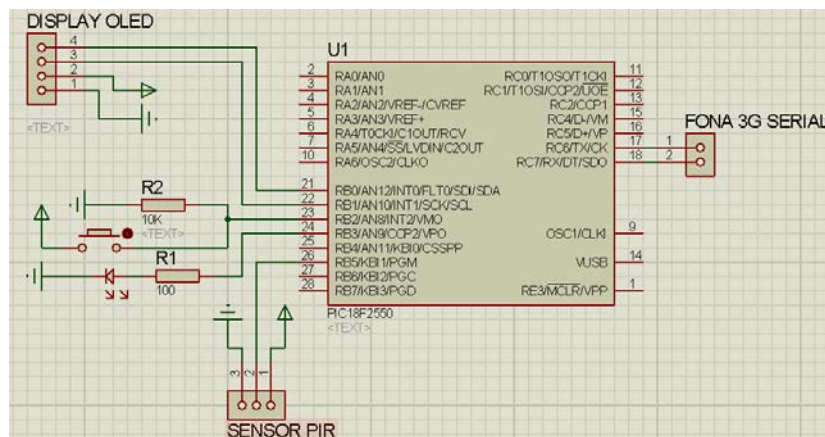


Figura 5 Esquemático de conexión.

Elaboración de Firmware

Para la programación correspondiente del microcontrolador se utilizó el IDE oficial del fabricante: Microchip MPLAB IDE X, utilizando el compilador de lenguaje C, aprovechando su conocida estructura por funciones y tomando como base ejemplos de programación hallados en [Deitel, 2003], [Eckell, 1999] y [Brassard, 2007]. Se programó el microcontrolador utilizando sus pines en modo digital para el uso con el display, sensor, botón y módulo FONA con lógica TTL. Se hizo uso de las interrupciones por flancos de subida en el puerto B para detectar rápidamente cualquier actividad del sensor PIR y del botón [Brooch, 1994]. El firmware programado en el PIC obedece el diagrama de flujo que aparece en la figura 6.

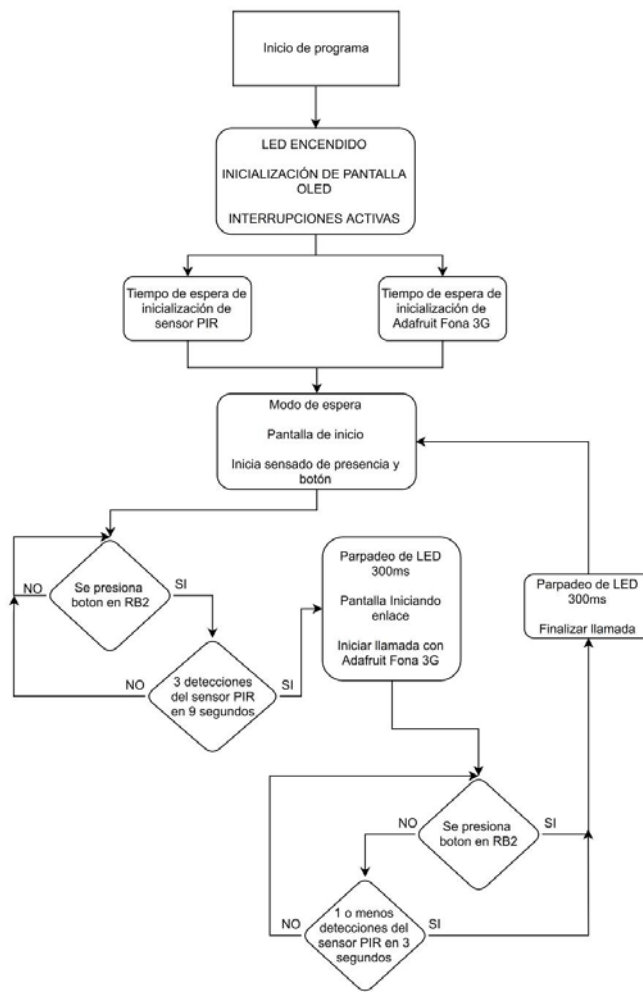


Figura 6 Diagrama de flujo de firmware del intercomunicador.

Construcción de Prototipo

- Circuito electrónico. El circuito electrónico correspondiente fue montado en una placa perforada para prototipos, la cual tiene una dimensión de 7.5 x 4.5 cm. Se colocó un zócalo de 28 pines para el microcontrolador de forma que éste pudiera ser fácilmente colocado o extraído de su sitio. Se hicieron las conexiones necesarias y se pusieron también pines o header para poder conectar fácilmente los cables destinados al display OLED, sensor, botón, LED y modulo FONA 3G. Se buscó integrar los elementos mencionados en forma compacta y funcional como se sugiere y explica en [Bueno-Soto, 2005].
- Carcasa e implementación de componentes adicionales. Para el completo funcionamiento del dispositivo, se requiere por parte del módulo Adafruit FONA 3G el uso de un altavoz de 8 ohm y de una batería LiPo recargable de 3.7 V. La bocina seleccionada es una bocina de 2" de diámetro, mientras que la batería utilizada es de 1300 mAh.

Se usó una carcasa de plástico para proyectos para resguardar el sistema electrónico del intercomunicador. Esta misma fue perforada para crear los espacios adecuados para el sensor, la bocina, pantalla, LED y un interruptor de encendido. El sistema del intercomunicador puede recargar su batería mediante el circuito de carga integrado en el módulo Adafruit FONA 3G, usando un puerto MicroUSB. En la figura 7 y 8 se observa la colocación de los componentes dentro de la carcasa plástica mientras que en la figura 9 se aprecia el dispositivo terminado.



Figura 7 Detalle de colocación de componentes dentro de la carcasa.



Figura 8 Detalle de colocación de componentes dentro de la carcasa.



Figura 9 Dispositivo intercomunicador terminado.

Funcionamiento y primeras pruebas

Cuando el sistema enciende, después del proceso de inicialización permanece en una pantalla de espera como se ve en la figura 10.



Figura 10 Pantalla de inicio en intercomunicador.

Tras presionar el botón, el dispositivo comprueba la presencia de un usuario por medio del sensor PIR, en caso de comprobarla, muestra una pantalla que valida el uso del intercomunicador, figura 11.



Figura 11 Pantalla que valida el uso del timbre y antecede a la llamada.

En seguida por medio del módulo FONA 3G, el dispositivo comienza a enlazar una llamada telefónica a un número previamente programado, lo cual se indica también en la pantalla. Entre cada cambio de pantalla el LED parpadea al permanecer apagado 300 ms como señal de cambio entre cada etapa, figura 12.



Figura 12 Pantalla de enlace que acompaña al establecimiento de la llamada telefónica.

Una vez establecida la llamada telefónica, el usuario puede escuchar a través del altavoz y hablar por medio del micrófono del dispositivo. En la pantalla entonces se muestra un mensaje que sugiere se presione nuevamente el botón para finalizar la llamada como se ve en la figura 13.



Figura 13 Mensaje en pantalla para finalizar llamada.

En caso de que el botón no sea presionado, el dispositivo colgará la llamada automáticamente si detecta que el usuario se ha retirado del lugar.

Al finalizar la llamada, el dispositivo regresa a la pantalla de inicio y queda en espera de un nuevo usuario.

3. Resultados

El sistema se colocó en una vivienda como puede notarse en la figura 14 se evaluó su desempeño considerando la efectividad del sensor de movimiento, la duración de la carga de la batería, y la utilidad de la intercomunicación por telefonía celular. Fue probado durante 5 días entre las 9 y 18 horas. El

intercomunicador realizó un promedio de tres llamadas efectivas por día, en todas esas ocasiones la comunicación transcurrió y finalizó sin problemas, siendo de gran utilidad para cuando el habitante no se encontraba en su domicilio.



Figura 14 Intercomunicador al exterior de una vivienda durante el periodo de pruebas.

La batería duró aproximadamente 6 horas sin necesitar recargarse. Siempre se estuvo pendiente de este hecho para recargar la batería y que el dispositivo siguiera funcionando durante el periodo de prueba. En las pruebas realizadas, se pudo constatar que solo 3 usuarios utilizaron el botón para finalizar la llamada, el resto se retiró cuando terminó de hablar por el intercomunicador, siendo este último el que detectara la ausencia del usuario y terminara automáticamente la llamada. Se tomó nota de lo anteriormente mencionado y de la calidad de la llamada con base en la opinión del visitante, calificando la intercomunicación con un número del 1 al 5, donde 1 es una baja calidad de audio y 5 es una excelente llamada sin ningún inconveniente. Estos datos se encuentran reunidos en la figura 15.

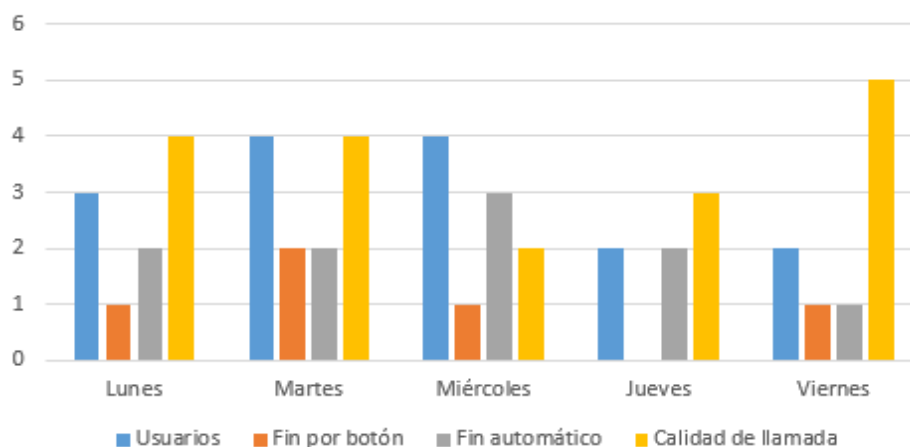


Figura 15 Resultados de pruebas del intercomunicador.

Por otra parte, en 7 ocasiones el intercomunicador no funcionó adecuadamente. En cuatro de ellas, el dispositivo no se había enlazado adecuadamente a la red celular, impidiendo la realización de una llamada cuando el timbre fue tocado. En las tres restantes, después de enlazar la llamada no se detectó la presencia del usuario y el dispositivo finalizó la llamada en curso automáticamente aun cuando el visitante permanecía en el lugar. Sin embargo, los siete usuarios hicieron un reintento al tocar nuevamente el timbre después de unos segundos, el intercomunicador pudo cumplir su función al enlazarse correctamente a la red celular. Los datos mencionados están reunidos en la tabla 1.

Tabla 1 Resultados obtenidos durante 5 días de prueba.

Resultado	Intercomunicaciones
Llamada exitosa	15
Falla en el sensor PIR. No se detecta presencia y finaliza la llamada.	3
Falla por baja señal de red celular. Llamada no enlazada	4

4. Discusión

Observando el funcionamiento del dispositivo se hace notar que incluso cuando hay personas dentro del inmueble se realizará la llamada al número celular predefinido, si bien el propietario puede designar un equipo celular para atender el timbre y decidir si dejarlo dentro de la casa o llevarlo afuera, dicha acción puede considerarse una deficiencia ya que debería cumplir también la función básica de un timbre avisando al interior de la residencia sobre una visita.

Debido a los datos obtenidos queda como principales aspectos a trabajar:

- Detección de presencia en el interior de la casa.
- Aviso al interior de la vivienda ante una visita antes de realizar la llamada.
- Autonomía | Duración de batería.
- Calidad de llamada | Uso de micrófono.
- Mejorar recepción de señal celular.
- Precisión del detector de presencia.

- Indicador de carga de batería.
- Aprovechamiento de energía solar.

Añadiendo, también será importante mejorar la calidad de la carcasa para hacerla más resistente y más agradable al usuario. Es posible disminuir el tamaño final del dispositivo al utilizar una bocina más pequeña y reducir el tamaño de la placa que albergue al circuito que acompaña al microcontrolador, ya que la pantalla y demás componentes ya son de pequeño tamaño. Esto haría al dispositivo más atractivo y práctico para su colocación en cualquier sitio. Se considera también modificar el sistema para utilizar un sensor ultrasónico en vez de un PIR, con la intención de mejorar la detección de presencia, así como modificar el firmware para que éste sea capaz de enviar mensajes de texto SMS al propietario con información sobre los eventos registrados, e incluso la opción de que el habitante envíe al intercomunicador un SMS con texto que se muestre en la pantalla para los usuarios.

Todas las anteriores son mejoras factibles para implementarse en el dispositivo, conservando su practicidad y funciones principales.

5. Conclusiones

El concepto tratado en este dispositivo resultó una idea útil y funcional, ayudando a quien lo posea a atender a cualquier persona que le visite incluso cuando no se encuentra en casa. Utilizar directamente las redes de telefonía celular es una gran ventaja para el intercomunicador al disponer de una amplia cobertura.

El hecho de que se disponga de un sensor de presencia, además de asistir al usuario para el finalizado de llamadas, también es una función que elimina la activación del dispositivo innecesaria ante bromas o equivocaciones de timbre, lo cual a menudo resulta molesto para el residente. En el concepto básico del intercomunicador, resultó altamente útil y funcional, cumpliendo su cometido sin mayor problema y dando solución a algunas problemáticas importantes.

6. Bibliografía y Referencias

- [1] ANGULO, José, MICROCONTROLADORES «PIC», Diseño práctico de aplicaciones. Primera parte: El PIC16F84. Lenguajes PBASIC y Ensamblador”, McGraw-Hill Interamericana de España, Tercera edición, España 2003.
- [2] Booch, Grady. Análisis y diseño orientado a objetos, con aplicaciones. 2da Ed. Addison Wesley–Diaz de Santos 1994.
- [3] De la Mora Medina, José; Pliego Mendoza, Nieves, Textos y contextos de la comunicación masiva, primer módulo. Edición interna del Colegio de Ciencias y Humanidades Plantel Sur. México, D. F., pp.175, 2004.
- [4] Brassard, G., Bratley, P., Fundamentos de Algoritmia. Madrid: Prentice-Hall, 1997.
- [5] Deitel, Harvey & Deitel, Paul, Cómo Programar en C++. Pearson Prentice Hall 4ta. Ed., 2003.
- [6] Desarrollo y construcción de prototipos electrónicos, Ángel Bueno Martín y Ana I. de Soto Gorroño, ISBN:84-267-1363-7. Ed. Marcombo.
- [7] DoorBell, 2017: <http://www.doorbellhome.org/reviews/doorbot/>, Consultada el 19 febrero 2017.
- [8] Eckell, Bruce, Thinking in C++. 2da. Ed. Prentice Hall Inc., 1999.
- [9] García Molina, J. J.; Montoya Dato, F. J.; Fernández Alemán, J. L.; Majado Rosales, M. J. (2005). Una introducción a la programación. Un enfoque algorítmico. Madrid: Thomson-Paraninfo, 2005.
- [10] Industries, A., Adafruit FONA 808 - Mini Cellular GSM + GPS Breakout ID: 2542 - \$49.95 : Adafruit Industries, Unique & fun DIY electronics and kits. Adafruit.com, Disponible en: <https://www.adafruit.com/product/2542> Consultada 19 febrero 2017.
- [11] Industries, Adafruit Industries, Unique & fun DIY electronics and kits. Adafruit.com. Disponible en: <https://www.adafruit.com/> Consultada 19 febrero 2017.
- [12] Jiménez, José Juan, Evolución e historia de la telefonía celular, Consultado el 13 de septiembre de 2007: <http://www.monografias.com/>.

- [13] Joyanes, L., *Problemas de Metodología de la Programación*. Madrid: McGraw-Hill, 1990.
- [14] Matpic.com, PIC-COMUNICACIÓN Serial PC–PIC: http://www.matpic.com/esp/microchip/com_serial_pc_pic.html, consultada 19 febrero 2017.
- [15] *Microcontroladores*; Vicente Torres, Servicio Publicaciones UPV, 1999.
- [16] *Microcontroladores PIC, La Solución en un Chip*; J. M. Angulo Usategui, E. Martín Cuenca, I. Angulo Martínez; Ed. Paraninfo, 1997.
- [17] Ring Products: <https://ring.com/videodoorbells>, consultada el 19 febrero 2017.
- [18] Smartbell, Wi-Fi doorbell for video chats to iOS and Android, 2017. Kickstarter. <https://www.kickstarter.com/projects/1256599792/smartbell-wi-fi-doorbell-for-video-chats-to-ios-an>, consultada el 19 febrero 2017.
- [19] SkyBell HD-SkyBell WiFi Doorbell, 2017, SkyBell WiFi Doorbell. <http://www.skybell.com/product/skybell-video-doorbell-hd/>, consultada el 19 febrero 2017.
- [20] Stroustrup, Bjarne, *El lenguaje de programación C++*. 3ra Ed. Addison-Wesley/Díaz de Santos, 1999.

DESARROLLO DE UN PROCESO DE AUTENTICACIÓN FACIAL EN UN SISTEMA ANDROID UTILIZANDO EL ALGORITMO LDA (ANÁLISIS DE DISCRIMINACIÓN LINEAL)

Francisco Emiliano Aguayo Serrano

Universidad Autónoma de Querétaro
microstudio.aguayo@gmail.com

Jesús Carlos Pedraza Ortega

Universidad Autónoma de Querétaro
caryoko@yahoo.com

Edgar Alejandro Rivas Araiza

Universidad Autónoma de Querétaro
erivas@uaq.mx

José Erik Rivas Araiza

Universidad Autónoma de Querétaro
jerivas@uaq.edu.mx

Resumen

En este trabajo de investigación se desarrolló un proceso de autenticación facial implementando primero los algoritmos PCA y LDA en una PC, evaluando sus respectivos desempeños en tiempo y nivel de autenticación con bases de datos públicas y posteriormente implementando el algoritmo LDA en una aplicación Android utilizando una base de datos propia donde las imágenes están bajo diferentes condiciones de iluminación, distancia, pose de la persona, fondo de la imagen, etc. Todo esto con el fin de seguir contribuyendo a los sistemas de reconocimiento y autenticación facial ya que esta área ha ido creciendo a lo largo de estas tres últimas décadas y se aplica en diversas áreas como la seguridad, la interacción entre hombre y máquina, video juegos, etc. Este proceso de

reconocimiento facial se divide a su vez en reconocimiento 1:n (reconocimiento facial) y reconocimiento 1:1 (autenticación facial).

Palabras Claves: Autenticación, cara, LDA, PCA, visión por computadora.

Abstract

In this work, the facial authentication process was developed by first implementing the PCA and LDA algorithms in a PC, evaluating their performance in time and level of authentication in public databases and later implementing the algorithm LDA in an Android application using an own database where the images are under different conditions of illumination, distance, pose of the person, background of the image, etc. Everything in order to continue contributing to facial recognition and authentication systems as this area has been growing throughout these three decades and is applied in various areas such as security, the interaction between man and machine, video games, etc. This facial recognition process is once divided into 1: n recognition (facial recognition) and 1: 1 recognition (facial authentication).

Keywords: Authentication, Computer Vision, Recognition, LDA, PCA.

1. Introducción

La biometría es una disciplina que estudia la identificación de una persona en base a sus características como huellas, la forma del rostro y su contorno, la voz, el iris de los ojos, cicatrices, etc. [Brumnik, 2011]. La biometría tiene muchas aplicaciones como la identificación de delincuentes, controles de acceso automático, entre muchas otras como se muestra en [Duró, 2001].

En los últimos años se ha incrementado el uso de dispositivos y tecnologías que le permiten a las personas llevar a cabo sus labores cotidianas como transacciones en la banca, acceso a un sistema mediante un usuario y contraseña e incluso con autenticación con huella o rostro, tal como se muestra en el trabajo presentado por [Hernández, 2010], donde se explica la importancia de contar con un buen sistema de autenticación facial, una metodología de trabajo y la presentación de los algoritmos más utilizados.

En la actualidad son muy pocos los sistemas que cuentan con autenticación facial, además se utilizan solamente en empresas donde se requiera un nivel de seguridad aceptable [Pentland, 2014], [Fuentes, 2011]. Con estas razones y con el objetivo de contar con más herramientas que permitan realizar autenticación facial de forma rápida y segura, se propone realizar una aplicación Android que permita realizar esta tarea y así pueda aplicarse como herramienta de autenticación móvil para validar transacciones, dar acceso a un lugar o ingresar a algún sistema.

El algoritmo PCA es muy parecido a LDA con la sutil diferencia que LDA hace una mejor reducción de la dimensionalidad de los datos y además una mejor separación de clases debido a que LDA coloca una etiqueta a los datos, lo que permite que se agrupen mucho mejor, además gracias a esta característica es considerado un algoritmo de aprendizaje supervisado, [Viola, 2001] adicionalmente explica la teoría matemática donde tomó como referencia una SVM (máquina de soporte vectorial) donde se reconoce que este algoritmo podría no resultar tan eficiente en comparación con la máquina de soporte vectorial, pero que en cuestión de costo computacional es mucho mejor [Pentland, 2014].

La primera implementación del algoritmo PCA (Análisis del componente principal) para autenticación facial fue propuesto por [Turk, 1991] donde se comprobó que este algoritmo es muy eficiente debido a que se reduce la dimensión de la imagen, una matriz formada por datos de varias imágenes a las que se les aplica una serie de operaciones matemáticas como el promedio, los valores propios, vectores propios, etc. Obteniendo así lo que se le denomina eigenfaces para clasificar las imágenes del rostro tal como se muestra en la figura 1, también se implementaron redes neuronales, al final se concluyó que gracias a la obtención de una cantidad menor de datos que representan en buena parte la matriz original, estos se pueden utilizar para autenticación, reduciendo así el costo computacional.

En [Delbracio, 2006] Se muestra una comparación entre LDA, PCA y ICA (análisis del componente independiente) se trabajó con una vasta colección de 1,176 imágenes, es decir, 49 personas con 24 imágenes cada una, se realizaron tres pruebas las cuales se basaron en seleccionar un conjunto reducido de imágenes que eran muy parecidas, también con un conjunto un poco más amplio y muy

parecidas y por ultimo un conjunto de imágenes pero con expresiones faciales distintas, este concluyo que el algoritmo LDA es mejor que PCA y ICA debido a que este algoritmo es el único que tiene un entrenamiento supervisado, también fue mejor a la hora de utilizar imágenes con distintas expresiones faciales.



Figura 1 Eigenfaces resultado del trabajo de [22].

La tesis presentada por [Mendoza, 2015] se enfoca únicamente en el algoritmo SURF modificado donde se destaca su implementación en dispositivos como teléfonos inteligentes y tabletas con sistema operativo Android y también se destaca la propuesta de una metodología que consta de seis etapas como se muestra en la figura 2 una de ellas es el procesamiento de las imágenes y su normalización, este algoritmo tiene la ventaja de usar un umbral de coincidencias, pero solo se comparan una a una cada par de imágenes lo que no resulta muy eficiente si la expresión facial de la persona va cambiando gradualmente, esto disminuye el número de coincidencias en la autenticación.

En el trabajo mostrado por [Hernández, 2010] se implementó un sistema de reconocimiento de rostros y se utilizó una metodología definida por seis etapas: Captura de la imagen, un pre procesamiento de imágenes, localización de la zona de interés, un escalamiento y ajuste, posteriormente la extracción de las

características faciales y por último la aplicación del algoritmo y la toma final de decisión.

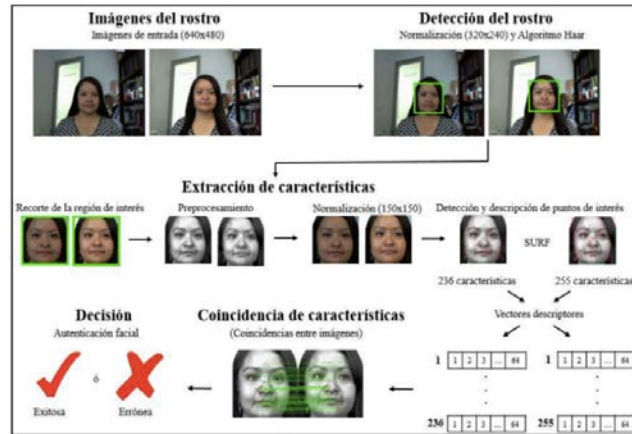


Figura 2 Metodología propuesta por [Mendoza, 2015].

2. Métodos

En este trabajo de investigación se implementaron los dos algoritmos principales LDA y PCA en la plataforma de Anaconda que permite ejecutar librerías open source y posteriormente la implementación de LDA en Android, donde la aplicación principal está escrita en java y manda llamar a las funciones escritas en C++ del algoritmo LDA el cual es un código nativo, la metodología para realizar el sistema de autenticación facial en Android es la que se muestra en la figura 3.

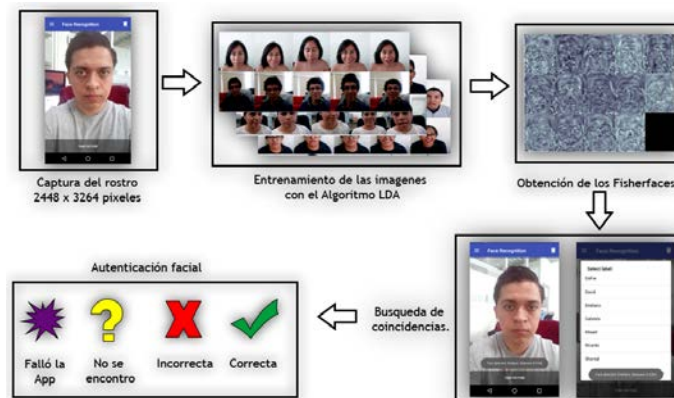


Figura 3 Metodología propuesta.

Donde se realizó la captura del rostro de todos los individuos, llevando a cabo el entrenamiento supervisado de LDA, después se obtienen las Fisherfaces y posteriormente se buscan las coincidencias con la matriz resultante dando así cuatro posibles resultados: Autenticación correcta, es decir, el usuario si se encontró; la autenticación incorrecta conocida también como falso positivo, es decir, se encontró coincidencia pero no con el usuario correcto; falso negativo donde no se encontró al usuario y si debió encontrarlo y por último donde se consideró la posible interrupción de la aplicación; el algoritmo LDA comprende los siguientes pasos:

- Construir la matriz X como en el método de eigenfaces, pero asignando a cada imagen una clase con la clase correspondiente a la matriz de clases c , ecuaciones 1 y 2.

$$X = \{X_1, X_2, \dots, X_c\} \quad (1)$$

$$X_i = \{x_1, x_2, \dots, x_n\} \quad (2)$$

- Proyectar la matriz X dentro de $(N - c)$ -dimensional sub-espacio a través de PCA con la matriz rotada. Donde N es el número de muestras en X y c es el número de clases o caras únicas, ecuación 3.

$$W_{PCA} = \operatorname{argmax}_W |W^T S_T W| \quad (3)$$

- Calcular la matriz de dispersión entre clase y la matriz de dispersión dentro de la clase, ecuaciones 4 y 5.

$$S_B = \sum_{i=1}^c N_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (4)$$

$$S_W = \sum_{i=1}^c \sum_{x \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \quad (5)$$

Donde μ es el promedio total de todas las clases y se calcula con la ecuación 6.

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (6)$$

Donde μ_i es el promedio de cada clase $i \in \{1, \dots, c\}$ y se calcula a partir de

la ecuación 7.

$$\mu_i = \frac{1}{|X_i|} \sum_{x \in X_i} x_j \quad (7)$$

- Aplicar el discriminante lineal de Fisher Y maximizar la relación entre el determinante de la matriz de dispersión entre clases y la matriz de dispersión dentro de la clase. La solución está dada por el conjunto de los vectores propios generalizados W_{FLD} de S_B y S_W , esto resulta $c - 1$ valores propios diferentes a cero, ecuación 8.

$$W_{FLD} = \operatorname{argmax}_W \frac{|W^T W^T_{PCA} S_B W_{PCA} W|}{W^T W^T_{PCA} S_W W_{PCA} W} \quad (8)$$

- Obtener las Fisherfaces, ecuación 9.

$$W = W^T_{FLD} W^T_{PCA} \quad (9)$$

3. Resultados

El análisis de discriminación lineal es la técnica más común usada para reducir la dimensionalidad en el pre-procesamiento para la clasificación de patrones y aplicaciones como máquinas de aprendizaje, así como en la autenticación facial. El objetivo es proyectar un conjunto de datos dentro de un espacio de dimensión con una buena separabilidad de clases en orden de evitar el sobre ajuste y reducir el costo computacional [Raschka, 2014]. El enfoque general de LDA es muy parecido a PCA, pero LDA además de encontrar los ejes de los componentes que maximizan la varianza de los datos (lo que hace PCA), además se interesa en los ejes que maximizan la separación entre múltiples clases [Welling, 2005]. PCA maximiza la varianza entre componentes mientras que LDA además encuentra un sub-espacio de características que optimizan la separación de los datos a través de clases tal y como se muestra en la figura 4.

En este trabajo primero se realizaron pruebas de con una base de datos publica conocida como Yale Facedatabase A comúnmente conocida como Yalefaces, se decidió utilizar esta base de datos ya que se puede apreciar con más claridad la diferencia y eficiencia de los dos algoritmos principales de LDA y PCA. Estas

imágenes están en escala de grises, son 40 personas con 10 expresiones faciales distintas y sus dimensiones son de 92x112 píxeles como se muestra en la figura 5.

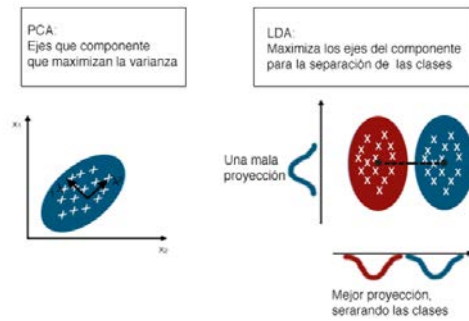


Figura 4 Comparación de PCA y LDA.

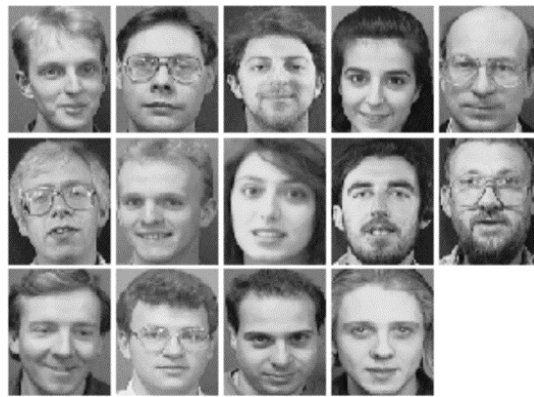


Figura 5 Rostros utilizados para las pruebas.

Para hacer las pruebas preliminares se utilizó el IDE Anaconda que es una plataforma de software libre que permite ejecutar librerías de data science en el lenguaje de programación Python con distintas librerías. En este caso se obtuvieron las 14 Eigenfaces que se muestran en la figura 6.

Debido a que el algoritmo PCA maximiza la variación entre las imágenes de entrenamiento con la imagen actual y a pesar de que las proyecciones de PCA a la hora de hacer una reconstrucción de los datos, este algoritmo no cumple con la discriminación de los datos, es decir, al tener un conjunto grande de imagen para comparar se empieza a perder información útil para poder hacer la discriminación necesaria y así lograr una autenticación más efectiva como se muestra en la figura 7, no alcanza a realizar todas las coincidencias de forma correcta. Como se explicó con anterioridad LDA maximiza la separación de los datos mediante

clases.

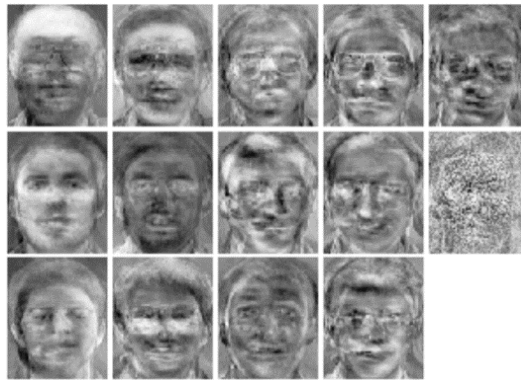


Figura 6 Eigenfaces obtenidas.



Figura 7 Coincidencias a partir del algoritmo PCA.

El método de Fisherfaces funciona a partir de una matriz de transformación específica de clase, por lo que no capturan la iluminación como el método de Eigenfaces. En cambio, el análisis discriminante lineal encuentra las características faciales para discriminar los datos entre las personas. El rendimiento de los Fisherfaces depende en gran medida de los datos de entrada, las Fisherfaces obtenidas se muestran en la figura 8.

Si los Fisherfaces se obtuvieron a partir de imágenes en condiciones bien iluminadas solamente y se intenta reconocer caras en escenas mal iluminadas, entonces el método encontrará componentes incorrectos, esto es algo lógico, ya que el método no tuvo oportunidad de aprender de la iluminación, esto no lo hace el método de Eigenfaces como ya se comprobó. De aquí la necesidad de utilizar este conjunto de imágenes de Yalefaces, ya que las 10 muestras de cada rostro tiene diferentes condiciones, es este caso la figura 9 muestra que efectivamente debido el método de Fisherfaces alcanza a encontrar todas las coincidencias dentro de la base de datos pública utilizada.

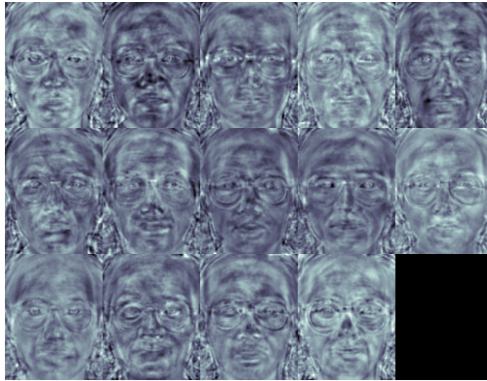


Figura 8 Fisherfaces obtenidas.

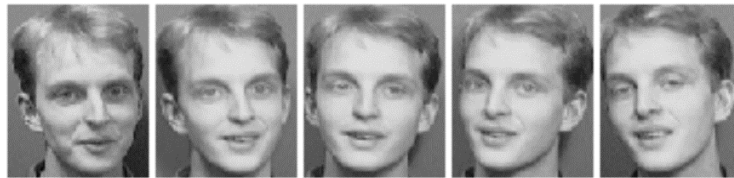


Figura 9 Coincidencias a partir del algoritmo LDA.

Como parte última de este trabajo de investigación se realizaron las pruebas de autenticación con una base de datos propia que se muestra en la tabla 1, en un teléfono inteligente con las características de hardware y software que se muestran en la tabla 2.

Tabla 1 Descripción de la base de datos propia.

Descripción
10 personas, 5 imágenes distintas por persona, Imágenes en formato JPG, resolución de 4160 x 3120 píxeles, diversos fondos, iluminación, expresiones faciales, diferentes distancias entre el rostro y la cámara, distintos colores de piel y uso de lentes en algunas personas.

Tabla 2 Descripción del hardware y software del teléfono inteligente.

Descripción
Procesador ARM Cortex-A53, 1300 MHz con 4 núcleos. Procesador gráfico ARM Mali-T720 MP2, 600 MHz con 2 núcleos. Memoria RAM de 2 GB, 640 MHz. Tamaño de 12.7 cm, 720 x 1280 píxeles. Cámara de 4160 x 3120 píxeles, sensor de proximidad, acelerómetro, lector de huella dactilar.

El proceso de adquisición de las imágenes fue el siguiente: usuario se toma una selfie, es decir, un auto retrato, la primera vez que se hace esto se le asigna una

nueva etiqueta para dicho usuario (su nombre), y cada vez que el mismo usuario se tome la selfie debe seleccionar la etiqueta creada con anterioridad, las cinco expresiones faciales fueron serio, alegre, cerrando los ojos, sorprendido, y enojado, tal como se muestra en la figura 10.

Cada vez que se capturan usuarios y expresiones faciales, la aplicación procede a realizar el entrenamiento como se explicó con anterioridad. Al ser una imagen tomada desde el celular, el movimiento o la distancia no pueden ser acotados de forma estricta por lo que la aplicación muestra un mensaje de autenticación y la distancia del rostro con respecto a la cámara.



Figura 10 Base de datos propia.

Posteriormente de realizar el entrenamiento, se realizaron 10 pruebas a cada uno de los usuarios, es decir 100 pruebas en total, en este punto no se implementaron la extracción de las características faciales y el pre-procesamiento de las imágenes. En la tabla 3 se muestran los primeros resultados obtenidos, donde la columna prueba refleja las 10 pruebas de cada uno de los usuarios. La columna usuario muestra el nombre del usuario en cuestión, la tercera columna NP el número de pruebas de cada uno de ellos, la columna AC muestra el número de

autenticaciones correctas, la columna AI el número de autenticaciones incorrectas, es decir, que la aplicación no encontró al usuario, la columna FR muestra el número de falsos reconocimientos, es decir, el caso en el que al hacer el reconocimiento mostró el nombre de otro usuario y por último la columna EA que muestra el número de errores de la aplicación, es decir, donde la aplicación se cerró de forma abrupta sin poder realizar el reconocimiento.

Tabla 3 Resultados de las pruebas sin pre-procesamiento.

Prueba	Usuario	NP	AC	AI	FR	EA
1	Paulina	10	8	1	1	0
2	Dafne	10	8	2	0	0
3	Ricardo	10	9	0	1	0
4	Shantal	10	6	3	1	0
5	David	10	5	0	3	2
6	Campa	10	10	0	0	0
7	Gabriela	10	10	0	0	0
8	Misael	10	10	0	0	0
9	Emiliano	10	7	0	1	2
10	Rodrigo	10	7	0	2	1
Porcentajes de los resultados:			80.00%	6.00%	9.00%	5.00%

Hasta este momento los resultados mostrados en la tabla 4 los resultados son aproximados a otros resultados de otros trabajos de investigación del algoritmo LDA donde el porcentaje de autenticación va del 80.00% al 88.75%.



Figura 11 Resultados con la aplicación Android.

En base a los resultados obtenidos en las pruebas del algoritmo PCA y LDA que se mostraron anteriormente, se obtuvieron los porcentajes de autenticación, autenticación incorrecta, falso reconocimiento y error de aplicación.

Tabla 4 Resultados de los algoritmos PCA y LDA con 2 bases de datos distintas.

Algoritmo	DB	NP	Porcentajes			
			AC	AI	FR	EA
LDA Python	Yalefaces	14	87	6	7	0
PCA Python	Yalefaces	14	79	9	12	0
LDA Python	Propia	10	87	7	6	0
LDA Android	Propia	10	80	6	9	5

4. Discusión

El algoritmo de análisis de discriminación lineal mostro claramente ser superior al algoritmo de Análisis del componente principal, principalmente en condiciones diferentes de iluminación y distancia, aunque en la aplicación Android se mostraron algunas deficiencias de reconocimiento ya que al ser un dispositivo móvil al momento de tomar una foto, la misma puede estar distorsionada o bien en condiciones de iluminación que no son las ideales.

En trabajos anteriores de los que se habló, se mostraron buenos resultados con PCA debido a que se aplicaron acciones como procesamiento de imágenes y técnicas de visión por computadora como el face tracking, con estas acciones y a la propuesta de una metodología, es claro que es necesario contar con estos elementos para formar un buen sistema de autenticación y reconocimiento facial.

5. Conclusiones

En este trabajo de investigación se muestra la propuesta de autenticación facial utilizando el algoritmo LDA, la metodología que se siguió fue la de implementar el algoritmo de LDA y PCA en Python para poder tener una idea preliminar y contar con los elementos necesarios e implementar posteriormente el algoritmo más eficiente en una aplicación Android.

Se obtuvieron tiempos de respuesta aceptables para este algoritmo y por ser un algoritmo que funciona bien en condiciones de iluminación y distancia diferentes, a pesar de esto, se llegó a la conclusión de ser necesario una mejor metodología de sistema de reconocimiento facial.

Es posible mejorar el trabajo mostrado ya que en trabajos futuros se plantea

seguir una metodología de autenticación facial, y pre procesamiento de imágenes, este trabajo se enfocó principalmente en implementar el algoritmo en un sistema Android y realizar pruebas para probar únicamente su eficiencia.

6. Bibliografía y Referencias

- [1] Aceves M. A. y J. M. Ramos, Fundamentos de Sistemas Embebidos (Ed.). Asociación Mexicana de Mecatrónica A.C. México, 2012.
- [2] Anton, Howard. Introducción al álgebra lineal/Por Anton, Howard. No. 512.897 A5.
- [3] Brumnik, R., Podbregar, I. y Ivanuša, T., Reliability of Fingerprint Biometry (Weibull Approach). En Z. Riaz, Biometric Systems, Design and Applications, 2011.
- [4] Chapra, Steven C. Canale, et al., Métodos numéricos para ingenieros. McGraw-Hill, 2007.
- [5] Delbracio, M., & Mateu, M., Trabajo Final de Reconocimiento de Patrones: Identificación utilizando PCA, ICA y LDA. Grupo de tratamiento de señales de la Universidad de la República-Instituto de Ingeniería Eléctrica, Montevideo, Uruguay, 2016.
- [6] Duró, V. E., Evaluación de sistemas de reconocimiento biométrico. Departamento de Electrónica y Automática. Escuela Universitaria Politécnica de Mataró, 2001.
- [7] Duc, N. M., & Minh, B. Q., Your face is not your password face authentication bypassing lenovo–asus–toshiba. Black Hat Briefings, 2009.
- [8] Embedinfo, Embedded System Development Specialist: http://www.embedinfo.com/en/ARM_Cortex-list.asp?id=15, 2014.
- [9] Fuentes H. A., Recognition systems base on the facial image, Universidad Industrial de Santander, 2011.
- [10] García, Gloria Bueno, et al., Learning Image Processing with OpenCV. Packt Publishing Ltd, 2015.
- [11] Grossman, Stanley I., and Fernando Piña Soto, Álgebra lineal. No. 512.5 G7A4 1996 QA184. G37 1996. Grupo Editorial Iberoamericana, 1983.

- [12] Hernández, R. G., Estudio de técnicas de reconocimiento facial., Departamento de Procesado de Señal y Comunicaciones. http://upcommons.upc.edu/pfc/bitstream/2099.1/9782/1/PFC_RogerGimeno.pdf, 2010.
- [13] Howse, Joseph. Android Application Programming with OpenCV 3. Packt Publishing Ltd, 2015.
- [14] Ifarraguerri A. y Chang C. I., Multispectral and hyperspectral image analysis with projection pursuit, IEEE Transactions on Geoscience and Remote Sensing. Approach). En Z. Riaz, Biometric Systems, Design and Applications, 2000.
- [15] Introna, L., & Nissenbaum, H., Facial Recognition Technology A Survey of Policy and Implementation Issues, 2010.
- [16] Martínez C., FACE recognition in the context of Smart Rooms, Department of Signal Theory and communications UPC, 2005.
- [17] Pentland A. y Adelson T., Face Recognition Demo Page, MIT Media Laboratory, 2014.
- [18] Raschka, Sebastian, Implementing a principal component analysis (pca) in python step by step, 2014.
- [19] Raschka, Sebastian, Linear Discriminant Analysis bit by bit, 2014.
- [20] Raschka, Sebastian, Python Machine Learning. Packt Publishing Ltd, 2015.
- [21] Sandoval, A. E. L., Mendoza, C., Martínez, L. Á. R. C., Rivas, E. A., Araiza, J. M. R. A., Carlos, J., & Ortega, P., Sistema de Autenticación Facial mediante la Implementación del algoritmo PCA modificado en Sistemas embebidos con arquitectura ARM. La Mecatrónica en México, 4, pp. 53-64, 2015.
- [22] Serratosa, F., La biometría para la identificación de las personas. Universitat Oberta de Catalunya, pp. 8-20, 2008.
- [23] Turk, M., & Pentland, A., Eigenfaces for recognition. Journal of cognitive neuroscience, 3(1), pp. 71-86, 1991.
- [24] Villalba, A., Artacho, J. M., Sanchez, D., & Bernués, E., Autenticuz: Sistema de reconocimiento facial para control de acceso automático [DISK]. Zaragoza: Universidad de Zaragoza, 2004.
- [25] Viola, P., & Jones, M., Rapid object detection using a boosted cascade of

simple features. In *Computer Vision and Pattern Recognition, 2001, CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on IEEE*, Vol. 1, pp. I-511, 2001.

[26] Vision and Modeling Group Vismod: <http://vismod.media.mit.edu/vismod/demos/facerec/basic.html>.

[27] Walpole, Ronald E., Raymond H. Myers, and Sharon L. Myers, *Probabilidad y estadística para ingenieros*. Pearson Educación, 1999.

[28] Welling, M., Fisher-Ilda. Technical report: http://www.ics.uci.edu/welling/classnotes/papers_class/Fisher-LDA. Pdf, 2005.

COMPARACIÓN DE LAS TÉCNICAS DE DETECCIÓN DE CRUCE POR CERO Y LA TRANSFORMADA Z-CHIRP PARA MEDIR FRECUENCIAS EN EL RANGO ULTRASÓNICO

Guadalupe Aguilar Cerda

Universidad Autónoma de Querétaro

lups_doll@hotmail.com

Luis Morales Velázquez

Universidad Autónoma de Querétaro

lmorales@hspdigital.org

Resumen

Este trabajo presenta una comparativa entre 2 métodos para detectar y medir la frecuencia con una alta precisión en el rango ultrasónico. La aplicación propuesta para esta investigación es la medición de velocidad con ultrasonido mediante efecto el Doppler en actuadores lineales, esta se desarrollará en un trabajo posterior. Se generaron señales sintéticas con distintos niveles de ruido, simulando la señal entregada por un sensor ultrasónico. Para la detección de frecuencia se diseñó una metodología para comparar las técnicas de detección de cruce por cero y la Transformada-Z Chirp. La Transformada-Z Chirp, tiene mejores resultados ya que se tiene una buena aproximación de la frecuencia real, y el error no incrementa en señales con ruido, en cambio en la detección de cruce por cero el error incrementa mostrando unos picos indeseables. Una vez que la metodología se perfeccione, se implementará en un sistema embebido para el procesamiento en tiempo real.

Palabras Claves: Alta resolución, detección de cruce por cero, efecto Doppler, estimación de frecuencia, transformada-Z Chirp, ultrasonido.

Abstract

This work presents a comparison between 2 methods to detect and measure the frequency with a high precision in the ultrasonic range. The proposed application for this research is the measurement of velocity with ultrasound by Doppler effect in linear actuators, this will be developed in a later work. Synthetic signals were generated with different levels of noise, simulating the signal delivered by an ultrasonic sensor. For frequency detection, a methodology was designed to compare zero crossing detection techniques and the Chirp Z-Transform. The Chirp Z-Transform, has better results since it has a good approximation of the real frequency, and the error does not increase in signals with noise, instead in the detection of crossing by zero the error increases showing some undesirable peaks. Once the methodology is perfected, it will be implemented in an embedded system for real-time processing.

Keywords: *Chirp z-transform, Doppler effect, frequency estimation, high resolution, ultrasound, zero crossing detection.*

1. Introducción

La frecuencia es un parámetro de gran importancia, por ello la precisión en la estimación de la misma es indispensable, la detección y medición de la frecuencia de una señal, ya sea pura o con ruido es un problema que ha sido estudiado en distintos trabajos para varias aplicaciones. Dentro de las metodologías propuestas están la Transformada Discreta de Fourier [Venkataramanan, 2006] que es una de las más utilizadas al igual que técnicas de estimación basadas en la interpolación de la Transformada Rápida de Fourier [Qi, 2004]. Se ha demostrado a pesar de tener un procesamiento rápido, la desventaja de utilizar la Transformada Rápida de Fourier, es que su rendimiento y resolución dependen de la relación señal ruido (SNR) y del número de muestras de la señal analizada. Para mejorar la precisión y resolución, [Yulan, 2007] combinaron una técnica de interpolación cuadrática y la Transformada Rápida de Fourier. Es importante destacar que las aplicaciones en las que utilizan la Transformada discreta de Fourier y la Transformada rápida de Fourier son mayormente para la estimación de parámetros como ángulo, fase, y

frecuencia entre otros en los sistemas eléctricos de potencia [Phadke et al, 1983]. La estimación de la frecuencia mediante detección de cruce por cero es ampliamente utilizada debido a su simplicidad [Friedman, 1994], el error presentado es discreto, el cual puede atenuarse aún más agregando un filtro. El algoritmo supervisado de Gauss-Newton (SGN) presenta la combinación de la Transformada Discreta de Fourier, el método de detección de cruce por cero y un filtro de respuesta infinita al impulso (IIR) mejorando así la estimación de errores [Xue, 2009]. La Transformada-Z Chirp presenta una resolución de frecuencia mucho más alta que la presentada por las técnicas anteriormente descritas (FFT y Detección de cruce por cero). La Transformada-Z Chirp permite la evaluación de la Transformada Z en M puntos equi-angularmente espaciados en los contornos que entran o salen en espiral desde un punto arbitrario en el plano Z. Dentro de las aplicaciones de este algoritmo se encuentra el análisis de frecuencia en alta resolución, la interpolación del tiempo de datos de una tasa de muestreo a cualquier otra tasa de muestreo y la mejora de los polos para su uso en el análisis espectral [Rabiner et al, 1969]. La importancia de encontrar la técnica adecuada para la medición de frecuencia en el rango ultrasónico con una alta resolución radica en la aplicación que se le desea dar al método propuesto. La aplicación consiste en resolver el problema que se tiene al medir velocidad. Uno de los métodos que usualmente se utilizan, es el de estimar la velocidad a partir de la posición del encoder esto debido a que el encoder es el sensor más común (típico) en el control de movimiento. La medición es simple, ya que se basa en la diferencia de recuentos sucesivos del encoder. A altas velocidades la estimación con este método proporciona resultados relativamente precisos, pero a bajas velocidades e inclusive velocidades extremadamente bajas, la estimación tiene una fiabilidad bastante baja. Un encoder de alta resolución podría proporcionar una estimación de velocidad mucho más precisa incluso para bajas velocidades, pero el coste de implementación, puede ser muy elevado tomando en cuenta que la aplicación no lo requiera [Jeon, 2007]. El uso de acelerómetros para la estimación de la velocidad ha ido en aumento gracias al desempeño que han mostrado y a la reducción de su coste. La velocidad se puede estimar integrando

la aceleración, pero debido a que no es una medición directa y se recurre a la integración, se presenta un error, el cual crece sin límites debido a la polarización y la deriva de la salida del acelerómetro. Para solucionar este problema también se utiliza la medición de la posición y así realizar una compensación para tener una estimación de la velocidad más precisa [Shim et al, 1998]. La propuesta de acelerómetros MEMS con sensores de baja resolución, utilizando el filtro cinemático de Kalman combina las mediciones de posición y aceleración para realizar la estimación de la velocidad. Este método puede presentar perturbaciones a la salida del acelerómetro [Tomizuka, 2001]. La implementación del efecto Doppler para distintas aplicaciones ha ido en aumento. La velocimetría láser de efecto Doppler (laser Doppler Velocimetry, LVD) puede ser usada para medir velocidades de flujo con exactitud y sin invasión. El método consiste en la observación de una luz reflejada, esta luz es generada originalmente por un láser y el estudio de las reflexiones permite derivar la velocidad local e instantánea. El espectro Doppler obtenido a partir del desplazamiento Doppler en frecuencia del rayo láser incidente al mover objetos sólidos o dispersores en flujos, se ha utilizado desde un inicio para propósitos de detección, incluyendo la flujometría y aplicaciones biomédicas [Mowla et al, 2014]. El uso del ultrasonido ha evolucionado de manera importante hasta volverse una de las herramientas de diagnósticas más importantes en el campo de la medicina para la detección de enfermedades gracias a que es una técnica no invasiva y por ello es la más utilizada en todo el mundo [Dávila et al, 2016]. Pero las aplicaciones del ultrasonido no se limitan únicamente al área médica. Se opta por utilizar el método Doppler por ultrasonido debido a la propagación de la onda, en el ultrasonido se propaga con un haz pequeño, de mayor concentración, el cual tiene poca dispersión, a su vez es direccional, en cambio el sonido, tiene una dispersión muy rápida, lo cual provoca la generación de ecos, que al momento de realizar la medición, pueden existir que provengan de distintos lados, que se vuelva señales indeseables, las cuales posteriormente sean discriminadas y llevar a cabo el filtrado [Pye et al, 1998]. En investigaciones actuales, el uso del efecto Doppler por ultrasonido, está enfocado exclusivamente para aplicaciones en áreas médicas,

como estimación no invasiva de la presión sistólica mediante la ecografía Doppler [Yock, 1984]. En los trabajos existentes, la medición de la frecuencia está enfocada para el monitoreo de la calidad en sistemas de potencia, lo cual implica la medición de la línea eléctrica, la cual se encuentra a frecuencias muy bajas, es decir, oscila en el rango de los 60 Hz.

La aportación de este trabajo es la medición de frecuencias en el rango ultrasónico (>20kHz) con una alta precisión y resolución. Una vez que la técnica de medición se encuentre sustentada como la más adecuada, el procesamiento del algoritmo se implementará en un arreglo de compuertas programables en campo (FPGA) debido a que como las frecuencias que se manejan son muy altas, será un cálculo pesado y es necesario hacerlo en un dispositivo de alto desempeño que trabaje a una alta velocidad.

2. Métodos

En este apartado se presenta la metodología para la detección y medición de la frecuencia que consiste en la comparación de 2 técnicas, la detección de cruce por cero y la Transformada-Z Chirp. La cual consiste en la generación de 3 señales sintéticas, de las cuales 2 contienen ruido, una en mayor proporción. Una vez que se tengan las señales generadas, se les aplican ambos métodos implementados en Matlab®. Teniendo los resultados de los 2 métodos, se procede a realizar la comparación entre ambos mediante el análisis estadístico del error. La figura 1 muestra el diagrama a bloques de la metodología propuesta para la detección y medición de la frecuencia en el rango ultrasónico de alta resolución.

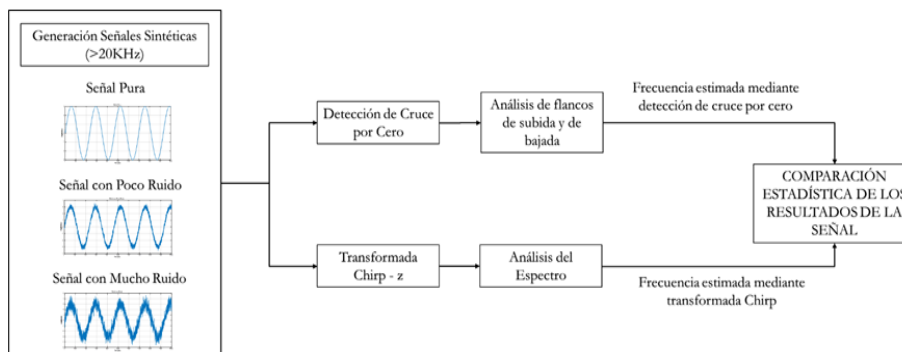


Figura 1 Diagrama a bloques de la metodología propuesta.

Primero se generaron las señales sintéticas, simulando la señal entregada por el sensor ultrasónico. La frecuencia de las señales oscila entre los 30 kHz hasta los 50 kHz, es decir, dentro del rango ultrasónico (>20 kHz). Además dos de las señales generadas contendrán ruido, el cual está presente en las señales reales. Con las señales generadas, se procedió a aplicar 2 de los métodos que más comúnmente se utilizan para medición de frecuencia, con la finalidad de comparar ambas técnicas. Las técnicas utilizadas fueron la de detección de cruce por cero y la Transformada-Z Chirp. En ambos métodos las señales mantuvieron restricciones y limitaciones basadas en la aplicación contemplada a futuro. Una de las restricciones es la resolución. Se restringió la resolución de las señales, y los cálculos con la finalidad de implementar en tiempo real el método en el procesador (FPGA) para tener un sensor en tiempo real y llegar a la medición de la velocidad mediante el efecto Doppler en el rango ultrasónico, lo cual es difícil ya que las frecuencias son muy altas y un dispositivo de baja gama como un microcontrolador, arduino, entre otros, no tiene la velocidad suficiente para realizar la adquisición y el procesamiento al mismo tiempo. La detección de cruce por cero consiste en el cálculo de un punto donde el signo de la función cambia ya sea de positivo a negativo o viceversa, representado por un cruce del eje, el cruce por cero es el punto donde no hay amplitud. El conteo de cruces por cero es un método utilizado en el procesamiento para estimación de la frecuencia. Para ello se diseñó un algoritmo que hace lo siguiente, a partir de la señal generada de manera sintética, se guarda la señal en un arreglo de 4096 valores. Ese arreglo se comparó con cero, si es mayor o menor a cero, en el caso de que sea mayor a cero se le asigna el valor de 1, si es menor a cero, se le asigna el valor de -1, así es la manera en que detecta los cruces por cero, en el cambio de signo de la señal. Así se tendrá un arreglo con unos y menos unos, el paso siguiente es hacer la diferencia de números consecutivos dando como resultado un arreglo de 0, +2, y -2. Los valores que interesan son únicamente los diferentes de 0, ya que ahí se indica el cambio de signo o cruce por cero, al utilizar el absoluto a los valores del arreglo únicamente quedan valores positivos. Posteriormente se obtiene la longitud de la señal dividida entre 2, cada uno de los valores que se tienen en el

arreglo se van a dividir entre ese valor, para este caso 2048. Al final se hace una sumatoria de todos los elementos del arreglo, dando como resultado un valor numérico. Para obtener la frecuencia, se multiplicó ese valor numérico obtenido con anterioridad por la frecuencia de muestreo en este caso 200 000 muestras por segundo (samples per second, Sps) y se dividió entre 2. La Transformada-Z Chirp permite el cálculo rápido de la Transformada Z en ciertos puntos dentro de una región de la circunferencia de radio. Es útil cuando no se desea evaluar la Transformada Discreta de Fourier en todo el intervalo, sino sólo en un rango de frecuencias [Albertí, 2006]. La Transformada-Z Chirp $X(k)$ de una secuencia de N puntos $x(n)$ para $n = 0, 1, 2, \dots, N - 1$ está dado por la ecuación 1.

$$X(k) = \sum_{n=0}^{N-1} x(n) Z_L^{kn} \quad (1)$$

En dicha ecuación $X(k)$ permite calcular los contenidos de frecuencia de $x(n)$ muestreados a una frecuencia f_s , en un conjunto de L frecuencias en el rango cubierto por el arco del círculo unitario que comienza en $\omega_0 = 2\pi f_0$ y termina en $\omega_1 = 2\pi f_1$ [Proakis, 1996].

En la ecuación 1 la transformación kernel Z_L^{kn} está dada por la ecuación 2.

$$Z_L^{kn} = \exp\left\{ -j \frac{2\pi n}{f_s} \left[f_0 + \frac{(f_1 - f_0)k}{L} \right] \right\} = \cos(\omega n) - j \sin(\omega n) \quad (2)$$

Donde:

$$\omega = \frac{2\pi n}{f_s} \left[f_0 + \frac{(f_1 - f_0)k}{L} \right] \text{ y } k = 0, 1, \dots, L - 1$$

La transformación kernel Z_L^{kn} se puede implementar como 2 funciones discretas recursivas descritas por la ecuación 3 y la ecuación 4 para la componente real Z_R y la componente imaginaria Z_I .

$$Z_R(n) = Z_R(n-1) \cos(\omega n) - Z_I(n-1) \sin(\omega n) \quad (3)$$

$$Z_I(n) = Z_I(n-1) \cos(\omega n) + Z_R(n-1) \sin(\omega n) \quad (4)$$

De igual manera que en la detección de cruce por cero, se diseñó un algoritmo que permitiera aplicar la Transformada-Z Chirp a la señal deseada. Se obtuvo la frecuencia a partir del análisis de su espectro.

3. Resultados

En este apartado se muestran las simulaciones correspondientes a la comparación de las 2 técnicas para la detección y medición de frecuencia. Al realizar las simulaciones de los algoritmos diseñados mediante la herramienta matemática Matlab ®.

En la gráfica de la figura 2 se puede observar que las señales sintéticas generadas a través del algoritmo cuentan con las restricciones específicas del sistema, como lo son el rango de frecuencias ultrasónicas (30-50 kHz) al que pertenecen, y una frecuencia de muestreo de 200 kSps.

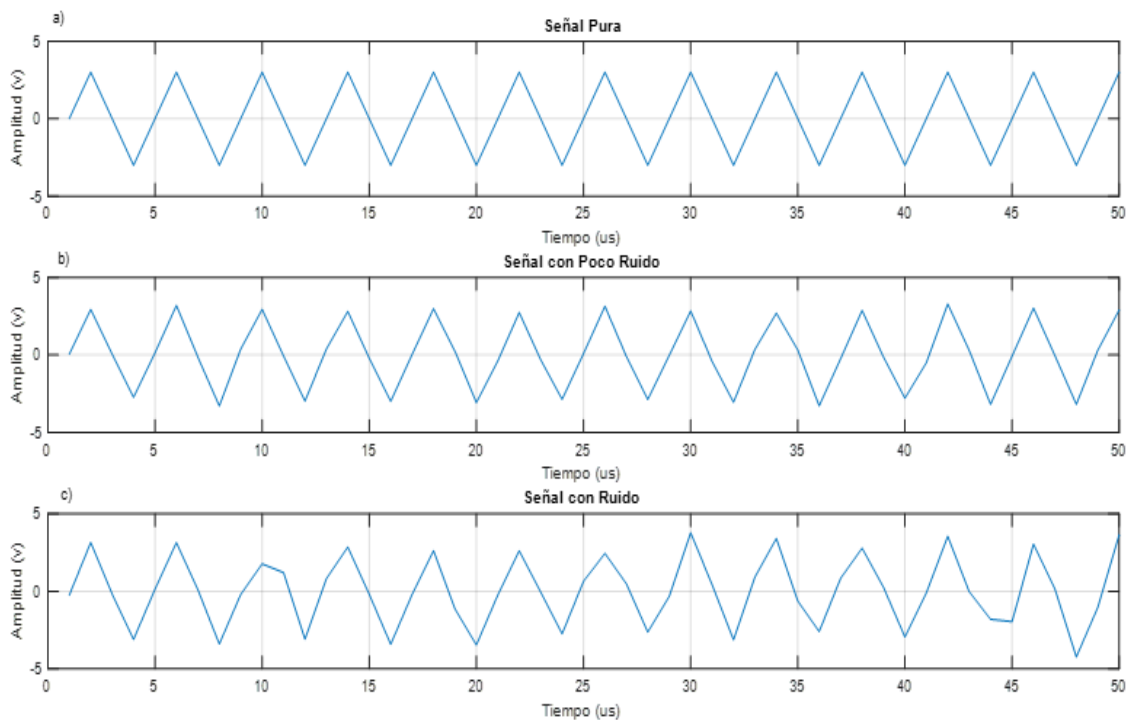


Figura 2 Generación de señales sintéticas (pura, poco ruido, con ruido) con una frecuencia en el rango ultrasónico.

De igual manera se observan 3 gráficas de las cuales la primera, es una señal pura, por el contrario, a las otras 2 señales, se les agregó ruido blanco Gaussiano con una relación señal-ruido por muestra de 10 y 20 dB, esto con la finalidad de simular las condiciones reales que estaría entregando un sensor ultrasónico. Una vez que se obtuvieron las señales sintéticas con las especificaciones requeridas, se procedió a aplicar los dos métodos descritos con anterioridad para una comparación entre ambos métodos.

El algoritmo para la detección de cruce por cero, es muy sencillo y entrega una buena estimación de la frecuencia. En el caso del algoritmo para detectar la frecuencia mediante Transformada-Z Chirp, se analiza el espectro para así poder obtener la frecuencia. En el caso de la Transformada-Z Chirp, el análisis se realizó a partir de su espectro, a continuación se muestran las gráficas.

La figura 3 presenta el espectro de la señal a la cual se le agregó ruido, es evidente como se empieza a notar la presencia de ese ruido en su espectro.

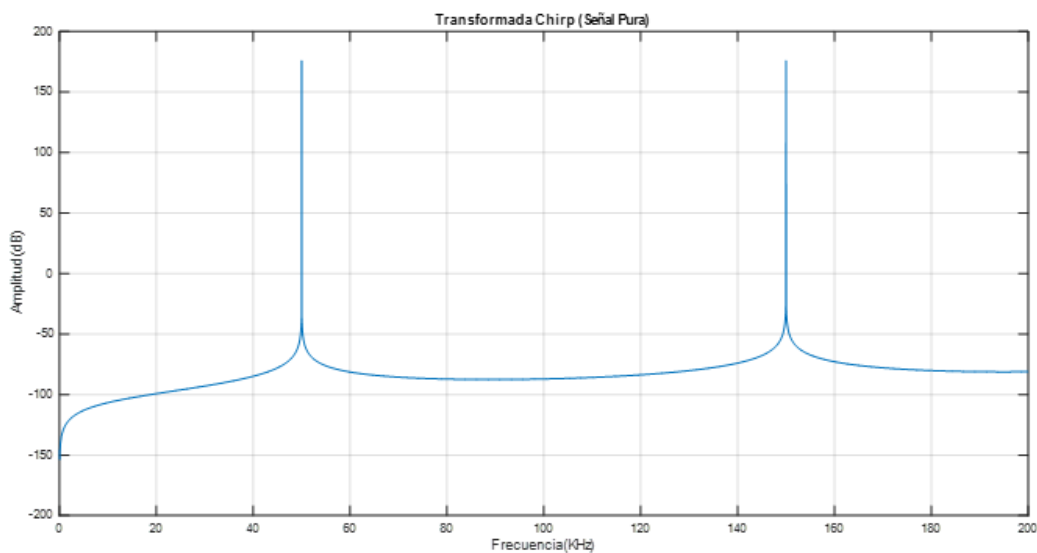


Figura 3 Espectro de la Transformada Z Chirp (Señal Pura).

En la figura 4 y en la figura 5, que son las señales a las que se les agregó ruido, se observa claramente que el ruido está presente en ambas señales, para la señal de la figura 4 se le inyectaron 20dB, en el caso de la figura 5, el ruido inyectado fue de 10dB.

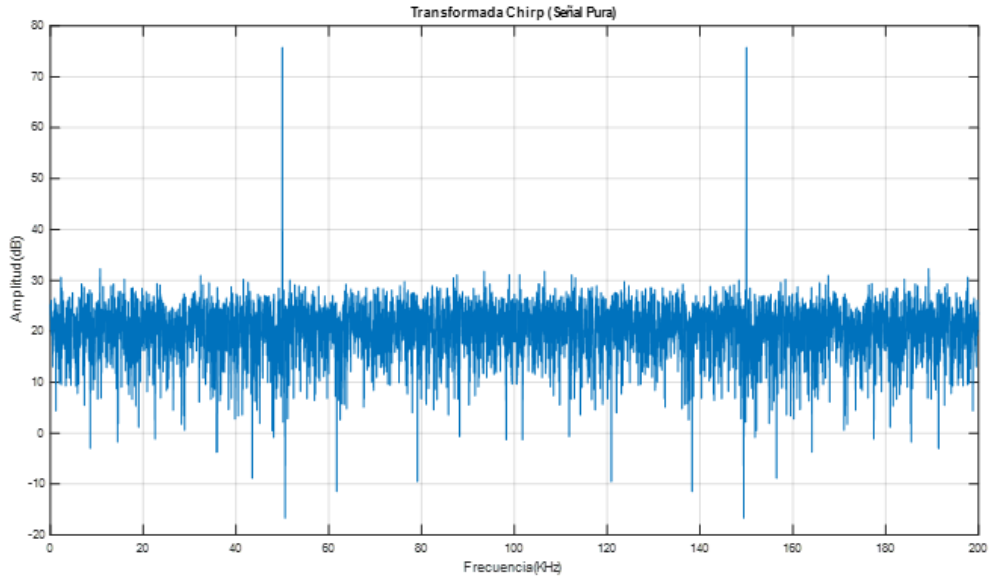


Figura 4 Espectro de la Transformada-Z Chirp (Señal con Poco Ruido).

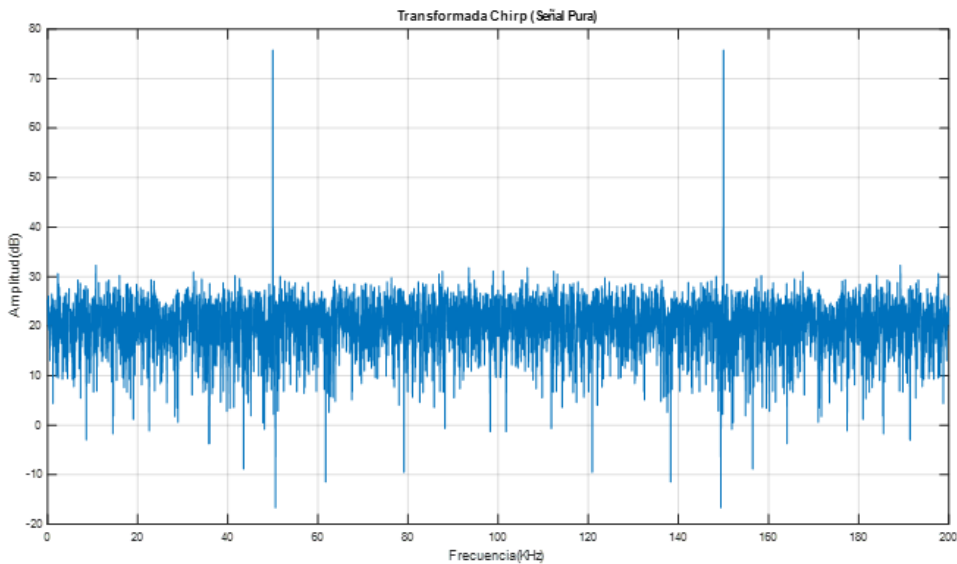


Figura 5 Espectro de la Transformada-Z Chirp (Señal con Ruido).

La figura 5 presenta el espectro de la señal que contiene la mayor cantidad de ruido. Se puede observar que la señal de ruido no afecta la medición del punto máximo.

Para realizar la comparación entre ambos métodos, se obtuvo el error entre la frecuencia de la señal de referencia y la frecuencia que entrega el algoritmo de detección mostrado al aplicar las dos técnicas: la Transformada-Z Chirp y la

detección de cruce por cero para los 3 tipos de señales generadas de manera sintética realizando un barrido de frecuencia desde los 30 hasta los 50 kHz.

En la figura 6 se muestra la comparativa de los errores presentados por ambas técnicas para cada una de las señales sintéticas generadas. En tablas 1 y 2 se muestran algunos parámetros obtenidos de las gráficas de error.

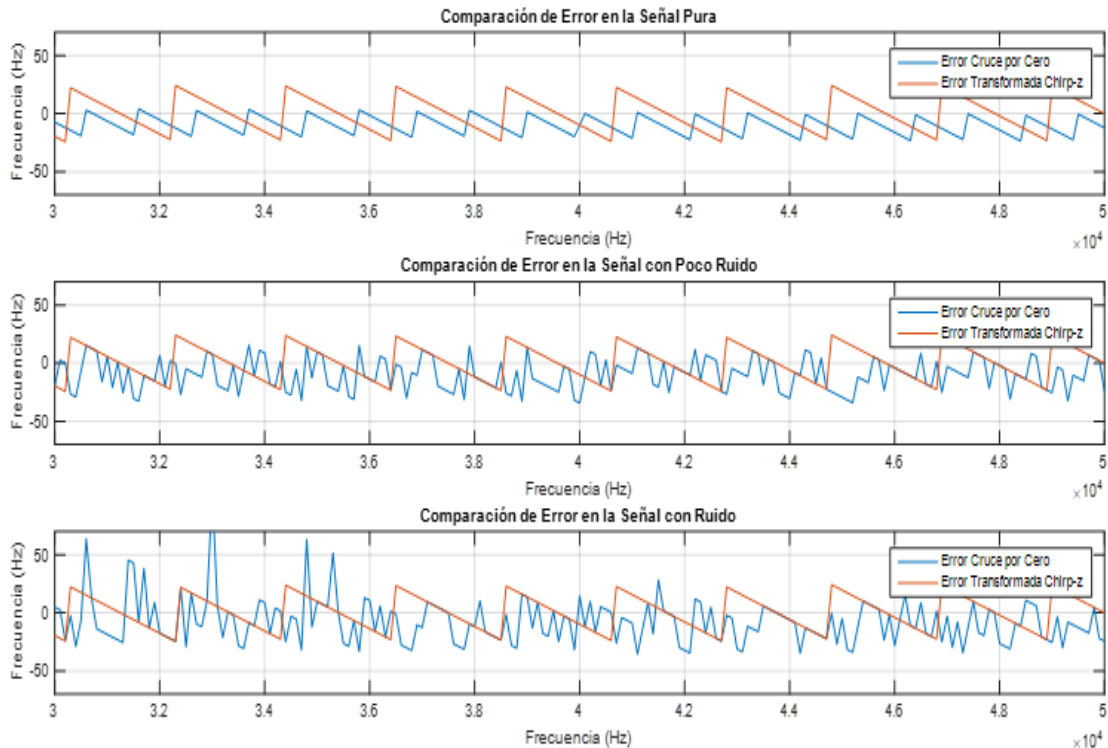


Figura 6 Comparativa de error de las técnicas de cruce por cero y Transformada-Z Chirp para las 3 señales sintéticas generadas.

Tabla 1 Parámetros obtenidos de las gráficas de error para el cruce por cero.

Error Señal Sintética				
Señal Sintética	Cruce por cero			
	Error cuadrático medio	Desviación estándar	Valor mínimo	Valor máximo
Pura	146.5892792	7.123553213	-4.00390625	23.53515625
Poco Ruido	484.8575592	20.28231855	-97.4609375	36.71875
Ruido	334.6538544	16.85483291	-83.7890625	37.6953125

Tabla 2 Parámetros obtenidos de las gráficas de error para la Transformada-Z Chirp.

Error Señal Sintética				
Señal Sintética	Transformada-Z Chirp			
	Error cuadrático medio	Desviación estándar	Valor mínimo	Valor máximo
Pura	202.6176453	14.22763111	-24.21875	24.21875
Poco Ruido	202.6176453	14.22763111	-24.21875	24.21875
Ruido	202.7130127	14.23639897	-24.21875	24.609375

4. Discusión

Al analizar las gráficas de las señales sintéticas mostradas en las figura 3, 4 y 5, es interesante observar que para identificar la frecuencia de la señal se parte del pico ignorando las demás señales mostradas, en otras palabras, al analizar mediante su espectro, el ruido es despreciado. Posteriormente, se obtuvieron los errores de cada una de las señales y se compararon ambos métodos. Para una mejor visualización, esos datos se muestran en las gráficas. Al comparar las gráficas de las figuras 6 se observa que en el caso de la figura 6a, que es la señal pura, el error mostrado por la técnica Transformada-Z Chirp, es mayor que el de detección de cruce por cero. De primera instancia se pudiera pensar que el método de detección de cruce por cero presenta un menor error. Por ello se decidió generar las señales con ruido y comprobar si el comportamiento seguía siendo el mismo. Pero al analizar las figuras 6b y 6c ambas con ruido (en diferentes porcentajes), se muestra que el error mostrado por la Transformada-Z Chirp se mantiene constante sin mostrar grandes cambios. En cambio el error mostrado por el cruce por cero es mucho mayor, inclusive muestra varios picos. Esto debido a lo que se mencionó anteriormente, al analizar el espectro, únicamente se toma el pico máximo e ignora las demás componentes. Por ello, pensando en la aplicación práctica, las señales reales entregadas por el sensor difícilmente serían señales puras. Con los datos obtenidos a partir de las gráficas, se analizaron algunos parámetros de error, los cuales son mostrados en la tabla 1 y tabla 2, se puede observar que el error cuadrático medio se mantiene en la Transformada-Z Chirp para los 3 tipos de señal por el contrario, los valores de

cruce por cero, se nota como se va deteriorando la señal al agregarle ruido. La desviación estándar, indica la variación que hay en la estimación de la frecuencia con la señal con o sin ruido, al tener ruido se va a ir alejando de la frecuencia estimada. Es importante la diferencia entre el valor mínimo y el máximo, ya que indica si existen picos en la señal, en el caso de la detección mediante cruce por cero, la diferencia es muy amplia y varía bastante, en el caso de la Transformada-Z Chirp la diferencia es prácticamente la misma en los 3 tipos de señales lo que indica que no existen tantos picos en la señal, lo cual es deseable debido a se puede tomar valores erróneos en la medición. Al comparar de cruce por cero y Transformada-Z Chirp para medición de frecuencia se puede concluir que cuando las señales son puras, la técnica de Transformada-Z Chirp no funciona de la manera esperada ya que existe una mayor cantidad de variación que con el cruce por cero. Ahora en señales que contienen ruido, el cruce por cero presenta la desventaja de que al incrementar el ruido, el error se va haciendo más, lo que significa que se tienen picos en la señal que pueden afectar su posterior procesamiento. Precisamente la ventaja de la Transformada-Z Chirp es que no importa el nivel de ruido que tenga la señal, el error se mantiene igual, no va creciendo, lo que se traduce en que no presenta picos ni variaciones indeseables. Para la aplicación propuesta de medición de velocidad por ultrasonido mediante efecto Doppler es necesario tener esa característica, ya que no se sabe que cantidad de ruido tenga la señal entregada por el sensor ultrasónico. Por ello lo ideal es utilizar una técnica que no tenga variaciones drásticas a pesar de tener ruido.

5. Conclusiones

Este trabajo presenta el desarrollo de un método de detección de frecuencias en el rango ultrasónico de alta resolución que servirá como base, para la futura aplicación de medición de velocidad por ultrasonido mediante el efecto Doppler en tiempo real para actuadores lineales. Por lo tanto la técnica más conveniente para la aplicación propuesta en el trabajo posterior, debido a la inmunidad que tiene al ruido es la Transformada-Z Chirp. Una vez que el método de detección se tenga

perfeccionado, se tiene pensado implementarlo en un dispositivo FPGA debido a que los cálculos y la velocidad necesaria para realizar la adquisición y el procesamiento solo lo puede hacer un dispositivo de alta gama.

Agradecimientos

Este trabajo ha sido parcialmente financiado por la beca CONACyT 742849 y el proyecto FIN201613 de la Universidad Autónoma de Querétaro.

6. Bibliografía y Referencias

- [1] Albertí, E. B., *Procesado digital de señales-II: Fundamentos para comunicaciones y control* Vol. 170, Universitat Politècnica de Catalunya. Iniciativa Digital Politècnica, 2006.
- [2] Dávila, F., Barros, L. A., Reynolds, J., Lewis, A. J., & Mogollón, I. R., *El ultrasonido: desde el murciélago hasta la cardiología no invasiva*. Revista Colombiana de Cardiología, 2016.
- [3] Friedman, V., *A zero crossing algorithm for the estimation of the frequency of a single sinusoid in white noise*. IEEE Transactions on Signal Processing, 42(6), pp. 1565-1569, 1994.
- [4] Jeon, S., & Tomizuka, M., *Benefits of acceleration measurement in velocity estimation and motion control*. Control Engineering Practice, 15(3), pp. 325-332, 2007.
- [5] Mowla, A., Nikolić, M., Lim, Y. L., Bertling, K., Rakić, A. D., & Taimre, T., *Effect of the optical numerical aperture on the Doppler spectrum in laser Doppler velocimetry*. In *Optoelectronic and Microelectronic Materials & Devices (COMMAD)*, pp. 72-74, December, 2014
- [6] Phadke, A. G., Thorp, J. S., & Adamiak, M. G., *A new measurement technique for tracking voltage phasors, local system frequency, and rate of change of frequency*. IEEE transactions on power apparatus and systems, pp. 1025-1038, 1983.
- [7] Proakis, J. G., & Manolakis, D. G., *Digital signal processing: principles, algorithms, and applications*, 1996.

- [8] Pye, J. D., & Langbauer Jr, W. R., Ultrasound and infrasound. In *Animal acoustic communication* Springer Berlin Heidelberg, pp. 221-250, 1998.
- [9] Qi, G. Q., & Jia, X. L., Accuracy analysis of frequency estimation of sinusoid based on interpolated FFT. *Acta Electronica Sinica*, 32(4), pp. 625-629, 2004.
- [10] Rabiner, L. R., Schafer, R. W., & Rader, C. M., The Chirp z-Transform Algorithm and Its Application. *Bell Labs Technical Journal*, 48(5), pp. 1249-1292, 1969.
- [11] Shim, H., Kochem, M., & Tomizuka, M., Use of accelerometer for precision motion control of linear motor driven positioning system. In *Industrial Electronics Society, 1998. IECON'98. Proceedings of the 24th Annual Conference of the IEEE, Vol. 4*, pp. 2409-2414, 1998.
- [12] Tomizuka, M., State/Parameter/Disturbance Estimation with Accelerometer in Precision Motion Control of Linear Motor. In *Proceedings 2001 ASME International Mechanical Engineering Congress, IMECE2001/DSC-24578*, 2001.
- [13] Venkataramanan, R., & Prabhu, K. M. M. Estimation of frequency offset using warped discrete-Fourier transform. *Signal Processing*, 86(2), pp. 250-256, 2006.
- [14] Xue, S. Y., & Yang, S. X., Power system frequency estimation using supervised Gauss–Newton algorithm. *Measurement*, 42(1), pp. 28-37, 2009.
- [15] Yock, P. G., & Popp, R. L., Noninvasive estimation of right ventricular systolic pressure by Doppler ultrasound in patients with tricuspid regurgitation. *Circulation*, 70(4), pp. 657-662, 1984.
- [16] Yulan, C., & Cunyang, F., A new method of frequency measurement of power system. In *Industrial Electronics and Applications, 2007. ICIEA 2007. 2nd IEEE Conference on*, pp. 2522-2525, May 2007.

REALIDAD AUMENTADA CON MARCADORES CUADRADOS Y NATURALES PARA NAVEGACIÓN QUIRÚRGICA

Eliana Margarita Aguilar Larrarte

Universidad del Cauca
eaguilar@unicauca.edu.co

Oscar Andrés Vivas Albán

Universidad del Cauca
avivas@unicauca.edu.co

José María Sabater Navarro

Universidad Miguel Hernández
j.sabater@goumh.umh.es

Resumen

Este artículo muestra el resultado de pruebas preliminares con librerías y entornos de desarrollo comerciales encaminados a la construcción de una aplicación para navegación quirúrgica en el área de laparoscopia, usando marcadores cuadrados y marcadores naturales extensibles a despliegues en dispositivos móviles y gafas de realidad virtual. Para la renderización de los objetos 3D se usó el Game Engine Unity junto a librerías especializadas para la visión por computador y realidad aumentada como OpenCV, Vuforia y Kudan. Las pruebas preliminares muestran resultados satisfactorios en el seguimiento de marcadores y en la construcción de la información aumentada útil para el cirujano.

Palabras Claves: Aplicaciones móviles, cirugía laparoscópica, realidad aumentada.

Abstract

This article shows preliminary tests with libraries and integrated development environments (IDE) aimed at building an application for laparoscopic surgical

navigation using square markers and natural feature markers extensible to displays, mobile devices and virtual reality goggles. For the rendering of 3D objects, the Game Engine Unity was used together with specialized libraries for computer vision and augmented reality such as OpenCV, Vuforia and Kudan. Preliminary tests show satisfactory results in the tracking of markers and in the construction of augmented information useful to the surgeon.

Keywords: *Augmented reality, laparoscopic surgery, mobile applications.*

1. Introducción

La cirugía laparoscópica requiere alto nivel de experiencia en el manejo de los instrumentos y buenas capacidades de interpretación visual en cuanto a la profundidad de campo se refiere, lo cual exige un nivel de abstracción superior para la ubicación del instrumental y el desarrollo de tareas en el mundo tridimensional real teniendo en cuenta que esa tarea se realiza a través de una limitada percepción 2D del entorno [Moreno, 2012]. Los cirujanos tienen el obstáculo de ubicarse en un campo visual limitado, con problemas de maniobrabilidad y movilidad que en muchos casos ponen en riesgo la precisión del procedimiento haciéndose aun más complejo en presencia de humo y sangrado. El estudio presentado en [Van Vellen, 2003] muestra un inventario de los problemas encontrados durante 12 operaciones endoscópicas realizadas en un hospital de la ciudad de Eindhoven en Países Bajos. Se distribuyó un cuestionario a todo el personal médico implicado, encontrándose que todo el personal tenía problemas físicos, perceptuales y cognitivos, especialmente los cirujanos, residentes y las enfermeras de operación estéril. Las principales causas fueron el posicionamiento de los aparatos, la ropa de trabajo y el alcance limitado de aparatos y / o instrumentos. Datos relevantes de este estudio son los siguientes: el 50% de los encuestados respondió que existe problemas de percepción y un 63% de que hay incomodidad física. Otra encuesta fue diseñada para examinar las barreras de adopción de la laparoscopia para ginecólogos en ejercicio. La encuesta se aplicó en 4.273 cirujanos ginecológicos a través de los Estados Unidos. De dicho estudio se puede abstraer que del numero total de cirujanos que

pueden aconsejar intervenciones mínimamente invasivas, el 62.50% no lo hace debido a las limitaciones visuales y el 65.60% por la pérdida de visión binocular. En los últimos años la tecnología en el campo quirúrgico ha tenido un avance cada vez mayor, ya sea ayudando en el entrenamiento de cirujanos o asistiéndoles durante la intervención. Se han creado una variedad de sistemas de entrenamiento, asistencia y navegación quirúrgica o también llamados sistemas guiados. Las aplicaciones de los sistemas guiados pueden ser usadas para crear diagramas tridimensionales de datos médicos en tiempo real, ampliando la percepción visual, resultando de gran ayuda en los procesos de navegación quirúrgica [Marescaux, 2015], posicionándose dentro de las tecnologías computacionales que mayor futuro tiene debido al aumento de la realidad y al acercamiento de ampliación con realidad mixta. Dentro de ellas se puede nombrar con toda seguridad la realidad aumentada (*Augmented Reality* - AR). Dicha tecnología basa su funcionamiento en el despliegue de imágenes superpuestas en tiempo real sobre video o fotogramas proporcionados por una cámara [Furts, 2014]. Esta técnica ha sido rápidamente adoptada en aplicaciones médicas debido a que proporciona una mezcla de información adicional a modo de extensión de la percepción visual, proporcionando resultados de mayor precisión por parte de los cirujanos, lo cual deriva en mayor éxito dentro de las intervenciones. En el caso de simulaciones en sistemas de entrenamiento, esta tecnología proporciona incremento de la respuesta mano ojo debido principalmente a dicha expansión de la realidad. La realidad aumentada es de mayor aceptación por el cerebro humano debido a que conserva gran parte del recorrido visual real, en contraste a otra tecnología muy popular llamada realidad virtual, que se caracteriza por crear un ambiente y una visión totalmente artificial [De Lacy, 2013]. En los últimos años las aplicaciones médicas de AR han tenido una rápida expansión, dirigida por avances en el hardware (interfaces hápticas y despliegues), al mismo tiempo que los teléfonos inteligentes y tabletas se han constituido en herramientas cada vez más populares para aplicaciones de AR en medicina, industria y educación [Cukovic, 2016]. La realidad aumentada tiene dos funciones: aumentar la percepción de la realidad (muestra la realidad pero elige qué se puede ver y qué

no), la segunda función crear un ambiente artificial (muestra lo que no es real permitiendo ver lo imaginario), con una percepción aumentada de información útil que ayuda a la toma de decisiones. Los sistemas de AR se caracterizan porque tiene elementos reales y virtuales en un entorno real con alto grado de interactividad en tiempo real, y donde se tiene la opción de registrar y posicionar la información virtual teniendo en cuenta la tridimensionalidad del mundo real [Cukovic, 2016], [Duchene, 2006].

La cirugía laparoscópica, de la misma manera que las restantes cirugías mínimamente invasivas, trabaja bajo la disminución de las inserciones en el paciente usando instrumentos quirúrgicos especiales y un sistema de visión compuesto por una cámara y una lámpara de iluminación fría. Cámara y lámpara se conocen como endoscopio y junto con los otros instrumentos son insertados a través de pequeñas incisiones en la cavidad abdominal. La imagen del área de cirugía es desplegada en una pantalla y el cirujano guiado principalmente por lo que se ve en pantalla manipula los instrumentos y efectúa el proceso operatorio. Este tipo de procedimiento quirúrgico evita cortes de gran extensión aunque debido a las características del método requiere alto nivel de experticia por parte del cirujano [Moreno, 2012]. Muchos procedimientos quirúrgicos tradicionales han sido remplazados por esta técnica laparoscópica, dando origen a otras técnicas derivadas como es el caso de NOTES y LESS. NOTES (*Natural Orifice Transluminal Endoscopic Surgery*) realiza el acceso a la cavidad abdominal a través de orificios naturales del cuerpo humano como la boca, la nariz, el ano y la vagina [Autorino, 2011], [Kipper, 2012]. LESS (*Laparo Endoscopic Single-Site Surgery*) es una técnica donde dicho acceso se realiza por una única incisión [Autorino, 2011], [Rané, 2008].

Entre los entrenadores basados en realidad aumentada se han desarrollado sistemas que permiten a los estudiantes de medicina hacer una primera aproximación al procedimiento de acceso venoso central en recién nacidos. Este es el caso de un sistema desarrollado en la Universidad Militar Nueva Granada en Colombia, el cual posee herramientas para el seguimiento de posición y orientación de un marcador 3D, lo que permite al usuario interactuar con modelos

de herramientas quirúrgicas tales como la jeringa, alambre guía, dispositivo de dilatación y el catéter, cada uno de ellos superpuestos como contenido virtual sobre el marcador. Este sistema también dispone de entradas de teclado con el fin de desplazarse en la escena y cambiar entre las vistas de la piel, el esqueleto o el sistema circulatorio del paciente. El prototipo está programado en Unity3D con el uso de la librería para RA Vuforia y un Oculus VR con una cámara web adjunta [Gutiérrez, 2015]. También se ha implementado un sistema para intervenciones endodónticas, consistente en un software que usa C++, Qt y la librería de procesamiento de imágenes de OpenCV, en la cual el diente intervenido es detectado en imágenes de video de una cámara intraoral, y a partir de un algoritmo de identificación se hace el reconocimiento del canal de raíces del diente. Esta información se sobrepone en la imagen real, donde la localización, el tamaño, la orientación y las distancias son guardadas para posteriores estudios morfológicos [Bruellmann, 2013] En operaciones de irradiación localizada con radiofrecuencia para el tratamiento del cáncer hepático, se utiliza un procedimiento que consiste en introducir una aguja hasta el tumor y aplicar una inyección de radiofrecuencia para hacer morir el tejido canceroso por hipertermia. La acción de ubicar la aguja en el sitio cercano al tumor es una tarea de alta dificultad, ya que para la guía se usa ultrasonido, tomografía computarizada o imágenes de resonancia magnética. En este caso los sistemas de realidad aumentada ayudan en la ubicación de la aguja y en las tareas de planeación preoperatoria, gracias a la visualización de modelos 3D de los órganos del paciente reconstruidos a partir de imágenes médicas [De Paolis, 2011].

En el caso de la cirugía ortopédica y de trauma, la tecnología AR se constituye en una ayuda en el cambio de tareas y en el entendimiento de la relación entre la anatomía, los implantes y las herramientas.

Se han realizado trabajos que permiten una visualización de la escena operatoria, los cuales mezclan las diferentes fuentes de información, entre ellas los datos proporcionados por un sensor Kinect. En este desarrollo se introduce un paradigma basado en aprendizaje de máquina, en el cual se identifican aspectos relevantes de la anatomía con el Kinect y los datos de rayos X por un lado, y por

otro lado se crea un mapa que mezcla las imágenes en video junto a las imágenes de rayos X en una sola vista, obteniendo excelentes resultados en el reconocimiento de la escena quirúrgica y la ampliación del campo de percepción [Pauly, 2015]. ARDental es un sistema de realidad aumentada construido por varias universidades para ayudar al entrenamiento de intervenciones en el campo odontológico, combina elementos reales y modelos 3D, lo cual resulta revolucionario en el sentido de que este tipo de entrenamiento era realizado a partir de imágenes 2D [Peng, 2015].

En cuanto a las mejoras en el sistema de reconocimiento y tracking para realidad aumentada para operaciones mínimamente invasivas los estudios van encaminados al uso de algoritmos más robustos que permitan renderizar los objetos aumentados pese a posibles pérdida en los marcadores naturales usando técnicas mixtas que permitan sobrepasar obstáculos visuales como obstrucciones o problemas de iluminación como en [Puerto-Souza, 2013] en el cual el sistema predice los puntos del marcador natural en el caso de limitaciones visuales, en [Collins, 2016] se propone un SURF con KLT de igual forma en [Mahmound, 2017] se propone algoritmos tipo SLAM para ambientes operatorios usando una cámara endoscópica monocular.

Este artículo muestra el primer grupo de pruebas dirigidas al desarrollo de una aplicación para navegación quirúrgica con realidad aumentada en procedimientos de cirugía mínimamente invasiva haciendo uso de dispositivos Android. El estudio fue enfocado a la identificación de las herramientas software o AR SDKs para Android, y a las pruebas de las mismas estableciendo el alcance en el campo médico.

2. Métodos

Unity es un motor de video juegos ampliamente usado para crear un buen número de ambientes virtuales interactivos 2D y 3D [Duchene, 2006]. Unity incluye también un ambiente integral de desarrollo (IDE) y maneja *scripts* orientados a objetos en tres lenguajes: Boo (un lenguaje propio de Unity language basado en Python), JavaScript y C#. Soporta múltiples plataformas tales como: aplicaciones

web, juegos para consolas como Xbox, Wii y PS3, así como dispositivos móviles como iPhone, iPad, equipos con sistema operativo Android, y otros sistemas operativos como Windows, Mac OS y Linux. Adicionalmente tiene un amplio soporte técnico [Simonetti, 2013] y su entorno de desarrollo es también ampliamente usado en aplicaciones de ingeniería y medicina [Bae, 2016], [Narahara, 2015], [Cristie, 2015], [Soto, 2015]. Vuforia es un paquete de desarrollo software (SDK) enfocado al desarrollo de aplicaciones de realidad aumentada para dispositivos móviles, que permite identificar, clasificar y seguir marcadores visuales. Vuforia, anteriormente llamado (Qcar), permite crear aplicaciones de realidad aumentada con soporte en plataformas IOS Android y Unity 3D [Peng, 2015]. La tarea principal de la aplicación futura a la cual se encamina este trabajo es usar como despliegue un dispositivo móvil con sistema operativo Android como una tableta o un sistema de visión basado en HoloLens, asistiendo en la visualización de la cirugía laparoscópica con información adicional de realidad aumentada como se muestra en la figura 1.

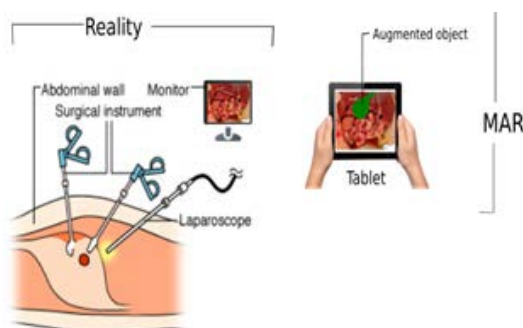


Figura 1 Esquema general del sistema propuesto.

La aplicación de realidad aumentada en el dispositivo móvil (*Mobile Augmented Reality* - MAR) está compuesta por dos partes: hardware y software. El hardware incluye: El dispositivo que despliega simultáneamente el ambiente real del usuario y los objetos virtuales que aumentan la realidad, en este caso una tableta con sistema operativo Android u algún tipo de HoloLens. La segunda parte la constituye el software que incluye reconocimiento y seguimiento; el cual reconoce el marcador, captura el ambiente que el usuario ve y toma la posición del

marcador y la cámara, y por ultimo renderiza los objetos 3D que hacen parte de la aumentación en tiempo real y los despliega también en tiempo real. La figura 2 muestra el diagrama de la creación de la imagen.



Figura 2 Diagrama de la creación de la imagen.

En este trabajo se usaron librerías (SDK: *software development kit*) para realidad aumentada en dispositivos móviles las cuales no son posibles de ser usadas por sí mismas. Es necesario un ambiente de desarrollo software o IDE, que para este caso se usó el motor de video juegos Unity.

Al importar la librería de realidad aumentada en Unity, esta funciona como un *asset* (acción) y todos los recursos para realidad aumentada pueden ser usados: reconocimiento de marcadores, marcadores de objetos pre hechos, seguimiento de cámara AR y demás relacionados.

3. Resultados

Como es necesario estimar la posición y orientación de la cámara con respecto al mundo real y viceversa. La combinación de posición y orientación es llamada “pose” y en este caso se utilizaron dos técnicas de seguimiento o “*tracking*”: seguimiento de marcadores cuadrados (*marker-based tracking*) y seguimiento de características naturales (*markerless*).

Marker Based AR

Uno de las áreas más trabajadas en el reconocimiento de marcadores es el reconocimiento de marcadores cuadrados (*border marker*) en tiempo real, realizándose incluso en situaciones de difícil reconocimiento como al encontrarse girado o sesgado el objeto. Por lo tanto este enfoque resulta ser el más popular y el primero a ser trabajado en los proyectos de realidad aumentada. Un marcador cuadrado es por lo general una imagen 2D impresa en una hoja de papel o superficie lisa. Ese tipo de marcadores son cuadrados y tienen un borde negro de tamaño visible. Durante la fase de seguimiento el sistema realiza una búsqueda de un rectángulo negro y solo si es identificado se procede a examinar el interior de la frontera para determinar el marcador real. Dependiendo de las características del marcador se puede determinar la posición, escala y orientación con respecto a la cámara. En este caso las librerías de realidad aumentada identifican marcadores cuadrados cuyas características son conocidas a partir de un proceso previo de extracción de características en tiempo real. A partir de esta identificación se estima la posición relativa de la cámara. Este sistema está basado en un *framework* de seguimiento que proporciona los datos del reconocimiento y posición del marcador con respecto a la cámara, y un *game engine* para la construcción del mundo virtual sobre Unity. Al desarrollar la aplicación móvil y fijar la cámara en el marcador se obtiene el despliegue en pantalla del objeto 3D relacionado a ese marcador. El objeto tridimensional es hijo del “*image target*”. La captura de pantalla mostrada en la figura 3 muestra un elemento tridimensional renderizado sobre un marcador cuadrado.

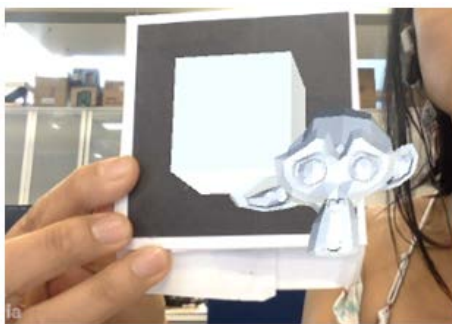


Figura 3 Prueba de deformación entre objetos en el entorno de realidad aumentada.

Posteriormente una vesícula virtual fue relacionada a un nuevo marcador sobre un hígado sintético en el mundo real. Se realizó el ejecutable de la aplicación móvil (apk) y se instaló en una tableta Samsung, la cual fue ubicada en un robot y la cámara alineada con la vista del hígado sintético. Dicho montaje se realizó en el laboratorio del grupo nBIO (Grupo de investigación en Neuroingeniería Biomédica) de la Universidad Miguel Hernández en España, sobre un robot de asistencia quirúrgica en una caja de entrenamiento laparoscópica como se puede ver en la figura 4.

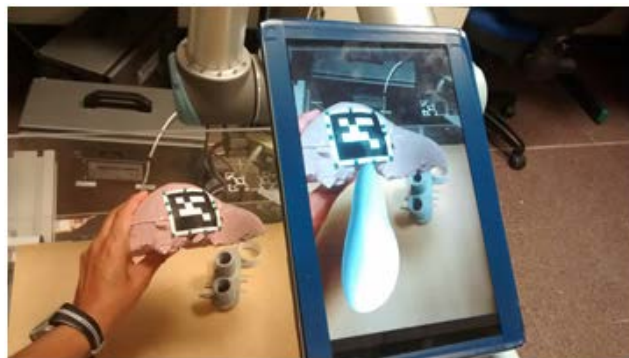


Figura 4 Montaje dispositivo de despliegue y marcador físico ubicado en hígado sintético.

Natural Feature Tracking

NFT (*Natural Feature Tracking Marker*) o marcadores de características naturales no necesitan el borde negro debido a que el reconocimiento del marcador se realiza a partir de características naturales de la imagen, lo cual permite que el marcador pueda ser cualquier porción de la imagen. Para el reconocimiento de características naturales se utilizaron tres herramientas técnicas:

- El *asset* Vuforia para Unity. Se tomó una imagen cualquiera, se usó la herramienta web de Vuforia “Target Manager” para crear una base de datos de las características del marcador. Dicha base de datos es importada en el IDE de Unity, de esta manera se realiza un pre procesamiento anterior a partir de una extracción de características con “*Fast corner detector*” para calcular rápidamente “*feature points*” en la

imagen de la cámara. La herramienta de Vuforia permite categorizar la imagen dependiendo de la facilidad de seguimiento de la misma a partir de la densidad de características encontradas en la imagen. Este tipo de aplicación funciona bien en el proceso de extracción de características naturales con una imagen que se conoce con anticipación, tal como se puede ver en la figura 6. La figura 7 muestra el proyecto en el IDE de Unity haciendo uso del complemento de realidad aumentada de Vuforia; mientras que la figura 8 muestra la aplicación corriendo en tiempo real en un dispositivo móvil con sistema operativo Android.

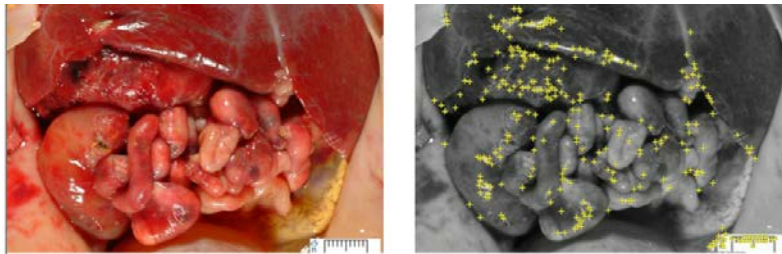


Figura 6 Extracción de características usando Target Manager de Vuforia.

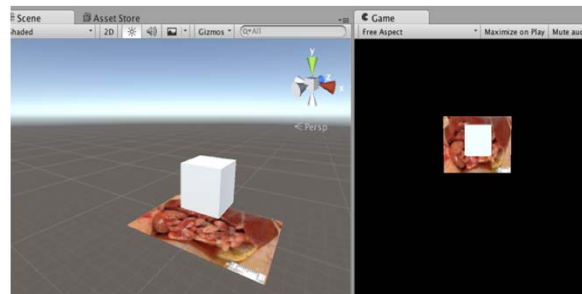


Figura 7 Proyecto en Unity usando el complemento de Vuforia.



Figura 8 Apk usando Vuforia y Android para AR con características naturales.

- La librería OpenCV con Python. Las limitaciones de extracción de características naturales y seguimiento de las mismas por parte de la librería Vuforia llevaron a pensar en el uso de la librería de visión por computador OpenCV que es una de las más ampliamente usadas en el mundo. Existe un complemento de dicha librería para Unity, lo cual resultaba de gran importancia dado que se quiere mantener el beneficio de crear aplicaciones multiplataforma con relativa facilidad. El *asset* de OpenCv para Unity debe ser comprado para su uso. Mientras se cerraba el proceso de adquisición de esa librería, se realizaron pruebas de las capacidades de la librería normal OpenCV para el reconocimiento de características naturales. Dado que el OpenCV clásico es una librería que puede ser usada con varios lenguajes de programación, para este caso particular el lenguaje seleccionado fue Python.

El siguiente paso fue la prueba de extracción de características naturales con la dupla OpenCV – Python como se observa en la figura 9 para lo cual se usó un clasificador Haar Cascade [Viola, 2001] entrenado a partir de imágenes de un objeto particular a ser reconocido. Posteriormente este entrenamiento fue cargado usando la librería OpenCV en el lenguaje de programación Python. Para el entrenamiento del Haar Cascade propio se realizó un proceso de preparación de datos en el cual:

- ✓ Se prepararon imágenes negativas, un banco de imágenes que no contienen el objeto de interés.
- ✓ Se prepararon las imágenes positivas las cuales contienen el objeto de interés.
- ✓ Se realiza la marcación del objeto en las imágenes positivas.
- ✓ Se creó un vector con la información de las imágenes positivas.
- ✓ Se ejecutó el algoritmo de entrenamiento del clasificador Haar Cascade.
- ✓ Se cargó el archivo con los resultados del entrenamiento con OpenCV. 7) Se hizo la prueba con un *script* de Python.

- ✓ Se enlazó con una aplicación de Unity enviando las coordenadas del objeto reconocido desde el programa en Python.

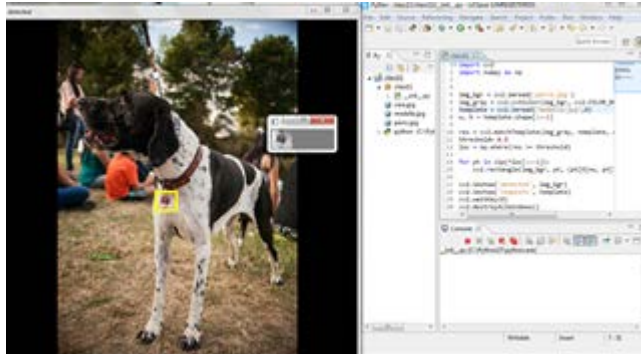


Figura 9 Prueba de extracción características clasificador Haar Cascade OpenCV–Python.

- El asset de OpenCV para Unity. En el caso del asset OpenCV la importación del mismo se realizó copiándolo a la carpeta de complementos del proyecto de Unity. A modo de prueba para verificar el funcionamiento de la librería de visión por computador en una aplicación móvil, se usó detección de círculos con la función `cv2.HoughCircles` como puede apreciarse en figura 10. Se obtuvo como resultado un correcto funcionamiento para los dos ejemplos en una tableta Samsung Galaxy Tab2 y un celular Moto X.



Figura 10 Extracción de características OpenCV–Unity.

Markerless

La técnica “Markerless” o sin marcadores y que también suele llamarse HoloLens, consiste en colocar manualmente un objeto virtual en una vista

particular en la que este parece estar bloqueado en el lugar que ocupa en el espacio.

Este sistema no depende únicamente de sensores (GPS, acelerómetros...) para mantener el objeto en su lugar, es probable que la información de dichos sensores se combine con la técnica SLAM (*Simultaneous Localization and Mapping*), la cual está compuesta por algoritmos que utilizan datos de sensores para construir un mapa de un entorno desconocido, al mismo tiempo que se utiliza para identificar la ubicación espacial. Para que SLAM funcione se necesita crear un mapa preexistente de su entorno y luego orientarse dentro de este mapa. Cuando se coloca el objeto, los algoritmos de Markerless extraen características desde atrás y alrededor del objeto, y utilizan esta información junto con los datos proporcionados por los sensores para ajustar la posición del objeto. Así que colocar objetos sobre un terreno con muchas características funciona mejor que colocarlo en un trozo de papel blanco. Para esta prueba de "Markerless" se utilizó la librería de AR para *markerless* Kudan, la cual se importó como un complemento en el IDE de Unity. En la figura 11 se muestra el anclaje de un elemento de realidad aumentada en un video de una laparoscopia.

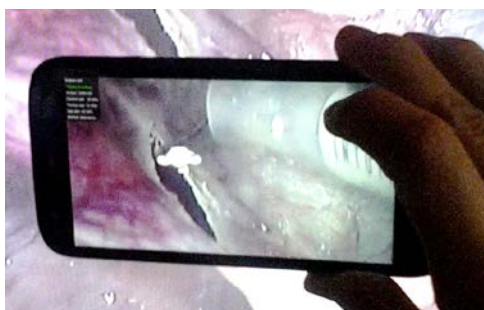


Figura 11 Realidad aumentada, complemento realidad aumentada Kudan para Unity.

4. Discusión

La librería de realidad aumentada Vuforia muestra un adecuado seguimiento de marcadores con características muy estructuradas, aunque existe dificultad en el uso de marcadores naturales en especial si esas características no corresponden a imágenes que se conocen con anticipación, ya que es necesario usar el *target manager* de Vuforia para tener una base de datos de las mismas. Lo anterior

conllevó a explorar otras librerías de AR que permitan la implementación de algoritmos de seguimiento y reconocimiento de marcadores naturales en video, por lo cual se optó por usar OpenCV.

OpenCV se abordó desde dos enfoques de implementación:

- Usando el lenguaje de programación Python.
- Como un complemento del IDE Unity que permita usar *scripts* en C#.

En ambos casos se abre todas las posibilidades para trabajar con clasificadores más especializados como el Haar Cascade el cual es ampliamente usado para reconocimiento de rostro debido a su capacidad de reconocer características naturales incluso si el objeto no es exactamente el mismo, lo cual permite extender su uso al reconocimiento de otros objetos y características. La dupla OpenCV–Python sirve como ejercicio práctico para probar algoritmos y clasificadores. Pero el enlace con Unity solo es posible a través de una comunicación externa, por ejemplo con el uso de *sockets*, lo cual agrega complejidad a la compilación de una aplicación para móviles u HoloLens. Caso contrario se presenta con el *asset* OpenCV para Unity, en el cual la librería es integrada al entorno de desarrollo permitiendo empaquetar una app de muy buen funcionamiento en teléfonos inteligentes de gama media y alta, sin necesidad de comunicaciones externas, lo cual conlleva a deducir que tendrá un comportamiento satisfactorio para HoloLens y demás interfaces de realidad virtual. El objetivo de reconocer y seguir características estructuradas en imágenes en movimiento llevó a la exploración de otro *asset* o complemento para Unity llamado Kudan, el cual presenta como gran ventaja la implementación de *markerless* usando el algoritmo SLAM (*Simultaneous Localization and Mapping*). Las capacidades de esta opción no se han trabajado más allá de realizar una aplicación con ubicación de un objeto en un entorno.

5. Conclusiones

La realidad aumentada es una técnica que puede ser usada como herramienta para facilitar el posicionamiento visual de cirujanos dentro de operaciones laparoscópicas, permitiendo sumar información creada de forma artificial

ubicándola dentro del campo visual que es proporcionado por la cámara endoscópica por lo cual los trabajos realizados en esta área resultan de especial interés debido a una representación más intuitiva.

En este trabajo se realizó una prueba estándar con marcadores cuadrados también llamados “fiducial” o “Border Markers” y marcadores naturales o “Natural Feature Marker” usando librerías comerciales de realidad aumentada y como base el entorno integrado de desarrollo (IDE) Unity y su motor multiplataforma de video juegos. Entre las librerías usadas se citan: Vuforia, Kudan y OpenCV.

Como resultado se tiene que la elección de Vuforia como librería de reconocimiento de marcadores para realidad aumentada y de Unity como IDE de desarrollo resulta ser satisfactoria para marcadores cuadrados y marcadores naturales. El seguimiento del marcador es rápido y la aplicación para sistema operativo android resulta ser muy estable. Uno de los mayores obstáculos en el uso de esta librería radica en la exigencia de extracción previa de las características del marcador, tanto para fiducial markers como para natural marker. Labor que se realiza usando una aplicación web proporcionada por los fabricantes de Vuforia y que es llamada “target manager”. Dicha aplicación entrega una base de datos de las características del marcador la cual es importada en Unity y que sin duda proporciona un reconocimiento rápido del marcador. Ante lo cual se propuso el uso de otra librería que permita mayor flexibilidad en el manejo de los algoritmos de visión por computador. El paso siguiente se enfocó al uso del asset OpenCV para Unity. Este asset al estar integrado a Unity permite compilar aplicaciones multiplataforma de muy buen funcionamiento en tiempo real.

En el caso de Markerless se trabajó con Kudan un complemento para Unity que trabaja bajo el algoritmo SLAM. Aunque las aplicaciones con Kudan resultar tener la misma facilidad que con Vuforia, el Markerless es altamente dependiente del posicionamiento con GPS, es de recalcar que no se descarta el uso de esta librería para trabajos próximos.

Dado que el trabajo actual estaba enfocado a la exploración de herramientas de realidad aumentada y a pruebas con *border markers* y de manera no muy profunda con *markerless*, una de las principales direcciones para trabajos futuros

es el uso de algoritmos más robustos tales algunas variaciones de SLAM, o algoritmos mixtos que vayan más allá, como SURF con KLT, tal como se propone en [Puerto, 2013]. [Collins, 2016], reajustando la metodología de reconocimiento de los marcadores naturales teniendo en cuenta obstáculos presentes en las cirugías laparoscópicas tales como oclusión y problemas de iluminación.

6. Bibliografía y Referencias

- [1] Autorino, R., Cadeddu, J. A., Desai, M. M., Gettman, M., Gill, I. S., Kavoussi, L. R., Lima, E., Montorsi, F., Richstone, L., Stolzenburg, J. U., Laparoendoscopic single-site and natural orifice transluminal endoscopic surgery in urology: a critical analysis of the literature. *European Urology* 59, 1, pp. 26-45, 2011.
- [2] Bae, J. H., Development of smart game based on multi-platform game engine. *International Journal of Multimedia and Ubiquitous Engineering* 11, 3, pp. 345-350, 2016.
- [3] Bruellmann, D. D., H. Tjaden, U. Schwanecke, and P. Barth, An optimized video system for augmented reality in endodontics: a feasibility study, *Clinical oral investigations* 17, no. 2, pp. 441-448, 2013.
- [4] Collins, T., Bartoli, A., Bourdel, N., & Canis, M., Robust, real-time, dense and deformable 3d organ tracking in laparoscopic videos. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer International Publishing, pp. 404-412, 2016.
- [5] Cukovic, S., Gattullo, M., Pankratz, F., Devedzic, G., Carrabba, E., Baizid, K. Marker based vs. natural feature tracking augmented reality visualization of the 3D foot phantom. *Electrical and Bio-medical Engineering, Clean Energy and Green Computing* 1, 2016.
- [6] Cristie, V., Berger, M., Bus, P., Kumar, A., Klein, B., City-heat: visualizing cellular automata-based traffic heat in unity3d, *Visualization in High Performance Computing*, ACM, 6, SIG-GRAPH, Asia 2015.
- [7] De Lacy, A. M., Rattner, D. W., Adelsdorfer, C., Tasende, M. M., Fernández, M., Delgado, S., Sylla, P., Martínez-Palli, G., Transanal natural orifice transluminal endoscopic surgery (notes) rectal resection: "down-to-up" total

- mesorectal excision (tme)—short-term outcomes in the first 20 cases. *Surgical Endoscopy* 27, 9, pp. 3165–3172, 2013.
- [8] De Paolis, L. T., Ricciardi, F., Dragoni, A. F., Aloisio, G. An augmented reality application for the radio frequency ablation of the liver tumors. In *International Conference on Computational Science and Its Applications*, Springer, Berlin, Heidelberg, pp. 572-581, 2011.
- [9] Duchene, D. A., Moinzadeh, A., Gill, I. S., Clayman, R. V., and Winfield, H. N., Survey of residency training in laparoscopic and robotic surgery. *The Journal of Urology* 176, pp. 2158–2167, 2016.
- [10] Llena, C., S. Folguera, L. Forner, and F. J. Rodríguez Lozano, Implementation of augmented reality in operative dentistry learning. *European Journal of Dental Education*, 2017.
- [11] Furst, J., Fierro, G., Bonnet, P., Culler, D. E., Busico 3d: building simulation and control in unity 3d. *12th ACM Conference on Embedded Network Sensor Systems*, ACM, pp. 326–327, 2014.
- [12] Gutiérrez Puerto, E.M., Sistema de Realidad Aumentada para la interacción con el Instrumental en el procedimiento de Acceso Venoso Central, Tesis, Universidad Militar Nueva Granada, 2015.
- [13] Jamali, S. S., Shiratuddin, M. F., Wong, K. W., Oskam, C. L., Utilising mobile-augmented reality for learning human anatomy. *Social and Behavioral Sciences* 197, pp. 659–668, 2015.
- [14] Kipper, G., Rampolla, J., *Augmented Reality: an emerging technologies guide to AR*. Elsevier, 2012.
- [15] Mahmoud, N., Hostettler, A., Collins, T., Soler, L., Doignon, C., & Montiel, J. M. M. SLAM based Quasi Dense Reconstruction For Minimally Invasive Surgery Scenes, 2017.
- [16] Pauly, O., Diotte, B., Fallavollita, P., Weidert, S., Euler, E., & Navab, N. Machine learning-based augmented reality for improved surgical scene understanding. *Computerized Medical Imaging and Graphics*, pp. 55-60, 2015.

- [17] Puerto-Souza, G. A., & Mariottini, G. L. An augmented-reality system for laparoscopic surgery robust to complete occlusions and fast camera motions, In International Conference on Robotics and Automotion, 2013.
- [18] Marescaux, J., Diana, M., Next step in minimally invasive surgery: hybrid image-guided surgery. *Journal of pediatric surgery* 50, pp. 30–36, 2015.
- [19] Moreno, M. R., Moraes, T. F., Amorim, P. H., da Silva, J. V. L., Rodriguez, C. A. Virtual open source environment for training and simulation of laparoscopic surgery. XII Work-shop de Informática Médica (WIM2012) XXXII Congresso da Sociedade Brasileira de Computacao, pp. 1–4, 2012.
- [20] Narahara, T., Abbruzzese, K. M., Foulds, R. A. Haptic collaboration: biomedical engineering meets digital design, SIGGRAPH 2015, 2015.
- [21] Peng, H., Application research on face detection technology based onopencv in mobile augmented reality. *International Journal of Signal Processing, Image Processing and Pattern Recognition* 8, pp. 249–256, 2015.
- [22] Rané, A., Rao, P., and Rao, P., Single-port access nephrectomy and other laparoscopic urologic procedures using a novel laparoscopic port (r-port). *Urology* 72, pp. 260–263, 2008.
- [23] Viola, P., Jones, M. Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition. CVPR 2001. IEEE Computer Society Conference*, 2010.
- [24] Simonetti, A., Paredes, J. Vuforia v1. 5 sdk, Analysis and evaluation of capabilities. Master's thesis, Universitat Politècnica de Catalunya, 2013.
- [25] Soto, C., Vargas, M., Uribe-Quevedo, A., Jaimes, N., and Kapralos, B. Ar stereoscopic 3d human eye examination app. In *Interactive Mobile Communication Technologies and Learning (IMCL)*, 2015 International Conference, IEEE, pp. 236–238, 2015.
- [26] Van Veelen, M., Nederlof, E., Goossens, R., Schot, C., Jakimowicz, J., Ergonomic problems encountered by the medical team related to products used for minimally invasive surgery. *Surgical Endoscopy and Other Interventional Techniques* 17, pp. 1077–1081, 2003.

DETECCIÓN DE NECESIDADES Y DEFINICIÓN DE CONTENIDOS PARA LA ENSEÑANZA DE LA METODOLOGÍA DEL PERIODISMO DE DATOS: EL CASO DE DATAÍSTA

Belén Alazañez Cortés

Universidad Autónoma Metropolitana Unidad Cuajimalpa
belenalazagnez@gmail.com

Zayra Monserrat Miranda Aguirre

Universidad Autónoma Metropolitana Unidad Cuajimalpa
zayry08@gmail.com

Jocelyn Lizbeth Molina Barradas

Universidad Autónoma Metropolitana Unidad Cuajimalpa
d9molina@gmail.com

Rocío Abascal Mena

Universidad Autónoma Metropolitana Unidad Cuajimalpa
mabascal@correo.cua.uam.mx

Resumen

En este artículo se presenta un sistema de información, llamado Dataísta, capaz de llevar a cabo un proceso de enseñanza-aprendizaje pertinente para los periodistas de investigación mexicanos que quieren incursionar en el Periodismo de Datos. Este sistema tiene como función principal proveer conceptos y herramientas prácticas que le permitan integrar el razonamiento detrás de la metodología del Periodismo de Datos a su propio conocimiento, sin ser expertos en computación. Aplicando procesos del Diseño Instruccional, así como de Diseño Centrado en el Usuario (DCU), se identificaron las necesidades de los usuarios y se definieron los contenidos y formatos en los que serían presentados los

contenidos. Asimismo, se desarrollaron y evaluaron prototipos para conocer la relevancia y efectividad de Dataísta para lograr enseñar las cuatro fases que conforman la metodología (obtención, limpieza o filtrado, análisis y visualización). El sistema es el primero en su tipo para la formación de periodistas.

Palabras Claves: Diseño Centrado en el Usuario, diseño Instruccional, entorno de aprendizaje, periodismo de datos, sistema de información.

Abstract

This article presents an information system, called Dataísta, able to carry out a relevant teaching-learning process for Mexican investigative journalists who want to learn about the Data Journalism. Its main function is to provide concepts and practical skills that allow them to integrate the reasoning behind the methodology of Data Journalism into its own knowledge without being computer experts. Applying processes of Instructional Design, as well as User-Centered Design (DCU), we have identified the needs of users and defined the contents and formats in which to present the information. Likewise, prototypes were developed and evaluated to know the relevance of Dataísta and the effectiveness of the system in order to teach the four phases of the methodology (obtaining, cleaning or filtering, analysis and visualization). The system is the first of its kind in the training of journalists.

Keywords: *Data journalism, information system, instructional design, learning environment, user centered-design.*

1. Introducción

Durante los últimos años, el periodismo ha enfrentado retos significativos, relacionados, principalmente, con el auge de las nuevas tecnologías. La inclusión de las Tecnologías de Información y Comunicación (TICs) en el mundo, ha transformado la forma en la que se ejerce el periodismo. De esta manera, en la actualidad, los periodistas tienen acceso a una gran variedad de fuentes de información, como las bases de datos publicadas por los gobiernos, reportes de Organizaciones no Gubernamentales (ONGs) e incluso *posts* y *tuits* provenientes

de medios sociales como Twitter o Facebook. A las fuentes de información mencionadas anteriormente, se les conoce como Datos Abiertos y, generalmente, son utilizados por los periodistas para iniciar sus investigaciones. El Open Data Handbook define a los Datos Abiertos como *“datos que pueden ser utilizados, reutilizados y redistribuidos de forma libre por cualquiera”* [Poikola, 2009]. Los Datos Abiertos deben estar disponibles, preferentemente vía Internet, y no deben ser fragmentados.

El acceso a diversas fuentes de información puede ser de gran ayuda para el periodista, la realidad es que se requiere de cierto tipo de especialización, generalmente en áreas técnicas o en computación, para ser capaz de procesar, analizar y visualizar toda la información que de ellas se obtiene. Es importante destacar que, debido a la cada vez mayor apertura de la información, los datos ya no son usados únicamente para hacer más atractivos los artículos periodísticos, ahora, toman un rol protagónico pues sirven de apoyo para encontrar historias en las bases de datos y sustentar las notas periodísticas. La mezcla entre la forma tradicional de hacer periodismo y el avance tecnológico ha dado pie al desarrollo de una nueva especialización del periodismo, denominada “Periodismo de Datos”. Con el fin de comprender qué es el Periodismo de Datos, se puede retomar la definición de Aron Pilhofer, quien dice que se trata de *“una disciplina cada vez más amplia, que usa herramientas, técnicas y enfoques para contar historias, e incluye desde las tradicionales técnicas del Periodismo Asistido por Computadora hasta las más novedosas aplicaciones de visualización. Con la finalidad de proporcionar información y análisis al público para acercar los temas más importantes del día a día”* [Flores, 2013]. Esta definición describe la forma en la que el periodismo hace uso de las nuevas tecnologías para hacer frente a la era digital. Es importante destacar que, el Periodismo de Datos no reemplaza la metodología seguida por los periodistas durante décadas, simplemente agrega nuevas técnicas que dotan a los periodistas de la posibilidad de acercarse a otras disciplinas con el propósito de lograr investigaciones más completas apoyándose del trabajo interdisciplinario.

La transición del periodismo tradicional al Periodismo de Datos se ha dado de manera paulatina, principalmente, porque los periodistas no cuentan con el tiempo, los conocimientos o la infraestructura necesaria para iniciar con su especialización [Zanchelli, 2013]. De acuerdo con el Global Investigative Journalism Network, el Periodismo de Datos se ha vuelto relevante principalmente porque *“gobiernos y empresas han incrementado el flujo de información y alrededor del mundo se tiene un mayor acceso a una gran cantidad de datos”* [Zanchelli, 2013]. Debido a lo anterior, es importante destacar que existe un nicho de oportunidad en la generación de herramientas y habilidades relacionadas con el Periodismo de Datos para el usuario hispano parlante. Muchas de las herramientas existentes no toman en cuenta necesidades propias, por ejemplo, del periodista mexicano.

El artículo está organizado de la siguiente manera: en la Sección 2, se introduce a un estado del arte de los manuales, artículos y herramientas que fueron creados para introducir al periodista a la metodología del Periodismo de Datos. De igual manera, la Sección 2, presenta el estudio contextual, la explicación del proceso de trabajo y la metodología seguida para desarrollar los prototipos, así como su evaluación con el usuario final: el periodista mexicano. En la Sección 3 se demuestra la relevancia de la propuesta y su contribución al proceso de enseñanza-aprendizaje de esta nueva especialización del periodismo a través de la presentación del prototipo y su funcionamiento. Finalmente, se aborda el trabajo futuro y las conclusiones en la Sección 4.

2. Métodos

Como se mencionó en la sección de Introducción, existe una nueva rama del periodismo la cual es *“una disciplina periodística que se nutre de otras muchas: de investigación, en profundidad, de precisión, asistido por computadora y analítico. En ella se trabaja con grandes volúmenes de datos, se aprovecha al máximo la visualización interactiva y se incorpora al programador al equipo periodístico”* [Crucianelli, 2013]. Sin embargo, el periodista actual no está preparado para dar el gran salto hacia la tecnología y, mucho menos, cuando cada una de las etapas del

Periodismo de Datos requiere de habilidades precisas. Por ejemplo, los periodistas deben buscar los datos y limpiarlos, aunque lamentablemente las bases de datos que proporciona el gobierno o que se encuentran en la red, están ‘sucias’ es decir, contienen espacios innecesarios, letras o símbolos que se repiten y esto hace que el análisis de los datos no sea óptimo. Para ello, es necesario realizar una adecuada limpieza de las bases de datos para poder avanzar al siguiente paso que es el análisis. En esta fase se pueden encontrar patrones que no se habían considerado antes y que pueden darle un giro a la investigación. Otra de las etapas del Periodismo de Datos es la visualización la cual es la presentación de todos los pasos anteriores, enfocándose en dotar a los lectores de la información pertinente y que puedan identificar claramente los puntos claves del proyecto, así como que lo entregado tenga un impacto relevante en la sociedad. Es así como el periodista tradicional se enfrenta a una serie de tareas para las que no estaba habilitado y que incluyen las etapas presentadas en la figura 1.



Figura 1 Metodología del Periodismo de Datos (Creación Propia).

Ante la necesidad de metodologías más abiertas, que ofrezcan herramientas para construir el proceso propio de enseñanza-aprendizaje, existen algunas opciones complementarias y/o alternativas al aprendizaje formal. El periodista puede buscar y encontrar dentro de la red una lista muy larga de herramientas tecnológicas que pueden ayudarle en su labor. Desde, por ejemplo, conocer tendencias en Internet como Google Trends¹ hasta para mapear información geográfica como CartoDB². Dentro del periodismo de datos también existe una gama muy importante de herramientas que son un recurso vital para la sistematización de datos. Para la

¹ <https://trends.google.com.mx/trends/>

² <https://carto.com/>

obtención de datos (Google Trends, Youtube³), limpieza de datos (Open Refine⁴, Tabula⁵), análisis de datos (R⁶, Tableau⁷) y visualización de datos (D3.js⁸, My Maps⁹). Además de la extensa variedad de herramientas auxiliares disponibles en la red, también están disponibles algunos cursos masivos gratuitos en línea (MOOCs, por sus siglas en inglés). Desafortunadamente, la gran mayoría están pensados en generar científicos de datos, más que en preparar a periodistas de datos. Por lo tanto, el objetivo del presente artículo se enmarca dentro de la siguiente hipótesis:

Ho: Los periodistas de investigación carecen de los conocimientos necesarios para realizar artículos basados en datos. Por ello, es importante identificar sus necesidades para generar contenidos que se puedan incluir en un sistema de información que les ayude a comprender la metodología propuesta por el Periodismo de Datos.

Para la propuesta de un sistema de información se utilizaron dos metodologías: Diseño Centrado en el Usuario y el Diseño Instruccional, se explican a continuación.

Diseño Centrado en el Usuario (DCU)

Una vez que se identificaron las diferentes herramientas y cursos disponibles, se realizó un estudio contextual aplicando la metodología del Diseño Centrado en el Usuario (DCU) con el objetivo de identificar las necesidades de los periodistas de investigación; es decir, los usuarios finales. Mediante la realización de observación participativa, entrevistas y encuestas, fue posible definir las características que debían estar incluidas en un entorno de aprendizaje que permitiera a los periodistas de investigación incursionar en la metodología del Periodismo de Datos.

³ <https://www.youtube.com/>

⁴ <http://openrefine.org/>

⁵ <http://tabula.technology/>

⁶ <https://www.r-project.org/>

⁷ <https://www.tableau.com/es-es>

⁸ <https://d3js.org/>

⁹ <https://www.google.com/maps/d/>

Se eligió aplicar la metodología del DCU debido a que ayuda en la obtención de información precisa que permitirá, posteriormente, el desarrollo de un sistema de información mucho más preciso. En el caso del sistema propuesto, la presencia del usuario, el periodista de investigación fue crucial para identificar sus necesidades y así producir un prototipo adecuado a sus necesidades. Con el fin de identificarlas, se decidió entrevistar a un grupo de profesionistas de datos mexicanos, bastante reconocidos dentro de su comunidad. Todos ellos, trece en total, forman parte de equipos interdisciplinarios que realizan periodismo de datos, tabla 1.

Tabla 1 Profesionales del periodismo de datos entrevistados.

No.	Nombre	Medio/Proyecto	Perfil Profesional
1	Israel Piña	Quién Compró	Periodista
2	Tania Montalvo	Animal Político	Periodista
3	Lilia Saúl	Universal Data	Periodista
4	Daniela Guazo	Universal Data	Periodista
5	Yosune Chamizo	Animal Político	Diseñadora de Información
6	David Hernández	UAM Cuajimalpa	Diseñador de Información
7	Gilberto León	Animal Político	Programador
8	Sergio Araiza	SocialTic	Programador
9	Irving Morales	Morlan	Programador
10	Emmanuel Landa	Morlan	Programador
11	Eduard Martín Borregón	Poder	Periodista
12	Susana Zabala	Independiente	Periodista
13	Daniel Lizárraga	Mexicanos Contra la Corrupción	Periodista

Las entrevistas arrojaron que la principal razón por la que existen escasos ejemplos de Periodismo de Datos en México es la falta de conocimientos en estadística y computación por parte de los periodistas de investigación. Esto les dificulta apropiarse de la metodología propuesta por el Periodismo de Datos, impidiéndoles incluirla en su trabajo diario. Este hallazgo, permite, además, definir al periodista de investigación, como el principal usuario.

Una vez que se definió al usuario, se buscó a un grupo de periodistas de investigación abierto a colaborar en la investigación. Se les aplicó una encuesta,

para conocer qué tanto sabían sobre Periodismo de Datos, qué tantas partes de la metodología conocían o habían aplicado en su trabajo y, finalmente, los principales problemas que enfrentaron al intentar aplicarla. Además, se les consultó sobre las fases de la metodología que les presentaban mayor complicación. La figura 2 muestra los términos más utilizados al contestar la encuesta haciendo énfasis en las principales barreras: cada una de las fases de la metodología del Periodismo de Datos.



Figura 2 Principales necesidades de los periodistas de investigación.

Además, se identificaron las 4 principales dificultades al hacer uso de las herramientas y cursos disponibles en la red:

- Contar con un equipo de cómputo y conexión a Internet.
- El idioma, la mayoría solo están disponibles en el idioma inglés.
- Requiere conocimientos específicos de disciplinas ajenas a la suya.
- No todas son gratuitas.

El cruce de la información sobre necesidades y barreras posibilitó definir que los contenidos del sistema deberían estar enfocados al entendimiento de la metodología del Periodismo de Datos y el proceso de trabajo a seguir cuando se quiere escribir un artículo basado en datos.

Diseño Instruccional

El desarrollo de las nuevas Tecnologías de la Comunicación e Información (TICs), el auge de las concepciones cognoscitivas y las necesidades pedagógicas de los últimos años han propiciado un desarrollo acelerado de nuevas formas de

interacción en el proceso instruccional que permite la integración de esas nuevas tecnologías al ámbito didáctico con una perspectiva amplia y con mayor eficiencia en lo concerniente al mejoramiento del aprendizaje [Córdova, 2002].

El Diseño Instruccional es necesario en cualquier modalidad para organizar de una manera sistemática, no solo la enseñanza, sino también el aprendizaje. Los profundos cambios que se han producido a raíz de los avances tecnológicos no dejan de lado la forma como se ha venido diseñando la instrucción; es por ello que los modelos instruccionales de hoy se caracterizan por ser procesos integrales y holísticos, dialécticos, creativos y flexibles [Romero, 2016]. Desde el principio, el Diseño Instruccional ha tenido una gran influencia en las TICs, así que la etiqueta de *tecnología instruccional* es un sinónimo [Seel et al., 2017]. De esta manera, existen algunos trabajos como el de [Balzhiser, 2015] en el que también se ve implícito el diseño participativo a la hora de construir una herramienta cuyo principal propósito es la de instruir mediante el uso de la tecnología.

Como se explicó en la Sección 1, en este caso el DCU cobra relevancia en el Diseño Instruccional al estar mediado por tecnologías. El usuario es muy importante y forma parte de todo el proceso de construcción de la solución. Así, tomando lo anterior como base, se puede decir que la función principal del sistema, es preparar al periodista de investigación en la aplicación de sus conocimientos a situaciones reales, nuevas y cambiantes; reforzando el aprendizaje significativo. Éste es el que se puede incorporar a las estructuras de conocimiento que tiene y que adquirirá significado a partir de la relación con conocimientos previos. De igual forma, es el periodista el que guía, a partir de sus propias lagunas de información, los temas y herramientas que son relevantes a aprender. El usuario está presente en el desarrollo del sistema permitiendo conocer la manera en la que se puede descomponer a la metodología del Periodismo de Datos en pequeños pasos o instrucciones.

Para la propuesta de un prototipo, se ha considerado que, en todas las modalidades de recursos didácticos en línea personalizados, los contenidos deben de tomar en cuenta lo siguiente:

- La organización de las temáticas como eje principal de acuerdo a la forma en la que aprenden los usuarios.
- Posibilidad de ingresar las preferencias de los usuarios. Es decir, el usuario puede guardar su perfil para retomar temas en el momento en que lo decida.
- Retroalimentación por parte del sistema o por parte de los participantes. En este caso, es importante contemplar la interacción entre la comunidad para potencializar la transmisión de conocimientos.

La aplicación de conceptos del Diseño Instruccional, junto con los puntos anteriores, a los contenidos del sistema permitirá al usuario aprovechar la información que en ellos se aborda. Con ello, el prototipo permitirá facilitar el entendimiento del razonamiento detrás de la metodología que se desea que los periodistas de investigación comprendan.

3. Resultados

Basados en la revisión bibliográfica, así como en la información proporcionada por los usuarios, se identificaron cuáles son la mayoría de los problemas que los periodistas de investigación enfrentan cuando tratan de aplicar la metodología propuesta por el Periodismo de Datos en sus investigaciones. Estos problemas están, principalmente, enmarcados dentro de la carencia de conocimientos previos en otras áreas, así como poco o nulo conocimiento acerca de la metodología propia del Periodismo de Datos. En la hipótesis, se plantea probar que la falta de sustento teórico, así como la falta de conocimientos sobre procesamiento de datos por parte de los periodistas de investigación, son los principales obstáculos por los cuales los periodistas no utilizan los Datos Abiertos en sus investigaciones. Por ello, la creación de un Sistema de Información basado en sus necesidades, les permitirá aprender la metodología del Periodismo de Datos, para así, ser capaces de contar historias a través de los datos.

Organización de temáticas: El prototipo contiene conceptos básicos sobre cada paso de la metodología, ver figura 3.



Figura 3 Vista del menú principal del sistema.

Mediante el uso de videos infográficos, el usuario puede identificar la forma en la que todas las etapas de la metodología están conectadas, figura 4. Algunos temas, contienen infografías con el paso a paso sobre cómo realizar cierta actividad, por ejemplo, el *scraping*. Además, cada parte de la metodología cuenta con su propia herramienta auxiliar, misma que servirá al usuario para practicar los conocimientos y habilidades adquiridos dentro del sistema. Asimismo, el usuario es capaz de regresar a cualquiera de las etapas que conforman la metodología, así como saltar las que considera que ya domina. Esto le permite al usuario ubicarse dentro del proceso.

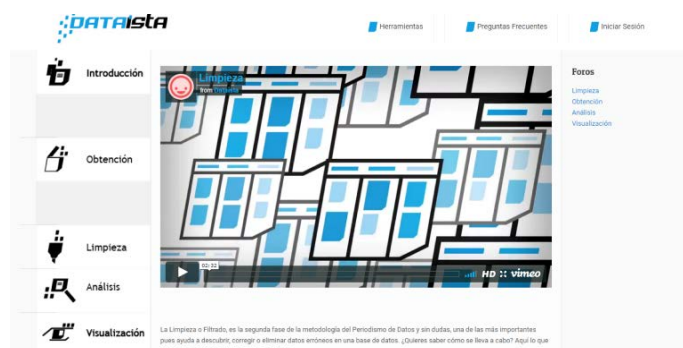


Figura 4 Vista de la sección de limpieza, video introductorio.

Preferencias y perfil del usuario: el prototipo permite logearse con el fin de contar con un sistema capaz de adaptarse al perfil del usuario. En todo momento, el usuario será capaz de guardar el estatus de su recorrido por la metodología, así como interactuar con la comunidad.

Uno de los hallazgos dentro de la investigación, fue el saber que los usuarios querían conocer más herramientas en línea que pudieran simplificar sus labores.

Por ello, dentro de cada etapa de la metodología se incluye un listado de aquellas herramientas más usadas por los Periodistas de Datos, figura 5. Los usuarios logeados pueden calificar las herramientas existentes.

Nombre	Link	Evaluación
OPEN REFINE	http://openrefine.org/	✓✓✓
EXCEL	https://products.office.com/en-us/excel	✓
GOOGLE FUSION	https://fusiontables.google.com	✓✓✓
DATA WRANGLER	http://uris.stanford.edu/wrangler	✓✓✓
TABLEAU PUBLIC	https://public.tableau.com/v/	✓

Herramienta Auxiliar

LIMPIEDATA

Figura 5 Vista de la sección de limpieza, listado de herramientas y herramienta auxiliar.

Retroalimentación e interacción con la comunidad: el prototipo cuenta, además, con un foro donde el usuario puede escribir sus dudas e interactuar con otros usuarios del sistema (ver figura 6). Se llegó a esta decisión porque durante la observación participativa, se identificó que esta comunidad es pequeña y comprometida con ayudar a otros durante su camino a la especialización.

Search forum

Este foro contiene 2 debates y lo actualizó [Zayra Monserrat Miranda Aguirre](#) hace 1 mes, 2 semanas.

Viendo 2 debates - del 1 al 2 (de un total de 2)

Debate	Usuarios	Publicaciones	Último mensaje
Búsquedas avanzadas Iniciado por Zayra Monserrat Miranda Aguirre	1	1	Zayra Monserrat Miranda Aguirre hace 1 mes, 2 semanas
Uso de OpenRefine Iniciado por Erick Money Cuevas	1	1	Erick Money Cuevas hace 1 mes, 2 semanas

Viendo 2 debates - del 1 al 2 (de un total de 2)

Debes estar registrado para crear debates nuevos.

Figura 6 Vista del foro del sistema.

Además, es posible compartir la evaluación de herramientas con el resto de la comunidad. Esto permite el intercambio de información y que el proceso de enseñanza-aprendizaje fluya de manera colaborativa.

4. Discusión

Una de las características esenciales de Dataísta es el uso de videos infográficos que permiten comunicar y generar un proceso de enseñanza-aprendizaje no solo auditivo sino también visual. Los videos contienen características importantes para poder lograr objetivos importantes en la comprensión y aprendizaje de los contenidos. Se diseñaron a partir de un guion que permitió lo siguiente:

- Clarificar ideas.
- Diseñar con un propósito.
- Sintetizar en pocos minutos el tema.
- Utilizar un lenguaje claro y conciso.
- Comunicar.

Así, se considera que una vez que el periodista de investigación pone en marcha el video es capaz de entender en pocos minutos cada una de las etapas de una forma lúdica.

Dataísta es un sistema basado en el Diseño Instruccional que incorpora a sus usuarios desde la concepción hasta la implementación del mismo detectando necesidades claras del periodista de investigación en su aprendizaje sobre la metodología del Periodismo de Datos. La innovación de Dataísta reside en la carencia de un sistema igual que introduzca al periodista de investigación en temas que considera complicados debido a sus características propias al provenir de disciplinas ajenas. La concepción de un sistema como Dataísta para el proceso de enseñanza-aprendizaje para usuarios con diferentes conocimientos previos y habilidades es un gran paso al lograr la introducción de estos y la puesta en marcha de trabajos especializados en el Periodismo de Datos.

Al estar enmarcada Dataísta dentro de un proceso de DCU ha sido evaluada y seguida de cerca por periodistas interesados en que ya se libere para que puedan utilizarla. Por el momento, la evaluación de Dataísta ha permitido mejorar aspectos de usabilidad que permitirán que el sistema sea más fácil de usar por el usuario no experto en aspectos tecnológicos.

5. Conclusiones

Los periodistas de investigación no cuentan con los conocimientos necesarios para hacer uso de los Datos Abiertos que tanto el gobierno mexicano, como otras instituciones ponen a disposición de todos. Aunque estudian por su cuenta, consultan libros y acceden a las herramientas disponibles en la red, no pueden llegar al entendimiento de la metodología del Periodismo de Datos, porque desconocen conceptos básicos relacionados con otras disciplinas que también la nutren, como lo son el Diseño y la Computación. Además, la falta de tiempo y el nulo apoyo de las redacciones donde laboran, impiden que aquellos que están interesados en esta especialización puedan hacerlo. Por ello, el desarrollo de Dataísta pensado en las necesidades específicas de los periodistas mexicanos, reducirá la barrera relacionada con la formación académica, hasta que las instituciones educativas y los medios de comunicación se den cuenta de la importancia que tienen en invertir en esta nueva forma de hacer periodismo.

La contribución de Dataísta será medida en los próximos meses al evaluar a los usuarios del sistema y su contribución en medios nacionales con notas e investigaciones que incorporan información proveniente de bases de datos abiertas cuando antes no lo hacían por no saber cómo hacerlo.

El trabajo a futuro requiere de la automatización de búsquedas y categorías para las temáticas que se generan en el foro con el fin de utilizar la participación de los usuarios como otro insumo para el periodista de investigación. Esto será de gran ayuda ya que es el mismo usuario quien podrá ser el generador de contenidos para Dataísta.

6. Bibliografía y Referencias

- [1] Balzhiser, D., Sawyer, P., Womack-Smith, S., & Smith, J. A. Participatory design research for curriculum development of graduate programs for workplace professionals. *Programmatic Perspectives*, 7(2), pp. 79–133, 2015.
- [2] Córdova, D. El Diseño Instruccional: Dos Tendencias y una Transición Esperada. *Docencia Universitaria*, III, 2002.

- [3] Crucianelli, S. ¿Qué es el Periodismo de Datos? *Cuadernos para Periodistas*, 106-124, 2013.
- [4] Flores, J. V., & Salinas, C. A. El periodismo de datos como especialización de las organizaciones de noticias en Internet. *Correspondencias y Análisis*, pp. 15-34, 2013.
- [5] Poikola, A., Villum, C., Dietrich, D., Gray, J., Tait, J., Rogers, K., Zijlstra, T. *The Open Data Handbook*: <http://opendatahandbook.org/>. [01/06/2017], 2009.
- [6] Romero, N. *Manual de Diseño Instruccional: Una Propuesta con Tareas Integradoras*. Digital UNID, 2016.
- [7] Seel, N. M., Lehmann, T., Blumschein, P., & Podolskiy, O. A. Research-Based Instructional Design. *Instructional Design for Learning* (pp. 109-175). Sense Publishers, 2017.
- [8] Zanchelli, M., Crucianelli, S. Integrando el Periodismo de Datos en las Salas de Redacción. *Knight International Center for Journalists*, Washington 2013.

METODOLOGÍA PARA LA IMPLEMENTACIÓN DEL MPM EN VHDL Y LA EMULACIÓN DE AMPLIFICADORES DE POTENCIA EN UNA TARJETA FPGA

Edgar Allende Chávez

Tecnológico Nacional de México/Instituto Tecnológico de Tijuana
edgar.allende@tectijuana.edu.mx

José Ricardo Cárdenas Valdez

Tecnológico Nacional de México/Instituto Tecnológico de Tijuana
jose.cardenas@tectijuana.edu.mx

José Alejandro Galaviz Aguilar

Instituto Politécnico Nacional, IPN-CITEDI
jgalaviz@citedi.mx

Andrés Calvillo Téllez

Instituto Politécnico Nacional, IPN-CITEDI
calvillo@citedi.mx

José Cruz Núñez Pérez

Instituto Politécnico Nacional, IPN-CITEDI
nunez@citedi.mx

Resumen

El presente trabajo muestra el diseño e implementación en VHDL del modelo polinomial con memoria que fue seleccionado para la emulación del comportamiento de amplificadores de potencia con el propósito de proporcionar una plataforma de pruebas y evaluación para el modelado matemático y su posterior uso en pre-distorsión digital. Las mediciones de un amplificador real modelo NXP de 10 W medido a 2 GHz se utilizaron para la obtención del modelo matemático el cual fue implementado en una tarjeta de evaluación y desarrollo

DSP-FPGA Altera Stratix III. Además el artículo describe el desarrollo de un conjunto de funciones, para la manipulación de números complejos, necesario para la implementación del modelo. Los resultados muestran un desempeño adecuado del modelo en VHDL el cual es capaz de emular las curvas de distorsión en amplitud y fase AM-AM y AM-PM. Finalmente a modo de validación la implementación se compara con una simulación en Matlab.

Palabras Claves: Amplificador de potencia, emulación, FPGA, modelo polinomial con memoria, VHDL.

Abstract

This paper shows the design and implementation in VHDL of the memory polynomial model which was selected for emulating the behavior of power amplifiers with the purpose of providing a test and evaluation test bed for mathematical modeling and its later use in digital predistortion. Measurements of a real power amplifier model NXP 10W at 2 GHz were used for obtaining the mathematical model which was implemented in the DSP-FPGA development kit, Stratix III Edition by Altera. This paper also describes the development of a function set for complex numbers manipulation which is needed for the implementation of the model. Results show a correct performance of the VHDL model which can emulate distortion curves for amplitude and phase AM-AM and AM-PM. Finally a comparison is done between VHDL model and Matlab simulation.

Keywords: Emulation, FPGA, Memory polynomial model, Power amplifier, VHDL.

1. Introducción

El amplificador de potencia para radiofrecuencia (RF-PA) es el circuito electrónico de entrada que permite incrementar el nivel de potencia de la señal RF que se desea enviar, antes de que la misma llegue a la antena. Lo anterior con el fin de que la señal sea transmitida y sobrepase la sensibilidad del receptor, lo que garantiza la demodulación de la información [Núñez, 2014]. Sin embargo el RF-PA es un elemento inherentemente no lineal sobre todo cuando se trabaja en la zona

de saturación y la amplificación lograda no corresponde a una ganancia lineal de la señal de entrada [Wood, 2016]. Aunado a esto, se han desarrollado nuevos esquemas de modulación que buscan hacer un uso más eficiente del ancho de banda disponible tales como acceso múltiple por división de código de banda ancha (WCDMA) y multiplexación por división de frecuencias ortogonales (OFDM) las señales resultantes de los anteriores esquemas tienen una envolvente no constante y un factor de potencia pico-promedio (PAPR) sumamente grande generalmente de 10 dB [Roblin, 2013]. Este tipo de multiplexaciones digitales lleva a trabajar el RF-PA en su región no lineal lo cual tiene efectos no deseados en la señal de salida tales como: productos de intermodulación (IMD), efectos de memoria, recrecimiento espectral así como interferencia en canales adyacentes; lo cual puede llevar a sanciones por parte de los organismos nacionales e internacionales reguladores de las telecomunicaciones [Wood, 2016], [Kiran, 2016], en el caso de México la comisión federal de telecomunicaciones (COFETEL).

Con el fin de corregir los efectos no deseados de la amplificación no lineal del RF-PA se han desarrollado varias técnicas de linealización entre ellas la linealización por anticipación, por retro alimentación y, una de las más estudiadas por su flexibilidad y exactitud, la pre-distorsión digital (DPD) [Wood, 2016], [Braithwaite, 2015]. La técnica de DPD consiste en que la señal de entrada sea acondicionada antes de ser aplicada al RF-PA, el tratamiento de la señal será entonces que pase por un elemento que tenga un comportamiento inverso del RF-PA [Hammi, 2014]. Para lograr obtener un bloque con el comportamiento inverso del RF-PA es necesario en primera instancia contar con un modelo matemático que describa adecuadamente el dispositivo [Liu, 2014], [Moon, 2011]. La literatura muestra una gran cantidad de modelos adoptados para modelado de RF-PAs entre ellos encontramos: modelos basados en series de Volterra, redes neuronales, sistemas neuro-difusos y recientemente modelos generados con programación genética [Fehri, 2014], [Mkadem, 2010], [Zhai, 2008], [Cárdenas, 2017].

Usualmente para la implementación del bloque de comportamiento inverso se elige usar tarjetas de desarrollo sobre todo basadas en FPGA, donde se

aprovechan las bondades de flexibilidad. La implementación de estos modelos en hardware generalmente se hace mediante el uso de tablas de búsqueda (LUT) como se muestra en [Gilabert, 2008] y [Cárdenas, 2015]. Sin embargo el uso de la metodología anterior no permite utilizar señal compleja la cual resulta necesaria al aplicar DPD como un solo modelo donde se incluya el comportamiento en fase y amplitud. Un factor a tomar en cuenta es la complejidad del modelo elegido ya que está impactará directamente en la cantidad de recursos necesarios ya sea por complejidad de computo o por necesidades de almacenamiento, este aspecto es sumamente importante dado que en el FPGA se tienen recursos limitados y la optimización resulta crucial [Renteria, 2016].

VHDL es el lenguaje de descripción de hardware de circuitos integrados de alta velocidad. Mediante este lenguaje se puede describir el comportamiento y la estructura de los sistemas electrónicos y es particularmente adecuado para describir la estructura de los diseños en hardware electrónico digital e implementarlos en plataformas tales como ASICs y FPGAs [Rushton, 2011].

El presente trabajo está organizado de la siguiente manera: se muestra la teoría del modelo polinomial con memoria (MPM) para la realización del modelado de un amplificador NXP 10W a 2GHz, se describe el proceso de codificación del MPM en VHDL así como el desarrollo de un conjunto de funciones que permitiera el uso de números complejos y la representación elegida para el uso de números fraccionales. En la sección 3 se muestra como resultado la simulación en Modelsim con una señal de entrada de amplitud modulada misma que se compara con gráficas de la simulación del modelo en Matlab. En la sección 4 se muestra la viabilidad de los módulos obtenidos para su uso en el modelado de RF-PAs así como DPD. Finalmente en la sección 5 se presentan las conclusiones de este trabajo de investigación.

2. Métodos

Dentro de las técnicas más difundidas para el modelado matemático de los RF-PAs se encuentran los modelos basados en series de Volterra debido a que consideran tanto la no linealidad del dispositivo como los efectos de memoria que

el mismo pudiera tener. Las series de Volterra describen la relación de una entrada con una salida en un sistema no lineal y para un caso discreto con datos del tipo complejos su representación matemática está dada por la ecuación 1.

$$y(n) = \sum_{k=1}^K \sum_{i_1}^Q \dots \sum_{i_{2k-1}}^Q h_{2k-1}(i_1, i_2, \dots, i_{2k-1}) \prod_{j=1}^{k+1} x(n - i_j) \prod_{k+2}^{2k+1} x^*(n - i_j) \quad (1)$$

Dónde:

$()^*$ - denota el complejo conjugado, $x(n)$ – es la entrada discreta de la muestra n , $y(n)$ – Es la salida discreta de la muestra n , K - es el orden de no linealidad del modelo, Q - es la profundidad de memoria del modelo y $h_{2k-1}(i_1, i_2, \dots, i_{2k-1})$ – Es el coeficiente de Volterra de orden k .

El modelo Polinomial con Memoria

A pesar de que las Series de Volterra modelan con exactitud el comportamiento del RF-PA, su uso no resulta práctico al momento de implementar el modelo en hardware debido a su gran complejidad computacional, la cual queda manifiesta cuando al incrementar el orden de no linealidad o la profundidad de memoria el número de coeficientes necesarios para el modelo crece de manera exponencial. Para evitar lo antes mencionado existen modelos derivados de ellas que permiten disminuir la complejidad del modelo sin sacrificar tanta exactitud, uno de estos modelos es el MPM. El MPM consiste en fases de retardo y solo considera los coeficientes de la diagonal principal de Volterra [Nuñez, 2014]. La ecuación 2 muestra el MPM utilizado.

$$y(n) = \sum_{q=0}^Q \sum_{k=1}^K a_{2k-1,q} |x(n - q)|^{2(k-1)} x(n - q) \quad (2)$$

Dónde:

$x(n)$ – es la entrada discreta de la muestra n , $y(n)$ – es la salida discreta de la muestra n , K - es el orden de no linealidad del modelo, Q - es la profundidad de memoria del modelo y $a_{2k-1,q}$ es el coeficiente correspondiente.

El MPM puede ser construido a partir de $Q + 1$ funciones con la estructura de la segunda sumatoria las cuales tendrán como estrada la señal retrasada q veces según sea su caso. En la figura 1, se muestra la estructura a bloques de una función con la estructura de la segunda sumatoria de la ecuación 2, mientras que en la figura 2 se muestra el diagrama a bloques del MPM completo.

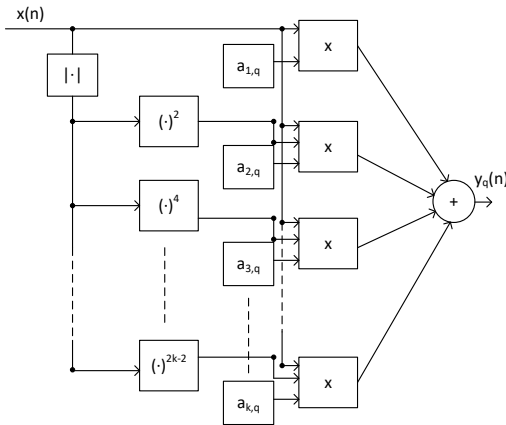


Figura 1 Diagrama a bloques de una función básica del MPM.

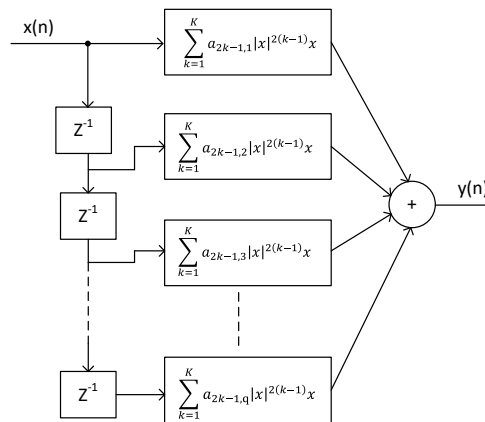


Figura 2 Diagrama a bloques del MPM completo.

Con el fin de extraer los coeficientes que mejor ajustan el modelo, se realiza una regresión lineal basada en mínimos cuadrados de la siguiente manera; la ecuación 2 se representa en forma matricial de modo que adquiere la forma de la ecuación 3 [Ku, 2003].

$$Y = H * a \tag{3}$$

Donde \mathbf{Y} es el vector de salidas, \mathbf{H} es la matriz de observación construida a partir de cada uno de los términos del MPM donde se tiene la unidad como coeficientes y \mathbf{a} es vector de coeficientes cuyo estimador puede ser calculado mediante la ecuación 4.

$$\hat{\mathbf{a}} = \mathbf{H}^+ * \mathbf{Y} \quad (4)$$

Donde \mathbf{H}^+ representa la pseudo-inversa de la matriz de observación.

Se utilizó un MPM con $K = 5$ y $Q = 2$ para el modelado de comportamiento de un amplificador NXP 10W medido a 2 GHz cuyas características eléctricas se muestran en la tabla 1.

Tabla 1 Características eléctricas del NXP 10W.

Parámetro	NXP 10W RF-PA @ 2 GHz
Ganancia (1dBm)	36 dBm
Clase	AB ($V_{ds}=50$ V, $I_{ds}=54$ mA)
Frecuencia de operación	500-2500 MHz
Eficiencia de drenado (η_d)	21%

Se cuenta con un total 65,536 muestras para este amplificador mismas que se usaron para la obtención de las curvas de distorsión, estas curvas caracterizan los efectos que tiene el RF-PA en la señal de entrada tanto en amplitud como en fase y permiten visualizar que también se ajusta el modelo a los valores reales medidos, las curvas de distorsión del NXP 10W fueron realizadas en Matlab y se muestran en los resultados. El cálculo de coeficientes para el MPM fue hecho en Matlab con el objetivo de insertar estos últimos en el modelo escrito en VHDL. Un parámetro numérico que permite cuantificar la calidad del modelo matemático es el error cuadrático medio normalizado (NMSE) el cual está dado por la ecuación 5.

$$NMSE = \frac{\sum_{n=0}^{N-1} (\hat{y}(n) - y(n))^2}{\sum_{n=0}^{N-1} (y(n))^2} \quad (5)$$

En donde: N es el número de muestras, $y(n)$ es el valor real de la salida en la muestra n y $\hat{y}(n)$ es el valor estimado por el modelo para la muestra n . El NMSE suele expresarse en decibeles para lo cual se utiliza la ecuación 6.

$$NMSE(dB) = 10 \log_{10}(NMSE) \quad (6)$$

Codificación del MPM en VHDL

Una de las necesidades para una implementación adecuada del MPM en VHDL es el uso de números fraccionales, por lo que se eligió la representación en punto fijo, esto debido a la ventaja en el tiempo de procesamiento con respecto a la representación en punto flotante.

La representación en punto fijo es una representación de un número fraccionario, el cuál se almacena en la memoria, en este caso el número se almacena como un entero con signo en el formato de dos complementos. Sobre lo anterior, se aplica una separación del vector localizando el punto base que separa la parte entera de la fraccional un número fijo de bits a la izquierda de su posición inicial, lo anterior se ilustra en el diagrama de la figura 3. Cuando se interpretan los bits del entero con signo almacenado en la memoria, se reposiciona el punto de base multiplicando el entero almacenado por un factor de escala fijo en este caso una potencia de dos ya sea positiva (parte entera) o negativa (parte fraccional).

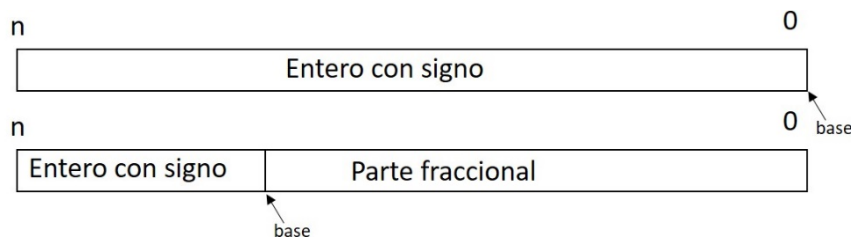


Figura 3 Representación de números fraccionales en punto fijo.

A partir de la versión del 2008, el estándar de VHDL se definen los tipos de datos de punto fijo y punto flotante, sin embargo el entorno de desarrollo QUARTUS II de Altera hace uso de la versión de 1993 por lo que fue necesario el uso de la librería de compatibilidad que se recomienda en [Rushton, 2011]. En dicha librería se tienen disponibles las diferentes funciones de operación para el trabajo con números de punto fijo con y sin signo, así como funciones de conversión y escalamiento.

Con el objetivo de que el módulo escrito pudiera procesar señal compleja, se escribió un conjunto de funciones en VHDL para el manejo de números complejos, en el que se define lo siguiente:

- Definición del tipo de dato complejo
- Suma de números complejos.
- Resta de números complejos.
- Multiplicación de números complejos.
- Absoluto de números complejos.
- Raíz cuadrada de números en representación de punto fijo.

Se planeó representar el tipo de dato complejo en su forma cartesiana, por lo que está constituido por un arreglo de dos vectores de 32 bits los cuales representan dos números fraccionales en punto fijo con signo, el número de bits para la parte entera y la fraccional es configurable con el fin de tener diversos niveles de exactitud y rangos de representación, para este trabajo se normalizó la señal compleja con el fin de tener valores de magnitud a la salida entre 0 y 1.

En el caso de la raíz cuadrada se escribió en VHDL el algoritmo mediante restas, descrito en [Paeth, 2014] y cuyo pseudocódigo aparece en la figura 4, las funciones de operación entre números complejos se escribieron utilizando las definiciones matemáticas para las formas binomiales de números complejos. Para todo lo anterior se eligió lo siguiente: en caso de un desbordamiento del número fraccional que hubiese un comportamiento de saturación y en el caso de que la resolución no fuera suficiente para la representación del número se emplea el redondeo al número más cercano que fuera posible representar. Todas las operaciones con números complejos codificadas tienen parámetros para el escalamiento del resultado.

Con el fin de mejorar el uso de recursos se modificó la parte del MPM en la que se realizan potencias de la magnitud de la entrada, la modificación hecha se muestra en la figura 5, ya que era deseable que la entrada se procesara en un solo ciclo de reloj y no era posible hacer multiplicaciones secuenciales se optó por un diseño donde se toma ventaja de las potencias pares y en lugar de usar 16

multiplicadores solo fueron necesarios 4 para cada una de estas etapas con una no linealidad de 5.

```

1   Definir tipo de dato "FXP" de 32 bits
2   con 2 bits para la parte entera
3   Sqrt(FXP x)
4   {FXP raiz, remHi, remLo, testDiv,
contador;
5   //inicializar parámetros
6   raiz = 0; remHi = 0; remLo = x;count =
30;
7   hacer {//obtener 2 bits del argumento
8     remHi = (remHi<<2) or (remLo>>30);remLo
        <<= 2;
9     raiz <<= 1; //preparar el próximo bit de
la raiz
    
```

Figura 4 Pseudocódigo para extraer la raíz cuadrada de un número de punto fijo de 32 bits.

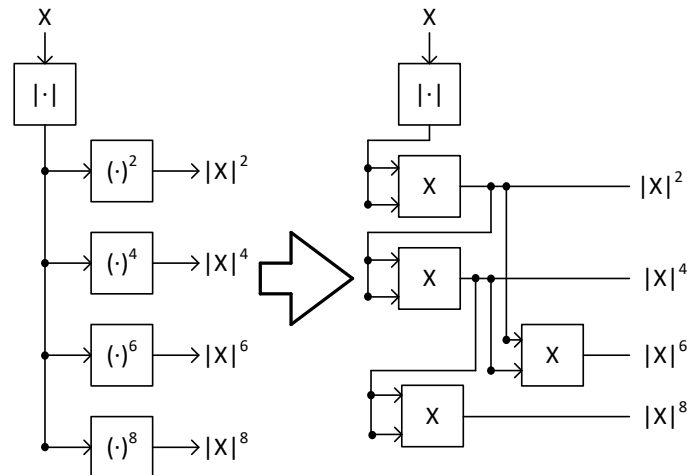


Figura 5 Optimización propuesta para el cálculo de las potencias.

Se desarrolló también el código para un bloque de retardo el cuál a cada ciclo de reloj entrega a la salida la entrada del ciclo anterior y hace una lectura de la nueva para tenerla disponible en memoria para la repetición del ciclo con un valor de 0 para la primera salida. Finalmente, se realizó el diseño de la entidad principal cuyo diagrama a nivel de transferencia de registros (RTL) puede verse en la figura 6 y en la que son claras las similitudes con el diagrama general del MPM mostrado

anteriormente. En la figura 6, puede observarse que el tipo de dato complejo en la entrada del sistema pasa hacia las diferentes funciones no lineales después de ser retrasada la cantidad de veces necesarias para finalmente realizar una suma de las salidas particulares de cada función y con ello lograr obtener la salida del MPM completo.

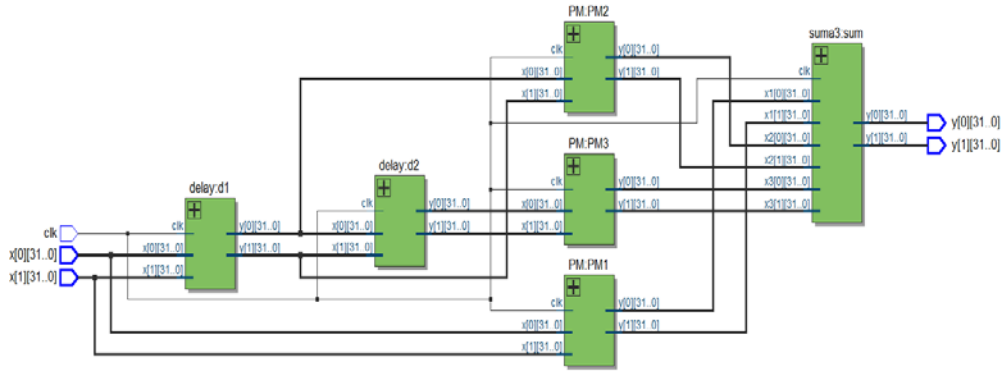


Figura 6 Diagrama RTL del MPM implementado mediante VHDL.

3. Resultados

La figura 7 muestra las curvas de distorsión AM-AM y AM-PM obtenidas por el MPM con $K = 5$ y $Q = 2$, mismo que fue implementado en VHDL. El NMSE obtenido por este modelo fue de -19.8256 dB y como puede observarse ajusta de manera correcta las mediciones del amplificador.

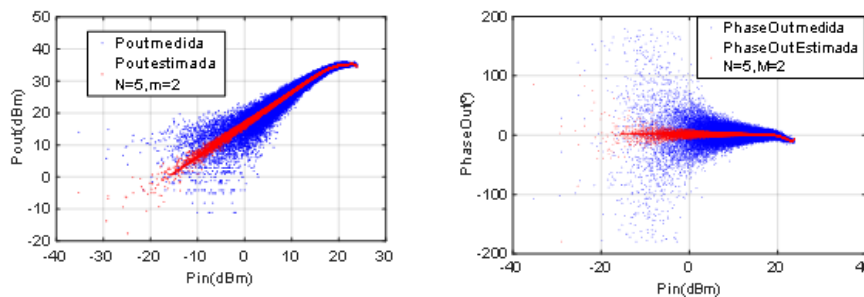


Figura 7 Curvas de distorsión AM-AM y AM-PM del amplificador NXP 10W.

Posterior al desarrollo de la implementación en VHDL del MPM se realizó un análisis de los recursos utilizados para la síntesis del mismo en una tarjeta de evaluación y desarrollo DSP-FPGA Stratix III de Altera, los resultados obtenidos se muestran en la tabla 2.

Tabla 2 Recursos utilizados por el MPM escrito en VHDL.

Recurso	Cantidad usada
ALUTs combinacionales	32,661 / 86,000 (38 %)
ALUTs de memoria	0 / 43,000 (0 %)
Registros lógicos dedicados	512 / 86,000 (< 1 %)
Total de registros	512
Total de pines	129 / 488 (26 %)
Total de pines virtuales	0
Total de bloques de memoria	0 / 4,303,872 (0 %)
Elementos DSP de 18 bits	214/ 288 (74 %)
Total de PLLs	0 / 4 (0 %)
Total de DLLs	0 / 4 (0 %)

Con el fin de conocer el desempeño del módulo escrito, se realizó la codificación de un banco de pruebas, en el que se suministra a la entrada una onda modulada en amplitud (AM) con portadora de 5 MHz y mensaje de 500 kHz muestreada a la frecuencia del reloj del FPGA, la cual es de 50 MHz. Para poder visualizar la salida se verificó la amplitud de la misma mediante el uso del valor absoluto del número complejo, en la figura 8 se muestra la simulación en Modelsim en donde se visualizan las magnitudes de la onda de la entrada y la onda a la salida del módulo que emula el comportamiento del RF-PA mediante el MPM.

Para validar los resultados obtenidos en el modelo MPM escrito en VHDL se realizó la codificación del mismo usando Matlab y se le aplicó la misma señal de amplitud modulada a la entrada, obteniendo los resultados observados en las figuras 9 y figura 10. Como puede observarse al comparar las gráficas tanto de la simulación en VHDL como de Matlab tienen exactamente la misma forma de onda. A nivel de valores en el binomio que representa el número complejo las variaciones con relación a la simulación fueron pequeños, puesto que en la representación en punto fijo se tenía una resolución mínima de 2^{-30} .

Finalmente, se realizó la programación de la tarjeta DSP-FPGA Stratix III para la visualización de los resultados, mismos que se muestran en las figuras 11 y 12, se acondicionó la señal resultante de modo que pudiera ser utilizada por el convertidor digital analógico de 14 bits de la tarjeta de adquisición de datos Terasic HSMC AD/DA.

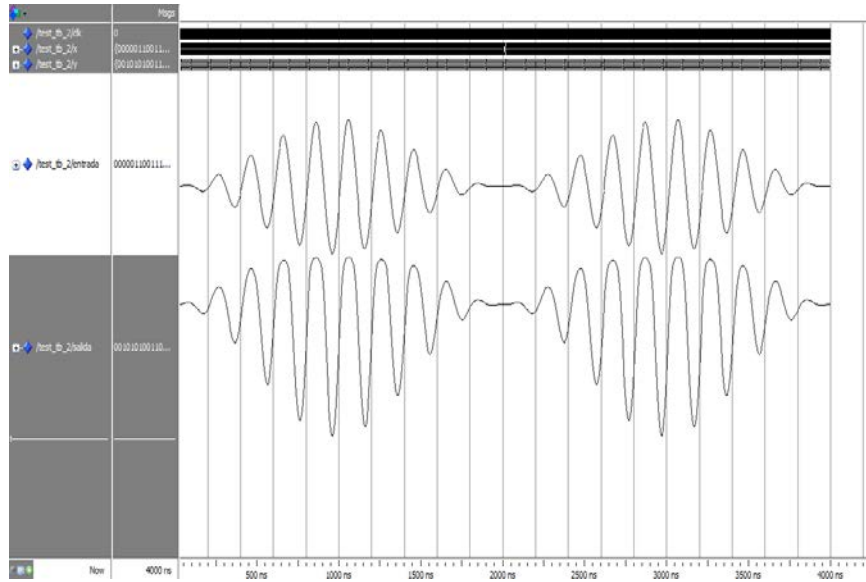


Figura 8 Simulación del modelo MPM a una entrada AM.

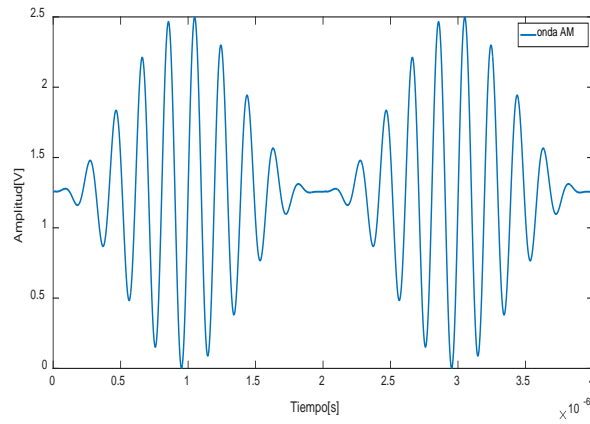


Figura 9 Entrada AM al modelo MPM escrito en Matlab.

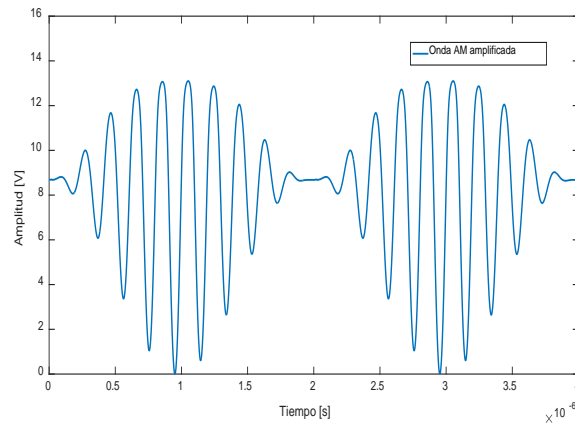


Figura 10 Salida del modelo MPM escrito en Matlab.

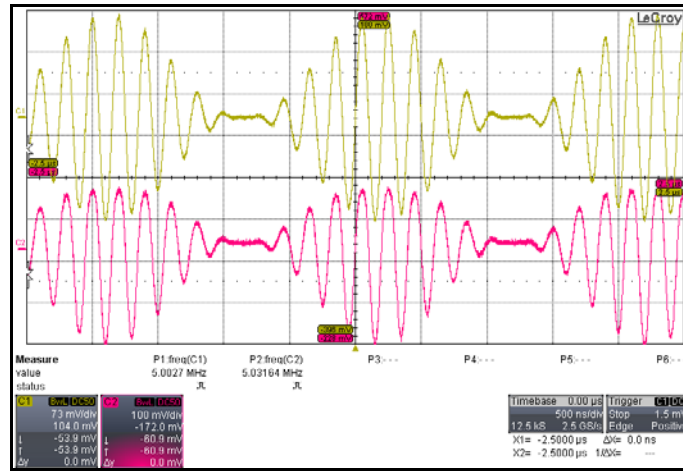


Figura 11 Entrada y salida del modelo MPM implementado en una tarjeta FPGA Stratix III.



Figura 12 Banco de pruebas con la tarjeta Stratix III emulando el amplificador NXP 10 W.

4. Discusión

Después de haber realizado la implementación del MPM completo en VHDL, es posible observar que a diferencia de aquellas implementaciones basadas en LUTs tiene un considerable aumento en el uso de recursos lógicos. Sin embargo tiene la capacidad de utilizar señal compleja tanto a la entrada como a la salida logrando tener la capacidad de modelar en una sola implementación, los efectos del RF-PA tanto en amplitud como en fase, además de poder trabajar en un rango mayor en la entrada ya que no está limitada a una cierta cantidad de direcciones como en el caso mencionado.

El MPM codificado tiene las características de no linealidad 5 y profundidad de memoria 2. Sin embargo, estas características pueden ser modificadas según se requiera para el modelado de otros modelos de RF-PA con lo que queda

manifiesta la flexibilidad del uso de la plataforma FPGA para este tipo de aplicaciones, si bien el MPM utilizado se optimizó en la parte de potenciación del valor absoluto de la entrada, para otras aplicaciones con un MPM de parámetros fijos, el sistema puede optimizarse de una manera más profunda con lo que se puede lograr un uso óptimo de los recursos disponibles en el FPGA.

5. Conclusiones

El MPM permite un modelado correcto del comportamiento del RF-PA tanto en amplitud como en fase logrando en el caso del NXP 10W un NMSE de -19.8256 dB con un ajuste de coeficientes mediante regresión lineal simple. Además de ser implementado de manera exitosa en la plataforma FPGA con un uso mediano de recursos, por lo que representa un modelo que tiene las características de complejidad moderada y exactitud aceptable, ambas deseables para la aplicación de DPD.

La codificación del MPM completo en VHDL ,permite una emulación más adecuada para el desarrollo de pre-distorsionadores que aquellas implementaciones basadas en LUTs , puesto que permite emular los efectos completos que induce el RF-PA en la señal de entrada, todo lo anterior con el fin de poder probarlos sin tener el RF-PA de manera física. Con la implementación del MPM completo en FPGA se abre la posibilidad del cálculo de coeficientes directamente en esta plataforma a través de diversos métodos de estimación particularmente la estimación por mínimos cuadrados secuenciales, además de poder realizar la DPD de manera adaptativa ajustando los parámetros del modelo inverso según cambie el comportamiento del RF-PA por calentamiento o envejecimiento de componentes.

6. Bibliografía y Referencias

- [1] Cárdenas-Valdez J. R. et al., Local Search Approach to Genetic Programming for RF-PAs Modeling Implemented en FPGA, Results of the Numerical and Evolutionary Optimization Workshop NEO 2015, Springer, pp. 67-88, 2017.

- [2] Cárdenas-Valdez J. R. et al., Modeling memory effects in RF power amplifiers applied to a digital pre-distortion algorithm and emulated on a DSP-FPGA board, *Integration, the VLSI Journal*, Volume 49, pp 49-64, 2015.
- [3] Fehri B. and Boumaiza S., Baseband Equivalent Volterra Series for Behavioral Modeling and Digital Predistortion of Power Amplifiers Driven With Wideband Carrier Aggregated Signals, *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 11, pp. 2594-2603, 2014.
- [4] Gilabert P.L. et al., Multi-Lookup Table FPGA Implementation of an Adaptive Digital Predistorter for Linearizing RF Power Amplifiers With Memory Effects, *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 2, pp. 372-384, 2008.
- [5] Hammi O. et al., A Digital Predistortion System With Extended Correction Bandwidth With Application to LTE-A Nonlinear Power Amplifiers, *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 12, pp. 3487-3495, 2014.
- [6] Kiran V., ACPR reduction for better power efficiency using adaptive DPD, 2016 International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, pp. 0495-0498, 2016.
- [7] Ku H. y Kenney J. S., Behavioral modeling of nonlinear RF power amplifiers considering memory effects, *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 12, pp. 2495-2504, 2003.
- [8] Liu Y.J. et al., A Robust Augmented Complexity-Reduced Generalized Memory Polynomial for Wideband RF Power Amplifiers, *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2389-2401, 2014.
- [9] Mkadem F. et al., Behavioral modeling and digital predistortion of Power Amplifiers with memory using Two Hidden Layers Artificial Neural Networks, 2010 IEEE MTT-S International Microwave Symposium, Anaheim, CA, pp. 656-659, 2010.
- [10] Paeth A., *Graphics Gems V (Macintosh Version)*, 1ra ed. Burlington, Elsevier Science, pp. 22-24, 2014.

- [11] Moon J. y Kim B., Enhanced Hammerstein Behavioral Model for Broadband Wireless Transmitters, *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 4, pp. 924-933, 2011.
- [12] Núñez Pérez J.C. et al., Flexible test bed for the behavioural modelling of power amplifiers, *COMPEL - The international journal for computation and mathematics in electrical and electronic engineering*, vol. 33, no. 1/2, pp. 355–375, 2014.
- [13] R. N. Braithwaite, A Comparison for a Doherty power amplifier linearized using digital predistortion and feedforward compensation, *2015 IEEE MTT-S International Microwave Symposium*, pp. 1-4, Phoenix, AZ, 2015.
- [14] Renteria J. et al., A novel configurable FPGA architecture for hardware implementation of multilayer feedforward neural networks suitable for digital pre-distortion technique, *2016 46th European Microwave Conference (EuMC)*, London, pp. 854-857, 2016.
- [15] Roblin P. et al, Concurrent linearization: The state of the art for modeling and linearization of multiband power amplifiers, *IEEE Microwave Magazine.*, vol. 14, no. 7, pp. 74–91, 2013.
- [16] Rushton A., *VHDL for logic synthesis*, 3ra ed. Chichester, John Wiley and Sons, 2011.
- [17] Wood J. et al., The Evolution of PA Linearization, *IEEE Microwave Magazine*, no. 2, pp. 32–40, 2016.
- [18] Zhai J. et al., Dynamic Behavioral Modeling of Power Amplifiers Using ANFIS-Based Hammerstein, en *IEEE Microwave and Wireless Components Letters*, vol. 18, no. 10, pp. 704-706, 2008.

METODOLOGÍA PARA LA INTEGRACIÓN DE UN MANIPULADOR MÓVIL BAJO SOFTWARE LIBRE

Oswaldo Alquisiris Quecha

Universidad del Istmo
oswaldoaq@gmail.com

Francisco Aguilar Acevedo

Universidad del Istmo
aguilar.afco@sangunga.unistmo.edu.mx

Ignacio Algreto Badillo

Universidad Politécnica de Tlaxcala
ignacio.algreto@uptlax.edu.mx

J. Jesús Arellano Pimentel

Universidad del Istmo
jjap@sandunga.unistmo.edu.mx

Resumen

Los manipuladores móviles combinan manipuladores y plataformas móviles para obtener sistemas con capacidades extendidas. Una de las ventajas que presentan los manipuladores móviles es su amplio espacio de trabajo. Con un propósito didáctico, en este artículo se presenta la integración de un manipulador móvil con ruedas a partir del robot móvil ERA-MOBI y el brazo robótico *Smart Robotics Arm*. El manipulador móvil que opera bajo Linux en su distribución Fedora 23, fue puesto a prueba bajo un modo de teleoperación. Los experimentos llevados a cabo muestran que nuestro manipulador móvil necesita alrededor de tres minutos para tomar y colocar una sola pieza a 2.5 metros de distancia.

Palabras Claves: Brazo robótico, manipulador móvil, robot móvil.

Abstract

Mobile manipulators combine manipulators and mobile platforms to obtain systems with extended capabilities. Their expanded working space is one of the mobile manipulators' advantages. The purpose of this article is educational. It presents the integration of a mobile manipulator on wheels, which uses an ERA-MOBI mobile robot and a Smart Robotics Arm. The mobile manipulator, which operates on Linux - Fedora distribution 23, was tested using a remote operative modality. The tests show that our mobile manipulator needs approximately three minutes to take and place a single piece at a distance of 2.5 meters

Keywords: *Mobile manipulator, mobile robot, robotic arm.*

1. Introducción

Un manipulador móvil consiste de uno o más manipuladores montados en la parte superior de una plataforma móvil [Li, 2013]. Un manipulador móvil combina las destrezas de manipulación ofrecida por los manipuladores de base-fija y la movilidad de las plataformas móviles. Estos sistemas tienen muchas aplicaciones potenciales en la manufactura, mantenimiento de reactores nucleares, construcción y exploración planetaria. Si bien, resulta obvio el potencial de los manipuladores móviles para realizar diversas tareas, también es cierto que su aumento en complejidad estructural trae consigo desafíos en su modelado, planeación de movimiento y control.

Alrededor de la conformación y estudio de sistemas manipuladores móviles diversas investigaciones se han encausado. En [Szrek, 2016] se presenta el modelado y análisis de un manipulador móvil *wheel-legged*, el cual es una combinación de una plataforma móvil con un sistema de suspensión especial y un manipulador. La estructura cinemática del manipulador, un modelo computacional del manipulador con la plataforma, el diseño del sistema de control, y el prototipo desarrollado son presentados. En [Deepak, 2016] se aborda la coordinación y control de un manipulador móvil con ruedas usando un sistema inmune artificial (AIS, *Artificial Immune System*). La metodología tiene como objeto la navegación autónoma en entornos industriales. Un robot manipulador de cuatro ejes y una

plataforma móvil con ruedas de tipo diferencial son integrados para conformar el manipulador móvil.

La aplicación de manipuladores móviles también ha sido motivo de investigación. El estudio de [Ding, 2017] se enfoca en el uso de un manipulador móvil para la apertura de puertas en cuatro fases: llegar a la puerta, tomar la manija de la puerta, girar la manija de la puerta y tirar de la puerta. Se emplean un control retroalimentado de fuerza/torque para el contacto con la manija, y un control adaptativo para el seguimiento de la trayectoria planeada para abrir la puerta. La implementación de la propuesta en un manipulador móvil que abre una puerta real cerrada, demuestra la validez del enfoque. Por su parte, en [Lu, 2017] se propone la conformación de un manipulador móvil denominado BOW (Baxter-on-Wheels) operable por personas con discapacidad motora pero cognitivamente sanas. BOW combina el amigable robot industrial Baxter de Rethink Robotics con una silla de ruedas comercial. Se emplea una estrategia de control compartido que combina el comando humano con mecanismos que proporcionan movimientos intuitivos y seguros. En [Li, 2017] se describe el desarrollo de un prototipo de sistema denominado TRINA (*Tele-Robotic Intelligent Nursing Assistant*) el cual consiste de un manipulador móvil, una consola para un operador humano, y algoritmos de asistencia al operador que automatizan parcial o totalmente las tareas tediosas y propensas a errores. Las capacidades del sistema para realizar tareas estándar de enfermería son evaluadas en un laboratorio de simulación.

No obstante la diversidad de trabajos relacionados, no es posible identificar una metodología explícita para la integración de manipuladores móviles de carácter didáctico. Para el caso de la plataforma móvil ERA-MOBI y el brazo robótico AX-18A *Smart Robotics Arm*, su uso individual ha sido motivo de estudio. En [Saez, 2010] se describe un sistema multirobot usando dispositivos ERA-MOBI para la asistencia a bomberos en casos de búsquedas y rescate en incendios donde el humo producido reduce drásticamente la visibilidad. Con ayuda de sus sensores el conjunto de robots móviles guían al bombero indicándole posibles obstáculos. El enfoque del prototipo es la generación de conductas básicas para permanecer en grupo, es decir, generar una formación y navegar mientras se mantiene ésta. En

[Gutiérrez, 2014] se presenta el desarrollo de una interfaz de comunicación capaz de reconocer una serie de comandos de voz, con el propósito de interactuar con el robot ERA-MOBI. La comunicación interfaz-robot se implementó mediante sockets orientados a conexión y el software *Player/Stage* que permite controlar los dispositivos de un robot y obtener información de sus sensores. Por su parte, en [Wang, 2016] un robot AX-18A *Smart Robotics Arm* es usado en un sistema de microondas de bajo costo con fines de diagnóstico y tratamiento médico. El robot que es posicionado en función del área de interés identificada mediante una cámara web, porta un par de antenas en su extremo final que permiten obtener información en 3 dimensiones. Finalmente en [Griggs, 2016] un robot AX-12 *Smart Robotics Arm* es empleado como plataforma de prueba de una propuesta de control cartesiano. El sistema se presenta como un enfoque de control alternativo de prótesis de brazos robóticos. Los componentes de software fueron desarrollados bajo el software comercial Matlab®.

Así, en este artículo se presenta la conformación de un manipulador móvil con ruedas a partir del uso de la plataforma móvil ERA-MOBI y el brazo robótico AX-18A *Smart Robotics Arm*. La contribución del trabajo se centra en integración de los diversos componentes, lo que antepone un carácter didáctico a la metodología que se describe. Se muestran los pormenores de la integración del hardware, y del software bajo Linux, así como pruebas funcionales del sistema.

2. Métodos

AX-18A *Smart Robotics Arm* es un brazo robótico de 5 grados de libertad (gdl) distribuidos en cuatro articulaciones, tres de rotación y una cilíndrica (rotación/traslación), que tiene como actuadores a los servomotores Dynamixel de la serie AX-18A. El robot puede ser controlado mediante una computadora vía USB bajo Windows o Linux, empleando su interfaz hardware denominada USB2Dynamixel y su correspondiente SDK (*Software Development Kit*). El SDK contiene APIs (*Application Programming Interface*) que permiten emplear VB.NET, C#, LabVIEW, MATLAB, JAVA, C/C++ y Python, entre otros entornos y lenguajes para la programación del robot.

ERA-MOBI es un robot móvil de tipo diferencial compacto equipado con un arreglo de sensores ultrasónicos e infrarrojos (IR), y una computadora integrada a la plataforma. A nivel de software, los movimientos del robot ERA-MOBI son programados con la herramienta *Player*, un servidor genérico de código abierto (*open source*) utilizado en el control de robots móviles, cuya versión 3.0.2 cuenta con soporte para 133 diversos controladores. La función de Player es permitir el control de los componentes de distintos dispositivos robóticos mediante un sencillo sistema de interfaces genéricas, independientes del hardware del robot [Whitbrook, 2010].

En la figura 1 se ilustra la metodología empleada para la integración del manipulador móvil denominado ERA-SRA en alusión al robot móvil ERA-MOBI y el manipulador *Smart Robotics Arm*, empleados para su conformación.

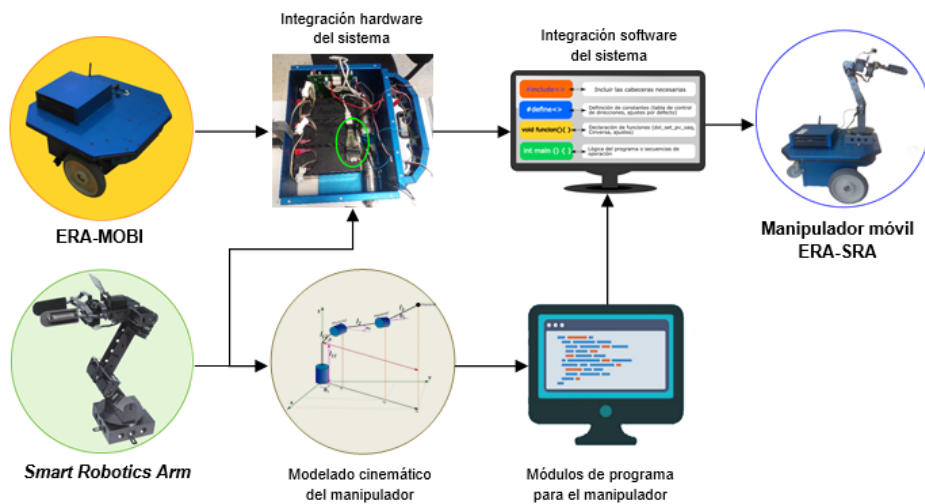


Figura 1 Metodología de integración del manipulador móvil ERA-SRA.

Por su parte, la **integración de hardware** consistió en incorporar al robot ERA-MOBI el brazo robótico, para lo cual fue necesario adecuar al interior del robot móvil conexiones para la energización de los servomotores del manipulador, e incluir perforaciones a la cubierta del móvil para sujetar la base del brazo robótico. La **integración de software** se realizó sobre la computadora a bordo del ERA-MOBI bajo Linux en su distribución Fedora 23. Dado que los componentes de software de Player son definidos en C++, los módulos de programa adicionales

son desarrollados en este lenguaje. Así, al diseñar un programa será necesario incluir las librerías <playerc++.h> y <dynamixel.h> que brindan el software base para el control del robot móvil y los servomotores del manipulador respectivamente. La librería dynamixel define métodos para controlar la comunicación con los servomotores, y definir/transmitir/recibir paquetes de datos, más no así para el control específico del *Smart Robotics Arm* bajo una configuración, para lo cual es necesario describir el movimiento del robot mediante un modelo.

El **modelo cinemático** se limitó a los primeros tres grados de libertad del manipulador, que son suficientes para definir la posición en el espacio del efector final. La obtención de la cinemática directa e inversa se realizó mediante un enfoque geométrico. En la figura 2 se muestra una representación del manipulador empleada para la obtención de la cinemática directa descrita por las ecuaciones 1, 2 y 3, y la cinemática inversa expresada en las ecuaciones 4, 5 y 6.

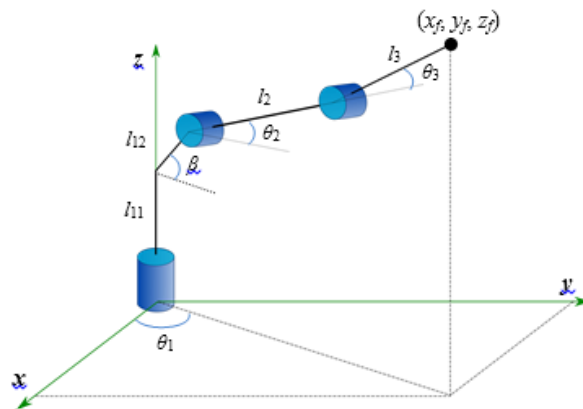


Figura 2 Diagrama de alambre para manipulador de 3 gdl.

$$x_f = \cos \theta_1 (l_{12} \cos \beta + l_2 \cos \theta_2 + l_3 \cos(\theta_2 + \theta_3)) \quad (1)$$

$$y_f = \sin \theta_1 (l_{12} \cos \beta + l_2 \cos \theta_2 + l_3 \cos(\theta_2 + \theta_3)) \quad (2)$$

$$z_f = l_{11} + l_{12} \sin \beta + l_2 \sin \theta_2 + l_3 \sin(\theta_2 + \theta_3) \quad (3)$$

$$\theta_1 = \arctan\left(\frac{y_f}{x_f}\right) \quad (4)$$

$$\theta_2 = \arctan\left(\frac{z'}{w}\right) - \arctan\left(\frac{l_3 \sin \theta_3}{l_2 + l_3 \cos \theta_3}\right) \quad (5)$$

Donde:

$$D = \frac{z'^2 + w^2 - l_2^2 - l_3^2}{2l_2l_3} \quad w = \sqrt{x_f^2 + y_f^2 - l_{12} \cos \beta} \quad z' = z_f - (l_{12} \sin \beta + l_{11})$$

Siendo el signo del argumento en la ecuación 5, el que definirá la configuración que describa la posición del brazo, el signo + denota la configuración codo abajo y con el signo - la configuración codo arriba [Siciliano, 2010].

Bajo un enfoque de programación modular se diseñaron cinco **módulos de programa para el manipulador** en lenguaje C, dos para el cálculo de las cinemáticas, dos para la operación y comunicación con el brazo robótico, y uno para relacionar los valores del modelo cinemático con los valores angulares que asumirán los servomotores del manipulador. Dichas posiciones resultan ser distintas debido a la forma en que los motores son anclados a cada una de las articulaciones del robot. En base a los esquemas mostrados en la figura 3, se pueden establecer las siguientes relaciones:

$$\theta_{1motor} = \theta_{1modelo} + 150^\circ, \theta_{2motor} = -\theta_{2modelo} + 240^\circ, \theta_{3motor} = -\theta_{3modelo} + 150^\circ$$

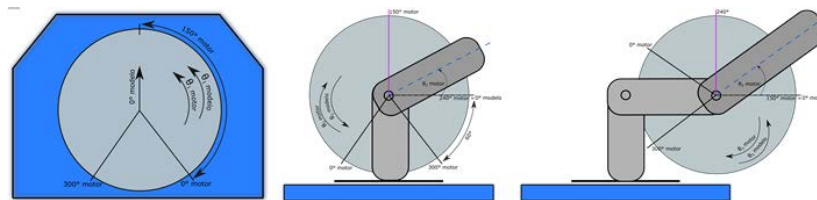


Figura 3 Diferencias entre valores angulares del modelo y los reales.

3. Resultados

Las pruebas sobre el manipulador se realizaron en dos sentidos. Verificar el modelo cinemático obtenido y los módulos de programa para el manipulador, y evaluar el desempeño del manipulador móvil.

Para el caso del brazo robótico se definió una tarea la cual consistió en programar el posicionamiento del extremo final del robot (con la pinza cerrada) para ir de una posición inicial a una objetivo (sin carga), modificando las velocidades de operación de los servomotores (de una resolución de 10 bits y 0.111 rpm por

unidad), con el propósito de ilustrar el error de posicionamiento del manipulador. En la figura 4 se esquematiza la prueba planteada y tabla 1 se muestran los resultados obtenidos.

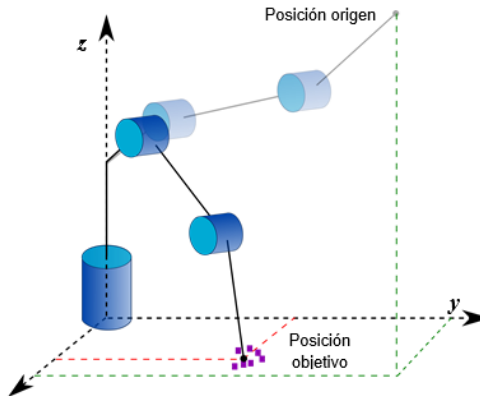


Figura 4 Esquema del error de posicionamiento.

Tabla 1 Error de posicionamiento.

Velocidad (rpm)	Error absoluto (mm)
2.77 (4%)	x = 1 mm, y = 2.5 mm
5.55 (4.8%)	x = 1 mm, y = 1 mm
8.32 (7.3%)	x = 1 mm, y = 4 mm
11.1 (9.7%)	x = 1 mm, y = 2.5 mm
13.87 (12.2%)	x = 3 mm, y = 5 mm
16.65 (14.6%)	x = 2 mm, y = 4.5 mm
19.42 (17.1%)	x = 2 mm, y = 3 mm
22.2 (19.5%)	x = 2 mm, y = 7.5 mm
24.97 (21.9%)	x = 1 mm, y = 9 mm
27.75 (24.4%)	x = 1 mm, y = 1 mm

% respecto a la máxima velocidad

El error absoluto máximo en el posicionamiento fue de 3 mm y 9 mm para x y y , respectivamente. Se observó que al usar los servomotores a una velocidad de 27.75 rpm (250/1023 unidades) el manipulador presento errores de posicionamiento similares a los obtenidos operando los servomotores a 5.55 rpm (50/1023), sin embargo, una velocidad alta trae consigo un mayor impacto de los efectos de inercia que pueden comprometer la integridad del robot.

Con el propósito de estimar el error del manipulador en la reubicación de un objeto (movimiento con carga), se realizó el montaje mostrado en la figura 5a. Para ello se programó la tarea de tomar un objeto (con forma de prisma hexagonal de 35mm de altura y un peso de 18.34 gramos) ubicado en una posición inicial, trasladarlo a velocidad constante y finalmente colocarlo en la casilla de una cuadrícula, En la figura 5b se observan los contornos de las piezas reubicadas, tras la realización de cuatro pruebas. El error máximo medido fue de 7 mm respecto a los centros de ambas figuras (posición objetivo y final). Durante las pruebas se observó que la rotación de la pieza durante su traslado se debía al movimiento no controlado en velocidad y sus consecuentes efectos de inercia, así como a la superficie lisa (aluminio) del objeto que impacta en la fuerza de agarre de la pinza.

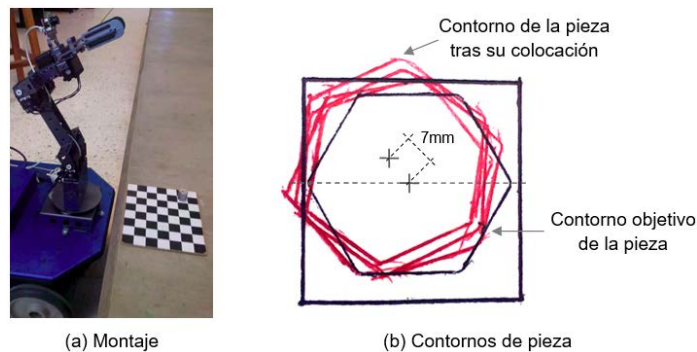


Figura 5 Error del manipular al colocar una pieza.

Respecto a la funcionalidad del manipulador móvil se trató al ERA-SRA como un sistema robótico teleoperado [Cerón, 2005]. Su operación se desarrolla mediante dos arquitecturas Cliente/Servidor de conexión inalámbrica Wifi, una por cada dispositivo. Para el móvil ERA-MOBI se ejecuta el servidor *player* en la computadora a bordo, mientras el cliente opera el robot a través de una herramienta de Player denominado *playerjoy*. Para el caso del manipulador *Smart Robotics Arm* se desarrolló un cliente/servidor en lenguaje C. Para validar la teleoperación del manipulador móvil se realizó una prueba que consistió en medir el tiempo empleado por el usuario para tomar un objeto y trasladarlo a otra posición recorriendo una distancia total de 2.5 m, ver figura 6.

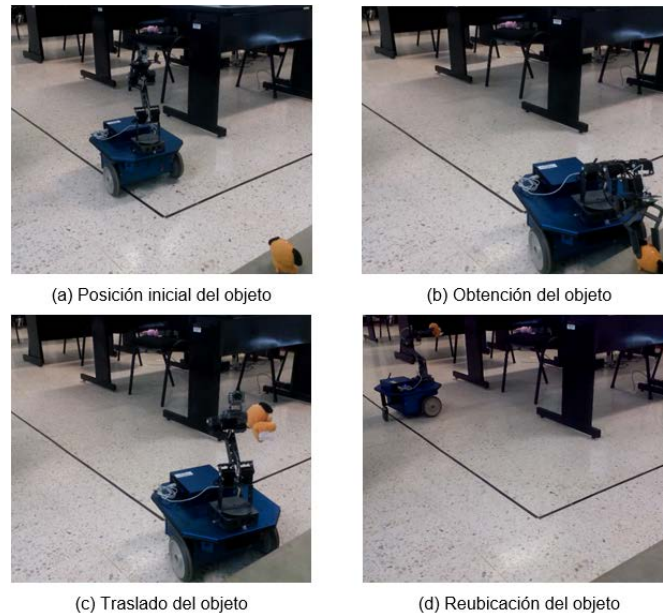


Figura 6 Prueba mediante teleoperación.

La interacción con el sistema se realizó a través de una interfaz de usuario ver figura 7, mediante la cual el extremo final del manipulador puede ser desplazado en el espacio cartesiano (± 1 cm de resolución) y abrir/cerrar la pinza, mientras el móvil puede avanzar/retroceder de manera constante al 30% de su velocidad máxima y realizar giros de 45° (a velocidad 0). El tiempo promedio empleado por diez usuarios para realizar la tarea fue de 3 minutos.



Figura 7 Interfaz gráfica para el control del manipulador móvil.

4. Discusión

La teleoperación de un manipulador móvil para la realización de tareas que puedan ser consideradas peligrosas para los humanos, es solo una aplicación dentro de la llamada robótica de servicio. En la salud, en particular estos sistemas pueden ser empleados para mejorar la calidad de vida de las personas al bríndales apoyo en sus actividades, como pueden ser tomar, trasladar y colocar objetos.

El manipulador móvil planteado cuenta con prestaciones que lo hacen candidato a ser considerado como una plataforma experimental para la conformación de un robot de servicio, no obstante, es necesario considerar dos supuesto bajo los cuales se realizaron las pruebas antes descritas: las texturas de los objetos empleados son blandas lo que facilita su sujeción con la pinza del manipulador, y los objetos utilizados son de un peso inferior al máximo estimado por el fabricante (1.58 kg). Por otra parte, el empleo de velocidades altas en la operación de los servomotores del manipulador trae consigo efectos de inercia, lo cual impacta en el desempeño de los motores y en la trayectoria descrita por el robot, al respecto una mejora inmediata al sistema es la definición de una curva de movimiento suave (por ejemplo una trayectoria polinómica cubica) que permita desplazar el objeto sujeto por el robot de una posición a otra mientras se satisface condiciones de velocidad y/o aceleración.

Es de mencionar que dado el propósito didáctico de este trabajo, no es posible establecer una comparativa con desarrollos existentes, no obstante cabe señalar, que el manipulador móvil ERA-SRA presenta características distinguibles con respecto a los desarrollos que hacen uso de cada dispositivo de manera individual, lo cual extiende su campo de aplicación.

5. Conclusiones

En el presente artículo se abordó la integración de un manipulador móvil con ruedas bajo plataforma Linux, empleando el robot móvil ERA-MOBI y el brazo robótico *Smart Robotics Arm*. La metodología descrita desde un enfoque didáctico, buscar definir un procedimiento reproducible para la integración de un manipulador

móvil bajo sistemas iguales o similares a los empleados. A nivel de software cabe señalar que todas las herramientas y librerías empleadas son de licencia libre.

El manipulador móvil es capaz de tomar objetos y trasladarlos a un destino establecido. La interfaz de usuario desarrollada permitió la teleoperación del dispositivo, validando su utilidad.

Como trabajo a futuro se sugiere: implementar trayectorias de movimiento suave para cada servomotor con el fin de mejorar la operación del robot; para el control simultáneo de los dispositivos se sugiere realizar una programación con un enfoque paralelo a través de la implementación de hilos dentro de las aplicaciones; integración de un sistema de visión por computadora que permita calcular la posición/orientación del objeto en el espacio cartesiano para su correcta sujeción.

6. Bibliografía y Referencias

- [1] Cerón, A., *Sistemas Robóticos Teleoperados*. Ciencia e Ingeniería Neogranadina, 15, 62-69, 2005.
- [2] Deepak, B. B. V. L., & Parhi, D. R., Control of an automated mobile manipulator using artificial immune system. *Journal of Experimental & Theoretical Artificial Intelligence*, 28(1-2), pp. 417-439, 2016.
- [3] Ding, L., Xia, K., Gao, H., Liu, G., & Deng, Z., Robust adaptive control of door opening by a mobile rescue manipulator based on unknown-force-related constraints estimation. *Robotica*, pp. 1-22, 2017.
- [4] Griggs, L., & Fahimi, F., Introduction and testing of an alternative control approach for a robotic prosthetic arm. *The open biomedical engineering journal*, 8, pp. 93-105. 2014.
- [5] Gutiérrez, H., Arellano, J. J., & Pacheco, D., Interacción humano-máquina por voz para la operación de plataformas robóticas móviles. *Pistas Educativas*, 108, pp. 1309-1328, 2014.
- [6] Li, Z., & Ge, S. S., *Fundamentals in Modeling and Control of Mobile Manipulators* (Vol. 49). CRC Press. Boca Raton, FL. 2013.

- [7] Li, Z., Moran, P., Dong, Q., Shaw, R. J., & Hauser, K., Development of a Tele-Nursing Mobile Manipulator for Remote Care-giving in Quarantine Areas. *IEEE Int. Conf. Robotics and Automation Proceedings*, Singapore, May, 2017.
- [8] Lu, L., & Wen, J. T. Baxter-On-Wheels (BOW), An Assistive Mobile Manipulator for Mobility Impaired Individuals. In *Trends in Control and Decision-Making for Human–Robot Collaboration Systems*, pp. 41-63, Springer International Publishing, 2017.
- [9] Saez-Pons, J., Alboul, L., Penders, J., & Nomdedeu, L., Multi-robot team formation control in the GUARDIANS project. *Industrial Robot: An International Journal*, 37(4), pp. 372-383, 2010.
- [10] Siciliano, B., Sciavicco, L., Villani, L., & Oriolo, G., *Robotics: modelling, planning and control*. Springer Science & Business Media. Girona, Spain, 2010.
- [11] Szrek, J., Muraszkowski, A., & Sperzyński, P., Type synthesis, modelling and analysis of the manipulator for wheel-legged robot. *acta mechanica et automatica*, 10(2), pp. 87-91, 2016.
- [12] Wang, F., Wu, X., & Arslan, T., Mobile-controlled portable robotic measurement setup for microwave imaging diagnosis. In *Antennas & Propagation Conference*, Loughborough, UK. November, 2016.
- [13] Whitbrook, A., *Programming mobile robots with Aria and Player*. Springer Science & Business Media, London, 2010.

APROXIMACIÓN AL RECONOCIMIENTO DE EMOCIONES FACIALES BASADO EN POSICIÓN DE PUNTOS DE INTERÉS

Víctor Manuel Álvarez Pato

Universidad Panamericana
valvarez@up.edu.mx

Ramiro Velázquez Guerrero

Universidad Panamericana
rvelazquez@up.edu.mx

Resumen

Con las técnicas actuales de reconocimiento facial, es posible descubrir automáticamente las emociones de una persona a través de una imagen de su rostro. Este estudio se vale de una aplicación en línea para detectar algunos puntos de interés en imágenes de rostros que expresan alguna emoción y compara sus posiciones con las de una expresión considerada neutral. Se busca establecer una relación entre el resultado obtenido y el propuesto por la herramienta FACS de Paul Ekman para determinar la viabilidad de un algoritmo de reconocimiento de emociones, así como posibles pautas para su desarrollo.

Palabras Claves: Face++, FACS, reconocimiento de emociones, reconocimiento facial.

Abstract

With the current facial recognition techniques, it is possible to automatically determine an individual's emotions through a digital image of his face. The present study employs an online API to detect certain landmarks in images of faces affected by some emotion and compares their positions with those of a neutral expression. We seek to establish a relationship between the obtained results and the one proposed by Paul Ekman's FACS tool to determine the viability of an

emotion recognition algorithm, as well as some possible guidelines for its development.

Keywords: *Emotion recognition, Face++, face recognition, FACS.*

1. Introducción

El rostro humano es capaz de comunicar grandes cantidades de información: nos permite distinguir a una persona de otras, conocer sus sentimientos e incluso prever sus reacciones inmediatas por medio de un rápido análisis facial.

La capacidad de procesar toda esta información ha sido exclusiva de algunos organismos vivos hasta hace relativamente poco tiempo. El trabajo de cientos de investigadores ha dado origen a algoritmos capaces de reconocer y aprovechar parte de esos datos presentes en el rostro [Zhao, 2003], [Liong, 2016], [Giannakakis, 2017], a pesar de que en muchas ocasiones sean tan sutiles que resultan difíciles de describir en el lenguaje común.

Actualmente, el desarrollo de la ciencia ha permitido que las computadoras puedan determinar si una imagen digital contiene o no un rostro humano, además de localizar su posición [Zhan, 2016], [Yan, 2014], [Wang, 2017]. Esto se conoce como detección facial y suele ser la primera fase en muchos algoritmos de análisis facial. También se ha conseguido identificar rostros de manera automática, lo cual resulta útil por ejemplo en sistemas de seguridad. La disciplina que engloba este tipo de algoritmos suele recibir el nombre de reconocimiento facial.

El reconocimiento de expresiones faciales, por otro lado, es una rama que estudia el modo de analizar y reconocer movimientos faciales y cambios en los rasgos de la cara a partir de información de tipo visual. Esto representa un apoyo fundamental para otro objetivo que requiere aún mayor conocimiento: el análisis de emociones [Sariyanidi, 2015], [Tarnowski, 2017].

El análisis automatizado de emociones permite tener una idea de los sentimientos que experimenta una persona, aun cuando ella misma no sea completamente consciente de ellos. Así, un video que capture las expresiones de un voluntario al probar un producto nuevo en un estudio mercadológico puede analizarse por medio de una computadora para conocer con mayor detalle sus reacciones,

mismas que podrían pasar inadvertidas si se le pidiera contestar una encuesta por escrito [Hamelin, 2017], [Yu, 2017]. Otras aplicaciones ya estudiadas incluyen diagnosticar desórdenes neuropsiquiátricos [Hamm, 2011], *E-learning* [Krithika, 2016] y monitorizar avances en terapias psiquiátricas, por citar algunos ejemplos. Este documento muestra una primera aproximación al reconocimiento automático de emociones en el rostro humano, con la intención de explorar su posible aplicación como apoyo en el diagnóstico de enfermedades.

Buena parte de los estudios realizados en esta materia, se basan los trabajos del psicólogo Paul Ekman [Ekman, 1971], quien sostiene que los principios de expresión facial son prácticamente innatos e independientes de la influencia cultural: esto permite crear un sistema de reconocimiento de emociones basado únicamente en la información visual del rostro.

Ekman también desarrolló el Sistema de Codificación de Acciones Faciales (FACS por sus siglas en inglés), que reduce cualquier emoción a una serie de movimientos aislados de los músculos faciales [Ekman, 1983]. Por ejemplo, una expresión de alegría en un rostro puede traducirse como una contracción de la porción orbitaria del músculo orbicular en combinación con el músculo cigomático mayor. Este tipo de movimientos cambian la posición relativa de algunos puntos de la cara –como la comisura de los labios, o los bordes de las cejas– que hoy en día resultan fáciles de detectar con la tecnología disponible.

2. Métodos

Para los experimentos realizados, se utilizó la interfaz de programación (API) de Face++ [Face++, 2017], una tecnología de servicios cognitivos en la nube gestionada por la empresa china Megvii, cuyo algoritmo principal está basado en redes neuronales convolucionales [Fan, 2014]. Dicha API permite enviar a los servidores de Face++ un conjunto de imágenes y recibir en formato JSON las coordenadas (x,y) de 83 puntos de interés para cada rostro detectado. Por medio de un programa escrito en Java se automatizó la transformación y envío de imágenes, y después se tradujeron los archivos JSON al formato .mat para poder analizar la información en Matlab.

El programa fue alimentado con una serie de imágenes obtenidas de la base de datos de Expresiones Faciales de Mujeres Japonesas (JAFFE) [Lyons, 1998], ya que ésta es de fácil acceso y contiene fotografías de varios rostros y diversas emociones. De este universo, se seleccionaron seis individuos con siete expresiones diferentes (una de ellas se considera neutral).

La teoría de Ekman postula la existencia de siete expresiones faciales básicas: alegría, tristeza, sorpresa, miedo, enojo, disgusto y desdén. En el estudio se buscó comprobar la relación entre los cambios en los puntos de interés provistos por Face++ y la clasificación realizada por humanos de las expresiones disponibles en la base de datos. De las siete expresiones básicas citadas, no se incluyó la última en las pruebas por no encontrarse explícitamente entre las imágenes disponibles. Una vez obtenidas las coordenadas de cada punto de interés, se normalizaron todos los conjuntos de acuerdo a dos puntos considerados menos sujetos a cambios a lo largo de todas las fotografías: los correspondientes al contorno superior de la nariz. Este proceso permite reducir las alteraciones de escala y rotación entre las imágenes utilizadas.

Para cada uno de los puntos de interés, se evaluó la diferencia entre su posición en el rostro de un individuo cuando su expresión se muestra neutral y su posición en cada una de las seis expresiones restantes, figura 1; finalmente se graficaron estas diferencias, promediadas entre los seis individuos. En la figura 2 se pueden apreciar los vectores de cambio para cada punto.

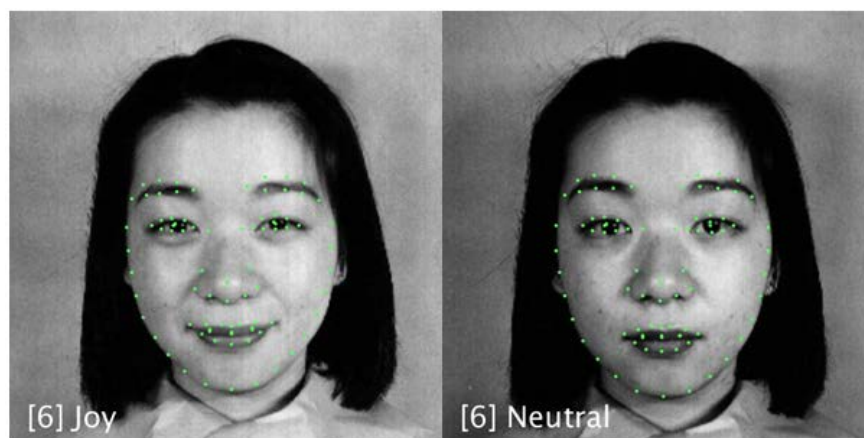


Figura 1 Localización de puntos de interés en dos expresiones del mismo individuo.

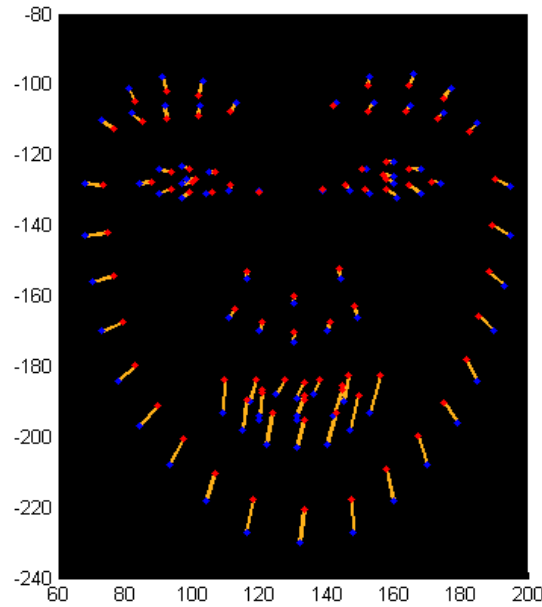


Figura 2 Diferencias en puntos de interés para las imágenes anteriores.

3. Resultados

Las gráficas obtenidas muestran por medio de una escala de colores el cambio en píxeles para la coordenada vertical de cada uno de los puntos de interés. Los puntos aparecen con sus descripciones respectivas y están agrupados por similitud en figuras 3 a la 7.

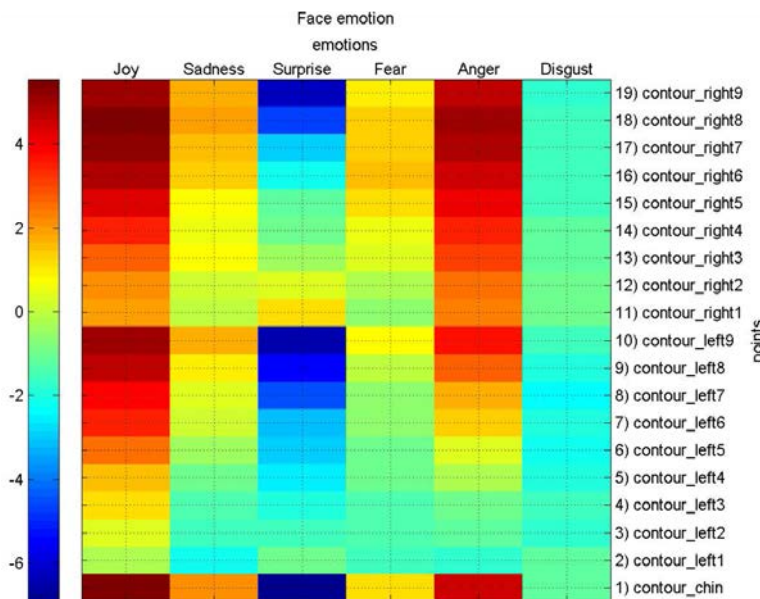


Figura 3 Desplazamientos verticales en los puntos asociados al contorno de la cara.

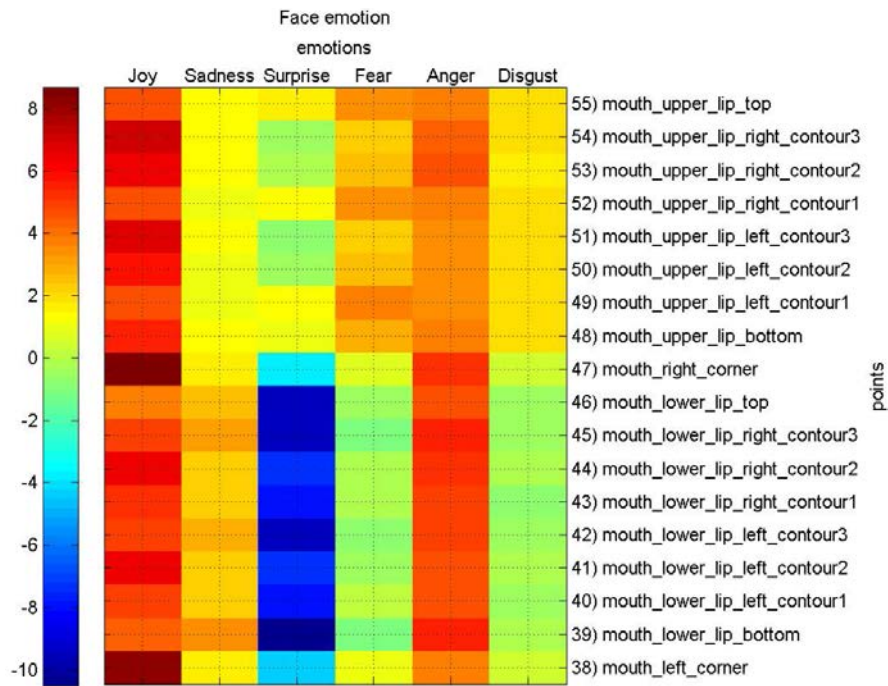


Figura 4 Desplazamientos verticales en los puntos correspondientes a la boca.

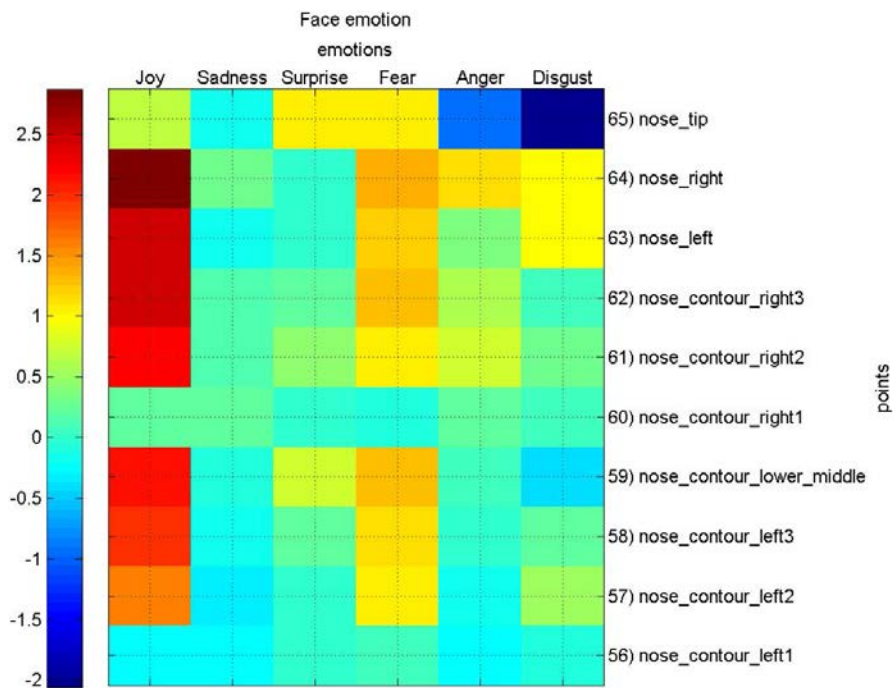


Figura 5 Desplazamientos verticales en los puntos correspondientes a la nariz.

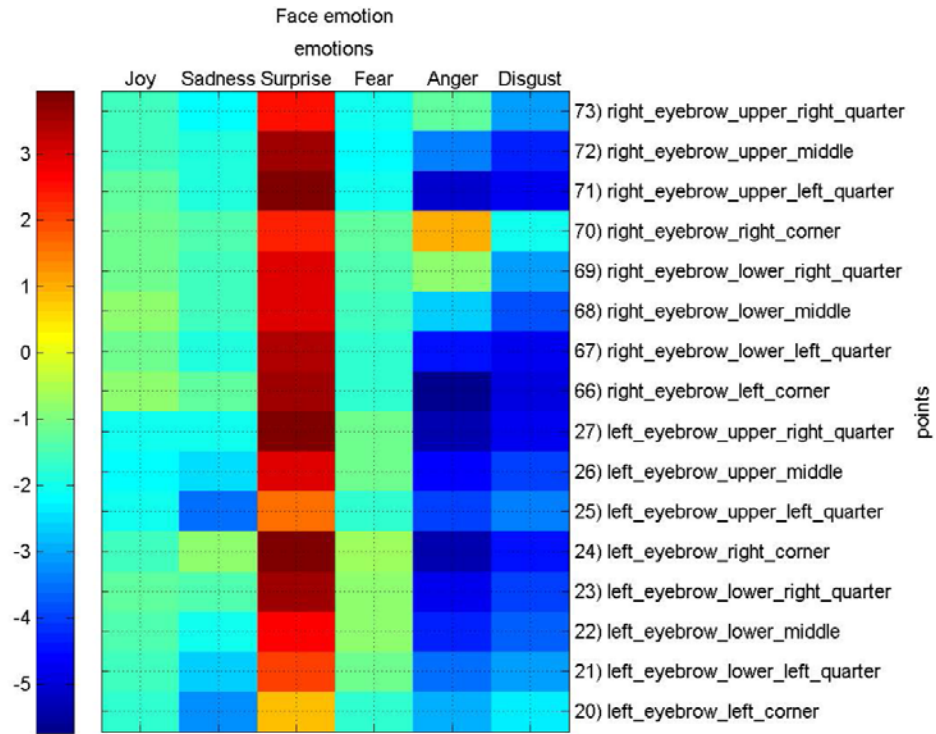


Figura 6 Desplazamientos verticales en los puntos correspondientes a las cejas.

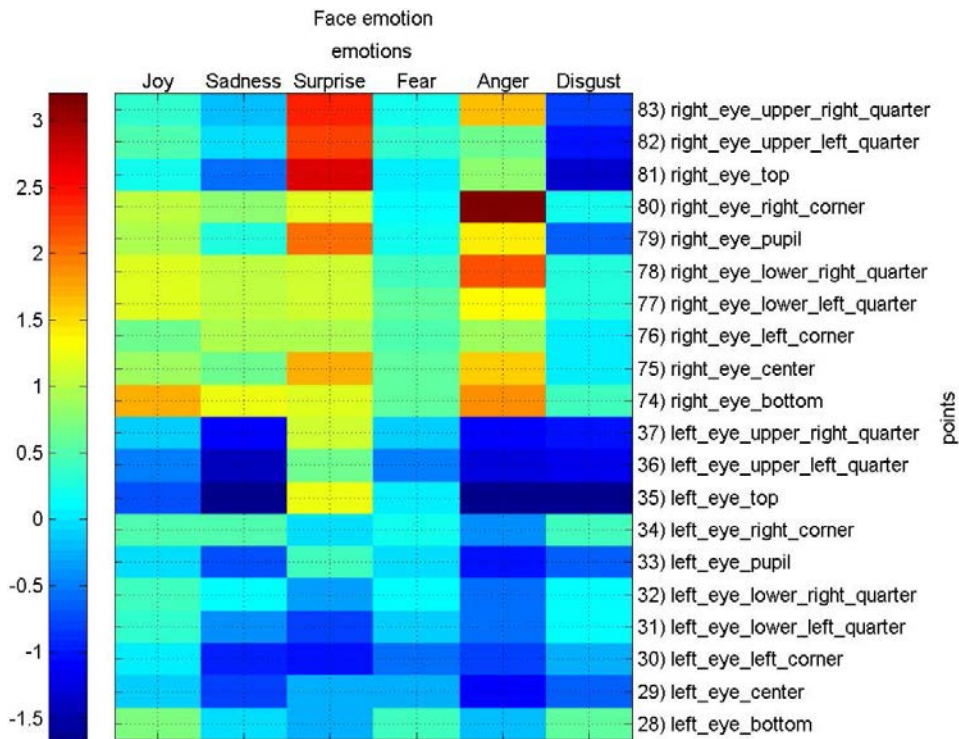


Figura 7 Desplazamientos verticales en los puntos correspondientes a los ojos.

4. Discusión

Puede observarse la congruencia entre los datos obtenidos y lo predicho por Ekman en las siguientes emociones:

- **Alegría:** se percibe un leve levantamiento de los párpados inferiores en los puntos 74 y 28 y levantamiento en las comisuras de los labios en los puntos 47 y 38. Los resultados coinciden con la predicción.
- **Tristeza:** el efecto del conjunto de músculos elevadores internos de las cejas puede apreciarse en los puntos 24 y 66, aunque deberían subir, parecen bajar ligeramente, es probable que por la acción del siguiente grupo muscular. Puede verse que los depresores de las cejas actúan en los puntos 26 y 25, los cuales claramente bajan, también en 73 y 72 se aprecia un descenso, aunque con menor claridad. Las comisuras de la boca 47 y 38 parecen subir ligeramente, cuando cabría esperar que bajaran por la actividad del grupo depresor de las comisuras.
- **Sorpresa:** los puntos 24 y 66 suben claramente, confirmando la previsión para los elevadores internos de las cejas. Lo mismo sucede con todos los puntos correspondientes a las cejas. Los músculos elevadores del párpado superior (35 y 81) presentan ascenso, con más claridad del lado derecho que del izquierdo. Esto concuerda con el supuesto de que los ojos se abren para expresar sorpresa, al igual que la boca, movimiento que puede distinguirse en los puntos del contorno facial: 1 a 19.
- **Miedo:** esta es la expresión más complicada de detectar, ya que actúan varios grupos de músculos simultáneamente, a veces cancelando entre sí los efectos observables. Por ejemplo, los músculos elevadores y depresores de las cejas afectan los puntos 24, 25, 26, 66 y 72 y 73 en sentidos opuestos, por lo que las lecturas obtenidas no son concluyentes. Lo mismo puede decirse de los puntos 28, 35, 74 y 81, que son jalados por los elevadores de párpados superiores y empujados por el músculo orbicular.
- **Enojo:** las cejas bajan claramente, como lo muestran los puntos 25, 26 y 72 los párpados superiores bajan en los puntos 35 y 81, de acuerdo con lo

previsto. Las coordenadas verticales no parecen ser suficientes para confirmar la tensión esperada en los labios ni en los orbiculares. Tal vez haga falta un análisis ligeramente distinto.

- **Disgusto:** el músculo transversal de la nariz hace subir los puntos 63 y 64 como estaba pronosticado, pero los puntos 38 y 47 en las comisuras de los labios parecen subir cuando deberían bajar, por otro lado, el depresor del labio inferior asociado al punto 39 no parece moverse.

5. Conclusiones

Los resultados obtenidos confirman la teoría para tres de las seis emociones estudiadas.

Para las otras tres, dada su complejidad (el método FACS requiere un entrenamiento especializado) y la aparente acción contraria de algunos grupos de músculos cabría replantear el tipo de análisis realizado. En el futuro sería interesante alimentar los puntos de interés a algún tipo de inteligencia artificial (p.e. redes neuronales) para obviar el estudio pormenorizado de cada músculo individual.

No todos los puntos de interés provistos por la API de Face++ tienen la misma importancia. Podrían obtenerse resultados similares con un conjunto más reducido.

Las imágenes utilizadas corresponden a unas pocas personas, mujeres exclusivamente, y sus expresiones son actuadas. Sería útil contar con una base de datos más amplia y con expresiones naturales.

La normalización aplicada a los puntos de interés no es suficiente para corregir los errores que los movimientos de la cabeza introducen en los cálculos. Si uno de los sujetos levanta la barbilla o voltea hacia un lado, afecta a todos los puntos de interés.

6. Bibliografía y Referencias

- [1] Ekman, P., & Friesen, W. V., Constants Across Cultures in the Face and Emotion. *Journal of Personality and Social Psychology*, 1971.

- [2] Ekman, P., Friesen, W. V., & Hager, J., The Facial Action Coding System (FACS): A technique for the measurement of facial action. Palo Alto, California: Consulting Psychologists Press, Inc. Ekman, P. Levenson. RW, & Friesen WV, Autonomic nervous system activity distinguishes among emotions. *Science*, 221, 1208-12, 1983.
- [3] Face++ - Cognitive Services - Leading Facial Recognition Technology: <https://www.faceplusplus.com/>, último acceso junio 2017.
- [4] Fan, H., Cao, Z., Jiang, Y., Yin, Q., & Doudou, C., Learning Deep Face Representation. *ArXiv preprint arXiv: 1404*, pp. 3840, 2014.
- [5] Giannakakis, G. et al. Stress and anxiety detection using facial cues from videos. *Biomedical Signal Processing and Control*, 31, pp. 89-101, 2017.
- [6] Hamelin, N., El Moujahid, O., Thaichon, P. Emotion and advertising effectiveness: A novel facial expression analysis approach. *Journal of Retailing and Consumer Services*. 2017. 36, pp. 103-111.
- [7] Hamm, J., Kohler, C.G., Gur, R.C., Verma, R. Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders. *Journal of Neuroscience Methods*, pp. 237-256, 2011.
- [8] Krithika L.B, Lakshmi P.G., Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric. *Procedia Computer Science*, 85, pp. 767-776, 2016.
- [9] Liong, et al. Spontaneous subtle expression detection and recognition based on facial strain, *Signal Processing: Image Communication*, 47, pp. 170-182, 2016.
- [10] Lyons, M.J. et al. Coding Facial Expressions with Gabor Wavelets. *Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE Computer Society. Nara, Japan, April 1998.
- [11] Sariyanidi, E. et al. Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(6), pp. 1113-1133, 2015.

- [12] Tarnowski, P., Kołodziej, M., Majkowski, A., Rak, R.J., Emotion recognition using facial expressions, *Procedia Computer Science*, 108, pp. 1175-1184, 2017.
- [13] Wang, N., et al., Facial feature point detection: A comprehensive survey. *Neurocomputing*, 2017.
- [14] Yan, J., Zhang, X., Lei, Z., Li, S. Face detection by structural models, *Image and Vision Computing*, 32 (10), pp. 790-799, 2014.
- [15] Yu, C. & Ko, C., Applying FaceReader to Recognize Consumer Emotions in Graphic Styles, 60, pp.104-109, 2017.
- [16] Zhan, S., Tao, Q. & Li, X. Face detection using representation learning, *Neurocomputing*, 187, 19-26, 2016.
- [17] Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. Face recognition: A literature survey, *ACM computing surveys (CSUR)*, 35(4), 399-458, 2003.

DISEÑO Y ANÁLISIS DE LA ω DE UN MOTOR DE CC MEDIANTE LA SELECCIÓN ÓPTIMA DE PARÁMETROS

Jesús Alejandro Álvarez Tostado Uribe

General Motors

ironalex2310@gmail.com

Irma Martínez Carrillo

Universidad Autónoma del Estado de México

imartinezca@uaemex.mx

Carlos Juárez Toledo

Universidad Autónoma del Estado de México

cjuarez@uaemex.mx

Resumen

El uso de motores en las actividades cotidianas del hombre ha sido una herramienta de gran ayuda para simplificar y/o automatizar procesos repetitivos donde se requiere de un gran esfuerzo humano. La existencia de diversos motores permite seleccionar el modelo adecuado de acuerdo a las funciones que se requieran, siendo así el motor de corriente continua (CC) uno de los motores más utilizados, actualmente uno de los temas principales en el ahorro de energía en los procesos industriales es la variación de la velocidad de los motores en tiempos picos y tiempos de baja producción para disminuir el consumo de energía eléctrica.

En este trabajo se presenta una metodología basado en el método de identificación de estabilidad de Routh-Hurwitz y el lugar geométrico de las raíces para identificar los parámetros y los valores que podrían cambiarse dentro de la función característica que define el modelo de estudio.

Se presentan dos resultados el primero con condiciones nominales del sistema original y el segundo con variación de parámetros.

Palabras Claves: Función característica, lugar geométrico, motor de CC, Routh-Hurwitz, variación de parámetros.

Abstract

The use of motors in the daily activities of man has been a great tool to simplify or to automate repetitive processes. The existence of several motors allows to select the appropriate model according to the functions that are required, this works studies one of the most popular motors (DC motor).

This paper presents a methodology based on the Routh-Hurwitz stability identification with the root locus analysis to identify the parameters and values that could be changed in the characteristic function.

Two results are presented the first with nominal conditions of the original system and the second with variation of parameters.

Keywords: characteristic function, DC motor, root locus, Routh-Hurwitz, variation of parameters.

1. Introducción

Un motor de corriente continua (CC) es un sistema que está constituido de una parte eléctrica y una mecánica para conformar un sistema electromecánico como se muestra en la figura 1.

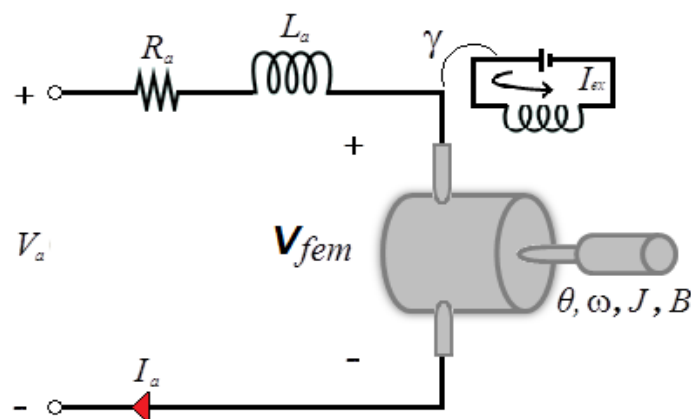


Figura 1 Diagrama esquemático del motor de CC [Dorf, 2008].

En estudios y análisis convencionales de los motores de CC se centran en los valores de placa proporcionados por el fabricante el cual ya mantiene las características funcionales del motor fijas, usualmente son diseñadas mediante softwares de Matlab (Simulink), Multisim, etc. los cuales proporcionan una respuesta de salida a partir de una señal de entrada.

En este trabajo se propone un modelo analítico para conocer la dinámica del motor de CC y encontrar los parámetros óptimos de algunos de sus elementos mediante el método de Routh-Hurwitz y el lugar geométrico de las raíces; Ya que ante los diversos cambios de demanda en la fabricación de productos industriales, donde los procesos de manufactura están gobernados por motores y que la velocidad del rotor se mantiene constante en intervalos de tiempo definidos, se consume la misma energía en el transcurso del proceso, por consiguiente se propone variar la velocidad para acelerar un proceso en horas pico, bajando la velocidad en ciertos tiempos, propiciando un ahorro de energía y evitando sobrecalentamiento en el motor como a continuación se describe.

2. Métodos

Para la implementación de un controlador dentro de la dinámica del comportamiento natural del sistema de estudio, considérese el modelo mostrado de la figura 1.

Las ecuaciones que describen el comportamiento dinámico del motor de CC mostrado en la figura 1, [Roldán, 2014], [Wildi, 2007] son:

$$V_a = R_a i_a + L_a \frac{di_a}{dt} + V_{fem} \quad (1)$$

$$J \frac{d\omega(t)}{dt} = T_r - B\omega(t) \quad (2)$$

La ecuación 1 y ecuación 2 representan el comportamiento eléctrico y mecánico respectivamente del motor de CC de la figura 1, donde cada uno de sus elementos se describe en la tabla 1.

El acoplamiento entre la parte eléctrica y mecánica del motor de CC están representados por las ecuaciones 3 a la 5.

$$V_{fem} = k_3 \omega(t) = k_3 \frac{d\theta(t)}{dt} \quad \text{cupla eléctrica-mecánica} \quad (3)$$

$$T_m(t) = k_1 \gamma i_a(t) \quad \text{cupla mecánica-eléctrica} \quad (4)$$

$$\gamma = k_{ex} i_{ex} \quad (5)$$

Tabla 1 Variables del motor de CC.

Símbolo	Definición
$V_a(t)$	Tensión aplicada al motor
$I_a(t)$	Corriente del motor
$L_a(t)$	Inductancia total equivalente en serie
R_a	Resistencia total
$\omega(t)$	Velocidad angular del motor
J	Momento de inercia
B	Coeficiente de rozamiento
T_r	Par resistente

Sustituyendo la ecuación 3 en la ecuación 1 y ecuación 4 en la ecuación 2 resultan ecuaciones 6 y 7.

$$V_a = R_a i_a + L_a \frac{di_a}{dt} + k_3 \omega(t) \quad (6)$$

$$T_m(t) = k_1 k_{ex} i_{ex} i_a(t) = k_2 i_a(t) \quad \text{con } k_2 = k_1 k_{ex} i_{ex} \quad (7)$$

Aplicando transformada de Laplace a las ecuaciones 2, 6 y 7 se obtienen las funciones de transferencia, ecuaciones 8 a 10.

$$\frac{\omega(s)}{T_r(s)} = \frac{1}{Js+B} \quad (8)$$

$$\frac{I_a(s)}{V_a(s) - k_3 \omega(s)} = \frac{1}{sL_a + R_a} \quad (9)$$

$$\frac{T_m(s)}{I_a(s)} = k_2 \quad (10)$$

La representación en diagramas de bloques de las ecuaciones 8, 9 y 10 se muestran en la figura 2.

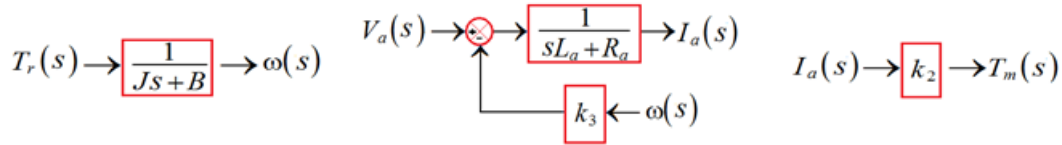


Figura 2 Diagramas de bloques de las ecuaciones 8, 9 y 10.

Relacionando los diagramas de bloques de la figura 2 para conformar un solo sistema de lazo cerrado con señal de entrada $V_a(s)$ y señal de salida $\omega(s)$ se ilustra en la figura 3.

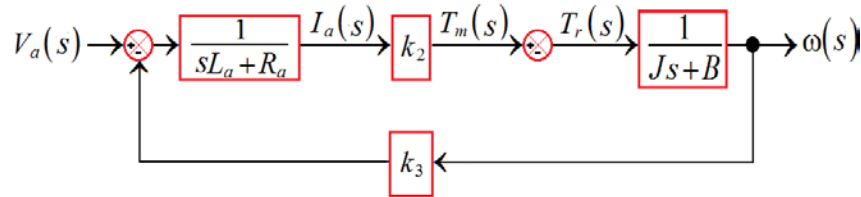


Figura 3 Diagrama de bloques de lazo cerrado de motor de CC.

Para lograr el equilibrio entre la cupla eléctrica-mecánica se considera que $T_r(s)=T_m(s)$ obteniéndose la función de transferencia:+

$$\frac{\omega(s)}{V_a(s)} = \frac{k_2}{JL_a} \frac{1}{s^2 + \left(\frac{JR_a + BL_a}{JL_a}\right)s + \left(\frac{BR_a + k_2k_3}{JL_a}\right)} \quad (11)$$

Selección de Parámetros Óptimos

Para la selección del o los parámetros se propone una metodología basada en los métodos que se describen a continuación.

Método de Estabilidad de Routh-Hurwitz

Proporciona una respuesta inmediata para conocer la estabilidad a partir del análisis de la función característica escrita como un polinomio de la forma [Ogata, 2010]:

$$a_0s^n + a_1s^{n-1} + a_2s^{n-2} + \dots + a_{n-1}s + a_n = 0 \quad (12)$$

Donde los coeficientes de la función característica se agrupan de acuerdo al arreglo de la figura 4.

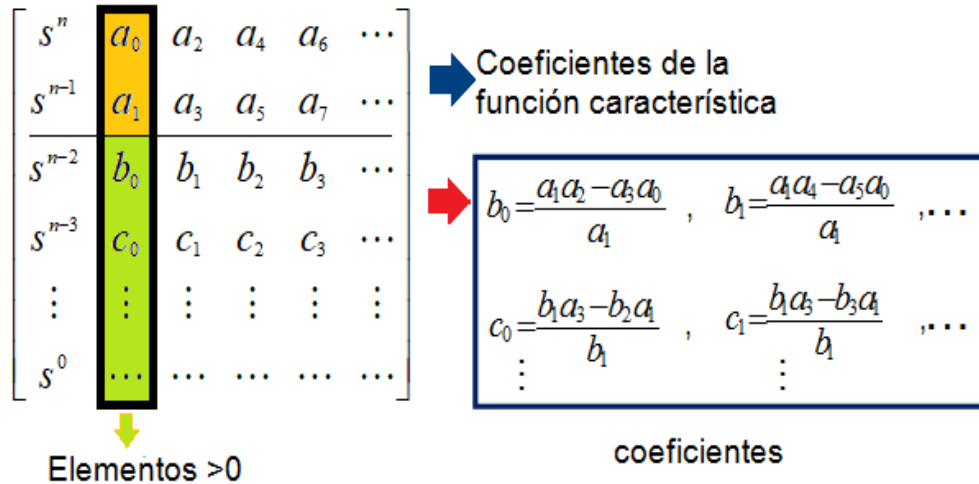


Figura 4 Arreglo de coeficientes por método de Routh-Hurwitz.

La condición necesaria para determinar si el sistema es estable con el método de Routh- Hurwitz, requiere que todos los coeficientes de la función característica de 12, tengan signo positivo o negativo, además, los elementos de $a_0, a_1, b_1, \dots, h_1$ sean positivos, [Navarro, 2004].

Método del Lugar Geométrico de las Raíces

La característica básica de la respuesta transitoria de un sistema de un lazo cerrado se relaciona estrechamente con la localización de los polos, los polos son los elementos que permiten la solución de la función característica de la ecuación (12), entonces la estabilidad del sistema se define si la ubicación de las soluciones se grafica en el lado izquierdo del plano complejo [Dorsey, 2005].

Selección de Parámetros Óptimos para Variar la ω

El uso del método de estabilidad de Routh-Hurwitz, conmutado con el lugar geométrico de las raíces, permite la obtención de un modelo propuesto para la selección de parámetros diferentes a los de la placa nominal del sistema que optimicen el modelo de estudio.

El método propuesto consiste en seleccionar adecuadamente los siguientes puntos:

- Identificar los parámetros candidatos a modificarse del sistema original.
- Mantener el o los parámetros a modificar como variables dentro de la función característica.
- Usar el método de Routh-Hurwitz para identificar rangos permisibles de estabilidad.
- Mapear la ubicación de raíces de la función característica de acuerdo a los valores sugeridos en 3.
- Calcular y graficar la respuesta de salida en el tiempo de acuerdo a los parámetros seleccionados en 3 y 4.
- Analizar y comparar la respuesta obtenida con respecto a la señal original.

El proceso para la selección de parámetros óptimos se resume en el diagrama de la figura 5.



Figura 5 Metodología para seleccionar parámetros óptimos.

La metodología propuesta para la selección de los parámetros a modificar dentro de la dinámica del sistema de estudio se ilustrará en el siguiente apartado.

3. Resultados

Para la implementación del método propuesto se realizarán dos casos de estudio bajo las siguientes condiciones:

- **Caso 1:** Sistema original usando valores de placa.
- **Caso 2:** Sistema con variación de parámetros

En la tabla 2 se muestran las condiciones nominales del sistema de estudio.

Tabla 2 Condiciones nominales del sistema de estudio [Dorf, 2008].

Valores de placa	
$V_a(t)=26Vu(t)$	$B=150$
$R_a=0.2\Omega$	$k_2=10$
$L_a=0.001H$	$k_3=50$
$J=30kg.m^2$	----

Identificación de Parámetros Por Modificar

A partir de la función de transferencia de la ecuación 11 se obtiene el arreglo de coeficientes de la función característica por el método de Routh-Hurwitz como se muestra en la figura 6.

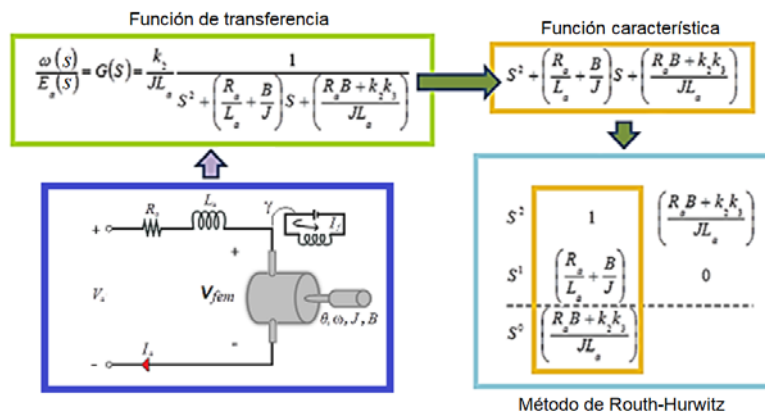


Figura 6 Identificación de estabilidad por el método de Routh-Hurwitz.

Usando los valores nominales de la tabla 2, las desigualdades originadas del método de Routh-Hurwitz resultan

$$\left(\frac{JR_a + BL_a}{JL_a}\right) = 205 > 0 \quad \text{y} \quad k_3 > \frac{R_a B}{k_2} \tag{13}$$

De acuerdo con el análisis de la ubicación de las raíces del sistema de la función característica de la ecuación 11 manteniendo como variable k_3 en un intervalo de $(-3, \infty)$, se obtiene la ubicación de los polos ubicados como se muestra en la figura 7.

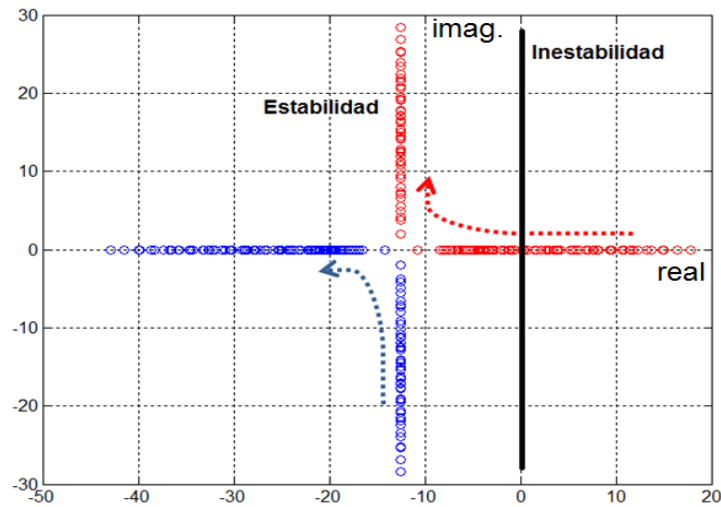


Figura 7 Tendencia de las raíces del sistema.

Reduciendo de forma aleatoria el intervalo para k_3 se puede visualizar polos ubicados en el lado izquierdo del plano complejo en la gráfica de la figura 8, los cuales son de interés ya que garantiza la estabilidad del sistema visto desde el método del lugar de las raíces.

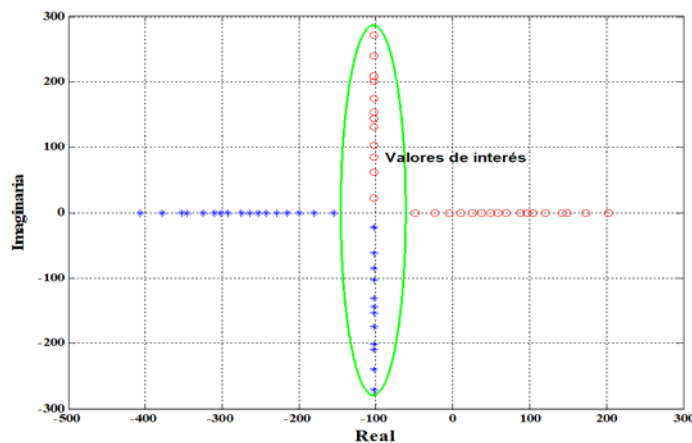


Figura 8 Reducción de intervalo variando k_3 .

Mapeo de Diversas Gráficas

Una vez identificados los posibles valores de k_3 los cuales optimizan el comportamiento con respecto a las condiciones originales del sistema de estudio, se procede a proponer un intervalo definido como el que se muestra en la figura 9, el cual será referencia de análisis para obtener la respuesta en el tiempo.

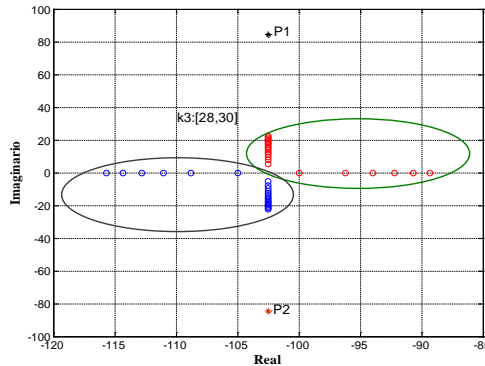


Figura 9 Identificación de valores óptimos de k_3 .

Obtención de Respuesta de Salida en el Dominio del Tiempo

Usando el intervalo identificado para $k_3 \in [28, 30]$ mediante el lugar geométrico de las raíces como se muestra en la figura 9 y mediante la elaboración diversos programas en lenguaje de Matlab se procesaron las señales de respuesta en el tiempo como se ilustra en la figura 10.

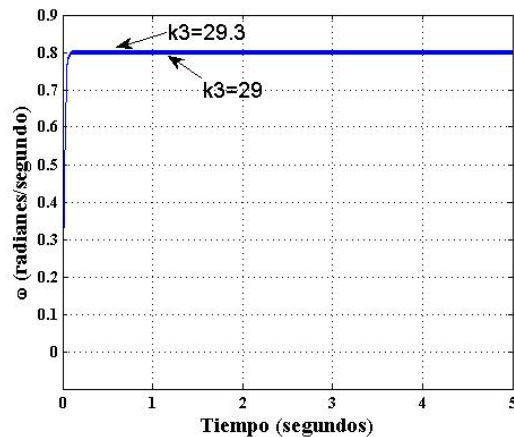


Figura 10 Respuesta en el tiempo de la ω en $k_3 \in [28, 30]$.

Para comprobar sobre la veracidad de los parámetros seleccionados para modificarse en las gráficas de las figuras 11 y 12 se comparan las respuestas de salida del modelo original contra el modelo con variación de parámetros con un valor fijo de $k_3=29$.

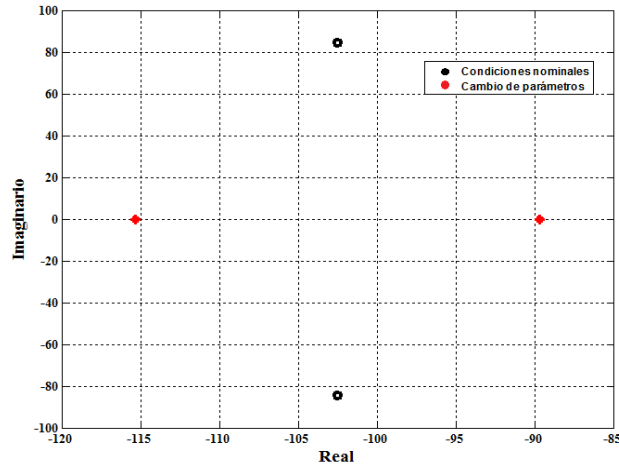


Figura 11 Ubicación de polos.

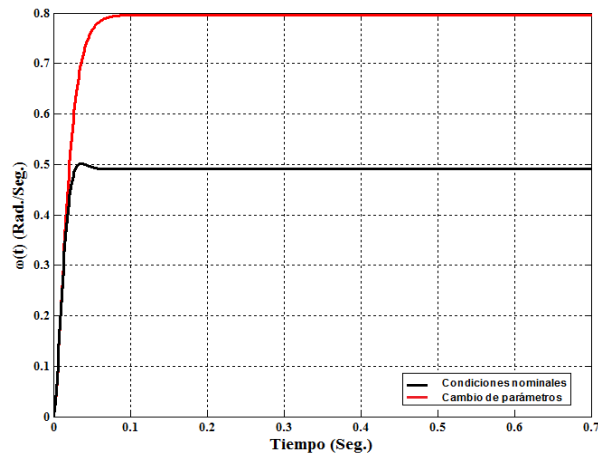


Figura 12 Respuesta en el tiempo de la ω .

4. Discusión

En este trabajo se modelo el motor de corriente CC, además se propone una metodología rápida y fácil de implementar basada en el método de estabilidad de Routh-Hurwitz y el lugar geométrico de las raíces para seleccionar parámetros dentro de la función característica que pueden ser modificados y optimizar la

funcionalidad del sistema original sin la implementación de un controlador adicional. En trabajos futuros se pretende incluir un modelo detallado del motor de CC considerando pérdidas en el hierro, pérdidas en el cobre, rendimiento de motor etc.

5. Conclusiones

Existen diferentes aplicaciones para el control de la velocidad de un motor CC, en la industria surgen diariamente los retos por gestionar y optimizar costos. En los tiempos llamados set off, y los horarios de menor demanda dentro de las empresas se deben generar las menores perdidas en cuanto a costo.

Este proyecto está planteado de manera práctica para conocer rápidamente la estabilidad del sistema y poder ubicar y seleccionar parámetros que fácilmente podrían cambiarse dentro de la dinámica original, también puede considerarse para la implementación de un controlador proporcional derivativo (PD) que permita cambiar automáticamente la velocidad del motor como el proceso lo requiera.

6. Bibliografía y Referencias

- [1] J. Dorsey, *Sistemas de control continuos y discretos*, Editorial Mc. Graw Hill, pp. 109, ISBN: 970104674-9, pp. 109, 2005.
- [2] J. Roldán, *Motores de corriente continua*, Ediciones Paraninfo, 2014, ISBN: 8428399018, pp. 218, 2014.
- [3] K. Ogata, *Ingeniería de control moderna*, Editorial Pearson, ISBN: 9701700481, 9789701700488, pp. 209, 2010.
- [4] R. Dorf, R. Bishop, *Sistemas de control moderno*, Editorial Pearson, España, ISBN: 978-84-205-4401-4, pp. 2, 8, 2008.
- [5] R. Navarro, *Ingeniería de control analógica y digital*, Editorial Mc. Graw Hill, ISBN: 970104677-3, pp. 101, 2004.
- [6] T. Wildi, *Maquinas eléctricas y sistemas de potencia*, Editorial Pearson, 2007, pp. 406, ISBN: 9702608147, 9789702608141, España, pp. 406, 2007.

DESIGN, CONSTRUCTION AND CHARACTERIZATION OF A THREE-CHANNEL COSMIC RAY DETECTOR BASED ON ALUMINUM BLOCKS ELECTRONICS

Luis Arceo

Universidad de Guanajuato, División de Ciencias e Ingenierías campus León
miquel@fisica.ugto.mx

Julián Félix

Universidad de Guanajuato, División de Ciencias e Ingenierías campus León
felix@fisica.ugto.mx

Resumen

Actualmente los físicos estudian e investigan los rayos cósmicos; en el mercado se puede adquirir la tecnología de detección de rayos cósmicos, pero con limitaciones en el espectro de detección y en la rapidez. El reto actual es diseñar, construir, y probar la electrónica de detección, y el sistema de adquisición de datos para adquirir continuamente (sin ventana de tiempo). En este trabajo reportamos el diseño, la construcción, la caracterización y las pruebas de las tarjetas electrónicas de un detector de rayos cósmicos de tres canales basado en tres barras de Aluminio de 2.54 x 5.08 x 20.32 cm (diseño novedoso y único) y en un fotodiodo Hamatsu MPPC, S12572-100P por canal. Para adquirir datos se usó el modelo compactRIO de National Instruments por su alta tasa de muestreo, embebido, y basado en tecnología FPGA. Se presentan detalles del diseño, de la construcción, y de la caracterización de la electrónica de detección y resultados físicos preliminares.

Palabras Claves: Embebido, FPGA, rayos cósmicos.

Abstract

Currently physicists study and research cosmic rays; cosmic ray detection technology can be purchased at the market, but with limitations in detection range

and speed. The challenge in these days is to design, build, and test the detection electronics and the data acquisition system for to continuously acquire data (no time window). In this work we report the design, construction, characterization and electronic board testing of a three-channel cosmic ray detector based on Aluminum blocks (innovative and unique design), each one attached to a MPPC, S12572-100P Hamamatsu photodiode. To acquire data it was used the National Instruments CompactRIO model for their high rate sampling, embedded, and based on FPGA technology. The details of the design, construction, and preliminary physical results are presented and discussed.

Keywords: *Cosmic ray, embedded, FPGA.*

1. Introduction

The scientists around of the world investigate cosmic rays; different types of cosmic rays detectors have been developed to achieve this goal [Morello, 2010]. The applications are mineralogy, spectroscopy, medicine (free X ray), radiography [Durham, 2015], security scanning, earthquake research, etc. In Mexico there are small scientific groups in experimental high energy physics. Particularly, the students and professors from Laboratory for elementary particles (laboratorio de partículas elementales [Félix, 2005]) design, construct, and run their own cosmic ray detectors [Félix, 2015].

The cosmic ray detector presented in this work is home designed and constructed with passive and active electronic parts to activate Hamamatsu avalanche photodiode S12572-100P [Hamamatsu, 2015], figure 1.

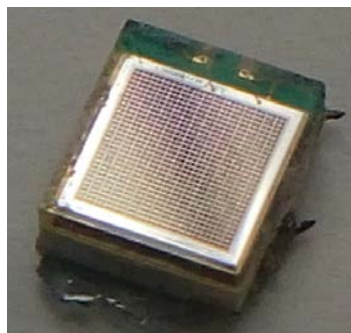


Figure 1 S12572-100P photodiode.

The data acquisition system selected was 9025 CompactRIO [National Instruments, 2015] cRIO of National Instrument, the core of this device is based with Field Programmable Gate Array "FPGA" technology.

2. Methods

In this work, were worked the following steps: design (electronics, using OrCAD), construction (local workshop), test (oscilloscope, LabVIEW and cRIO), characterization (power suppliers, LabVIEW and cRIO), and run (power suppliers, LabVIEW and cRIO).

Design

The one-channel cosmic ray detector is assembled with five stages and its material detection –Aluminum block for this report-, device sensitive to light -avalanche photodiode-, passive electronics board, discriminator board, and data acquisition system with PC-Host. The figure 2 displays the diagram block of one-channel cosmic ray detector.

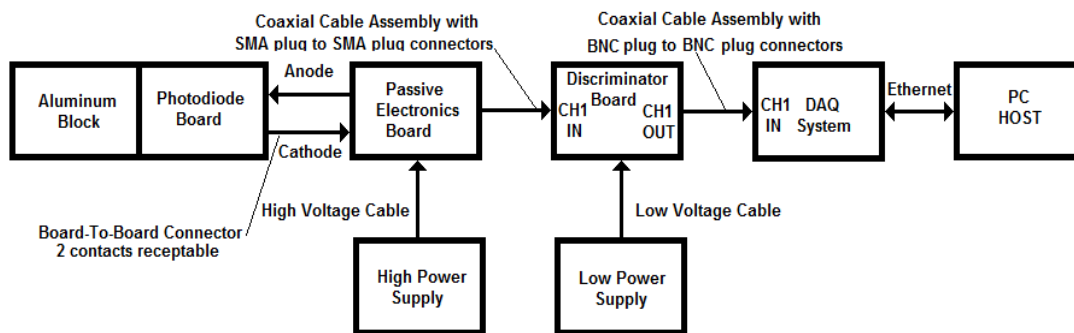


Figure 2 Diagram block of one-channel cosmic ray detector.

Electronic Boards

Photodiode board, the photodiode is soldered on top layer, and connection terminal anodes and cathodes are soldered on bottom layer. Its function is to attach photodiode to the Aluminum block on one of its polished 2.54 x 5.08 cm ends.

The figure 3 and 4 display the bottom and top layer, respectively.



Figure 3 Photodiode board bottom layer.



Figure 4 Photodiode board top layer.

The figure 5 displays photodiode board isolated optically with one layer of Aluminum tape 3311, the figure 6 displays one front end 2.54 x 5.08 cm Aluminum block polish and the figure 7 displays the Aluminum block attached with a photodiode board and isolated optically with four layers of Aluminum tape 3311.

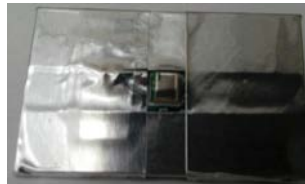


Figure 5 Photodiode board isolates optically.

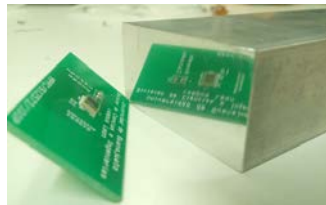


Figure 6 Aluminum block polished.



Figure 7 Aluminum block and photodiode isolated optically.

Passive electronic board, the photodiode is enabled by high voltage, and for the readout of the signal was implemented a RC circuit. The photodiode requires reverse polarity for working. The C3 capacitor is charged when the cosmic ray hits Aluminum block and generates Cerenkov radiation and photodiode detects it, and the C3 capacitor discharges through R7 resistor, analogue signal is obtained, positive decay exponential form is obtained. The readout signal is obtained at J1 SMA connector or a test point.

The figure 8 displays passive electronic board diagram schematic and figure 9 displays the passive electronic board.

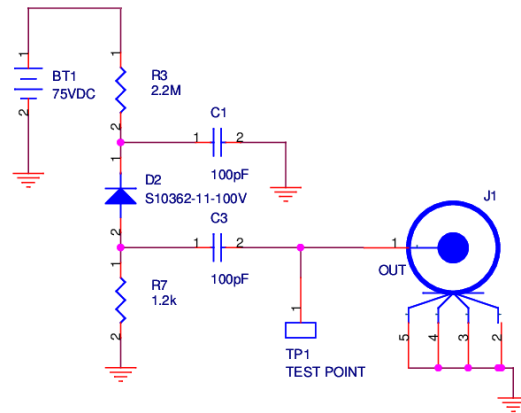


Figure 8 Passive electronic board diagram.

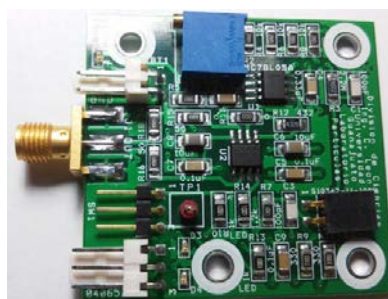


Figure 9 Passive electronic board with photodiode connector.

Discriminator board, its function is to compare analog signal coming from the passive electronic board, with a fixed trigger voltage defined by the final user and give out the digital signal. If the input signal is higher in amplitude than the fixed trigger, the discriminator turned on -one logical state-, otherwise, turns off -zero

logical state-. For this board was selected a single integrated circuit CMP401 of Analog Device [Analog Devices, 2002], with 23 ns propagation delay, quad comparators and compatible with 5 V logic. The figure 10 and 11 display the one channel discriminator board diagram schematic and four channel discriminator board, respectively.

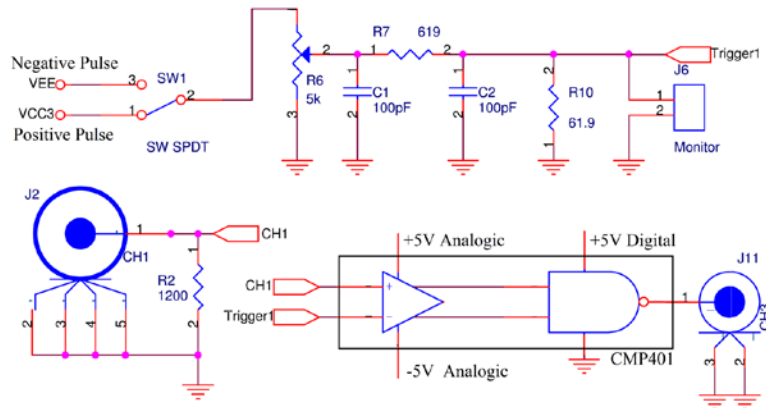


Figure 10 One channel discriminator board schematic diagram.



Figure 11 Four channel discriminator board.

The trigger voltage circuit is made up of three stages: First, polarity voltage selector -the SW1 switch selected only one voltage 3 or -5 V-. Second, voltage adjustment -R6 variable resistor-. Third, voltage divider by ten -divided voltage 3 or - 5 V, obtaining fine adjustment in mV-, and go to the comparator invert input, similarly, J2 SMA connector is for analogue signal from to passive electronic board to comparator non-invert input.

The trigger has wide input range positive of 0 volts to 300 or -470 mV, the end user can measured trigger voltage with J6 board-to-board connector and 2 contacts header.

The J11 BNC connector is for the digital signal output, obtained by comparison of the analogue signal input and trigger voltage defined by final user.

Data Acquisition System

The cRIO is assembled with 9025 embedded controllers, with NI-9402 four-channel, LVTTTL digital input/output module [National Instruments, 2016] and one rack. The figure 12 displays assembled cRIO. Finally, the cRIO is connected to the PC HOST via Ethernet cable with one number port.



Figure 12 Assembled cRIO with 32-channel digital input.

The cRIO requires programming in LabVIEW-FPGA and the PC HOST accept only LabVIEW program.

The LabVIEW-FPGA program reads for each channel the digital input signal every 25 ns from C module, submits the digital signal to an edge detector with rise polarity configured for searching digital rise pulse (the digital rise pulse is a possible cosmic rays). If the digital rise pulse is detected, the unsigned 64-bits (U 64-bits) counter increases by one and the digital rise pulse complement is discriminated (to avoid count the same digital signal), this is a parallel process and individual for each channel. The figure 13 displays a block diagram of LabVIEW-FPGA program.

Every 1 ms, LabVIEW-FPGA program copies the three counters results and the time in ms to Direct Memory Access First Input First Output (DMA FIFO) memory

with interleaving method, clears the three counters and LabVIEW-VHDL program begins again. Automatically, the DMA FIFO memory information is transferred for Ethernet port to PC HOST by embedded controller.

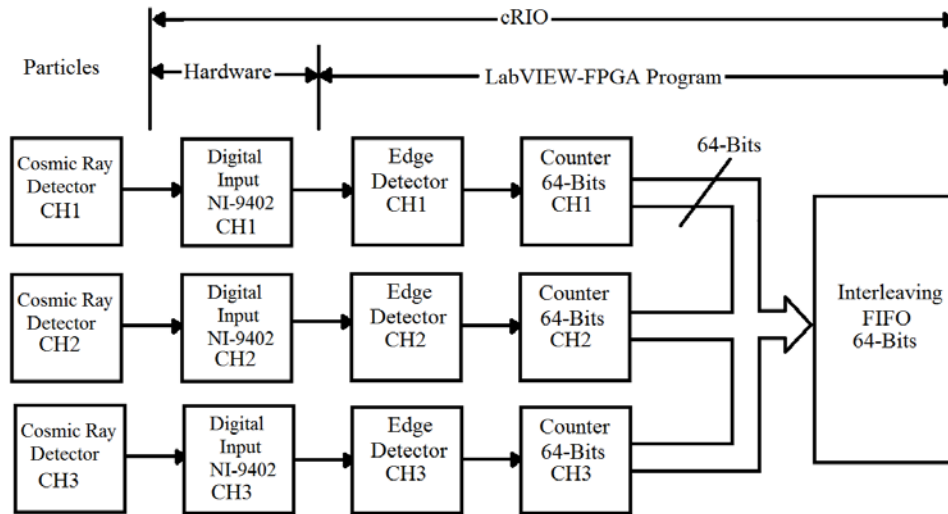


Figure 13 Block diagram of LabVIEW-FPGA program.

Construction

There were implemented three 2.54 x 5.08 x 20.32 cm Aluminum blocks, with only one 2.54 x 5.08 cm end polished to mirror for each Aluminum block. Each photodiode board was attached to the Aluminum block on the polished end, and isolated optically with four layers of Aluminum tape 3311.

A 25 x 35 cm Aluminum plate was used as a main board for some three Aluminum blocks and electronics assemble. On the main board was installed the first Aluminum block and it was connect with the first passive electronic board, it was fixed with four-inch screws and nuts. This process is repeated for the next two channels in stack arrangement and fixed only with nuts.

The discriminator board was fixed at the main board one edge. The top, middle and bottom output channels were connected to 1, 2 and 3 input channels discriminator board with SMA plug to SMA plug coaxial cable. The figure 14 displays the three channels detector final assembly.

The XLN10014 BK Precision and 72-8335A Tenma were used high and low power supply, respectively. The low and high voltage cables were implemented 22 AWG

size with wire-board receptacle MC34 series Molex, three contacts for low power and two contacts for high power.

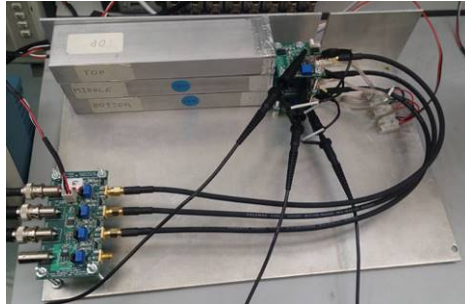


Figure 14 Three channel detector final assembly.

Test

The correct operation of each channel of the cosmic ray detector was verified using TDS1001C-EDU Tektronix oscilloscope. The oscilloscope channel one was connected to the passive probe to passive electronic board test point and oscilloscope channel two was connected to the BNC plug to BNC plug coaxial cable to discriminator board signal output. The figure 15 displays the oscilloscope test to three channel detector.

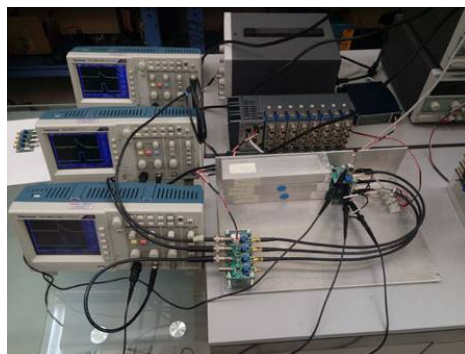


Figure 15 Oscilloscope test to three channel detector.

Characterization

The PC-host has a total control for automatic characterization process. The PC-Host uses a USB2.0 interface to control the high power supply -XLN10014 from BK Precision [BK PRECISION, 2014].

To define the high voltage range and the maximum current, to turn on the PC-Host program is necessary. The sequence is defined in five step: first, output turn on of the high voltage supply with the minimum defined voltage; second, to stabilize the new high voltage, a time delay of 10 seconds from the PC-Host is applied; third, a text file and records for ten minutes is created; fourth, the text file is closed; fifth, the PC-Host program checks if it has reached the maximum voltage defined. If it is true, the PC-Host terminates its execution and disables the output high power supply; if it is false, the PC-Host increases the voltage by 5 volts, and jumps to second step. The figure 16, displays hardware automatic characterization system blocks diagram.

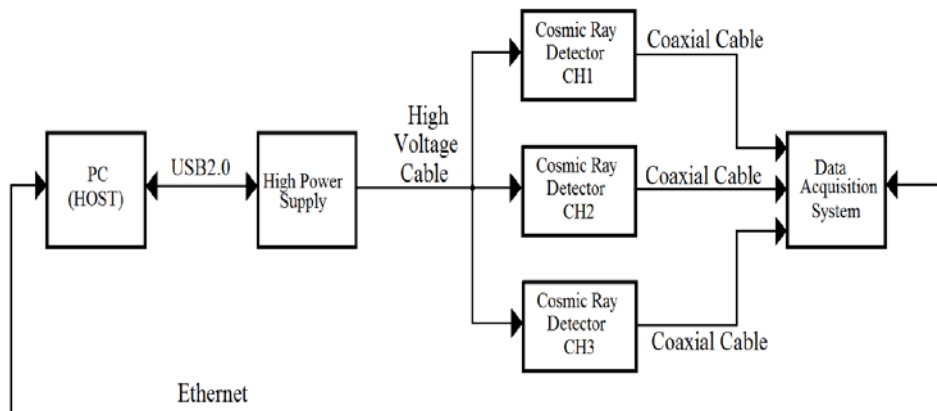


Figure 16 Characterization process block diagram.

3. Results

Test Results

The analogue signal noise was 20 mVpp and the configuration parameters were 75 Volts for operation voltage and 100 mV of threshold for the discriminator board. The analogue signal amplitude is variable, with duration of 200 ns pulse width, and exponential decay positive form.

The figure 17, 18 and 19 display top, middle and bottom Aluminum block output signal tests respectively. The oscilloscope channel one displays analogue output signal of a passive electronic board; the oscilloscope channel two displays the digital version of the analogue signal coming from discriminator board.

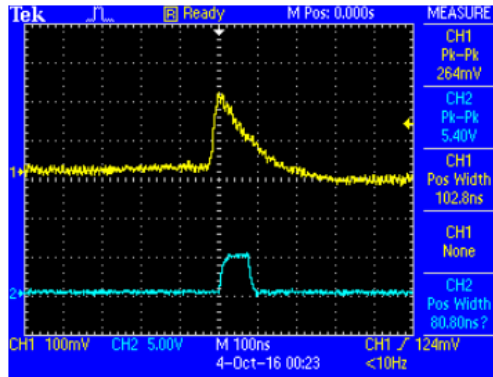


Figure 17 Top Aluminum block oscilloscope output signal test.

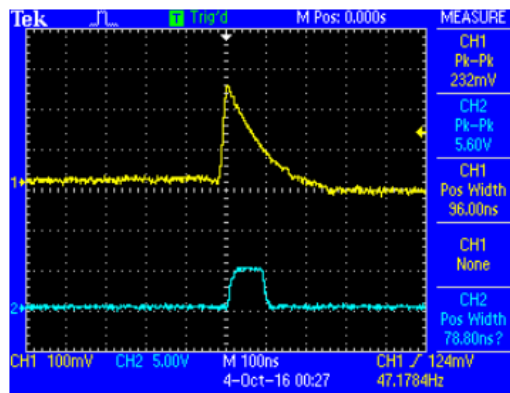


Figure 18 Middle Aluminum block oscilloscope output signal test.

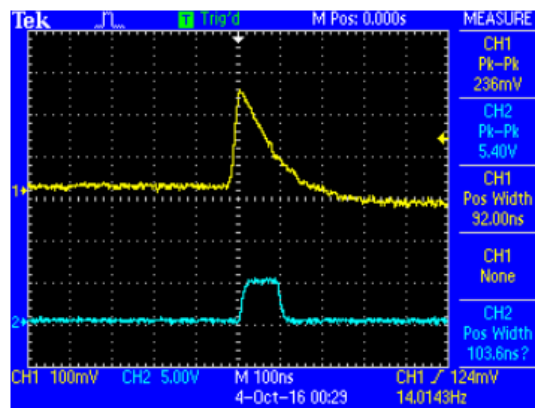


Figure 19 Bottom Aluminum block oscilloscope output signal test.

Characterization Results

The output counts as function of applied high voltage for each channel are displays in figures 20, 21 and 22. The recording interval began at 60 volts and

finished at 100 volts in steps of 5 volts each. Nine ten-minute-text files were generated. Figure 20 corresponds to the results of top Aluminum block. Figure 21 corresponds to the result of middle Aluminum block. And figure 22 corresponds to the result of bottom Aluminum block. For 75 volts and higher the number the counts increases lineally with applied voltage.

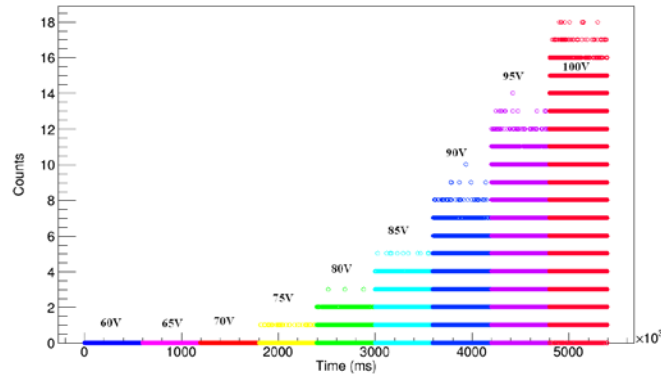


Figure 20 Top Aluminum block characterization result.

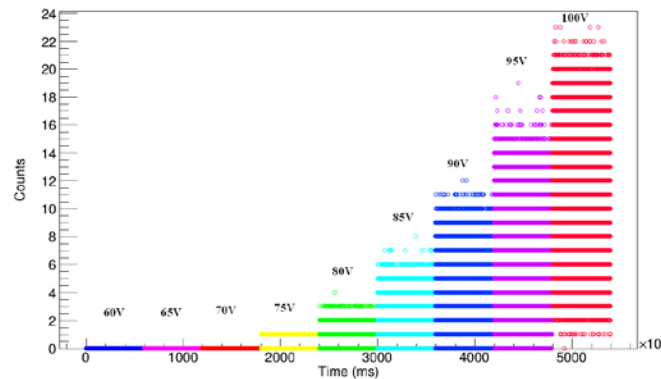


Figure 21 Middle Aluminum block characterization result.

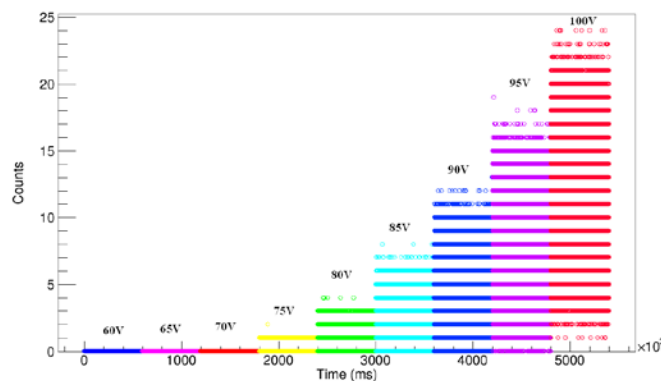


Figure 22 Bottom Aluminum block characterization result.

Results

The configuration parameters were 75 Volts for operation voltage and 100 mV of threshold for the discriminator board.

The figure 23, 24 and 25 display the number of counts vs time for top, middle and bottom channels, respectively. The figure 26, 27 and 28 display frequency vs counts histograms for top, middle and bottom channels, respectively. The recording time length was 30 minutes for each file.

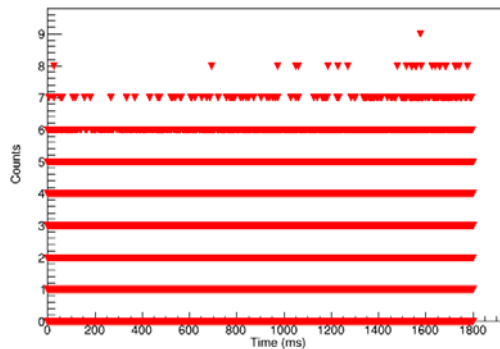


Figure 23 Counts vs time of top Aluminum block.

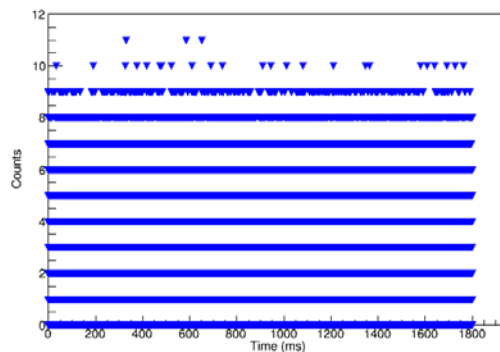


Figure 24 Counts vs time of middle Aluminum block.

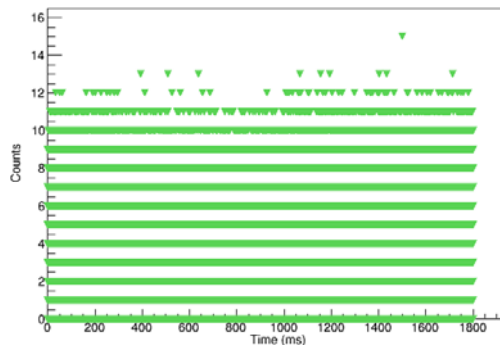


Figure 25 Counts vs time of Bottom Aluminum block.

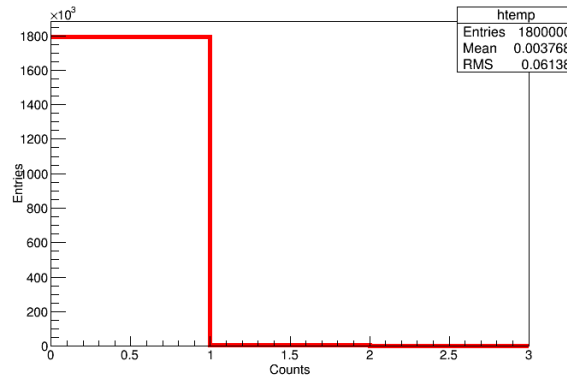


Figure 26 Frequency vs counts of top Aluminum block.

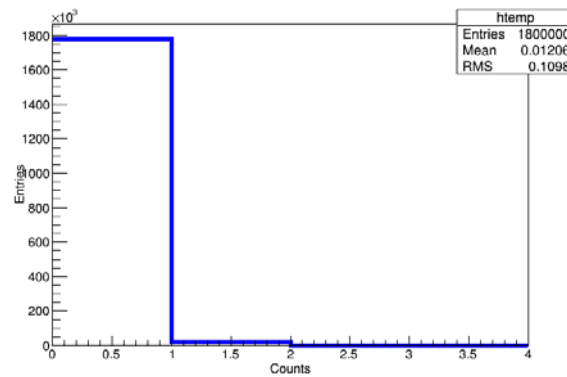


Figure 27 Frequency vs counts of middle Aluminum block.

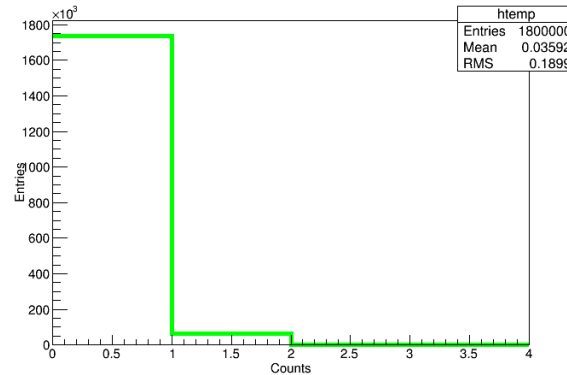


Figure 28 Frequency vs counts of bottom Aluminum block.

4. Discussion

We have planned, designed, constructed, tested and characterized a 3 Aluminum bar cosmic ray detector readout electronic board. It works fine. And the collected data, with a cRIO from NI, is reasonable good.

5. Conclusions

It was designed, constructed, tested, and characterized a three channel cosmic ray detector. The photodiode board, passive electronic board, discriminator board, the data acquisition system work properly. It was characterized the cosmic ray detector to obtain a linear function of the counts vs applied high voltage. The distributions of counts vs time are almost flat, for the three channels of the cosmic ray detector. The distributions of frequency vs counts are almost a Poisson distribution, for the three channels of the cosmic ray detector.

6. Bibliography and References

- [1] Analog Devices, datasheet, 2002, http://www.analog.com/media/en/technical-documentation/datasheets/CMP401_402.pdf.
- [2] BK PRECISION, user manual, 2014 https://bkpmedia.s3.amazonaws.com/downloads/manuals/en-us/XLN_manual.pdf.
- [3] Durham M, Guardincerri E, Morris C, Bacon J, Fabritius J, Fellows S, Poulson D, Plaud-Ramos K, and, Renshaw J, Tests of cosmic ray radiography for power industry applications, AIP Advances, Volume 5, Issue 6, May 2015.
- [4] Félix Julián, Design, construction, and operation of small cosmic rays detectors at Universidad de Guanajuato, Mexico, August 2016, <https://indico.cern.ch/event/432527/contributions/1072033/attachments/1321179/1981335/posterICHEP2016JFelix.pdf>.
- [5] Félix Julián, Laboratorio de Partículas Elementales, 2005, <http://laboratoriodeparticulaselementales.blogspot.mx/>.
- [6] Hamamatsu, datasheet, December 2015, <http://www.hamamatsu.com/jp/en/S12572-100P.html>.
- [7] Morello Carlo, Cosmic Ray Detectors, Proceedings of Science, PoS CRA School, 2010, http://inspirehep.net/record/912798/files/CRA%20School_027.pdf?.
- [8] National Instruments, datasheet, NI-9402, January 2016, http://www.ni.com/pdf/manuals/374614a_02.pdf.

DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA DE MEDICIÓN COMBINADA DE 4-PUNTAS

Francisco Javier Arizaga Ayala

Universidad de Sonora
farizagaayala@gmail.com

Arturo III Espinoza Duarte

Universidad de Sonora
arturoiii@gmail.com

José Antonio Gallardo Cubedo

Universidad de Sonora
a_n_t_1212@hotmail.com

Armando Gregorio Rojas Hernández

Universidad de Sonora
arojas@cifus.uson.mx

Resumen

Se ha desarrollado un sistema con diseño automatizado por medio de software para la medición de resistividad de materiales, en este caso semiconductores, por medio de la técnica de 4 puntas. El método Rymaszewski es utilizado para eliminar la dependencia geométrica de las muestras y sólo configurar la posición de las puntas conforme al método. Los resultados de resistividad obtenidos para las muestras CdS, PbS y ZnO son 11.8×10^6 , 7.7×10^6 y 5.3×10^6 $\Omega \cdot \text{cm}$, respectivamente. Las películas han sido crecidas por síntesis y con grosores diferentes lo cual resultará en pequeñas pequeñas diferencias entre los resultados obtenidos y los resultados en la literatura: la síntesis del CdS y PbS ha sido por el método de depósito en baño químico (CBD, por sus siglas en inglés) y el ZnO por el método de depósito de capas atómicas (ALD, por sus siglas en inglés).

Palabras Claves: Cuatro puntas, Método Rymaszewski, resistividad, semiconductor.

Abstract

A software-automated system has been developed to measure the materials resistivity, in this case semiconductors, by the four point-probe technique. We neglected the sample geometry by using the Rymaszewski method but positioned the probes carefully according to the method. The resistivity results obtained for the samples CdS, PbS and ZnO are: 11.8×10^6 , 7.7×10^6 & 5.3×10^6 $\Omega \cdot \text{cm}$, respectively. The thin films have been grown by different synthesis and thickness that means that the resistivity results will show some differences between the results obtained and the results in the literature: The CdS y PbS samples was grown by Chemical bath deposition (CBD) method and the ZnO sample was grown by Atomic Layer Deposition (ALD) method.

Keywords: *Four point-probe, Rymaszewski method, resistivity, semiconductor.*

1. Introducción

El avance de dispositivos a base de película delgada como transistores de efecto de campo metal-óxido-semiconductor (MOSFET, por sus siglas en inglés), algunos otros basados en tecnología de integración a gran escala (VLSI, por sus siglas en inglés) ha impulsado el desarrollo y mejoramiento de los métodos de medición de parámetros de semiconductores, tal como la técnica de 4-puntas, método utilizado en la manufactura de materiales semiconductores, implementándose en procesos de caracterización [Dieter, 2006].

La medición de resistividad de materiales semiconductores se puede realizar con la técnica de 4-puntas [Haibin, 2010], [Rymaszewski, 1969]. El método de 4-puntas depende de aspectos como la geometría de la muestra, la posición de las puntas y la separación de puntas entre ellas. Estas condiciones afectan la calidad de la medición en la muestra, dónde el parámetro de factor de corrección se ve influenciado. Sin embargo, se tienen modificaciones en la técnica que garantizan la independencia de algunos de estos factores como la geometría de la muestra. Se recurre al método de Rymaszewski [Haibin, 2010], [Rymaszewski, 1969] que implementa una combinación de mediciones, modificando la inyección de la corriente con la medición del voltaje entre las puntas.

Las resistencias de contacto pueden ser un problema serio cuando se hace contacto eléctrico utilizando los micros manipuladores puesto que el área de contacto es muy pequeña. Rymaszewski describe un método experimental de calibración lineal de un micro-arreglo de distribución de electrodos (cuatro puntas) [Rymaszewski, 1969], ver la figura 1:

- Las puntas están en contacto en un arreglo de línea recta en la muestra.
- Las puntas están distribuidas arbitrariamente a lo largo de la línea perpendicular al límite no conductor en el medio plano.
- Las puntas están distribuidas a lo largo del límite no conductor del medio plano.

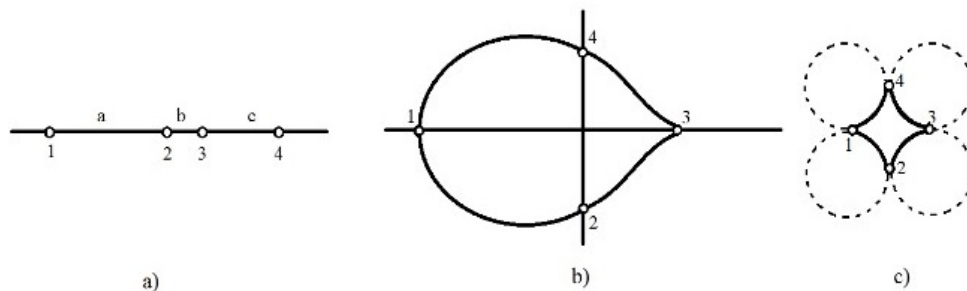


Figura 1 Posición de puntas para diferentes arreglos en el método Rymaszewski.

En este trabajo, se lleva a cabo el método de Rymaszewski mediante el diseño de un sistema de medición utilizando un Keithley 2400 SourceMeter, controlado a través de instrumentación virtual con el software LabVIEW y una tarjeta DAQ NI 6009 como dispositivo de interfaz y circuitería externa para la conmutación de las puntas haciendo uso del circuito integrado CD4052.

2. Métodos

Principio de Funcionamiento

La resistividad puede afectar la resistencia serie de un dispositivo, voltaje de umbral, capacitancia y otros parámetros [Keithley, 2017]. La resistividad se puede representar con ecuación 1.

$$\rho = R_s \cdot \delta \cdot F \quad (1)$$

Donde R_s es la resistencia de hoja, δ es el grosor de la muestra y F es el factor de corrección para el grosor. Para R_s se tiene:

$$R_s = \frac{\pi}{\ln 2} \cdot \left(\frac{V_1 + V_2}{I} \right) \cdot f \left(\frac{V_2}{V_1} \right) \quad (2)$$

Donde $f(V_2/V_1)$ es la función de Van der Pauw, que es una función implícita.

El diseño propuesto del sistema de medición de resistencia de hoja R_s y resistividad ρ consiste del método de cuatro puntas haciendo uso, a la vez, del método de Rymaszewski que consiste de dos medidas: para la primera medida, la corriente I pasa por la punta 1 y sale por la punta 4, el voltaje V_1 se toma de las puntas 2 y 3. Posteriormente se toma una segunda medida donde la corriente I pasa por las puntas 1 y 2 y el voltaje V_2 es medido entre las puntas 3 y 4, como se muestra en la figura 2.

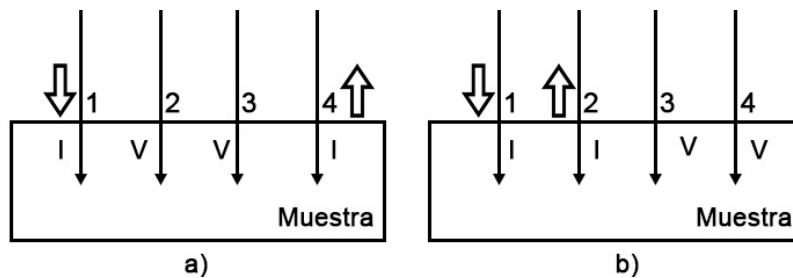


Figura 2 Arreglo de las puntas para el sistema de medición de resistividad de 4-puntas.

3. Resultados

Trabajo Experimental

El sistema desarrollado tiene como objetivo medir la resistividad de distintas muestras en función de la corriente. Se realizan mediciones de películas delgadas de sulfuro de cadmio (CdS), sulfuro de plomo (PbS) y óxido de zinc (ZnO) mediante una configuración de arreglo lineal en la posición de las puntas como se observa en la figura 2.

En la figura 3 se tiene un esquema del sistema que se compone de una PC con LabVIEW 2015 Windows 2010 como software controlador para el Keithley 2400. Como comandos de control se utiliza el diagrama de la figura 5. En el sistema,

originalmente se tiene una configuración para mediciones I-V, compuesto de dos puntas, por lo que se recurre a construir dos estructuras para soporte de las dos puntas restantes. Mediante el uso de un centro de maquinado CNC se fabricaron dos estructuras de aluminio 6061 para estación de prueba “Probing Solution Inc.” modelo EMZ-5TR. Las dimensiones de la estructura son 38.1 milímetros por 38.1 milímetros de base por 115.824 milímetros de largo, cuenta con una cavidad de 115.824 milímetros por 48.006 milímetros y una profundidad de 20.32 milímetros de profundidad. Estas medidas se tomaron con respecto a la altura de la mesa de la estación de pruebas. En la figura 4a se observa el diseño de la estructura fabricada en el centro de maquinado CNC, y en la figura 4b se muestra cómo se adapta la estructura al mecanismo de posicionamiento de la punta de la estación.

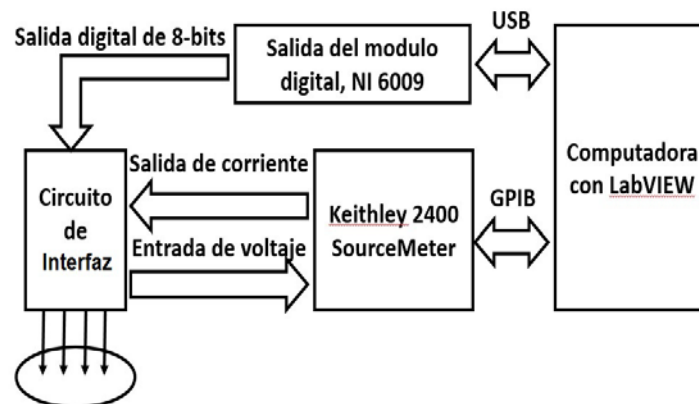


Figura 3 Esquema del sistema de medición combinada de 4-puntas.

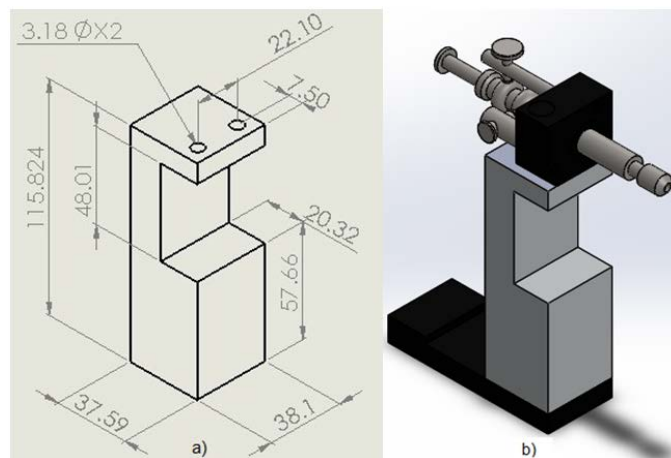


Figura 4 Estructura fabricada para soporte de puntas de medición, estructura ensamblada con el mecanismo de posicionamiento de la punta.

El modelo 2400 de Keithley es un sistema diseñado para poder excitar mediante corriente o voltaje a un dispositivo bajo prueba y posteriormente poder medir su respuesta en corriente o voltaje de manera automatizada mediante algún dispositivo controlador en este caso una PC, por lo cual usaremos en adelante esta definición como (fuente-medidor).

Para el diseño de un sistema fuente-medidor de corriente-voltaje utilizando Keithley 2400 se recurre al manual del mismo equipo para conocer los comandos y su funcionamiento. Los comandos necesarios para el uso del Keithley 2400 para ser utilizado como fuente-medidor observan en la tabla 1.

Tabla 1 Comandos de programación de Keithley 2400.

N°	Acción	Comandos	Comentarios
1	Selección de medidor; función de fuente	:SOUR:FUNC CURR :SOUR:CURR:MODE:FIXED	Función fuente de corriente Configura modo fuente corriente
2	Selección de fuente y compliancia	:SENS:FUNC "VOLT" :SOUR:CURR:RANG MIN :SOUR:CURR:LEV 0	Medición de voltaje Rango mínimo de fuente
3	Selección de medición de rango de voltaje	:SENS:VOLT:PROT 25 :SENS:VOLT:RANG 20 :FORM:ELEM:VOLT	Nivel de fuente Selección de compliancia 25V Rango de voltaje 20V Lectura solo de voltaje
4	Encender salida	:OUTP ON	Salida encendida antes de medir
5	Leer datos	:READ?	Disparo, obtención de lectura
6	Apagar salida	:OUTP OFF	Salida apagada después de medir

Para realizar correctamente el método Rymaszewski, se adapta un código para automatizar la medición para una primera configuración como se menciona en el apartado *Principio de funcionamiento*. El circuito multiplexa las puntas para que ahora la corriente entre por 1 y salga por 2, mientras la medición de voltaje se realiza entre 3 y 4. La función de multiplexar las puntas con la que cuenta el

circuito integrado CD4052 se realiza mediante la tarjeta de National Instruments modelo 6009, en la cual se envían 6 señales digitales codificadas para cada una de las configuraciones. Se crea un código que realiza la función de incrementos iguales de corriente, desde un punto inicial a un punto final especificados, leyendo en cada uno de ellos el voltaje medido para cada muestra. Posteriormente, los resultados son procesados como se indica en la ecuación 1 y ecuación 2, figura 5.

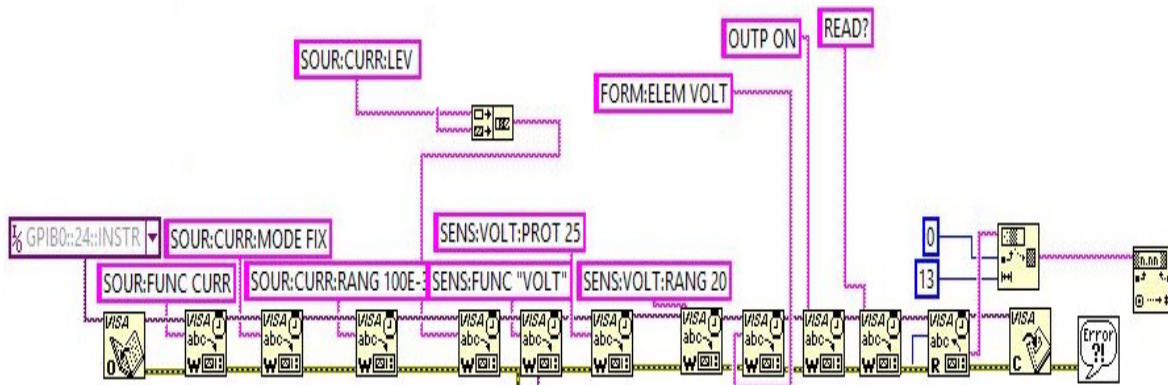


Figura 5 Código de LabVIEW, código en lenguaje para realizar un fuente-medidor.

En la figura 6 se tiene una representación del panel de control de LabVIEW. En él se especifican las condiciones para la inyección de corriente en la muestra para medir el voltaje y obtener curvas I-V y posteriormente obtener la resistividad del material.

La figura 7 muestra los resultados de resistividad para CdS, PbS y ZnO.

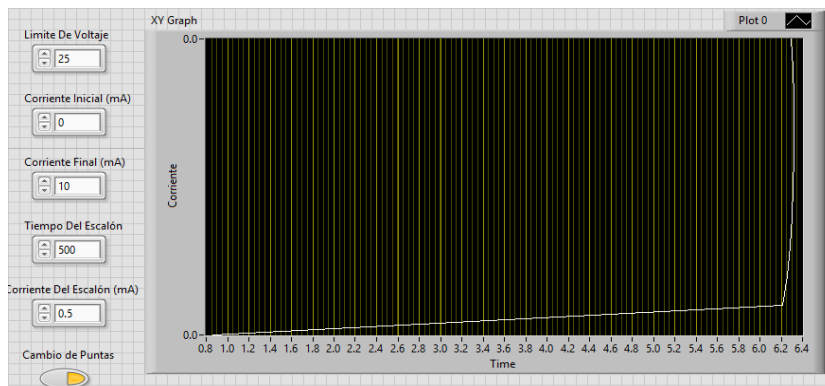


Figura 6 Panel de control LabVIEW con especificaciones voltaje y corriente para sistema

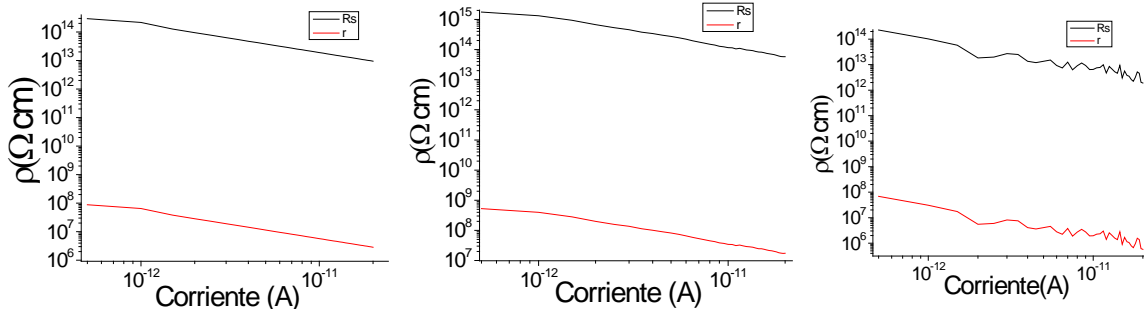


Figura 7 Comparación de resultados de resistividad (ρ) muestras: CdS, PbS y ZnO.

4. Discusión

Tres diferentes muestras se someten a medición en el sistema propuesto. La tabla 2 muestra la información obtenida de las tres muestras medidas y la tabla 3 resultados de diferentes autores que usaron diferentes síntesis y grosores para los mismos materiales. Se realiza un barrido de corriente de 0 a 20 pA, obteniendo los diferentes voltajes para cada muestra, implementando estos resultados para calcular la resistividad a partir de la ecuación 1 y ecuación 2. Se observa que los comportamientos de las mediciones de resistividad de las tres muestras presentan leves fluctuaciones y cambios poco representativos en comparación, así como también se tienen rangos de medición en el orden de $10^6 \Omega \cdot \text{cm}$ con el sistema aquí desarrollado, *Chemical Spray Pyrolysis (CSP).

Tabla 2 Información de las muestras de películas delgada de CdS, PbS y ZnO.

Muestras propias			
Muestra	Síntesis	Grosor (nm)	Resistividad promedio ($\Omega \cdot \text{cm}$)
CdS	CBD	300	11.8×10^6
PbS	CBD	300	7.7×10^6
ZnO	ALD	50	5.3×10^6

Tabla 3 Muestras de película delgada de CdS, PbS y ZnO encontradas en la literatura.

Otros autores		
Síntesis	Grosor	Resistividad promedio ($\Omega \cdot \text{cm}$)
CSP*	---	Rango de 10^3 y 10^5 [Santiago, 2008]
CSP	300nm	1.1×10^3 [Ghaffar, 2015]
PEALD	23nm	5.9×10^3 [Jian, 2013]

5. Conclusiones

Se desarrolló un sistema automatizado de la técnica de 4 puntas (Four-Point Probe) por medio de software para hacer medidas de resistividad en semiconductores. En conjunto con la técnica de cuatro puntas se usa el método de Rymaszewski para eliminar dependencias geométricas de las muestras y sólo tomar en cuenta el posicionamiento de las puntas. El valor del factor de corrección F de la ecuación 1 está relacionado con la tasa entre el grosor de la película y el espaciamiento entre las puntas. Cuando la tasa es menos que 0.6 entonces $F=1$. Por facilidad de cálculos y métricas se toma el factor de corrección F igual a 1. Los resultados obtenidos de resistividad de las mediciones son 11.8×10^6 , 7.7×10^6 y $5.3 \times 10^6 \Omega \cdot \text{cm}$.

Se observa que en promedio la resistividad de la muestra de ZnO es menor en comparación a las otras muestras, esto es debido a que el grosor es un factor clave en la ecuación 1, esperando entonces mayor resistividad en películas gruesas.

Se comparan los resultados obtenidos con resultados de diferentes trabajos los cuales muestran relativa diferencia debido a las diferentes técnicas de síntesis de películas delgadas usadas y distintos grosores. Sin embargo, el comportamiento de resistividad es similar en todos los casos.

6. Bibliografía y Referencias

- [1] Ghaffar Faraj M., Effect of Thickness on the Structural and Electrical Properties of Spray Pyrolysed Lead Sulfide Thin Films. *American Journal of Condensed Matter Physics*, 2015.
- [2] Haibin Pan, Jianning Ding, et.al., Design of thin film resistivity measurements system based on virtual instrumentation technology. *Journal of Jiangsu University, Natural science edition*, 2010.
- [3] Haibin Pan, Boquan Li, et.al., Design, Implementation, and Assessment of a High-precision and Automation Measurement System for Thin Film Resistivity. Department of Measurement and Control Technology, 2010.

- [4] Dieter K. Schroder, *Semiconductor Material and Device Characterization*. Wiley IEEE Press, 2006.
- [5] Jian Zhang, Hui Yang, Qi-long Zhang, Shurong Dong, J.K. Luo, *Structural, optical, electrical and resistive switching properties of ZnO thin films deposited by thermal and plasma-enhanced atomic layer deposition*. ELSIEVER, 2013.
- [6] Rymaszewski R., Relationship between the correction factor of the four-point probe value and the selection of potential and current electrodes. *Journal of Scientific Instruments*, Vol. 2, 1969.
- [7] Santiago Tepantlán C., Structural, optical and electrical properties of CdS thin films obtained by spray pyrolysis. *Revista Mexicana de física*, Vol. 2, 2008.
- [8] Measuring the Resistivity and Determining the Conductivity Type of Semiconductor Materials Using a Four-Point Collinear Probe and the Model 6221 DC and AC Current Source: www.keithley.com.

COMPARACIÓN DE TARJETAS ARDUINO UNO ORIGINALES Y CLONES COMO INSTRUMENTO DE MEDICIÓN

Miguel Angel Bañuelos Saucedo

Universidad Nacional Autónoma de México

miguel.banuelos@ccadet.unam.mx

Resumen

La tarjeta Arduino UNO es una opción económica de uso muy difundido entre aficionados y expertos en electrónica. Una de sus principales aplicaciones es la de realizar mediciones de variables físicas, utilizando la gran cantidad de sensores analógicos disponibles para esta plataforma. En el mercado se consiguen versiones originales, y clones de la tarjeta Arduino. Este trabajo compara ambos tipos, primero evaluando los niveles de ruido presentes al utilizar como fuente de alimentación el puerto USB de una computadora portátil. Posteriormente, se compara la estabilidad de la salida del convertidor-analógico digital ante una entrada constante, utilizando la referencia de voltaje propia de la tarjeta y una referencia externa de 4.096 V. Finalmente se analiza la operación del convertidor, con su referencia de voltaje, cuando se alimenta la tarjeta con una pila de respaldo de celular, y una batería alcalina de 9 V. Los resultados muestran que no existen diferencias fuera de las especificaciones, entre mediciones realizadas con tarjetas originales y clones.

Palabras Claves: Arduino UNO, conversión analógica-digital, instrumentación electrónica.

1. Introducción

En el mercado hay una gran variedad de tarjetas de desarrollo basadas en microcontrolador, dentro de ese conjunto, las tarjetas Arduino han crecido en popularidad debido a que el hardware es de diseño abierto, y cuentan con

herramientas de programación de uso libre compatible con los sistemas operativos Windows, Linux y Mac OS X. Existen diferentes modelos de tarjeta Arduino, los cuales están basados en diferentes procesadores, y por lo tanto, cuentan con diferentes capacidades de cómputo y de dispositivos periféricos; sin embargo, todos ellos son programables con el mismo lenguaje y ambiente de desarrollo. Una de las tarjetas más populares es la Arduino UNO, debido a su bajo costo. Las tarjetas Arduino cuentan con un arreglo de conectores que se han convertido en un estándar en la industria. Varias compañías han desarrollado módulos de expansión compatibles con estos conectores, y la popularidad de la familia ha hecho que tarjetas de desarrollo de otras arquitecturas también cuenten con módulos compatibles con Arduino. El éxito comercial de la arquitectura Arduino ha hecho que varios fabricantes hayan desarrollado copias compatibles (clones), que se comercializan a precios inferiores al modelo original. Existen varias tarjetas clones, de procedencia mayormente china, donde el fabricante no está especificado, y diversos modelos que utilizan sus propios nombres comerciales, pero que son funcionalmente compatibles con la tarjeta Arduino UNO, tales como: Diavolino, Freeduino, Nanode, Freakduino-Chibi, Illuminato, Boarduino, Geekcreit UNO, Sparkfun RedBoard, RobotDyn UNO, etc [Addicore, 2017], [Geekcriet, 2017], [Torrone, 2012].

Una de las consecuencias más importantes de la popularidad de la plataforma Arduino es que la comunidad de desarrolladores ha elaborado docenas de bibliotecas, que facilitan la utilización de una gran variedad de sensores. Es por ello que resulta atractivo utilizar una tarjeta Arduino para desarrollar un sistema de medición, y el presente trabajo se centra en la evaluación de la tarjeta Arduino UNO original y una versión clon, como un sistema de medición de una señal analógica.

La tarjeta Arduino UNO está basada en un microcontrolador de 8-bits ATmega328 de Atmel, el cual cuenta con 32 kB de memoria Flash, 2 kB de memoria SRAM, 1 kB de memoria EEPROM, 14 líneas de entrada/salida, y opera con un reloj de 16 MHz [Arduino, 2017a]. También cuenta con un convertidor analógico-digital de 10 bits con 6 canales de entrada. Este convertidor es la base del proceso de medición

de todos los sensores que producen una señal analógica. La conversión a 10 bits significa que se cuenta con 1024 niveles de cuantización, lo que podemos aproximar a una medición de una parte en mil. Esta resolución puede ser útil para muchos sistemas, por ejemplo, para la medición de temperatura y humedad ambiente.

2. Métodos

Dada la existencia en el mercado de versiones originales y clones de la tarjeta Arduino UNO (figura 1), se decidió evaluar el desempeño de ambas, en diferentes condiciones de operación y utilizando diferentes suministros de energía. Para la comparación se utilizan dos tarjetas Arduino originales y dos clones.

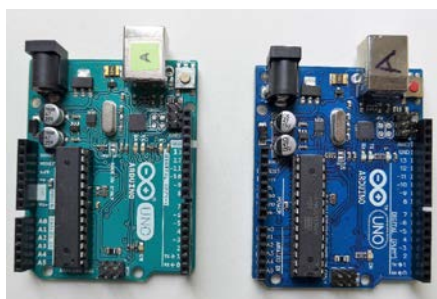


Figura 1 Tarjeta Arduino UNO original (a la izquierda) y clon (a la derecha).

La tarjeta Arduino UNO se basa en el microcontrolador Atmel ATmega328P. Su convertidor analógico digital (CAD) cuenta con una terminal independiente para suministro de voltaje de alimentación (AV_{CC}). Por omisión, el voltaje de referencia para el CAD se toma de la terminal AV_{CC} , y en el caso de la tarjeta Arduino UNO, dicha terminal está conectada directamente a la alimentación de +5 V [Arduino, 2017b]. La tarjeta puede obtener los 5 V de alimentación a través de la conexión USB, o mediante un conector para eliminador de baterías, que acepta voltajes de 7 a 12 VCD y que utiliza un regulador de voltaje lineal para producir +5 V. Por lo anterior, existe una clara dependencia entre el voltaje de alimentación de la tarjeta y la operación del CAD, ya que dicho voltaje puede contener un nivel de ruido suficiente para afectar el proceso de cuantización.

Para evaluar la influencia del voltaje de suministro en la operación del CAD, se midió el nivel de ruido presente en las conexiones de alimentación de las tarjetas Arduino UNO, tanto originales como clones, cuando éstas eran alimentadas mediante una conexión USB a un puerto de una computadora portátil (Lenovo ideapad 510S). Este tipo de computadoras utiliza una fuente conmutada para la generación del voltaje de 5 V, por lo que no es una fuente de bajo nivel de ruido. Para las mediciones se empleó un multímetro de banco GW Instek modelo GDM-8351, que en su escala de 10 V cuenta con una resolución de 100 μV , una exactitud de 0.012 % +5 dígitos y una impedancia de entrada de 11.1 M Ω . Para esta prueba se cargó en el Arduino un programa vacío, lo que evita cambios en el consumo de corriente del microcontrolador. La medición se realizó utilizando cable coaxial para reducir el acoplamiento de ruido y en un área libre de tráfico de personas, para evitar que la electricidad estática de personas moviéndose en las proximidades de las tarjetas pudiera afectar los registros.

Posteriormente, para evaluar el desempeño del convertidor analógico-digital de la tarjeta, se suministró una señal de corriente directa constante al canal A0. Esta señal fue generada mediante un divisor de voltaje resistivo, a partir del voltaje de alimentación de la tarjeta Arduino, figura 2.

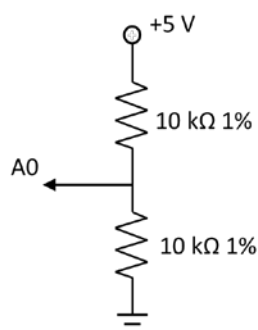


Figura 2 Divisor resistivo.

Se considera que el divisor resistivo presenta un caso general, ya que el voltaje que produce queda sujeto a las variaciones de la fuente de alimentación, lo cual suele ocurrir con algunos sensores analógicos, por ejemplo, termistores, fotoresistencias, etc. En el divisor se utilizan resistencias de 10 k Ω , de película

metalizada y 1% de tolerancia. Este tipo de resistencias tienen una menor deriva térmica, con relación a las resistencias de carbón. Las mediciones se realizaron en un ambiente estable térmicamente, donde la temperatura no cambió más de 0.5 °C. Las resistencias utilizadas tienen una deriva térmica de 100 ppm/°C; sin embargo, su cercanía física permite considerar que se encuentran a la misma temperatura, por lo que las variaciones de voltaje en la terminal A0 debidas a los cambios de temperatura mencionados se vuelven indetectables en la escala a la que fueron realizadas las mediciones. Se selecciona el valor de 10 kΩ, para no producir un exceso de carga de corriente sobre la fuente y no contribuir mediante efecto Joule a su calentamiento.

Se evaluaron dos casos donde la tarjeta se alimenta mediante conexión USB a la computadora portátil, primero utilizando la referencia de voltaje de la tarjeta (5 V), y luego una referencia externa de 4.096 V (el circuito integrado MCP1541). También se consideró el caso en el que el divisor resistivo se encontraba 15 cm alejado de la tarjeta, pero utilizando la referencia de la tarjeta (5 V). Finalmente, se comparan los resultados que entrega el CAD al utilizar la referencia de la tarjeta, pero usando como alimentación la pila de respaldo de teléfono celular y la batería de 9 V. La pila de respaldo para teléfono celular es una opción útil en los casos en que se quiere realizar alguna medición sin depender de la conexión a una computadora. Por otro lado, una batería de 9 V constituye una fuente de voltaje de bajo nivel de ruido. El análisis consistió en adquirir conjuntos de 500 o 1000 datos, y mostrar el histograma resultante. Como el voltaje que se aplica es constante, se espera que el resultado que entrega el CAD también lo sea. Mientras mayor sea el número de datos que se desvían de la moda estadística, menor será la confiabilidad de la medición.

3. Resultados

La primera prueba consistió en medir el voltaje de operación de la tarjeta Arduino mientras era alimentada por la computadora portátil (Lenovo ideapad 510S). Se tomaron 1000 registros para cuatro tarjetas distintas, dos tarjetas originales y dos tarjetas clones. Los registros se realizaron mediante un multímetro

GW Instek, modelo GDM-8351 (figura 3). La toma de datos se automatizó utilizando el software proporcionado por el fabricante. Los resultados se muestran en la tabla 1.

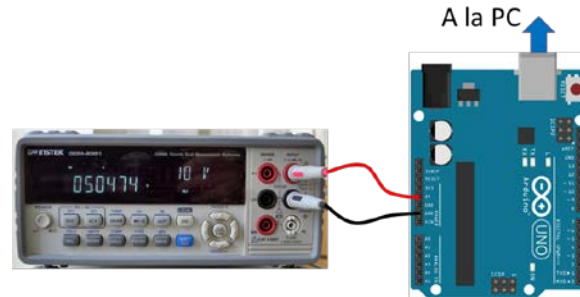


Figura 3 Evaluación del voltaje de alimentación proporcionado mediante la conexión USB con una computadora portátil.

Tabla 1 Parámetros de operación de la fuente de una tarjeta Arduino UNO.

Parámetro	Tarjeta Arduino UNO			
	Original A	Original B	Clon A	Clon B
Media	5.048196	5.0403232	5.0441606	5.0406187
σ	0.000181	0.0001726	0.0001118	0.0003898
máx	5.0494	5.0409	5.0445	5.0436
mín	5.048	5.0401	5.0439	5.0395
SNR _{dB}	88.92	89.31	93.08	82.23

El multímetro entrega datos con 4 decimales, por lo que los valores promedio observados en la tabla no deben considerarse con una precisión mayor; sin embargo, se anotan con 7 decimales, pues es el valor que se utilizó para el resto de los cálculos. Para las cuatro tarjetas el voltaje de alimentación es el mismo, pero se observan pequeñas variaciones entre cada una de las tarjetas. Esto se explica porque las mediciones se realizan mientras la tarjeta está operando, por lo tanto, la influencia de la calidad de los componentes y su montaje afecta el voltaje de la fuente. En todos los casos, la variación entre los valores mínimos y máximo del voltaje de alimentación es menor a 1 mV. Dado que la resolución del convertidor analógico-digital es de 5 mV, estas variaciones son inferiores al error de cuantización del convertidor de 10 bits. Adicionalmente se determina la relación

señal-a-ruido (SNR, *signal-to-noise ratio*) como parámetro de comparación. En este caso, las tarjetas originales presentan un valor muy similar y cercano a los 89 dB, mientras que las versiones clones tienen una mayor variabilidad, que oscila de 82 a 93 dB. La SNR fue calculada utilizando ecuación 1.

$$SNR = 20 \log_{10} \frac{\mu}{\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}} \quad (1)$$

Donde μ es la media del conjunto de datos y N es el número de datos.

Se ha establecido que las variaciones en el voltaje de la fuente de alimentación (proveniente de una conexión USB a una computadora personal) son menores a la resolución del CAD. Ahora se realizará una serie de pruebas aplicando un voltaje constante a la entrada del CAD y registrando la variabilidad del resultado de la conversión. Para ello se utiliza el circuito divisor de voltaje de la figura 2, conectado al canal A0. Se obtuvieron 1000 lecturas del CAD de cada una de las tarjetas y los histogramas resultantes se muestran en la figura 4. A todas las tarjetas se les conectó el mismo divisor de voltaje, mediante una tarjeta de expansión (*shield*) que se muestra en la figura 5. Esta tarjeta fue elaborada por el autor, a partir de una tarjeta comercial PCB de prototipaje (Prototype shield V.5). Se realizó un programa que toma mil muestras y las envía por el puerto serial, donde pueden registrarse mediante un programa de terminal serial. Todas las tarjetas, salvo la original B, proporcionan una conversión perfecta, es decir, entregan todas las veces el mismo dato. Estos resultados están dentro de las especificaciones que da el fabricante para el error en el convertidor analógico-digital, que es de ± 2 cuentas.

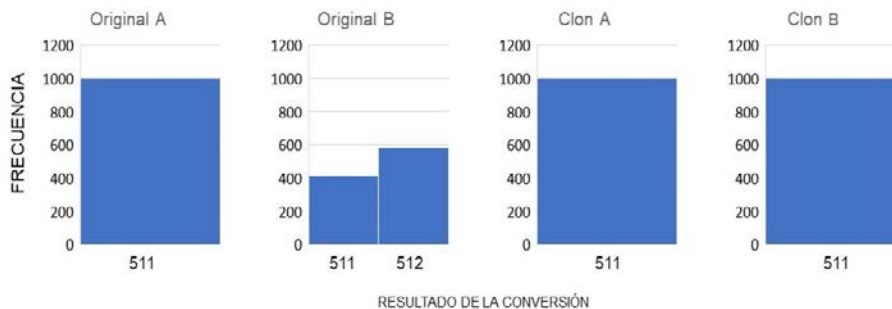


Figura 4 Valores que produce el convertidor A/D, cuando la tarjeta Arduino se alimenta mediante una conexión USB con una computadora portátil.

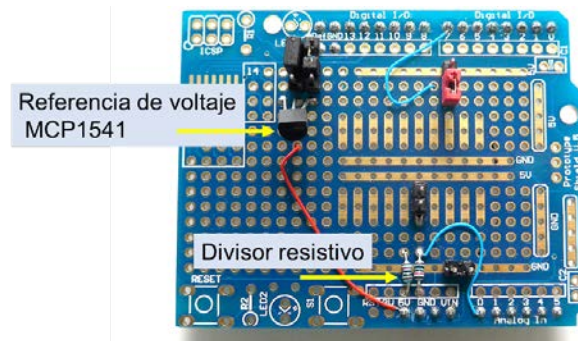


Figura 5 Tarjeta de expansión para evaluar conversión analógico-digital, tarjetas Arduino.

En muchas situaciones se recomienda conectar una referencia de voltaje externa, que puede presentar una mayor exactitud inicial, un menor corrimiento térmico y menor dependencia del voltaje de suministro. Se procedió a conectar una referencia de voltaje externa MCP1541 de 4.096 V [Microchip, 2012], y los resultados se presentan en la figura 6. Contrario a lo que se podría esperar, la precisión del resultado de conversión a disminuido. Esto se puede atribuir a la ubicación de la referencia de voltaje. Las tarjetas Arduino tienen conexiones para la utilización de una referencia externa, pero está unos 4 cm alejada de los pines correspondientes del microcontrolador, lo cual posibilita el acoplamiento de ruido [Arduino, 2017b]. Por otro lado, el divisor resistivo que está montado en la tarjeta de expansión se ubica en la posición más próxima posible a las líneas de alimentación y al convertidor A/D. Estos resultados sugieren que no es adecuado conectar una referencia de voltaje externa a la tarjeta Arduino UNO.

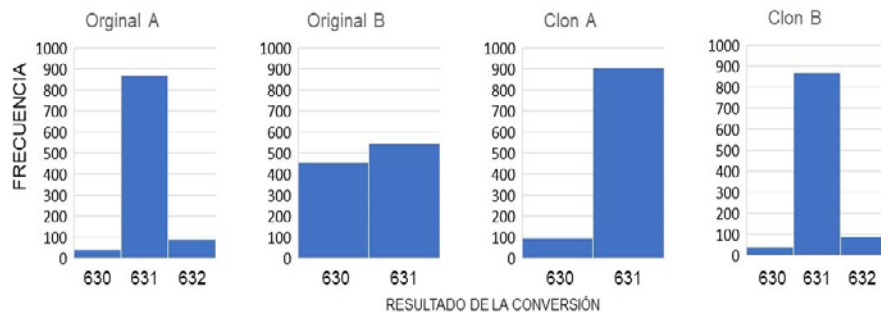


Figura 6 Comparación de los valores que produce el convertidor A/D cuando se utiliza una referencia de voltaje externa de 4.094 V y alimentación mediante conexión USB a una computadora portátil.

Como comparación, se conectó un divisor resistivo idéntico, en una tarjeta protoboard externa, y se conectó a la tarjeta Arduino mediante cables de 15 cm de longitud. En este caso, el acoplamiento de ruido produce una reducción en la precisión de la conversión según se muestran en la figura 7.

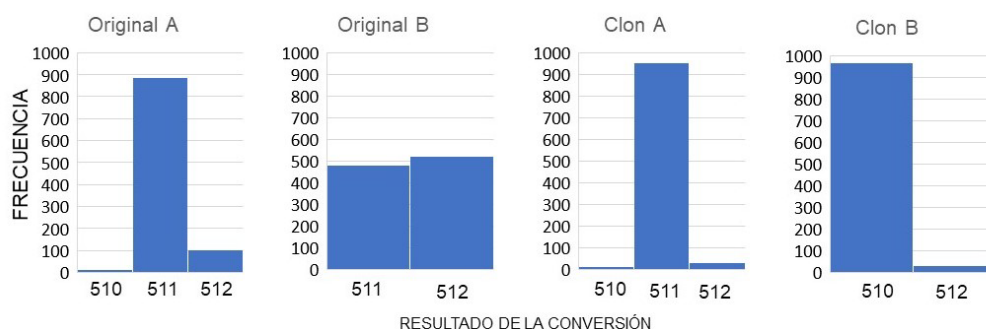


Figura 7 Comparación de los valores que produce el convertidor A/D cuando se utiliza una referencia de voltaje interna y un divisor resistivo a 15 cm de la tarjeta Arduino. Alimentación mediante conexión USB a una computadora portátil.

En algunas ocasiones se desea operar la tarjeta Arduino sin estar conectada a una computadora. Una opción es utilizar como fuente de energía una pila de respaldo de un teléfono celular. Si la pila de respaldo se conecta al puerto USB de la tarjeta Arduino, entonces no podrán enviarse los datos utilizando este puerto. Para evitar este inconveniente se desarrolló un programa que graba los datos en la memoria EEPROM del microcontrolador. La tarjeta Arduino UNO solo tiene 1 kB de memoria EEPROM, por lo que únicamente se pueden registrar 500 datos de 10 bits cada uno. Una vez grabados, otra sección del programa permite su envío por el puerto serial para su registro en un archivo de datos. Los resultados de la utilización de la pila de respaldo se muestran en la figura 8. Se observa que la precisión ha disminuido, lo cual se debe a que la pila de celular entrega un nivel de voltaje con mayor ruido. La pila de respaldo está constituida por batería de litio de 3.7 V y una fuente conmutada para producir los 5 V, figura 9.

Para evitar el ruido de la pila de respaldo, ésta se reemplazó por una pila alcalina de 9 V, la cual se conecta al regulador de voltaje lineal de la tarjeta Arduino. Al no tener una fuente conmutada, el ruido disminuye considerablemente y los resultados se presentan en la figura 10. Se observa un valor de conversión

constante para las tarjetas original A y las dos clones. Por otro lado, la tarjeta original B parece haber mejorado su desempeño con respecto al que presentó al ser alimentada mediante la conexión USB con una computadora (figura 4), dado que presenta una menor frecuencia de lecturas fuera del valor esperado de 511. Se podría decir que las tarjetas clones funcionan mejor que las originales, pero el número de tarjetas probadas no es suficiente para generalizar los resultados.

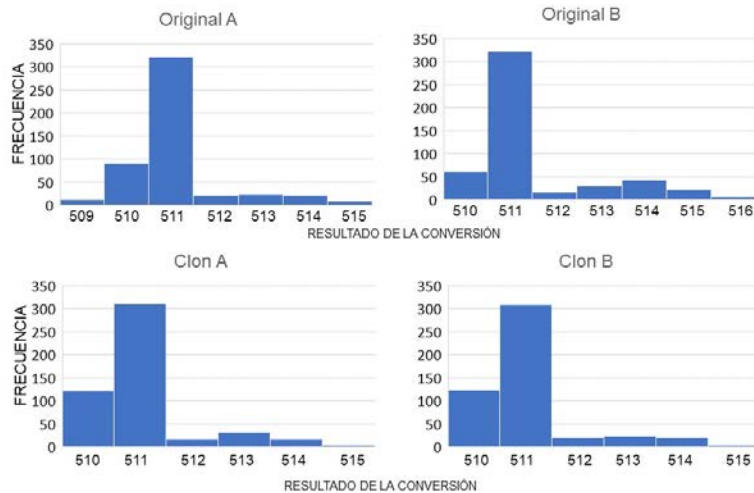


Figura 8 Comparación de los valores que produce el convertidor A/D cuando se utiliza una referencia de voltaje interna y se alimenta la tarjeta mediante una pila de respaldo de celular.



Figura 9 Interior de una pila de respaldo de teléfono celular.

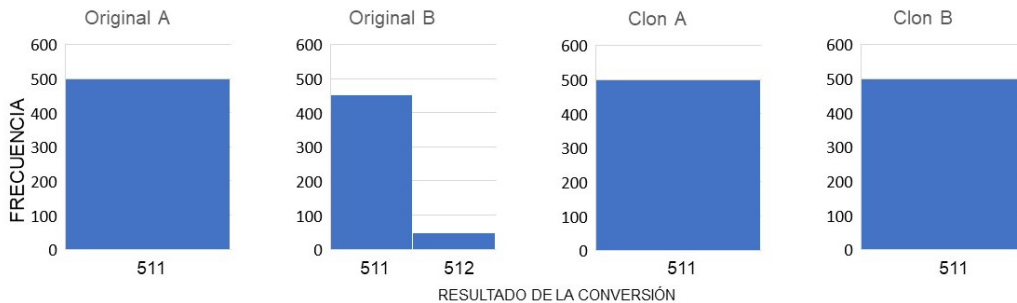


Figura 10 Comparación de los valores que produce el convertidor A/D cuando se utiliza una referencia de voltaje interna y se alimenta mediante una pila de 9 V.

4. Discusión

La tarjeta Arduino se ha convertido en una herramienta muy útil tanto para aficionados como para expertos en electrónica. La gran variedad de dispositivos, módulos y bibliotecas disponibles han contribuido a su popularidad. Resulta entonces muy sencillo crear un instrumento de medición basado en dicha tarjeta. Cuando se conecta un sensor analógico, el desempeño está limitado por las características del convertidor digital-analógico interno del microcontrolador, pero también se ve influenciado por factores externos.

Se han evaluado tarjetas originales y versión clon sin encontrar diferencias significativas. En la muestra que se tomó, realizando pruebas de medición de un divisor de voltaje conectado de forma próxima a la entrada del convertidor analógico-digital, las tarjetas originales y clones tuvieron un desempeño dentro de las especificaciones. Se encontró que la calidad de la medición disminuye si se aleja el sensor, puesto que los cables de conexión actúan como antenas donde se puede acoplar el ruido. Lo mismo ocurre si la tarjeta se alimenta mediante una pila de respaldo de teléfono celular, si esta contiene un convertidor de voltaje CD-CD. El uso de una referencia de voltaje externa no presentó una ventaja en la medición, debido a que su posición no es óptima dentro del conjunto de conectores de la tarjeta.

Finalmente, el mejor desempeño se obtiene cuando se siguen las recomendaciones generales de un sistema de instrumentación, es decir, mantener las conexiones cortas, utilizar una fuente de voltaje con poco ruido, etc.

5. Conclusiones

La tarjeta Arduino UNO es un dispositivo económico que se puede utilizar para desarrollar un sistema de medición de manera sencilla. En el mercado existen versiones clones de la tarjeta que pueden costar entre 30% y 50% del costo de la tarjeta original que fue desarrollada en Italia. Para analizar las tarjetas originales y clones se realizaron conjuntos de 500 o 1000 muestras, del voltaje en un divisor resistivo, mientras las tarjetas eran alimentadas con una conexión USB a una computadora portátil, una pila de respaldo de teléfono celular o una pila de 9 V.

También se hicieron algunas pruebas utilizando un circuito integrado como referencia de voltaje externa de 4.096 V. Las pruebas no mostraron ninguna diferencia de precisión entre ambas versiones, que cayera fuera de las especificaciones del convertidor analógico-digital.

Muchos aficionados a la electrónica hacen uso de estas tarjetas; sin embargo, para aprovechar al máximo sus capacidades, es necesario subrayar la importancia de seguir algunas recomendaciones básicas de la instrumentación electrónica: se deben mantener conexiones cortas o blindadas y una fuente de alimentación estable. Se encontró que el uso de una referencia de voltaje externa, también recomendado usualmente, parece no representar una mejora en el desempeño de estas tarjetas.

Los errores en el resultado de la conversión analógica-digital fueron el principal objetivo de este trabajo; sin embargo, otras características que se pueden evaluar en el futuro son el cruce de señales (*crosstalking*) y la velocidad máxima de muestreo.

6. Bibliografía y Referencias

- [1] Addicore, UNO R3 ATmega328P (RobotDyn UNO), agosto 2017: <https://www.addicore.com/Black-UNO-R3-p/ad308.htm>.
- [2] Arduino, Arduino UNO & Genuino UNO, 2017: <https://www.arduino.cc/en/Main/ArduinoBoardUno>, junio 2017.
- [3] Arduino, Arduino UNO Reference Design: <https://www.arduino.cc/en/uploads/Main/arduino-uno-schematic.pdf>, junio 2017.
- [4] Geekcriet, Geekcreit® UNO: <http://www.geekcreit.com>, agosto 2017.
- [5] Microchip, MCP1525/41 2.5V and 4.096 Voltage References, 2017, <http://ww1.microchip.com/downloads/en/DeviceDoc/21653C.pdf>, junio 2017.
- [6] Sparkfun, SparkFun ReadBoard, 2017: <https://www.sparkfun.com/products/13975>. agosto 2017.
- [7] Torrone, P., Top 10 Arduino-compatibles: <http://makezine.com/2012/04/24/soapbox-my-top-10-favorite-arduino-compatible-clones-and-derivatives/>.

ANÁLISIS DE RENDIMIENTO DE LA PC ODROID C2 PARA SU USO EN ESQUEMAS DE CIUDADES INTELIGENTES

Ernesto Bardales Hernández

Universidad Autónoma del Estado de México

ebardalesh.7@gmail.com

Saúl Lazcano Salas

Universidad Autónoma del Estado de México

slazcanos@uaemex.mx

Resumen

Dentro de las llamadas ciudades inteligentes, las redes de sensores para el monitoreo de diversas métricas como calidad del aire, densidad vehicular, entre otros, cobran particular importancia por la información que proporcionan. Estas redes envían sus datos a un punto central de procesamiento para su análisis e interpretación; sin embargo, en muchas ocasiones el envío de estos datos se realiza a través de canales de comunicaciones poco seguros y con baja relación de señal a ruido (*signal-to-noise-ratio*, *SNR*), tales como canales inalámbricos. En el presente trabajo, se propone el uso de la PC Odroid C2 como herramienta para implementar un codificador de convolución como esquema de codificación de canal que permita que los datos transmitidos sean menos susceptibles al ruido, así como para otro tipo de procesos relacionados con el área. Se estudia de la misma forma una tecnología similar: la PC Raspberry Pi versión 3, anexando pruebas de benchmarking de ambas PC.

Palabras Clave: Arduino UNO, ciudades inteligentes, codificación de canal, código de convolución, Odroid C2, Raspberry Pi, redes de sensores.

Abstract

Within the so-called intelligent cities, networks of sensors for the monitoring of various metrics such as air quality, vehicular density, and others, are

particularly important because of the information they provide. These networks send their data to a central processing point for analysis and interpretation; However, this data is often sent through unsafe and low signal-to-noise-ratio (SNR) communications channels, such as wireless channels. In the present work, it is proposed to use the PC Odroid C2 as a tool to implement a convolution encoder as a channel coding scheme that allows the transmitted data to be less susceptible to noise, as well as to other types of processes related to the area. A similar technology is studied in the same way: PC Raspberry Pi version 3, appending benchmarking tests to both PC.

Keywords: *Arduino UNO, Channel encoding, Convolution codes, Odroid C2, Raspberry Pi, Sensor Networks, Smart Cities.*

1. Introducción

El crecimiento de las grandes urbes conlleva una problemática compleja en diversos ámbitos como salud, transporte, trabajo y servicios. El garantizar calidad de vida y sustentabilidad de manera simultánea implica retos cuya solución no es única. En este contexto, una alternativa que ha cobrado particular fuerza en la última década es el llamado paradigma de las ciudades inteligentes, que en esencia, se puede describir como un gran sistema compuesto por pequeños subsistemas, cada uno de los cuales es responsable de una tarea en particular [Movuna, 2016]. El objetivo en una ciudad inteligente es lograr una mejora en la calidad de vida de las personas tomando como punto de partida el uso inteligente, inclusivo y de manera sustentable de las tecnologías de la información [Arroub, 2016].

Dentro de la infraestructura para las ciudades inteligentes, destaca el uso de las tecnologías de la información como apoyo a tareas de monitoreo automatizado en cámaras, sistemas GPS, sensores de camino (detección de baches y objetos) entre otros. Este monitoreo se hace por lo general con apoyo de redes de sensores, que recaban la información deseada y la transmiten a un punto de control para su análisis y toma de decisiones correspondiente [Arroub, 2016], [Benamrou, 2016].

La transmisión de datos puede llevarse a cabo aprovechando la infraestructura existente, como la red eléctrica o empleando canales inalámbricos. Sin embargo, tanto la red eléctrica como los canales inalámbricos presentan el inconveniente de ser canales con una figura de ruido muy agresiva [Tseng, 2014], [Lui, 2014] y en consecuencia, la calidad de los datos en el extremo receptor se puede ver seriamente comprometida en caso de no tener un esquema de codificación de canal adecuado.

En este sentido, desde la publicación del trabajo de Claude E. Shannon en 1948 [Shannon, 1948] hasta la actualidad, se ha realizado un gran esfuerzo en buscar esquemas de codificación de canal que permitan acercarse al límite teórico establecido en dicho trabajo. Uno de los esquemas de codificación de amplia implementación en diversos sistemas son los códigos de convolución [Morelos, 2006], predecesores de los turbo códigos los cuales junto a los códigos LDPC (de sus siglas en inglés Low Density Parity Check), presentan un desempeño muy cercano al límite de Shannon.

Existen algunas herramientas tecnológicas de bajo costo, capaces de realizar tareas requeridas por los esquemas de ciudades inteligentes. Un ejemplo de ellas son las PC de placa reducida con arquitectura ARM.

Las PC analizadas en este trabajo son: la PC Raspberry Pi versión 3, de la cual se han realizado diversos trabajos; por citar algunos se tiene que [González, 2015] menciona el uso de la PC Raspberry Pi versión 2 como servidor de datos, el cual mediante el uso del microcontrolador Arduino conectado en el puerto serial USB de la PC, se encarga de hacer lecturas y recolectar información de sensores y actuadores, almacenándolos en una base de datos para futuros análisis.

Otro trabajo al respecto, es el presentado por [Liscano, 2014] el cual propone mejoras al tránsito vehicular mediante una red de dispositivos electrónicos, diseñando variantes en infraestructura de TICS (Tecnológicas de la Información y Comunicación) enfocadas a entornos de ciudades inteligentes, tomando como base de procesamiento el uso de la PC Raspberry Pi versión 2.

Por otra parte, [Solarte, 2017] generan una propuesta basada en una plataforma que recibe información de redes de sensores diversas. El trabajo de dicha

plataforma es homogeneizar la información recibida, procesarla y ponerla a disposición de los usuarios a través de internet, permitiendo de este modo contar con información en tiempo real vinculada con el entorno, como ejemplo la densidad vehicular o el clima.

Por otro lado, se analiza la PC Odroid C2, la cual es otra tecnología con capacidades similares, incluso superiores a la PC Raspberry Pi; ejemplo de estas son: memoria RAM, CPU, compresión de datos, transferencia por el adaptador de red Ethernet, encriptado mediante SSL (por sus siglas en inglés Secure Sockets Layer), y acceso a almacenamiento de disco.

Actualmente la PC Odroid C2 no ha sido utilizada en esquemas de ciudades inteligentes como lo ha sido la Raspberry Pi, sin embargo al contar con características técnicas superiores, puede ser utilizada bajo este tipo de esquemas.

En el presente trabajo, se analiza la factibilidad de emplear la PC Odroid C2 como plataforma base para la implementación de un codificador de canal de convolución el cual se aplica a los datos provenientes de una red de sensores. El objetivo con lo anterior es brindar a dichos datos robustez para ser transmitidos en canales con SNR bajas, como es el caso de un canal inalámbrico. De igual forma, se compara este mismo esquema de codificación de canal con la PC Raspberry Pi versión 3 obteniendo el desempeño de ambas plataformas. Adicionalmente se realizará a ambas PC pruebas de benchmarking en aspectos relacionados a las ciudades inteligentes, para tener de este modo un punto comparativo base.

2. Métodos

Para realizar la codificación de canal, se selecciona un codificador de convolución como se define en [Morelos, 2006], con una tasa de codificación de $R = \frac{1}{2}$ con polinomios generadores definidos por ecuaciones 1 y 2.

$$P_1(x) = x^2 + x + 1 \quad (1)$$

$$P_2(x) = x^2 + 1 \quad (2)$$

El esquema general del codificador se muestra en la figura 1. Se selecciona este codificador de convolución como referencia ya que a partir del mismo se pueden

implementar esquemas más robustos de codificación de canal tales como los turbo códigos [Berrou, 1993].

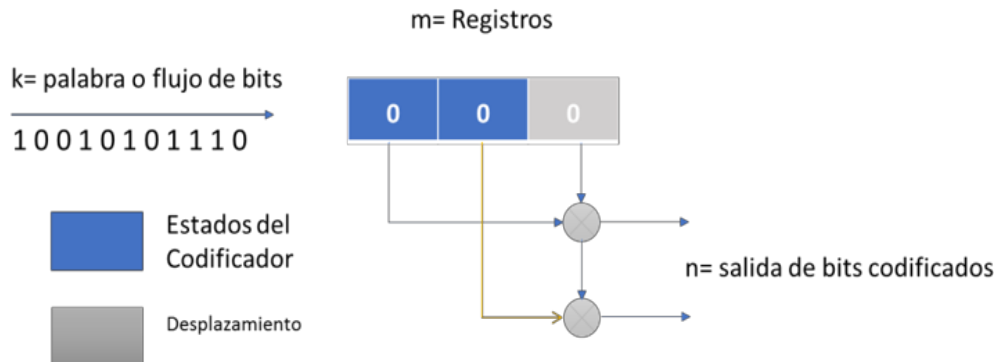


Figura 1 Estructura del codificador convolucional.

El esquema de configuración para pruebas a emplear se muestra en la figura 2.

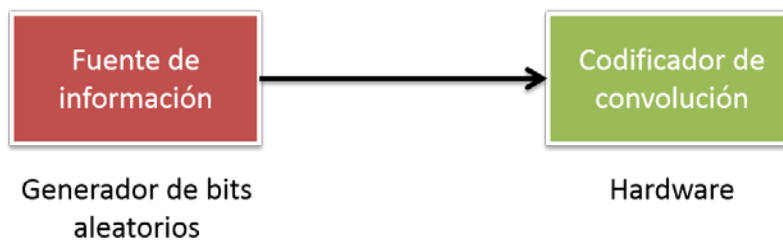


Figura 2 Esquema de transmisión con fuente de información al hardware propuesto.

Como fuente de información se programa un simulador de generador de bits aleatorios a diferentes tasas de transmisión, basado en la plataforma Arduino UNO cuyo algoritmo base se muestra a continuación.

- Inicio
- Iniciando la configuración de transmisión del puerto serial a una tasa de bits determinada.
- Iniciando iteraciones sin fin.
- Imprimiendo en el puerto serial un valor entero aleatorio entre 0 y 1.

Este simulador de generador de bits aleatorios se enlaza al hardware a través del puerto serial de este. La tasa de bits usada se muestra en la tabla 1, es necesario

resaltar que cada valor de bits por segundo correspondiente a la programación de la fuente de información es modificada a diferentes tasas de bits para cada iteración de pruebas.

Tabla 1 Tasas de bits por segundo usadas para el puerto serial del hardware.

Tasa de bits por segundo
1200 bps
2400 bps
4800 bps
9600 bps
19200 bps
38400 bps

Para analizar los niveles de carga de procesamiento del hardware, se somete a pruebas de 4 horas de trabajo continuo con monitoreo cada dos segundos, conforme a las velocidades de transmisión del puerto serial de la tabla 1. El algoritmo para la lectura del bit a ser codificado, se muestra a continuación:

- Inicio
- Iniciar variable identificador del puerto serial y tasa de transmisión
- Iniciando iteraciones sin fin
- Leyendo valor entero del puerto serial mediante variable identificador
- Enviando valor leído al algoritmo del codificador

El diagrama de flujo correspondiente al algoritmo del codificador convolucional de la figura 1 se presenta en la figura 3.

Para analizar la factibilidad de uso de la PC Odroid C2 en un entorno de trabajo real, se analizan tres variables importantes de comportamiento: temperatura, tasa de flujo de bits y porcentaje de carga de CPU. Esta metodología será implementada de la misma forma con la placa Raspberry pi versión 3.

Adicionalmente se usará el software tipo benchmark [Phoronix, 2017] para pruebas a ambas tecnologías, para resaltar las capacidades de la PC Odroid C2 en operaciones a nivel hardware. Los factores sometidos a las pruebas son:

- ✓ Memoria caché
- ✓ Velocidad de RAM
- ✓ Carga de CPU
- ✓ Compresión de datos
- ✓ Algoritmo de encriptado SSL para transporte de datos
- ✓ Velocidad del adaptador de red Ethernet
- ✓ Velocidad de acceso al disco duro

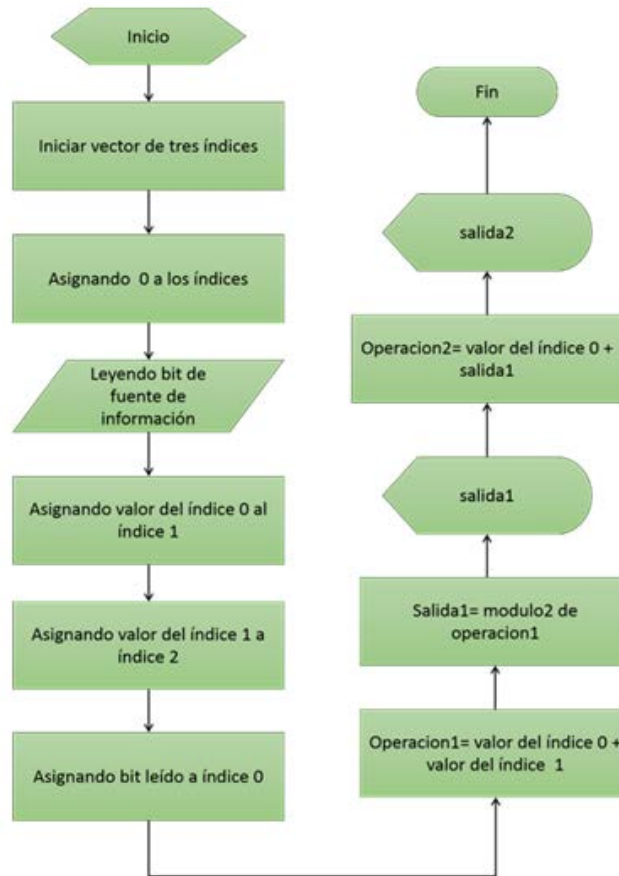


Figura 3 Diagrama de flujo de datos del algoritmo de figura 1.

3. Resultados

- De acuerdo con las pruebas realizadas en ambas tecnologías (con las tasas mostradas en la tabla 1) se obtuvieron los resultados que se muestran en la tabla 2 y tabla 3. Cabe mencionar que la información mostrada hace referencia al promedio de los valores generales.

Tabla 2 Carga de CPU promedio del hardware estudiado.

Tasa de bits	%CPU - Raspberry Pi	%CPU - Odroid C2
1200	1.66	1.41
2400	2.96	2.39
4800	5.65	4.55
9600	9.57	8.33
19200	17.11	14.19
38400	17.3	14.3

Tabla 3 Temperatura de la zona térmica promedio del hardware estudiado.

Tasa de bits	C° Raspberry Pi	C° Odroid C2
1200	37	27.86
2400	45.17	41.13
4800	45.2	42.7
9600	46.46	42.13
19200	46.71	43.63
38400	47	43.2

Ambas tecnologías presentan un comportamiento similar, aunque a tasas de bits altas la PC Odroid C2 comienza a despuntar su desempeño contra la PC Raspberry Pi. En la figura 4 se representa la curva de comportamiento de carga de la CPU con respecto a la tasa de bits determinada a ambas PC.

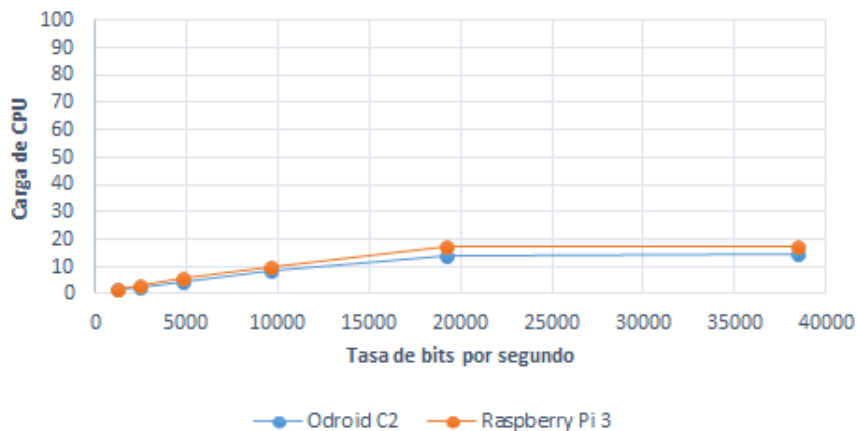


Figura 4 Representación de la carga de CPU con respecto a la tasa suministrada PC.

En la figura 5 se muestra la curva de comportamiento de temperatura de la zona térmica de ambas PC con respecto a la tasa de bits determinada.

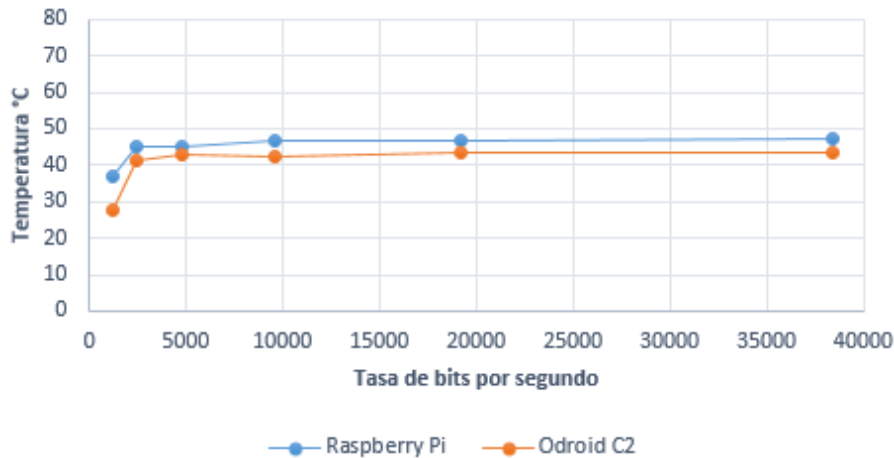


Figura 5 Representación de temperatura con respecto a la tasa suministrada.

- Dentro de las pruebas de benchmarking realizadas a ambas PC, se obtuvieron los siguientes resultados. Cabe mencionar que las pruebas están sostenidas a los tiempos predeterminados asignados por el software:
 - ✓ **Tiempo de compresión** de información con la herramienta Gzip figura 6. La PC Odroid C2 destaca con un mejor desempeño por su rapidez de compresión, ya que PC Raspberry Pi se tomó el doble de tiempo para realizar dicha tarea.

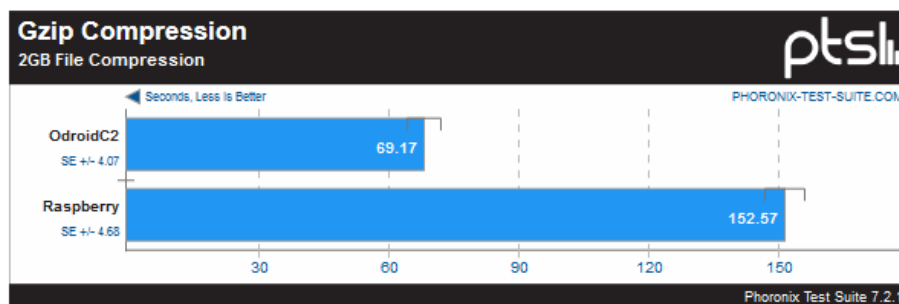


Figura 6 Tiempos de compresión con Raspberry Pi y Odroid C2.

- ✓ **Compresión** de información con la herramienta 7-zip figura 7. En esta prueba se analiza la cantidad de instrucciones por segundo que cada PC puede realizar para una compresión de datos efectiva. La PC Odroid C2 destaca nuevamente al realizar poco más del doble de MIPS (millones de operaciones por segundo).

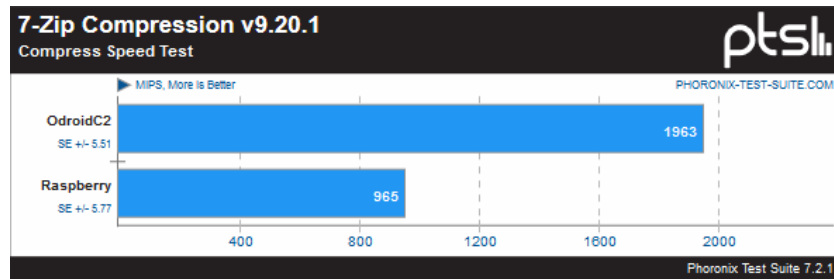


Figura 7 Velocidad de compresión por MIPS.

- ✓ **Velocidad de RAM** promedio de operaciones con valores enteros y punto flotante figura 8 y figura 9. Se analiza la cantidad de MB/s que las operaciones pueden manejar. La PC Odroid C2 muestra ventaja considerable en velocidad de RAM, siendo casi el doble su desempeño.

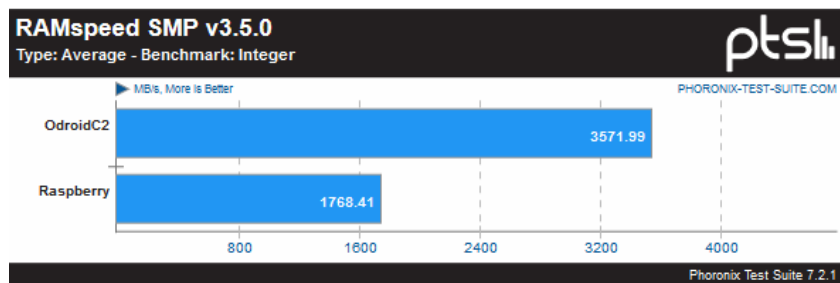


Figura 8 Velocidad de RAM promedio con valores enteros.

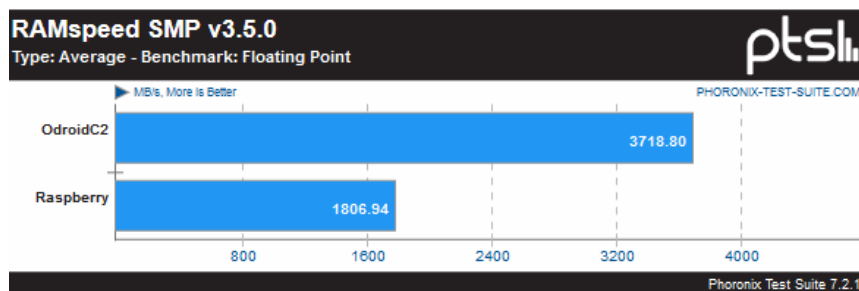


Figura 9 Velocidad de RAM promedio con valores de punto flotante.

- ✓ **Velocidad de transmisión del adaptador de red Ethernet** figura 10. La PC Odroid C2 tuvo un tiempo de transferencia de información sobresaliente, al transmitir en menor tiempo con respecto a PC Raspberry Pi.

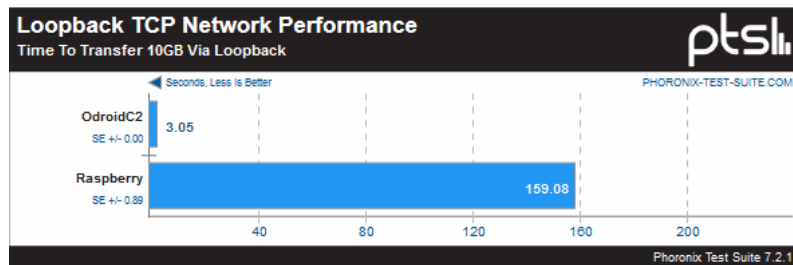


Figura 10 Velocidad de adaptador de red.

- ✓ **Velocidad de escritura del medio de almacenamiento** figura 11 en diferentes sectores. La PC Odroid C2 presentó de manera considerable una mejor tasa de escritura de información a diferencia de la PC Raspberry Pi.

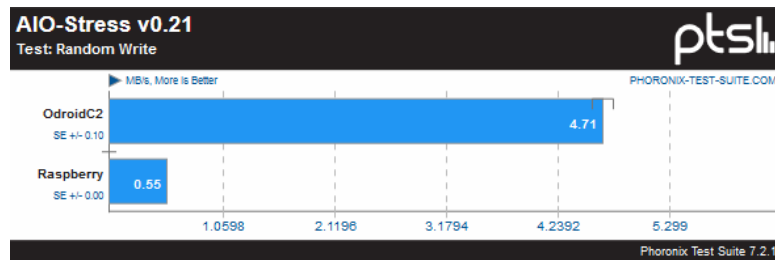


Figura 11 Velocidad de escritura en el medio de almacenamiento.

- ✓ **Velocidad de memoria caché medido en MB/s para diversas operaciones** figura 12. La memoria caché de la PC Odroid C2 obtuvo la ventaja al operar a más del doble de velocidad, que la tecnología competidora.

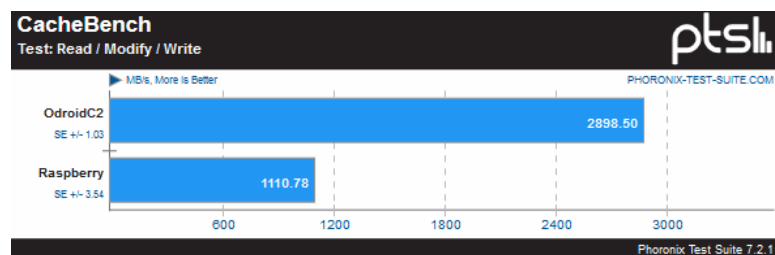


Figura 12 Velocidad de memoria caché.

- ✓ **Velocidad de codificación de video MPEG** figura 13. La PC Odroid C2 obtuvo una mejor marca de tiempo en codificación de video, frente

a la PC Raspberry Pi, la cual se tomó poco más del doble de tiempo que la PC Odroid C2 en operación de codificación.

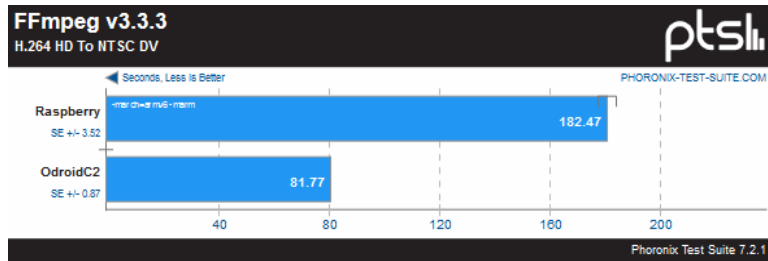


Figura 13 Tiempos de codificación de video.

- ✓ **Rendimiento de CPU** para cálculos de punto flotante figura 14. La PC Odroid C2 muestra ventaja al operar en tiempo menor que la tecnología competidora, ya que la PC Raspberry Pi se tomó más del doble de tiempo para realizar sus respectivas operaciones.

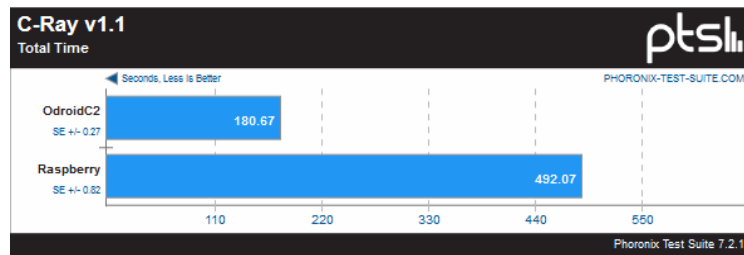


Figura 14 Tiempo de cálculo de punto flotante en CPU.

- ✓ **Rendimiento de encriptado** mediante el protocolo SSL con firma digital RSA-4096 medido en símbolos/segundo, figura 15. La PC Odroid C2 mantiene ventaja en capacidad de encriptado, por lo que la PC Raspberry Pi se queda por debajo de la mitad.

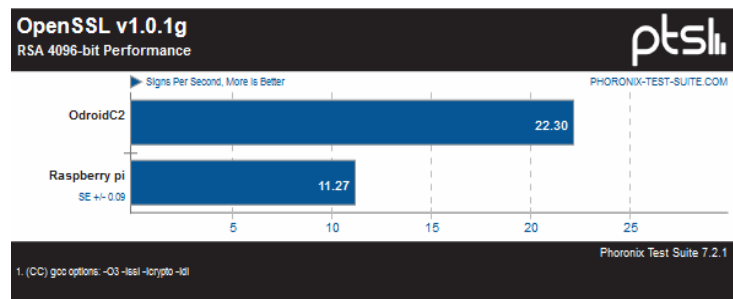


Figura 15 Rendimiento de encriptado mediante SSL.

4. Discusión

A partir de los resultados mostrados en la sección anterior, se puede observar que para tasas de datos de hasta 38400 bps, ambas plataformas de hardware presentan un comportamiento muy estable, y sobre todo confiable. Sin embargo, se debe resaltar que la PC Odroid C2 consume menos recursos de CPU en relación con la PC Raspberry Pi versión 3.

Tema de particular interés es la temperatura, pues la PC Odroid C2 incorpora en su arquitectura un disipador de calor que le permite tener un comportamiento mucho más estable, lo cual resulta en una ventaja significativa si se desea conectar dispositivos diversos. La demanda de corriente tiene relación directa con el incremento de temperatura en cualquier plataforma.

Adicionalmente, la PC Odroid C2 incorpora un mecanismo de protección en caso de que la temperatura de trabajo exceda los límites de operación, consistente en disminuir las operaciones realizadas por el procesador hasta que la temperatura regrese a los umbrales de operación [Roy, 2015].

Con los resultados obtenidos de las pruebas de benchmarking realizadas a ambas PC, y tomando como referencia algunas de las tareas que se llevan a cabo dentro de una infraestructura de una ciudad inteligente, se observa que la PC Odroid C2 presenta ventaja de potencia de operaciones, como es el caso del encriptado de datos mediante el protocolo SSL, del cual la PC Odroid C2 tuvo un desempeño mayor con respecto a la Raspberry Pi. Asimismo, la transferencia de datos mediante el puerto de red Ethernet de la PC Odroid C2 es mucho más veloz, al incorporar un puerto Gigabit Ethernet con lo que puede transferir una cantidad de información mayor en menor tiempo, obteniendo de esa manera un mejor uso de los recursos de la PC Odroid C2.

Otro factor importante dentro de las tareas de una ciudad inteligente es el análisis de compresión de datos. En este sentido, al transmitir información por un medio de comunicación inalámbrico es necesario considerar que mientras mejor compactados vayan los datos, se hará uso de un menor ancho de banda, dejando espacio suficiente para que más canales de comunicación ejecuten otros procesos. La PC Odroid C2 presentó un nivel superior de compresión de datos al

operar a más MIPS (millones de instrucciones por segundo) en menor cantidad de tiempo que la tecnología competidora.

Ante las notables diferencias entre ambas tecnologías se puede observar que la PC Odroid C2 muestra mejor desempeño en las tareas a las que fue sometida.

5. Conclusiones

A partir de los resultados obtenidos, se observa que la PC Odroid C2 presenta un comportamiento en desempeño superior a otras plataformas similares, concretamente la PC Raspberry Pi versión 3.

Se destaca la capacidad de la PC Odroid C2 para realizar diversas tareas de alta importancia como lo es la codificación de canal, la compresión de datos, entre otros, y de este modo garantizar que los datos a ser transmitidos llevan robustez en caso de errores en la transmisión de los mismos y por otra parte, los datos no viajan expuestos sobre el medio de comunicación elegido.

Otras plataformas como la PC Raspberry Pi, presentan un desempeño aceptable en tareas individuales y paralelas, pero a su vez presentan tiempos de respuesta más elevados al momento de implementar tareas más complejas, por lo cual se limita el uso de la misma en entornos de mayor demanda de procesamiento.

Llevando estas características al entorno de ciudades inteligentes, emplear la PC Odroid C2 se vuelve una tarea muy atractiva, dadas las características y desempeño analizados. Esta plataforma es una herramienta con prestaciones muy sobresalientes, haciendo muy atractivo su uso en este entorno. Queda como trabajo a futuro, implementar esquemas prototipo como:

- ✓ Análisis de algoritmos de compresión de imágenes y video, pensando en reducir la cantidad de datos a ser enviados y procesados.
- ✓ Análisis de esquemas de codificación de canal más robustos, como turbo codificación y esquemas de perforado, aptos para canales inalámbricos.
- ✓ Análisis de complejidad de algoritmos de cifrado, particularmente cifrado de curvas elípticas que permiten un cifrado muy robusto de los datos a un costo razonable de complejidad.

6. Bibliografía y Referencias

- [1] A. Arroub, B. Zahi, E. Sabir & M. Sadik, A literature review on Smart Cities: Paradigms, opportunities and open problems, International Conference on Wireless Networks and Mobile Communications (WINCOM), Fez, Morocco, 2016.
- [2] B. Benamrou, B. Mohamed, A. S. Bernoussi & O. Mustapha, Ranking models of smart cities, de 4th IEEE International Colloquium on Information Science and Technology (CiSt), Tangier, 2016.
- [3] C. Berrou, A. Glavieux & P. Thitimajshima, Near Shannon limit error-correcting coding and decoding: Turbo codes, Proc. IEEE Int. Conf. Commun, 1064-1070, 1993.
- [4] C. E. Shannon, A Mathematical Theory of Communication, The Bell System Technical Journal, vol. 27, 379–423, 623–656, 1948.
- [5] D. F. Tseng, F. G. Mengistu, Y. S. Han, A. M. Mulatu & T. R. Tsai, Robust Turbo Decoding in a Markov Gaussian Channel, in IEEE Wireless Communications Letters, vol. 3, nº 6, 633-636, 2014.
- [6] E. González, Red de sensores, Internet de las cosas, Tesis de grado, Escola Tècnica Superior d'Enginyeria Informàtica, Universitat Politècnica de València, 2015.
- [7] J. Muvuna, T. Boutaleb, S. B. Mickovski & K. J. Baker, Systems engineering approach to desing and modelling of smart cities, de International Conference for Students on Applied Enginnering (ICSAE), Newcastle upon Tyne, UK , 2016.
- [8] Phoronix, Phoronix test suite: <https://www.phoronix-test-suite.com/>. último acceso: 15 Agosto 2017.
- [9] R. H. Morelos, The Art of Correcting Coding, San Jose Francisco University USA: WILEY, 2006.
- [10] R. Roy & V. Bommakanti, User Manual Odroid C2, South Korea: Hardkernel, 2015.
- [11] T. A Liscano, D. Montoya, Mejoramiento de la movilidad y el tránsito en la ciudad de Santiago de Cali a través de la planeación y diseño de dos

servicios basados en TIC, Tesis de grado, Facultad de ingeniería departamento académico de tecnologías de información y comunicaciones, Universidad ICESI, 2014-2015.

- [12] Y. Liu, Z. Tan, H. Hu, L. J. Cimini & G. Y. Li, Channer Estimation for OFDM, IEEE Communications Surver & Tutorials, vol. 16, nº 4, 1891-1908, 2014.
- [13] Z. Solarte , L. Peña, D. Almario y C. Loaiza, Urbaneyes: plataforma de la gestion de información de ciudad, El hombre y la maquina, nº 48, 54-61, 2017.

SISTEMA DE ADQUISICIÓN DE DATOS DE BAJO COSTO PARA UN INVERNADERO BASADO EN TECNOLOGÍA DE ACCESO LIBRE

Felipe de Jesús Becerra Woo

Universidad Politécnica de Aguascalientes

felipe.becerra@upa.edu.mx

Araceli Gárate García

Universidad Politécnica de Baja California

araceli.garate@gmail.com

Tania Aglaé Ramírez del Real

Universidad Politécnica de Aguascalientes

tania.ramirez@upa.edu.mx

Ervin Jesús Alvarez Sánchez

Universidad Veracruzana Campus Xalapa

ervin.alvarezs@gmail.com

Resumen

En este trabajo se presenta un sistema de adquisición de datos desarrollado para un invernadero clásico cenital que se basa en el microcontrolador ESP826612E y la computadora de bolsillo Raspberry Pi 3, los cuales son plataformas de hardware libre. Los parámetros obtenidos son la temperatura y la humedad. En el método se incluye la integración de los componentes al sistema de adquisición de datos, en particular el sensor de temperatura y humedad (DHT11), el servidor (Mosquitto y Node-RED), utilizando los protocolos de comunicación inalámbrica (WiFi y MQTT). Los resultados muestran la factibilidad para utilizar un conjunto de dispositivos inalámbricos para la integración de un sistema donde se requiere procesar información de manera remota, en este caso un invernadero.

Palabras Claves: Adquisición de datos, invernadero, mosquito, Node-RED, Raspberry Pi.

Abstract

A data acquisition system for a classical zenith greenhouse is developed in this paper. It is based on the ESP826612E microcontroller and the Raspberry Pi 3, which are open source hardware. The humidity and temperature are the parameters to acquire. The methodology includes the integration of some key components, such as the DHT11 sensor, the Mosquitto and Node-RED server, using the wireless communication protocols (WiFi and MQTT). The results show the possibility to use a set of wireless devices in order to process the information in a remote connection, in this case a greenhouse.

Keywords: *Data acquisition, greenhouse, mosquito, Node-RED, Raspberry Pi.*

1. Introducción

El efecto del cambio climático ha tenido un impacto en las actividades de campo abierto y en su práctica alrededor del mundo [Kang et al, 2009], [Ramirez, 2013]. Es por ello que el uso de invernaderos se ha incrementado en los últimos años con el propósito de poder producir cultivo todo el año [Moulton, 2016]; sin embargo, para que un agricultor tenga acceso a la tecnología que le permite obtener este tipo de beneficio, requiere de una fuerte inversión económica. En [De Anda, 2017] se estima que el costo promedio para la tecnificación básica de un invernadero en México es de \$5 dólares por metro cuadrado; la cual consiste en mallas para sombra, cubiertas de plástico, estructuras metálicas y algunos otros elementos estructurales, mientras que uno completamente automatizado es de un costo promedio de \$115 dólares por metro cuadrado.

A pesar del costo, existen diversos trabajos de investigación que se dedican a realizar la automatización o sensado de los parámetros considerados como fundamentales para el crecimiento ideal de una planta, tales como la temperatura [Márquez et al, 2016], humedad relativa [Outanoute, 2016], humedad del suelo [Sharma et al, 2017], ventilación [Makhlouf et al, 2016], iluminación [Chang et al,

2016], radiación solar [Cossu et al, 2014], dióxido de carbono [Xu et al, 2017], oxígeno, entre otros [Leal Iga et al, 2006]. Las interfaces de dichos sistemas de supervisión remota en invernaderos, según la literatura, han sido desarrolladas utilizando diferentes tecnologías, las cuales normalmente son costosas y/o complicadas de implementar al requerir módulos de conversión y acoplamiento de señales [Pawlowski et al, 2016]. Entre los dispositivos de automatización más populares se encuentran los de la compañía National Instruments™ en conjunto con el software LabVIEW [Guofang et al, 2010], [Fang, 2011], [Juárez et al, 2016], sin embargo, el costo de esta tecnología es alto y puede reducirse con el uso de plataformas de arquitectura abierta. Actualmente, el uso de licencias libres ha ganado popularidad y aunado a ello, el hardware libre se le ha sumado permitiendo la realización de desarrollos accesibles y económicos que podemos encontrar aplicados en el sensado, supervisión, análisis y control de diferentes sistemas [Khot, 2016], [Ferrarezi et al, 2015].

Las condiciones de humedad, temperatura y radiación dentro de un invernadero dificultan la adquisición y la transmisión de datos empleando protocolos de comunicación cableada, como son el CAN-bus y el RS485, es por ello que su uso en este tipo de sistemas ha disminuido en los últimos años [Du et al, 2013]. Lo anterior ha permitido la proliferación del uso de tecnologías inalámbricas para el envío de información, dominando el uso de protocolos de área personal (WPA, por sus siglas en inglés), como son ZigBee y sobre todo Wi-Fi [Juárez et al, 2016], [Mad et al, 2014], [Xiaoyan et al, 2013], [Fezari et al, 2011].

La contribución principal de este trabajo es la integración de diversas tecnologías para el desarrollo de un sistema de adquisición de datos que utiliza dispositivos de tecnología libre. A lo largo del artículo se explicará cómo fue posible implementar el protocolo de comunicación WiFi en conjunto con el protocolo MQTT en este sistema utilizando sensores digitales para la lectura de las variables de temperatura en un invernadero clásico cenital. Cabe aclarar que en el presente artículo se muestran los resultados únicamente para un sensor, pero se muestra que la extensión a una red es sencilla gracias a la combinación en el servidor del uso de Mosquitto y Node-RED.

2. Métodos

En la figura 1 se puede ver la conexión de todos los elementos que intervienen en el proceso de adquisición de temperatura y humedad dentro de un invernadero de 6mx18m, el cual está ubicado dentro de las instalaciones de la Universidad Politécnica de Aguascalientes. A continuación, se describe cada una de las partes que conforma al sistema para después explicar el proceso de adquisición de datos de forma inalámbrica.

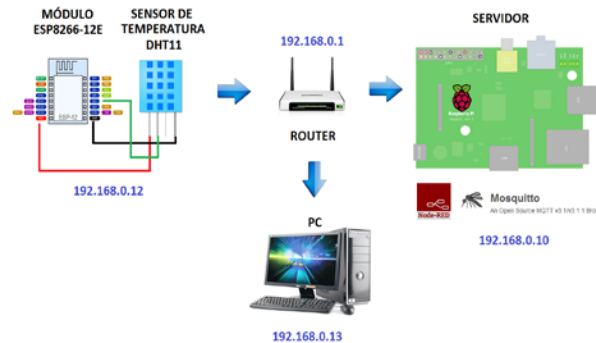


Figura 1 Sistema de adquisición de datos.

El sensor utilizado es el DHT11, el cual cuenta con las características necesarias para hacer la prueba de lectura de temperatura y humedad del invernadero pues tiene una etapa de instrumentación analógica-digital interna; asegurando la calibración del sensor, para después hacer uso del sistema de comunicación integrado por comunicación serial. El sensor DHT11 está dotado por una serie de características expresadas en la tabla 1.

Tabla1 Tabla de parámetros del sensor DHT11 [AOSONG, 2017].

Parámetro	Valor
Alimentación	3.5 a 5.5 V
Precisión en Humedad Relativa	±5% RH
Precisión en Temperatura	±2% °C
Rango de Lectura de Temperatura	0 – 50°C
Rango de Lectura de Humedad	20 – 90% RH

El microcontrolador ESP826612E cuenta con una serie de características expuestas en la tabla 2; de entre las cuales destaca la capacidad de poder hacer lectura y escritura de datos por comunicación serial y Wi-Fi.

Tabla 2 Parámetros del microcontrolador ESP826612-E [Espressif Systems, 2017].

Parámetro	Valor
Protocolo de comunicación Wi-Fi	802.11 b/g/n
Antena	Salida de potencia +19.5 dBm en modo 802.11b
MCU	32-bit
Puertos de comunicación	SDIO 2.0, SPI, UART
Consumo	215 mA
Fuente de alimentación	5 V

La tarjeta Raspberry Pi es la parte donde se tiene una serie de servicios gestionados por un sistema operativo basado en Linux para arquitectura ARM llamado Raspbian. Cuenta con una serie de características que la hacen ideal para la aplicación que se plantea en este trabajo, a continuación en la tabla 3 se exponen los datos [Vishnukumar et al, 2016], [Raspberry-Pi-Foundation, 2016].

El elemento principal de comunicación para la red de sensores es el protocolo MQTT, por sus siglas en inglés (MQ Telemetry Transport) el cual está diseñado para el uso de adquisición de datos sobre dispositivos de forma remota como sensores, teléfonos, computadoras, etc. [DC-Square, 2016]. El funcionamiento del protocolo está basado en 2 elementos el broker y el usuario. El broker es el servidor que administra todas las comunicaciones y los usuarios envían y reciben información directamente de este. Esto se hace mediante la suscripción a un tema, estar suscrito te permite enviar y recibir información sobre todo lo que se publique en ese tema. Es así como varios sensores se pueden conectar a un tema y un usuario puede ver todas las publicaciones en un solo lugar.

Tabla 3 Parámetros de Raspberry Pi 1 B [Raspberry-Pi-Foundation, 2016].

Parámetro	Valor
SoC	Broadcom BCM2835 (CPU + GPU + DSP + SDRAM + USB)
CPU	ARM 1176JZFS a 700 MHz
RAM	512 MB
Almacenamiento	SD
USB	2 x USB 2.0
Redes	Ethernet 10/100
Consumo	3.5 W
Fuente de alimentación	5 V micro USB

El servidor NodeRED, mostrado en la figura 2, es una plataforma gráfica de gestión de dispositivos e información en red, ya sea local o abierta. Esta

plataforma es muy versátil y permite hacer modificaciones a la red de forma gráfica y compacta. A los bloques que modifican el funcionamiento del proceso dentro de la plataforma se le conocen como nodos. Los nodos pueden ser de entradas, salidas, procesamiento, análisis, etc. Estos nodos se pueden conseguir prefabricados y modificarlos o también se pueden desarrollar desde cero para aplicaciones específicas que desee el usuario [Blackstock, 2014].

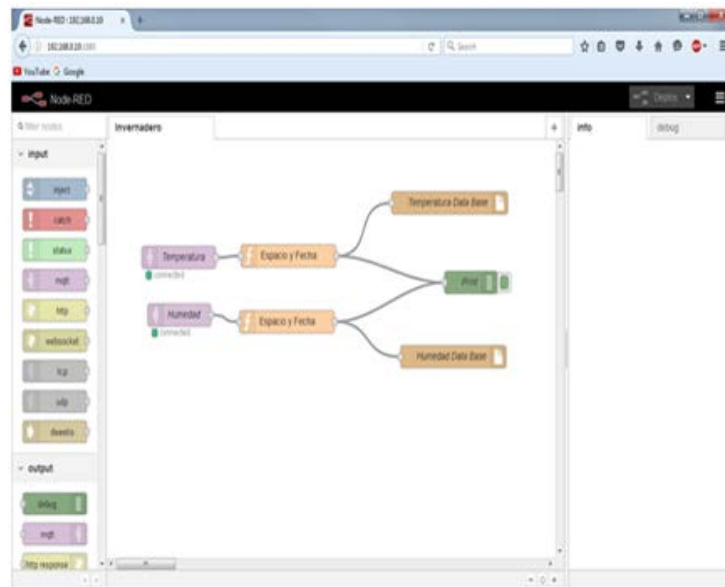


Figura 2 Enlace de datos en Node-RED.

El proceso de adquisición de datos está dividido esencialmente en tres partes: El momento en el que el ESP826612-E hace la adquisición del dato, la cual es cada 5 minutos, por recomendación del fabricante del sensor DHT11 [AOSONG, 2017]. El segundo proceso es cuando se establece una conexión entre el módulo y el servidor para después mandar la información, la cual se almacena en este último. El tercer proceso es cuando el usuario accede a la información para poder hacer análisis. En este momento se cuenta con un esquema general del sistema completo, a continuación se detalla cada parte:

- Proceso 1: El microcontrolador mediante una referencia de un reloj en tiempo real (RTC, por sus siglas en inglés) solicita una muestra al sensor DHT11 por medio de comunicación serial. La información se concatena; hora de muestreo, temperatura y humedad, y se envía por red inalámbrica.

Para poder mandar la información por Wi-Fi es importante primero pertenecer a una red con una dirección, en este caso el módulo hace una petición de conexión a un router y el DHCP (Protocolo de Configuración de Host Dinámico) y se le concede una dirección IP (Protocolo de Internet). La velocidad por el medio inalámbrico está sujeta a la tecnología que acepta el router y el microcontrolador; en este caso está ligado a la norma IEEE 802.11b que transmite hasta 11 Mbit/s usando la banda de 2.4 GHz. Después de estar dentro de la red se hace una configuración para el envío y recepción de información por MQTT, suscribiéndonos a un tema y publicando la cadena de datos que se creó anteriormente.

- Proceso 2: La información llega del microcontrolador a la Raspberry; esta cuenta con una conexión al router de la misma forma que el ESP, por TCP/IP. La velocidad de transferencia de datos de la Raspberry es por medio cableado y la velocidad es de hasta 100 Mbps. La Raspberry es un servidor multiservicio que gestiona el broker MQTT con la plataforma libre Mosquitto, el servidor de gestión gráfica NodeRED y la base de datos. La información llega al tema del broker seleccionado y se almacena en un documento de texto plano en un espacio de memoria de 2 Gb.
- Proceso 3: La última etapa es en la que se solicita la información para que el usuario pueda hacer la representación gráfica de los resultados. La información se solicita por FTP (Protocolo de Transferencia de Archivos) desde cualquier cliente FTP hacia la Raspberry y se descarga el archivo. Después de descargar el archivo con las lecturas se optó por graficar los datos en MS Access. Una vez que se tenían las gráficas se obtuvo una curva de tendencia, para ver si era posible tener una ecuación que se acercara al comportamiento de la temperatura y la humedad.

3. Resultados

El resultado del proceso de adquisición fue un archivo de texto con los parámetros de temperatura y humedad separados por comas, el cual se importó, por medio de FTP, en la computadora del usuario final, para después cargar la

información en la herramienta Access para visualizar los datos. En la figura 3 se puede ver la temperatura y humedad del día 20 de marzo del 2016 junto con sus líneas de tendencia.

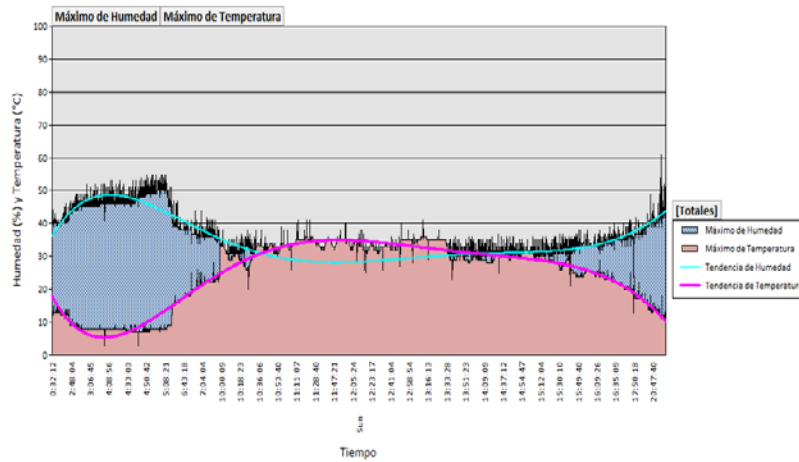


Figura 3 Gráfica del 20 de marzo de 2016 con líneas de tendencia.

El proceso de adquisición de datos de forma remota utiliza los elementos descritos en el método y nos otorga los valores que se esperaban en comparación con la estación Granja Elsa, Aguascalientes, ver figura 4.

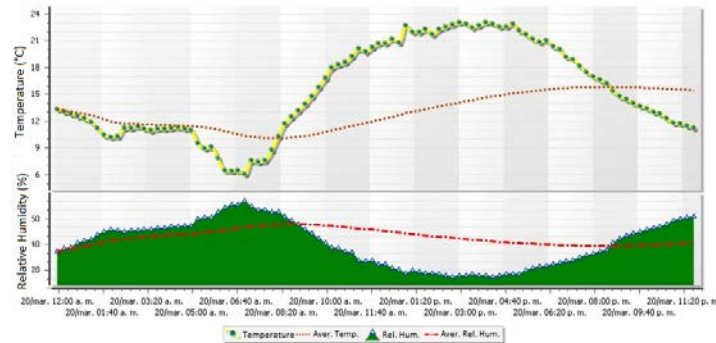


Figura 4 Datos climáticos 20/marzo/2016, estación meteorológica Granja Elsa, Ags.

4. Discusión

El propósito de combinar diferentes herramientas es el de simplificar la expansión y mantenimiento de la red. El protocolo MQTT será pieza clave en la expansión del sistema ya que una vez que se establecen los paramentos de comunicación entre dos dispositivos; refiriéndonos a dispositivos como sensores,

actuadores, computadoras y más, solamente hay que unirse en forma de suscripción a la red de transferencia de datos y con esto poder mandar y recibir información.

El servidor Node-RED es una herramienta gráfica que nos va a permitir dar de alta nuevas conexiones y funciones, agregando bloques. Los bloques aportan nuevas funciones para conectar más dispositivos, con configuraciones menores que incrementan las capacidades del sistema. La Raspberry está basada en Linux, un sistema operativo libre, que nos permite hacer numerosas pruebas con diferentes herramientas de red sin la necesidad de pagar derechos por el uso del software. En un futuro, si se decide desarrollar el sistema con exigencias que el sistema no soporte, será muy fácil migrar a un equipo con mayor capacidad haciendo uso de Linux como sistema y migrando los servicios instalados en la Raspberry. Se sabe que existen servidores como Bluemix que cubren todo el marco que se pretende en este proyecto y se ve como una alternativa para la expansión del sistema.

5. Conclusiones

La implementación del sistema de adquisición de datos ha sido satisfactoria, dado que se logró utilizar un sensor dentro de un invernadero y almacenar los datos de temperatura y humedad obtenidos en un servidor dentro de una red local, por medio de redes inalámbricas, utilizando hardware y software libre, acompañado de protocolos y sistemas de implementación amigables con el administrador y el usuario.

El análisis de los resultados nos permite ver que se pueden hacer algunos cambios para mejorar el sistema en cuanto a las capacidades del sensor y la herramienta de representación gráfica y las capacidades de almacenamiento.

En trabajos futuros se planea incluir más sensores; un administrador de bases de datos, como MySQL con una capacidad de almacenamiento superior a los 2Gb que se establecieron, un servidor de representación gráfica que nos permita hacer consultas de diferentes periodos de tiempo, además de graficar el estado en el momento de supervisión, en la Raspberry, como sustituto de MS Access y llevar la plataforma de comunicación de una red local, a una red a través de internet.

6. Bibliografía y Referencias

- [1] AOSONG. Especificaciones del módulo DHT11: <https://akizukidenshi.com/download/ds/aosong/DHT11.pdf>, 2017.
- [2] Blackstock M. & Lea R. Toward a Distributed Data Flow Platform for the Web of Things. Proceedings of the 5th International Workshop on Web of Things, October 2014.
- [3] Chang, C.L., Chang, K.P., & Song, G.B. Design and Implementation of a Cloud-Based LED Lighting Control System for Protected Horticulture. *Applied Engineering in Agriculture*, 32(6), pp. 697-706, 2016.
- [4] Cossu, M., Murgia, L., Ledda, L., Deligios, P.A., Sirigu, A., Chessa, F., & Pazzona, A. Solar radiation distribution inside a greenhouse with south-oriented photovoltaic roofs and effects on crop productivity. *Applied Energy*, 133, pp. 89–100, 2014.
- [5] DC-Square, HiveMQ, 2016: <http://www.hivemq.com/> el 2 de abril de 2016.
- [6] Du Y., Xue Z., Zhu Q., Liu X., Feng Y., and Zhang S. Design and Application of Intelligent Control System for Greenhouse Environment Based on CAN bus. Proceedings of International Conference on Modeling, Identification & Control, 2013.
- [7] Fang, J., & Wang, F. Design of greenhouse remote monitoring system based on LabVIEW. *Computer Science and Automation Engineering*, pp. 536-539, 2011.
- [8] Ferrarezi, R.S., Dove, S.K., & Van Lersel, M.W. An automated system for monitoring soil moisture and controlling irrigation using low-cost open-source microcontrollers. *HortTechnology*, 25(1), pp. 110-118, 2015.
- [9] De Anda, J. & Shear, H. Potential of Vertical Hydroponic Agriculture in Mexico. *Sustainability*, 9(1), pp. 140, 2017.
- [10] Espressif Systems, Especificaciones ESP-12E, 2017: <http://d1jy6p8pks3hof.cloudfront.net/datasheets/ESP12Espec.pdf>, 2017.
- [11] Fezari M., Khati A. and Boumaza M.S. Implementation of Wireles Sensor Network for Automatic Greenhouse Monitorign. *Communications, Computing and Control Applications*, 2011.

- [12] Guofang, L., Lidong, C., Yubin, Q., Shengtao, L., & Junyu, X. Remote Monitoring System of Greenhouse Environment Based on LabVIEW. International Conference on Computer Design and Applications, Vol. 2, 2010.
- [13] Juárez-Gutiérrez, S.S., Gárate-García, A., Ramírez del-Real, T. A., & Álvarez-Sánchez, E.J. Temperature Modeling of a Greenhouse Environment, Handbook of Research on Military, Aeronautical, and Maritime Logistics and Operations, IGI Global, pp. 257, 2016.
- [14] Márquez-Vera, M.A., Ramos-Fernández, J.C., Cerecero-Natale, L.F., Lafont, F., Balmat, J.F. & Esparza-Villanueva, J.I. Temperature control in a MISO greenhouse by inverting its fuzzy model. Computers and Electronics in Agriculture, 124, pp. 168–174, 2016.
- [15] Khot, S.B., & Gaikwad, M.S. Development of cloud-based Light intensity monitoring system for greenhouse using Raspberry Pi. International Conference on Computing Communication Control and automation, pp. 1-4, 2016.
- [16] Makhlouf, S., Laghrouche, M. & Adane, A.E.H. Hot Wire Sensor-Based Data Acquisition System for Controlling the Laminar Boundary Layer Near Plant Leaves Within a Greenhouse. IEEE Sensors Journal, 16(8), pp. 2650-2657, 2016.
- [17] Mad S. S., Munirah L., Kamarudin K., Mohd W., Syed M. M., Muhammad S., Zakaria A. and Nor M. Real-Time Greenhouse Monitoring System for Mango with Wireless Sensor Network. International Conference on Electronic Design, 2014.
- [18] Ramirez-Villegas, J., Jarvis, A. & Läderach, P. Empirical approaches for assessing impacts of climate change on agriculture: The EcoCrop model and a case study with grain sorghum. Agricultural and Forest Meteorology, 170, pp. 67-78, 2013.
- [19] Outanoute, M., Lachhab, A., Ed-dahhak, A., Guerbaoui, M., Selmani, A. & Bouchikhi, B. Synthesis of an Optimal Dynamic Regulator Based on Linear Quadratic Gaussian (LQG) for the Control of the Relative Humidity Under

- Experimental Greenhouse. *International Journal of Electrical and Computer Engineering*, 6(5), pp. 2262-2273, 2016.
- [20] Kang, Y., Khan, S. & Ma, X. Climate change impacts on crop yield, crop water productivity and food security – A review. *Progress in Natural Science*, 19 (12), pp. 1665-1674, 2009.
- [21] Leal Iga, J., Alcorta García, E., & Rodríguez Fuentes, H. Modelado del clima en invernaderos: respuesta de la temperatura a cambios de humedad. *Ingenierías*, 9(33), pp. 7-13, 2006.
- [22] Moulton, A.A. & Popke, J. Greenhouse governmentality: Protected agriculture and the changing biopolitical management of agrarian life in Jamaica. *Environment and Planning D: Society and Space*, pp. 1-19, 2016.
- [23] Pawlowski, A., Beschi, M., Guzmán, J.L., Visioli, A., Berenguel, M. & Dormido, S. Application of SSOD-PI and PI-SSOD event-based controllers to greenhouse climatic control. *ISA Transactions*, 65, pp. 525-536, 2016.
- [24] Raspberry-Pi-Foundation. (s.f.). Raspberry Pi. Recuperado de <https://www.raspberrypi.org/> el 2 de abril de 2016.
- [25] Sharma, H., Shukla, M.J., Bosland, P.W. & Steiner, R. Soil moisture sensor calibration, actual evapotranspiration, and crop coefficients for drip irrigated greenhouse chile peppers. *Agricultural Water Management*, Vol. pp. 179, 81–91, 2017.
- [26] Vishnukumar, V., Kumar, S., and Madhusoodanan, K. N. Wireless sensor networks for internet of things based laboratory automation system. *International Journal of Research in Engineering and Technology*, 5 (22), 2016.
- [27] Xu, Y.H., Wu, W.L., Xu, Y., Tham, M.L. & Ramli, N. A framework of fuzzy control-based intelligent control system for greenhouse. *Artificial Intelligence Research*, 6(1), pp. 1-5, 2017.
- [28] Xiaoyan Z., Xiaoyan Z., Chen D., Zhaohui C., Shangming S., and Zhaohui Z. The Design and Implementation of the Greenhouse Monitoring System Based on GSM and RF Technologies, *Computational Problem-solving (ICCP)*, 2013.

ONLINE PARAMETRIC IDENTIFICATION OF MASS- SPRING-DAMPER MECHANICAL SYSTEMS USING ACCELERATION MEASUREMENTS

Francisco Beltrán Carbajal

Universidad Autónoma Metropolitana, Unidad Azcapotzalco

fbeltran@azc.uam.mx

Gerardo Silva Navarro

Centro de Investigación y de Estudios Avanzados del IPN

gsilva@cinvestav.mx

Luis Gerardo Trujillo Franco

Centro de Investigación y de Estudios Avanzados del IPN

ltrujillo@cinvestav.mx

Resumen

Implementación de esquemas de control activo de vibraciones, detección de fallas o tareas de monitoreo de la operación adecuada de estructuras mecánicas flexibles pueden requerir el uso de técnicas de identificación paramétrica ejecutadas en línea. Mediciones de señales de aceleración se usan en varias aplicaciones de identificación de parámetros en sistemas mecánicos vibratorios. En este artículo se propone un enfoque para estimación de parámetros en línea en el dominio del tiempo para sistemas mecánicos del tipo masa-resorte-amortiguador de n grados de libertad, usando únicamente mediciones de aceleración. Se usa integración por partes en la síntesis del método de identificación de parámetros propuesto. De esta manera, conocimiento previo de las condiciones iniciales del sistema son innecesarias. El método de estimación propuesto se puede extender para estimación paramétrica en tiempo real para sistemas mecánicos vibratorios no lineales, completamente actuados o subactuados. Se incluyen algunos resultados de simulación numérica para mostrar la efectividad del enfoque de estimación de parámetros de masa, rigidez y

amortiguamiento, combinado con tareas de seguimiento de trayectorias de referencia en lazo-cerrado especificadas para el sistema mecánico vibratorio.

Palabras Claves: control activo de vibraciones, identificación de parámetros, sistemas mecánicos vibratorios, sistemas masa-resorte-amortiguador.

Abstract

Implementation of active vibration control schemes, failure detection and monitoring tasks of the suitable operation of flexible mechanical structures can require the use of on-line parametric identification techniques. Measurements of acceleration signals are preferred in several applications of parameter identification of vibrating mechanical systems. In this article, an on-line parameter estimation approach in time domain is proposed for linear mass-spring-damper mechanical systems of n degrees of freedom using acceleration measurements solely. Integration by parts is properly used in the synthesis of the proposed parameter identification method. In this fashion, a priori knowledge of the initial conditions of the system becomes unnecessary. The introduced identification method can be extended for real-time parametric estimation of nonlinear fully actuated or under-actuated nonlinear vibrating mechanical systems. Some numerical results are provided to show the effectiveness of the on-line estimation approach of the mass, stiffness and damping parameters combined with closed-loop reference trajectory tracking tasks specified for the vibrating mechanical system.

Keywords: Active vibration control, mass-spring-damper systems, mechanical vibration systems, parameter identification.

1. Introduction

Identification of vibration mechanical systems is an active research subject and its results admit several practical applications. Diverse methodologies have been mainly proposed for off-line estimation of parameters in the time domain or in the frequency domain [Soderstrom, 1989], [Isermann, 2011], [Ljung, 1987]. Some off-line estimation methods of modal parameters for mechanical systems are also described in [Heylen, 2003], [Le, 2013] and [Yang, 2013].

Recently, an on-line parametrical identification method for continuous-time constant linear systems has been proposed in [Fliess, 2003]. This algebraic approach is based on powerful mathematical tools of module theory, differential algebra and operational calculus. It is assumed that a mathematical model of the dynamic system is available for the synthesis of some parameter identifier/estimator. Thus, a suitable structure of a mathematical model describing the system dynamics is employed for the algebraic estimation of its parameters in a small time interval. Reasonably slow changes of some parameter values are also admitted during the system operation.

Algebraic identification has been successfully applied for on-line estimation of parameters and signals for vibrating mechanical systems using position measurements in [Beltran, 2015a], [Beltran, 2014], [Beltran, 2013] and [Beltran, 2013]. Harmonic forces can also be reconstructed on-line by employing algebraic system identification techniques [Beltran, 2015b]. Algebraic estimation of the frequency and amplitude of exogenous harmonic excitations in damped Duffing systems with an autoparametric pendulum vibration absorber has been introduced in [Silva, 2013].

This paper presents an on-line parameter estimation approach in continuous time domain for mass-spring-damper linear mechanical systems of n degrees of freedom using acceleration measurements solely. The presented results constitute a natural extension of previous works based on parameter identification using position measurements [Beltran, 2015a] and Mikusinski operational calculus [Mikusinski, 1983]. Integration by parts is properly used in the synthesis of the proposed parameter identification method. In this way, position and velocity measurements and priori knowledge of the initial conditions of the system are avoided. Thus, algebraic estimators for mass, stiffness and damping parameters can be reseeded and updated continuously for operation scenarios where slow changes of the parameter values are expected. The introduced identification method can be extended for real-time parametric estimation of nonlinear fully actuated or under-actuated nonlinear vibrating mechanical systems. In fact, algebraic identification has been applied to sequentially estimate the parameters of

nonlinear mass-spring-damper systems using position measurements [Beltran, 2014]. Some numerical simulation results are included to show the effectiveness of the on-line algebraic identification approach combined with adaptive-like reference trajectory tracking control for a Multiple-Input-Multiple-Output (MIMO) mechanical system of 3 degrees of freedom.

2. Methods

In the present section, the proposed algebraic parametric identification method for linear mass-spring-damper mechanical systems of n degrees of freedom using acceleration measurements is developed in detail.

Parametric Identification of a Mass-Spring-Damper System of n Degrees of Freedom

Firstly, consider the n DOF linear vibrating mechanical system schematically described in figure 1. Here, $x_i, i=1,2,\dots,n$, are the position coordinates, u_i the force control inputs, and m_i, k_i and c_i denote mass, stiffness and viscous damping parameters associated to the i -th DOF. The mathematical model of this MIMO flexible mechanical system is given by equation 1.

$$\begin{aligned}
 m_1 \ddot{x}_1 + c_1 \dot{x}_1 + k_1 x_1 + k_2 (x_1 - x_2) &= u_1 \\
 m_2 \ddot{x}_2 + c_2 \dot{x}_2 + k_2 (x_2 - x_1) + k_3 (x_2 - x_3) &= u_2 \\
 &\vdots \\
 m_{n-1} \ddot{x}_{n-1} + c_{n-1} \dot{x}_{n-1} + k_{n-1} (x_{n-1} - x_{n-2}) + k_n (x_{n-1} - x_n) &= u_{n-1} \\
 m_n \ddot{x}_n + c_n \dot{x}_n + k_n (x_n - x_{n-1}) + k_{n+1} x_n &= u_n
 \end{aligned}
 \tag{1}$$

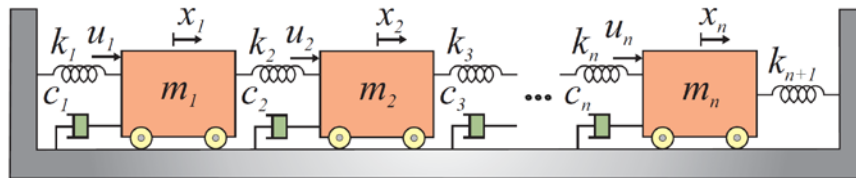


Figure 1 Schematic diagram of a n DOF mass-spring-damper system.

In the synthesis of the proposed parametric identification scheme, it is considered that only measurements of the acceleration output variables $y_i = \ddot{x}_i$ and the control

inputs u_i are available. Therefore, Equation 1 are differentiated twice with respect to time and then multiplied by $\Delta^2 = (t-t_0)^2$ in order to avoid dependence on initial conditions of the system, where t_0 is the start time when the parameter identification process is performed, resulting equation 2.

$$\begin{aligned}
 m_1 \Delta^2 x_1^{(4)} + c_1 \Delta^2 x_1^{(3)} + k_1 \Delta^2 \ddot{x}_1 + k_2 \Delta^2 (\ddot{x}_1 - \ddot{x}_2) &= \Delta^2 \ddot{u}_1 \\
 m_2 \Delta^2 x_2^{(4)} + c_2 \Delta^2 x_2^{(3)} + k_2 \Delta^2 (\ddot{x}_2 - \ddot{x}_1) + k_3 \Delta^2 (\ddot{x}_2 - \ddot{x}_3) &= \Delta^2 \ddot{u}_2 \\
 &\vdots \\
 m_{n-1} \Delta^2 x_{n-1}^{(4)} + c_{n-1} \Delta^2 x_{n-1}^{(3)} + k_{n-1} \Delta^2 (\ddot{x}_{n-1} - \ddot{x}_{n-2}) + k_n \Delta^2 (\ddot{x}_{n-1} - \ddot{x}_n) &= \Delta^2 \ddot{u}_{n-1} \\
 m_n \Delta^2 x_n^{(4)} + c_n \Delta^2 x_n^{(3)} + k_n \Delta^2 (\ddot{x}_n - \ddot{x}_{n-1}) + k_{n+1} \Delta^2 \ddot{x}_n &= \Delta^2 \ddot{u}_n
 \end{aligned} \tag{2}$$

Double integration by parts of equation 2 with respect to time yields to equation 3.

$$a_{11,i}(t)m_i + a_{12,i}(t)c_i + a_{13,i}(t)k_i + a_{14,i}(t)k_{i+1} = b_{1,i}(t), \quad i = 1, 2, \dots, n \tag{3}$$

With

$$\begin{aligned}
 a_{11,i} &= \Delta^2 y_i - 4 \int_{t_0}^t \Delta y_i(\tau_1) d\tau_1 + 2 \int_{t_0}^t \int_{t_0}^{\tau_2} y_i(\tau_1) d\tau_1 d\tau_2 \\
 a_{12,i} &= \int_{t_0}^t \Delta^2 y_i d\tau_1 - 2 \int_{t_0}^t \int_{t_0}^{\tau_2} \Delta y_i(\tau_1) d\tau_1 d\tau_2 \\
 a_{13,i} &= \int_{t_0}^t \int_{t_0}^{\tau_2} \Delta^2 [y_i(\tau_1) - y_{i-1}(\tau_1)] d\tau_1 d\tau_2 \\
 a_{14,i} &= \int_{t_0}^t \int_{t_0}^{\tau_2} \Delta^2 [y_i(\tau_1) - y_{i+1}(\tau_1)] d\tau_1 d\tau_2 \\
 b_{1,i} &= \Delta^2 u_i - 4 \int_{t_0}^t \Delta u_i(\tau_1) d\tau_1 + 2 \int_{t_0}^t \int_{t_0}^{\tau_2} u_i(\tau_1) d\tau_1 d\tau_2
 \end{aligned} \tag{4}$$

Where $y_i = \ddot{x}_i$, and $y_0 = y_{n+1} \equiv 0$.

Equation 3, after three more integrations, leads to the linear system of equations 5.

$$\theta_i = A_i^{-1} B_i = \frac{1}{\Delta_i} [\Delta_{1,i} \quad \Delta_{2,i} \quad \Delta_{3,i} \quad \Delta_{4,i}]^T, \quad i = 1, 2, \dots, n \tag{5}$$

Where $\theta_i = [m_i, c_i, k_i, k_{i+1}]^T$ is the vector of positive constant parameters to be identified, A_i and B_i are 4×4 and 4×1 matrices, respectively, described by

$$A_i = \begin{bmatrix} a_{11,i} & a_{12,i} & a_{13,i} & a_{14,i} \\ a_{21,i} & a_{22,i} & a_{23,i} & a_{24,i} \\ a_{31,i} & a_{32,i} & a_{33,i} & a_{34,i} \\ a_{41,i} & a_{42,i} & a_{43,i} & a_{44,i} \end{bmatrix}, B_i = \begin{bmatrix} b_{1,i} \\ b_{2,i} \\ b_{3,i} \\ b_{4,i} \end{bmatrix}$$

Whose components are considered as output signals of the dynamic system, equation 6.

$$\begin{aligned} \dot{a}_{kh,i} &= a_{k-1h,i} \\ \dot{b}_{k,i} &= b_{k-1,i} \end{aligned} \tag{6}$$

With $i = 1, 2, \dots, n$, $k = 2, 3, 4$, $h = 1, 2, 3, 4$, and zero initial conditions at $t = t_0$.

Hence, the estimators (equation 7) are proposed for the algebraic estimation of the mass, damping and stiffness parameters using measurements of acceleration signals y_i :

$$\begin{aligned} \widehat{m}_i &= \frac{\int_{t_0}^{(2)} |\Delta_{1,i}|}{\int_{t_0}^{(2)} |\Delta_i|}, & \widehat{c}_i &= \frac{\int_{t_0}^{(2)} |\Delta_{2,i}|}{\int_{t_0}^{(2)} |\Delta_i|} \\ \widehat{k}_i &= \frac{\int_{t_0}^{(2)} |\Delta_{3,i}|}{\int_{t_0}^{(2)} |\Delta_i|}, & \widehat{k}_{i+1} &= \frac{\int_{t_0}^{(2)} |\Delta_{4,i}|}{\int_{t_0}^{(2)} |\Delta_i|}, \forall t > t_0 > 0 \end{aligned} \tag{7}$$

Where $\widehat{(\cdot)}$ denotes estimated parameter and $\int_{t_0}^{(2)} \phi(t)$ the iterated integral of the form $\int_{t_0}^t \int_{t_0}^{\tau_2} \phi(\tau_1) d\tau_1 d\tau_2$.

3. Results

Main results of the present contribution are constituted by the algebraic formulas or estimators (7) to compute mass, stiffness and damping parameters in flexible mechanical systems using measurements of acceleration signals.

In this section, some numerical simulation results are included to depict the effectiveness of the parameter identification approach on a 3 DOF MIMO mechanical system characterized by the set of parameters described in table 1.

Table 1 Parameters of the vibrating mechanical system.

Mass (kg)	Damping (N.s/m)	Stiffness (N/m)
$m_1 = 2.0$	$c_1 = 5.5$	$k_1 = 1000$
$m_2 = 2.5$	$c_2 = 5.0$	$k_2 = 900$
$m_3 = 3.0$	$c_3 = 4.5$	$k_3 = 900$
		$k_4 = 700$

The main interest of the present work resides on the fast identification of the system parameters. Thus, constant or variable forces u_i can be applied to the mechanical system to get estimates of the mass, damping and stiffness parameters in a small period of time. Nevertheless, the output feedback control scheme proposed in [Beltran, 2015a] was used to assess the dynamic performance of the estimation approach for reference trajectory tracking tasks, equations 8 a la 10.

$$u_i = \widehat{m}_i v_i + \widehat{c}_i \widehat{\dot{x}}_i + \widehat{k}_i (x_i - x_{i-1}) + \widehat{k}_{i+1} (x_i - x_{i+1}) \tag{8}$$

With

$$v_i = \ddot{y}_i^* - \beta_{2,i} (\widehat{\dot{x}}_i - \dot{x}_i^*) - \beta_{1,i} (x_i - x_i^*) - \beta_{0,i} \int_0^t (x_i - x_i^*) dt \tag{9}$$

$$\begin{aligned} \widehat{\dot{x}}_i = & -\frac{\widehat{c}_i}{m_i} x_i - \frac{\widehat{k}_i}{m_i} \int_0^t (x_i - x_{i-1}) dt - \frac{\widehat{k}_{i+1}}{m_i} \int_0^t (x_i - x_{i+1}) dt \\ & + \frac{1}{m_i} \int_0^t u_i dt \end{aligned} \tag{10}$$

Where $i=1,2,3$.

Figures 2 and 3 display the accurate and fast estimation of the mass, stiffness and damping parameters using the proposed algebraic formulas 7. The satisfactory tracking of position reference trajectories planned for the vibration mechanical system is clearly manifested as well. In the next section some important discussions about the analytical and numerical results are described.

4. Discussion

A very good closed-loop estimation of the system parameters using the algebraic estimators 7 and control forces 8 was verified in figure 2. At the beginning, in the control implementation the mass values were fixed at 1 kg and

the other parameter values at 0. Next, at $t > 0.1$ the estimates were replaced in the tracking controllers. The Runge-Kutta-Fehlberg 4/5 method with step time of 1×10^{-3} s was used in the numerical tests. Hence, accurate and fast estimates of the mass, damping and stiffness parameters were computed before 0.1 s. Thus, parameter estimators can be reused and updated continuously for operation scenarios where slow changes of the parameter values are expected.

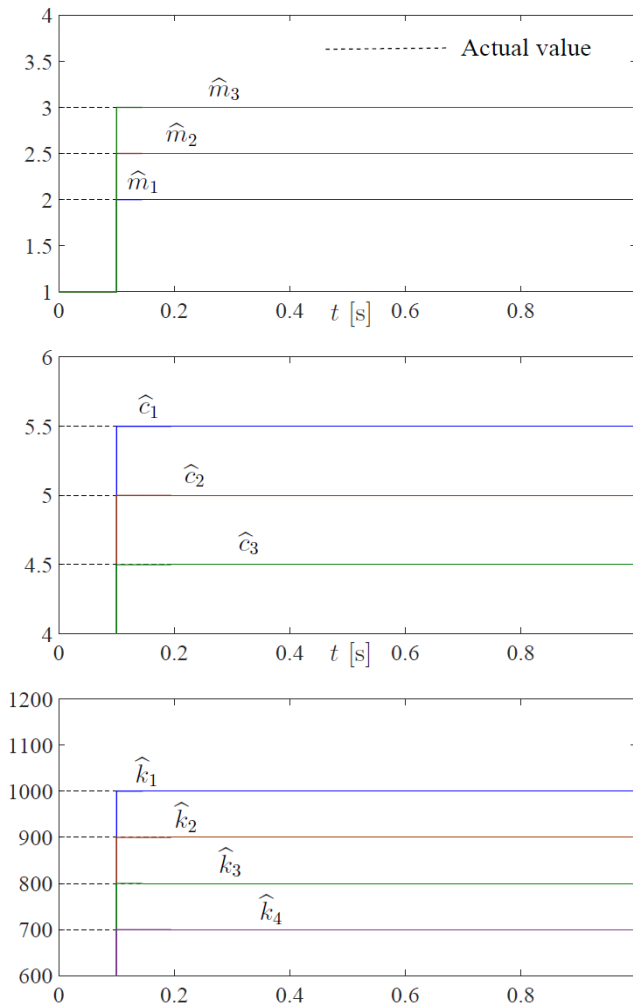


Figure 2 Closed-loop algebraic estimation of mass, damping and stiffness parameters.

A satisfactory closed-loop tracking of the reference trajectories using the estimates of the mechanical system parameters is displayed in figure 3. The desired motion profiles are described by Bézier interpolation polynomials which were defined to firstly transfer the mechanical system from a rest equilibrium state to another for

$\bar{x}_1 = 0.005$ m, $\bar{x}_2 = 0.005$ m and $\bar{x}_3 = -0.005$, and next to the rest equilibrium state again in 5 s. The controlled forces applied to the vibration system are shown in figure 4. A reasonable momentary overshoot in the force responses is presented when the estimated parameters are substituted in the control algorithms (8) at $t > 0.1$ s.

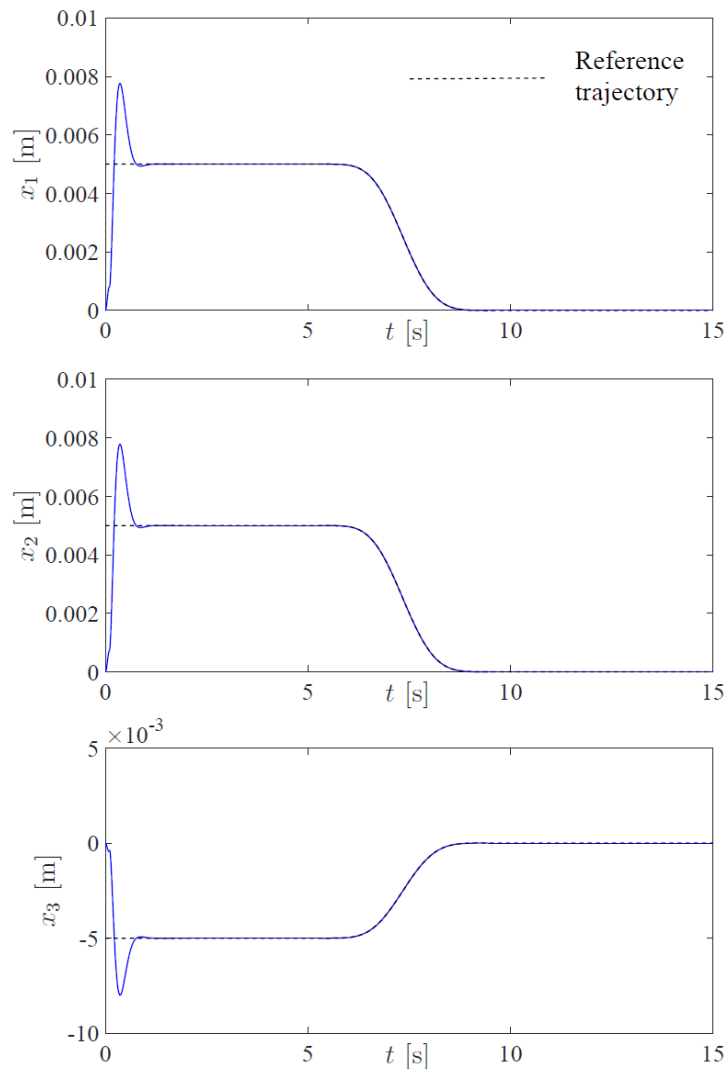


Figure 3 Control forces applied to the mechanical system for trajectory tracking tasks.

Therefore, the analytical and numerical results have confirmed the effectiveness of the on-line estimation approach of mass, stiffness and damping parameters combined with closed-loop reference trajectory tracking tasks on multiple-input multiple-output vibrating mechanical system of n degrees of freedom.

5. Conclusions

An algebraic identification method for mass, damping and stiffness parameters for linear MIMO mechanical systems of n degrees of freedom using acceleration measurements has been proposed. The presented identification approach constitutes an extension of the parametric identification method for linear mass-spring-damper mechanical system using position measurements introduced in [Beltran, 2015a]. The parameter values are estimated accurately and algebraically into a small windows of time depending on the numerical integration method and the velocity and precision of the computer processor employed for the implementation of the algebraic estimators. The dynamic performance of the proposed parameter estimators was numerically evaluated for closed-loop tracking tasks of reference trajectories on a 3 DOF MIMO mechanical system. Preliminary computer simulation results show a good estimation of the unknown parameters of the linear vibration mechanical system. Future studies will include the application of the algebraic identification approach to the on-line parametric estimation problem of nonlinear vibrating mechanical systems.

6. Bibliography and References

- [1] Beltran Carbajal, F. and Silva Navarro, G., On the Algebraic Parameter Identification of Vibrating Mechanical Systems, *International Journal of Mechanical Sciences*, Vol. 92, pp. 178-186, 2015a.
- [2] Beltran Carbajal, F. and Silva Navarro, G., and Trujillo-Franco, L.G., Evaluation of on-line algebraic modal parameter identification methods, In R. Allemang (ed.), *Topics in Modal Analysis II*, Vol. 15, No. 8, Springer, NY, pp. 145-152, 2014.
- [3] Fliess, M. and Sira-Ramírez, H., An algebraic framework for linear identification, *ESAIM: Control, Optimization and Calculus of Variations*, Vol. 9, pp. 151-168, 2003.
- [4] Le, T. P. and Paultre, P., Modal identification based on the time-frequency domain decomposition of unknown-input dynamic tests, *International Journal of Mechanical Sciences*, Vol. 71, pp. 41-50, 2013.

- [5] Heylen, W., Lammens, S. and Sas, P., *Modal Analysis, Theory and Testing*, Katholieke Universiteit Leuven, Belgium, 2003.
- [6] Beltran Carbajal, F. and Silva Navarro, G., On-line harmonic force estimation in active vibration control of mass-spring-damper systems, *Proceedings of the 22th International Congress on Sound and Vibration*, Florence, Italy, 12-16 July, 2015b.
- [7] Beltran Carbajal, F. and Silva Navarro, G., On-line algebraic parametric identification of uncertain nonlinear vibrating mechanical systems, *Proceedings of the 21th International Congress on Sound and Vibration*, Beijing, China, 13--17 July, 2014.
- [8] Beltran Carbajal, F. and Silva Navarro, G., Adaptive-like Vibration Control in Mechanical Systems with Unknown Parameters and Signals, *Asian Journal of Control*, Vol. 15, No. 6, pp. 1613-1626, 2013.
- [9] Beltran Carbajal, F., Silva Navarro, G. and Arias-Montiel, M., Active Unbalance Control of Rotor Systems Using On-line Algebraic Identification Methods, *Asian Journal of Control*, Vol. 15, No. 6, pp. 1613-1626, 2013.
- [10] Isermann, R. and M. Munchhof, *Identification of Dynamic Systems*, Springer Verlag, Berlin, 2011.
- [11] Ljung, L., *Systems Identification: Theory for the User*, Prentice-Hall, NJ, 1987.
- [12] Mikusinski, J., *Operational Calculus*, 2nd Ed., Vol. 1, PWN & Pergamon, 1983.
- [13] Silva Navarro, G., Beltran Carbajal, F. and Vazquez-Gonzalez, B. Synthesis of an adaptive-like autoperametric pendulum absorber for damped Duffing systems, *Proceedings of the 20th International Congress on Sound and Vibration*, Bangkok, Thailand, pp. 7-11 July, 2013.
- [14] Soderstrom, T. and Stoica, P., *System Identification*, Prentice-Hall, NY, 1989.
- [15] Yang, Y. and Nagarajaiah S., Output-only modal identification with limited sensors using sparse component analysis, *Journal of Sound and Vibration*, Vol. 332, pp. 4741-4765, 2013.

ROBOT MÓVIL 3.0 UNA EVALUACIÓN DE RENDIMIENTO

Saul Enrique Benítez García

Instituto Politécnico Nacional, CIDETEC

sbenitezg1100@egresado.ipn.mx

Jorge Luis de la Cruz Osorio

Instituto Politécnico Nacional, ESIME Unidad Culhuacan

jdelacruz1100@egresado.ipn.mx

Miguel Gabriel Villarreal Cervantes

Instituto Politécnico Nacional, CIDETEC

mvillarrealc@ipn.mx

Resumen

En la actualidad existe una gran demanda de sistemas robóticos que presenten un alto grado de precisión y repetibilidad con el propósito de obtener productos de mejor calidad. Uno de los aspectos claves para desarrollar sistemas con un buen desempeño es el sistema de posicionamiento y control. Por tal motivo en el presente trabajo se expone la evaluación de rendimiento de un sistema de posicionamiento con base en odometría y el sistema de control de un robot móvil 3.0. La evaluación del sistema de control se obtiene con base en el cálculo de la repetibilidad, obtenido a través del sistema de odometría del robot móvil. Por otra parte, la exactitud del sistema de localización con base en odometría se compara con los resultados obtenidos de un sistema de localización de faros activos basado en cámaras. Bajo el estándar internacional ISO-9283 se observa que la repetibilidad y la exactitud del robot móvil son apropiadas.

Palabras Claves: Exactitud, ISO-9283, repetibilidad, robot móvil 3.0, sistema de localización.

Abstract

Nowadays, robotic systems with a high accuracy and repeatability are highly used due to the requirement to obtain a superior quality in the final product. One of the main issue to carry out systems with “good” performance is the positioning and control system. For this reason, a positioning performance evaluation based on both odometry and control system of a 3.0 mobile robot, is presented in this work. The control performance evaluation is obtained by computing the repeatability using the obtained result from the odometry system of the mobile robot. On the other hand, the performance of the location system based on odometry is compared by the results obtained with a camera based active beacon location system in order to determine its accuracy. Using the International Standard ISO-9283, the repeatability and accuracy of the mobile robot is suitable.

Keywords: *Accuracy, ISO-9283, location system, mobile robot 3.0, repeatability.*

1. Introducción

Actualmente gracias a los avances tecnológicos en sistemas de medición, actuadores, sistemas de cómputo, sistemas de comunicación y sistemas de energía, han permitido que se mantenga un fuerte interés por la robótica móvil terrestre [Guerrero, et al., 2014], y en particular en el robot móvil 3.0. Este particular interés se debe a que es un robot holónimo que presenta tres grados de movilidad y no contempla grado de direccionalidad, por tal motivo posee total movilidad en el plano, es decir, que tiene la capacidad de moverse en cada instante de tiempo a cualquier dirección independientemente de su orientación [Campion, et al., 1996].

Por otro lado, en robótica móvil, se considera el problema de localización como la parte clave de dichos sistemas, dado que para que un robot móvil navegue de forma autónoma debe tener la capacidad de determinar o estimar su posición y orientación en relación con su entorno, además de conocer la posición de otros objetos o características de interés en el ambiente del robot. Con el propósito de realizar estas acciones los robots móviles utilizan un sistema de posicionamiento que puede ser de tipo absoluto o relativo [Borenstein, et al., 1996] o ambos con la

finalidad de obtener mayor fiabilidad su localización como se muestra en [Fu, et al., 2013]. En [Sin, 2007] se hace uso del GPS (del inglés Global Positioning System) para desarrollar un robot móvil que opere de manera autónoma y manual con el fin de patrullar las calles. Para la localización de un robot AIBO en [Huang, et al., 2006] se propone un método de posicionamiento absoluto configurando 4 altavoces como faros activos de sonido en una posición conocida. En [Guerrero, et al., 2014] se utiliza como sistema de localización absoluta con base en odometría, abordando el problema de seguimiento de trayectorias para un robot móvil basado en su modelo cinemático. Por tal motivo, en este trabajo se presenta la evaluación del sistema de posicionamiento con base en odometría y el sistema de control de un robot móvil 3.0 en un ambiente controlado, asumiendo que se conoce la posición cartesiana inicial y que las ruedas no deslizan. El rendimiento que presenta el robot móvil se determina con base en los parámetros de repetibilidad y exactitud que exhibe dicho sistema.

El resto del documento se encuentra organizado de la siguiente forma: en la sección 2 se describe el modelo cinemático del robot móvil 3.0 y el controlador empleado. En la sección 3 se realiza un análisis del sistema de posicionamiento global con base en odometría. En la sección 4 se exponen las características de la plataforma experimental. La evaluación del rendimiento que presenta el sistema de control y el sistema de posicionamiento del robot móvil 3.0 se presenta en la sección 5 y finalmente en la sección 6 se exhiben las conclusiones del trabajo.

2. Métodos

El desarrollo y explicación del modelo cinemático y dinámico del robot móvil 3.0 es obtenido de [Canudas, et al., 1996] y [Villarreal, 2015] donde se considera que el robot móvil está fabricado como una estructura rígida y que sus ruedas se desplazan sobre un plano horizontal, sin presentar deformaciones y deslizamiento en el punto de contacto.

Modelo Cinemático

La representación esquemática del robot móvil se presenta en la figura1, donde la posición y orientación del robot móvil se representa con el sistema de

coordenadas relativo de movimiento $\{m_f\}$, que se encuentra fijo en el centro de geométrico del robot móvil; y para representar el desplazamiento absoluto del robot móvil se utiliza el sistema de coordenadas $\{\omega_f\}$, que se encuentra fijo al plano. Siendo $\dot{\eta}_m = [\dot{x}_m, \dot{y}_m, \dot{\phi}_m]^T$ y $\dot{\eta}_w = [\dot{x}_w, \dot{y}_w, \dot{\phi}_w]^T$ las velocidades lineales y angulares en el sistema de coordenadas de movimiento y en el sistema de coordenadas absoluto respectivamente. Considerando que las ruedas 1 y 3 tienen un ángulo simétrico de $\delta = \frac{\pi}{3}$ con respecto al eje Y_m y la rueda 2 se encuentra alineado con el eje X_m , es posible representar el modelo cinemático como se muestra en la ecuación 1.

$$\dot{\eta}_w = \begin{bmatrix} \cos(\phi_w) & -\sin(\phi_w) & 0 \\ \sin(\phi_w) & \cos(\phi_w) & 0 \\ 0 & 0 & 1 \end{bmatrix} \dot{\eta}_m \quad (1)$$

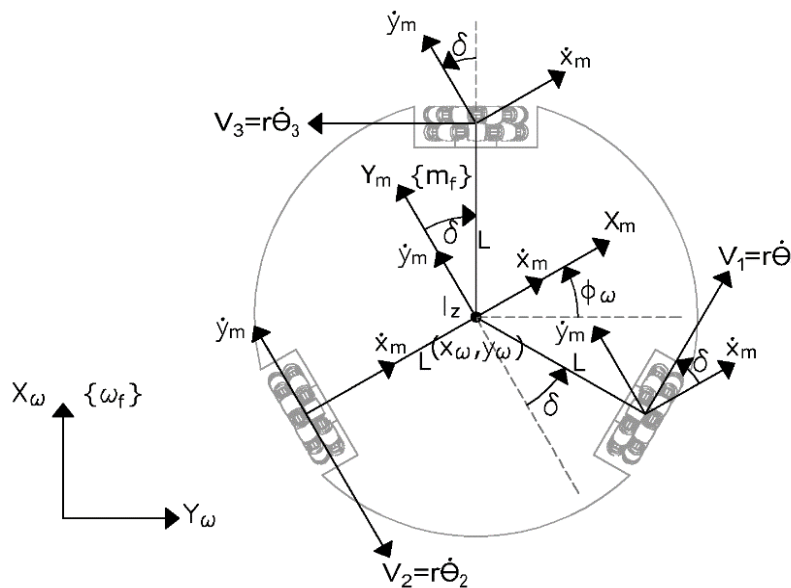


Figura 1 Diagrama esquemático del robot móvil 3.0.

El mapeo entre la velocidad lineal de las llantas y la velocidad angular y lineal en el sistema coordinado del robot móvil está dada por la ecuación 2, donde $\dot{\theta} = [\dot{\theta}_1, \dot{\theta}_2, \dot{\theta}_3]^T$ es la velocidad angular de cada rueda, $R = [r, r, r]^T = [0.0508m., 0.0508m., 0.0508m.]^T$ el radio de cada rueda y

$L = 0.1847m$. la distancia del centro geométrico del robot móvil al centro de la rueda.

$$R\dot{\theta} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} & L \\ 0 & -1 & L \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} & L \end{bmatrix} \begin{bmatrix} \dot{x}_m \\ \dot{y}_m \\ \dot{\phi}_m \end{bmatrix} \quad (2)$$

Controlador PD

Considerando el problema de regulación, sea $\dot{x} = f(x) + g(x)u$ la representación en el espacio de estados del modelo dinámico del robot móvil y $x = [x_1, x_2, x_3, x_4, x_5, x_6]$ como el vector de estados del robot móvil, se propone el controlador presentado en la ecuación (3), donde el vector de par de fuerzas de entrada en las ruedas esta definida por $u = [u_1, u_2, u_3]^T$ y considerando $\bar{v} = k_p e + k_d \dot{e}$, donde $e = [\bar{x}_d - x_1, \bar{y}_d - x_2, \bar{\phi}_d - x_3]^T$ es el error de posición lineal y angular del móvil entre la posición deseada $\bar{x}_d, \bar{y}_d, \bar{\phi}_d$ y los estados reales del sistema x_1, x_2, x_3 , $\dot{e} = [\dot{\bar{x}}_d - x_4, \dot{\bar{y}}_d - x_5, \dot{\bar{\phi}}_d - x_6]^T$ es la velocidad de cambio del error y $k_p = \text{diag}(k_{p1}, k_{p2}, k_{p3}) \in R^{3 \times 3}$, $k_d = \text{diag}(k_{d1}, k_{d2}, k_{d3}) \in R^{3 \times 3}$ son matrices diagonales definidas positivas que almacenan las ganancias proporcional y derivativa del controlador.

$$u = J^T \bar{v} \quad (3)$$

Cabe mencionar que las ganancias proporcionales $k_p = \text{diag}(11, 11, 2.4)$ y derivativa $k_d = \text{diag}(3, 3, 2.3)$, fueron obtenidas bajo el enfoque heurístico mostrado en [Villarreal, 2012], donde se realiza la sintonización del controlador mediante experimentación y la observación de los mejores resultados.

Sistema de Posicionamiento Absoluto con Base en Odometría

En el presente trabajo para proporcionar de autonomía al robot móvil, se ha implementado un sistema de posicionamiento absoluto con base en odometría

debido a que brinda una buena precisión a corto plazo, es de bajo costo y permite altas frecuencias de muestreo, como es el caso que se presenta en [Guerrero, et al., 2014], donde se ha utilizado dicho sistema para la localización del robot móvil. Este sistema de posicionamiento es sencillo de implementar debido a que esta con base en ecuaciones que transforman el número de revoluciones de las ruedas en desplazamientos lineales relativos al suelo, considerando una condición de partida definida previamente por el usuario.

Asumiendo que el robot móvil se encuentra en un ambiente controlado, es decir, el espacio de trabajo es una superficie totalmente plana y sin inclinaciones, no existe deslizamiento de las ruedas, y los parámetros cinemáticos del robot móvil presentan una adecuada caracterización, el sistema de posicionamiento con base en odometría es una opción factible. Es bien conocido que al no tomar en cuenta la consideración anteriormente comentada y al no contemplar referencias externas de posicionamiento se genera una acumulación de error considerable. Para la corrección de este error se requiere la implementación de sistemas de procesamiento y medición de alta velocidad y precisión, o bien implementar métodos recientemente desarrollados en múltiples investigaciones, como es el caso que se presenta en [Xu, 2009], donde hace uso de redes neuronales para la calibración y corrección de errores en el sistema de odometría y a su vez realiza un reconocimiento del ambiente en el que se encuentra operando.

Para que el sistema de odometría determine la posición del robot móvil, es requerida una lectura continua de los pulsos generados por los codificadores rotatorios para estimar el desplazamiento lineal de cada rueda a partir de sus respectivos desplazamientos angulares. El esquema mostrado en la figura 2 se representa el funcionamiento general del giro de un motor, el efecto de reducción y el desplazamiento lineal de la rueda.

Dónde: q_i representa el desplazamiento angular del i -ésimo motor, $M = 51$ la relación de movimiento entre la entrada y la salida de la caja de engranajes, r el radio de la rueda, θ_i y ΔS_i es el desplazamiento angular y lineal de la i -ésima rueda, respectivamente. Para el caso de interés se tiene que $i = 1, 2, 3$.

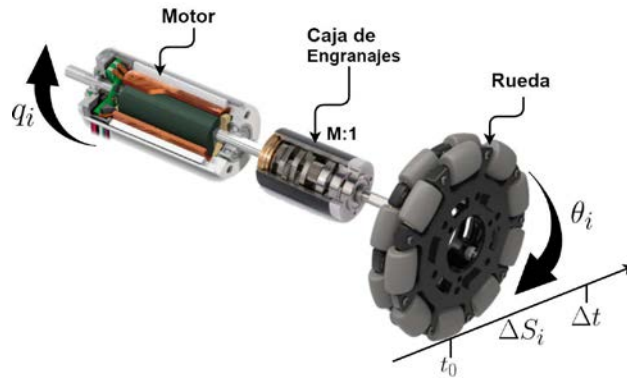


Figura 2 Diagrama esquemático del acoplamiento rueda, caja de engranes y motor.

Por otro lado, se ha considerado que la relación M y el radio de la rueda r , es el mismo para cada motor y rueda respectivamente.

El desplazamiento angular θ_i es obtenido con base en el desplazamiento angular q_i de i -ésimo motor, o de pulsos N_p^i de la i -ésima rueda en un intervalo de tiempo Δt , como se muestra en la ecuación 4 donde $P_v = 256$ es el número de pulsos por vuelta generados por el codificador rotatorio.

$$\theta_i = \frac{q_i}{M} = 2\pi \left(\frac{N_p^i}{P_v M} \right) \quad (4)$$

La transformación de desplazamiento angular a desplazamiento lineal esta dado por la ecuación 5.

$$\Delta S_i = r [\theta_i(t) - \theta_i(t - \Delta t)] = r \Delta \theta_i(t) \quad (5)$$

De las ecuaciones 1 y 2 del modelo cinemático del robot móvil, se obtiene la ecuación 6, que representa las velocidades del robot móvil en el espacio de operación en función del desplazamiento lineal de cada una de las ruedas.

$$\dot{\eta}_w = \begin{bmatrix} \cos(\phi_w) & -\sin(\phi_w) & 0 \\ \sin(\phi_w) & \cos(\phi_w) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} & L \\ 0 & -1 & L \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} & L \end{bmatrix}^{-1} R \dot{\theta} \quad (6)$$

Realizando una integración numérica, se obtienen las ecuaciones 7, 8 y 9 para determinar la posición absoluta del robot móvil el sistema coordenado $\{\omega_f\}$ para cada intervalo de tiempo. Donde x_{wi} , y_{wi} , y ϕ_{wi} representan las condiciones iniciales.

$$x_1(t + \Delta t) = x_{wi} - \frac{\Delta S_1}{3} (\sin \phi_w - \sqrt{3} \cos \phi_w) - \frac{\Delta S_3}{3} (\sin \phi_w - \sqrt{3} \cos \phi_w) + \frac{2\Delta S_2}{3} \sin \phi_w \quad (7)$$

Plataforma Experimental

EL robot móvil cuenta con una estructura base rígida, que fue manufacturada en aluminio debido a que es un material resistente y de alta durabilidad; además de ser ligero, de bajo costo y presenta una alta resistencia a la corrosión. Por otro lado el robot cuenta con una cubierta de acrílico negro de **6mm**, proporcionándole protección y estética. Las dimensiones características del robot móvil (figura 3) en altura y radio son de **0.23 m** y **0.215 m** respectivamente.



Figura 3 Plataforma Experimental.

El prototipo experimental cuenta con una interfaz de comunicación y cuatro subsistemas, figura 4:

- Interfaz de comunicación: La interfaz de comunicación emplea una computadora con MATLAB y un dispositivo de radio frecuencia XBee-Serie-1 para realizar la inicialización de variables y el registro del comportamiento de la plataforma experimental de forma inalámbrica.
- Subsistema de energía: El subsistema de energía cuenta con dos modos de operación: con baterías internas o con fuente externa de alimentación de Corriente Directa (CD), donde el modo de operación es seleccionado mediante un interruptor eléctrico. Durante el modo de operación por fuente externa es posible realizar la carga correspondiente de las baterías internas, esto con la finalidad de seguir operando el prototipo y evitar contratiempos. Por otro lado, se utilizó un convertidor tipo Buck Pololu

D24V22F5 de 5 V a 2.5 A, que proporciona el voltaje y corriente adecuados para los elementos electrónicos

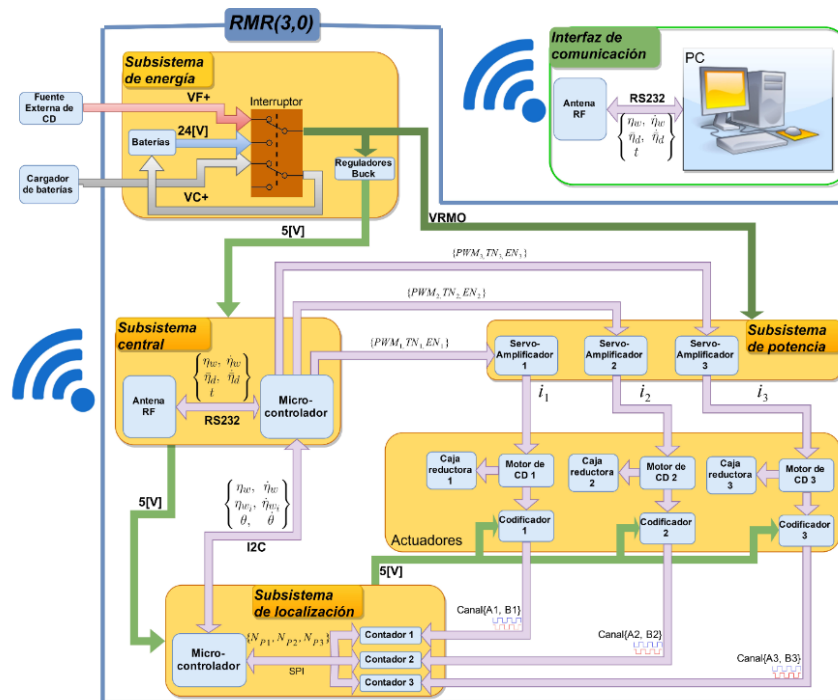


Figura 4 Diagrama del sistema de control.

- Subsistema de localización: El funcionamiento de este subsistema esta con base en dos etapas:
 - ✓ Etapa de adquisición de datos, donde se realiza la lectura de las tres señales de los codificadores ópticos acoplados a los motores y por medio del circuito integrado (C.I.) LS7366R, se convierten en un número entero de 16 bits, el cual contiene el número de pulsos generados por el codificador. El número de pulsos es enviado a la siguiente etapa de estimación de estado mediante el protocolo de comunicación SPI (Serial Pheriferal Interface por sus siglas en inglés).
 - ✓ Etapa de estimación de posición, en donde se estiman los estados actuales del robot móvil a partir del número de pulso N_p^i obtenidos en la etapa anterior. El vector de estados actuales x es enviado

mediante el protocolo de comunicación I2C (Inter-Integrate Circuit por sus siglas en inglés) al subsistema central. Para este proceso se hace uso del micro-controlador (μC) ATmega328P.

- Subsistema central: Dividido en dos etapas que son realizadas en un μC ATmega328P:
 - ✓ La etapa de procesamiento, hace uso de un μC ATmega328P encargado de solicitar de datos al sistema de localización, prosigue con el cálculo de la señal de control necesaria para realizar una tarea específica y transformar este resultado en señales PWM equivalente a la corriente necesaria para producir dicho par con base en la ecuación 10 que resulta del modelo estático del motor, donde $K_m = 0.0389 \text{ Nm/A}$ es la constante de fuerza del motor de CD, i_a es la corriente de armadura y $M = 51$ es la reducción de la caja de engranes.
 - ✓ La etapa de envío de datos, hace uso de una antena RF Xbee-Serie-1 para la transmisión de los estados reales x a la interfaz de monitoreo.

$$\tau = K_m i_a M \quad (10)$$

- Subsistema de potencia: Este proporciona la potencia eléctrica necesaria para producir movimiento en los motores de CD y que a su vez produzcan un desplazamiento del robot móvil en el plano. Este subsistema se compone por tres servo-amplificador ESCON 50/5 que realizan la interpretación de la señal PWM (generada por el subsistema central) y convertirla en corriente que es suministrada a cada motor de CD del robot móvil.

3. Resultados

En esta sección se evalúa el rendimiento del sistema de control y el sistema de posicionamiento con base en las ecuaciones de repetibilidad y exactitud que se encuentran en el estándar ISO-9283 [Standarization, 1991]. El rendimiento del

sistema de control se evaluó con base en el parámetro de repetibilidad, que a su vez es calculado a partir de los resultados del sistema de odometría. Por otro lado evaluación del sistema de posicionamiento se determinó a partir de la exactitud calculada de los datos adquiridos por un sistema de localización de faros activos basado en cámaras. Para evaluar dichos parámetros se consideró un tiempo de muestreo de $\Delta t = 5\text{ms}$ y condiciones iniciales del robot móvil $x_{wi} = [0, 0, 0, 0, 0, 0]$.

Resultados del Sistema de Posicionamiento

Se propone posicionar al robot móvil en cuatro coordenadas cartesianas con una orientación desfasada $\phi_w = \frac{\pi}{2} \text{rad}$ entre cada una de ellas, de manera que estas cuatro coordenadas cartesianas formen los vértices de un cuadrado de 1m por lado. Los estados correspondientes a cada una de las coordenadas deseadas son:

$$P_1: \bar{x}_d = [0.5, 0.5, 0, 0, 0, 0], \quad P_2: \bar{x}_d = [-0.5, 0.5, \frac{\pi}{2}, 0, 0, 0],$$

$$P_3: \bar{x}_d = [-0.5, -0.5, \pi, 0, 0, 0] \quad \text{y} \quad P_4: \bar{x}_d = [0.5, -0.5, \frac{3\pi}{2}, 0, 0, 0].$$

Cabe mencionar que el tiempo máximo asignado para posicionarse en cada coordenada es de 8s. En la figura 5a se muestran las coordenadas cartesianas deseadas en superposición con el desplazamiento realizado por el robot móvil para posicionarse en cada una de las cuatro coordenadas cartesianas y en la figura 5b se presenta el comportamiento del robot móvil en orientación con su respectiva orientación deseada.

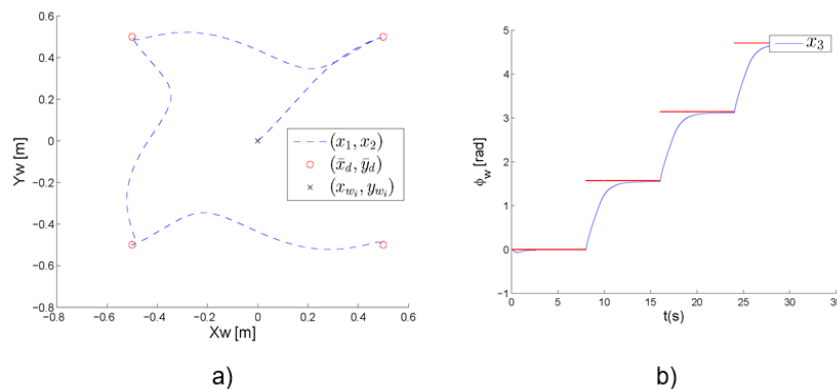


Figura 5 Comportamiento del robot móvil.

En la figura 6 se observa el error en la regulación dada por el sistema de control de $e = [\pm 0.002m, \pm 0.001m, \pm 0.032rad]$ en x_1, x_2 y x_3 respectivamente.

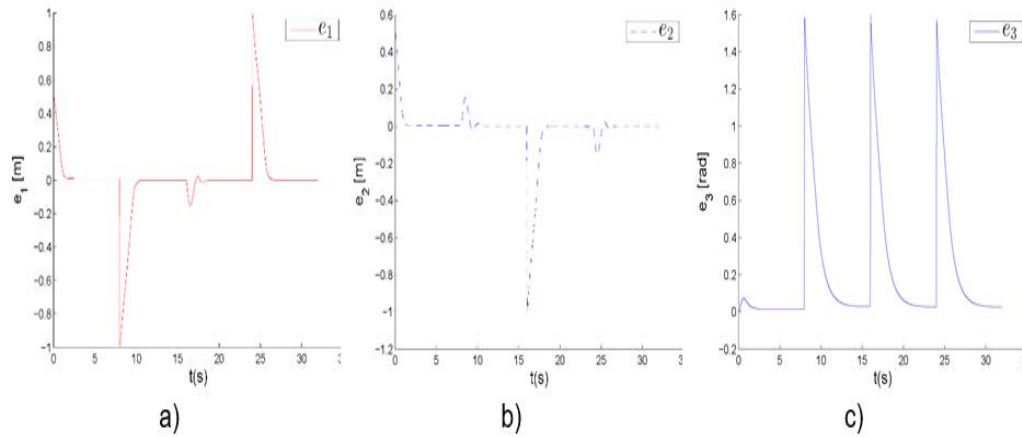


Figura 6 Error en x_w , Error en y_w . Error en ϕ_w .

Repetibilidad

La repetibilidad expresa el grado de concordancia entre las posturas alcanzadas por el robot móvil, después de m repeticiones, y la postura de referencia. Para m posturas conocidas, la repetibilidad se obtiene de acuerdo con las ecuaciones 11 y 12, donde RP_l y RP_α es la repetibilidad en posición y orientación respectivamente. Cabe mencionar que para el caso de interés se ha omitido el eje coordenado Z_w dado que el robot móvil solo trabaja en los dos ejes coordenados $X_w - Y_w$.

$$RP_l = \bar{l} + 3S_l \quad (11)$$

$$RP_\alpha = \pm 3S_\alpha \quad (12)$$

considerando:

$$\bar{l} = \frac{1}{m} \sum_{i=1}^m l_i \quad (13)$$

$$S_l = \sqrt{\frac{\sum_{i=1}^m (l_i - \bar{l})^2}{m-1}} \quad (14)$$

$$S_{\alpha} = \sqrt{\frac{\sum_{i=1}^m (x_3^i - \hat{\phi})^2}{m-1}} \quad (15)$$

$$l_i = \sqrt{(x_3^i - \hat{x}) + (x_3^i - \hat{y})^2} \quad (16)$$

Donde x_1^i , x_2^i y x_3^i representan el i -ésimo resultado de las m repeticiones para cada estado x_1 , x_2 y x_3 respectivamente, \hat{x} , \hat{y} y $\hat{\phi}$ representan el valor promedio de las m repeticiones del experimento.

Por consiguiente para obtener los parámetros de repetibilidad se ha considerado, bajo las mismas condiciones de experimentación, realizar un número de repeticiones de $m = 30$ y como estados deseados a $x_d = [0.5, 0.5, 0, 0, 0, 0]$. Los resultados de las m repeticiones se muestran en la tabla 1.

El grado de repetibilidad calculado con base en los resultados de la tabla 1, son: $RP_l = \pm 0.0011m$ y $RP_{\alpha} = \pm 0.0136rad$ en posición y en orientación respectivamente, donde estos resultados verifican la fiabilidad del sistema de control para realizar una tarea en específico bajo las mismas condiciones de operación.

Tabla 1 Resultados en estado estacionario para $m = 30$ repeticiones.

# Experimento \ GDL	x_1^i	x_2^i	x_3^i	# Experimento \ GDL	x_1^i	x_2^i	x_3^i
1	0.4963	0.4982	-0.0056	16	0.4955	0.4978	-0.0049
2	0.4961	0.4979	-0.0045	17	0.4958	0.4988	-0.0040
3	0.4962	0.4985	-0.0055	18	0.4956	0.4976	-0.0042
4	0.4955	0.4977	-0.0050	19	0.4954	0.4980	-0.0052
5	0.4960	0.4982	-0.0047	20	0.4952	0.4979	-0.0047
6	0.4960	0.4984	-0.0045	21	0.4956	0.4978	-0.0051
7	0.4957	0.4984	-0.0050	22	0.4958	0.4980	-0.0060
8	0.4960	0.4981	-0.0054	23	0.4960	0.4977	-0.0053
9	0.4952	0.4978	-0.0047	24	0.4960	0.4976	-0.0046
10	0.4959	0.4978	-0.0040	25	0.4964	0.4984	-0.0018
11	0.4953	0.4977	-0.0049	26	0.4954	0.4980	-0.0052
12	0.4957	0.4977	-0.0050	27	0.4960	0.4985	-0.0057
13	0.4957	0.4977	-0.0050	28	0.4954	0.4981	-0.0053
14	0.4955	0.4978	-0.0049	29	0.4958	0.4979	-0.0045
15	0.4958	0.4988	-0.0040	30	0.4959	0.4979	-0.0043

Exactitud

La exactitud expresa la desviación entre la postura deseada y la media de m posturas obtenidas, cuando la postura final deseada se dirige en la misma dirección. La exactitud en postura se encuentra dividido exactitud en posición (AP_p) y exactitud en orientación (AP_α), dadas por las ecuaciones 17 y 18

$$AP_p = \sqrt{(\hat{x}_e - \bar{x}_d)^2 + (\hat{y}_e - \bar{y}_d)^2} \quad (17)$$

$$AP_\alpha = (\hat{\phi}_e - \bar{\phi}_d) \quad (18)$$

Donde \hat{x}_e , \hat{y}_e y $\hat{\phi}_e$ representan el valor promedio en los estados x_1 , x_2 y x_3 para las m repeticiones del experimento.

El registro de datos para el cálculo de la exactitud fue obtenido a partir de un sistema de faros activos basado en cámaras, localizado en la Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas del Instituto Politécnico Nacional (UPIITA-IPN). Este sistema de localización cuenta con 16 cámaras OptiTrack que operan en un espacio de trabajo útil de $1.5 \times 1.5m$ y una PC con sistema operativo Windows 7 con procesador Core i7 a 3.5GHz y 32GB de memoria RAM. Este sistema obtiene la posición del robot móvil mediante la triangulación de tres marcadores retro-reflectantes que se colocan en la cara superior del robot. Cabe mencionar que este sistema presentó un margen de error $M_e = 0.07m$. Este margen de error se obtuvo mediante el siguiente experimento: se elaboró un cuadrado de $1m$ por lado, donde en cada arista y en su centro geométrico se colocaron marcadores. Se registraron las posiciones cartesianas de las marcas a través del sistema de cámaras en donde el origen era el centro del cuadrado. Se observó que las distancias del centro del cuadrado a cada vértice, presentaba una variación entre $0.04m$ a $0.07m$ entre la posición real y la medida por el sistema de cámaras.

El experimento propuesto para obtener el parámetro de exactitud consiste en obtener los estados correspondientes a P_1 , P_2 , P_3 y P_4 . Hay que mencionar además, en el estándar ISO-9283 se establece que el número de muestras mínimo es de $m = 30$. Sin embargo, debido al proceso de calibración necesarios

para el sistema de cámaras y el tiempo que conlleva su ejecución, se realizó el registro de datos correspondientes a dos experimentos, resultando en $m = 2$. En la tabla 2 se muestran los resultados de exactitud en posición y orientación.

Evaluando los resultados anteriores que se obtuvieron del sistema de cámaras y con base en las ecuaciones 17 y 18 se determinó que el sistema de posicionamiento del robot móvil presenta una exactitud de $AP_p = 0.0779m$ y $AP_a = 0.1137rad$ en posición y orientación respectivamente.

Tabla 2 Resultados de cámaras en estado estacionario.

Numero de Corrida	Coordenada GDL (m)	P_1	P_2	P_3	P_4
1	x_1	0.4176	-0.5131	-0.5112	0.4436
	x_2	0.4729	0.4446	-0.5300	-0.5485
	x_3	0.0094	1.4810	3.0381	4.5797
2	x_1	0.4176	-0.5229	-0.5132	0.4436
	x_2	0.4837	0.4319	-0.5418	-0.5590
	x_3	0.0193	1.4740	3.06921	4.6176

En la figura 7 se muestra en superposición el desplazamiento realizado por el robot móvil en los dos experimentos realizados, las posiciones deseadas, el radio de exactitud $AP_p = 0.0779m$, y el radio del margen de error $M_e = 0.07m$ de las cámaras. Haciendo referencia a la figura 5a se observa que los resultados registrados por las cámaras y los resultados obtenidos por el sistema de localización por odometría, se comportan de forma similar y además las posiciones alcanzadas se encuentran dentro del radio M_e , presentando por lo tanto un buen desempeño.

Para mostrar el funcionamiento físico del robot, se ha realizado la grabación del experimento realizado en el apartado 3, extrayendo las imágenes (figura 8) en las que el robot se encontraba en estado estacionario en cada coordenada deseada y de inicio. En es posible observar que el robot móvil exhibe un mayor grado de error en las coordenadas alcanzadas 1 y 3, teniendo un error máximo aproximado de 0.04 m y 0.02 m en X_w y Y_w respectivamente, donde este parámetro fue determinado al superponer un mallado sobre las imágenes obtenidas.

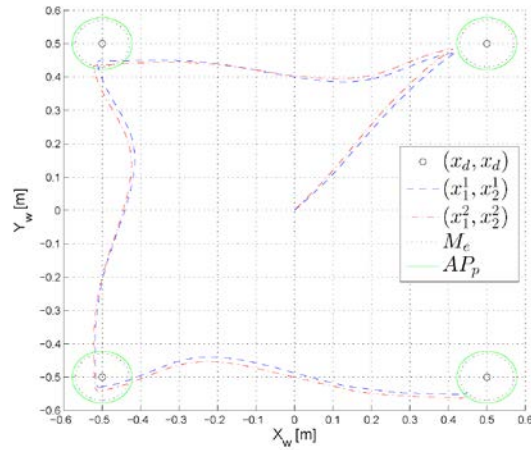


Figura 7 Posición deseada y comportamiento del robot móvil en sistema coordenado.

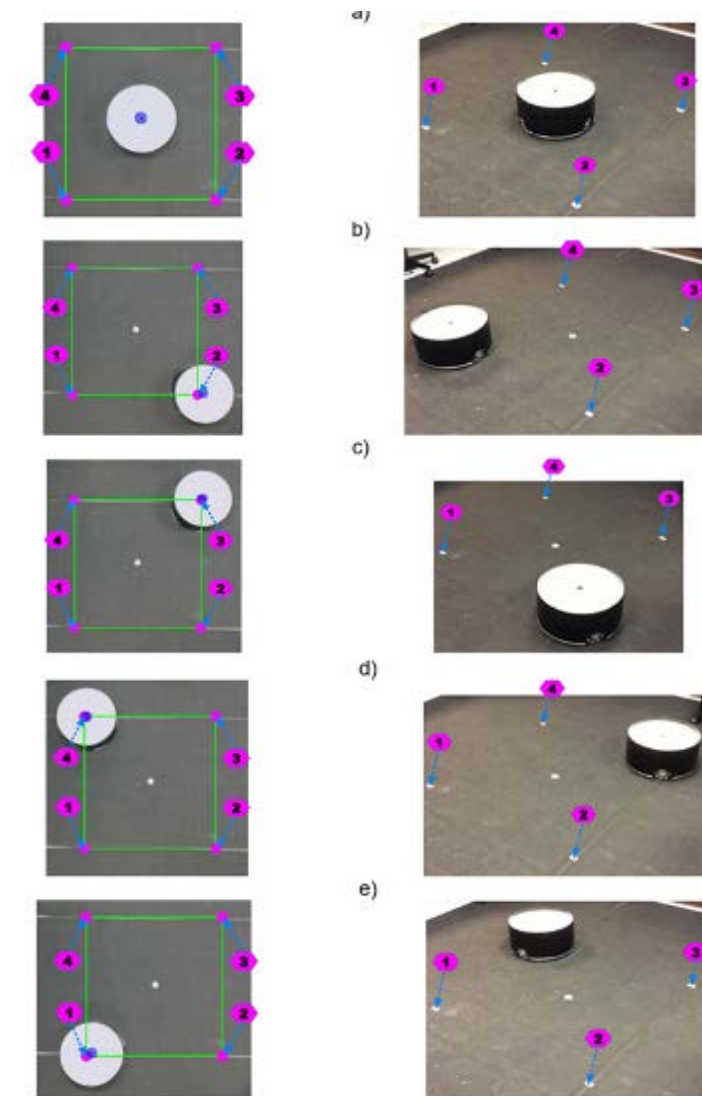


Figura 8 Capturas del posicionamiento del robot móvil 3.0.

5. Conclusiones

En el presente trabajo se presentó un robot móvil 3.0 y la evaluación de su sistema de control y de posicionamiento. Esta evaluación fue realizada al determinar el grado de repetibilidad y exactitud en posicionamiento y en orientación con base en el estándar ISO-9283.

Considerando el sistema de odometría como sistema de medición, la repetibilidad que presenta el robot móvil después de realizar 30 veces el experimento fue de $RP_t = \pm 0.0011$ m y $RP_o = \pm 0.0136$ rad en posición y en orientación respectivamente. Estos resultados verifican que el sistema de control presenta un buen grado de fiabilidad para realizar una tarea en específico bajo las mismas condiciones de operación.

Al comparar la posición absoluta del robot móvil obtenida a través de un sistema de faros activos basado en cámaras con la posición absoluta del sistema de odometría presente en el trabajo, se obtuvo una exactitud de $AP_p = 0.0779$ m y $AP_o = 0.1137$ rad en posición y orientación respectivamente, a pesar de presentar un margen de error $M_e = 0.07$ m en el proceso de calibración del sistema de cámaras.

Considerando la intersección del radio M_e con el radio de la exactitud AP_p , se puede observar que existe una diferencia de ± 0.0079 m, lo que indica que el sistema de posicionamiento absoluto con base en odometría presenta un buen grado de fiabilidad en la localización del robot móvil.

6. Bibliografía y Referencias

- [1] Cox, I., Blanche-an experiment in guidance and navigation of an autonomous robot vehicle. s.l.:IEEE Transactions on Robotics and Automation, 1991.
- [2] Guerrero-Castellanos, J. F., Villarreal-Cervantes, M. G., Sánchez Santana, J. P. & Ramírez-Martínez, Seguimiento de Trayectorias de un Robot Móvil 3.0 mediante control acotado. Revista Iberoamericana de Automática e Informática Industrial, pp. 426-434, 2014.

- [3] Dierks, T. & Jagannathan, S., Control of Nonholonomic Mobile Robot Formations: Backstepping Kinematics into Dynamics. Singapore, Singapore: IEEE International Conference on Control Applications, 2007.
- [4] Fu, G. y otros, Precise Localization of Mobile Robots via Odometry and Wireless Sensor Network. s.l.:International Journal Advance Robotics Systems, 2013.
- [5] Huang, J. y otros, Robot Position Identification by Actively Sound Beacons. Instruments and Measurement Technology Conference, pp. 1908-1912, 2006.
- [6] Siegwart, R. & Nourbakhsh, I. R., Introduction to Autonomous Mobile Robots. s.l.:Massachusetts Institute of Technology, 2004.
- [7] Sin, S., Kwon, D. & Myung, J., New Tag Arrangement Pattern for a Differential Driving Mobile Robot Based on RFID System. International Conference on Control, Automation and Systems, pp. 1228 -1233, 2007.
- [8] Standarization, I. O. o., Manipulating Industrial Robots-Performance Criteria and Related Test Methods. s.l.:ISO-9283, 1991.
- [9] Villarreal-Cervantes, M. G., Notas de Lectura: Modelo cinemático y dinámico del robot móvil 3.0. s.l.:s.n, 2015.
- [10] Villarreal-Cervantes, M. G. & Pantoja-García, J. S., Análisis comparativo entre un control heurístico y un PID para un sistema mecatrónico. s.l.:3th International Supercomputing Conference, 2012.
- [11] Xu, H. & Collins, J. J., Estimating the Odometry Error of a Mobile Robot by Neural Networks. s.l.:International Conference on Machine Learning and Applications, 2009.

BLOOD PRESSURE MEASUREMENT SYSTEM BASED ON OSCILLOMETRIC METHOD

Jessica Bolaños Olvera

Universidad Autónoma de Querétaro

jbo0206@gmail.com

Roque A. Osornio Rios

Universidad Autónoma de Querétaro

raosornio@hspdigital.org

Rosalía Reynoso Camacho

Universidad Autónoma de Querétaro

rosalia.reynoso@uaq.mx

Resumen

Este trabajo presenta una metodología para la medición no invasiva de la presión arterial, el algoritmo para la obtención de la presión arterial (presión sistólica, media y diastólica) es basado en el método oscilométrico. En el método oscilométrico las presiones son determinadas aplicando un criterio matemático al índice de pulso oscilométrico. En este trabajo es usado el criterio de pendientes para calcular la presión arterial sistólica y presión arterial diastólica. El sistema desarrollado es capaz de obtener los parámetros de presión arterial tanto para humanos como para ratas tipo Wistar, donde las mediciones de la presión arterial en las ratas Wistar se realizan como parte del desarrollo de otras investigaciones científicas en el campo de Química.

Palabras claves: Método oscilométrico, no invasivo, presión arterial, presión diastólica, presión sistólica

Abstract

This work presents a methodology for the non-invasive measurement of blood pressure, the algorithm for the obtention of blood pressure (systolic, mean and

diastolic pressures) is based on the oscillometric method. In oscillometric method the pressures are determined by applying a mathematical criterion to the oscillometric index pulse. In this work the slope criteria to calculate systolic and diastolic pressures is used. The developed system is capable of obtaining the blood pressure parameters for humans and rats, where blood pressure measurements in Wistar rats are realized as part of the development of other scientific research in Chemistry field.

Keywords: *Blood pressure, diastolic pressure, non-invasive, oscillometric method, systolic pressure.*

1. Introduction

The determination of Blood Pressure (BP) is a very important element in medicine and biological sciences due to that, information about the heart condition is provided [Ball et al, 2003]. In this sense, animal based researches have a significant impact in the development of new treatments for human diseases [Cong et al, 2009]. Mice share about 98% of the deoxyribonucleic acid (DNA) with humans [16]; these animals have the tendency to be affected for many health problems that affect the humans. According to [Broten et al, 1997], among all biological signals, blood pressure is one of the most important signal because of this is correlated with a long number of ailments and disorders, so that a suitable system to obtain BP signal is highly supportive for the analysis on the effects of new drugs for diseases control, clinical analysis as well as monitoring the behavior of mice. There are many devices and techniques dedicated to BP detection; the most common for rats is based on the use of an invasive catheter-tip inserted into an artery, as well as a tail-cuff device in rats or a cuff attached at the limbs in humans [Cong et al, 2009], [Monassier et al, 2006], [Malakoff, 2015]. On the other hand, in humans, there are several indirect methods for BP monitoring, such as, pulse method, auscultatory method, ultrasonic/Doppler method, oscillometric method, pulse transit time (PTT), among others [Cuesta, 2004], [Fernández, 1985]. One of the most common ways to obtain the BP parameters is through the auscultatory method [González, 2008], in this method, the BP is taken by an

expert. In [Escobar, 2012], a system for non-invasive measure continuous BP without cuff is mentioned, in this paper the PTT is the base for the development of the system. [Ball et al,2003] presents an algorithm based on oscillometric method for the estimation of BP parameters, in this work height criteria are mentioned for the estimation of the systolic and diastolic pressure, as well as the use of a commercial equipment to perform the validation of the results, the DOCTUS IV [Valdés, 2011]. The aforementioned works present a non-invasive way to obtain the BP parameters; however, this does not exclude the invasive methods as in [Ramírez, 2001], where an algorithm is used to acquire the BP parameters by means of a catheter inserted in a blood vessel of a human being. In animals, the most commonly indirect method for BP monitoring is the cuff technique, in which BP is measured using a variety of methods for sensing the changes in blood flow over the tail or limb, such as photoelectric sensors, oscillometric sensors, Doppler sensors, chamber volume sensors and acoustic sensors [Pickering et al, 2005]. The BP measurements by means of direct method use a radio telemetry technique or via indwelling catheters externally connected over mounted transducers [Monassier et al, 2006], [Pickering et al, 2005]. Different works have been carried out by BP monitoring through the direct method, such as the work presented in [Kramer et al, 1999], that describes one of the first possibilities for recording systolic, diastolic and mean BP, as well as the heart rate, and locomotor activity in freely moving mice, using a commercial telemetry and a data acquisition system, the system presented in the aforementioned work is achieved by an invasive method, in which, the paper describes the surgical technique for implanting a small radio-telemetry transmitter; likewise, the paper presents the differences among the indirect tail-cuff plethysmography method, the direct measurements by fluid-filled arterial catheters and the radio-telemetry used in the developed system. Related with the invasive method, in [Mills et al, 2000] a radio telemetry device is described; the implantable device provides measurements of systolic, diastolic and mean BP, as well as the heart rate and locomotor activity. On the other hand, in [Cong et al, 2009] the authors present the development of a real time wireless implantable blood pressure sensing microsystem for laboratory mice. The

aforementioned BP measurement modes are based on invasive methods, in which a surgical procedure and therefore the use of anesthesia is necessary in order to place the implant, which often causes distortions over the BP measurements. Different works about indirect BP measurements are mentioned below. A BP monitoring system using a photoconductive cell, to illuminate mice's tail by a small electric lamp is presented in [Van Nimwegen et al, 1973], in this system, the measurements are based on the sphygmomanometer method, in which, the pulsations in the arteries of the tail are converted into an electrical signal, after that, BP signal is displayed on an oscilloscope; the use of anesthesia to obtain the measurements is mentioned in this paper. In [Feng et al, 2009], BP measuring in mice is realized with a non-invasive BP monitoring system named CODA; the authors of this work provide an experimental protocol for accurately measure the tail-cuff blood pressure with this commercial system. In the same way, in [Infante et al, 1997] is reported a system that uses tail-cuff BP and an electrocardiogram (ECG), where plethysmographic pulse and BP signal are used in order to calculate systolic and diastolic BP parameters. The aforementioned system is based on sphygmomanometric method to obtain systolic and diastolic BP. According to [Pickering et al, 2005] it should be emphasized that regardless of the method used for measuring BP, systemic anesthesia should be avoided whenever feasible/possible because of the well-documented effects of anesthetics on cardiovascular function. In conclusion the invasive method is more accurate in relation to the tail-cuff, due to the BP measurements that can be continuously achieved and they are obtained directly from the artery, however, the cost of the implementation of this method, due to the implants should be analyzed according to the type of research performed; in addition, some surgical skills and training should be considered especially for small species such as mice. On the other hand, tail-cuff methods have served a valuable role in experimental hypertension research [Pickering et al, 2005], these methods are under development because some of these do not provide diastolic BP. Another aspect to consider in the indirect method is that the rat should be trained in order to avoid stress when the cuff is placed; besides, indirect methods have the advantage of they do not need a

surgical procedure and they are less expensive in relation direct methods. Different commercial equipment for BP measurement are presented in the literature, however the cost of these is high and they are completely closed in hardware and software architecture. When an investigation is carried out, the analysis is required and probably the implementation of new elements for the measurement of new variables, therefore, control over the hardware and software of the equipment allows habilitation, development and continuous improvements in this process, in order to obtain a suitable equipment for the needs of the investigation that is carried out. In this sense, and due to the need to obtain the blood pressure parameters in rats for the development of several investigations that are executed in the Universidad Autónoma de Querétaro, this work presents the development of a non-invasive blood pressure system for rats in which the methodology is based on the oscillometric method.

Oscillometric method

The most employed BP measure method in automatic devices is the Oscillometric method [Gamboa et al, 2007]. This method consists on inflating a cuff 20 to 30 mmHg above systolic pressure to ensure the occlusion of the artery, further to this, the cuff is deflated at rate of 3 mmHg per second. The aim of this method is to find the oscillations at the time the cuff is deflated. In the Oscillometric technique [Valdés,2011], the cuff is slowly deflating, due to this, the walls on the artery begin to vibrate as the blood flows through the artery partially occluded, these vibrations are recorded by the electronic transducer which oversee blood flow. On the other hand, while the cuff is slowly deflated, the oscillations are increased to a maximum amplitude and then decrease until they disappear in the point the cuff is completely deflated and the blood flow returns to normal. In the Oscillometric method the systolic pressure and diastolic pressure are determined by applying a mathematical criterion to the envelope curve that is produce by the oscillometric index pulse of the oscillations that were recorded by the pressure transducer. Figure 1 shows the oscillometric signal, and the parameters involved. In the oscillometric method, the only parameter that is measured is the mean BP,

and the systolic and diastolic BP are estimated. Likewise, the way to obtain systolic and diastolic BP parameters according to [Gamboa et al, 2007] are based on height or the slope analysis. Figure 2 shows the way to relate the oscillometric index pulse (OIP) and BP signal in order to find systolic and diastolic BP by the height and slope analysis. In the height-based method, the desired pressure values are determined as the pressure of the cuff at which the ratio of the oscillometric index pulse at the peak relative to the maximum index pulse is equal to certain predetermined values, while, the slope-based criterion is determined by looking at the maximum and minimum slope points in the envelope curve. Figure 3 shows the relation between the BP cuff signal and the oscillometric index signal.

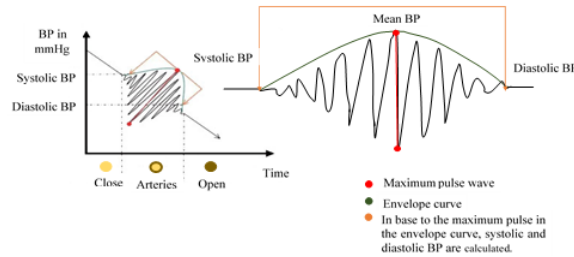


Figure 1 Oscillometric signal.

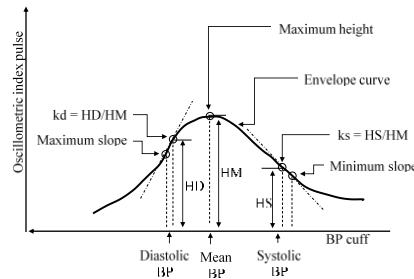


Figure 2 Oscillometric index pulse.

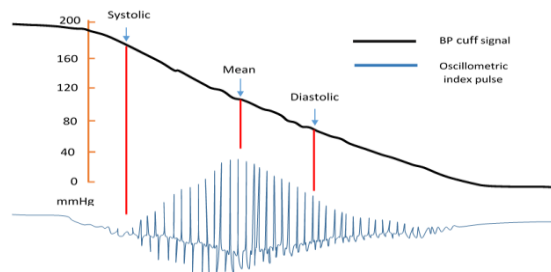


Figure 3 BP cuff signal when it is deflated vs Oscillometric index pulse.

2. Methods

In this section proposed methodology for the obtaining of BP is presented. First, arterial vibration is obtained by means of the occlusion cuff. These vibrations were recorder by a pressure transducer coupled to the occlusion cuff. Next, obtained vibration signal is submitted to a pre-processing stage in which a low pass filter to 10 Hz was used in order to remove high frequency noise induced by the inflation process of the cuff. Filter signal was digitized by a 12 bit-ADC embedded in a Beaglebone black board at a sampling frequency of 1 kHz. Once obtained the digital signal, it was filtered by a pass band digital filter in order to obtain the oscillometric index pulse signal. For this case, digital filter type IIR Butterworth was used. In this point of methodology, the systolic, mean and diastolic pressures was obtained by applying the slope criterion on the envelope of the OIP signal calculated by a peak detection algorithm.

The previously paragraph describe the methodology and the next section show in more detail all the sections before mentioned; first figure 4 shows the methodological diagram of the system for the measurement of blood pressure, which is composed of 4 main modules, "Module of acquisition of the blood pressure signal", this module was composed of a pump (5 V supply voltage), solenoid valve (5 V supply voltage), an occlusion cuff (for the case of the rat, this was placed in the tail because in this is the distal continuation of the aorta is presented [Olds, 1979], whereas in the humans a occlude band of 22-32 cm was placed in the arm), a MP3V5050DP pressure transducer with typical 3 V supply voltage as well as a maximum pressure range of 50 kPa (375.03095 mmHg) according to the maximum blood pressure range required for both rats an humans this element was suitable for the application. For the control of the pump, the solenoid valve and the digitization of the BP signal was used the low power computer board Beaglebone Black. Regarding the frequency of sampling, according to the American Heart Association (AHA) [AHA Committee,1975] committee, the Task force of the Europe Society of Cardiology and the North American Society of Pacing and Electrophysiology recommend that Electrocardiogram (ECG) sampling frequency range of 250-500 Hz is optimum for

accurate characterization of heart rate variability [Ahmad et al,2010]. Follows the Shannon sampling theorem and in relation to the typical frequency spectrum range from heart rate of the signals in humans and in Wistar rats the sample of 1 kHz was sufficient to provide correct estimation for the OIP. The next module corresponds to the “Analog filter”, how was describe previously it was a 10 Hz low pass analog filter; the typical frequency spectrum of human heart rate is 1-1.666 Hz whereas the typical frequency spectrum of Wistar rats is 6.33-7.4833 Hz so the analog filter does not eliminate information from the oscillations caused due to the walls of the artery.

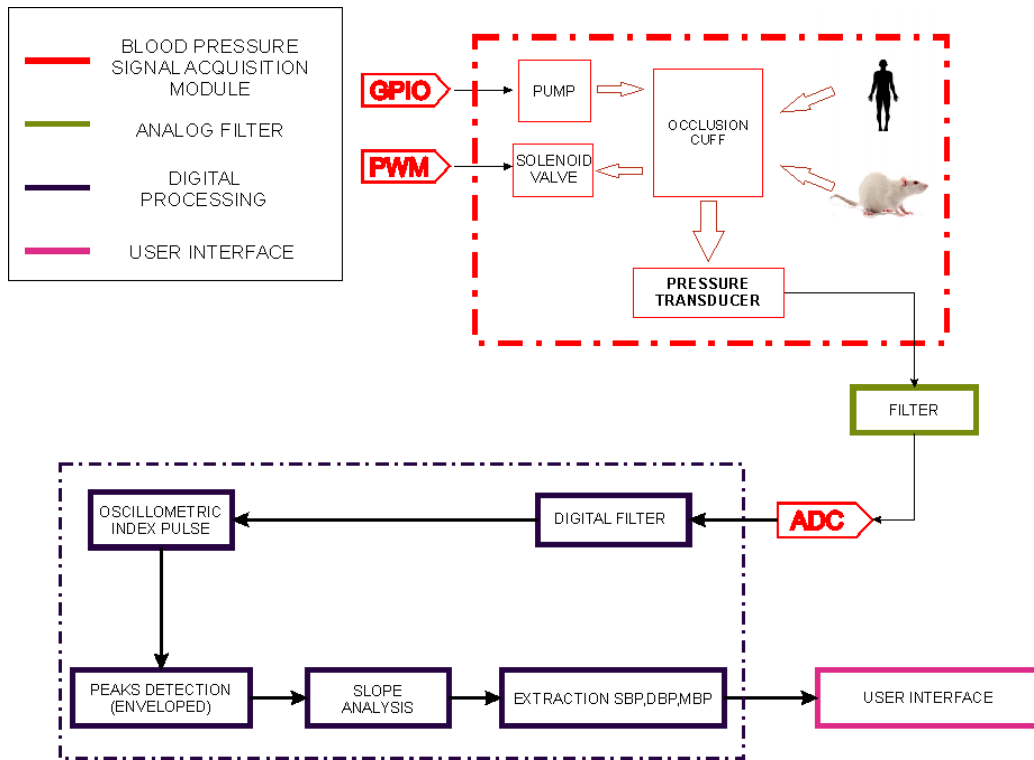


Figure 4 Methodological diagram for the BP measurement.

The “Digital processing” this module corresponds to the software analysis of the BP signal, this module was divided in two fundamental parts: the acquisition of the BP signal and an algorithm of analysis of the pressure signal of the deflation cuff. The acquisition of the BP was performed in a real time by the interaction of the Beaglebone Black board and a user interface in Matlab® platform, in the user

interface the data analysis option was available, which corresponds to the analysis of the deflation BP cuff signal, the processing that was applied to the signal was by software in the Matlab platform, the first paragraph of this section that describe the methodology describe the elements of this module. Finally, the last modulo “User interface” includes the BP cuff signal and shows the BP parameters. The methodology aforementioned was used for calculate the BP parameters in 2 different cases of study: humans and rats, the digital filter is the only module that was different due to the heart rate is different in both species, so the results are presented for both cases. The experimental set up for the human case was performed in base to the Mexican Norm [PROY-NOM-030-SSA2-217] the norm describe the basic procedure for take the blood pressure, while for the second case the measurements were performed in a Wistar rat weighing 442 grams, the procedure to maniple the rat was in base to the Official Mexican Norm [NOM-062-ZOO-1999], for the test, three occlusion cuff of different internal diameter were use and five test were performed by each occlusion.

3. Results

In this section, the obtained results for both cases of study using the proposed methodology are presented, first section shows the results for tests realized on humans; figure 5 shows the BP cuff signal, where the inflation signal is depicted in red and the deflation signal of the occlusion cuff is depicted in blue. As mentioned in the methodology the signal of interest is the signal of deflation occlusion so the analysis is performed on this signal; following the steps described in the methodology figure 6 shows the OIP depicted in green, envelope curve depicted with circles in red and the derivate for the envelop curve in black. Finally figure 7 shown the values for BP parameters in both signals, deflation cuff signal and OIP, the SBP is depicted in a red circle, the MBP with a black square and finally the DBP with a blue triangle; the red diamond, black cross and blue asterisk represent the moment in which the SBP, MBP and DBP appear in the slope analysis that was made over the envelope curve. For each tests performed the same procedure was followed.

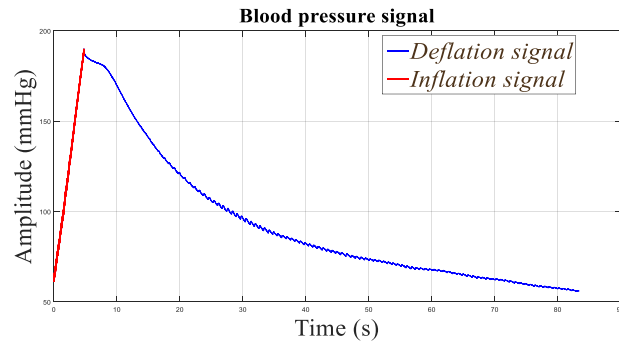


Figure 5 BP cuff signal.

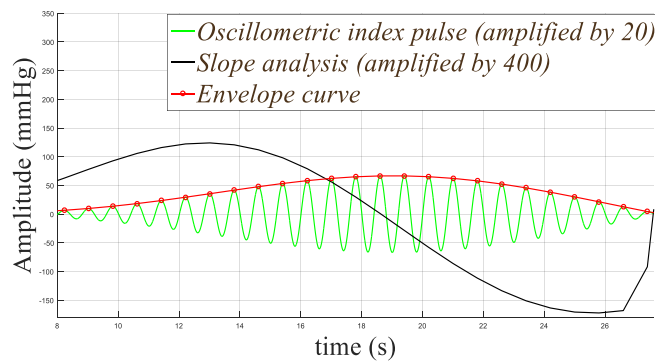


Figure 6 Analysis for obtain BP parameters.

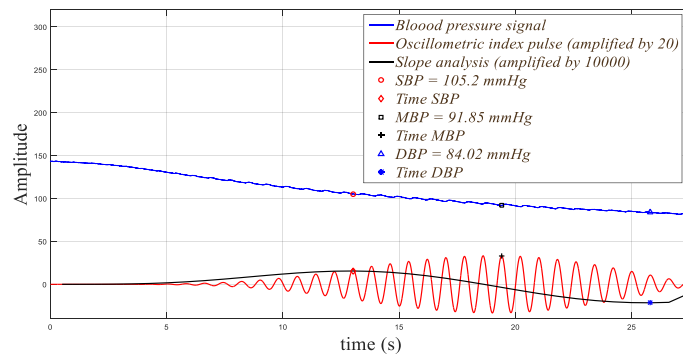


Figure 7 Establishment to the BP parameters in deflation cuff signal and OIP.

In the next section the total results are presented. Figure 8 shown the BP signal for ten samples that are presented and over each signal the BP parameters; SBP is depicted in a red circle, MBP is depicted in a yellow square, and finally DBP is depicted in a blue diamond. Table 1 shows the BP parameters for each test.

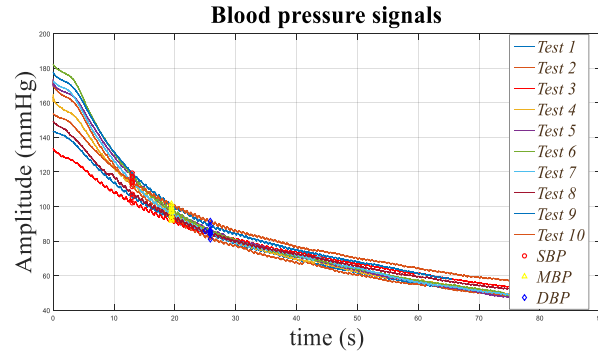


Figure 8 BP parameters obtained for the human case.

Table 1 Results for each test in humans.

Test	SBP (mmHg)	MBP (mmHg)	DBP (mmHg)
Test 1	105.2	91.85	84.02
Test 2	111.4	92.7167	85.8
Test 3	102.1	92.3212	85.8
Test 4	113.8	96.8553	86.29
Test 5	116.6	97.5336	85.64
Test 6	117.6	98.4497	84.84
Test 7	113.9	95.0563	84.92
Test 8	107.2	94.4983	84.44
Test 9	119.5	100.021	88.94
Test 10	114.8	101.5107	91.57

In the next section Wistar rat results are presented, figure 9 shows the BP cuff signal, the inflation signal depicted in red and the deflation signal depicted in blue, while the figure 10 shows the oscillometric index pulse depicted in magenta, the derivate to the envelope curve is represent in black and envelope curve is depicted in red circles. Figure 11 shows the BP parameters, SBP is depicted in a red circle, MBP is depicted with a black square and DBP is depicted with a blue triangle, the value for each parameter is shown in the figure, as well as the moment in which each parameter appears in the slope analysis that was made over the envelope curve. Figures 12a, 12b and 12c, shown the results for the BP parameters with the occlusion cuff A, B and C respectively; the figures show the BP signal for each sample and over each signal the BP parameters, SBP is depicted in a red circle, MBP is depicted in a black square, and finally DBP is depicted in a blue triangle. As presented in the methodology for each occlusion cuff five test were performed, the BP parameters are shown in table 2.

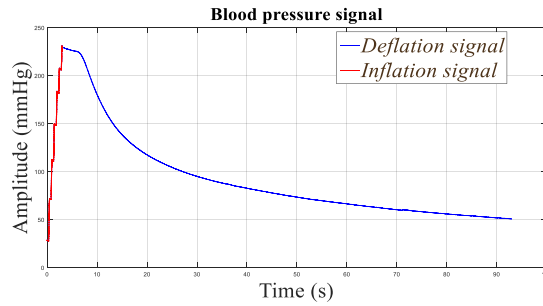


Figure 9 BP cuff signal.

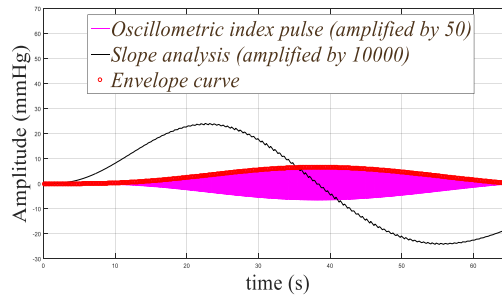


Figure 10 Slope analysis over the envelope curve for obtain the BP parameters.

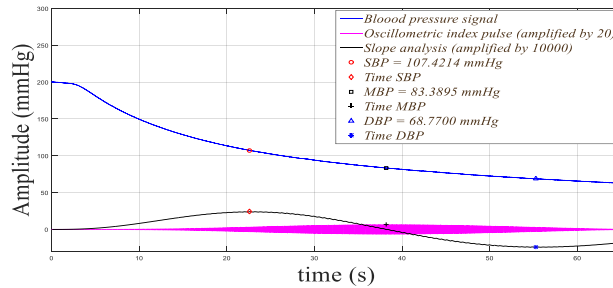


Figure 11 Establishment to the BP parameters in deflation cuff signal and OIP.

4. Discussion

In order to validate the methodology, both cases were compare, for the case in humans, the results were compared with the commercial equipment OMRON, for each sample taken with the development equipment another was taken with the commercial equipment. As shown in table 1 and table 3, the results obtained with the proposed methodology are congruent with the commercial equipment. By means of the arithmetic mean and standard deviation, the ranges of BP parameters for OMRON equipment and the proposed methodology are shown in table 4.

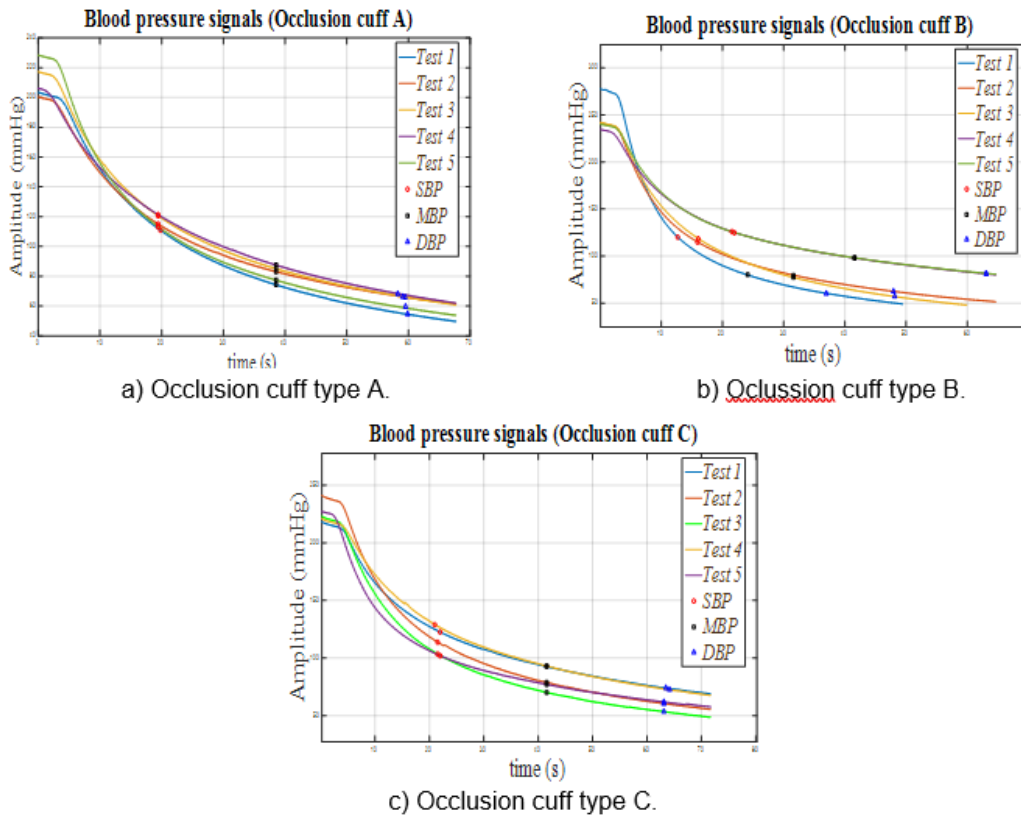


Figure 12 BP parameters for each occlusion cuff.

Table 1 Results for each test in Wistar rat.

Occlusion Cuff A (15 mm internal diameter)			
Test	SBP (mmHg)	MBP (mmHg)	DBP (mmHg)
Test 1	101.8263	84.6687	77.0766
Test 2	107.421	83.389	68.77
Test 3	112.96	82.474	67.231
Test 4	108.552	83.449	70.898
Test 5	103.027	81.719	71.298
Occlusion Cuff B (12 mm internal diameter)			
Test	SBP (mmHg)	MBP (mmHg)	DBP (mmHg)
Test 1	94.5619	77.765	69.3665
Test 2	94.941	78.33	69.603
Test 3	98.4	78.97	69.734
Test 4	107.514	84.094	71.864
Test 5	118.223	83.445	66.056
Occlusion Cuff C (15 mm internal diameter)			
Test	SBP (mmHg)	MBP (mmHg)	DBP (mmHg)
Test 1	106.8976	82.486	70.2802
Test 2	108.53	83.135	68.052
Test 3	99.322	79.102	69.453
Test 4	97.317	85.207	79.152
Test 5	99.862	79.152	76.8

Table 2 Results from OMRON equipment.

Test	SBP (mmHg)	MBP (mmHg)	DBP (mmHg)
Test 1	106	92	85
Test 2	109	93.666	86
Test 3	105	92.333	86
Test 4	114	95.333	86
Test 5	116	98.666	90
Test 6	117	95	84
Test 7	114	94	84
Test 8	106	90	82
Test 9	118	97.333	87
Test 10	114	98.666.91	91

Table 3 Ranges for BP parameters.

OMRON equipment		
SBP(mmHg)	MBP (mmHg)	DBP (mmHg)
111.9±4.931	94.232 ± 3.928	86.1 ± 2.7264
Proposed methodology		
SBP(mmHg)	MBP (mmHg)	DBP (mmHg)
112.210 ± 5.683	95.081 ± 4.232	86.226 ± 2.317

As shown in the table 4, the deviation between the proposed methodology and the OMRON equipment is very small, therefore it is established that the results are congruent. For the Wistar rat, in [Wang et al,2013] the author presents BP measurements which were obtained from the caudal ventral artery and femoral artery in Wistar rats the results are presented in table 5, both methods aforementioned are invasive procedures. Comparing the results presented in table 2 with the results obtained from [Wang et al,2013] table 5, it is perceived that the pressure parameters obtained from the proposed methodology are within the ranges of the invasive method caudal except for test 3 in occlusion cuff A and test 5 occlusion cuff B, which present an error of 4.16 and 9.425 mmHg respectively, on the maximum estimate of the systolic blood pressure, whereas for the diastolic pressure the error was 0.769 and 1.9744 mmHg respectively and were below the estimate established by the invasive method. To delimit the BP ranges for the proposed methodology, the arithmetic mean and the standard deviation were obtained, the results are presented in table 6. Although the results obtained by the system are not completely accurate to those presented by the direct method, these

are within the range; the variability of the presented results can be due to several factors, among them that the measurement of the BP by indirect methods allows the animal to remain awake and place the animal in the holder and position the occlusion cuff in the tail, can cause stress in the rat (although the rat was previously trained), whereas with the direct method the rat is kept asleep and therefore the stress is avoided; another factor that may influence variations in BP is the rat itself, since the comparative values were extracted from results reported [Wang et al,2013]; however, to observe the dispersion of the system presented in this work as compared to the caudal method (direct method) the BP must be measured with both methods; but the material and human resources were not qualified to carry out this invasive procedure and only a comparative was realized with results reported in the literature.

Table 4 BP Parameters for invasive methods.

Caudal ventral artery		
SBP (mmHg)	MBP	DBP (mmHg)
108.8 – 85.6	100.9 – 78.9	92.2 – 68
Femoral artery		
123.1 – 105.3	107.2 – 85.8	96.1 – 74.3

Table 6 Ranges for the BP parameters.

SBP (mmHg)	MBP (mmHg)	DBP (mmHg)
103.957 ± 6.803	81.85 ± 2.483	71.042 ± 3.772

5. Conclusions

In conclusion, the oscillometric method is a suitable methodology for the obtention of BP parameters in both Wistar rats and humans. In the work developed, the results obtained for both case studies were compared, in the case of humans with a commercial system having favorable results while in the case of Wistar rats the results obtained were compared by invasive methods were compared, although in this case obtained some variations the next step of the research is to compare the system developed with a commercial system to be able to have the same test subject (Wistar rat) and to take the BP with both systems and in this sense to know the variability of the proposed system.

6. Bibliography and References

- [1] Aha Committee, Recommendations for standardizations of leads and specifications for instruments in ECG and VCG. *Circulation*, 52, 11-251975, 1975.
- [2] Ahmad, S., Bolic, M., Dajani, H., Groza, V., Batkin, I., & Rajan, S., Measurement of heart rate variability using an oscillometric blood pressure monitor. *IEEE transactions on instrumentation and measurement*, 59(10), 2575-2590, 2010.
- [3] Ball-Llovera, A., Del Rey, R., Ruso, R., Ramos, J., Batista, O., & Niubo, I., An experience in implementing the oscillometric algorithm for the noninvasive determination of human blood pressure. In *Engineering in Medicine and Biology Society*, 2003. Proceedings of the 25th Annual International Conference of the IEEE (Vol. 4, pp. 3173-3175). IEEE. September, 2003.
- [4] Broten, T. P., Kivlighn, S. D., Harvey, C. M., Scott, A. L., Schorn, T. W., & Siegl, P. K. S., Techniques for the measurement of arterial blood pressure. *Measurement of cardiovascular function*, 1997.
- [5] C. A. Ramírez, Algoritmo para el cálculo de la presión sistólica y diastólica en el ventrículo izquierdo. Presented in *Memorias II Congreso Latinoamericano de ingeniería biomédica*, Habana, Cuba, May. 2001.
- [6] Cong, P., Chaimanonart, N., Ko, W. H., & Young, D. J., A wireless and batteryless 10-bit implantable blood pressure sensing microsystem with adaptive RF powering for real-time laboratory mice monitoring. *IEEE Journal of Solid-State Circuits*, 44(12), pp. 3631-3644, 2009.
- [7] Cuesta, *Medición de la Tensión Arterial*, Dept. Enfermería, Valencia Univ., España, Sep. 2004.
- [8] Escobar-Restrepo, B. Y., Sistema para la medición de la presión arterial continua no invasiva sin brazalete, Doctoral dissertation, *Biomédica, Mecatrónica y Mecánica*, 2014.
- [9] Feng, M., & DiPetrillo, K., Non-invasive blood pressure measurement in mice. *Cardiovascular Genomics: Methods and Protocols*, pp. 45-55, 2009.

- [10] Fernández-Abascal, E. G., El tiempo de tránsito del pulso: un índice de cambios en la presión arterial. *Estudios de Psicología*, 6(21), pp. 21-33, 1985.
- [11] Gamboa, W., Rodríguez, L., Cháves, A., de Colombia, F. C., & de Bioingeniería, G., Dispositivo Digital para el registro continuo de presión arterial de forma no invasiva y ambulatoria. In VII Congreso de la Sociedad Cubana de Bioingeniería, 2007.
- [12] González, Significación de los ruidos de la presión sanguínea. Sociedad Mexicana para el estudio de la hipertensión arterial, 2008.
- [13] Hernández, El modelo en las investigaciones biomédicas. *Biomedicina*, vol. 2, no. 3, pp. pp. 252-256, 2006.
- [14] Infante-Vázquez, O., Sánchez-Torres, G., Martínez-Memije, R., Flores-Chávez, P., Pastelin-Hernández, G., & Sánchez-Miranda, M., Medición de la presión arterial utilizando el retardo en el pulso distal, *Rev Bras Eng Bioméd*, 13, pp. 81-92, 1997.
- [15] Kramer, K., Voss, H. P., Grimbergen, J. A., Mills, P. A., Huetteman, D., Zwiers, L., & Brockway, B., Telemetric monitoring of blood pressure in freely moving mice: a preliminary study, *Laboratory Animals*, 34(3), pp. 272-280, 2000.
- [16] Leong, X. F., Ng, C. Y., & Jaarin, K., Animal models in cardiovascular research: hypertension and atherosclerosis. *BioMed research international*, 2015.
- [17] Malkoff, J., Non-invasive blood pressure for mice and rats. *Animal Lab News*, Kent Scientific Corporation, pp. 1-12, 2005.
- [18] Mexicana, N. O. NOM-062-ZOO-1999, Especificaciones Técnicas para la producción, cuidado y uso de los animales de Laboratorio. México: Diario Oficial de la Federación, 1999.
- [19] Mills, P. A., Huetteman, D. A., Brockway, B. P., Zwiers, L. M., Gelsema, A. M., Schwartz, R. S., & Kramer, K., A new method for measurement of blood pressure, heart rate, and activity in the mouse by radiotelemetry. *Journal of Applied Physiology*, 88(5), pp. 1537-1544, 2000.

- [20] Monassier, L., Combe, R., & El Fertak, L., Mouse models of hypertension. *Drug Discovery Today: Disease Models*, 3(3), pp. 273-281, 2006.
- [21] Olds, R. J. A., colour atlas of the rat: dissection guide. Wolfe Medical Publications Ltd, 1979
- [22] Pickering, T. G., Hall, J. E., Appel, L. J., Falkner, B. E., Graves, J., Hill, M. N., & Roccella, E. J., Recommendations for blood pressure measurement in humans and experimental animals. *Circulation*, 111(5), pp. 697-716, 2005.
- [23] Prevención, S., de la Salud, P., de Integración, S., & del Sector Salud, D. proyecto de norma oficial mexicana proy-nom-030-ssa2-2017, para la prevención, detección, diagnóstico, tratamiento y control de la hipertensión arterial sistémica.
- [24] Valdés, M. J. G., & Kuchinskaia, D. V., Mejoras del método oscilométrico de medición de la presión no invasiva en el monitor de paciente DOCTUS VI. *Revista Ingeniería Electrónica, Automática y Comunicaciones* ISSN: 1815-5928, 31(1), pp. 55-59. 2011.
- [25] Van Nimwegen, C. H. R., Van Eijnsbergen, B., Boter, J., & Mullink, J. W. M. A., A simple device for indirect measurement of blood pressure in mice. *Laboratory animals*, 7(1), pp. 73-84, 1973.
- [26] Wang, Y., Cong, Y., Li, J., Li, X., Li, B., & Qi, S. Comparison of invasive blood pressure measurements from the caudal ventral artery and the femoral artery in male adult sd and wistar rats, *PloS one*, 8(4), e60625, 2013.
-

INGENIERÍA ONTOLÓGICA APLICADA EN EL DISEÑO DE UN SISTEMA DE ONTOLOGÍAS PARA LA GESTIÓN DE HORARIOS

Maricela Claudia Bravo Contreras

Universidad Autónoma Metropolitana
mcbc@correo.azc.uam.mx

Francisco Pavón Gutiérrez

Universidad Autónoma Metropolitana
teorema06@gmail.com

José Alejandro Reyes Ortiz

Universidad Autónoma Metropolitana
jaro@correo.azc.uam.mx

Roberto Alfonso Alcántara Ramírez

Universidad Autónoma Metropolitana
raar@correo.azc.uam.mx

Resumen

En este artículo se describe el proceso de ingeniería ontológica aplicado en el diseño y construcción de un sistema de ontologías para la gestión de horarios. Se presenta la metodología detallada que se implementó, la cual consistió de las siguientes etapas: especificación de las preguntas de competencia, diseño modular y basado en dominios del sistema de ontologías, diseño y evaluación de cada ontología individualmente, integración de las ontologías y evaluación general del sistema de ontologías mediante preguntas de competencia. Como resultado se obtuvo un sistema de ontologías que cumple con los principios de diseño de claridad, coherencia, modularidad, y usabilidad. Finalmente se evaluó la competencia de la ontología ejecutando las preguntas de competencia a través de la definición de reglas de inferencia lógica.

Palabras Claves: Ingeniería ontológica, principios de diseño de ontologías, sistema de ontologías.

Abstract

In this paper we describe the ontological engineering process applied in the design and construction of an ontology system for academic schedule management. We present the detailed methodology that was implemented, which consisted of the following stages: specification of competency questions, modular and domain-oriented design of the ontology system, design and evaluation of each ontology, integration of ontologies and general evaluation of the ontology system. As a result, a system of ontologies was obtained that complies with the design principles of clarity, coherence, modularity, etc. Finally, the competence of the ontology was evaluated by executing the competency questions through the definition of inference and query rules.

Keywords: *Ontology engineering, ontology design principles, ontology system.*

1. Introducción

La ingeniería ontológica es el proceso mediante el cual se diseñan y aplican métodos, técnicas y principios de diseño con el objetivo de producir una ontología o sistema de ontologías. Una pregunta frecuente durante el diseño de una ontología es decidir sobre la incorporación de otras ontologías en una sola. En este artículo nos referimos al concepto de **sistema de ontologías** como una ontología que incluye o incorpora otras ontologías, dentro de la cual se definen relaciones semánticas que existen tanto entre conceptos como entre individuos de las diferentes ontologías importadas. El objetivo de la ingeniería ontológica es apoyar todas las etapas del diseño y desarrollo de manera eficiente para lograr ontologías de calidad. Un factor clave para lograr la eficiencia durante el diseño de ontologías es la incorporación de principios de diseño.

Los principios de diseño son criterios de calidad que guían y orientan el diseño y construcción de las ontologías con el objetivo principal de lograr ontologías reutilizables, fáciles de mantener y actualizar a lo largo de su vida útil. Asimismo, a

partir de los principios de diseño es posible proponer mecanismos para evaluar la calidad del diseño de una ontología. En esta sección se presenta una revisión de los principios de diseño que han sido propuestos y discutidos por renombrados autores en la literatura especializada.

Thomas Gruber [Gruber, 1993] fue uno de los primeros autores que describió principios de diseño aplicables a las ontologías. Gruber propuso un conjunto de criterios de diseño cuyo propósito era compartir el conocimiento y facilitar la interoperabilidad entre programas mediante una conceptualización compartida. Los criterios de diseño de ontologías descritos por Gruber son los siguientes:

- Claridad. Especifica que una ontología debe comunicar efectivamente el significado intencional de los términos definidos. Las definiciones deben ser objetivas e independientes del contexto social o computacional. Siempre que sea posible, una definición debe estar completa con las condiciones necesarias y suficientes, en lugar de ser parcialmente definida. Todas las definiciones deben documentarse en lenguaje natural.
- Coherencia. Establece que una ontología debe generar inferencias que sean consistentes con las definiciones. Esto implica que todos los axiomas definidos en la ontología deben ser consistentes lógicamente. La coherencia también debe aplicarse a los conceptos descritos en lenguaje natural en la documentación. Si un enunciado que puede inferirse de los axiomas contradice una definición dada informalmente, entonces la ontología es incoherente.
- Extensibilidad. Establece que una ontología debe diseñarse anticipando los posibles usos del vocabulario. Debe ofrecer el fundamento conceptual para un rango de tareas anticipadas, y la representación debe de construirse de tal forma que se pueda extender y especializar la ontología monotónicamente. En otras palabras, debe permitir la definición de nuevos términos para usos especiales basados en el vocabulario existente.
- Tendencia de codificación mínima. La conceptualización debe ser especificada a nivel de conocimiento sin depender de una codificación particular a nivel de símbolos.

- Compromiso ontológico mínimo. Una ontología debe poseer el compromiso ontológico mínimo y suficiente para soportar las actividades requeridas de conocimiento compartido. Esto significa que una ontología debe hacer tan pocas aclamaciones como sea posible sobre el espacio del mundo que está modelando, permitiendo que las partes adheridas a la ontología tengan la libertad de especializar e instanciar la ontología como sea necesario.

En [Fox, 1998] los autores propusieron seis características para evaluar un Modelo Empresarial. Estas características fueron propuestas para responder a la pregunta de ¿Cómo se puede determinar cuál ontología es la correcta para determinada tarea? Para dar una pauta sobre la operatividad de estas características, los autores definen el concepto de competencia del modelo de la siguiente forma: dado un modelo apropiadamente instanciado y un demostrador de teoremas, la competencia de un modelo es el conjunto de preguntas que el modelo puede responder. Otro significado de competencia se refiere a la expresividad requerida de la ontología para representar las preguntas de competencia y para caracterizar sus soluciones. Con base a este concepto de competencia, explican las características que se listan a continuación:

- Completitud funcional. Esta característica es determinada por la competencia de la ontología. Esto es, el conjunto de preguntas que puede responder con un modelo instanciado apropiadamente.
- Generalidad. La generalidad de una ontología puede determinarse verificando si la unión de preguntas de un amplio conjunto de funciones, inclusive provenientes de diferentes sectores, es reducible al conjunto de preguntas de competencia de la ontología en cuestión.
- Eficiencia. La eficiencia de una ontología puede medirse calculando el número de inferencias lógicas por segundo requeridas para responder una pregunta. Sin embargo, los autores toman en cuenta que existe más de una forma para representar el mismo conocimiento y que cada representación no tiene la misma complejidad al responder a una clase específica de

preguntas. Por lo tanto proponen calcular la complejidad promedio de las preguntas de competencia para estimar la eficiencia.

- Claridad. La claridad de una ontología se logra mediante la axiomatización. Esto es, proporcionar definiciones formales de los objetos, sus relaciones y sus atributos.
- Precisión. Esta característica se refiere al grado en el que las definiciones de los conceptos se declaran distintos entre sí. Por ejemplo: si un concepto está incluido en otro, o qué conceptos se encuentran en la intersección o en la unión de dos o más conceptos.
- Granularidad. Esta característica se refiere a la capacidad de una ontología de representar conceptos en diferentes niveles de abstracción.
- Minimalista. Esta característica se determina utilizando los axiomas para determinar que por cada objeto o concepto en la ontología, no existe otro objeto que sea lógicamente equivalente.
- En [Morbach, 2009] los autores establecieron que una ontología debe cumplir con dos objetivos prioritarios: ser usable y reutilizable:
 - a) La reutilización la definen como el grado en el que un módulo de software u otro producto puede ser utilizado en más de un programa de cómputo o sistema de software. En particular, la reutilización de ontologías puede definirse como la capacidad de adaptación de una ontología a contextos de aplicación arbitrarios, incluyendo aquellos contextos que no fueron previstos al momento de la creación de la ontología.
 - b) La usabilidad, por otro lado denota el grado en el que un componente de software es útil para una tarea o aplicación específica. El término también tiene la connotación de facilidad de uso, refiriéndose al esfuerzo requerido por un usuario para utilizar un sistema de software dado. El objetivo de la usabilidad de una ontología es minimizar el esfuerzo requerido para adecuar la ontología de tal forma que pueda ser usada por humanos o máquinas en un contexto

de aplicación dado. En este sistema de ontologías se atendieron los principios de Modularidad, Coherencia, Claridad y Usabilidad.

En relación con las metodologías para el diseño y construcción de ontologías, [Gómez, 3] establece que una metodología se compone de métodos, técnicas, procesos y actividades. En esta sección se presenta una revisión de las metodologías de diseño de ontologías.

La metodología para construir la ontología CyC fue presentada por [Lenat, 1990]. Los autores proponen tres métodos para la construcción de una ontología: extracción manual y codificación del conocimiento, extracción semi-automática y codificación del conocimiento, y extracción completamente automática y representación del conocimiento.

[Uschold, 1995] presentaron una metodología para desarrollar y evaluar ontologías, la cual fue refinada y mejorada después por [Uschold, 1996]. Esta metodología incluye las siguientes etapas: identificar el propósito, construir la ontología – captura de la ontología, codificación de la ontología, integración de ontologías existentes -, evaluación y documentación.

En [Grüninger, 1995] los autores describen el proceso de construir una ontología para el proyecto TOVE (Toronto Virtual Enterprise). La metodología consiste de las siguientes etapas: describir el escenario de motivación, definir las preguntas de competencia, definir la terminología de la ontología (objetos, atributos y relaciones), y especificar las definiciones y restricciones, finalmente la evaluación de la ontología se realiza por el teorema de completitud.

En [Berner et al., 1996] los autores presentaron una metodología para construir ontologías para el proyecto KACTUS. El objetivo de la metodología era reutilizar las ontologías existentes y evaluar la factibilidad de reutilizar conocimiento en sistemas complejos. Esta metodología estableció los siguientes procesos generales: desarrollo de la lista de requerimientos (especificación de la aplicación), identificación de los términos generales para generar un modelo preliminar, estructuración y refinamiento de la ontología, y búsqueda de ontologías pre-existentes para reutilizarlas.

METHONTOLOGY fue presentada por primera vez por [Gómez, 1996]. Posteriormente refinada y detallada por [Fernández et al, 1997], [Fernández, 1981], y [Fernández, 1999]. El objetivo de esta metodología es facilitar la construcción de ontologías. METHONTOLOGY define un proceso de desarrollo y un ciclo de vida el cual consiste de las siguientes actividades: especificación, conceptualización, formalización, implementación, y mantenimiento. La conceptualización consiste en: construir el glosario de términos, construir la taxonomía de conceptos, construir el diagrama de relaciones binarias, y construir el diccionario de conceptos.

CommonKADS [Schreiber, 2000] es una metodología que reconoce la necesidad de reutilizar elementos del modelo de conocimiento o una combinación de éstos. La metodología CommonKADS incluye las siguientes etapas: identificación del conocimiento, especificación del conocimiento, y refinamiento del conocimiento.

La metodología NEON [Suárez, 2010] se basa en el uso de patrones de diseño. Esta metodología propone la reutilización de ontologías de un repositorio público de ontologías y un conjunto de patrones de diseño conocidos para integrarlos mediante procesos de re-ingeniería. Las etapas generales de esta metodología son: identificación de requerimientos, identificación de los patrones de diseño disponibles, dividir y transformar el problema en problemas parciales, encontrar las coincidencias de los problemas parciales con los patrones de diseño, seleccionar el patrón de diseño, aplicar el patrón de diseño para hacer una composición, evaluar las soluciones parciales, e integrar las soluciones parciales.

A pesar de que se han presentado varias propuestas metodológicas para la construcción de ontologías, éstas no consideran la incorporación de principios de diseño desde la especificación de la competencia de la ontología, hasta su evaluación final.

En este artículo se presenta una metodología integral para el diseño y construcción de un sistema de ontologías para la gestión de horarios. La metodología descrita se basa en los principios de diseño de coherencia, inteligibilidad y modularidad. La evaluación de la ontología se realiza con base a su competencia y atendiendo una lista inicial de preguntas de competencia.

2. Métodos

En esta sección, se describe detalladamente la metodología integral que se implementó para el diseño, construcción y evaluación del sistema de ontologías para la gestión de horarios. Uno de los aspectos más importantes de esta metodología es que se aplica para el diseño y construcción de ontologías de dominio de aplicación.

De acuerdo con [Morbach, 2009] una ontología de dominio tiene como objetivo capturar el conocimiento de todo un dominio de aplicación como la física, la química o la ingeniería.

Las principales características que fueron incorporadas en esta metodología son las siguientes:

- a) Diseño del sistema de ontologías con orientación a dominios.
- b) Centrado en la reutilización de las ontologías mediante el diseño modular.
- c) Lograr que las ontologías resultantes sean coherentes.
- d) Metodología iterativa e incremental.
- e) Evaluación del sistema de ontologías basada en principios de diseño de calidad.
- f) Especificación y evaluación de la competencia del sistema de ontologías por medio de preguntas de competencia.

Esta metodología está prevista para realizarse con un equipo de trabajo integrado por un grupo de expertos en el dominio de aplicación, un grupo de programadores con experiencia en el desarrollo de aplicaciones que explotan ontologías, y un grupo de ingenieros ontológicos. La figura 1 presenta la metodología dividida en cuatro etapas.



Figura 1 Etapas de la metodología para la construcción del sistema de ontologías.

Etapas 1. Diseño General del Sistema de Ontologías

Durante esta etapa se realiza el diseño general del sistema de ontologías, identificando primero los conceptos clave, a partir de estos conceptos definir cuáles son las clases candidatas y separando por dominios los módulos del sistema de ontologías:

- Elicitación de términos. El diseño general de sistema de ontologías para la gestión de horarios se inició con la elicitación de términos; para ello se implementó una técnica de Preguntas de Competencia, la cual consiste en la definición de una lista de preguntas que la ontología debe ser capaz de contestar con todos los conceptos (clases), propiedades entre conceptos, propiedades entre conceptos y datos, los axiomas y reglas definidas en la ontología. Se solicitó a los participantes en el equipo de diseño de la ontología que indicaran las preguntas que les interesaba que la ontología fuera capaz de contestar con respecto al escenario que se planteó al inicio del ejercicio. De esta actividad la lista de preguntas de competencia resultante es la siguiente:

¿En qué salón se encuentra profesor “Pérez”?

¿Cuáles son los cursos que se imparten en el salón F305 los días martes?

¿Qué horario tiene el profesor “Suárez”?

¿Cuáles son los cursos que tienen en su nombre la palabra “programación”?

¿Cuáles son los cursos que se imparten los martes de 10:00 a 12:00?

¿Cuáles son los cursos que pertenecen al Tronco General?

¿Cuáles son los cursos que pertenecen al Tronco Básico Profesional?

¿Cuántos créditos tiene el curso “Teoría de la Computación”?

¿Cuáles son los cursos que pueden llevar corrección?

¿Qué profesores imparten la materia de “Compiladores”?

¿Cuántos créditos suman todos los cursos del área de Sistemas de Información?

¿Cuáles son las claves de los cursos del área de Sistemas Embebidos?

¿Cuáles son las horas y los días en que se imparten los cursos del tronco Inter y Multidisciplinar?

¿Cuáles son los cursos que requieren autorización para inscribirse a ellas?

¿Quiénes son todos los profesores que pertenecen a la División de Ciencias Básicas?

Como resultado de la elicitación de términos, se identificaron los siguientes conceptos como clases candidatas: institución, profesor, estudiante, curso, salón. Antes de finalizar la lista de conceptos, es necesario determinar si se requiere más de una ontología, para ello se deben responder a las siguientes preguntas:

¿La lista de términos identificados podría especializarse de tal forma que su manejo se facilite en una ontología separada?

¿Es posible clasificar los términos en más de una ontología con un dominio particular claramente diferenciado en cada ontología?

¿De qué tamaño será la población de cada ontología y qué tan dinámica será dicha población?

- Identificación de módulos del sistema de ontologías. Esta actividad consiste en realizar un análisis de la lista de términos para identificar los conceptos principales del dominio de la aplicación, los cuales serán representados por ontologías separadas. El objetivo de esta actividad es aplicar el principio de diseño de Modularidad, para obtener ontologías individuales, auto contenidas, reutilizables y más fáciles de actualizar. En particular, para este sistema se identificaron las siguientes ontologías: Persona, UEA (curso), Salón, y Grupo. Las cuales se integrarían en un sistema de ontologías para la gestión de horarios.

Etapas 2. Diseño y Evaluación de las Ontologías Individuales

Durante esta etapa se realiza el diseño y evaluación de cada una de las ontologías que se definieron en la etapa anterior. Como resultado de la identificación de módulos, se decidió que los siguientes conceptos fueran implementados como ontologías individuales: Persona, UEA (curso), Salón, y

Grupo. Por cada una de estas ontologías se debe realizar el proceso iterativo que se muestra en la figura 2, el cual señala las actividades de: definir y verificar las relaciones taxonómicas; definir las relaciones entre objetos y las relaciones entre tipos de datos; realizar la axiomatización de clases, de propiedades y de individuos; y finalmente evaluar cada una de las ontologías individuales.



Figura 2 Proceso del diseño y evaluación de las ontologías individuales.

Descripción de la Ontología Persona:

- Paso 1. En este paso se realizan dos actividades: la definición de la jerarquía de clases y la validación de las relaciones taxonómicas. Como resultado del proceso de diseño de la ontología Persona, se definió la jerarquía de clases que se muestra en la figura 3. La jerarquía de clases consiste de tres clases principales llamadas Persona, GradoAcademico y Departamento. La clase Persona se sub-clasifica en Alumno, Empleado y Visitante, la Clase Empleado se sub-clasifica en Academico y Administraivo, y la clase Academico se sub-clasifica en Ayudante y Profesor. Todas estas clases se definieron para representar a los diferentes tipos de personas que pueden existir en un entorno académico. Para validar que la jerarquía de clases es “semánticamente” correcta se deben enunciar las relaciones taxonómicas; por ejemplo: un Profesor es un Academico, es un Empleado, y es una Persona. Si el enunciado suena congruente para el ser humano y experto del dominio del tema modelado, entonces la jerarquía es correcta.

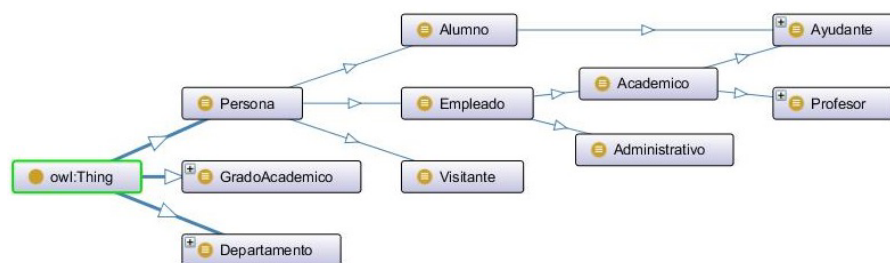


Figura 3 Jerarquía de clases de la ontología *Persona*.

- Paso 2. En este paso se realizan las siguientes actividades: definición de las relaciones entre clases y tipos de datos (dataProperties), así como las relaciones entre clases (objectProperties). Las propiedades de datos representan atributos o características de cada concepto. En una jerarquía de clases se definen primero todos los atributos que son comunes para todos los conceptos; las características comunes deben especificarse en el nivel de clase correcto. Así mismo, es muy relevante especificar una característica distintiva para cada concepto (sub-clase) de la jerarquía de clases para facilitar la tarea de clasificación de individuos del razonador. Las propiedades entre clases que se definieron en la ontología *Persona* son *tieneDepartamento* y *tieneGradoAcademico*, que se utilizaron específicamente con la clase de *Profesor*. Para cada relación entre objetos se debe decidir si la relación es de alguno de los siguientes tipos: funcional, inversa funcional, transitiva, simétrica, reflexiva, etc.
- Paso 3. Definición de axiomas de clases, propiedades e individuos. Durante este este paso se realiza la axiomatización de clases, la cual requiere que por cada clase se determine el conjunto de propiedades (dataProperties, objectProperties) necesarias y suficientes que cada individuo debe cumplir para pertenecer a la clase. También se definen otros tipos de axiomas como: clases disjuntas, cierre axiomático sobre propiedades, y axiomas de cobertura sobre clases. Por cada propiedad definida se deben identificar el tipo de restricciones a utilizar: restricciones de cardinalidad, restricciones existencial o universal, o restricciones de valor. La axiomatización o conceptualización de ontologías puede definirse como el proceso mediante

el cual todos los conceptos (o clases) de una ontología se definen explícitamente estableciendo las condiciones (o restricciones) necesarias y suficientes sobre el conjunto de propiedades de datos o de objetos. En una jerarquía de clases todas las propiedades comunes se heredan en las sub-clases, por lo tanto, se deben definir un conjunto mínimo de propiedades específicas y distinguibles en cada sub-clase para poder identificar de forma inequívoca a cada concepto dentro de la ontología.

En la figura 4 se muestran varios ejemplos de la implementación de axiomas sobre la clase Profesor. Por ejemplo: se observa que un Profesor por el hecho de ser Persona hereda las restricciones sobre las propiedades de tieneGenero y tieneNombrePersona. Así mismo se observa que por ser Empleado hereda la restricción sobre la propiedad de tieneNumEconomico y que por ser Academico hereda las restricciones de tieneCategoriaNivel, tieneCorreoElectronico y tieneDepartamento. Adicionalmente se estableció una restricción más sobre la propiedad tieneGradoAcademico. Por todas las restricciones anteriores, un individuo de la clase Profesor, para pertenecer a esta clase, debe cumplir con todo el conjunto de restricciones, tanto las heredadas como las propias.

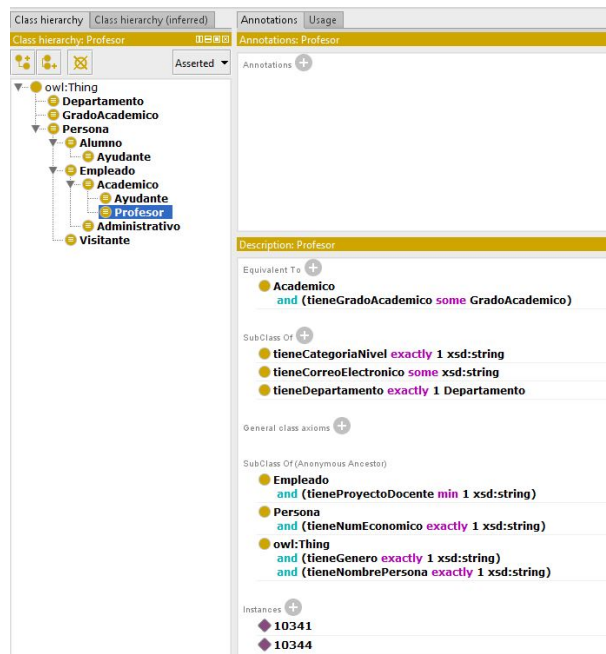


Figura 4 Descripción de la clase Profesor.

- Paso 4. En este paso se realizan las actividades de poblado y corrección de la ontología. El propósito de instanciar individuos en la ontología es para verificar el uso de los conceptos definidos, verificar si los tipos de datos especificados para las propiedades son correctos e intuitivos. En esta actividad se realiza la evaluación de la “usabilidad” de la ontología para la creación de nuevos individuos. Si alguna propiedad de datos o de clases no es clara ni útil, ésta se modifica o se elimina definitivamente.
- Paso 5. En este paso se ejecutan las tareas de chequeo de consistencia lógica del razonador para verificar que la ontología poblada sea lógicamente congruente. En la figura 5 se muestran las propiedades de objetos y de datos de un individuo instanciado en la clase de Profesores. La ontología fue evaluada verificando que fuera consistente lógicamente.

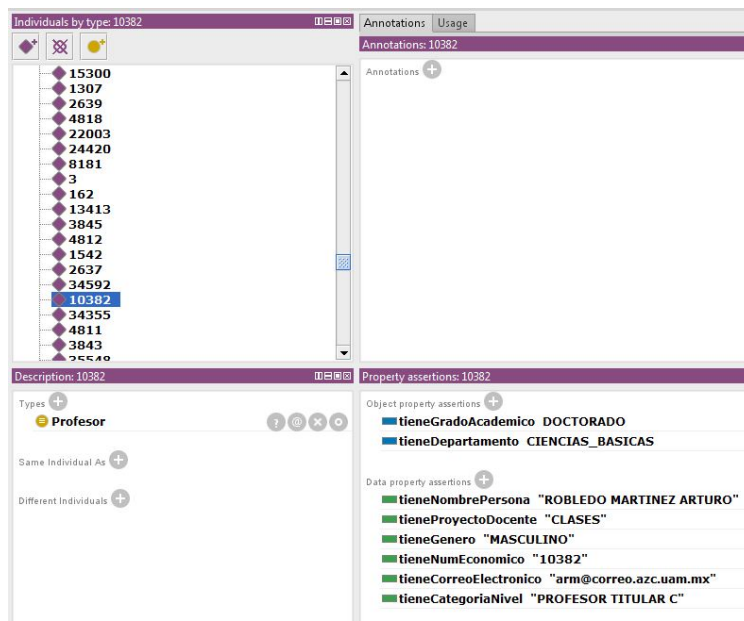


Figura 5 Propiedades instanciadas de un individuo de la clase Profesor.

Etapa 3. Integración del Sistema de Ontologías

Durante esta fase se realiza la importación de las ontologías individuales dentro de una ontología general para la gestión de horarios. La figura 6 muestra las ontologías importadas en la ontología Horario.

La figura 7 muestra los IRIs de las ontologías importadas en la ontología Horario.

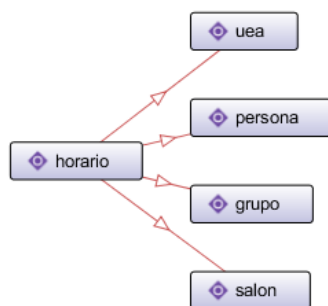


Figura 6 Ontologías importadas en el sistema de ontologías.

```
Ontology: <http://www.academia_uam.com/informacion_academica/ontologies/horario.owl>
Import: <http://www.academia_uam.com/informacion_academica/ontologies/grupo.owl>
Import: <http://www.academia_uam.com/informacion_academica/ontologies/persona.owl>
Import: <http://www.academia_uam.com/informacion_academica/ontologies/salon.owl>
Import: <http://www.academia_uam.com/informacion_academica/ontologies/uea.owl>
```

Figura 7 Ontologías importadas en la ontología **Horario**.

El sistema de ontologías Horario especifica que cada individuo de la clase Horario deberá tener exactamente un grupo, deberá tener exactamente una UEA, y que será identificado exactamente por un único Id. También hereda las restricciones de que los días de clases deberán ser de la clase Dia, que tendrá máximo un salón asignado y que tendrá máximo un profesor asignado, que debe definir un cupo del curso, y una hora de clase. Los axiomas de clase definidos son todos de tipo disyunción. La clase Horario es disjunta de las siguientes clases: Grupo, Dia, Departamento, GradoAcademico, Persona, Salon, Area, Exigencia, Tronco y Uea.

Etapas 4. Evaluación del Sistema de Ontologías

La evaluación del sistema de ontologías se realiza de dos formas: la primera se refiere a la verificación del cumplimiento de los principios de diseño y la segunda se refiere a la competencia del sistema de ontologías, es decir, si ésta es capaz de responder a cada una de las preguntas de competencia que se especificaron al principio de la metodología. La verificación de consistencia lógica del sistema de

ontologías se ejecuta para verificar que ninguna de las definiciones y axiomas de clase tienen contradicciones lógicas, o que algún individuo instanciado en la ontología contradice los axiomas definidos. Esta actividad final de razonamiento consiste en ejecutar las tareas de clasificación taxonómica, derivación de tipos y verificar consistencia.

3. Resultados

Como resultado de la ejecución de la metodología de diseño, en la figura 8 se muestran las métricas del sistema de ontologías para la gestión de horarios. Como se puede observar, el sistema consta de más de 20 clases o conceptos principales, un total de 12 propiedades entre objetos, 23 propiedades de datos y un total de 1069 individuos. Por el tipo de axiomas y restricciones empleado, el nivel de expresividad alcanzado es ALCOQ(D).

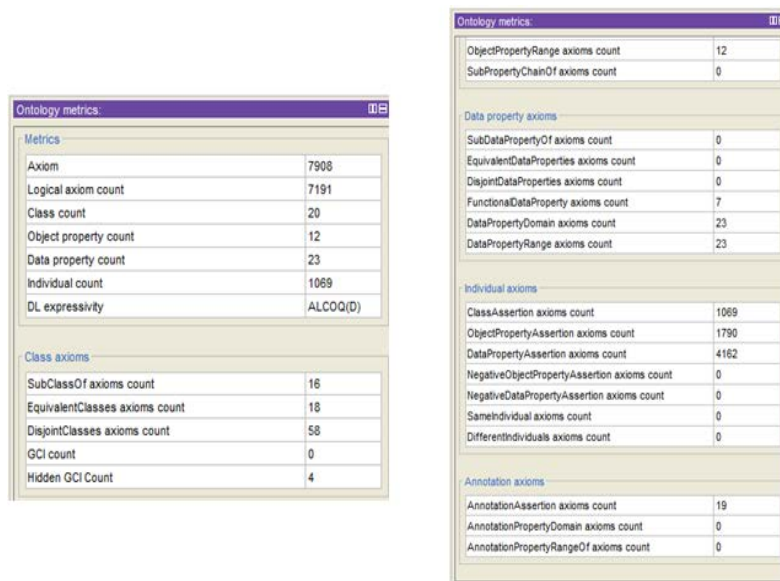


Figura 8 Métricas del sistema de ontologías.

En relación con la evaluación de los principios de diseño, la ontología cumple satisfactoriamente con los siguientes:

- Principio de diseño de claridad. Aseguramos que no hay una interpretación incorrecta al definir sólo las igualdades en las definiciones de las clases axiomáticas.

- Principio de diseño de coherencia. Después de la ejecución del Fact++, todos los individuos fueron clasificados correctamente.

Los principios de diseño más importantes se consideraron y verificaron a través de herramientas de Protege tales como los razonadores y el DLQuery.

Con respecto a la **evaluación de la competencia** del sistema de ontologías se tradujeron las preguntas de competencia al lenguaje de consulta SWRL. La tabla 1 muestra algunos ejemplos de preguntas de competencia traducidas al lenguaje de reglas.

Tabla 1 Traducción de las preguntas de competencia a reglas.

Pregunta	SWRL	Resultado
¿Qué UEAs imparte el profesor "Pérez" indicando salón, días y horas?	Profesor(?prof) ^ tieneNombrePersona(?prof, ?nombre) ^ swrlb:stringEqualsIgnoreCase(?nombre,"Pérez") ^ Horario(?hor) ^ tieneProfesor(?hor, ?prof) ^ tieneDia(?hor, ?dia) ^ tieneHora(?hor, ?horas) ^ tieneSalon(?hor, ?salon) -> sqwrl:select(?prof, ?nombre, ?salon, ?dia, ?horas)	Considerando que esta regla se ejecuta proporcionando el nombre del profesor, devuelve el horario que tiene actualmente el profesor.
¿Cuáles son los cursos que se imparten en el salón F305 los días martes?	Horario(?hor) ^ tieneSalon(?hor, ?salon) ^ swrlb:stringEqualsIgnoreCase(?salon,"F305") ^ tieneUea(?hor, ?uea) ^ tieneNombreUea(?uea, ?nombreUea) ^ tieneDia(?hor, ?dia) ^ tieneHora(?hor, ?horas) -> sqwrl:select(?salon, ?nombreUea, ?dia, ?horas)	Cuando esta regla se ejecuta devuelve la lista de Ueas que se imparten en el salón especificado en la consulta.
¿Cuáles son los cursos que tienen en su nombre la palabra "programación"?	Uea(?uea) ^ tieneNombreUea(?uea, ?nombreUea) ^ swrlb:contains(?nombreUea, "programación") -> sqwrl:select(?uea, ?nombreUea)	Esta regla devuelve las claves y nombres de las Ueas que contienen la palabra especificada en la consulta.
¿Cuáles son los cursos que se imparten los martes de 10:00 a 12:15?	Horario(?hor) ^ Dia(?dia) ^ tieneDia(?hor, ?dia) ^ swrlb:stringEqualsIgnoreCase(?dia,"MARTES") ^ tieneUea(?hor, ?uea) ^ tieneNombreUea(?uea, ?nombreUea) ^ tieneHora(?hor, ?horas) ^ swrlb:stringEqualsIgnoreCase(?horas,"10:00 - 12:15") -> sqwrl:select(?salon, ?nombreUea, ?dia, ?horas)	Esta regla devuelve el salón y los nombres de las Ueas que se imparten el día y la hora especificada en la consulta.

Adicionalmente se ejecutaron consultas utilizando el razonador Pellet y el lenguaje SPARQL en *Protege*. La siguiente pregunta es un ejemplo de la traducción de las pregunta: ¿Qué UEA imparte la Dra. Maricela Bravo, en qué horas y qué días? En la figura 9 se muestra el código de la pregunta en lenguaje SPARQL y la respuesta dada por el razonador Pellet.

Como se puede apreciar en la figura 9, la respuesta del razonador indica que la Dra. Bravo imparte dos UEA: Arquitectura e integración de aplicaciones empresariales los martes y jueves de las 14:30 a las 16:45 horas, y Programación orientada a servicios, también, los martes y los jueves, pero de las 16:45 a las 19:00 horas.

nomProf	nomUea	día	hora
"BRAVO CONTRERAS MARICELA CLAUDIA"	"ARQUITECTURA E INTEGRACION DE APLICACIONES EMPRESARIALES"	JUEVES	"14:30 - 16:45"
"BRAVO CONTRERAS MARICELA CLAUDIA"	"ARQUITECTURA E INTEGRACION DE APLICACIONES EMPRESARIALES"	MARTES	"14:30 - 16:45"
"BRAVO CONTRERAS MARICELA CLAUDIA"	"PROGRAMACION ORIENTADA A SERVICIOS"@	MARTES	"16:45 - 19:00"
"BRAVO CONTRERAS MARICELA CLAUDIA"	"PROGRAMACION ORIENTADA A SERVICIOS"@	JUEVES	"16:45 - 19:00"

Figura 9 Evaluación de la competencia del sistema de ontologías.

4. Discusión

El sistema de ontologías para la gestión de horarios se diseñó y construyó con una metodología que incorpora características de calidad desde la inepción de la ontología. En este artículo las ontologías son consideradas como módulos reutilizables, el objetivo de este enfoque de diseño es que los dueños y desarrolladores de las ontologías las puedan reutilizar de manera independiente

dentro de sus empresas para más aplicaciones. Otro de los requisitos relevantes que debe atender una metodología para el diseño y construcción de ontologías es que incorpore métodos y técnicas para abarcar todos los conceptos y relaciones semánticas necesarios para satisfacer las necesidades de los usuarios y de las aplicaciones de la ontología. Por ello, en esta metodología se hace énfasis en el uso de las preguntas de competencia para la identificación y construcción de relaciones. Durante el proceso de diseño se deben tomar decisiones que eventualmente facilitarán o afectarán la traducción de las preguntas a reglas, así que se debe incorporar un proceso de evaluación continua mediante el poblado y el uso de consultas ya sea con DL-Query, con SWRL o con SARQL.

5. Conclusiones

En este artículo se describió la metodología implementada para el diseño y construcción de un sistema de ontologías para la gestión de horarios. La metodología presentada es holística en el sentido de que abarca desde el diseño de clases hasta la integración y evaluación del sistema de ontologías. Esta metodología incorpora actividades para atender los principios de claridad, coherencia, modularidad y usabilidad. El sistema de ontologías implementado muestra que una metodología que incorpora características de calidad desde el principio logra producir ontologías consistentes, usables y reutilizables.

6. Bibliografía y Referencias

- [1] Fernández, M.; Gómez-Pérez, A., Juristo, N. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. Symposium on Ontological Engineering of AAAI. Stanford (California), March 1997.
- [2] Gómez-Pérez, A., Ontological engineering: A state of the art. Expert Update: Knowledge Based Systems and Applied Artificial Intelligence, 2(3), pp. 33-43, 1999.
- [3] Gruber, Thomas R., Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies 43.5, pp. 907-928, 1995.

- [4] Bernaras, A., Laresgoiti, I., & Corera, J., Building and Reusing Ontologies for Electrical Network Applications. In ECAI pp. 298-302, PITMAN, 1996.
- [5] Gómez Pérez, A. Knowledge Sharing and Reuse. In J. Liebowitz (Editor) Handbook of Expert Systems. CRC, 1998.
- [6] Gómez-Pérez, A., Fernández, M., & Vicente, A. D., Towards a method to conceptualize domain ontologies, 1996.
- [7] Grüninger, M., & Fox, M. S., Methodology for the Design and Evaluation of Ontologies, 1995.
- [8] Lenat, D. B., & Guha, R. V., Building large knowledge-based systems; representation and inference in the Cyc project. Addison-Wesley Longman Publishing Co., Inc, 1989.
- [9] M. Fernández-López, A. Gómez-Pérez, A. Pazos-Sierra, J. Pazos-Sierra, Building a chemical ontology using METHONTOLOGY and the ontology design environment, IEEE Intelligent Systems & their applications 4 (1), pp. 37–46, 1999.
- [10] Morbach, Jan, Andreas Wiesner, and Wolfgang Marquardt, OntoCAPE—A (re) usable ontology for computer-aided process engineering, Computers & Chemical Engineering 33.10, pp. 1546-1556, 2009.
- [11] Schreiber, G., Knowledge engineering and management: the CommonKADS methodology. MIT press, 2000.
- [12] Suárez-Figueroa, M. C., NeOn Methodology for building ontology networks: specification, scheduling and reuse, Doctoral dissertation, Informatica, 2010.
- [13] Uschold, M., & Gruninger, M., Ontologies: Principles, methods and applications. Knowledge engineering review, 11(2), pp. 93-136, 1996.
- [14] Uschold, M., & King, M., Towards a methodology for building ontologies, Edinburgh: Artificial Intelligence Applications Institute, University of Edinburgh, pp. 15-30, 1995.

SELF-ORGANIZING MOBILE ROBOTS BASED ON MULTI-AGENT COORDINATION TECHNIQUES IMPLEMENTED WITH AERIAL VISION AND COMMUNICATION GATEWAY BETWEEN WIFI AND RF

Cynthia Daniela Briones Valencia

Universidad Autónoma de Guadalajara
danycdbv@gmail.com

Zandor Alfredo Machaen Terriquez

Universidad Autónoma de Guadalajara
zandor92@gmail.com

Luis Martin del Castillo

Universidad Autónoma de Guadalajara
luismartin.925@gmail.com

Gustavo Alejandro Torres Blanco

Universidad Autónoma de Guadalajara
gustavo.blanco@edu.uag.mx

Resumen

Este artículo presenta el desarrollo de robots móviles que poseen la capacidad de búsqueda y recuperación de obstáculos en un entorno de laberinto. El algoritmo incorporado en los robots fue diseñado con base en principios de coordinación y autoorganización, es decir, un grupo de agentes autónomos coordinan sus acciones para buscar y recuperar obstáculos del entorno a través de la cooperación. Para ello, se diseñaron dos tipos de agentes, organizadores y operadores. Los organizadores tratan de coordinar las acciones de los operadores, y estos últimos, tratan de recuperar todos los obstáculos en el medio ambiente. Cinco robots de cuatro ruedas fueron construidos desde cero utilizando Arduino Uno para los operadores, y Arduino Nano y NXP i.MX53 Quick Start

Boards para los organizadores. Además, se utilizó una cámara aérea (fijada al techo) para proporcionar percepción visual a los robots. La comunicación se realizó a través de una pasarela entre el canal de 8bit RF y WiFi, para los operadores y los organizadores, respectivamente.

Palabras Claves: Autoorganización, robots móviles, visión computacional.

Abstract

This paper presents the development of mobile robots that have the abilities of search and retrieval of obstacles in a maze-like environment. The algorithm embedded in the robots was designed based upon principles of coordination and self-organization, i.e., a group of autonomous agents coordinate their actions in order to search and retrieve obstacles from the environment through cooperation. To do this, two types of agents were designed, organizers and operators. Organizers try to coordinate the actions of the operators, and these last, try to retrieve all obstacles in the environment. Five four-wheeled robots were built from scratch using Arduino Uno for the operators, and Arduino Nano plus NXP i.MX53 Quick Start Boards for the organizers. Also, an aerial camera (attached to the ceiling) was used to provide visual perception to the robots. The communication was made through a gateway between 8bit channel RF and WiFi, for the operators and organizers respectively.

Keywords: Computer vision, mobile robots, self-organization.

1. Introduction

The solution of agent societies problems is a very complex task, it requires coordination techniques, policies and diverse communication standards. Fulfilling these requirements guarantees that the system will accomplish the objectives for which it was designed. This can be accomplished through the contribution of the individual knowledge of each of the members within the system because it allows to leverage the abilities and knowledge of each member that comprises the system. So, the general purpose of solutions like the one presented in this document is to divide complex tasks into single and simple sub-tasks, which will

reduce the computing order on space and time in each member of the system. Thus, these kinds of techniques improve the performance when a society of agents tries to achieve a global objective [Dimopoulos, 2006], [Chen, 2005], [Turner, 2013].

The self-organization principle states that the agents have a partial knowledge of the environment and can define, in an individual manner, the rules that regulate the interaction between members. It is believed that these interactions improve the performance of the agents within emerging environments. Therefore, they have abilities to sense and modify the environment and communicate with the rest of the agents within the system. Regardless of the environment configuration, the agent local interactions are performed to integrate all this knowledge in order to completely learn the current environment [Di Marzo, 2005], [Mataric, 1993]. Some methodologies and rules within software engineering have defined the coordination, scalability and redundancy in order to ensure the fulfillment of the given tasks [Yoshida, 2007], [Gomez-Sanz, 2006], [Criado, 2013].

The proposed approach in this document is to work with a decentralized system, [Muñoz, 2005], [Jacyno, 2013] which allows, unlike [Peng, 2013], to eliminate the need of relying on a leader to assign the tasks to each one of the agents. The design of the society is based on local interactions, through direct communication by message passing that allows the agents to determine their actions. Moreover, the whole idea is based principally in studies developed by Marco Dorigo at Iridia Labs. [Brambilla, 2012].

This paper is structured as follows. The description of the components of the system and the definition of the abilities of the agents is presented in section 2. The functions that define the interactions of each agent are described in section 3. The implementation and the results of the proposal appear in section 5. Finally, the conclusions and the future work are mentioned in section 6.

2. Methods

The maze-like environment is integrated by a finite number of obstacles, five mobile robots (agents) that are randomly distributed throughout the environment:

operator agents and organizer agents, three and two members, respectively, which must collaborate in order to achieve the following three main objectives: obtain a maze free of obstacles, organize obstacles outside the maze, and organize themselves into a formation.

The physical environment and the agents may be observed in figure 1. The maze had to be adjusted to the range of vision of the aerial camera, so the robots were to turn at 45 degrees through the perpendicular aisles. This was because a robot with a forklift was not able to completely turn to perform routes at right angles.

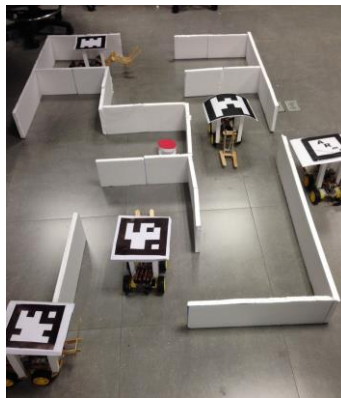


Figure 1 Maze configuration, the robots have a tag, the obstacle is a red dot.

Organizer Agent

These agents' duty is to coordinate the operators when they leave the environment and to relocate the obstacles that are already outside of the maze, their main functions are:

- Detect an agent outside the maze.
- Relocate obstacles towards the exit.
- Maintain the control of the agent's exit.
- Inform the agents that they may start the formation.

Operator Agent

Agents in charge of removing the obstacles that are immersed within the environment and evacuate through the nearest exit available, their main capabilities are the following:

- Detect obstacles in the environment.
- Remove obstacles from the environment.
- Request the organizing agent to relocate the obstacles away from the exit.
- Receive input notifications from the organizing agent.
- Initiate the formation outside of the maze.
- Search for maze exit.
- Avoid obstacles.

Mobile Robots

Five four-wheeled robots were used in this project, each of them built with the following components:

- Four 5 V DC motors, bridged in pairs.
- One H-Bridge to control the two pairs of motors.
- A forklift to grasp the obstacles in the maze.
- A unique tag with different patterns for aerial vision control.
- A 5 V servomotor for the forklift.

The two types of mobile robots were implemented using two different microcontroller architectures. First, the operator agents were built using Arduino Uno microcontrollers based on ATmega328P with the following extra components:

- A multichannel RF for wireless SPI protocol.
- Eight AA batteries to feed H-Bridge and Arduino Uno.

Secondly, the organizer agents were built using an i.MX53 QSB based on 32 bits ARM Cortex-A8 with the following extra components and software:

- A WiFi USB dongle.
- Arduino nano to control H-Bridge through serial communication (i.MX53 doesn't have GPIO ports).
- GNU/Linux Yocto based operating system.
- JADE framework to run agents.

Gateway Between WiFi and RF Communication Protocols

A Laptop PC with GNU/Linux operating system was used to create a gateway through WiFi and RF. WiFi interface was used to create a connection with the organizers, and serial communication with an Arduino Nano was used to create connections with the operators.

Maze

The maze was built with polystyrene sheets. It was made only to present a visual representation of the maze to human observers. The maze is hardcoded in the aerial vision system, so all mobile robots control is made on the laptop PC through artificial vision.

Aerial vision System

An Android smartphone camera, connected through WiFi using IP streaming based application, was used to obtain visuals from the environment. The vision system is made up of the following software components:

- OpenCV library:
 - ✓ ARma framework for robot tags recognition.
 - ✓ PatternDetector Library from George Evangelidis [Evangelidis, 2017].
 - ✓ Color recognition module for obstacle detection.
- JADE framework for the agent's container.
- A JADE agent that acts as the interface to the WiFi to RF gateway.
- A Java based application interface to Arduino Nano SPI master using
- USB serial communication between the PC and the Arduino Nano.
- A Java based application to perform all coordination algorithms for the robots. Thus, it sends actions to all robots in the environment.
- Path planning is performed using A* algorithm.
- Additionally, the vision system is responsible of the following actions:
 - Identify every robot in the environment through pattern recognition.
 - Provide vision to all robots, in order to tell them where to move.
 - Recognize the obstacles in the environment.

- Detect when a robot is able of lift an obstacle.

The pattern detector algorithm is as follows: first, detect pattern corner in order to establish pattern's position in every captured frame. Then, normalize the ROI for every detected pattern and compare it with loaded patterns. Once a pattern match with a known one, estimate de transformation between camera's cardinal system and the pattern rotation. Finally, use the transformation matrix for pattern location, rendering, etc. [Evangelidis, 2017].

In the following section, the description of the algorithms for the operators and organizers are presented. Then, the implementation issues and results are detailed.

The algorithms described in the following paragraphs were analyzed for completeness and soundness. All algorithms are complete because there are only if-else statements that covers up every choice in the space. Concerning to the soundness, it is affected by correct pattern detection and reliable communication, because of the hardware used in the implementation it can be said that there are not sound. But, in a future work it will be interesting to perform a study, concerning distributed systems and image processing algorithms, which was not the main focus of this research.

Operator Agent Functions

Forward: Function that allows the agent to move inside the maze until it finds a barrier that hinders it, which should be analyzed to determine which action to perform afterwards:

```
If(exit_found&&bring_obstacle)Then  
    Send request to the organizer to get out  
Else if(exit_found &&!bring_obstacle)Then  
    Avoid Exit  
Continue forward  
Else if(barrier_found)Then  
    Identify type of barrier  
Else  
    Continue forward  
End if
```

Identify Barriers: This function allows the agent to determine the actions that it must take according to the type of barrier that it finds within the environment:

```
If(obstacle)Then
    If(bring-obstacle)Then
        Avoid obstacle
    Continue forward
Else
    Get obstacle
    Find Exit
End if
Else if(wall) Then
    Avoid barrier
    Continue forward
Else if(agent)Then
    Stop for a time t
    Continue Forward
End if
```

Exit the Maze: The operator gets ready to start formation when it receives a notification to exit the environment by the organizer.

```
Request approval to get out of the labyrinth
If(request_approved) Then
    Place obstacle out of the labyrinth
    Wait for notification of organizing agent to get out
If(notification_received) Then
    Gets out of the labyrinth
    Start formation
End if
Else
    Find another exit
End if
```

Start Formation: Function that allows the agent to integrate into the formation once it is out of the environment:

```
If (agents_in_formation == 0) Then
    Start formation
Else
    Request re-organization of the agents
    Integrate into formation
End if
```

Functions of the Organizing Agent

Exit Authorization: Once the organizer receives an exit request from an operator agent, it will analyze if there is still a request in the formation in order to send an answer to the received request.

```
If (number_of_elements <= 3) Then  
    Send approval  
If (detecta_obstaculo) Then  
    Get the obstacle  
    Relocate the obstacle  
End if  
Else  
    Reject exit.  
End if
```

Relocate obstacle: Allows the organizer to detect an obstacle outside of the maze in order to relocate it with the rest of the obstacles that are already outside.

```
If(obstacle_ouside != 0) Then  
    Move obstacles that are outside  
End if  
Relocate obstacle  
Return to initial position  
    Send notification to the operator agent to start the formation.
```

3. Results

The simulation can be divided into three major phases: obstacle search, obstacle removal and outskirts formation. Every phase its explained in the following paragraphs. Before explaining phase's results, a discussion of some characteristics of the implementations is needed.

The most complex location is the one that is most distant to all the robot operators because the calculation of the roads is more complex. Moreover, time is not relevant, because the goal is not to finish in a time limit, but note that the mobile robots exhibit, through its search and retrieval capabilities of obstacles in a maze environment, cooperative behaviors. If there were more obstacles to be removed, as the system is parallel, it should work if there are the same number of obstacles

than robots. If there were more obstacles than robots, it should not complete the scenario.

Obstacle search: First, the robots try to find obstacles in the environment, performing a random search controlled by the aerial vision system. The vision system sends movement commands to adjust the travel of every robot.

When an obstacle is found, the robot approaches it, then it lift its forklift in order to retrieve the obstacle. Then, it communicates with an organizer robot to request to exit the maze, figure 2 shows these actions.

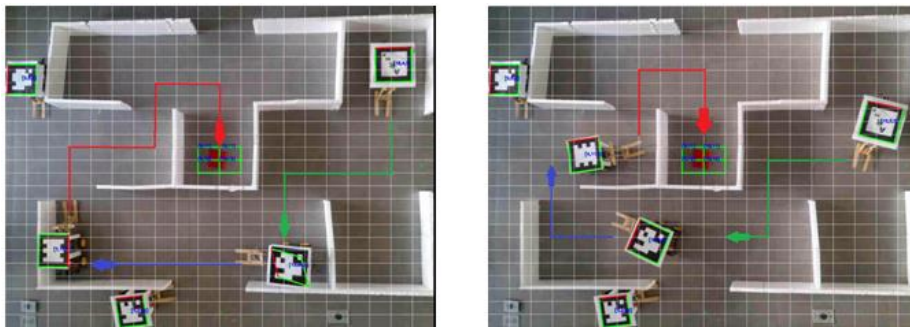


Figure 2 Obstacle search, the obstacle is identified by the aerial system.

Obstacle removal: Secondly, the robots that does not retrieve any obstacle receive a request from operator to exit the maze. Meanwhile, the robot that retrieved the obstacle tries to get out of the maze, aided by the vision system using an A* algorithm to generate the path to the nearest exit with an organizer robot. figure 3 shows these actions.

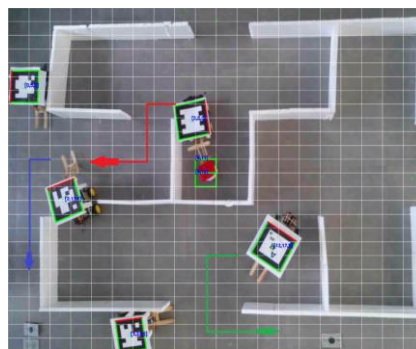


Figure 3 Obstacle removal, the nearest robot lifts the obstacle in order to remove it from the maze.

Outskirts formation: Thirdly, when the retriever robots gets to the exit, the organizer robot tells him to put down the object, removes the obstacle and tells the retriever to start the formation. The retriever robot exits the maze. figure 4 shows these actions.

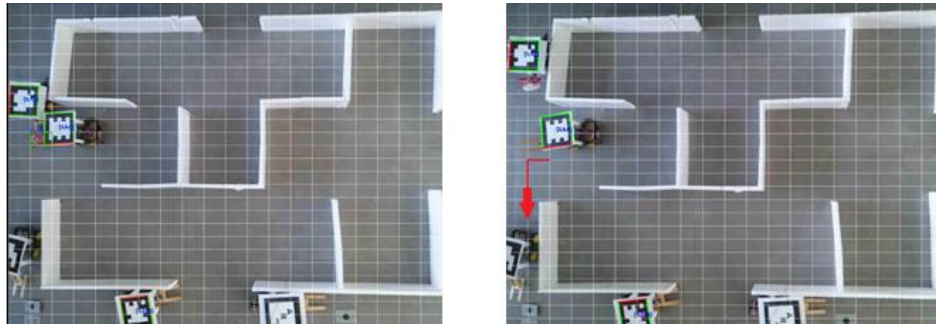


Figure 4 Organizer robot job, the organizer robot tells the retriever to leave the obstacle and begin with the formation.

Finally, the retriever robot accommodates in a formation on the outskirts of the maze. Meanwhile, the other robots keep searching for obstacles in the environment. figure 5 shows these actions.

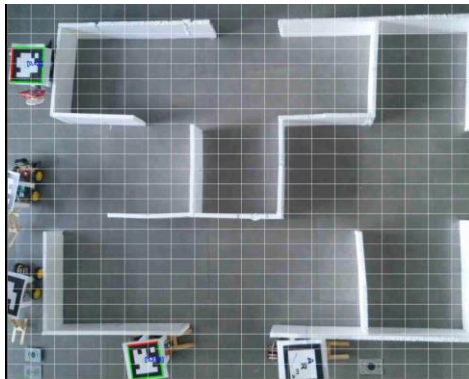


Figure 5 Formation, the retriever accommodates in a formation.

4. Discussion

The construction and implementation of the robots and the algorithms were not straightforward. The following list details most of the issues that the team solved:

1. Robot movement control through vision, caused by floor slipperiness.
2. Room illumination variation that affects vision.

3. Power issues caused by the battery packs.
4. Forklift design.
5. Integration of components, caused by power feeding issues.
6. RF Communication, caused by modules malfunction.
7. Robots free rotating wheels.

The solutions are listed as follows:

1. Diminish the distance traveled by the robots, and vision correction through a squared threshold in the virtual environment.
2. Blocking sunlight from the environment.
3. It was not resolved very well, in the future it is expected to use better performance battery packs.
4. Forklift design needs improvement, maybe using a 3D printer.
5. Power feed the H-bridge and motors independently of the microcontroller.
6. The issue was resolved by switching RF modules whenever there is a malfunction due to power issues.
7. Change free rotating wheels with an extra pair of DC motors and wheels.
8. It is important to note that the solutions need to be improve implementing more adequate methods, encountered in robotics theories. Most of the solutions presented here were made because of the pricing of more accurate components.

5. Conclusions

This paper detailed the coordination algorithms, the technologies for computer vision and communication gateway. The results were not as planned because of the several issues that the team faced in the construction and implementation of the robots. All major issues were explained in the corresponding section, and the solutions were not the best. But these can be taken as improvement opportunities to be prepared as future work.

The team has now the responsibility to improve the performance, first by trying to get more adequate materials and hardware, like switching forklift made up of wood

with either 3D printed or acrylic ones. Switching the NXP i.MX53 that lack of GPIO ports with microcontrollers that did have ports. Improving the traction of the robots using a less slippery floor. Improving the maze with a more professional material than polystyrene, and so on. It is expected to fulfill these improvements as soon as possible.

6. Bibliography and References

- [1] Chen, G., Yang, Z., He, H., and Goh, K. M., Coordinating multiple agents via reinforcement learning, *Autonomous Agents and MultiAgent Systems*, vol. 10, no. 3, pp. 273–328, 2005.
- [2] Criado, N., Using norms to control open multi-agent systems, *AI Commun.*, pp. 317–318, 2013.
- [3] Brambilla, M., Ferrante, E. Birattari, M., Dorigo, M. Swarm Intell Swarm robotics: a review from the swarm engineering perspective. *Swarm Intelligence*, Springer, 7 (1), pp.1-41, 2013.
- [4] Di Marzo, K. M., Serugendo, M., Gleizes, P., and Karageorgos, A., Self-organization in multi-agent systems, *Knowl. Eng. Rev.*, vol. 20, no. 2, pp. 165–189, 2005.
- [5] Dimopoulos, Y. and Moraitis, P., Multi-agent coordination and cooperation through classical planning, in *2006 IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, pp. 398–402, 2006.
- [6] Evangelidis, G. Arma library: Pattern Tracking for Augmented Reality: <https://sites.google.com/site/georgeevangelidis/arma> at June 2017.
- [7] Gomez-Sanz, J. J., and Pavon, J., Defining coordination in multi-agent systems within an agent oriented software engineering methodology, in *Proceedings of the 2006 ACM Symposium on Applied Computing*, pp. 424–428, 2006.
- [8] Jacyno, M., Bullock, S., Geard, N, Payne, T. R., and Luck, M., Self-organizing agent communities for autonomic resource management, *Adaptive Behavior - Animals, Animats, Software Agents, Robots, Adaptive Systems*, vol. 21, no. 1, pp. 3–28, 2013.

- [9] Mataric, M. J., Designing emergent behaviors: From local interactions to collective intelligence, in *Proceedings of the Second International Conference on From Animals to Animats 2: Simulation of Adaptive Behavior: Simulation of Adaptive Behavior*, pp. 432–441, 1993.
- [10] Muñoz Salinas, R., Aguirre, R., García-Silvente, M., and Gomez, M., A multi-agent system architecture for mobile robot navigation based on fuzzy and visual behaviour, *Robotica*, vol. 23, no. 6, pp. 689–699, 2005.
- [11] Peng, Z., Wen, G., and Rahmani, A., Leader-follower formation control of multiple nonholonomic robots based on backstepping, in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pp. 211–216, 2013.
- [12] Turner, J. R., Multiagent systems as a team member, *International Journal of Technology, Knowledge & Society*, vol. 9, no. 1, pp. 73–90, 2013.
- [13] Yoshida, T., Cooperation learning in multi-agent systems with annotation and reward, *Int. J. Know. -Based Intell. Eng. Syst.*, vol. 11, no. 1, pp. 19–34, 2007.

COMPARATIVA KINECT VS MYO APLICANDO LA PRUEBA NASA-TLX EN UN ENTORNO DE RVI PARA INSPECCIÓN EN AEROGENERADORES

Daniel Cantón Enríquez

Universidad del Istmo

daniel_ce92@outlook.com

J. Jesús Arellano Pimentel

Universidad del Istmo

jjap@sandunga.unistmo.edu.mx

Miguel Ángel Hernández López

Universidad del Istmo

mahl@sandunga.unistmo.edu.mx

Francisco Aguilar Acevedo

Universidad del Istmo

aguilar.afco@sandunga.unistmo.edu.mx

Resumen

Este artículo presenta un estudio comparativo entre dos dispositivos de interacción natural de usuario que son frecuentemente utilizados en sistemas de realidad virtual inmersiva: Kinect y Myo. Cada uno de estos dispositivos es usado en un sistema de realidad virtual inmersivo enfocado a la inspección de aerogeneradores. Se realiza una evaluación de ambos empleando como principal instrumento la prueba NASA-TLX, también llamado índice de carga mental. Los resultados permiten corroborar la potencial aplicación de cualquiera de los dos dispositivos. No obstante, el dispositivo Myo aventaja ligeramente al dispositivo Kinect como el más idóneo.

Palabras Claves: Energía eólica, interacción natural de usuario, NASA-TLX, realidad virtual inmersiva.

Abstract

This paper presents a comparative study between two natural user interaction devices that are frequently used in immersive virtual reality systems: Kinect and Myo. Each of these devices is used in an immersive virtual reality system focused on the inspection of wind turbines. An evaluation of both using the NASA-TLX test, also called mental load index, is the main instrument. The results allow to corroborate the potential application of either device. However, the Myo device slightly outstrips the Kinect device as the most suitable.

Keywords: *Immersive virtual reality, NASA-TLX, natural user interaction, wind energy.*

1. Introducción

De acuerdo con [Montuschi, 2014] la Interacción Humano Computadora (IHC) involucra el diseño, implementación y evaluación de nuevas interfaces para mejorar la interacción entre los dispositivos electrónicos y los usuarios. Una interfaz adaptable, intuitiva y natural puede reducir en gran medida el modelo mental entre un humano y la forma en la que una máquina o sistema computacional puede realizar una tarea determinada. La investigación dentro de la IHC ha permitido la construcción simuladores de Realidad Virtual (RV) empleados en la capacitación de personas para que sean actores efectivos en ciertos sectores como el industrial o el energético, entre otros [Denning, 2006].

Uno de los dispositivos de visualización e interacción de mayor uso, para sistemas de RV Inmersiva (RVI), es el casco Oculus Rift, de hecho algunos investigadores como [Hilfert, 2016] lo consideran un elemento primordial en los sistemas RVI, y algunos otros como [Freina, 2015] enlistan una gran cantidad de publicaciones científicas que involucran el uso de este dispositivo en sistemas de RVI con propósitos educativos. Aunado al uso de los cascos como el Oculus Rift, estudios como los de [Steed, 2016] y [Valkov, 2016] han concluido que emplear avatares incrementa y mejora la percepción de inmersión por parte de los usuarios.

No obstante, no todo gira en torno a la inmersión, también la interacción juega un papel preponderante en los sistemas de RVI cuando son empleados en la

instrucción o capacitación de personas para que se desempeñen en ciertos sectores, sobre todo aquellas actividades que durante el proceso de enseñanza presentan un potencial riesgo físico para los involucrados, por ejemplo, en el sector energético donde se trabaja con altos voltajes [Flores, 2014].

En investigaciones recientes sobresalen tres dispositivos que permiten la interacción natural de usuario a través de gestos con entornos de RVI, estos son: Kinect, Myo y Leap Motion. Por ejemplo, [Holmes, 2016] emplean el casco Oculus Rift DK1 como dispositivo de inmersión y los dispositivos Kinect V2, Leap Motion y Myo como dispositivos de interacción natural de usuario, su sistema de RVI está enfocado a la rehabilitación de personas con accidentes cardiovasculares. En [Han, 2017] utilizan el casco Oculus Rift DK2 como dispositivo de inmersión y el Leap Motion como medio de interacción natural de usuario para instruir a futuros operarios de máquinas CNC. Por su parte [Jiménez, 2016] desarrollaron un entorno de RVI que integra el casco Oculus Rift DK2 y Kinect V1 con el propósito de interactuar en tiempo real con modelos 3D altamente detallados de sitios culturales arquitectónicos.

Respecto al campo del sector eólico no son tan comunes los trabajos que aborden el uso de la RVI, ya sea con fines comerciales o didácticos, a pesar de que según [APREAN, 2007] los parques eólicos requieren de personal altamente calificado para su operación y mantenimiento, ya que existen riesgos que ponen en peligro la integridad física de quienes ahí laboran. En la industria eólica se sabe de dos casos donde aprovechan el potencial de la RV con fines de difusión y comercialización: la aplicación móvil "ACCIONA Virtual Experience" [ACCIONA, 2015] y la "experiencia de realidad virtual 100% inmersiva" [ACCIONA, 2016]; ambos fueron desarrollados por la empresa española de energías renovables ACCIONA. En el primer caso se utilizó un teléfono inteligente conjuntamente con unos lentes de realidad virtual, en el segundo caso los dispositivos utilizados fueron las gafas de realidad virtual y dos mandos inalámbricos para moverse por un recorrido virtual.

Otros dos trabajos relacionados con el sector eólico son [Trujillo, 2015], en donde se presenta la virtualización tridimensional de un parque eólico con fines

didácticos, empleando el casco Oculus Rift DK2 como medio de inmersión, conjuntamente con el control de la consola Xbox 360 como medio de interacción, además de [Hernández, 2015] que presenta un videojuego didáctico para la manipulación virtual de un aerogenerador a través de gestos reconocidos por el SDK del Kinect V1. Este último trabajo si bien no consiste en un RVI si cuenta con una visualización 3D del aerogenerador, siendo la idea principal la interacción natural de usuario a través de gestos de manos y brazos.

Además de los trabajos anteriores, también es importante mencionar algunas investigaciones cuyo eje central es contrastar el uso de diferentes dispositivos de interacción natural de usuario en entornos 3D. Por ejemplo, en [Vokorokos, 2016] comparan los resultados de un experimento sobre la eficiencia en el reconocimiento de cinco diferentes gestos, en cinco videojuegos distintos diseñados para alguno de los tres dispositivos antes mencionados: Leap Motion, Kinect & Myo. Por su parte, la investigación realizada por [Sánchez, 2017] prueban la viabilidad de utilizar comandos gestuales y de voz en una sala de operaciones estéril, reportan resultados de ocho aspectos comparativos entre los dispositivos Kinect, Leap Motion y Myo con comandos de voz, dado que se trata de una laparoscopia en una intervención quirúrgica dichos aspectos son: exactitud, confort, desconexión, inicialización, intuitividad, esfuerzo físico, velocidad y facilidad de uso.

Con base en lo anterior, el propósito del presente artículo es realizar un estudio comparativo entre dos de los tres sensores más utilizados en la interacción natural de usuario: el Kinect y el Myo; esto dentro de un entorno de RVI que emplea el casco Oculus Rift DK2 como dispositivo de inmersión, además de avatares para reforzar la sensación de presencia durante la inspección de un aerogenerador eólico. Dicho estudio comparativo toma como principal recurso la prueba NASA-TLX, también llamada índice de carga de trabajo o índice de carga mental [NASA-TLX, 2011]. Además, en el presente trabajo se realiza un conteo de la repetitividad de los gestos realizados por los usuarios al momento de interactuar con el sistema de RVI, esto para verificar la facilidad de uso en cada gesto. Los resultados, en concordancia con algunos trabajos relacionados, permiten corroborar la potencial

aplicación de los dispositivos de interacción natural de usuario en los sistemas de realidad virtual inmersivos, tomando en cuenta las principales ventajas y desventajas inherentes a la propia tecnología de cada dispositivo y a la valoración subjetiva de los usuarios finales. Las diferencias encontradas entre los dispositivos estudiados muestran que el dispositivo Myo aventaja ligeramente al dispositivo Kinect.

2. Métodos

La figura 1 esquematiza la metodología que se empleó en el presente trabajo. Se realizaron cuatro fases generales: desarrollo del entorno de RVI en dos versiones, preparación de la prueba NASA-TLX, uso de los gestos para la interacción natural de usuario (entrenamiento y tareas), así como valoración de la prueba NASA-TLX y la generación de resultados. En las siguientes subsecciones se describirán a más a detalle cada una de estas fases.

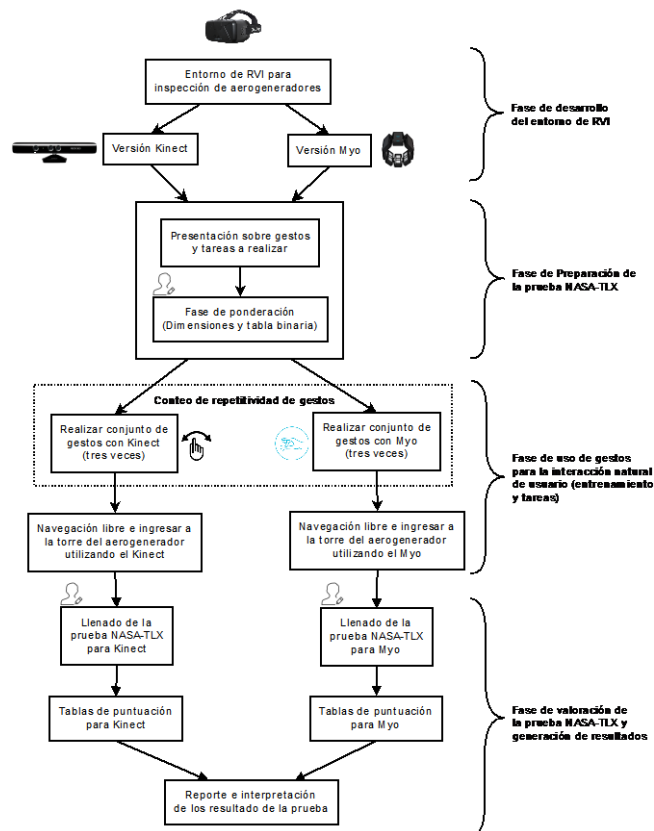


Figura 1 Metodología de trabajo.

Desarrollo del Sistema de RVI

Durante la primera fase de la metodología de trabajo se empleó el modelo incremental de desarrollo de software [Pressman, 2010]. El primer incremento consistió en crear el ambiente de RVI empleando la plataforma Unity conjuntamente con el casco Oculus Rift, ver figura 2a y figura 2b, para ello fue necesario importar el modelo genérico de un aerogenerador 3D diseñado en la herramienta SketchUp. El segundo incremento integró un par de avatares articulados al ambiente (figura 2C), estos fueron diseñados en Adobe Fuse CC y animados con la plataforma en línea Mixamo, para finalmente importarlos en la plataforma Unity, listos para ser manejados en primera persona.

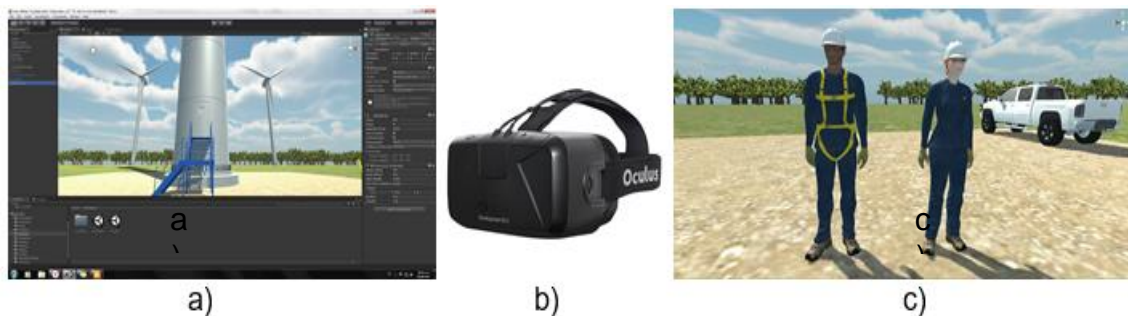


Figura 2 Sistema de RVI en su fase de desarrollo.

El tercer incremento incorporó los gestos del Kinect para manipular las acciones de uno de los avatares en el ambiente, ver figura 3a, con este incremento se generó la versión de Kinect. Por último, el cuarto incremento adicionó al segundo incremento la capacidad de reconocer los gestos del Myo para manipular el avatar, de esta forma se obtuvo la versión de Myo. Una vez probadas y depuradas ambas versiones se generaron las condiciones para realizar las pruebas con usuarios finales.

Preparación de la Prueba NASA-TLX

La prueba NASA-TLX distingue seis dimensiones de carga mental [NASA-TLX, 2011]. En la tabla 1 se presentan cada una de las dimensiones con una breve descripción de lo que subjetivamente miden.

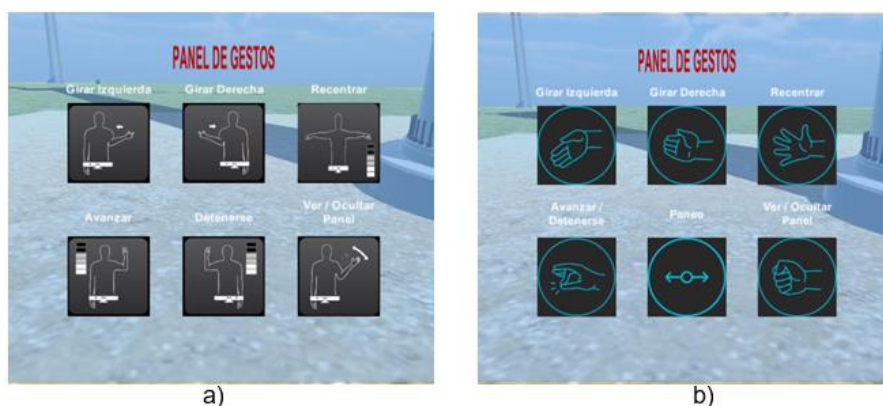


Figura 3 Paneles de gestos.

Tabla 1 Dimensiones de la prueba NASA-TLX, Fuente: [Rubio, 2008].

Dimensión	Descripción de lo medido
Exigencia Mental (M)	Cantidad de actividad mental y perceptiva que requiere la tarea (por ejemplo: pensar, decidir, calcular, recordar, mirar, buscar, etc).
Exigencia Física (F)	Cantidad de actividad física que requiere la tarea (por ejemplo: pulsar, presionar, girar, deslizar, etc).
Exigencia Temporal (T)	Nivel de presión temporal percibida. Razón entre el tiempo requerido y el disponible.
Rendimiento (R)	Hasta qué punto el sujeto se siente insatisfecho con su nivel de rendimiento.
Esfuerzo (E)	Grado de esfuerzo mental y físico que tiene que realizar el sujeto para obtener su nivel de rendimiento.
Nivel de Frustración (Fr)	Hasta qué punto el sujeto se siente inseguro, estresado, irritado, descontento, etc. durante la realización de la tarea.

En la segunda fase se tomó como referencia el trabajo de [Rubio, 2008] respecto al empleo del procedimiento de la prueba NASA-TLX en sus dos etapas de aplicación: una de obtención de la importancia inicial que tiene cada dimensión de carga mental para cada individuo (usuario), y otra de evaluación. En la primera etapa, los usuarios realizan todas las comparaciones binarias entre las seis dimensiones, señalando cuál de las dos les parece mayor fuente de carga mental en una tabla binaria, ver figura 4, en total siempre se realizan 15 comparaciones. No obstante, para, que el usuario tenga una mejor comprensión de lo que va a realizar antes, durante y después de usar el sistema de RVI, primeramente se le presentan los paneles de gestos de las dos versiones del sistema, ver figura 3, y agrandes rasgos se describen las taras a realizar en él. También se le da una breve explicación de la prueba NASA-TLX y sus dimensiones. Esto contribuye a

facilitar el llenado de la tabla binaria. La segunda etapa de evaluación de la prueba se aborda posteriormente.

M - F	F - T	T - R	M: Exigencia Mental
M - T	F - E	T - Fr	F: Exigencia Física
M - E	F - R	E - R	T: Exigencia Temporal
M - R	F - Fr	E - Fr	R: Rendimiento
M - Fr	T - E	R - Fr	E: Esfuerzo
			Fr: Nivel de Frustración

Figura 4 Tabla binaria de las seis dimensiones (la llena el usuario).

Uso de Gestos en Entrenamiento y Tareas

La fase de uso de gestos para la interacción natural de usuario también consistió en dos etapas: entrenamiento de los gestos y tareas a realizar. Durante la etapa de entrenamiento se solicitó a los usuarios que realizaran cada uno de los gestos tres veces, contabilizando el número de intentos fallidos para lograr que el sistema los reconociera de forma correcta. Esto se realizó para ambas versiones del sistema de RVI con todos los usuarios. Durante la etapa de tareas a realizar se pidió a los usuarios ejecutar dos tareas, en la primera deberían recorrer el ambiente de RVI libremente, en la segunda su misión consistiría en entrar a la base del aerogenerador, para lo cual era necesario ubicar las escaleras en la base, subirlas e introducirse en la torre pasando por la puerta de acceso. Ambas tareas se realizaron con la versión Myo y con la versión de Kinect, figuras 5a y 5b.



a) Usando el Myo.



b) Usando el Kinect.

Figura 5 Tareas realizadas por los usuarios: navegación libre e introducirse a la torre.

Valoración de la Prueba NASA-TLX

Una vez realizadas las etapas de llenado de la tabla binaria y las tareas dentro del sistema de RVI, el protocolo de la prueba NASA-TLX señala que los usuarios deben estimar, en una escala del 0 al 100, dividida en intervalos de 5 unidades, la carga mental de cada una de las 6 dimensiones, ver figura 6 (usuario llena una por cada dispositivo). Con los datos obtenidos a través de la tabla binaria y la etapa de evaluación de la prueba es posible calcular un índice global de la carga mental de la tarea aplicando la ecuación 1.

$$IC = (\sum p_i X_i) / 15 \quad (1)$$

Donde:

- IC Índice de Carga
- p_i Peso obtenido para cada dimensión en la tabla binaria (ponderación)
- X_i Puntuación obtenida por la dimensión en la etapa de evaluación
- 15 Número de comparaciones binarias realizadas

Exigencia Mental			
¿Cuánta actividad mental y perceptiva fue necesaria? ¿Es una tarea difícil o fácil, simple o compleja, pesada o ligera?			
<table border="1" style="width: 100%; height: 20px;"> <tr> <td style="width: 50%;"></td> <td style="width: 50%;"></td> </tr> </table>			
Baja	Alta		
Exigencia Física			
¿Cuánta actividad física fue necesaria? ¿Se trata de una tarea difícil o fácil, lenta o rápida, relajada o cansada?			
<table border="1" style="width: 100%; height: 20px;"> <tr> <td style="width: 50%;"></td> <td style="width: 50%;"></td> </tr> </table>			
Baja	Alta		
Exigencia Temporal			
¿Cuánta presión de tiempo sintió debido al ritmo al cual sucedían las tareas o elementos de las tareas? ¿Era el ritmo lento y pausado, o rápido y frenético?			
<table border="1" style="width: 100%; height: 20px;"> <tr> <td style="width: 50%;"></td> <td style="width: 50%;"></td> </tr> </table>			
Baja	Alta		
Rendimiento			
¿Hasta qué punto cree que ha tenido éxito en los objetivos establecidos por el investigador (o por Ud. mismo)? ¿Cuál es su grado de satisfacción con el nivel de ejecución?			
<table border="1" style="width: 100%; height: 20px;"> <tr> <td style="width: 50%;"></td> <td style="width: 50%;"></td> </tr> </table>			
Bueno	Malo		
Esfuerzo			
¿En qué medida ha tenido que trabajar (física o mentalmente) para alcanzar su nivel de resultados?			
<table border="1" style="width: 100%; height: 20px;"> <tr> <td style="width: 50%;"></td> <td style="width: 50%;"></td> </tr> </table>			
Bajo	Alto		
Nivel de Frustración			
Durante la tarea, ¿en qué medida se ha sentido inseguro, desalentado, irritado, tenso o preocupado?			
<table border="1" style="width: 100%; height: 20px;"> <tr> <td style="width: 50%;"></td> <td style="width: 50%;"></td> </tr> </table>			
Bajo	Alto		

Figura 6 Tabla de puntuación para evaluación de las seis dimensiones.

3. Resultados

La prueba se aplicó a una muestra de once usuarios, cuatro mujeres y siete hombres, con edades que van desde los 20 a los 40 años de edad. De ellos, tres son profesores-investigadores de la Universidad del Istmo, cuatro son estudiantes de la Maestría en Ciencias en Energía Eólica, y cuatro son estudiantes de licenciatura y posibles aspirantes a ingresar a dicha maestría. Todos ellos son integrantes y colaboradores del proyecto P08 "Diseño y construcción de un aerogenerador experimental con capacidad menor que 5 kW y desarrollo de software de simulación en realidad virtual, con fines didácticos" del CEMIE-Eólico [CEMIE, 2016]. La elección de estos usuarios se debe a que en un futuro el sistema será utilizado como material didáctico en asignaturas como Introducción a la tecnología de los aerogeneradores y Seminario de tecnología de aerogeneradores [UNISTMO, 2017].

La repetitividad, es decir, el número de fallos al realizar un gesto correctamente se muestra en la figura 7. Se puede observar que, a excepción del gesto para el paneo, todos los gestos relacionados con el dispositivo Kinect requirieron un menor número de intentos para ser reconocidos. Este resultado puede atribuirse a que los gestos asociados al Kinect son claramente diferenciables por el usuario y en consecuencia por el sistema de RVI, mientras que los gestos asociados al Myo dependen en gran medida de la habilidad del usuario para controlar el dispositivo, ya que en diversas ocasiones los usuarios generaban el gesto correcto e inmediatamente después otro gesto incorrecto haciendo que el sistema de RVI procesara el último gesto.

La figura 8 presenta las medias del índice ponderado global de la carga mental al emplear cada uno de los dispositivos de interacción (Kinect y Myo). En dicha figura se puede observar que la carga mental en promedio (considerando las tres categorías de usuarios) es ligeramente mayor cuando se utiliza el dispositivo Kinect que cuando se utiliza el dispositivo Myo. Algo interesante de observar es que conforme aumenta el nivel de formación académica, disminuye la percepción de la carga mental sin importar del dispositivo de interacción natural de usuario que se utilice. Evidentemente el nivel de formación está relacionado con la edad, y

en esta prueba se observó que a mayor edad del usuario, más atención prestó en las indicaciones previas y durante el uso del sistema de RVI, lo cual parece influir significativamente con la percepción de carga mental.

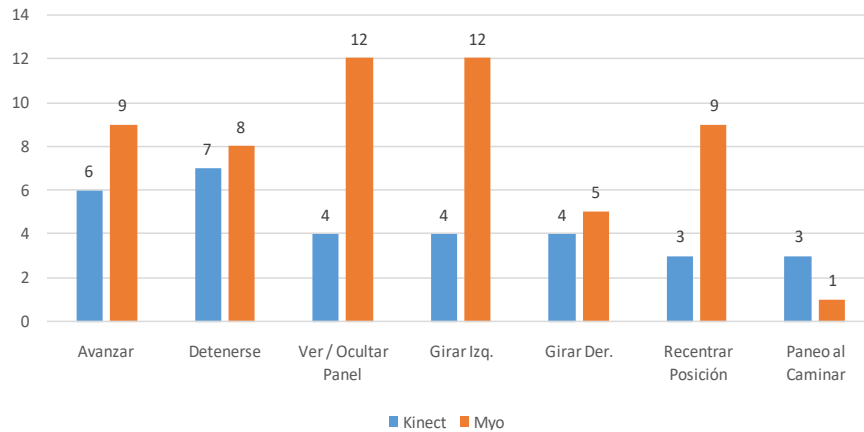


Figura 7 Repetitividad (fallos) al realizar los gestos.

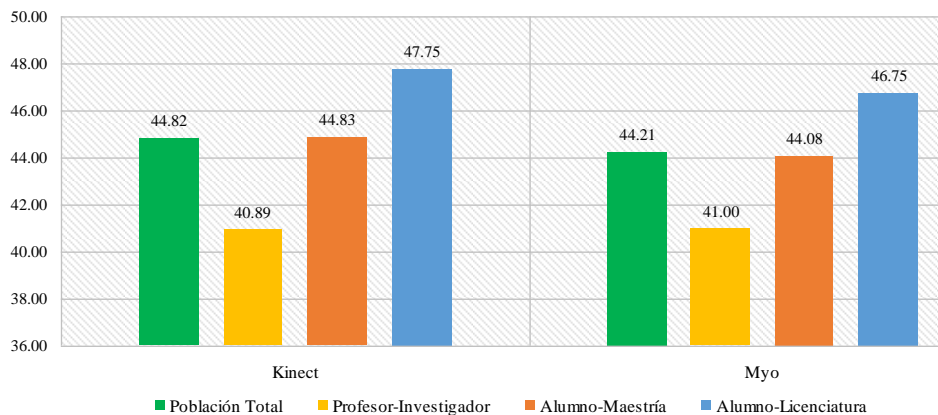


Figura 8 Resultados del índice ponderado global de la carga mental.

La figura 9 presenta más a detalle las medias en cada una de las dimensiones de la carga mental considerando la totalidad de los usuarios. Es claro que la exigencia mental es la dimensión que presenta una mayor carga mental, seguida de la dimensión del esfuerzo, sin importar el tipo de dispositivo. Estos resultados son comprensibles si se toma en consideración que los usuarios probaron por primera vez el sistema de RVI, por tal motivo necesitaban recordar cuáles eran los gestos y qué acción se asociaba a cada uno. Además, en la segunda tarea tenían que buscar las escaleras y tomar decisiones sobre qué gestos utilizar y en qué

instante ejecutarlos para lograr el objetivo de introducirse a la torre. Por otro lado, las dimensiones de exigencia física y rendimiento resultaron con una menor carga mental, por lo tanto, se puede inferir que ambos dispositivos de interacción natural de usuario facilitaron el desempeño de los usuarios para lograr las tareas encomendadas. En este sentido, el dispositivo Kinect presentó ligeramente una menor carga mental que el dispositivo Myo, esto es consistente con los resultados de la repetitividad.

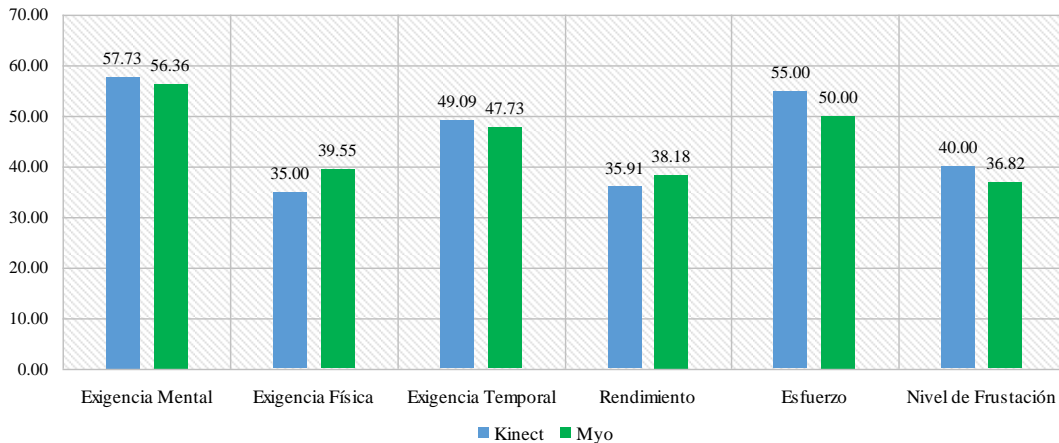


Figura 9 Resultados por cada dimensión de la prueba NASA-TLX.

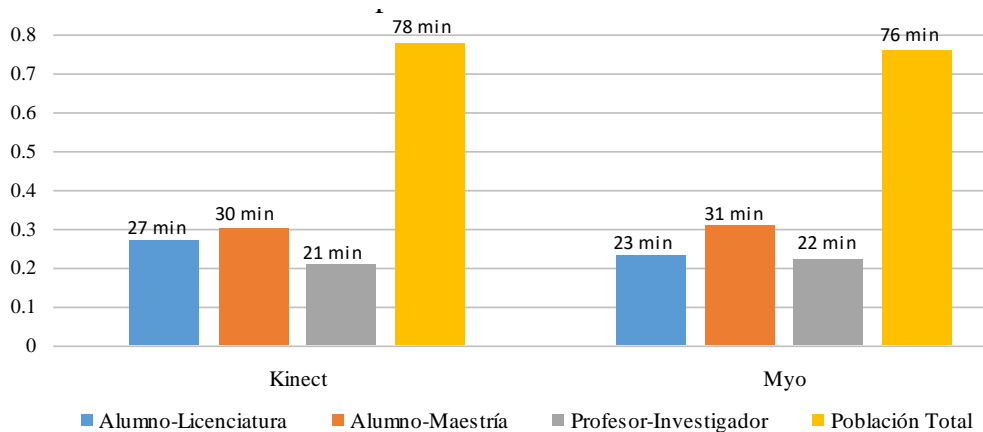


Figura 10 Tiempos promedio de los usuarios para realizar las tareas.

Por otro lado, si se observan los resultados de la dimensión del nivel de frustración de la figura 9 parecerían no coincidir si se contrastan con los resultados de la repetitividad. Sin embargo, cobran sentido si se considera el tiempo promedio

invertido por los usuarios para realizar los gestos, ver figura 10, ya que estos requirieron ligeramente una menor cantidad de tiempo para realizar los asociados al Myo, a pesar de que fallaron en mayor frecuencia para que los gestos se reconocieran correctamente. Esto indica que es más rápido para el usuario realizar un gesto con el Myo que con el Kinect.

4. Discusión

El propósito del sistema de RVI sobre el que se realizaron las pruebas es de tipo didáctico, orientado a la capacitación de personas para realizar la tarea de inspección de aerogeneradores de forma segura. Por tal motivo se utiliza el casco de RVI Oculus Rift, tal como lo sugiere la investigación realizada por [Freina, 2015]. Además, de acuerdo con [Flores, 2014], el uso de la RV como recurso didáctico y de capacitación posee un potencial enorme al proveer de un medio seguro de aprendizaje los estudiantes, sobre todo cuando lo que se debe aprender representa un riesgo físico. De manera específica, se puede mencionar que el estar físicamente cerca o dentro de un aerogenerador conlleva el riesgo de sufrir un accidente mortal [APREAN, 2007], [Hernández, 2015].

Por otro lado, cuando se utiliza un sistema de RVI el usuario tiene colocado sobre su cabeza un casco que le impide ver el entorno físico que lo rodea, lo que hace difícil la interacción a través de dispositivos convencionales, aquí es donde los dispositivos de interacción natural de usuario a través de gestos se convierten en la solución ideal a este problema. Sin embargo, pueden surgir algunas cuestiones respecto a ¿cuál dispositivo usar de los mencionados en los trabajos relacionados, el Kinect, el Myo, o el Leap Motion?, ¿cuáles son las ventajas y desventajas de elegir uno de estos? En el caso particular de este trabajo solo se contaba con dos de ellos (Kinect y Myo), por tal motivo el estudio se centró en tratar de responder, a través de la repetitividad de gestos y la prueba NASA-TLX, a las dos preguntas anteriores. Dicha prueba se eligió tomando como referencia, entre otras publicaciones relacionadas, a [Barrera, Díaz, Busto, Romero & Domínguez, 2014], ya que abordan la valoración de un entorno de RV e interacción humano-robot empleando la prueba NASA-TLX para evaluarlo.

Además, los trabajos encontrados que realizan algún análisis comparativo de los dispositivos de interés, por ejemplo [Vokorokos, 2016] y [Sánchez, 2017] no emplean sistemas de RVI, ni tampoco están enfocados al sector eólico. Es por ello que resultan de utilidad los resultados aquí obtenidos. En este sentido, con base en los tiempos promedios requeridos para realizar las tareas y los resultados la valoración de la prueba NASA-TLX, en primera instancia pareciera que cualquiera de los dos dispositivos representa una buena elección, ya que no existe una gran diferencia significativa entre ambos. Sin embargo, el Myo dio ligeramente mejores resultados en cuanto a tiempo para terminar las tareas y carga mental se refiere, no así en los resultados de la repetitividad. Aunque este último resultado puede mejorar si los usuarios utilizan el dispositivo Myo con mayor frecuencia.

También existen ciertas ventajas y desventajas que son inherentes a la propia tecnología de cada dispositivo. Por un lado, una de principales ventajas observadas al usar el Kinect es que se adapta con mayor facilidad a la fisonomía del usuario, tan solo habrá que colocarse más cerca o más lejos del dispositivo para un reconocimiento de gestos apropiado, mientras que con el Myo el brazo del usuario no puede ser muy delgado o muy grueso, lo que dificulta portarlo de forma correcta para un reconocimiento certero. Por otro lado el dispositivo Myo ofrece la ventaja de no ser sensible a la luz ambiental, ni a la presencia de otros usuarios cerca del dispositivo, mientras que para usar el dispositivo Kinect la iluminación debe estar controlada y no sufrir variaciones, además de que solo el usuario que va a emplear el sistema de RVI debe estar frente al dispositivo, de lo contrario podrían reconocerse los gestos de las personas que no están portando el casco Oculus Rift.

Cabe mencionar que en el estudio realizado por [Sánchez, 2017], también resulta mejor valorado el uso del dispositivo Myo respecto a los dispositivos Kinect y Leap Motion, aunque el Myo no se empleo solo, este fue utilizado conjuntamente con comandos de voz. Mientras que en el trabajo realizado [Vokorokos, 2016], sus resultados dejan mejor posicionado al Leap Motion, pero el Myo sale mejor posicionado con respecto al Kinect. No obstante, concluyen que la elección del

dispositivo idóneo está en función de la aplicación final y de la valoración subjetiva de los usuarios.

5. Conclusiones

En este trabajo se presentó un estudio comparativo entre los dispositivos de interacción natural de usuario Kinect y Myo usados en un entorno de realidad virtual inmersiva enfocado a la inspección de aerogeneradores. Aplicando para tal fin la prueba NASA-TLX en sus dos etapas. Los resultados, en concordancia con algunos trabajos relacionados, permiten corroborar la potencial aplicación de los dispositivos estudiados y su uso en función de la aplicación final, aunado las principales ventajas y desventajas inherentes a la propia tecnología de cada dispositivo.

No obstante, el dispositivo Myo aventaja ligeramente al Kinect en cuatro de las seis dimensiones de la carga mental, donde dos son atribuidas a las exigencias impuestas a los usuarios: exigencia mental y exigencia temporal; y dos son atribuidas a la interacción de los usuarios con la tarea: esfuerzo y nivel de frustración [Barrera, 2014]. En tanto que el dispositivo Kinect solo aventaja ligeramente al Myo en dos de las seis dimensiones: exigencia y rendimiento; una atribuida a las exigencias y otra a la interacción, respectivamente. Una ventaja más del Myo ante el Kinect que quedó de manifiesto en el conteo de repetitividad es el uso del giroscopio integrado en el Myo, el cual es de gran ayuda en la navegación del avatar al momento de caminar. Sin embargo, una desventaja que se presenta es controlar el dispositivo Myo para el resto de los gestos, la cual puede ser superada de manera gradual si el usuario utiliza frecuentemente el dispositivo.

Es importante señalar que ambos sistemas pueden mejorarse por separado, o incluso unir lo mejor de ambos. Una mejora que se sugiere en ambos sistemas es la creación y reconocimiento de nuevos gestos más intuitivos y fáciles de recordar para los usuarios. Así mismo, para el sistema del Kinect se propone analizar la altura y la masa del usuario permitiendo que el avatar se ajuste a sus características físicas. En lo que respecta al sistema del Myo se propone utilizar

dos brazaletes en lugar de solo uno, esto permitiría controlar más acciones tanto del avatar como del ambiente que lo rodea. Todas estas mejoras se sugieren como trabajos a futuro.

Además, cabe mencionar que la metodología planteada en este trabajo permite esbozar futuras comparaciones con otros dispositivos como el Leap Motion y los mandos inalámbricos que acompañan a algunos cascos de RVI. No obstante, y a pesar de que el Test de índice de carga ha sido probado por más de veinte años y citado en más de 4,400 estudios [NASA-TLX, 2011], contrastar los resultados de dicho Test con otros estudios de usabilidad contribuiría a obtener resultados más contundentes. Sin embargo, la comparativa con otros métodos está fuera del alcance de este artículo.

6. Bibliografía y Referencias

- [1] ACCIONA, Experiencias Inmersivas, 2015: <https://www.accion.com/es/salaprensa/afondo/2015/diciembre/experiencias-inmersivas/>.
- [2] ACCIONA E. ACCIONA, Energía presenta una experiencia de realidad virtual inmersiva en la feria eólica más importante de Norteamérica. ACCIONA, 2016: <http://www.accion.com/es/noticias/accion-presenta-experiencia-realidad-virtual-inmersiva-feria-eolica-importante-norteamerica/>.
- [3] APREAN, Guía de Buenas Prácticas Preventivas en el Sector de la Energía Eólica. Fundación para la Prevención de Riesgos Laborales. Sevilla, España 2007.
- [4] Barrera, G., Díaz, L., Busto, J., Romero, L., & Domínguez, O., Realización de una evaluación de un sistema de interacción físico Hombre - robot con base en el protocolo nasa tlx. Educación y Salud Boletín Científico de Ciencias de la Salud del ICESA. Vol. 3, No. 5, 2014: <https://repository.uaeh.edu.mx/revistas/index.php/ICESA/article/view/792/791>.
- [5] CEMIE-Eólico. P08, Diseño y construcción de un aerogenerador experimental con capacidad menor que 5 kW y desarrollo de software de simulación en realidad virtual, con fines didácticos. CEMIE-Eólico, 2017: <http://cemiee.iie.org.mx/Proyectos/Proyecto-P08>.

- [6] Denning, P. J., & Malone, T. W., Coordination. *Interactive Computation: The New Paradigm*. Springer. 415-439. Berlín, Alemania, 2006. Doi:10.1007/3-540-34874-3_16
- [7] Flores, J. A., Camarena P., & Avalos, E., La realidad virtual, una tecnología innovadora aplicable al proceso de enseñanza de los estudiantes de ingeniería. *Revista Apertura*. Vol. 6, No. 2, 2014.
- [8] Freina, L. & Ott, M., A Literature Review on Immersive Virtual Reality in Education: State Of The Art and Perspectives. eLSE Conference, Bucharest, April 2015.
- [9] Han, I., Ryu, J., & Kim, M., Prototyping Training Program in Immersive Virtual Learning Environment with Head Mounted Displays and Touchless Interfaces for Hearing-Impaired Learners. *Educational Technology International*. Vol. 18, No. 1, pp. 49-71, 2017.
- [10] Hernández, M., Hernández, M., Arellano, J., & Toledo, G., Videojuego didáctico empleando el Kinect para la manipulación virtual de un aerogenerador. *Pistas Educativas*. No. 112, 2015.
- [11] Hilfert, T., & König, M. Low-cost virtual reality environment for engineering and construction, *Visualization in Engineering*. 4:2, 2016. Doi: 10.1186/s40327-015-0031-5.
- [12] Holmes, D., Charles, D., Morrow, P., McClean, S., & McDonough, S. Usability and performance of Leap Motion and Oculus Rift for upper arm virtual reality stroke rehabilitation. 11th International Conference on Disability, Virtual Reality & Associated Technologies. Los Angeles, California, USA, 2016.
- [13] Jiménez, B., Morabito, D., & Remondino, F. Access to complex reality-based 3D models using virtual reality solutions. *Journal of Cultural Heritage*, 2016. Doi: 10.1016/j.culher.2016.09.003.
- [14] Montuschi, P., Sanna, A., Lamberti, F., & Paravati, G. Human-Computer Interaction: Present and Future Trends. *Computing Now*, 7(9), online, 2014.
- [15] Pressman, R., *Ingeniería del Software: un enfoque práctico*. Séptima Edición. Mc Graw-Hill. México, D.F, 2010.

- [16] NASA-TLX. Online NASA-TLX Beta, 2011: <http://www.nasatlx.com>.
- [17] Rubio, S., Díaz, E., Martín, J., & Luceño, L. Carga mental en vigilantes de seguridad. Diferencias por sexo y capacidad atencional. *EduPsykhé. Revista de psicología y educación*. Vol. 7, No. 2, pp. 213-230, 2008.
- [18] Sánchez, F., Sánchez, J., Moyano, J., Pérez, E., & Maestre, J., Use of natural user interfaces for image navigation during laparoscopic surgery: initial experience. *Minimally Invasive Therapy & Allied Technologies*, 2017. Doi: 10.1080/13645706.2017.1304964.
- [19] Steed, A., Pan, Y., Zisch, F., & Steptoe, W. The Impact of a Self-Avatar on Cognitive Load in Immersive Virtual Reality. *IEEE Virtual Reality Conference 2016*, 67-76, 2016.
- [20] Trujillo, K., Toledo, G., Arellano, J., & Hernández, M., Virtualización tridimensional interactiva de un parque eólico con fines didácticos. *Pistas Educativas*. No. 112, 2015.
- [21] UNISTMO, Maestría en Ciencias en Energía Eólica. Universidad del Istmo, 2017. Url: http://www.unistmo.edu.mx/m_eolica.html
- [22] Valkov, D., Martens, J. & Hinrichs, K., Evaluation of the Effect of a Virtual Avatar's Representation on Distance Perception in Immersive Virtual Environments. *IEEE Virtual Reality Conference 2016*, pp. 305-306, 2016.
- [23] Vokorokos, L., Mihal'ov, J., & Chovancová, E., Motion Sensors: Gesticulation Efficiency Across Multiple Platforms. *20th Jubilee IEEE International Conference on Intelligent Engineering Systems*. Budapest, Hungría, 2016.

ANÁLISIS DE ATAQUES DE RED DEL TIPO DHCP SPOOFING, TCP SYN FLOOD Y PAQUETES MALFORMADOS

Josué Cirilo Cruz

Universidad Autónoma Metropolitana, Azcapotzalco
Ingeniero.josuecc@gmail.com

Arturo Zúñiga López

Universidad Autónoma Metropolitana, Azcapotzalco
azl@azc.uam.mx

Carlos Avilés Cruz

Universidad Autónoma Metropolitana, Azcapotzalco
caviles@azc.uam.mx

Juan Villegas Cortez

Universidad Autónoma Metropolitana, Azcapotzalco
juanvc@azc.uam.mx

Resumen

Hoy en día las compañías, empresas e instituciones almacenan su información en bases de datos que están en alguno de los servidores de su red, y han tenido que abrir el acceso a dicha información para que los usuarios puedan conectarse a ella desde su intranet, esto las hace vulnerables a los ataques de los intrusos. Para detectar estas amenazas, es necesario conocer cómo funcionan, y encontrar patrones característicos que son implementados en tablas de aprendizaje de dispositivos, tales como firewalls, routers, etc., siendo éstas deducidas con base en el análisis del comportamiento del tráfico en la red, teniendo así la conformación de un patrón característico que identifica a la intrusión. En este artículo, se analiza el tráfico circulante en una intranet, con el objetivo de

caracterizar y formar un patrón de rasgos para cada uno de los ataques del tipo DHCP spoofing, TCP SYN flood y de paquetes malformados.

Palabras Claves: Ataques de red, DHCP spoofing, paquetes malformados, seguridad en redes, TCP SYN flood.

Abstract

Nowadays, companies and business offices store their information in databases all over on the servers of their computer networks, and they have had to open the access to this information, so users connected from their intranet, are vulnerable to the attacks of intruders. In order to detect these threats, it is necessary to know how they work, and find the characteristic patterns which are implemented in networking devices, which learn data base tables such as firewalls, routers, etc; the patterns are conformed based on the analysis of the communication traffic behavior. In this article, we analyze the traffic over an intranet in order to of characterize and conform patterns for each DHCP spoofing, TCP SYN flood, and tools which generate simulated attacks using malformed packets.

Keywords: DHCP spoofing, malformed packet, network attacks, network security, TCP SYN flood.

1. Introducción

En una red de cómputo local bajo el protocolo TCP/IP, se tiene que durante el intercambio de información generado por un equipo fuente (pc-usuario), y un equipo destino (servidor), como se aprecia en la figura 1, ésta es codificada para evitar que personas ajenas a la comunicación tengan acceso a la información o se deniegue el acceso a la misma. Con la introducción de las computadoras y servidores, se hizo evidente la necesidad de disponer de herramientas automatizadas para la protección de los archivos de información almacenadas en estos, la disponibilidad de los servicios ofrecidos por los servidores o la seguridad para realizar alguna actividad entre otras, esto añade un concepto referido en términos de seguridad de redes [Stallings, 2004].



Figura 1 Modelo simplificado de una comunicación entre computadoras.

La seguridad en redes de computadoras, se refiere a cualquier actividad diseñada para proteger la integridad de una red, manteniendo el intercambio de información, libre de riesgos y proteger los recursos informáticos de compañías, empresas o escuelas [Cisco, 2017], es por ello que cuando se habla de seguridad en redes se consideran como riesgos los ataques de códigos maliciosos, personas no autorizadas (hackers), denegación de servicios y amenazas combinadas. Un ataque se define como *una secuencia de operaciones que ponen en riesgo la seguridad de un sistema*, y por otro lado una anomalía o amenaza es *una actividad sospechosa desde la perspectiva de la seguridad* [García, 2009]. El modelo simplificado de un ataque a una red se muestra en la figura 2, donde se muestra la comunicación entre dos computadoras y un atacante que realiza operaciones que comprometen la comunicación; los ataques más comunes son enfocados a: la conectividad, la denegación de servicios, el consumo de ancho de banda, etc. Dichos ataques pueden ser mitigados conociendo su funcionamiento, formando así un patrón característico. Este trabajo de investigación se centra en el análisis del tráfico de una intranet en un ambiente simulado para ataques del tipo DHCP spoofing, TCP SYN flood, y paquetes malformados, para conformar patrones característicos que se podrán utilizar en tablas de aprendizaje de equipos de seguridad en redes. En la siguiente sección se explican los detalles de estos conceptos y su finalidad.

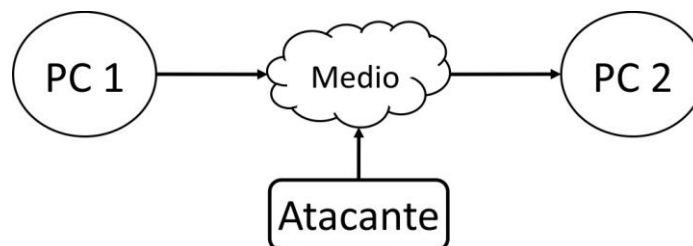


Figura 2 Modelo simplificado de un ataque de red.

DHCP Spoofing

El protocolo DHCP (Dynamic Host Configuration Protocol) es un componente integral para la funcionalidad del protocolo de internet (IP) de las redes actuales. Su función es configurar automáticamente equipos clientes con direcciones IP y algunos otros parámetros relevantes para la red e.g. la máscara de red, la puerta de enlace (Gateway), o los servidores DNS (Domain Name System) [Mukthar, 2012].

Un ataque del tipo DHCP spoofing consiste en capturar mensajes del tipo **DHCPDISCOVER**, esto se logra instalando un servidor falso de DHCP o con un software que emula las mismas funciones, de tal manera que conteste a las peticiones DHCPDISCOVER de los clientes, ver figura 3, dándole parámetros de configuración de tal forma que usurpa funciones, e. g. puede cambiar la dirección de la puerta de enlace (gateway), dando la dirección de él mismo, y realizar un ataque mejor conocido como *Man in the middle*, en el cual un atacante puede leer, insertar y modificar mensajes entre dos usuarios o sistemas [Symantec, 2017].

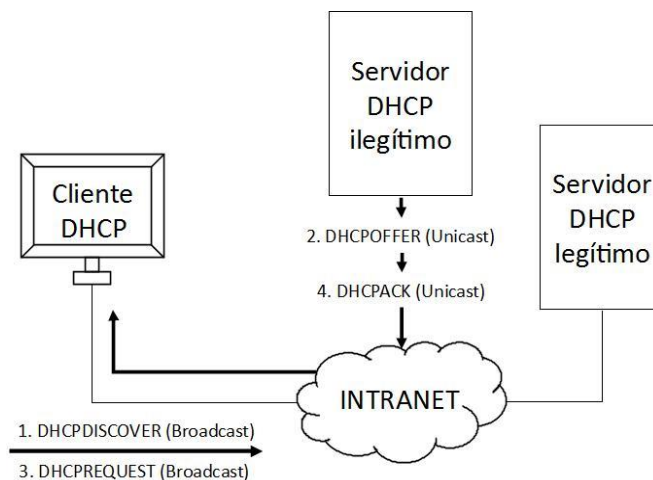


Figura 3 Esquema general del ataque DHCP Spoofing.

TCP SYN flood

El protocolo de control de transporte (TCP) especifica el formato de los datos y de los acuses de recibo usados en la transferencia de datos. TCP, es un protocolo orientado a conexión dado que los participantes en la comunicación deben

establecer una conexión previa, antes de que los datos puedan ser transferidos, realizando el control de flujo, corrección de errores, garantías TCP confiables y la entrega secuencial de los paquetes. Se considera un protocolo confiable porque si se corrompe o se pierde un paquete, TCP pedirá uno nuevo y correcto, hasta recibirlo [Cisco, 2017].

Un ataque TCP SYN flood, es llamado *flood* (Inundación, en español) porque afecta al ancho de banda que es necesario en una comunicación donde fluyen grandes cantidades de paquetes a frecuencias y tamaños significantes, de esta forma saturan las tarjetas de red al grado de detener su funcionamiento. Los ataques TCP SYN flood son diseñados para tomar ventaja de la metodología utilizada por una nueva conexión TCP (ver figura 4). De esta manera el atacante genera falsos paquetes que pretenden establecer una nueva conexión válida (SYN). Estos paquetes son recibidos por el servidor, el cual intenta responder (SYN-ACK), pero nunca es completada una conexión satisfactoriamente, dado que nunca recibe un mensaje de confirmación con la bandera ACK activa por parte del cliente. Esto hace que se agote el número máximo de clientes que el servidor puede atender, logrando la denegación de una nueva solicitud de conexión por parte de un nuevo cliente.

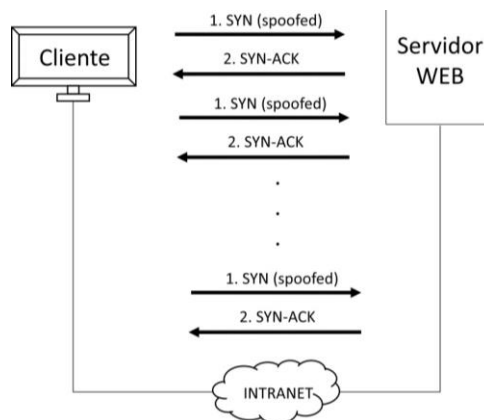


Figura 4 Esquema típico del ataque TCP SYN flood.

Paquetes Malformados

Un disector es un módulo encargado de decodificar convenientemente paquetes de una red, con características establecidas por los protocolos

[Wireshark, 2017], e. g. IP, TCP, DHCP, etc; al hablar sobre paquetes malformados, se refiere a que el disector elegido para fragmentar paquetes de algún protocolo no puede diseccionar correctamente el contenido de éstos [Wireshark, 2017]. Existen cuatro razones por las cuales no se puede diseccionar bien un paquete, estas son:

- **Disector está equivocado:** el sniffer erróneamente ha elegido el disector de protocolo incorrecto para este paquete [Wireshark, 2017].
- **El paquete no se puede rearmar:** El paquete excede los límites del tamaño del fragmento del paquete de red y no se puede reensamblar o rearmar [Wireshark, 2017].
- **El Paquete es incorrecto:** El paquete es realmente malo (tiene una malformación), lo que significa que una parte del paquete no es justo lo que esperábamos (no cumple con las especificaciones del protocolo) [Wireshark, 2017].
- **El disector tiene errores:** El disector del protocolo correspondiente, todavía está incompleto, es decir no es un disector correcto [Wireshark, 2017].

Generalmente un paquete malformado se debe a que es construido sin cumplir las reglas estipuladas por el protocolo en cuestión. Existen distintas formas de realizar ataques con paquetes malformados, comúnmente se suelen utilizar programas (software) que crean paquetes de distintos protocolos e inyectan éstos masivamente a equipos víctimas. Un ataque mediante Paquetes Malformados, es un ataque en el que el atacante puede utilizar múltiples equipos (zombis), a los que ordena enviar paquetes formados incorrectamente al sistema de la víctima con el fin de bloquearlo, e.g. en un ataque de direcciones IP, el paquete contiene las mismas direcciones IP de origen y de destino, esto puede confundir a los sistemas operativos de las víctimas y causar que se bloqueen. Otro ejemplo, es un ataque de opciones de paquetes IP, con ello se pueden asignar al azar los campos opcionales dentro de un paquete IP y establecer todos los bits de calidad de servicio en uno, para que el sistema de la víctima deba utilizar un tiempo de

procesamiento adicional para analizar el tráfico. Si este ataque se multiplica, puede agotar la capacidad de procesamiento de los sistemas de las víctimas [Spech, 2004].

2. Métodos

Para realizar la implementación y el análisis de los ataques propuestos en éste trabajo, se hizo uso de la topología de la intranet mostrada en la figura 5, en ella se simularon 3 subredes: LAN 1, LAN 2 y el resto de la INTRANET. Se implementaron tres servidores: DHCP, SYSLOG y NTP en la LAN 2; y en la LAN 1 se encuentra el equipo atacante, el equipo atacado y el sensor (analizador de protocolos), el cual ayudó a visualizar y detectar eventos y con ello analizar el tráfico circulante en la red. Para analizar el tráfico de la red, se apoyó de los eventos registrados en el servidor SYSLOG y el sensor de captura de tráfico. Por otra parte, para desarrollar la simulación se utilizó el simulador de redes: GNS3 [GNS3, 2017], para emular la topología de la red (routers, switches e integración de las máquinas virtuales), y el software de virtualización: VirtualBox [VirtualBox, 2017] para las emular las máquinas virtuales. Como herramientas de ataques de red se utilizó Ettercap [Ettercap, 2017], para generar el ataque DHCP spoofing, Dsniff [Dsniff, 2017], para el ataque con Paquetes Malformados, y Hping3 [Hping3, 2017], para el ataque TCP SYN flood.

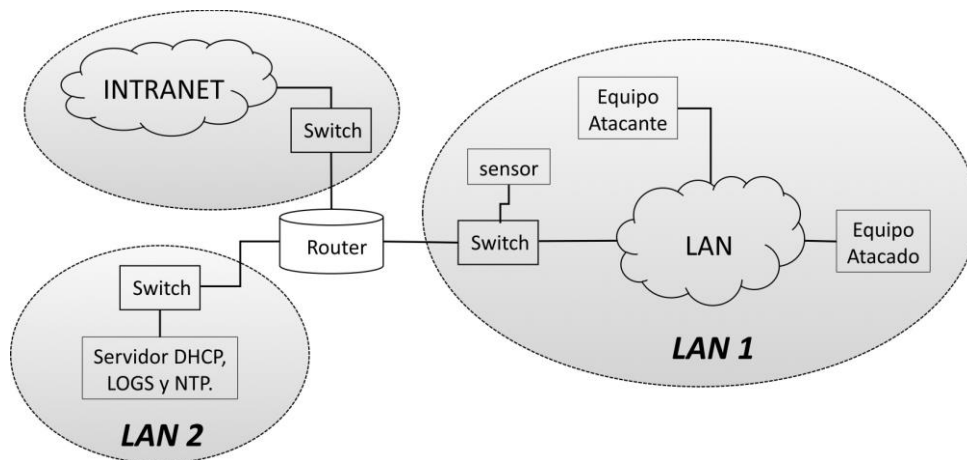


Figura 5 Topología de la intranet.

Implementación del DHCP spoofing

La condición inicial de la maqueta primera de pruebas es, cuando se realiza la asignación de una dirección IP válida por parte del servidor DHCP legítimo a la computadora que posteriormente será **atacada**. Para ello, si observamos el archivo de registros de salida del servidor SYSLOG, ver figura 6, primero se realiza la liberación de la dirección por parte del cliente (DHCPRELEASE), después éste envía un mensaje DHCPDISCOVER para volver a encontrar al servidor DHCP (con dirección IP: 192.168.2.2), posteriormente el servidor contesta con un mensaje DHCPOFFER ofreciendo la dirección IP: 192.168.1.10 al equipo con MAC-ADDRESS: 08:00:27:12:ed:14 (cliente), y enseguida el cliente contesta con un DHCPREQUEST (en modo unicast, que es contraria a su naturaleza, ver figura 3, aceptando la dirección ofrecida por el servidor, y finalmente el servidor asigna la dirección IP: 192.168.1.10 al equipo con la MAC-ADDRESS: 08:00:27:12:ed:14, mediante el mensaje DHCPACK. Por otro parte, en el cliente se ejecuta un analizador de protocolos, y su salida se muestra en la figura 7, en ella se observa que aparecen los mismos tipos de mensajes que el servidor DHCP genera, en la asignación de una dirección IP, por lo cual se deduce que; con sólo observar el archivo de registros del servidor SYSLOG se tiene la certeza de que se asignaron direcciones IP válidas.

```
Jun 1 14:11:46 ubuntu servidores dhcpd[2265]: DHCPRELEASE of 192.168.1.10 from 08:00:27:12:ed:14 (shadow_lite_sp3) via enp0s3 (found)
Jun 1 14:11:52 ubuntu servidores dhcpd[2265]: DHCPDISCOVER from 08:00:27:12:ed:14 via 192.168.1.1
Jun 1 14:11:53 ubuntu servidores dhcpd[2265]: DHCPOFFER on 192.168.1.10 to 08:00:27:12:ed:14 (shadow_lite_sp3) via 192.168.1.1
Jun 1 14:11:53 ubuntu servidores dhcpd[2265]: DHCPREQUEST for 192.168.1.10 (192.168.2.2) from 08:00:27:12:ed:14 (shadow_lite_sp3) via 192.168.1.1
Jun 1 14:11:53 ubuntu servidores dhcpd[2265]: DHCPACK on 192.168.1.10 to 08:00:27:12:ed:14 (shadow_lite_sp3) via 192.168.1.1
```

Figura 6 Asignación legítima de una dirección IP, por parte del servidor.

6	7.87324700	192.168.1.10	192.168.2.2	DHCP	342	DHCP Release	- Transaction ID 0xcc1c1cc1
8	13.6387900	0.0.0.0	255.255.255.255	DHCP	348	DHCP Discover	- Transaction ID 0x8c6c7a1e
10	14.6584870	192.168.1.1	192.168.1.10	DHCP	342	DHCP Offer	- Transaction ID 0x8c6c7a1e
11	14.6597810	0.0.0.0	255.255.255.255	DHCP	373	DHCP Request	- Transaction ID 0x8c6c7a1e
13	14.6894500	192.168.1.1	192.168.1.10	DHCP	342	DHCP ACK	- Transaction ID 0x8c6c7a1e

Figura 7 Asignación legítima de una dirección IP, en el cliente.

La configuración del servidor DHCP falso se muestra en la figura 8. El ataque se puede realizar en los siguientes casos:

- El equipo atacado realiza una renovación de los parámetros de la red.
- El equipo solicita una nueva dirección IP.
- El equipo se apaga o se reinicia.

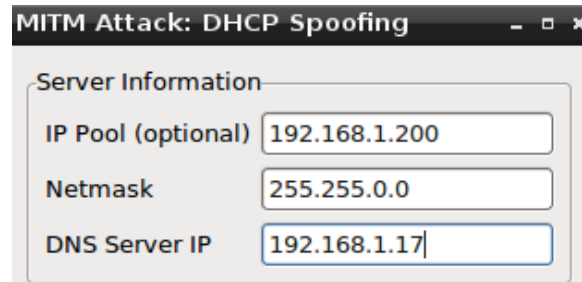


Figura 8 Configuración del falso servidor DHCP.

Es en éstas condiciones donde el ataque puede implementarse. Ahora, si se coloca un analizador de protocolos en la maquina atacante, se observan los mensajes que intercambia con una víctima, ver figura 9. En dicha figura se observa que el atacante responde a mensajes del tipo DHCPDISCOVER y asigna configuraciones similares a los que asigna el DHCP legítimo. De igual manera se observa la suplantación del servidor DHCP, poniéndose como una puerta de enlace válida para el equipo atacado. Para este ejemplo el equipo atacado solicita una nueva dirección IP al servidor DHCP legítimo, pero el falso servidor escucha el mensaje DHCPDISCOVER y responde primero con un DHCPOFFER, adelantándose al servidor legítimo, ofreciendo los parámetros de configuración realizados en la figura 8.

```
DHCP: [192.168.2.2] ACK : 192.168.1.17 255.255.255.0 GW 192.168.1.1 DNS 148.206.79.82
DHCP spoofing: using specified ip_pool, netmask 255.255.0.0, dns 192.168.1.17
Unified sniffing already started...
DHCP: [08:00:27:12:ED:14] DISCOVER
DHCP spoofing: fake OFFER [08:00:27:12:ED:14] offering 192.168.1.200
DHCP: [192.168.1.17] OFFER : 192.168.1.200 255.255.0.0 GW 192.168.1.17 DNS 192.168.1.17
DHCP: [08:00:27:12:ED:14] REQUEST 192.168.1.10
DHCP spoofing: fake ACK [08:00:27:12:ED:14] assigned to 192.168.1.10
```

Figura 9 Suplantación del servidor DHCP con Ettercap.

Por consiguiente, en la figura 10, se observa en el equipo cliente que el falso servidor DHCP logró su objetivo y generó un ataque *Man in the middle*, ya que

logró colocarse como puerta de enlace predeterminada y ahora se encuentra en medio de la comunicación entre la computadora cliente y la puerta de enlace.

```
C:\Documents and Settings\Administrador>ipconfig /renew
Configuración IP de Windows

Adaptador Ethernet Conexión de área local 3           :
    Sufijo de conexión específica DNS :
    Dirección IP. . . . . : 192.168.1.10
    Máscara de subred . . . . . : 255.255.0.0
    Puerta de enlace predeterminada : 192.168.1.17
```

Figura 10 Asignación de parámetros de red del servidor DHCP falso.

Implementación del TCP SYN flood

En la figura 11 se observa el cumplimiento del procedimiento del *Three-Way Handshake*. En dicha figura, el equipo cliente (con dirección IP: 192.168.1.10) solicita una nueva conexión al servidor (con dirección IP: 172.217.7.37), enviando un mensaje con la bandera SYN activada (ver el primer paquete marcado en color negro), posteriormente el servidor responde al cliente con un mensaje con la bandera SYN-ACK activada (ver el segundo paquete marcado en color negro), y finalmente el cliente finaliza el procedimiento enviando un mensaje con la bandera ACK activa al servidor (ver el tercer paquete marcado en color negro). Es importante mencionar que el procedimiento del *Three-Way Handshake*, no se efectúa de manera consecutiva forzosamente.

195	78.3942610	192.168.1.10	172.217.7.37	TCP	66	iascontrol-oms > http [SYN] Seq=0
197	78.4054740	192.168.1.10	107.167.110.211	TCP	66	iascontrol > http [SYN] Seq=0 win=
198	78.4090350	192.168.1.10	107.167.110.211	TCP	66	dbcontrol-oms > http [SYN] Seq=0 w
199	78.4156720	192.168.1.10	172.217.7.37	TCP	66	oracle-oms > http [SYN] Seq=0 win=
207	79.9896690	192.168.1.10	37.228.108.171	TCP	62	olsv > https [SYN] Seq=0 win=6424
212	81.3630880	192.168.1.10	172.217.7.37	TCP	66	iascontrol-oms > http [SYN] Seq=0
213	81.3631930	192.168.1.10	107.167.110.211	TCP	66	iascontrol > http [SYN] Seq=0 win=
214	81.3632450	192.168.1.10	107.167.110.211	TCP	66	dbcontrol-oms > http [SYN] Seq=0 w
215	81.3632930	192.168.1.10	172.217.7.37	TCP	66	oracle-oms > http [SYN] Seq=0 win=
216	81.4995140	172.217.7.37	192.168.1.10	TCP	66	http > iascontrol-oms [SYN, ACK] Seq=
217	81.4996360	192.168.1.10	172.217.7.37	TCP	54	iascontrol-oms > http [ACK] Seq=1

Figura 11 Three Way-Handshake (TCP).

El ataque inicia comúnmente, cuando el equipo atacante envía de manera indefinida, mensajes de tipo TCP con la bandera SYN activa al equipo víctima. En la figura 12, el atacante genera estos mensajes, inundando al servidor (equipo

victima con dirección IP: 192.168.1.10). Es importante mencionar que la generación de estos paquetes es de forma malformada, i.e. los paquetes son contruidos con direcciones IP falsas e inválidas, con direcciones IP que no están asignadas en equipos reales o que son direcciones IP del tipo multicast o de clase E (observar el último paquete, con dirección IP: 236.85.33.11 de la figura 13).

```
root@debianatacante:~# hping3 -I eth0 -p 80 --flood -S --rand-source 192.168.1.10
HPING 192.168.1.10 (eth0 192.168.1.10): S set, 40 headers + 0 data bytes
hping in flood mode, no replies will be shown
```

Figura 12 Implementación de ataque TCP SYN Flood.

```
3020 0.796583 115.224.191.90 -> 192.168.1.10 TCP 54 5347-80 [SYN] Seq=0 Win=512 Len=0
3021 0.796598 213.88.220.152 -> 192.168.1.10 TCP 54 5348-80 [SYN] Seq=0 Win=512 Len=0
3022 0.796614 183.151.26.152 -> 192.168.1.10 TCP 54 5349-80 [SYN] Seq=0 Win=512 Len=0
3023 0.796630 206.99.253.112 -> 192.168.1.10 TCP 54 5350-80 [SYN] Seq=0 Win=512 Len=0
3024 0.796645 149.161.234.208 -> 192.168.1.10 TCP 54 5351-80 [SYN] Seq=0 Win=512 Len=0
3025 0.796661 12.213.38.156 -> 192.168.1.10 TCP 54 5352-80 [SYN] Seq=0 Win=512 Len=0
3026 0.796676 41.145.31.182 -> 192.168.1.10 TCP 54 5353-80 [SYN] Seq=0 Win=512 Len=0
3027 0.796692 222.38.27.167 -> 192.168.1.10 TCP 54 5354-80 [SYN] Seq=0 Win=512 Len=0
3028 0.796707 236.85.33.11 -> 192.168.1.10 TCP 54 5355-80 [SYN] Seq=0 Win=512 Len=0
```

Figura 13 Mensajes TCP con la bandera SYN activa, generados por el atacante.

Implementación de los Paquetes Malformados

En la figura 14 se muestra la implementación de un ataque con paquetes malformados. Como se observa, se implementa una inundación de direcciones MAC al equipo con dirección IP: 192.168.1.10, por parte del equipo atacante con dirección IP 192.168.1.17. Cabe mencionar que este ataque es generado construyendo direcciones MAC falseadas, es decir: inválidas (direcciones MAC de tipo unicast y multicast), y no reales (que no están asignadas a tarjetas de red reales). En la figura 15, se observan estos paquetes clasificados como paquetes malformados, debido a que el analizador de protocolos no los puede diseccionar y las direcciones MAC de origen son inválidas.

```
root@debianatacante:~# macof -i eth0 -s 192.168.1.17 -d 192.168.1.10 -n 10
7f:b9:9b:4:40:44 31:24:f:9:d0:4e 192.168.1.17.61429 > 192.168.1.10.15669: S 2096545159:2096545159(0) win 512
73:4b:b6:1f:78:f8 56:9c:7e:62:2:6e 192.168.1.17.61923 > 192.168.1.10.26690: S 1579647407:1579647407(0) win 512
b4:3:c2:7b:b5:d1 42:7f:fa:2e:95:9 192.168.1.17.23151 > 192.168.1.10.2826: S 515529850:515529850(0) win 512
51:87:5c:74:55:42 79:5c:96:1d:91:94 192.168.1.17.15021 > 192.168.1.10.62491: S 464922677:464922677(0) win 512
64:b:f8:2c:2c:71 5d:88:3b:3f:51:a8 192.168.1.17.44641 > 192.168.1.10.33302: S 2144723376:2144723376(0) win 512
a7:20:e8:59:5a:e0 25:85:68:2:72:ec 192.168.1.17.51356 > 192.168.1.10.26468: S 1697697134:1697697134(0) win 512
46:bf:d3:79:db:60 e3:a6:c5:2a:89:13 192.168.1.17.27621 > 192.168.1.10.19678: S 1236564629:1236564629(0) win 512
5c:d4:d6:3a:59:3e 20:50:4:2c:df:c9 192.168.1.17.45684 > 192.168.1.10.49168: S 49299384:49299384(0) win 512
bd:c4:7f:6c:41:a 67:54:bc:3e:9e:5f 192.168.1.17.35927 > 192.168.1.10.29865: S 1388095085:1388095085(0) win 512
1d:b4:e6:20:b7:80 20:14:3a:5f:a1:9d 192.168.1.17.26651 > 192.168.1.10.9721: S 769800147:769800147(0) win 512
```

Figura 14 Implementación del ataque con paquetes malformados.

86	2.9569776	192.168.1.17	192.168.1.10	TCP	54 [Malformed Packet]
87	2.9573636	192.168.1.17	192.168.1.10	TCP	54 [Malformed Packet]
88	2.9577426	192.168.1.17	192.168.1.10	TCP	54 [Malformed Packet]
89	2.9581016	192.168.1.17	192.168.1.10	TCP	54 [Malformed Packet]

Frame 86: 54 bytes on wire (432 bits), 54 bytes captured (432 bits) on interface 0
Ethernet II, Src: 93:cc:24:15:43:af (93:cc:24:15:43:af), Dst: AvantecM_60:de:56 (00:12:bd:60:de:56)
Internet Protocol Version 4, Src: 192.168.1.17 (192.168.1.17), Dst: 192.168.1.10 (192.168.1.10)
[Malformed Packet: TCP]
[Expert Info (Error/Malformed): Malformed Packet (Exception occurred)]
[Malformed Packet (Exception occurred)]
[Severity level: Error]
[Group: Malformed]

Figura 15 Vista del ataque con paquetes malformados, desde un analizador de protocolos.

3. Resultados

Del ataque DHCP spoofing implementado, se observó que comúnmente para detectar este ataque basta con verificar que la direcciones IP fuente de los mensajes tipo DHCPOFFER y DHCPACK sean del servidor DHCP legítimo. Además, por cuestiones de la topología de la red, ver figura 5, al poner el servidor DHCP en una red diferente en la que están los equipos clientes, e.g. LAN 1, se agrega implícitamente un mecanismo de seguridad ya que se cambia el modo de trabajo del mensaje DHCPREQUEST, ejecutándolo en modalidad unicast, esto se debe a que al colocar el servidor en una red diferente, se fuerza a que el mensaje DHCPREQUEST trabaje en modalidad unicast, debido a que el dispositivo router no propaga mensajes tipo broadcast, y para poder entregar este mensaje y llevarlo hacia el DHCP legítimo tendrá que forzarlo a trabajar en una comunicación tipo unicast en lugar de una en modalidad broadcast. Así también, se observó que al colocar el servidor DHCP dentro de la misma LAN donde se realizarán los ataques, este es más vulnerable, ya que se puede averiguar su dirección IP y suplantarlos.

Con la implementación de un servidor SYSLOG dentro del servidor DHCP, se dedujo que; se pueden predecir ataques del tipo DHCP “starvation” (Agotamiento de direcciones IP), esto se logra verificando el archivo de registros de eventos que se genera en el servidor con las relaciones de direcciones IP y MAC válidas o asignadas legítimamente, ver figura 16, por lo que podemos predecir y detectar cuando algún intruso se infiltró en la red y está intentando agotar nuestro rango de direcciones IP válidas, al revisar las direcciones MAC del mensaje

DHCPDISCOVER. De igual forma, también se obtiene una relación de direcciones MAC - IP asignadas y válidas, del archivo de registro de eventos.

```
12:32:32 ubuntu servidores dhcpd[2242]: DHCPDISCOVER from 08:00:27:12:ed:14 via 192.168.1.1
12:32:34 ubuntu servidores dhcpd[2242]: DHCPDISCOVER from 08:00:27:c6:de:db via 192.168.1.1
12:33:25 ubuntu servidores dhcpd[2242]: DHCPDISCOVER from 08:00:27:ff:4f:94 via 192.168.1.1
12:34:37 ubuntu servidores dhcpd[2242]: DHCPDISCOVER from 08:00:27:c6:de:db (debianatacante)
12:34:39 ubuntu servidores dhcpd[2242]: DHCPDISCOVER from 08:00:27:c6:de:db via 192.168.1.1
```

Figura 16 Verificación de mensajes DHCPDISCOVER en el servidor DHCP.

Por otra parte, de ataque TCP SYN flood se observó que puede mitigarse con la realización de filtros de paquetes malformados (direcciones IP falseadas y con mensajes TCP con la bandera SYN activa), utilizando un analizador de protocolos. Esto es importante ya que en una tabla de aprendizaje donde son caracterizados patrones intrusivos, entra como una característica a considerar para ser parte de un patrón característico de algún ataque no conocido. De igual manera, se observó en el equipo atacado que cuando está bajo un ataque TCP SYN flood, envía mensajes del protocolo ARP en broadcast intentando contestar los mensajes generados por los paquetes malformados. Por lo que lo hace un comportamiento característico para este ataque, visto del lado del atacado. Otra característica vista en el ataque TCP SYN flood es que: en el equipo atacado, la carga de la CPU, el uso de la memoria RAM y el desempeño de la tarjeta de red se incrementa, por lo que, al intentar ejecutar alguna otra aplicación, el equipo atacado se bloquea.

Por último, del ataque con paquetes malformados se observa que depende del equipo de enlace de datos que se tenga (*switch administrable* y *no administrable*), ya que al inundar con direcciones MAC aleatorias, si el equipo de red es administrable y tiene implementada alguna medida de seguridad, detecta las direcciones MAC inválidas y no las coloca dentro de su tabla CAM y no afecta al servidor. Pero si son direcciones MAC válidas, ver figura 17, son enviadas por el switch al servidor y agota el número máximo de clientes que se pueden atender por parte del servidor; de la misma manera trabajaría un *switch no administrable*. Dado todo lo anterior, se crean tres tablas, tablas 1, donde se describen los patrones característicos de los ataques analizados.

f8e0.092c.9c69	Dynamic	1	FastEthernet1/2
006c.b001.2d01	Dynamic	1	FastEthernet1/2
d0cb.3b69.0dfb	Dynamic	1	FastEthernet1/2
54d8.361b.b51b	Dynamic	1	FastEthernet1/2
4c37.8902.d6ea	Dynamic	1	FastEthernet1/2
96ad.9853.4463	Dynamic	1	FastEthernet1/2
78af.7455.288e	Dynamic	1	FastEthernet1/2
6880.8d0b.74b5	Dynamic	1	FastEthernet1/2

Figura 17 Tabla CAM bajo un ataque con paquetes malformados.

Tabla 1 Descripción de los ataques.

Tipo de ataque	Descripción	Forma de detección
DHCP Spoofing	Proviene de una dirección IP obtenida del servidor DHCP legítimo. Utiliza los mensajes DHCPREQUEST, con direcciones IP de tipo broadcast en el campo de destino. Modifica la dirección de la puerta de enlace en el equipo víctima.	Para la topología propuesta, colocando un sniffer, y verificando que los mensajes DHCPREQUEST sean del tipo unicast, no broadcast. Verificar que dirección IP fuente del mensaje DHCP OFFER Y DHCPACK sean del servidor DHCP legítimo.
DHCP Starvation	Agota el rango de direcciones válidas del servidor DHCP. Es generado por equipos que cambian su dirección MAC aleatoriamente (inválidas) y solicitan nuevas direcciones IP.	Se detecta instalando un servidor de logs en el servidor DHCP, y verificando que la dirección MAC del equipo solicitante sea válida. Limitando el rango de direcciones del conjunto válido. Asignación de direcciones IP estáticas.
Paquetes malformados	Es generado con direcciones MAC-ADDRESS e IP aleatorias inválidas e incorrectas (tipo unicast y multicast), caen dentro de la categoría de ser paquetes malformados, por los analizadores de protocolos.	Se verifica que provengan de direcciones MAC e IP válidas. Creación de filtros para verificar la legitimidad de las direcciones MAC e IP en los paquetes entrantes.
TCP SYN Flood	Tienen las banderas SYN o ACK activas dentro del mensaje TCP o están construidas sin tener las banderas TCP activas. Son generados con malformaciones: direcciones IP y MAC inválidas, aunque también pueden ser generados con direcciones IP y MAC válidas. Aumentan la carga del CPU y la tarjeta de red, y el uso de memoria RAM del equipo atacado.	Instalando un servidor de logs dentro del servidor, e. g. WEB, y realizando un aprendizaje supervisado acerca de los mensajes entrantes: con los mensajes TCP con las banderas ACK, SYN activas o sin banderas. Verificar que el Three-Way Handshake se cumpla. Con la creación de filtros de paquetes malformados: revisando que las direcciones IP sean válidas.

4. Discusión

Cuando se utiliza Ettercap como herramienta para implementar un falso servidor DHCP, en el momento de que un equipo solicita una nueva dirección IP, Ettercap averigua su dirección IP anterior, mediante la captura del mensaje DHCPREQUEST, y asigna esa misma dirección IP. Para mitigar este ataque se propone cambiar la manera de comunicación de un mensaje DHCPREQUEST, de tipo broadcast a unicast, esto se logra instalando el servidor DHCP en un segmento de red distinto, al de los equipos restantes de la intranet. Para mitigar variantes de ataques de DHCP, e.g. el DHCP starvation; se instala un servidor SYSLOG para observar, analizar y realizar un control sobre los mensajes DHCPDISCOVER registrados en el servidor DHCP, tomando lapsos de tiempo en minutos y segundos, e. g. mensajes DHCPDISCOVER registrados por minuto.

Un equipo al estar bajo un ataque TCP SYN flood tiene cuatro efectos principales, estos son: la memoria RAM se agota, la carga del CPU se incrementa drásticamente y la tarjeta de red se satura: al enviar mensajes ARP en Broadcast tratando de responder los mensajes recibidos y atendiendo a los nuevos paquetes. Es decir, haciendo que el procesador priorice el procesamiento de cada uno de los paquetes recibidos, a la par de necesitar más memoria RAM para procesarlas. Finalmente, para todos los ataques generados, si el atacante se colocara en otra parte de la intranet, el patrón es muy similar y puede detectarse su comportamiento, solo que, por cada red adicional de la intranet, se debe agregar un sensor para analizar el tráfico circulante de la red.

5. Conclusiones

En este trabajo hemos presentado una investigación sobre algunos de los ataques más frecuentes en una red tipo intranet, que puede comprometer la seguridad de la misma, tanto en la información que comparte como en la operación hacia los nodos de la red. Hemos presentado una metodología para conformar la plataforma de simulación de los ataques y hemos hecho uso de las herramientas más comunes para generar este tipo de ataques. Podemos finalmente emitir una mejor valoración de la seguridad y estado de funcionamiento

de la red interna, intranet y proporcionar una recomendación para una auditoría a bajo nivel, gracias a nuestra metodología. Más aún, en la creación de reglas de seguridad o medidas de implementación para poder mitigar este tipo de ataques. A futuro será necesario implementar estos patrones encontrados en tablas de aprendizaje, para poder robustecer y agregar nuevas características más relevantes, ya que estos ataques suelen volverse más complejos en su programación y más sofisticados en su implementación. Así también, podremos incluir otro tipo de ataques o variantes de éstos, todo para conocer mejor el estado de la seguridad de una intranet y mejorarla dinámicamente acorde a las nuevas amenazas para la seguridad de una red de cómputo.

6. Bibliografía y Referencias

- [1] Cisco, Configurar el protocolo TCP. http://www.cisco.com/cisco/web/support/LA/111/1116/1116270_iap-tcp.pdf, 25 de mayo de 2017.
- [2] Cisco, What Is Network Security? <http://www.cisco.com/c/en/us/products/security/what-is-network-security.html>, 08 de marzo de 2017.
- [3] Dsniff, <https://www.monkey.org/~dugsong/dsniff/>, 05 de junio de 2017.
- [4] Ettercap, <https://ettercap.github.io/ettercap/>, 05 de junio de 2017.
- [5] García P., Díaz J., Maciá G., Vázquez E., Anomaly-based network intrusion detection Techniques, systems and challenges, *Computers and Security* 28, Elsevier, pp. 18-28, 2009.
- [6] GNS3, <https://www.gns3.com/>, 05 de junio de 2017.
- [7] Hping, <http://www.hping.org/manpage.html>, 05 de junio de 2017.
- [8] Mathworks, Supervised Learning. <https://www.mathworks.com/discovery/supervised-learning.html>, 23 de mayo de 2017.
- [9] Montero G. Implementación de un NIDS en un sistema embebido para el análisis de tráfico de una red. Proyecto Tecnológico. UAM, CBI, Departamento de Sistemas, Ingeniería en Computación, 2014.
- [10] Mukhtar H., Salah K., Iraqui Y. Mitigation of DHCP starvation attack, *Computers and Electrical Engineering* 38, Elsevier, pp. 1115-1128, 2012.

- [11] Oracle, Establecimiento de una conexión TCP. <https://docs.oracle.com/cd/E19957-01/820-2981/ipov-36/index.html>, 24 de mayo de 2017.
- [12] Spech S., Lee R., Distributed Denial of Service: Taxonomies of Attacks, Tools and Countermeasures. Proceedings of the 17th International Conference on Parallel and Distributed Computing Systems, 2004 International Workshop on Security in Parallel and Distributed Systems, pp. 543-550, September 2004.
- [13] Stallings W., Fundamentos de Seguridad en Redes: Aplicaciones y Estándares, Pearson-Prentice Hall, Segunda Edición, 2004.
- [14] Symantec, Man-in-the-middle attack (ataque de tipo "Man in the middle"). https://www.symantec.com/es/mx/security/_response/glossary/define.jsp?Letter=m&word=man-in-the-middle-attack, 24 de mayo de 2017.
- [15] VirtualBox, <https://www.virtualbox.org/>, 05 de junio de 2017.
- [16] Wireshark, Malformed Packet. https://www.wireshark.org/docs/wsug_html_chunked/AppMessages.html, 24 de mayo de 2017.
- [17] Wireshark, Packet dissection. https://www.wireshark.org/docs/wsdg_html_chunked/ChapterDissection.html#ChDissectWorks, 30 de mayo de 2017.

DISEÑO DE UN DEMODULADOR DE FM MEDIANTE PLL PARA LA INTERROGACIÓN DE SENSORES INTERFEROMÉTRICOS DE FIBRA ÓPTICA

Jesús Lorenzo Cisneros Hernández

Universidad Autónoma de Ciudad Juárez
al160585@alumnos.uacj.mx

Alejandro Rodríguez Antonio

Universidad Autónoma de Ciudad Juárez
al98621@alumnos.uacj.mx

Abimael Jiménez Pérez

Universidad Autónoma de Ciudad Juárez
abimael.jimenez@uacj.mx

José Mireles Jr. García

Universidad de Guadalajara
jmireles@uacj.mx

Rafael E. González Landaeta

Universidad Autónoma de Ciudad Juárez
rafael.gonzalez@uacj.mx

Ángel Saucedá Carvajal

Universidad Autónoma de Ciudad Juárez
angel.sauceda@uacj.mx

Resumen

En este trabajo se diseñó y se construyó un sistema de interrogación de sensores interferométricos. El sistema está constituido por una etapa que emula la señal interferométrica típica de un sensor de este tipo: Primeramente, una etapa de acondicionamiento que convierte esta señal en una señal de FM convencional

y finalmente una etapa de demodulación de frecuencia; mediante el uso de la técnica de amarre de fase PLL, (del inglés: *Phase Lock Loop*). El proceso de demodulación, denominado en la literatura como “heterodino sintético”, utiliza un par de osciladores locales sintonizados a la frecuencia de la señal portadora y al doble de ésta. Así mismo, se requirieron una serie de filtros pasabanda tipo Butterworth de segundo orden para acotar el espectro de las señales de interés centrados en la frecuencia de la armónica necesaria para realizar el proceso de mezclado. Finalmente, la señal acondicionada se usó como entrada a un demodulador de FM mediante un PLL. Se consiguió recuperar señales del orden de miliradianes en el rango de 90 a 260 Hz. Se observó que este rango dependió del ancho de banda de los filtros pasabanda utilizados en el circuito. Se optó por esta técnica de demodulación basada en un PLL, pues logra la sintonización de una amplia gama de frecuencias, al ser también sintonizable el PLL a través de su VCO.

Palabras Claves: Demoduladores de FM, fase óptica, sensores interferométricos, PLL.

Abstract

In this work, an interrogation system of interferometric sensors was designed and constructed. The system consists of a stage emulating the interferometric signal typical of such sensor: First a conditioning stage that converts the above signal into a conventional FM signal and finally a frequency demodulation stage, based in the Phase Lock Loop technique o demodulate FM signals (PLL). The demodulation process used here, referred in the literature as "synthetic heterodyne", uses a pair of local oscillators, one tuned to the frequency of the carrier signal and the other one tuned at twice of the carrier frequency. It also requires a series of second-order Butterworth bandpass filters to limit the signals of interest and maintain a constant amplitude in the passband. As well as a trimmer to minimize the amplitude changes, in the final part of the conditioning stage. Finally, the conditioned signal was used as input to an FM demodulator via a PLL and signals of the order of miliradianes were achieved; with frequencies of modulating

signals in the range of 90 to 260 Hz. It was observed that this range depended on the bandwidth of the bandpass filters used in the circuit.

Keywords: *FM demodulators, Interferometric sensors, optical phase, PLL.*

1. Introducción

En el transcurso de las últimas décadas, ha ido en aumento el uso de sensores ópticos basados en fibra óptica debido a que son utilizados para medir variables físicas como distancias, desplazamientos, así como amplitud, frecuencia, intensidad y fase en movimientos oscilatorios. De estos resaltan los sensores interferométricos, que además permiten realizar mediciones de rugosidad en superficies, pueden medir desplazamientos horizontales o verticales y también encuentran amplia aplicabilidad en la medición de vibraciones con cierta frecuencia, entre otras.

Estos sensores basados en el fenómeno de interferencia óptica [Fang, 2015], operan como sigue: Una de las ondas que componen al sensor interferométrico, es expuesta a las características físicas del mesurando, y las variaciones de este modifican en fase a dicha onda, obteniendo una señal modulada en FM. Esta señal modulada interfiere con una onda de referencia estable sin modular, y la información de interés, es decir las variaciones presentes en el mesurando, se encuentra en la fase del patrón de intensidad resultante, y para extraer esta información se requiere de un demodulador de FM. Este trabajo se enfoca en el diseño, construcción y prueba de un demodulador de fase óptica con las características mencionadas anteriormente.

Antecedentes

Existen diversos métodos de demodulación de una señal modulada en fase óptica, uno de ellos, el método DCM (del inglés: *differential-and-cross-multiply*) que utiliza dos fotodetectores y una serie de operaciones matemáticas a través de amplificadores operacionales (OpAmp) configurados como diferenciador, derivador, multiplicador y una doble integración. Los resultados reportados con este método tienen una relación señal a ruido (SNR) de 80dB [Kumar, 2012], las

señales demoduladas no presentan distorsión y es factible tanto para señales débiles como fuertes.

Otro método de demodulación es el que usa un PLL (del inglés: *phase lock Loop*). Un PLL es un sistema de retroalimentación que comprende un comparador de fase, un filtro pasa bajas y un amplificador de error en la trayectoria de la señal hacia adelante y un VCO (*del inglés: voltage controlled oscillator*) en la trayectoria de la retroalimentación [Schlecker, 2013]. Cuando el PLL es enganchado a una señal de frecuencia modulada (FM), el VCO rastrea la frecuencia instantánea de la señal de entrada. La tensión de error filtrada fuerza al VCO a mantenerse unida con la señal de entrada, luego se convierte en la salida de FM demodulada [Silver, 2008]. La ventaja de este método estriba en la sintonización de una gama amplia de frecuencias al ser también sintonizable el PLL a través de su VCO [Udd, 2007]. En esta investigación se implementará el método de demodulación homodina, ya que es la que mejor resolución y sensibilidad ofrece. En este método se detectan las diferencias de fase existentes entre dos señales que se interfieren y se elimina la señal de desvanecimiento causada por las grandes derivas ambientales; esto se logra mediante la introducción de un gran desplazamiento de fase una frecuencia fuera de la banda de la señal [Dandridge, 2010]. Estas señales de gran amplitud llevan como bandas laterales las señales de interés.

Los resultados mostrados con la demodulación homodina han sido un alto rango dinámico, una buena linealidad y una sensibilidad de 10^{-6} radianes [Schlecker, 2013]. El esquema utilizado para este tipo de demodulación se muestra en la figura 1.

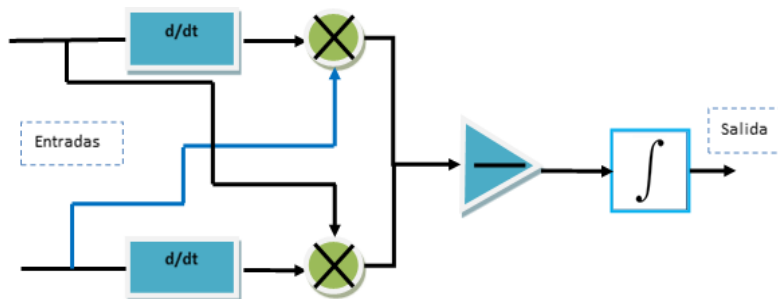


Figura 1 Esquemático de demodulador homodino.

2. Métodos

Para desarrollar el demodulador de FM, en primera instancia no se utilizó información procedente de algún mesurando, se decidió utilizar señales interferométricas controladas y reproducibles, tanto en amplitud como en frecuencia, para así poder verificar los resultados arrojados por el demodulador, y esto se logró implementando un emulador de interferogramas.

El método de demodulación seleccionado se simuló en el software de aplicación Cadence PSpice utilizando señales moduladoras de baja frecuencia, se observaron y se analizaron las formas de onda resultantes de cada etapa. Posteriormente se implementó el circuito electrónico del demodulador, y se verificó experimentalmente el rango de frecuencias y la amplitud mínima y máxima de la señal moduladora que el demodulador puede manejar, sin comprometer la calidad de la información obtenida:

- **Emulador de interferogramas.** Se implementó un emulador de interferogramas, debido a que durante el proceso de caracterización y prueba del sistema para extraer la señal de interés, se requerirían varias pruebas con interferogramas reproducibles y controlables en cuanto a la amplitud de la señal moduladora y de la portadora.

Este emulador se realizó con el generador de funciones trigonométricas AD639. La función de transferencia que sintetiza este IC, es la ecuación 1, en la que U representa la amplitud de la señal generada, y es la diferencia de voltaje entre la entrada $U_1 - U_2$, y estará en el rango de los 10mV a 10V. U_1 es polarizado con voltaje positivo, mientras que U_2 estará conectado a tierra. X_1 , X_2 , Y_1 , y Y_2 son las entradas diferenciales de los voltajes X y Y escaladas a $50^\circ/V$. LA

$$W = U \frac{\text{sen}(X_1 - X_2)}{\text{sen}(Y_1 - Y_2)} \quad (1)$$

Para generar el interferograma se introdujo una señal de 1 kHz con un amplitud de 5.8 Vpp como señal portadora (X_1), y una señal de 100 Hz y una amplitud de 200 mVpp como moduladora (X_2).

- **Selección del Método de Demodulación.** La elección del método de demodulación fue con base en las características que los diversos métodos presentaban. Se enfocó en cualidades como el ancho de banda en el cual el circuito opera y la posibilidad de que este fuera modificable para ser sintonizado en el rango de frecuencias de interés, además que la circuitería fuera simple sin dejar de un lado la eficiencia y la calidad de operación, por lo tanto, que se tuviera un control electrónico confiable.

En esta investigación se seleccionó el método de Detección Heterodina Sintética, porque como se mencionó antes, es el que mejor resolución y sensibilidad ofrece. Este método a partir de un solo interferograma genera las dos señales mediante un proceso de mezclado con osciladores trabajando a la frecuencia de la portadora y al doble de ésta. Posteriormente, mediante un filtraje pasa banda se selecciona la tercera armónica de la señal resultante del proceso de mezclado para después realizar la suma de las señales generadas en cada una de las ramas del detector; generándose así una señal resultante que después de aplicársele un proceso de eliminación de picos y un filtraje pasabanda adicional, se encuentra lista para introducirse al proceso de demodulación mediante el uso de un PLL. El diagrama a bloques de este demodulador se presenta en la figura 2.

- **Caracterización del demodulador para obtener sus especificaciones.** El circuito construido se caracterizó introduciendo una señal modulada en frecuencia con un generador de funciones marca Agilent 33220A con las siguientes características: una portadora de 1 kHz, una moduladora de 100 Hz, amplitud de 1 mV y un índice de modulación de 2 y se observó la señal de salida del demodulador en un osciloscopio Agilent DSO-X-2012, para observar tanto la forma de onda como su espectro de frecuencias. Se realizaron cambios de valores de la portadora y moduladora para el rango de frecuencia en el cual el demodulador deja de funcionar o presenta distorsión en su salida.

- **Construcción y Caracterización del Demodulador Seleccionado.** Se modificó el ancho de banda de los filtros del demodulador, y ser utilizado en un rango de baja frecuencia, para señales de interés que no sobrepasan los 300 Hz.

Todo el esquema de demodulación se muestra en el diagrama de bloques de la figura 2 donde BPF es filtro pasa banda, LPF es filtro pasa bajas y OSC es el oscilador local a ω_H y $2\omega_H$.

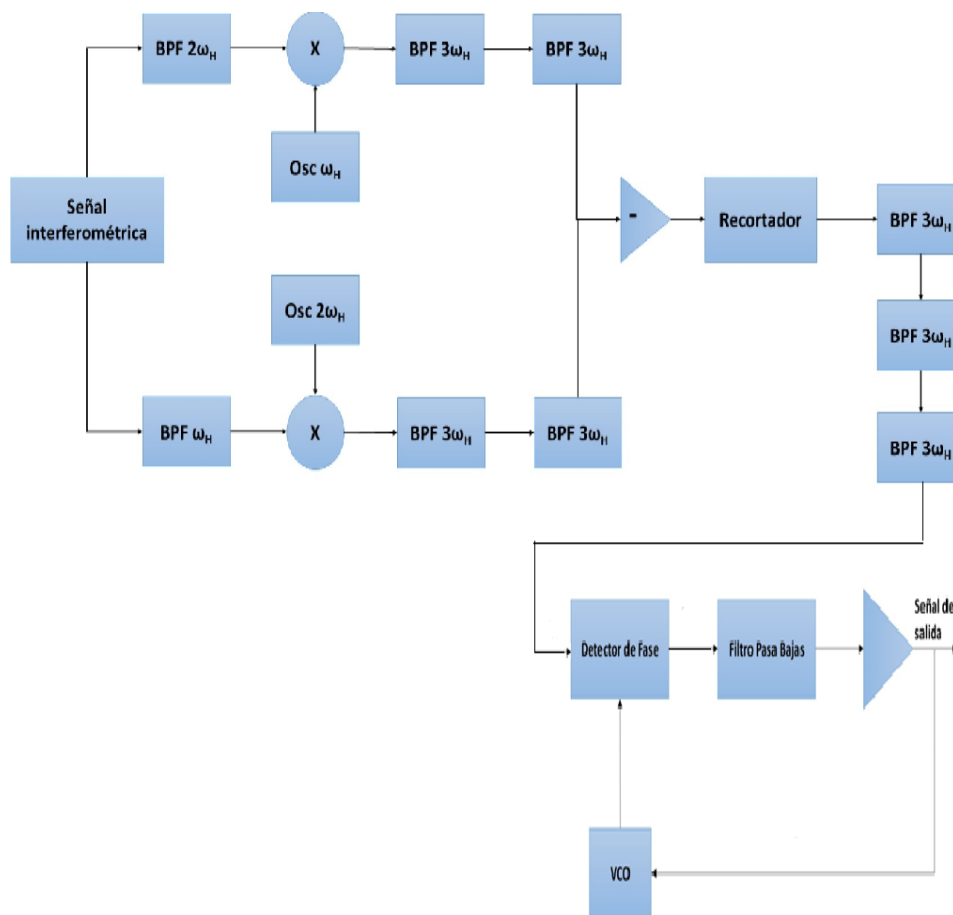


Figura 2 Diagrama a bloques del circuito demodulador utilizando PLL.

Filtros Pasabanda

Los filtros pasabanda se diseñaron con el circuito integrado UAF42 de Burr-Brown, son de segundo orden y poseen una arquitectura de filtros de variable de estado.

Multiplicadores

Los multiplicadores se construyeron con el circuito integrado AD633 cuya función de transferencia está dada por ecuación 2.

$$W = \frac{(X_1 - X_2)}{10V} \quad (2)$$

X_1 y X_2 representan voltajes de entrada que se convierten en fases, como se mencionó anteriormente.

Diseño del Oscilador Local

El diseño del oscilador local a ω_H utilizado corresponde a un oscilador de puente de Wien. Mientras que el oscilador local requerido a $2\omega_H$ se realizó utilizando un doblador de frecuencias a través de un multiplicador. La frecuencia del oscilador se define por la ecuación 3.

$$f_o = 1/2\pi R_{12} C_2 = \frac{1}{2\pi(158000)(1E^{-9})} = 1007 \quad (3)$$

Sumador

El sumador inversor permite combinar múltiples entradas, es decir, permite añadir algebraicamente dos (o más) señales o voltajes para formar la suma de dichas señales. La ganancia utilizada en nuestro diseño fue unitaria, por lo que el voltaje de salida del circuito es el mostrado en la ecuación 4, donde V_1 y V_2 representan las señales de entrada que serán sumadas.

$$V_{out} = -(V_1 + V_2) \quad (4)$$

Recortador

El recortador utiliza un diodo Zener BZX55C2V7, cuyo voltaje de operación es de 2.7 V, este limita la amplitud de la señal y con ello se logra una amplitud constante en toda la señal, minimizando los efectos de la variación de amplitud y así la distorsión adicional ocasionada por variaciones de amplitud indeseables en la señal que entrará al PLL para su posterior demodulación. Cuando la salida trata de exceder el voltaje zener, el zener entra en la región de avalancha y la salida queda recortada.

PLL

La parte central del demodulador la constituye el circuito mostrada en la figura 3, el demodulador de FM mediante PLL. La frecuencia del VCO está determinada por la ecuación 5 y en este caso se diseñó para una frecuencia de portadora de 3000 Hz para que pudiera engancharse con la señal modulada y se produjera la demodulación de FM.

$$VCO = \frac{0.3}{RT * CT} = \frac{0.3}{(4545 \Omega)(.022 \mu F)} = 3.0 \text{ kHz} \quad (5)$$

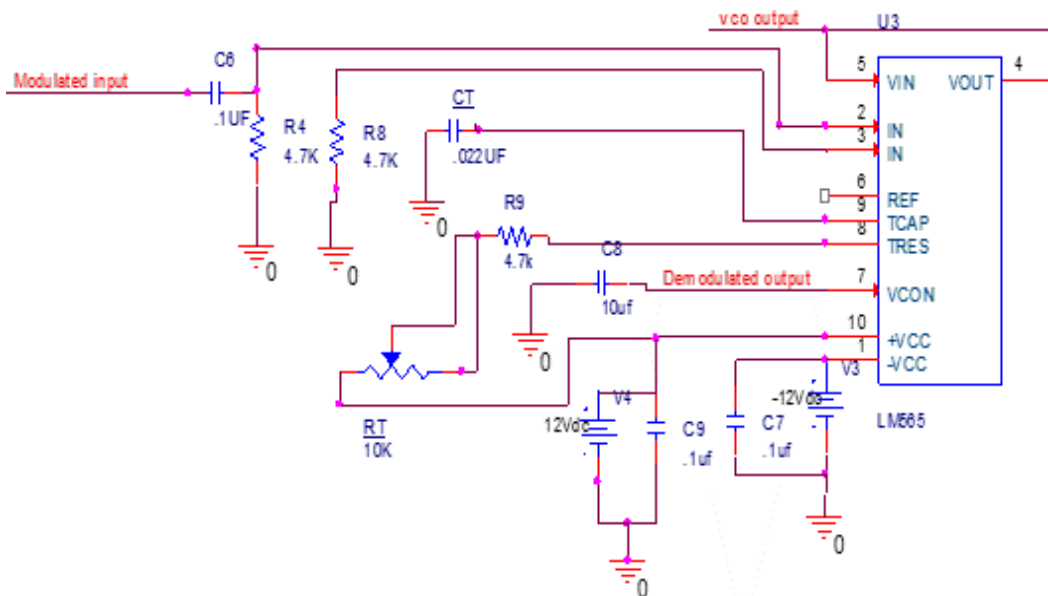


Figura 3 Diagrama del circuito demodulador de FM usando PLL.

Una vez ajustado el VCO a la frecuencia de la portadora que posee la señal modulada se procede al amarre de fase y a la demodulación de la señal. Se utilizó el IC LM565 al ser un circuito integrado que nos brinda sencillez en el diseño del circuito demodulador, sin dejar a un lado la calidad de este.

Filtro Pasabajos

El filtro pasabajos colocado en la etapa final del demodulador se construyó usando el circuito integrado UAF42. Su frecuencia de corte se calculó con base en la ecuación 6 y considerando un valor de $C=1000 \text{ pF}$.

$$F_c = \frac{1}{2\pi RC} = \frac{1}{2\pi(530500)(1E^{-9})} = 300 \text{ Hz} \quad (6)$$

Las características principales de este filtro fueron, un ancho de banda de 350 Hz y una banda de paso con caída de 6 dB por década.

3. Resultados

En este apartado se presentan los resultados experimentales obtenidos mediante la prueba del sistema construido.

Emulador de Interferogramas

La figura 4 muestra la señal generada por el emulador de interferogramas y la figura 5 muestra el espectro de frecuencias correspondiente a esta. En el espectro de Fourier de esta señal se puede ver que los valores de la portadora en 1 kHz y sus respectivos armónicos, así como también los valores de la moduladora con valor de 100 Hz presentes como bandas laterales de la portadora.

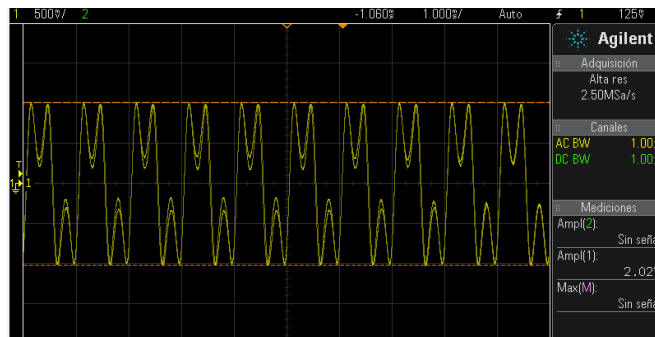


Figura 4 Interferograma obtenido con el emulador.

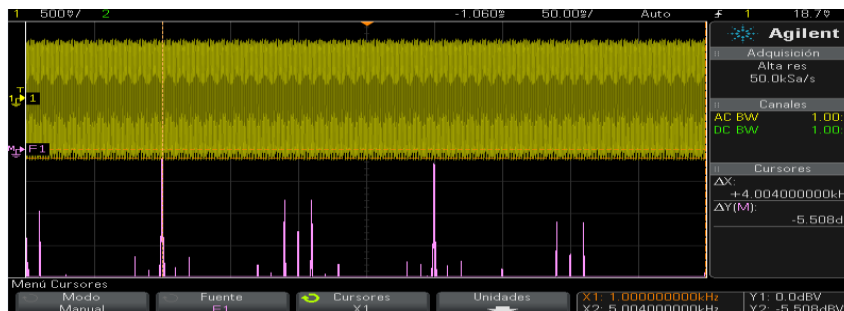


Figura 5 Espectro de frecuencias correspondiente a señal del simulador interferométrico.

Reproducción del Método de Demodulación a Utilizar

En la figura 6 se muestra la señal modulada en FM con un aportadora de 1 kHz y una moduladora de 100 Hz la cual fue introducida en el circuito PLL con una amplitud de 0.1 mV.

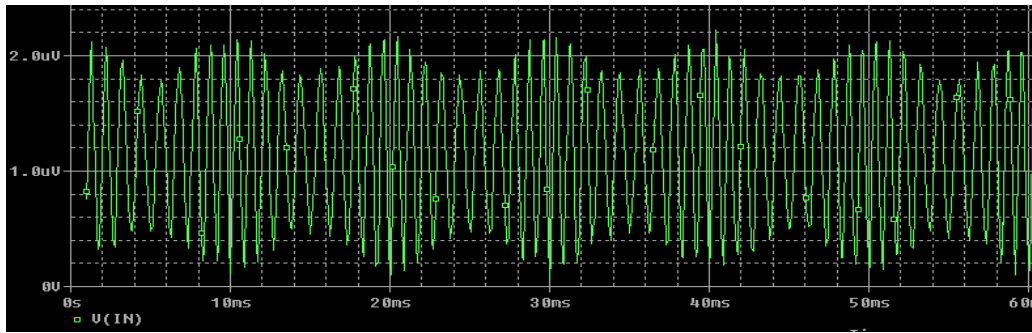


Figura 6 Señal modulada en frecuencia introducida en el demodulador de tipo PLL.

En la figura 7 se muestra la forma de onda obtenida en el simulador para una señal demodulada a través del PLL, corresponde a una senoidal con una frecuencia de 100 Hz.

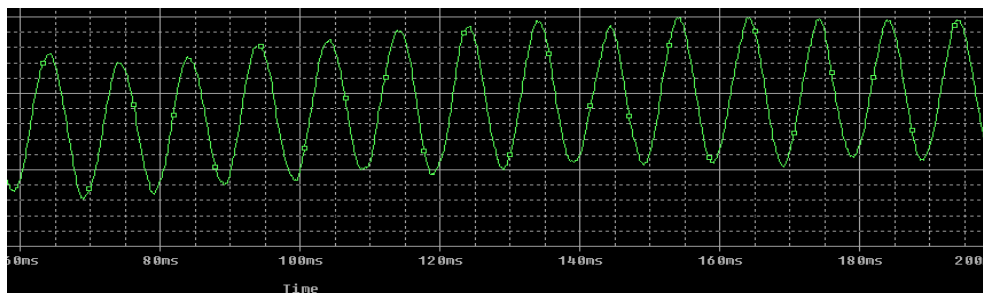


Figura 7 Forma de onda de la señal demodulada de 100 Hz.

Rediseño, Construcción y Caracterización del Demodulador Seleccionado, Mejorando la Relación Señal a Ruido

A continuación, se presentan los resultados obtenidos cuando la señal del interferograma correspondiente a la figura 5 y es procesada en cada etapa del demodulador. Por cuestiones de espacio se muestran solo las señales experimentales y solo en algunos casos se muestran las señales obtenidas con el simulador.

Filtros Pasabanda Centrados en 1 y en 2 kHz

En las figuras 8a y 8b se presenta la señal obtenida a la salida del filtro pasabanda centrado en 1 kHz.

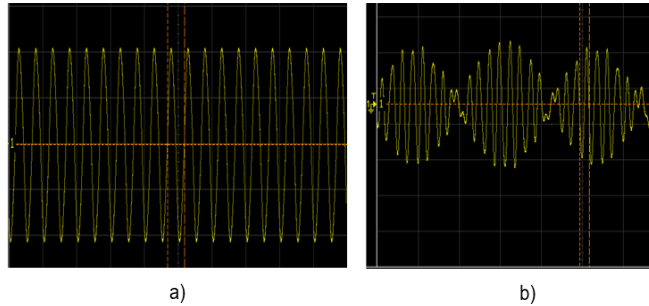


Figura 8 Señal entregada por el filtro pasabanda centrado en 1 kHz

Multiplicador del Canal 1

Como se mencionó anteriormente, existen dos etapas de mezclado o multiplicación, es decir, dos canales: el canal 1 situado en la parte superior del diagrama esquemático y el canal 2 situado en la parte inferior.

A continuación, en la figura 9a y figura 9b se presentan las señales obtenidas al multiplicar la señal de la figura 8a o figura 8b con una señal senoidal de 1 kHz.

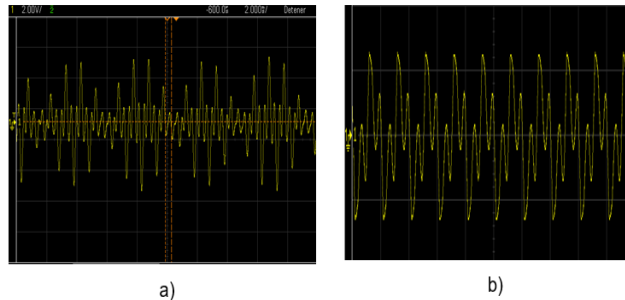


Figura 9 Señal entregada por los multiplicadores a) canal 1 y b) canal 2.

Etapas del Filtro Pasabanda en 3 kHz del Canal 1 y Canal 2

En la figura 10a y 10b se muestran las señales correspondientes a la salida de la etapa del filtro pasabanda centrado en 3 kHz de ambos canales.

Etapas del Sumador y del Recortador

La figura 11 se presenta la señal correspondiente a la salida tanto del sumador como de la etapa recortadora.

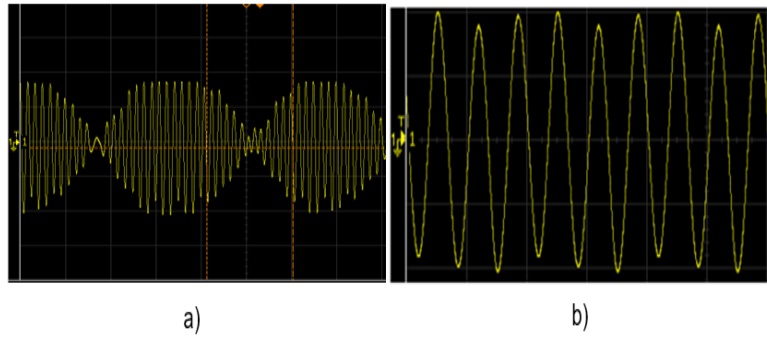


Figura 10 Señal entregada por filtro pasabanda a) canal 1 centrada en 3 kHz y b) canal 2.

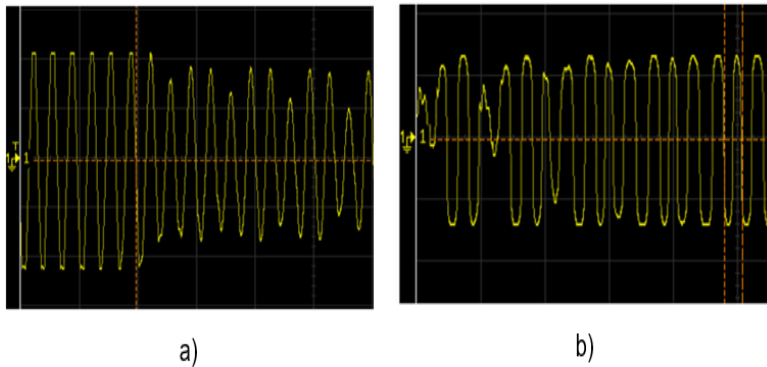


Figura 11 a) Señal entregada por el sumador y b) señal obtenida salida del recortador.

Etapa del Filtro Pasabanda en 3 kHz Previo a la Etapa del PLL

En la figura 12 se muestra la señal correspondiente a la salida del filtro pasabanda centrado en 3 kHz.

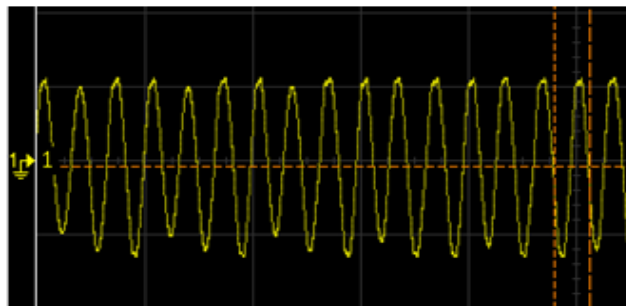


Figura 12 Señal entregada por filtro pasabanda de 3 kHz.

Etapa del PLL

La etapa medular del sistema de interrogación la constituye el demodulador de FM. Se presenta en la figura 13a la señal correspondiente a la salida del PLL.

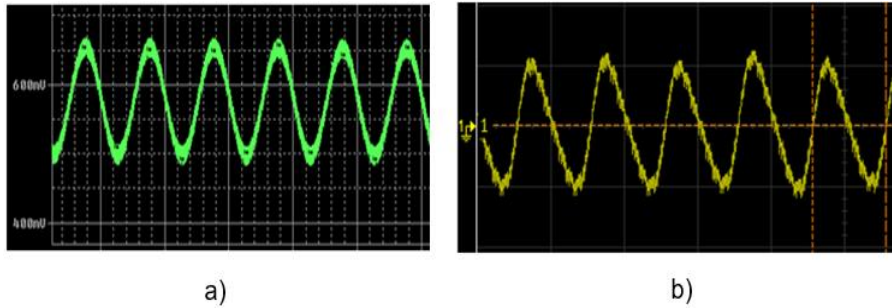


Figura 13 Señal correspondiente a salida del PLL. a) con simulador y b) con circuito físico.

Etapa del Filtro Pasabajas

La etapa del filtro pasabajas se desarrolló con frecuencia de corte en 350 Hz. Usando el emulador creamos 3 interferogramas, con señales moduladoras de diferentes valores de frecuencia 98.150 y 263 Hz. Al ser extraídas del circuito demodulador de fase óptica estas se muestran en los incisos de la figura 14 respectivamente. Se aprecian la forma de onda experimental y su comparación con la forma de onda original y se observan las distorsiones que se presentan a diferentes frecuencias para determinar el rango de operación óptimo. Señal recuperada (color amarillo) correspondiente a la moduladora introducida al inicio del circuito demodulador (color verde) y se compara la señal extraída respecto a su forma de onda para observar si existen deformaciones.

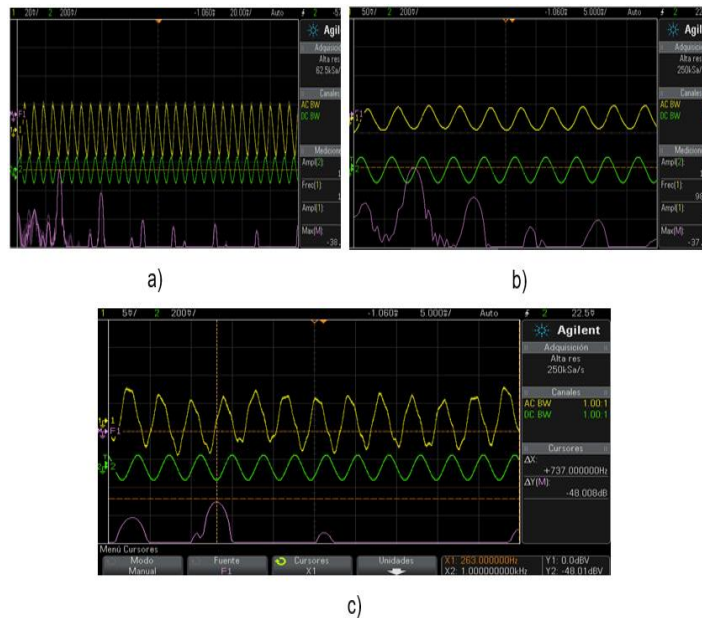


Figura 14 Señal moduladora: a) 150 Hz, b) 98 Hz y c) 263 Hz.

4. Discusión

La aplicación de las técnicas de demodulación de FM, a la interrogación de sensores interferométricos ha permitido que se pueda demodular o recuperar la fase óptica y consecuentemente la información codificada en la fase de un interferograma utilizando un circuito eléctrico. En particular, el uso de un PLL ofrece una gran oportunidad de diseñar sistemas de medición que operen en tiempo real con una resolución y una sensibilidad suficientemente buenas como para detectar cambios de fase en el orden de los miliradianes, aunque ya se han reportado la detección de hasta microradianes.

5. Conclusiones

Mediante el diseño y construcción del demodulador de fase óptica se concluye lo siguiente:

- Se diseñó y se construyó un circuito electrónico basado en el circuito integrado AD639 para generar una señal eléctrica que emula la señal generada por un biosensor óptico interferométrico.
- Mediante este simulador de interferogramas fue posible variar sistemáticamente y de manera controlada las variables de interés de un interferograma, como son amplitud y frecuencia y la fase estática de cada una de ellas; tanto de la señal portadora como de la moduladora, con la posibilidad de ingresar ruido aleatorio al sistema.
- Mediante este método demodulación se logró demodular señales del orden de miliradianes mediante la simulación en bloques de las etapas tanto de acondicionamiento previo al PLL para generar una señal típica de FM como la demodulación a través del PLL, esto a través de las librerías ABM de Orcad 16.6. lite
- La caracterización del demodulador seleccionado se logró mediante el espectro de frecuencias analizado a través de Orcad 16.6 lite y se concluye que es posible demodular señales de diversos rangos de frecuencias modificando tanto los rangos de operación de los osciladores locales para igualar la frecuencia de la portadora, como modificando el ancho de banda

de los filtros y el rango del VCO del PLL, así el conjunto de estas modificaciones permite sintonizar el rango de operación.

- Ajustando el conjunto de rangos señalados en el punto anterior referente a la caracterización (rango de operación del VCO, filtro pasabanda y osciladores locales) a 3 *kHz* se lograron demodular señales en el rango de frecuencias de moduladora de 80-260 *Hz*. Este rango de operación es modificable con base en la modificación del ancho de bando de los filtros pasabanda utilizados en el circuito y al rango de captura del PLL.

6. Bibliografía y Referencias

- [1] Dandridge, A. y A. B. Tveten. Homodyne Demodulation Scheme For Fiber Optic Sensors Using Phase Generated Carrier, *IEEE Photonics Soc.*, vol. 18, no. 10, pp. 1647–1653, 2010.
- [2] Fang, W. Jia, Q. Zhen, S. Chen, J. Cheng, X. y Yu, B. Low Coherence Fiber Differentiating Interferometer And Its Passive Demodulation Schemes, *Opt. Fiber Technol.*, vol. 21, no. 2015, pp. 34–39, 2015.
- [3] Kumar, R. Barrios, E. MacRae, A. Cairns, E. Huntington, E. H. y Lvovsky, A. I. Versatile Wideband Balanced Detector for Quantum Optical Homodyne tomography, *Opt. Commun.*, vol. 285, no. 24, pp. 5259–5267, 2012.
- [4] Malacara, D. *Óptica Básica*, Segunda edición, México: Fondo de Cultura Económica, pp. 532, 2004.
- [5] Malacara, D. *Optical Shop Testing*, Tercera edición, New Jersey: John Wiley & Sons, pp. 862, 2007.
- [6] McKinney, J. D. Colladay, K. y Williams, K. J. Linearization of phase-modulated analog optical links employing interferometric demodulation, *J. Light. Technol.*, vol. 27, no. 9, pp. 1212–1220, 2009.
- [7] Nolte, D. *Optical Interferometry for Biology and Medicine*, West Lafayette, IN, USA: Springer Science+Business Media, 2012.
- [8] Schlecker, B. Ortmanns, M. Anders, J. and Fantner, G. PLL-based high-speed demodulation of FM signals for real-time AFM applications, *Proc. - IEEE Int. Symp. Circuits Syst.*, pp. 197–200, 2013.

- [9] Silver, H. W. Experiment #68 — Phase Locked, Loops the Basic, hands-on radio, pp. 2, 2008.
- [10] Udd, E. Fiber Optic Sensors An Introduction for Engineers and Scientists, Segunda edición, New Jersey: John Wiley & Sons, 2007.
- [11] Zibar, D. Johansson, L. A. Chou, H. F. Ramaswamy, A. Rodwell, M. and Bowers, J. E. Novel optical phase demodulator based on a sampling phase-locked loop, *IEEE Photonics Technol. Lett.*, vol. 19, no. 9, pp. 686–688, 2007.

REVISIÓN DE MÉTODOS PARA LA ESTIMACIÓN DE LOS ESTADOS DE CARGA Y SALUD DE UNA BATERÍA

Alina Araceli Contreras Sillero

Tecnológico Nacional de México en Celaya

M1603053@itcelaya.edu.mx

Nimrod Vázquez Nava

Tecnológico Nacional de México en Celaya

n.vazquez@ieee.org

Claudia Verónica Hernández Gutiérrez

Tecnológico Nacional de México en Celaya

cvhg@ieee.org

Jeziel Vázquez Nava

Instituto Tecnológico Superior del Sur de Guanajuato

j.vazquez@itsur.edu.mx

Joaquín Vaquero López

Universidad Rey Juan Carlos

joaquin.vaquero@urjc.es

Resumen

Actualmente las baterías juegan un papel importante en el uso de energía eléctrica, dichos dispositivos tienen aplicaciones en pequeña y grande escala; para las aplicaciones de baja potencia (baja escala) las baterías son utilizadas en dispositivos electrónicos portátiles tales como teléfono celular, computadoras, ventiladores, etc.; para las aplicaciones de alta potencia (grande escala), éstas son usadas como reserva de energía para aplicaciones automotrices, inyección de energía a la red, entre otras.

La batería es el dispositivo más utilizado para almacenar energía por su practicidad y eficiencia que otorga al usuario. Una batería envejece en proporción

a los ciclos de carga y descarga; dichos procesos degradan las sustancias químicas que componen al dispositivo de almacenamiento; una baja carga tiene como consecuencia efectos de sulfatación y estratificación que acortan la duración de la batería, mientras que la sobrecarga provoca gases y pérdidas de agua.

Debido a que la energía almacenada en una batería es limitada se vuelve importante contar con la habilidad de determinar la capacidad disponible, el estado de carga (SOC, por sus siglas en inglés) y el estado de salud (SOH, por sus siglas en inglés) de dicho dispositivo; esto asegura que la batería tenga la energía disponible para ser utilizada. En este artículo se realiza una revisión y análisis comparativo sobre los principales métodos de estimación de SOC y SOH de las baterías en general.

Palabras Claves: AHC, carga, corriente, descarga, SOC, SOF, SOH, tensión.

Abstract

Batteries currently play a significant role in the use of electric power. Such devices have applications on a small and large scale; for low power applications (low scale) batteries are used in portable electronic devices such as cell phones, computers, fans, etc.; On the other hand, for high power applications (large scale) these sources are used as energy reserve for automotive applications, energy injection to the network, among others.

The battery is the most used device to store energy by its practicality and efficiency that it grants to the user. A battery ages in proportion to the loading and unloading cycles; Such processes degrade the chemicals that make up the storage device; a low charge results in stratification and sulphation effects that shorten battery life, while overloading causes gas and water loss.

Because power is stored in a limited battery, it is important to have the ability to determine the available capacity, the state of charge (SOC), and health status (SOH) of said device; this ensures that the battery has the power available to be used. In this paper, a review and a comparative analysis of the primary methods of estimation of SOC and SOH of general batteries is presented.

Keywords: AHC, charge, current, discharge, SOC, SOF, SOH, voltage

1. Introducción

Actualmente las diferentes áreas de investigación se han centrado en la utilización de recursos renovables para la generación y conservación de energía, sin embargo, su principal inconveniente es la intermitencia en la captación de la misma, por lo cual es necesario almacenarla en los períodos en los que no sea producida [Coleman, et al., 2008]. La batería es un dispositivo que se encarga de almacenar el excedente instantáneo de energía proveniente de fuentes intermitentes y proveerla cuando éstas son incapaces de satisfacer la demanda.

Las baterías han cobrado importancia en varias aplicaciones, como sistemas híbridos para la generación de energía, pues estos necesitan dispositivos de respaldo dónde almacenar la energía recogida; recientemente en la industria automotriz los acumuladores cumplen con la función de reducir la emisión de dióxido de carbono producido por el combustible, pues el objetivo a futuro es construir automóviles que usen energía de una batería en lugar de combustible tal como la gasolina o el diesel [Galeotti, et al., 2015].

El ciclo de carga y descarga de una batería tiene impacto en la vida de la misma ya que la sobrecarga de ésta provoca gases y pérdida de agua; por el contrario, la baja carga causa sulfatación, lo que reduce el área activa de las placas y puede incluso causar deformación en éstas.

Debido a la importancia de su papel en el funcionamiento de los sistemas que soportan, las baterías se han convertido en parte importante en la gestión de los mismos, por lo que la monitorización de su estado es relevante pues ayuda a evitar condiciones peligrosas de operación, a aumentar la vida útil de la batería, hacer eficiente su carga y descarga y prolongar su vida útil [Coleman, et al., 2008]. La caracterización de una batería se considera completa cuando se pueden medir sus 3 parámetros principales [Marchildon, et al., 2015].

Capacidad Ampere-Hora

(AHC, por sus siglas en inglés): Es la carga total que se le puede demandar a una batería completamente cargada bajo condiciones de carga definidas [Coleman, M., et al. 2008].

Estado de carga:

Se entiende como la cantidad de carga aún disponible en relación con la capacidad (AHC) de la batería. Sin embargo, la capacidad del acumulador no es del todo constante; algunos parámetros como temperatura, corriente de descarga, voltaje de corte y el estado de salud de la batería influyen en la capacidad de la misma. El estado de carga puede definirse de formas diferentes [Sauer, D. U., et al., 1999]. Tradicionalmente estado de carga (SOC) es la relación entre la diferencia de la capacidad nominal y la capacidad restante por un lado y la capacidad nominal por otro lado. El estado de carga es 1 cuando se alcanza el estado completo de carga y 0 después de una descarga completa.

$$SOC = \frac{AHC_{NOM} - Q_B}{AHC_{NOM}} \quad (1)$$

Donde:

- SOC Estado de carga de la batería
- Q_B Carga restante
- AHC_{NOM} Capacidad nominal Hora- Ampere

La capacidad práctica es siempre menor que la capacidad medida, ésta es usada comúnmente en sistemas fotovoltaicos.

Estado de salud (SOH, por sus siglas en inglés)

Se define como la razón entre la capacidad medida y la capacidad nominal. El estado de salud es 1, cuando ambas capacidades son iguales. Por definición, una batería está al final de su vida útil en un estado de salud de 0,8. [Sauer, et al., 1999]

$$SOH = \frac{AHC_{Aged}}{AHC_{Nom}} \quad (2)$$

La determinación del estado de carga de una batería no es suficiente para predecir el perfil general de la misma debido a su envejecimiento y deterioro con el paso del tiempo. La determinación del SOC no reconoce el cambio en la máxima capacidad útil de una batería; por lo tanto, se produce un error en la predicción de la potencia suministrada, éste se propaga conduciendo a errores mayores a

medida que disminuye el estado de salud; por tanto, la determinación de SOH de una batería es esencial en un sistema pues ayuda de una manera más eficaz a determinar las capacidades de energía de la batería. Adicionalmente el SOH permite establecer el régimen de carga adecuado y ayuda en la determinación de la capacidad de descarga o del perfil de potencia que queda en la batería

A lo largo de este artículo se dan a conocer los principales métodos para la determinación de estado de carga y salud de una batería en general, así mismo se realiza un análisis comparativo entre dichos métodos mencionando sus ventajas, desventajas y algunas aplicaciones de los mismos. Se pretende facilitar al usuario el conocimiento de estos métodos para que invierta la mayor parte del tiempo en la técnica que cumpla con los requerimientos de su sistema.

2. Métodos

Actualmente se han propuesto diferentes maneras para la determinación del estado de carga y el estado de salud de una batería. A continuación, se describen las principales técnicas para la estimación de dichos parámetros, así como algunas de sus aplicaciones.

Estimación por modelo eléctrico

Cada batería puede considerarse como un circuito eléctrico y los parámetros eléctricos se pueden modelar por diferentes métodos de acuerdo a la teoría de operación, todos ellos se basan en circuitos eléctricos equivalentes [Marchildon, et al., 2015].

Para representar las características eléctricas de las baterías se usan distintos modelos eléctricos: El modelo más simple consta de una fuente de voltaje ideal en serie con una resistencia interna, sin embargo éste modelo no tiene en cuenta el estado de carga de la batería. Otro modelo se basa en una tensión de circuito abierto en serie con resistencia y en paralelo con circuitos RC paralelos con la llamada impedancia Wartburg [Coleman, et al. 2008]. La identificación de todos los parámetros de este modelo se basa en una técnica bastante complicada llamada impedancia espectroscópica [Tremblay, et al., 2007].

Por otro lado, el modelo dinámico de la batería de la figura 1 considera el deterioro y el efecto de la temperatura. Se puede apreciar que este modelo se compone de una serie de resistencias y arreglos RC que representan distintas tensiones y/o impedancias internas del dispositivo de almacenamiento.

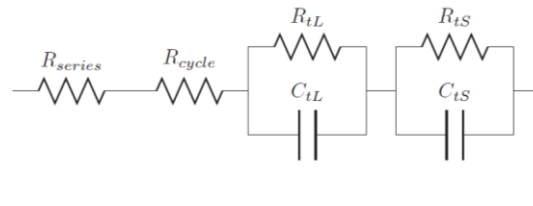


Figura 1 Modelo dinámico de la batería.

En la figura 1 R_{Series} representa la tensión de caída instantánea, R_{Cycle} representa una resistencia cíclica y R_{tL} , C_{tL} , R_{tS} , C_{tS} representan impedancias transitorias larga y corta respectivamente.

Los componentes RC en paralelo son convenientes para tomar en cuenta las pérdidas en la batería. Cabe mencionar que de acuerdo al valor de la carga en la batería, los valores de componentes de la figura 1 van cambiando [Purwadi, et al., 2014].

Los diferentes modelos eléctricos comparten una importante característica y es que hay una gran cantidad de simuladores que pueden realizar un análisis detallado de estos (de los más conocidos son Matlab, PowerSim, etc.), lo cual es una ventaja sobre otros modelos. Sin embargo, tener distintos modelos (uno para cada tipo de batería) representa un problema de estandarización; adicionalmente el modelado de cada uno de los parámetros del circuito eléctrico requiere de conocimientos electroquímicos y matemáticos, agregando cierto grado de complejidad en comparación con otras técnicas, lo cual es una desventaja para el uso de ésta [Piller, et al., 2001].

Técnica por Resistencia Interna

Las técnicas de impedancia han sido utilizadas para investigar la dinámica de las baterías, así como para determinar su estado de carga (SOC) y de salud

(SOH). La impedancia es denotada como un parámetro eléctrico, también llamada resistencia interna, (cuyo significado depende de la técnica de medición) ésta es definida como la función de transferencia entre la diferencia de potencial y la corriente, que usualmente es una cantidad compleja y se mide usando un analizador de respuesta de frecuencia [Huet, 1998].

La resistencia interna varía con el estado de carga de la batería y los cambios más grandes son notados en baterías de níquel. En la figura 2 se observa la resistencia interna de una batería de níquel-metal cuando está descargada, durante la carga y a plena carga después de un período de reposo de 4 horas. Los niveles de resistencia son más altos en estado bajo de carga e inmediatamente después de cargar. Contrariamente a la creencia popular, el mejor rendimiento de la batería no se logra inmediatamente después de una carga completa, sino después de un período de descanso de unas pocas horas. Durante la descarga, la resistencia interna de la batería disminuye, alcanza el punto más bajo a la mitad de carga y empieza a subir de nuevo (línea punteada).

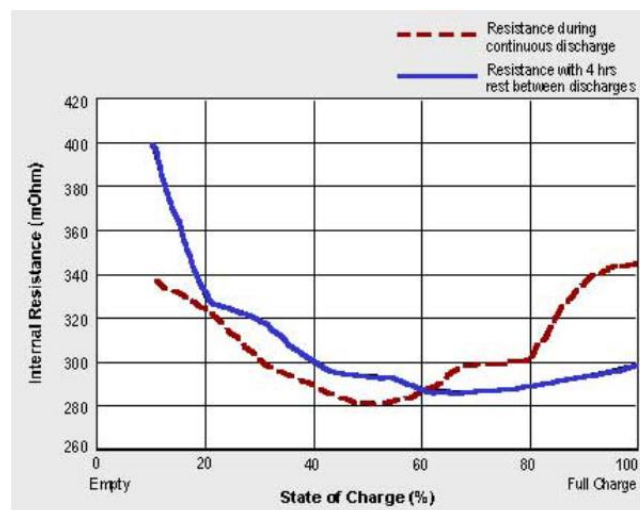


Figura 2 Resistencia interna de una batería de níquel-metal.

La resistencia interna proporciona información valiosa sobre una batería, como indicaciones de lectura altas al final de su vida útil (SOH). Esto es especialmente cierto con los sistemas basados en níquel [Battery University, 2017], [Coleman, et al., 2008].

La prueba de resistencia interna consiste en aplicar una carga breve a la batería y medir los cambios de voltaje y corriente para determinar su resistencia interna; ésta aumentará con la edad debido a la degradación química del material activo, a medida que aumenta la resistencia en el acumulador disminuye el SOC. Cabe hacer mención que esta técnica no considera los cambios de temperatura, por lo cual las lecturas de corriente y voltaje se ven afectadas si dichos parámetros suelen variar constantemente [Piller, et al., 2001].

Prueba de Descarga Completa

El modelo de descarga es usado por muchos fabricantes, pues es la prueba de calidad que se realiza a una nueva batería para determinar su capacidad AHC nominal. Este tipo de prueba implica aplicar una descarga a la batería cuando se encuentra completamente cargada y medir la carga entregada, ésta se compara con la obtenida de una prueba de descarga completa cuando la batería era nueva [Coleman, et al., 2008].

La batería es descargada usando su índice de corriente (CR, por sus siglas en inglés) mientras que el voltaje de la batería se mide durante las 'n' horas requeridas para realizar la prueba. Si la tensión de la batería, al final de la prueba, es más alta que la mínima especificada, entonces se reconoce que la batería tiene la clasificación CR. Las especificaciones más elaboradas incluirán varios gráficos según diferentes clasificaciones CR.

Las estimaciones de SOC se realizan comparando con el gráfico suministrado por el fabricante. Aunque esto parece sencillo y directo, hay tres dificultades en el uso de este método: la primera es que los gráficos proporcionados por el fabricante muestran el voltaje bajo una carga específica; se requiere hacer extrapolaciones si la carga real es diferente de la proporcionada por la hoja de datos. La segunda es que los gráficos no proporcionan información sobre el voltaje de la batería sin carga. Finalmente, este modelo no considera el SOH de las baterías.

Esta técnica normalmente incluye una recarga consecutiva, la cual es demasiado lenta para ser considerada en la mayoría de las aplicaciones, además durante la prueba la función del sistema es interrumpida [Marchildon, et al. 2015].

Método de 2 Pulsos

Los parámetros tales como AHC, SOC y SOH de una batería están relacionados con la caída de tensión después de cada pulso de descarga de corriente. El primer pulso resetea a la batería de su historia anterior y el segundo establece parámetros que tienen una relación directa con el estado de carga y salud de ésta.

El método consta de tres pasos (figura 3).

- Se comienza con una batería de historia desconocida, la cual debe estar en circuito abierto, la duración mínima de éste dependerá del perfil de carga de la batería y de los pulsos de corriente.
- Un pulso conocido de corriente de carga se aplica a la batería durante 10 s, aunque puede ser tan corto como 3 s, pero los estudios anteriores demuestran que uno de 10 s da resultados consistentes [Coleman, et al. 2008]. La tensión ΔV_1 en el transcurso del primer pulso decae, después de que el pulso es eliminado el voltaje se recupera durante otros 10 s a V_{MAX}
- Un segundo pulso idéntico al primero se aplica, ΔV_2 decae de la misma manera que ΔV_1 recuperándose y dando como resultado V_{MIN} .

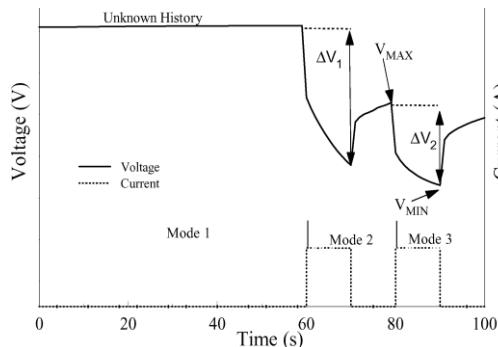


Figura 3 Gráfica para el método de 2 pulsos.

Una vez obtenidos los valores de I , V_{MAX} y Δv_2 , estos pueden ser interpretados para la obtención del estado de la batería, como se muestra a continuación:

- Paso 1.- El voltaje de equilibrio V_{EMF} es deducido de V_{MAX} y la hoja de datos del fabricante.

- Paso 2.- El SOC de la batería es deducido de V_{EMF} y de la hoja de datos del fabricante.
- Paso 3.- El C_R es deducido de Δv_2
- Paso 4.- El AHC se obtiene del paso 1.
- Paso 5.- El SOH está dado por el paso 3

La relación entre los parámetros obtenidos por medio de la gráfica mostrada en la figura 3 y los parámetros de estimación SOC y SOH, pueden expresarse de la siguiente manera:

$$SOC = \frac{V_{M\acute{a}x} + \beta - EMF_{Min}}{\sigma} \quad (3)$$

$$SOH = \frac{AHC_{Aged}}{AHC_{Nom}} \quad (4)$$

Donde:

σ Pendiente entre el % SOC y el voltaje en la batería.

β "Offset" que el voltaje de equilibrio puede tener.

β y σ dependen del tipo de batería, pero de manera general la relación se mantiene [Colleman, et al. 2008].

Un aspecto importante de la técnica de 2 pulsos es que se puede realizar en menos de 5 minutos una vez que la batería haya sido caracterizada. Hay que tener en cuenta que para esto se requieren de pruebas preliminares donde es posible utilizar otras técnicas tales como el conteo de Coloumb, descarga completa, etc. con el fin de obtener los coeficientes β y σ por lo que se requiere más tiempo para dicho proceso [Marchildon, et al., 2015], [Coleman, et al., 2008].

Conteo de Coulomb

Computadoras, equipos médicos y otros dispositivos portátiles utilizan el conteo de Coulomb para estimar el estado de carga de una batería, midiendo la corriente de entrada y salida. El voltaje y la resistencia interna de una batería son dos parámetros que se pueden obtener fácilmente y por lo tanto son convenientes para la estimación de SOC, sin embargo, estos no sólo cambian irregularmente

con la profundidad de descarga (DOD, por sus siglas en inglés), la velocidad (carga / descarga) y la temperatura ambiente, sino que también dependen en gran medida del estado de salud (SOH) de las baterías [Battery University, 2017].

El método de conteo de Coulomb calcula la capacidad restante acumulando la carga transferida dentro o fuera de la batería.

En la figura 4 se muestra el diagrama de flujo para la estimación de SOC. Inicialmente se recuperan los datos históricos de la batería (SOC, SOH); para una batería nueva, se suponen al 100%.

La estimación se inicia probando el voltaje en circuito abierto con el fin de verificar que la batería no se encuentre por debajo del voltaje de corte establecido por el fabricante.

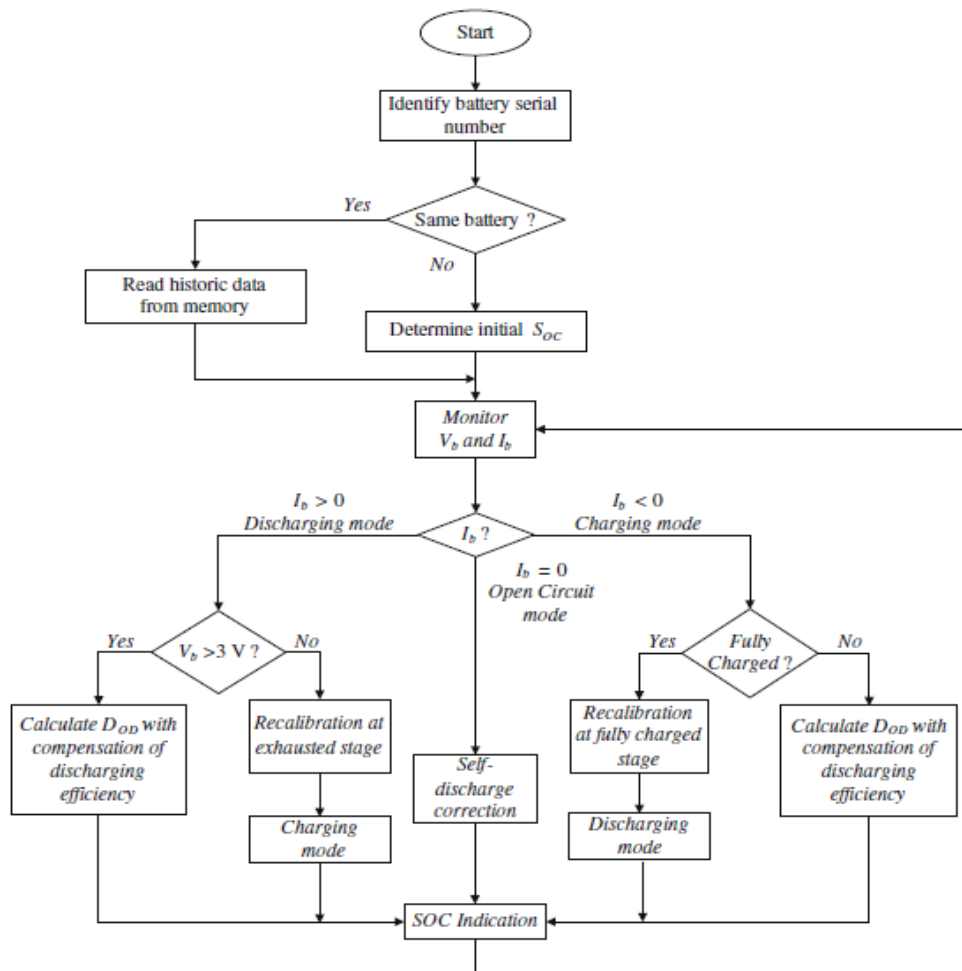


Figura 4 Diagrama de flujo del método de estimación de conteo de Coulomb mejorado.

El proceso de estimación se basa en el control de la tensión (V_B) y corriente (I_B) de la batería; su modo de funcionamiento puede ser conocido de acuerdo a la cantidad y dirección de la corriente. Cuando la batería está en circuito abierto con corriente cero, se realiza una compensación de auto descarga para la realización de una corrección de los estados iniciales de carga y salud de la batería.

Se dice que cuando el voltaje V_B sea menor al voltaje de corte (especificado en la hoja del fabricante) durante la descarga, la batería ya no puede ser usada y se debe recargar. Por otro lado, la batería está completamente cargada si V_B alcanza el límite superior especificado en la hoja de datos e $I_B=0$.

Así mismo el SOH y SOC son calculados con:

$$SOC_{(t)} = SOH_{(t)} - DoD_{(t)} \quad (5)$$

$$SOH = \frac{Q_{M\acute{a}x}}{Q_{Rated}} \quad (6)$$

Donde:

$DoD_{(t)}$: Profundidad de descarga

$Q_{M\acute{a}x}$: Carga Mxima

Q_{Rated} : Carga Nominal

Al final del ciclo de carga se obtiene un nuevo SOH, acumulando la suma de la carga total puesta en la batera.

Si bien sta es una solucin elegante a un problema retador, las prdidas presentadas en el proceso reducen la energa total entregada y lo que est disponible al final es siempre menor de lo que se tena en un inicio. A pesar de esto, la tcnica del conteo de Coulomb funciona bien, especialmente con bateras de Li-ion que ofrecen alta eficiencia colombina y baja auto-descarga. [Ng, et al., 2009].

Tcnica de Voltaje en Circuito Abierto

Medir el estado de carga por voltaje es simple, pero puede ser impreciso porque los materiales de las celdas y la temperatura se ven afectados. El error ms evidente del estado de carga basado en voltaje se produce al alterar una batera

con una carga o descarga. La agitación resultante distorsiona el voltaje y ya no representa una referencia SOC correcta. Para obtener lecturas precisas, la batería necesita descansar en el estado de circuito abierto durante al menos cuatro horas; aunque los fabricantes de baterías recomiendan 24 horas para el ácido de plomo. Esto hace que el método SOC basado en voltaje sea poco práctico para una batería en servicio activo [Battery University, 2017].

La curva de descarga de una batería depende de su química. Mientras que el SOC basado en voltaje funciona razonablemente bien para una de plomo-ácido, para las de níquel y litio el método de voltaje es impráctico.

Cuando se mide la SOC mediante voltaje de circuito, el voltaje de la batería debe ser "flotante" sin carga. Sin embargo, para aplicaciones automotrices las cargas parásitas para las funciones de limpieza ponen la batería en una condición de voltaje de circuito casi cerrado (CCV, por sus siglas en inglés).

A pesar de las imprecisiones, la mayoría de las mediciones SOC dependen, en parte o completamente, del voltaje debido a la simplicidad. El SOC basado en voltaje es popular en sillas de ruedas, patinetas y coches de golf. Algunos sistemas de gestión de baterías (BMS por sus siglas en inglés) innovadores usan los períodos de descanso para ajustar las lecturas SOC como parte de una función de "aprendizaje".

La medición de voltaje de circuito abierto suele combinarse con otras técnicas para asegurar una indicación continua de SOC, en tal combinación, la medición de voltaje de circuito abierto puede utilizarse para ajustar las otras técnicas. Finalmente, el desgaste químico que se produce en la batería debido a la carga y descarga de la misma, tales como la concentración de ácido y las estratificaciones ácidas pueden generar resultados inexactos [Battery University, 2017].

3. Resultados

Análisis Comparativo de Métodos

Para el análisis entre los diferentes métodos de estimación se toman en cuenta los siguientes criterios: complejidad, exactitud y monitoreo. En la tabla 1 se muestra la comparación entre cada una de las técnicas anteriormente descritas.

Tabla 1 Comparación de Métodos de estimación de estados de una batería.

Método de Estimación	Ventajas	Desventajas	Parámetros Obtenidos	Monitoreo	Aplicaciones
Modelo Eléctrico	*Cuenta con varias plataformas de software para su simulación. *Suelen dar una buena aproximación.	*Problemas de estandarización *Requiere de conocimientos electroquímicos y eléctricos para la extracción de los parámetros	SOC, SOH	N/A (Sim.)	Aplicaciones de diseño eléctrico y software
Impedancia Interna	*Fácil implementación	*Carece de exactitud pues considera parámetros que varían con la temperatura.	SOH, SOC	Fuera de línea	Baterías comerciales
Descarga Completa	*Método Estándar elegido por los fabricantes comerciales por su sencilla aplicación.	, *Demanda mucho tiempo. *Son necesarias extrapolaciones si la carga difiere a las consideradas por el fabricante.	SOH	Fuera de línea	Baterías comerciales.
Métodos 2 Pulsos	*Una vez encontrados los coeficientes es rápido de implementar. *Resultados consistentes	*Requiere de tiempo, así como de apoyo de otras técnicas de estimación para encontrar los coeficientes.	SOC, SOH	Fuera de línea	Banco de baterías
Conteo de Coulomb	*Implementación Sencilla *Alta Eficiencia Colombiana.a potencia media	*Es sensible a desbalances de carga, temperatura y pérdidas internas.	SOC, SOH	En línea	Computadoras, equipos médicos, elementos portátiles.
Técnica de voltaje en circuito abierto	*Fácil Implementación	*Solo funciona en circuito abierto y después de un tiempo de reposo. *Interrupción de las funciones del sistema.	SOC	Fuera de línea	Aplicaciones domésticas

4. Discusión

Para poder considerar que un método es mejor que otro, es importante conocer la aplicación del mismo, pues las técnicas más sencillas, aunque carezcan de exactitud, poseen la ventaja de su facilidad y pueden ser útiles para aplicaciones sencillas donde el usuario sólo requiera un estimado de los estados de carga y salud; por el contrario, no podemos tener el mismo criterio en aplicaciones fotovoltaicas o en micro redes donde las baterías juegan un papel crítico, pues se

usan como almacenadores de energía, cuando se carece de ésta, en esos casos se aprecia la exactitud y el monitoreo incluso cuando el método sea complejo.

Sin embargo, la técnica más utilizada actualmente es el conteo de Coulomb pues es el método más directo, transparente y cuyos resultados son satisfactorios a pesar de sus desventajas [Battery University,2017].

5. Conclusiones

En aplicaciones donde se utilizan baterías es importante conocer el estado de las mismas, para asegurar que la energía requerida por el sistema se encuentra disponible. A lo largo de este documento se realizó una revisión de las técnicas existentes para el cálculo del estado de carga y del estado de salud de la misma.

La determinación del SOC de una batería no es suficiente para determinar su estado, pues éste no reconoce el cambio en la máxima capacidad útil del acumulador; por lo tanto, se produce un error en la predicción de la potencia suministrada, este error se propaga conduciendo a otros mayores a medida que disminuye el estado de salud, de ahí la importancia de la obtención del mismo. La estimación del SOH de una batería es esencial pues ayuda de una manera más eficaz a determinar las capacidades de energía de la batería. Adicionalmente el SOH permite establecer el régimen de carga adecuado y ayuda en la determinación de la capacidad de descarga o del perfil de potencia que queda en la batería.

Entre los métodos descritos, el de conteo de Coulomb es la técnica que presenta mayores prestaciones, debido a que es funcional para todo tipo de baterías; el cálculo es bastante eficiente en media potencia a pesar de factores cambiantes tales como la temperatura y la impedancia, permite conocer el estado de la batería en todo momento incluso cuando está en funcionamiento, es de fácil aplicación y de bajo costo, sin contar que este método puede ser complementado por métodos que le permitan “re-calibrarse” en alguno de sus estados .

Con base en el análisis comparativo realizado se concluye que la selección de la mejor técnica para estimar el SOC y SOH de una batería depende la aplicación donde éste se requiera y de los parámetros a considerar de la misma.

6. Bibliografía y Referencias

- [1] Battery University, Battery University: 2017: http://batteryuniversity.com/learn/article/how_to_measure_state_of_charge_
- [2] Coleman, M., Hurley, W. G., & Lee, C. K., An improved battery characterization method using a two-pulse load test. *IEEE Transactions on energy conversion*, 23(2), pp. 708-713, 2008.
- [3] Ehret, C., Piller, S., Schroer, W., & Jossen, A., State-of-charge determination for lead-acid batteries in PV-applications. In *Proceedings of the 16th European Photovoltaic Solar Energy Conference, Glasgow (Vol. 2486, pp. 2489, 2000.*
- [4] Galeotti, M., Giammanco, C., Cinà, L., Cordiner, S., & Di Carlo, A., Synthetic methods for the evaluation of the State of Health (SOH) of nickel-metal hydride (NiMH) batteries. *Energy Conversion and Management*, 92, pp. 1-9, 2015.
- [5] Huet, F., A review of impedance measurements for determination of the state-of-charge or state-of-health of secondary batteries. *Journal of power sources*, 70(1), pp. 59-69, 1998.
- [6] Leksono, E., Haq, I. N., Iqbal, M., Soelami, F. N., & Merthayasa, I. G. N. (2013, November). State of charge (SoC) estimation on LiFePO 4 battery module using Coulomb counting methods with modified Peukert. In *Rural Information & Communication Technology and Electric-Vehicle Technology (rICT & ICeV-T), 2013 Joint International Conference on IEEE*, pp. 1-4, 2013.
- [7] Marchildon, J., Doumbia, M. L., & Agbossou, K., SOC and SOH characterisation of lead acid batteries. In *Industrial Electronics Society, IECON 2015-41st Annual Conference of the IEEE*, pp. 001442-001446, 2015.
- [8] Ng, K. S., Moo, C. S., Chen, Y. P., & Hsieh, Y. C., Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries. *Applied energy*, 86(9), pp. 1506-1511, 2009.

- [9] Piller, S., Perrin, M., & Jossen, A., Methods for state-of-charge determination and their applications. *Journal of power sources*, 96(1), pp. 113-120, 2001.
- [10] Purwadi, A., Rizqiawan, A., Kevin, A., & Heryana, N., State of Charge estimation method for lithium battery using combination of Coulomb Counting and Adaptive System with considering the effect of temperature. In *Power Engineering and Renewable Energy (ICPERE), 2014 International Conference on IEEE*, pp. 91-95. 2014.
- [11] Sauer, D. U., Bopp, G., Jossen, A., Garche, J., Rothert, M., & Wollny, M., State of Charge—What do we really speak about. In *The 21st international telecommunications energy conference*, pp. 6-9, 1999.
- [12] Tremblay, O., Dessaint, L. A., & Dekkiche, A. I., A generic battery model for the dynamic simulation of hybrid electric vehicles. In *Vehicle Power and Propulsion Conference, 2007. VPPC 2007. IEEE*, pp. 284-289, 2007.

APLICACIÓN MÓVIL PARA EL CÁLCULO DE RUTAS “LOBOBICI” EN CIUDAD UNIVERSITARIA BUAP BASADA EN BÚSQUEDAS

Eliúh Cuecuecha Hernández

Benemérita Universidad Autónoma de Puebla
eliuhcueh@gmail.com

José Javier Martínez Orozco

Benemérita Universidad Autónoma de Puebla
javier.35.93.01@gmail.com

Daniel Méndez Lozada

Benemérita Universidad Autónoma de Puebla
daniel.mendezl@alumno.buap.mx

Adán Zambrano Saucedo

Benemérita Universidad Autónoma de Puebla
zambranos.adan@gmail.com

Aldrin Barreto Flores

Benemérita Universidad Autónoma de Puebla
aldrin.barreto@correo.buap.mx

Verónica Edith Bautista López

Benemérita Universidad Autónoma de Puebla
vbautista@cs.buap.mx

Salvador Eugenio Ayala Raggi

Benemérita Universidad Autónoma de Puebla
saraggi@ece.buap.mx

Resumen

En este documento se presenta la implementación de los algoritmos de búsqueda no informada: amplitud (BFS) y costo uniforme (UCS), mediante el

desarrollo de una aplicación móvil en la plataforma *MIT App Inventor*. La aplicación se enfoca en el cálculo del camino más corto y el de menos transbordos que el usuario puede realizar dependiendo de su estación de origen y destino. Con ello se obtuvo una herramienta para comparar el desempeño de ambos algoritmos a la vez que se le brinda al estudiante un mejor servicio en el programa LOBOBICI.

Palabras Claves: Aplicación móvil, app Inventor, búsquedas, búsqueda de costo uniforme (USC), búsqueda en amplitud (BFS).

Abstract

This paper presents the implementation of uninformed search algorithms: Breadth-first (BFS) and Uniform cost (UCS), by a mobile application development on the MIT App Inventor platform. The application focuses on the estimation of the shortest path and the less transfers the user can take depending on his origin and destination stations. This provides a tool to compare the performance of both algorithms while providing the student with a better service in the LOBOBICI program.

Keywords: App Inventor, breadth-first search (BFS), mobile application, searches, uniform cost search (UCS).

1. Introducción

A lo largo de décadas se han estudiado problemas donde el objetivo es encontrar la ruta más corta [Guerriero, 2001]. Por muchos años diversos algoritmos de inteligencia artificial fueron propuestos para resolver este tipo de problemas que ahora se vuelve un reto determinar cuál es mejor para cada aplicación en específico [Kumar, 1988]. Las técnicas de búsqueda pretenden encontrar una solución válida dentro de un espacio de estados [Brooks, 1991].

En este documento se abordan las técnicas específicas de búsqueda en amplitud y costo uniforme, las cuales forman parte de las búsquedas no informadas. Son denominadas así debido a que el problema a resolver no ofrece ninguna

información adicional que ayude a encontrar una solución más rápido, es por esta razón que se han elegido estos algoritmos para el desarrollo de esta investigación. Uno de los primeros artículos enfocado a resolver problemas del camino más corto es [Gallo, 1986]. En este documento se describen diferentes algoritmos de búsquedas desde un enfoque puramente matemático con el objetivo de dar una pequeña clasificación de los mismos sin ahondar en alguna aplicación en específico.

Se han elaborado múltiples trabajos comparativos de algoritmos de búsquedas informadas y no informadas. En [Chiong, 2008] se determina la ruta más corta en la isla de Borneo en Malasia con diferentes algoritmos de inteligencia artificial destacando la eficiencia del algoritmo Dijkstra. En este documento se resalta la importancia en estos días de encontrar la ruta más corta para llegar de un lugar a otro en poco tiempo debido a que la planeación de viajes por carretera es un asunto que cobra cada vez más relevancia gracias al exponencial aumento de vehículos. Se denota además cómo es que estos algoritmos son de cuantiosa aplicación en la resolución y soporte de muchos de los problemas cotidianos que tiene una sociedad.

En otros casos se optó por la comparación entre búsquedas no informadas contra heurísticas para la resolución de un rompecabezas y obtener una idea de cuál algoritmo tuvo mejor desempeño en base a la memoria consumida y el tiempo de ejecución [Kuruvilla, 2014]. En este mismo trabajo se presentan los parámetros que deben ser evaluados para poder discernir el algoritmo adecuado para cada aplicación y abre un panorama amplio de posibilidades para su implementación.

Hay trabajos que sólo abordan el plano teórico y nos brindan el *know-how* de cada algoritmo para enfocarlos en distintas aplicaciones, en ellos se denotan las características sobresalientes de estos métodos y dejan al usuario una primera aproximación de su desempeño como en [Chandel, 2014].

A pesar del vasto trabajo y desarrollo de estos métodos, no hay un documento registrado dónde se hayan puesto en marcha algoritmos de búsquedas en una aplicación móvil, debido principalmente a la baja capacidad de procesamiento que

tenían estos dispositivos. Ahora contamos con teléfonos celulares de iguales características que una computadora portátil convencional.

Cabe destacar que el buen desempeño de un método depende en mayor grado al tipo de aplicación para la cual se esté ejecutando, un algoritmo es eficiente sólo para ciertos casos y cumpliendo ciertos requisitos [Korf, 1999].

La aplicación desarrollada aborda el problema del desconocimiento de los caminos y rutas que la universidad tiene, cuando se es alumno de recién ingreso. El campus principal de la Benemérita Universidad Autónoma de Puebla es muy extenso y cuenta con pistas para circular en bicicleta a lo largo del mismo, este servicio se llama LOBOBICI. Muchas veces es difícil escoger el mejor camino para llegar de un punto a otro pues existen diversas posibilidades, el hecho de tomar una mala decisión puede llevar al estudiante a perder mucho tiempo.

La aplicación móvil ataca este problema directamente y compara la distancia a recorrer contra el número de estaciones por las que tiene que pasar el usuario. Lo anterior permite una manera interesante de contrastar el desempeño de los dos algoritmos previamente mencionados pudiendo destacar la eficiencia de la búsqueda por amplitud (BFS) frente a la de coste uniforme (UCS), para este caso en específico.

2. Métodos

En primer lugar, se generó un diagrama con las estaciones de LOBOBICI y la distancia entre cada una de ellas. Estas últimas se obtuvieron mediante el uso de la aplicación para *smartphone* SHealth de la empresa Samsung, que mediante el conteo de pasos estima la distancia recorrida con ayuda del GPS del dispositivo. El diagrama de la figura 1 muestra los datos recabados. Todas las distancias se expresan en metros y cada estación se encuentra numerada del 1 al 18.

Programación de la Aplicación para Android

El diseño e implementación de la aplicación, se realizó en MIT App Inventor. Esta es una herramienta de programación para la creación de aplicaciones en Android, la característica principal de ella es el uso de bloques que pueden

conectarse entre sí, simplificando el proceso de programación a una tarea más visual.

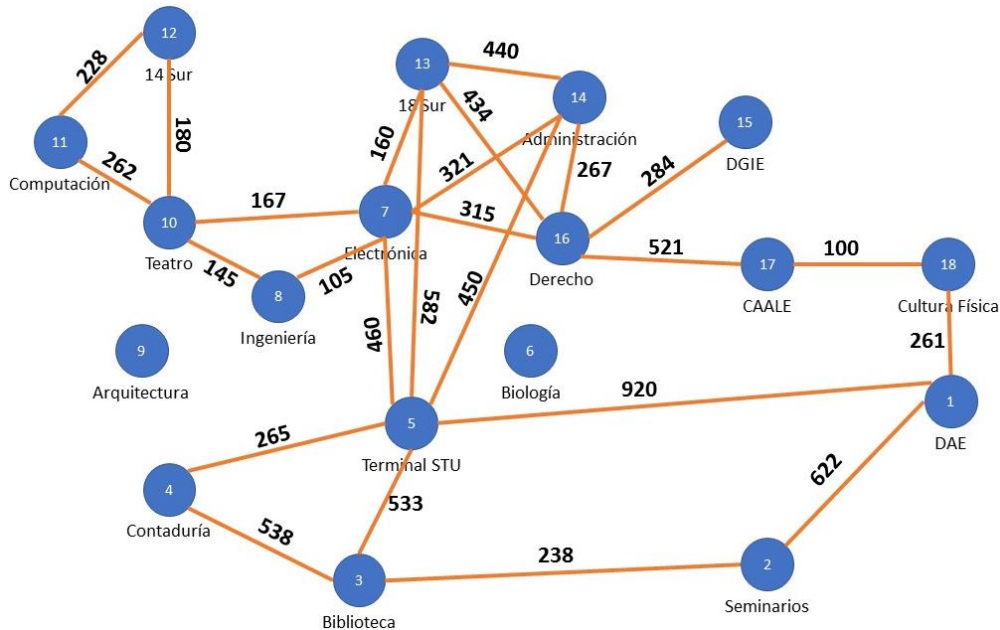


Figura 1 Esquema de las estaciones de LOBOBICI y distancia entre ellas.

Interfaz de Usuario

Esta consiste en una pantalla con el mapa de las estaciones de LOBOBICI de ciudad universitaria de la BUAP, se elige el origen y el destino para proceder a calcular la ruta más corta entre ellas, correspondiente a una búsqueda con costo uniforme (USC), o la ruta con menos estaciones que corresponde a una búsqueda por amplitud (BFS).

La figura 2 presenta el diagrama de bloques con el procedimiento para el uso de la aplicación.

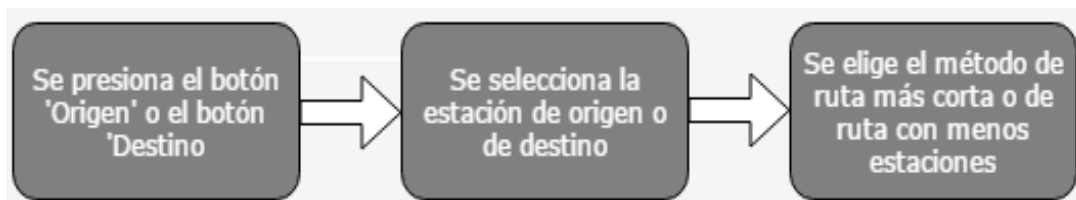


Figura 2 Diagrama de bloques de uso de la aplicación.

En la figura 3 se presenta la interfaz de usuario, los colores de las diferentes estaciones fueron establecidos por los administradores del programa LOBOBICI e indican las diferentes facultades que existen en el campus. Lo anterior sirve como guía visual para el usuario y no presenta relevancia para el desarrollo de este trabajo.



Figura 3 Interfaz de usuario.

Implementación de Algoritmo de Búsqueda por Amplitud (BFS)

El pseudocódigo mostrado en la figura 4, correspondiente al algoritmo de BFS [Serrano, 2012], hace uso de 2 listas principales, la de nodos frontera y la de nodos visitados. En la lista de nodos frontera se incluirán todos aquellos nodos que no han sido visitados ni repetidos. Mientras que la lista de nodos visitados contendrá a todos aquellos nodos que ya han sido visitados y que no son la solución [Korf, 1985].

Es importante mencionar que para el caso específico de las rutas de LOBOBICI, se descartan en primera instancia las estaciones de Arquitectura y de Biología pues ambas se encuentran cerradas debido a reparaciones que se llevan actualmente en el campus de la Universidad.

```
nodo_inicial = estado inicial
nodos_frontera = Cola FIFO
nodos_visitados = Lista
almacenar nodo_inicial en nodos_frontera
mientras nodos_frontera no vacío:
  nodo_actual = extraer un nodo de nodos_frontera
  si nodo_actual == solución:
    salir con solución

  introducir nodo_actual en nodos_visitados
  por cada operador:
    nodo_hijo = operador(nodo_actual)
    si nodo_hijo no en nodos_visitados ni nodos_frontera:
      introducir nodo_hijo en nodos_frontera
```

Figura 4 Pseudocódigo del algoritmo de búsqueda por amplitud–BFS [Serrano, 2012].

Además de las listas de nodos visitados y nodos frontera existe una lista que se encarga de almacenar los nodos padre con sus respectivos hijos. Esta lista sirve para obtener, a partir de una función programada en MIT App Inventor, el vector con los nodos que conforman el resultado.

Implementación de Algoritmo de Búsqueda por Costo Uniforme (UCS)

La organización de los datos en MIT App Inventor se realizó mediante listas, al igual que en el algoritmo BFS. En la figura 5 se presenta un fragmento de la organización de datos del mapa de la figura 1 en una estructura de tipo lista. La lista principal llamada datos_estaciones contiene sublistas que muestran las conexiones entre cada estación y la distancia que corresponde.

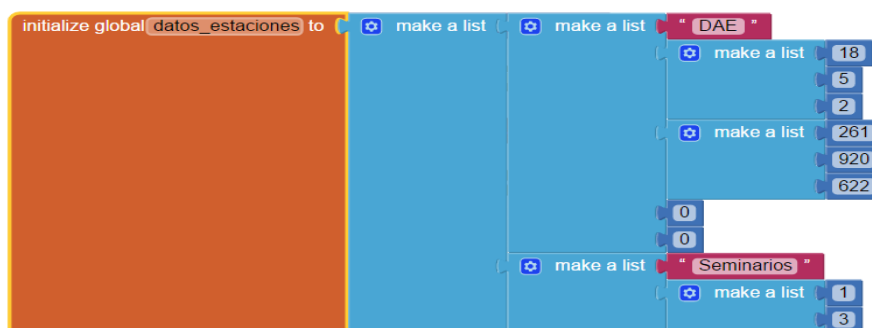


Figura 5 Datos del mapa en forma de lista en MIT App Inventor.

El pseudocódigo del algoritmo de UCS se muestra en la figura 6 [Serrano, 2012], en él se distingue el uso de dos listas, `nodos_frontera` y `nodos_visitados`. La primera corresponde a una cola con prioridad, cuyo uso marca la diferencia principal entre los métodos mostrados en el presente trabajo [Betali, 1999].

```
nodo_inicial = estado inicial
nodos_frontera = Cola con prioridad
nodos_visitados = Lista
almacenar nodo_inicial en nodos_frontera
mientras nodos_frontera no vacío:
    ordenar la lista de nodos_frontera según el costo
    nodo_actual = extraer el primer nodo de nodos_frontera
    si nodo_actual == solución

    introducir nodo_actual en nodos_visitados
    por cada operador:
        nodo_hijo = operador(nodo_actual)
        si nodo_hijo no en nodos_visitados:
            si nodo_hijo en nodos_frontera:
                si coste de nodo_hijo < nodo en nodos_frontera:
                    sustituir nodo_hijo en nodos_frontera
            si no
                introducir nodo_hijo en nodos_frontera
```

Figura 6 Pseudocódigo de búsqueda por costo uniforme—UCS [Serrano, 2012].

3. Resultados

Una vez finalizada la implementación en MIT App Inventor, se procedió a verificar que los resultados obtenidos fueran los esperados; las pruebas se realizaron tomando en cuenta el mapa mostrado en la figura 1.

En las tablas 1 y 2 se muestran principalmente casos críticos, es decir, los casos cuyas estaciones de origen y destino se encuentran separadas de extremo a extremo en el campus. Estas pruebas también sirven para dar ejemplo de las diferencias que presentan ambos algoritmos.

Se evalúa también el tiempo de procesamiento de cada algoritmo para verificar cual llega a una solución más rápido, consumiendo menos recursos.

Las pruebas 1 a 7 describen casos críticos; en las últimas 3 se muestran ejemplos donde ambos métodos arrojan los mismos resultados, pero se puede observar una diferencia en el tiempo de procesamiento. En general cada prueba presenta una diferencia de 200 ms entre la búsqueda por amplitud y la búsqueda de costo uniforme, lo podemos notar en la tabla 2.

Tabla 1 Resultados obtenidos por ambos algoritmos para 10 rutas propuestas.

Prueba N°	Origen	Destino	Ruta más corta (Búsqueda por costo uniforme-UCS)			Ruta con menos estaciones (Búsqueda por amplitud-BFS)			Diferencia entre distancia recorrida [m]	Diferencia entre n° de estaciones	Diferencia entre tiempo de procesamiento [ms]
			Distancia [m]	N° de estaciones	Tiempo de procesamiento [ms]	Distancia [m]	N° de estaciones	Tiempo de procesamiento [ms]			
1	Acceso 14 Sur (12)	DAE (1)	1544	7	283	1727	5	42	183	2	241
2	DAE (1)	Electrónica (8)	1197	5	258	1380	3	47	183	2	211
3	Ingenierías (8)	DAE (1)	1302	6	322	1485	4	59	183	2	263
4	Computación (11)	Cultura Física (18)	1365	6	267	2070	6	70	705	0	197
5	Acceso 18 Sur (13)	Seminarios (2)	1353	4	364	2124	4	62	771	0	302
6	Computación (11)	DAE (1)	1626	7	268	1809	5	46	183	2	222
7	Seminario (2)	Acceso 14 Sur (12)	1578	6	292	2349	6	85	771	0	207
8	Computación (11)	Derecho (16)	744	4	136	744	4	46	0	0	90
9	Terminal STU (5)	Cultura Física (18)	1181	3	345	1181	3	47	0	0	298
10	Ingenierías (8)	DGIE (15)	704	4	247	704	4	65	0	0	182

Tabla 2 Valoración de usuarios.

Estudiante	Efectividad	Eficiencia	Satisfacción	Promedio
1	10	10	10	10
2	10	10	9	9.67
3	9	10	9	9.33
4	10	10	10	10.00
5	9	9	8	8.67
6	10	10	9	9.67
7	9	10	10	9.67
8	9	10	10	9.67
9	9	9	9	9.00
10	9	10	10	9.67
Calificación promedio de App				9.53

En la figura 7 se observa el ejemplo 1, la aplicación presenta la ruta sugerida, la distancia y el tiempo de procesamiento para cada uno de los algoritmos.



Figura 7 Resultados de la prueba número 1.

Por otro lado, se pidió la evaluación de la aplicación a 10 estudiantes de nuevo ingreso, después del uso de la aplicación por una semana, tomando en cuenta 3 aspectos principales:

- Efectividad: ¿La búsqueda de la ruta más corta se realiza con éxito?
- Eficiencia: ¿El tiempo de ejecución es el idóneo?
- Satisfacción: ¿La aplicación agrada al usuario?

En base a los aspectos anteriores se demandó calificar la aplicación en escala del 1 al 10 para cada una de las 3 características mencionadas, siendo 1 el incumplimiento de dicha propiedad y 10 la satisfacción total del usuario en dicho aspecto.

Se puede decir que se tuvo una buena aprobación por parte de los estudiantes. Tuvimos algunas recomendaciones destacando que sería idóneo que se proporcionara esta aplicación móvil al momento de realizar la inscripción al servicio LOBOBICI.

4. Discusión

A partir de los datos recopilados en la tabla 1 y 2 se observa que cada algoritmo implementado en MIT App Inventor cumple con su cometido. La ruta más corta, dada por el algoritmo de BFS, entrega el resultado sin tomar en cuenta el número de estaciones. La ruta con menos estaciones, dada por el algoritmo de UCS, arroja el resultado sin tener en cuenta la distancia de la ruta.

Se tomarán 3 ejemplos de la tabla 1 para discutir los resultados obtenidos. En primer lugar, se tienen los casos similares a la prueba 1. El origen es la estación Acceso 14 Sur (estación número 12 en la figura 1), el destino es la estación DAE (estación número 1 en la figura 1). Aplicando ambos métodos se obtienen resultados que se diferencian entre sí de manera notable tanto en distancia como en número de estaciones.

También existen los casos parecidos al de la prueba número 4. Se desea ir de la estación Computación (11) a la estación Cultura Física (18). Las rutas entregadas por los algoritmos si bien son distintas, tienen el mismo número de estaciones, la

diferencia se remarca en la distancia recorrida para cada una. Esto da a entender que aunque una ruta tenga el mínimo de estaciones, no asegura que la distancia sea menor; de manera parecida, el obtener la ruta con la distancia más corta no asegura que el número de estaciones sea el mínimo, como se observa en el ejemplo anterior.

Finalmente se toma el ejemplo de la prueba 9 que busca la ruta entre Terminal STU (5) y Cultura Física (18). Los casos semejantes a éste entregan resultados idénticos para ambos algoritmos, es decir, la ruta propuesta es la solución óptima sin importar si el criterio es la distancia o el número de estaciones. Estas soluciones se dan especialmente cuando las estaciones de origen y de destino se encuentran cercanas.

5. Conclusiones

Los métodos evaluados en este trabajo presentan características similares, de hecho, el algoritmo de búsqueda por amplitud es un caso particular del algoritmo de costo uniforme donde todos sus nodos son iguales. A pesar de ello, la búsqueda por amplitud tuvo mejor desempeño ya que la propiedad principal a evaluar es la distancia que se debe recorrer y este algoritmo tiende a arrojar la menor, esto se refleja en menos consumo de memoria y tiempo de procesamiento en el dispositivo móvil.

De nuestro estudio, podemos destacar la resolución de un problema cotidiano con la ayuda de algoritmos desarrollados mucho tiempo atrás, brindándole al universitario una herramienta eficaz que evita pérdidas de tiempo en su rutina, aunado a su portabilidad y el bajo consumo de recursos en el celular. Del mismo modo nos permite observar el comportamiento de cada método en una aplicación móvil que no había sido evaluada de esta forma hasta ahora en dispositivos de esta índole.

El siguiente paso es desarrollar una aplicación en la que se resuelva un problema de mayor dificultad con el objetivo de comparar más algoritmos de búsquedas no informadas e informadas. Solo así se podrá brindar un mayor panorama para que los interesados puedan optar por una u otra opción.

6. Bibliografía y Referencias

- [1] Betali, J. Lecture Notes on Cognitive Science and Search Algorithms. University of California, San Diego. Noviembre, 1999.
- [2] Brooks, R. Intelligence without representation. *Artificial Intelligence*, Volume 47. Enero, 1991.
- [3] Chandel, A., & Sood, M., Searching and optimization techniques in Artificial Intelligence: A comparative study & complexity analysis. *International Journal of Advanced Reserch in Computer Engineering and Technology*. Marzo, 2014.
- [4] Chiong, R., & Hadi, J., & Japutra, W., A comparative study on informed and uninformed search for intelligent travel planning in Borneo Island. Sarawak, Malasya. 2008.
- [5] Gallo, G., & Pallottino, S., Shortest path methods: a unified approach. *Mathematical Programming Study*. 1986.
- [6] Guerriero, F., & Musmanno, R., Label correcting methods to solve multicriteria shortest path problems. *Journal of Optimization Theory and Applications*. 2001.
- [7] Korf, R. E. Depth-first iterative-deepening, An optimal admissible tree search. *Artificial Intelligence*. 1985.
- [8] Korf, R. E., Scientific paper on Artificial Intelligence Search Algorithms. University of California. Junio, 1999.
- [9] Kumar, V., & Ramesh, K., & Rao, V., Parallel Best-First Search of State-Space Graphs: A summary of results. Vol. 88, 1988.
- [10] Kuruvilla, M., & Mujahid, T., & Mohana, R., Experimental comparison of uninformed and heuristic AI Algorithms for N puzzle solution. Swinburne University of Technology. Kuching, Malasya. Enero, 2014.
- [11] Serrano, A. *Inteligencia Artificial: Fundamentos, prácticas y aplicaciones*. RC Libros, 2012.

SISTEMA DE RECONOCIMIENTO DE VOCALES DE LA LENGUA DE SEÑAS MEXICANA

Eliúh Cuecuecha Hernández

Benemérita Universidad Autónoma de Puebla
eliuhcueh@gmail.com

José Javier Martínez Orozco

Benemérita Universidad Autónoma de Puebla
javier.35.93.01@gmail.com

Daniel Méndez Lozada

Benemérita Universidad Autónoma de Puebla
daniel.mendezl@alumno.buap.mx

Adán Zambrano Saucedo

Benemérita Universidad Autónoma de Puebla
zambranos.adan@gmail.com

Aldrin Barreto Flores

Benemérita Universidad Autónoma de Puebla
aldrin.barreto@correo.buap.mx

Verónica Edith Bautista López

Benemérita Universidad Autónoma de Puebla
vbautista@cs.buap.mx

Salvador Eugenio Ayala Raggi

Benemérita Universidad Autónoma de Puebla
saraggi@ece.buap.mx

Resumen

La lengua de señas es un medio de comunicación tan importante como la lengua hablada para el desarrollo social del ser humano. Tras la aparición de

sensores de reconocimiento de gestos, como *Kinect*, surge especial interés por utilizarlos para interpretar la lengua de señas.

La finalidad del presente trabajo consistió en interpretar las vocales de la lengua de señas mexicana identificadas por gestos estáticos con la mano. Para ello se utilizó el sensor *Leap Motion Controller*, ideal para esta aplicación al detectar y seguir las manos con tal precisión sin necesidad de entrar en contacto con el usuario. Para lograr el reconocimiento de los gestos correspondientes a las vocales se utilizó el modelo de perceptrón multicapa junto a una interfaz visual en tiempo real. La red fue entrenada y calibrada por un experto en lenguaje de señas, logrando así una razón de reconocimiento de hasta 100%.

Palabras Claves: Leap Motion Controller, Lengua de Señas Mexicana (LSM), perceptrón multicapa, reconocimiento de Imágenes.

Abstract

The Sign Language is a communication mean as important as the speaking language, for the social development of the human being. Thus, after gesture recognition sensors emerged such as Kinect, interest arises for utilize them to interpret the Sign language.

The present work aims to interpret the Mexican Sign Language vowels identified by one-hand static gestures. In order to do so, the Leap Motion Controller was used, ideal for this application while detecting and tracking hand position with such accuracy, without the need to physically interact with the user.

To achieve the vowels' gesture recognition, the multilayer perceptron model was used along with a real time visual interface. The network was trained and calibrated by a sign language expert, achieving a recognition ratio up to 100%.

Keywords: *Image Recognition, leap motion controller, Mexican sign language, multilayer perceptron model.*

1. Introducción

El lenguaje de señas es un medio de comunicación para gente con problemas de audición y de habla. Este lenguaje usa información visual mediante

movimientos con los dedos, la mano y el brazo. El lenguaje de señas mexicano tiene el alcance de representar las 27 letras del alfabeto con una sola mano y así proveer un medio de comunicación de ayuda a personas con discapacidad del habla o audición [Serafín, 2011].

Sin embargo, interpretar el lenguaje de señas resulta complicado para la gran mayoría de personas no relacionadas con esta lengua. Por lo tanto, ante la imposibilidad de que las personas con discapacidad auditiva utilicen la palabra hablada, se plantea desarrollar un sistema confiable y fácil de usar para el reconocimiento de la Lengua de Señas Mexicana, y así proveer una plataforma de interfaz natural que facilite la comunicación. Más aún, esta plataforma también sirve para facilitar el proceso enseñanza-aprendizaje de la lengua de señas.

Múltiples trabajos de investigación se han llevado a cabo con el *Leap Motion Controller* para desarrollar aplicaciones de traducción del lenguaje de señas. En particular, [Naglot, 2015] presenta un notable trabajo con razón de reconocimiento de 96.15%. Aquí se plantea un sistema de reconocimiento en tiempo real del lenguaje de señas americano (ASL). Para ello, se procesan 520 muestras de letras del alfabeto mediante una red neuronal. Los datos de entrada comprenden distancias euclidianas entre diferentes partes de la mano, y a la salida se obtienen activaciones correspondientes a las 26 letras del alfabeto a identificar.

En [Fok et al., 2015] se muestra que utilizando más sensores es posible mejorar el reconocimiento de la Lengua de Señas. En este trabajo se usan dos sensores que en conjunto determinan la posición estática de la mano correspondiente a los 9 dígitos del ASL. La información se procesa mediante el Modelo Oculto de Markov (HMM) y muestra un mínimo de 68.78% de reconocimiento para un sólo sensor y un mínimo de 84.68% utilizando dos sensores. El experimento demuestra que es posible desarrollar un sistema de reconocimiento robusto utilizando más sensores. Por otra parte, en [Chuan, 2014] se muestra el uso de Machine Learning para el reconocimiento del alfabeto del ASL. No obstante, en este estudio la máxima razón de reconocimiento es de 72.78% y la aplicación del sistema está pensada en auxiliar a instructores en la enseñanza de la Lengua de Señas. Aquí se plantea, además, incorporar una cámara web para mejorar el reconocimiento de gestos.

De los trabajos anteriores resulta que la mayor razón de reconocimiento se obtiene con una red neuronal de perceptrón multicapa. Por esta razón, en el presente trabajo se decidió utilizar este modelo en conjunto con el *Leap Motion Controller* para el reconocimiento de vocales de la Lengua de Señas Mexicana (LSM), en vez del ASL como la mayoría de trabajos plantean.

2. Métodos

El Leap Motion Controller es un pequeño dispositivo USB diseñado para que el usuario pueda manejar una computadora mediante gestos manuales. El dispositivo, mostrado en la figura 1, consiste en un sensor de movimiento en 3D capaz de detectar la trayectoria de las manos, dedos y objetos similares, reportando su posición y movimiento.

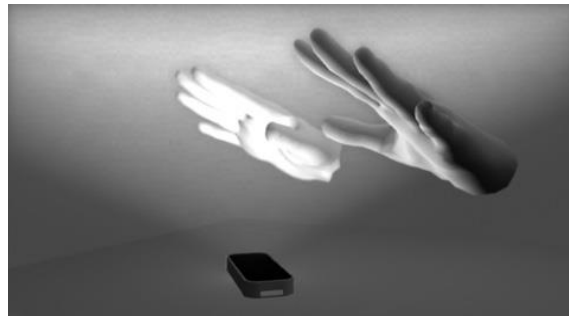


Figura 1 Detección del Leap Motion Controller [Leap Motion, 2017].

La figura 2 muestra las principales características que se pueden obtener a través del ambiente de programación de *Leap Motion Controller*. Entre estas se encuentran:

- Mano: Este modelo entrega información sobre el tipo (derecha, izquierda, ambas manos), posición (la posición central de la palma en milímetros), velocidad de movimiento, etc.
- Dedos: Características relacionadas con dirección (descrita por un vector unitario), longitud del dedo (en milímetros), ancho, posición de la punta, posición de la falange proximal (primera falange), posición de la falange distal (última falange) y la posición del metacarpo.

- Gestos: Algunos patrones de movimiento son reconocidos por el sensor, por ejemplo, el trazado de un círculo (*circle*) con un dedo, o el movimiento lineal que se interpreta como un deslizamiento (*swipe*), o el gesto de teclear (*key tap*).

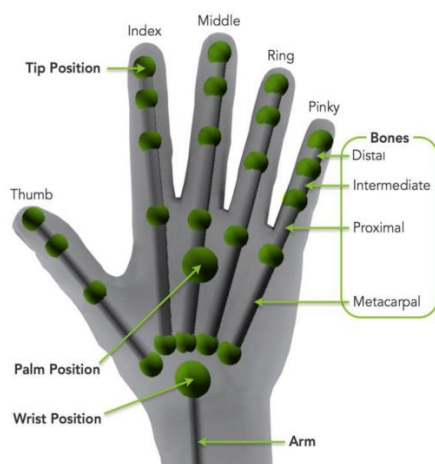


Figura 2 Partes de la mano identificadas por Leap Motion Controller [Davis, 2017].

Gracias a las múltiples variables de reconocimiento del *Leap Motion*, varios trabajos relacionados a identificación y clasificación se han llevado a cabo con este dispositivo [Erdogan et al., 2016], [Funasaka, 2015] y [Nájera et al., 2016].

De especial interés son las posiciones de los dedos (*tip position*), pues no importando la orientación de la mano, la representación de vocales en el LSM depende únicamente de la configuración espacial de los dedos. Por esta razón se decidió tomar como datos a ciertas distancias clave, pues resultan ser suficientes para poder discernir entre las distintas vocales. El *Leap Motion* calcula con gran precisión las puntas de los dedos gracias a los algoritmos de reconocimiento implementados en este dispositivo. Algunos algoritmos remarcables se pueden consultar en [Langford et al., 2016] y [Priego, 2012].

La información proporcionada a la red consiste entonces en los siguientes datos:

- Las distancias euclidianas entre la posición de la punta de cada dedo y la posición de la palma de la mano (5 datos).
- Las distancias euclidianas entre la posición de la punta de cada dedo, de manera consecutiva (4 datos).

El conjunto de datos recabados consta de 100 muestras (20 para cada vocal), tomadas de 4 autores del proyecto y un intérprete traductor experto en lenguaje de señas, trabajador del DIF municipal de la ciudad de Puebla. La mitad de estas muestras sirve como conjunto de aprendizaje y la otra mitad como conjunto de validación. El conjunto de datos es normalizado antes de ser utilizado para el entrenamiento utilizando ecuación 1.

$$N(x) = \frac{1}{\max_x - \min_x} (x - \min_x) \quad (1)$$

La tabla 1 muestra los datos obtenidos para el gesto correspondiente a la vocal A. Estos datos corresponden a las distancias clave correspondientes de la mano, expresadas en milímetros. Este conjunto de datos junto con el del resto de las vocales se deben normalizar para poder ingresar a la red mediante la función dada por la ecuación 1. Los valores máximos y mínimos se obtienen del conjunto de datos de todas las vocales.

Tabla 1 Datos de posición para la vocal A obtenidos con el Leap Motion Controller.

<i>Patrón</i>	Palma Pulgar	Palma Índice	Palma Medio	Palma Anular	Palma Menique	Pulgar Índice	Índice Medio	Medio Anular	Anular Menique
1	166.09	49.90	49.34	51.70	56.16	153.66	18.23	29.44	22.97
2	178.47	49.08	48.56	50.98	50.27	160.31	12.06	27.47	23.64
3	158.74	51.75	50.48	55.92	57.39	144.52	17.14	30.19	25.18
4	168.79	48.01	45.13	46.14	56.65	149.69	18.30	31.94	26.78
5	185.49	56.70	46.65	43.74	44.66	149.92	24.88	27.88	22.25
6	168.83	48.31	48.15	50.57	54.36	154.29	15.29	31.16	24.12
7	173.08	45.66	41.89	42.90	51.40	151.38	16.72	30.20	26.52
8	176.20	46.25	44.54	48.98	52.84	156.47	14.68	29.65	26.39
9	179.37	47.98	43.58	39.42	46.78	150.09	13.96	30.42	28.41
10	182.66	49.65	50.69	49.17	49.90	167.71	13.53	27.67	22.65
11	177.26	53.06	56.44	57.40	56.08	172.62	13.41	26.08	22.24
12	174.21	46.17	44.75	46.54	50.00	155.42	15.88	27.58	24.43
13	183.74	48.48	49.08	48.58	49.24	167.35	14.39	27.88	22.73
14	184.56	50.75	51.90	50.07	49.81	166.95	14.87	28.02	23.19
15	183.15	59.02	58.12	55.29	50.43	162.69	15.58	26.24	24.12
16	175.37	57.01	53.00	47.51	46.96	146.45	17.36	29.12	25.11
17	168.78	56.17	50.80	45.62	43.56	140.81	18.88	25.47	25.57
18	169.64	58.02	49.56	41.05	40.71	133.41	19.41	27.32	27.39
19	173.22	63.04	56.32	47.71	44.56	137.15	19.95	29.86	25.77
20	149.15	53.75	34.61	27.31	44.42	103.92	25.29	32.25	31.15

Para la clasificación se usa una red neuronal de perceptrón multicapa cuyo aprendizaje es supervisado y cuyo modelo consiste en prealimentar la red (*feedforward*). Esto es, se usan las características del conjunto de entrada y se obtiene una salida particular para la vocal que ha sido introducida. Para el entrenamiento se recurrió al uso del algoritmo de retropropagación de los errores, observando la evolución del error variando la razón de aprendizaje. Este consiste en introducir un conjunto inicial a la red y asignar valores aleatorios a cada uno de los pesos. Con base en estos pesos, se obtiene cada una de las salidas en las diferentes capas hasta llegar a la capa de salida, calculando así el error obtenido a la salida mediante la ecuación 2 y propagando este valor hacia cada una de las capas anteriores (ocultas y de entrada), modificando así los valores de los pesos de cada una de las interconexiones correspondientes mediante las ecuaciones 3, 4, 5 y 6 utilizando la función de activación sigmoideal.

Cálculo del error a la salida para un patrón n siendo $y(n)$ las salidas de la red y $s(n)$ las salidas deseadas, ecuación 2.

$$e(n) = \frac{1}{2} \sum_{i=1}^{n_c} (s_i(n) - y_i(n))^2 \quad (2)$$

Cálculo de los pesos capa oculta $C - 1$ a la capa de salida C , ecuación 3.

$$w_{ji}^{C-1}(n) = w_{ji}^{C-1}(n-1) + \alpha \delta_i^C(n) \alpha_j^{C-1}(n) \quad (3)$$

Para $j = 1, 2, \dots, n_{C-1}$ y $i = 1, 2, \dots, n_C$

Cálculo de umbrales de la capa de salida, ecuación 4.

$$u_i^C(n) = u_i^C(n-1) + \alpha \delta_i^C(n) \quad (4)$$

Para $i = 1, 2, \dots, n_C$

Con:

$$\delta_i^C(n) = -(s_i(n) - y_i(n))y_i(n)(1 - y_i(n))$$

Cálculo de los pesos de la capa c a la capa $c + 1$, ecuación 5.

$$w_{kj}^c(n) = w_{kj}^c(n-1) + \alpha \delta_j^{c+1}(n) \alpha_k^c(n) \quad (5)$$

Para $k = 1, 2, \dots, n_c$, $j = 1, 2, \dots, n_{c+1}$ y $c = 1, 2, \dots, C - 2$

Cálculo de umbrales de las neuronas de la capa $c + 1$ para $c = 1, 2, \dots, C - 2$, ecuación 6.

$$u_j^{c+1}(n) = u_j^{c+1}(n - 1) + \alpha \delta_j^{c+1}(n) \quad (6)$$

Para $j = 1, 2, \dots, n_{c+1}$ y $c = 1, 2, \dots, C - 2$

Con:

$$\delta_j^{c+1}(n) = \alpha_j^c(n) (1 - \alpha_j^c(n)) \sum_{i=1}^{n_{c+1}} \delta_i^{c+2}(n) \omega_{ji}^c$$

La red neuronal artificial en cuestión consta de 3 capas: una capa de entrada, una capa oculta y una capa de salida, i.e. $C = 3$. Las neuronas de la capa de entrada se conectan a las neuronas de la capa oculta y éstas a la capa de salida, lo que lleva a la interconexión de los pesos en la red.

El tamaño de la capa de entrada depende de los datos que serán introducidos a la red, en este caso 9, es decir 9 neuronas en la capa de entrada ($n_1 = 9$). De manera similar para la capa de salida se tienen 5 posibilidades (5 vocales) por lo tanto 5 neuronas en la capa de salida ($n_3 = 5$). Para la capa oculta no existe una regla establecida que calcule el número de neuronas a usar, sin embargo, debido a la gran tasa de reconocimiento alcanzada en [Naglot, 2015] parece prudente utilizar la regla empírica expuesta allí: se suma el número de neuronas de la capa de entrada más el número de neuronas de la capa de salida y este resultado se divide entre dos. Es decir, $n_2 = \frac{5+9}{2} = 7$. La arquitectura de la red neuronal queda entonces cómo se muestra en la figura 3.

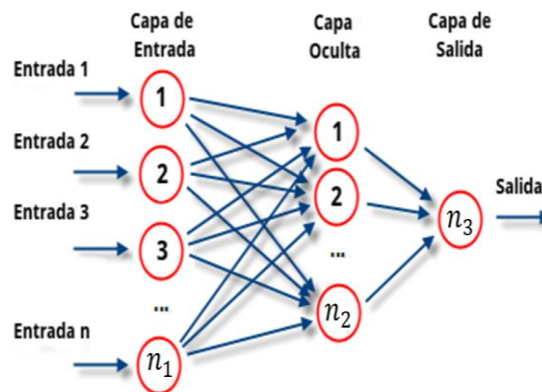
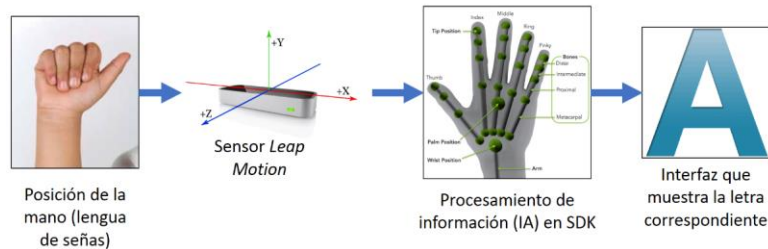


Figura 3 Modelo del perceptrón multicapa utilizado

3. Resultados

El sistema propuesto fue implementado en lenguaje Java. La figura 4 muestra un diagrama con el proceso utilizado para reconocer la Lengua de Señas.



[Chuan, 2014], [Langford et al., 2016], [Serafín, 2011]

Figura 4 Diagrama a bloques del flujo de información.

En particular, el programa de computadora desarrollado pretende ser fácil de usar. Para ello se muestra en pantalla una representación espacial de la mano vista por el *Leap Motion* (figura 5), con la finalidad de que el usuario evalúe si el reconocimiento del sensor es adecuado. Luego, para auxiliar en la realización del gesto, se muestra una imagen con la posición de la mano a realizar para cada vocal. Finalmente, a un lado se muestran las vocales que cambian de color azul a rojo, en dónde la letra más roja es la vocal identificada por el programa. El reconocimiento de la vocal se complementa con la reproducción del sonido correspondiente.

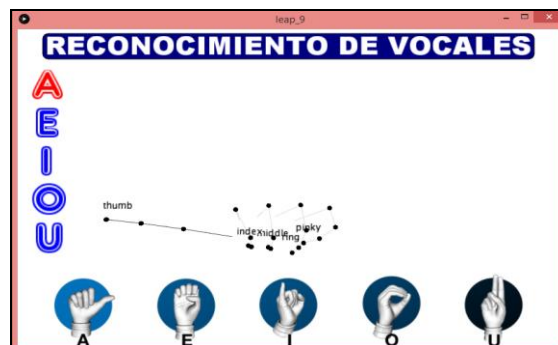


Figura 5 Interfaz del programa.

Cómo se mencionó anteriormente, del conjunto de datos obtenidos, se clasificaron las muestras en dos subconjuntos: entrenamiento y prueba. Durante el proceso de

entrenamiento, se realiza un super ajuste sobre dicho conjunto, concluyendo el proceso cuando se alcanza el primer mínimo de la función del error de validación. Cabe mencionar que para la estimación del error no se utilizó ni un método como validación cruzada o hold-out. El conjunto de entrenamiento se utilizó para determinar los parámetros del clasificador neuronal, mientras que el conjunto de prueba fue utilizado para estimar el error de generalización.

Una herramienta que permite visualizar el desempeño de un algoritmo que usa aprendizaje supervisado, como la red propuesta en este caso, es una matriz de confusión. La tabla 2 muestra una matriz de confusión para dos clases donde VP significa verdadero positivo, FP para falso positivo, VN verdadero negativo y FN falso negativo.

Tabla 2 Matriz de confusión para una red de dos clases.

		Valor predicho		
		Valor	Positivo	Negativo
Valor real	Positivo	VP	FN	P
	Negativo	FP	VN	N
	TOTAL	P'	N'	P+N

En esta matriz cada columna representa el número de predicciones de cada clase, mientras que la fila representa las clases reales. El principal beneficio de esta herramienta es que permite ver si la red está confundiendo dos clases.

La exactitud del sistema es el porcentaje de muestras que fueron correctamente clasificadas por la red. La precisión mide el grado de certitud del sistema, es decir, el error presente al momento de clasificar muestras de una sola clase, ecuación 7 y ecuación 8.

$$exactitud = \frac{VP+VN}{P+N} 100\% \tag{7}$$

$$precisión = \frac{VP}{VP+FP} 100\% \tag{8}$$

Para este caso en particular, las señas que representan a las vocales el LSM, no presentan un parecido notable entre ellas. Por esta razón es que para exactitud y precisión se tienen valores de 100%. Sin embargo, para mayor número de letras a identificar, estas razones disminuirán notablemente, tabla 3.

Tabla 3 Matriz de confusión del sistema propuesto.

		Valor predicho					TOTAL	VP	FP	Precisión
		A	E	I	O	U				
Valor real	A	10	0	0	0	0	10	10	0	100%
	E	0	10	0	0	0	10	10	0	100%
	I	0	0	10	0	0	10	10	0	100%
	O	0	0	0	10	0	10	10	0	100%
	U	0	0	0	0	10	10	10	0	100%
	TOTAL	10	10	10	10	10	50	50	0	100%
Razón de Reconocimiento = 100%										

4. Discusión

Como se estableció anteriormente, la exactitud y precisión del sistema resulta ser de 100% para ambos casos. Esto es, con base en la arquitectura diseñada del sistema, la cual busca identificar las vocales del LSM, la identificación de cada vocal presenta un error prácticamente nulo, lo cual dice que no existe confusión del sistema para clasificar las entradas.

Con base en trabajos similares sobre reconocimiento de lenguaje de señas, se observó que el caso de mayor razón de reconocimiento se da mediante el uso de una metodología similar: una red neuronal de arquitectura perceptrón multicapa. Cabe mencionar que dicha razón es muy alta, y por ende el sistema es muy eficiente.

Finalmente, el sistema tratado en el presente trabajo obtuvo una razón de reconocimiento de hasta el 100%. Esto se debió a diferentes factores como lo son: un número preciso y calculado de capas ocultas, validación y calibración adecuada de los patrones de entrada, capacidad de generalización alta, entre otros. Además, la representación en lenguaje de señas de las vocales en el LSM disminuye la malinterpretación de estas ya que no presentan similitudes mayores entre ellas, a diferencia de otras letras del alfabeto. En conjunto, estos factores propiciaron una eficacia casi absoluta del sistema, concluyendo que el uso de redes neuronales, en específico una arquitectura perceptrón multicapa, da resultados más precisos y exactos.

A diferencia de trabajos anteriores donde se utilizan otros métodos cuyos resultados, si bien son completamente funcionales, no se asemejan a los obtenidos en este trabajo. Además, la importancia de este trabajo radica en que el

reconocimiento se basa en el lenguaje de señas mexicano, el cual no ha sido tratado en gran detalle en otros trabajos, ya que la gran mayoría se basan en el lenguaje de señas americano.

Los resultados obtenidos dan pie a que se utilicen redes neuronales para la solución de problemas que requieran de un aprendizaje continuo del sistema basado en ciertos resultados deseados, y que sean aplicables en diferentes áreas.

5. Conclusiones

El reconocimiento de lenguaje de señas mexicano mediante el sensor *Leap Motion* brinda una interfaz amigable y práctica para poder interpretar el lenguaje de personas con discapacidad auditiva y verbal, y de esta manera darles un método de comunicación que simule una conversación hablada.

En conjunto con el sensor, la arquitectura de perceptrón multicapa fue esencial para la obtención de resultados. La eficacia de este método radica en el cálculo preciso de número de capas ocultas, así como en el uso del método *backpropagation* para su entrenamiento, propiciando que el sistema obtenga una capacidad de generalización alta. Esto se refleja en la facilidad de relacionar la entrada con la salida adecuada, a pesar de algunas similitudes que puedan presentar distintas entradas.

Gracias a la razón de reconocimiento obtenida, se comprueba que el *sensor Leap Motion* resulta ideal para el tipo de aplicación, debido a su facilidad de uso y precisión de reconocimiento 3D de la mano.

Algunas aplicaciones en las que se puede implementar dicho trabajo son: reconocimiento de enfermedades en la mano, fisioterapia personalizada de la mano [Santos et al., 2015], plataforma de voz para personas sordomudas, corrección de posturas, enseñanza de lenguaje de señas, entre otras.

6. Bibliografía y Referencias

- [1] Chuan, C. H. y Guardino, C., American Sign Language Recognition Using Leap Motion Sensor, International Conference on Machine Learning and Applications, 2014.

- [2] Davis A., Getting Started with the Leap Motion SDK. Leap Motion. <http://blog.leapmotion.com/getting-started-leap-motion-sdk/>.
- [3] Erdogan, K., Durdu, A., y Yilmaz, N., Intention Recognition Using Leap Motion Controller and Artificial Neural Networks, CoDIT, Malta, 2016.
- [4] Fok, K. Y., Ganganath, N., Cheng, C. T., y Tse, C. K. A Real-Time Recognition System Using Leap Motion Sensors. Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, 2015.
- [5] Funasaka, M., Ishikawa, Y., Takata, M., y Kazuki, J., Sign Language Recognition using Leap Motion Controller, PDTA, 2015.
- [6] Langford Cervantes, J., Alencastre Miranda, M., Muñoz Gómez, L., Navarro Hinojosa, O., Echeverría Furio, G. Manrique Juan, C., y Maqueo, M. Detección y seguimiento de palmas y puntas de los dedos en tiempo real basado en imágenes de profundidad para aplicaciones interactivas, Research in Computing Science, pp. 137-149, 21 Marzo 2016.
- [7] Leap Motion. API Overview: https://developer.leapmotion.com/documentation/csharp/devguide/Leap_Overview.html.
- [8] Naglot, D., y Kulkarni, M., Real Time Sign Language Recognition Using the Leap Motion Controller, Vishwakarma Institute of Technology.Pune, India, 2015.
- [9] Nájera Romero, L. O., López Sánchez, M., González Serna, J. G., Pineda, T. R., y Arana Llanes, J. Y. Recognition of Mexican Sign Language through the Leap Motion Controller, CSC, 2016.
- [10] Priego Pérez, F. P. Reconocimiento de Imágenes del Lenguaje de Señas Mexicano, Centro de Investigación en Computación IPN, México, 2012.
- [11] Santos, A., Guimaraes, V., Matos, N., Cevada, J., Ferreira, C., y Sousa, I., Multi-sensor Exercise-based Interactive Games for Fall Prevention and Rehabilitation, 9th International Conference on Pervasive Computing Technologies for Healthcare, 2015.
- [12] Serafín de Fleischmann, M. E., González Pérez, R., Manos con Voz, CONAPRED, México, 2011.

PROPUESTA DE UN ENTRENADOR MIOELÉCTRICO BASADO EN UNA APLICACIÓN MÓVIL

Humberto de la Cruz Regalado

Universidad Autónoma de Aguascalientes

betocrur666@hotmail.com

Carlos Edgar López Barrera

Universidad Autónoma de Aguascalientes

atlascelba@hotmail.com

Eduardo Emmanuel Rodríguez López

Universidad Autónoma de Aguascalientes

edwral@hotmail.com

Luis Mariano Sandoval González

Universidad Autónoma de Aguascalientes

mariano.sandovalg@hotmail.com

Alfredo Ramírez García

Universidad Autónoma de Aguascalientes

argarcia@correo.uaa.mx

Resumen

En este artículo se desarrolló la propuesta de un entrenador mioeléctrico basado en una aplicación móvil. La importancia de este tipo de sistemas en el área de rehabilitación es su utilidad en la adaptación de un usuario para la manipulación de prótesis mioeléctricas. En particular la propuesta de este artículo está orientada a prótesis mioeléctricas de mano de un grado de libertad.

El sistema utiliza la actividad eléctrica de los músculos o también llamada señal mioeléctrica, y mediante su procesamiento se obtiene la envolvente de la señal cuyo valor se utilizó como variable de control.

Se consideró la importancia de este tema debido a que las señales electromiográficas de cada individuo pueden variar, ya que las fibras musculares de una persona pueden generar más energía que las de otra, y por lo tanto la necesidad para el usuario no debe depender de un estándar establecido. Se describirán las metodologías variadas, las estructuras de procesamiento de la señal, así como el diagrama a bloques del circuito propuesto, los resultados y las conclusiones donde se discute la viabilidad del prototipo para su uso en el futuro en el área de rehabilitación y en el manejo de prótesis.

Palabras Claves: Electromiografía, envolvente, procesamiento de señales, servomotor.

Abstract

In this article a proposal of a myoelectric trainer based on mobile application is presented. This kind of systems is useful in the rehabilitation of users of prosthetic devices. The trainers facilitate the use of myoelectric prostheses. In particular the proposed myoelectric trainer is oriented to hand prostheses with one degree of freedom. In the system developed the myoelectric signal recorded from upper limb muscles was used as source of information and the envelope obtained by means of signal processing was used as a control variable.

The importance of this topic was considered because of the electromyographic signals of each individual can vary, since a person's muscle fibers can generate more energy than others, and therefore the need for the user should not depend on an established standard. The varied methodologies will be presented, the signal processing structures, as well as the block diagram of the proposed circuit, the results and the conclusion where it will be mentioned the viability of the prototype for its future use in the rehabilitation and handling of prostheses.

Keywords: *Electromyography, envelope, servomotor, signal processing.*

1. Introducción

En 1849, Du Bois Reymond fue la primera persona en demostrar la actividad eléctrica del músculo humano durante la contracción voluntaria, al conectar la

mano de un sujeto a las agujas de un galvanómetro. Observó que cuando el sujeto flexionaba su brazo, la aguja se deflataba, el grado de deflexión aumentaba con la fuerza de contracción [Ramírez, 2002].

Por otro lado, en 1929, a la electromiografía o registro de la actividad eléctrica de los músculos se le dio un sentido diagnóstico, cuando Adrian y Bronk lo utilizaron para estudiar la organización funcional de los movimientos, en el diagnóstico diferencial de atrofas neurógenas y miógenas [Ramírez, 2002].

Los usos de la electromiografía hoy en día son muy variados, contemplando, por ejemplo, la valoración del dolor lumbar [JL, 2009], comparación de rendimiento físico en ciertos ejercicios [Morant-Arilla, 2015], el procesamiento de la señal como medio de control de una prótesis [Campo, 2007], o el estudio de la actividad muscular en general. La señal electromiográfica (EMG) refleja el reclutamiento de las unidades motoras que se activan, esto es, cuando hay una mínima actividad voluntaria se activa un número pequeño de unidades y mientras el esfuerzo muscular se va incrementando el número de unidades motoras se incrementa. Esto muestra que la respuesta muscular es función del número de unidades motoras activas [Guyton, 2006].

Un entrenador mioeléctrico es un equipo que utiliza las señales EMG como fuente de información con el objetivo de establecer una retroalimentación relativa a la actividad muscular resultante de la actividad física que esté desarrollando un usuario.

Las aplicaciones de estos equipos se orientan principalmente a la rehabilitación de grupos musculares que presentan algún tipo de atrofia, como es el caso de los músculos involucrados en parálisis faciales, en problemas de incontinencia, o bien en la marcha.

Otra área de aplicación muy importante dentro de la rehabilitación es el entrenamiento para uso de dispositivos protésicos de extremidad superior [Barraza, 2010], [Ramírez, 2006], [Clingman, 2014], [Dupont, 1994]. En esta área es hacia donde se dirige la propuesta del sistema desarrollado en este artículo.

La importancia de entrenar a los músculos de los miembros amputados con la finalidad de activar una prótesis mioeléctrica movida por un servomotor se debe a

que las señales mioeléctricas no son lo suficientemente fuertes como para realizar el movimiento controlado del motor. Debido a que el musculo puede llegar a fatigarse con mucha facilidad, la perdida de la señal causaría un mal funcionamiento de la prótesis [Barraza, 2010], [Ramírez, 2006], [Clingman, 2014], [Dupont, 1994].

Antiguamente, en Grecia daban gran importancia la terapia de la dieta, ejercicios corporales, masajes y baños de mar. [Loreto, 2014]. Posteriormente, se fueron descubriendo muchos métodos de terapia como fue la hidroterapia, donde Cornelio Celso empezó a escribir un libro de este tema. Seguido por los científicos Luis Galvani y Volta en la electroterapia. La gimnasia que fue empleada como terapia por Jerónimo Mercuriale. Comenzó la primera mitad del siglo XX y fue así cuando la fisioterapia fue base de la medicina física, llegando hacer una especialidad médica. [Loreto, 2014].

En las últimas décadas ha crecido la tecnología, mostrando también progreso en la rehabilitación de extremidades, llegando a crear terapia con robots o mecanismos que monitorean el avance en la movilidad del paciente. [Bundhoo, 2009].

Los robots de entrenamiento son usados para extremidades superiores e inferiores, que ayudan al paciente a practicar movimientos específicos para incrementar la movilidad y corregir los movimientos de la extremidad.

El MIT MANUS therapy robot es un brazo robótico que ayuda al paciente a mover el brazo a través de una tabla para mover un cursor en la pantalla. [Bundhoo, 2009].

Los dispositivos de monitores son autónomos y portables, y ayudan a simplificar el proceso de captura de datos fisiológicos, a la vez que ayudan a ejercitar el musculo con mínima intervención de un terapeuta. [Bundhoo, 2009].

Aplicaciones móviles en salud. Una aplicación móvil o app es un software diseñado para funcionar en smartphones o tablets. En los últimos años, el desarrollo de estas app ha ido entrando en el campo de la medicina, tanto para profesionales como para pacientes. De las “health apps” más populares en la población, son de la categoría de “dieta y fitness” [San, 2014].

Otro concepto importante de conocer, relacionado al área de la salud, es el de Exergames, los cuales pretenden estimular la movilidad del cuerpo entero mediante el uso de ambientes interactivos que simulan diferentes sensaciones de presencia [Trujillo, 2013].

En este trabajo se propone el desarrollo de un entrenador mioeléctrico basado en una aplicación móvil con el objetivo futuro de que sea utilizado y ayude en el entrenamiento de pacientes candidatos al uso de prótesis mioeléctricas de mano de un grado de libertad. Una ventaja importante de este sistema es la portabilidad que tiene al estar basado en una aplicación móvil. Así el posible usuario tiene la oportunidad de seguir su terapia de forma continua, aspecto importante en el proceso de adaptación del paciente con el dispositivo protésico.

2. Métodos

La implementación del sistema propuesto en este artículo se compone de un módulo de hardware y un módulo de software. El módulo de hardware se refiere a las etapas de la instrumentación electrónica necesarias para la captación y registro de la señal EMG del grupo muscular de interés. Este módulo se muestra en la figura 1. El módulo de software incluye los algoritmos implementadas tanto en el sistema embebido como en el dispositivo móvil.

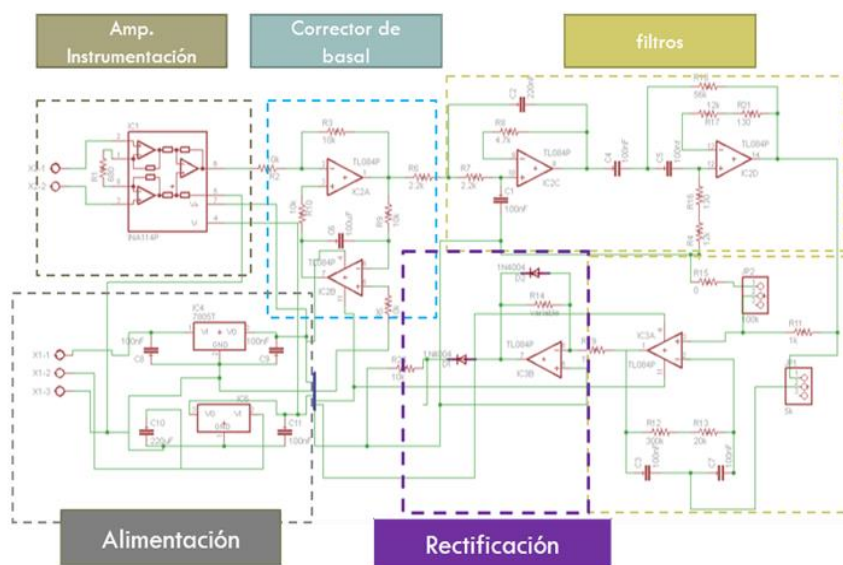


Figura 1 Etapas de la instrumentación electrónica del sistema propuesto.

Para la etapa de amplificación se usó el circuito integrado INA114AP, por su alta impedancia de entrada y un alto rechazo de modo común, mejorando la señal de entrada. Para su configuración se hizo uso de la ecuación 1 recomendada en la hoja de datos para obtener una ganancia de 11.6 [Texas, 2017].

$$G = 1 + \frac{50k\Omega}{4.7k\Omega} \quad (1)$$

En el diseño también se incluye un corrector de basal basado en un diferenciador y un integrador, que se utilizó para evitar un nivel de corriente continua en la señal, debido a los artefactos, y este se corrige cada segundo. Con los componentes estimados de acuerdo a la ecuación 2.

$$t = R * C = 10k\Omega * 100\mu F = 1 s \quad (2)$$

La etapa de filtrado, se basó en el ancho de banda de la señal EMG superficial que cae en el rango de 20 a 500 Hz. Por lo cual se implementaron filtros tipo Butterworth a -40 dB, un filtro pasa bajas de 500 Hz y un pasa altas de 20 Hz, al igual que un filtro tipo Notch para 60 Hz [Ramírez, 2005].

Dado que el procesamiento de la señal EMG incluye la envolvente de la misma, se determinó utilizar un rectificador de precisión, al mismo tiempo dándole una amplificación, regulada por un potenciómetro de precisión. Cabe mencionar que la ganancia se calibra en función de la señal EMG que proporciona cada persona.

Para la etapa de alimentación se utilizaron dos reguladores de +-5 V, para lograr la alimentación del microcontrolador atmega328p y la instrumentación electrónica desarrollada para el registro de la señal EMG.

Para el procesamiento de la señal se utilizó el microcontrolador atmega328p, donde el módulo ADC se configuró con una frecuencia de reloj de 250 kHz, a una frecuencia de muestreo de 19.230 kHz y un tiempo de muestreo de 52 μ s.

Dado que la información se transmitió a un dispositivo móvil por medio de bluetooth, fue necesaria la comunicación serial por medio del protocolo RS-232 por lo que se inicializó la comunicación serial a una velocidad de 115200 baudios, modo asíncrono a doble velocidad, sin bit de paridad y 1 bit de stop [Atmel, 2017].

A partir de la envolvente de la señal se controló un servomotor generando una señal de modulación por ancho de pulso (PWM). Esta última señal se generó con el microcontrolador. Para lo cual, se inicializó el PWM en modo rápido, modo no invertido, y un preescalador de 1024 [Atmel, 2017].

El diagrama de flujo del algoritmo implementado para el sistema se muestra en la figura 2, el cual se basa en una máquina de estados. El sistema se inicializa con dos variables, una llamada "Auxiliar" cuya función es realizar un bucle en los tres estados como se observa en la figura 2, y una variable "Max" con la cual se calibra el sistema; los estados son seleccionados por una serie de botones dentro de una aplicación móvil, figura 6.

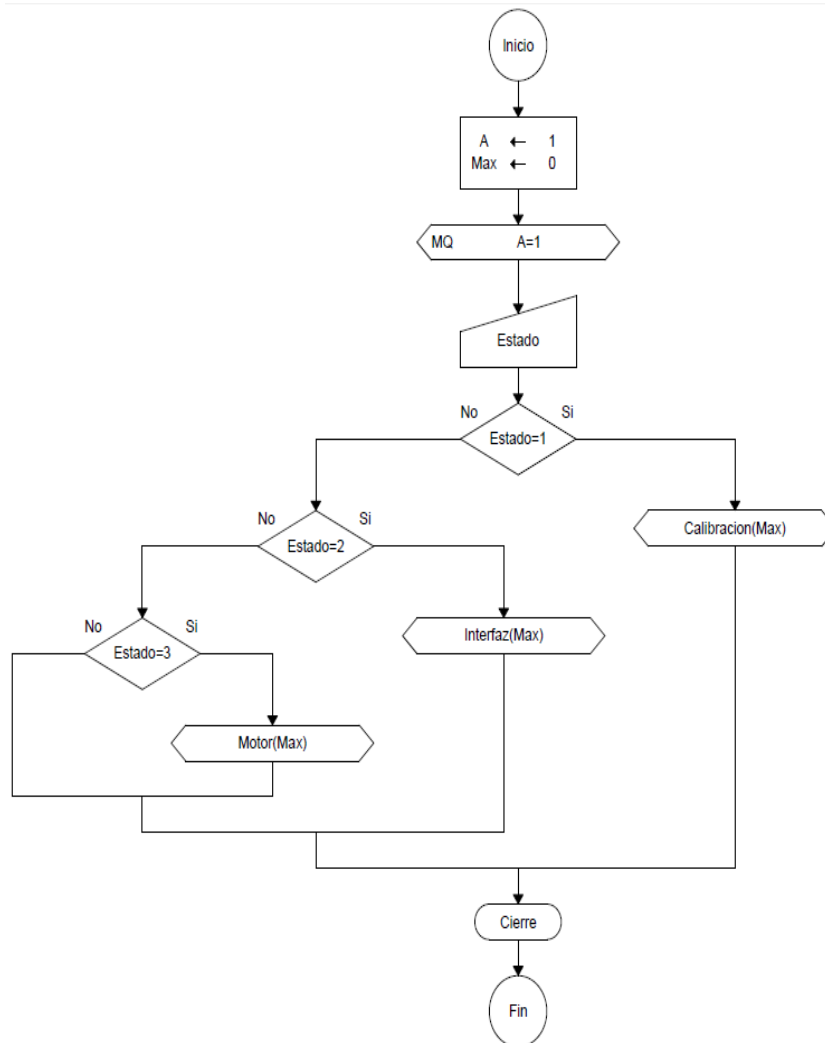


Figura 2 Máquina de estado para validar las diferentes funciones del sistema.

Al acceder a la función “Calibración”, figura 3, se procede a limpiar las variables de entrada y salida, y se inicializa el contador a doscientos; se procede a realizar la adquisición de muestras, calcular el promedio de ellas y posteriormente se almacenan en la variable “Max”.

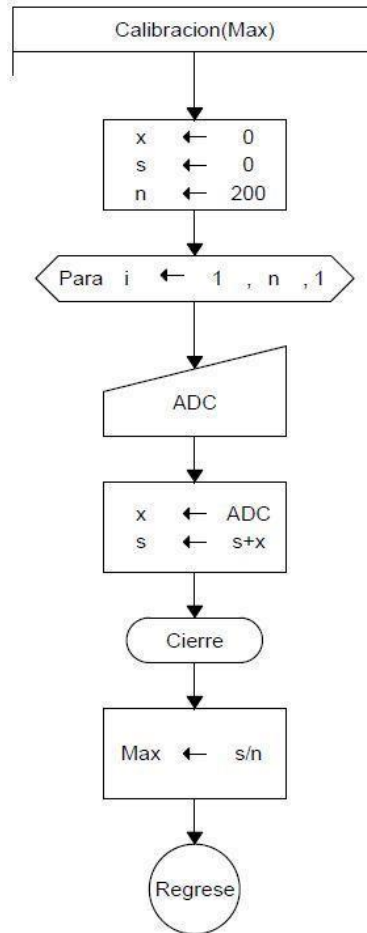


Figura 3 Primera función llamada Calibración.

Al acceder a la función “Interfaz”, figura 4, se procede a limpiar la variable “Envolvente” y el puerto de entrada, se declara un bucle el cual puede ser roto si la entrada es diferente de cero, dentro de él se adquieren los datos del ADC, en base a ellos se calcula la envolvente de la señal (valor RMS) la cual entrega como resultado la fuerza del usuario, se envía el valor por el puerto serial a través de la comunicación bluetooth a un dispositivo móvil (Android) en el cual se visualiza dicha fuerza implementada por el usuario en una barra de progreso, figura 5.

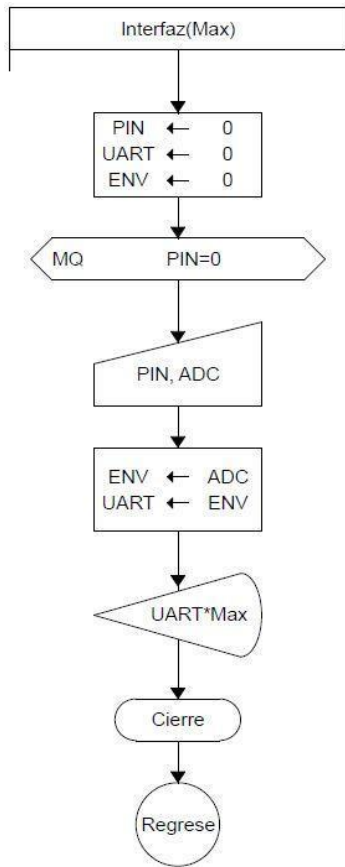


Figura 4 Segunda función llamada Interfaz.

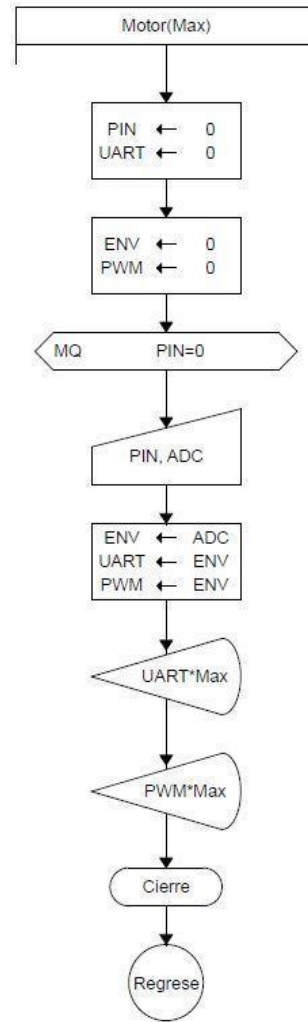


Figura 5 Tercera función llamada Motor.

Al acceder a la función “Motor”, figura 5, se procede a limpiar la variable “Envolvente”, el registro PWM y el puerto de entrada, se declara un bucle el cual puede ser roto si la entrada es diferente de cero, dentro de él se adquieren los datos del ADC, en base a ellos se calcula la envolvente de la señal (valor RMS) la cual entrega como resultado la fuerza del usuario, se envía el valor por el PWM el cual acciona un servomotor y simula físicamente el movimiento de una pinza; a su vez se logra visualizar la información y la simulación virtual de la pinza en un dispositivo móvil (Android) a través de una comunicación bluetooth, figura 6.

Para llevar a cabo el registro de la señal EMG se utilizó un canal diferencial con electrodos de plata cloruro de plata (AgCl/Ag) y con una separación entre ellos de

3 cm centro a centro. La colocación de los electrodos se hizo sobre el grupo de músculos superficiales (flexor superficial de los dedos) de la cara anterior del antebrazo relacionados a función de cierre y apertura de la mano.

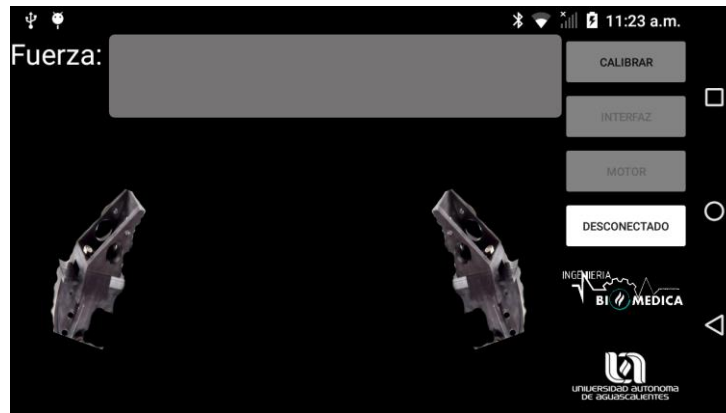


Figura 6 Esqueleto de la interfaz gráfica dentro de la aplicación móvil.

3. Resultados

La colocación de los electrodos sobre el sujeto de pruebas se muestra en la figura 7, el sujeto de prueba es una persona sana con sus extremidades superiores completas. Para posicionar los electrodos se le pidió que abriera y cerrara la mano y por medio de palpación se ubicó el lugar donde colocar los electrodos. Esto se determinó a partir de la actividad muscular observada en las señales EMG registradas.



Figura 7 Colocación de electrodos en la cara anterior del antebrazo.

Una vez teniendo el circuito de monitoreo de EMG se procedió a hacer pruebas, en la figura 8 se puede observar un ejemplo de la señal EMG del sujeto de

pruebas filtradas y amplificadas. Estas señales se observaron desde un osciloscopio Tektronix TED2012.

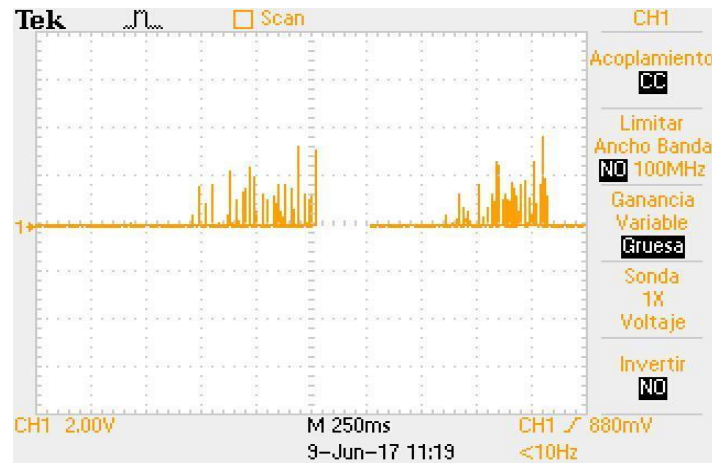


Figura 8 Salida final del circuito EMG.

En la figura 9 se le pidió al sujeto de pruebas que ejerciera una máxima contracción para obtener el máximo voltaje. Esto con el objetivo de hacer una calibración al sistema.

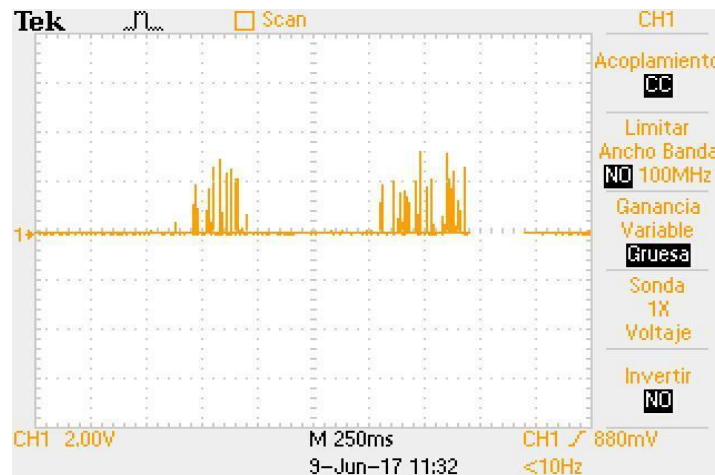


Figura 9 Amplitud de la señal depende de la fuerza de contracción del sujeto de prueba.

Esta misma señal fue observada mediante el monitor serial de Arduino visualizada en una computadora, como se observa en la figura 10, esto sirvió para comprobar que los datos adquiridos se estuvieran procesando por el microcontrolador.

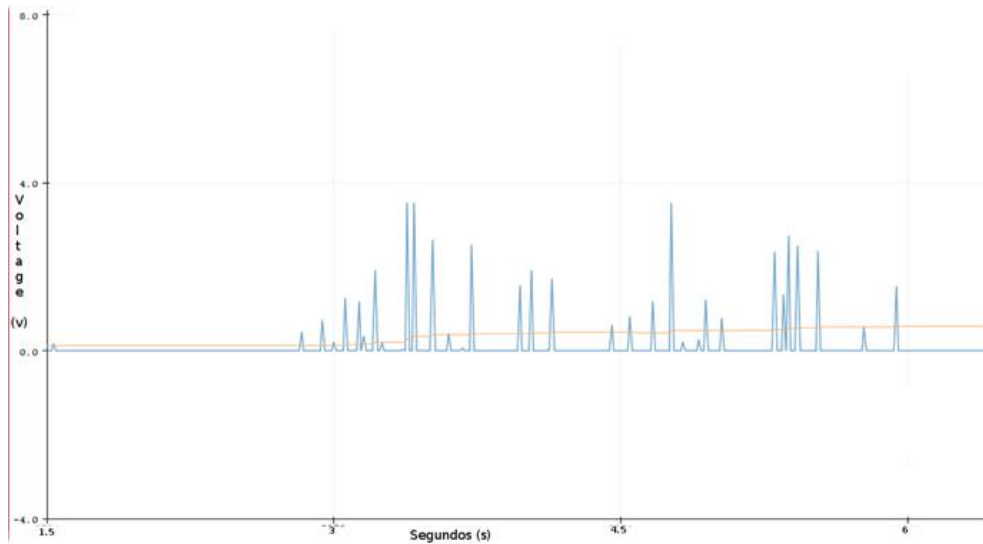


Figura 10 Visualización de la señal final del circuito EMG.

La señal es enviada a través del bluetooth a un celular con una aplicación móvil de diseño propio, en la cual se puede observar el esfuerzo y una representación de cómo se cerraría una pinza controlada por dicha señal. En esta interfaz se tienen los botones de la máquina de estados para calibrar, observar solo la interfaz, y mover el motor mientras se observa la interfaz (véase en la figura 6). En la figura 11 se observa una prueba en la cual el sujeto de prueba hizo una contracción leve la cual alcanzó un 20% del máximo adquirido.

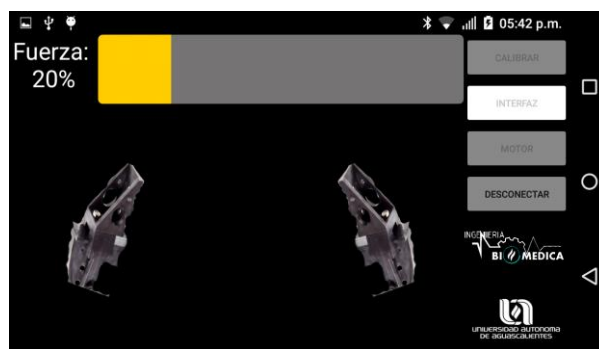


Figura 11 Captura de pantalla de la app funcionando.

Por último, se pidió al sujeto de prueba que generara una máxima contracción. En la figura 12 se observa la interfaz gráfica del celular donde muestra la adquisición de la señal al 100% y la pinza completamente cerrada.

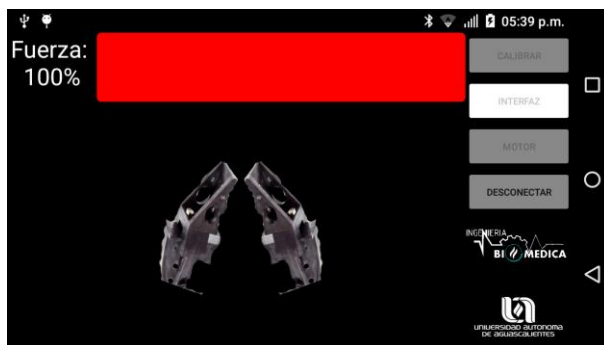


Figura 12 El sujeto de prueba ha alcanzado su contracción máxima al llegar al 100%, a la par, la pinza se cierra por completo.

4. Discusión

Tal cual se describe en la fisiología humana, cuando se realiza una contracción sostenida se reclutan más unidades motoras en el músculo para lograr un mayor esfuerzo físico [Gayton, 2006]. En base a esta premisa se desarrolló esta propuesta con el objetivo de ayudar a personas candidatas a prótesis mioeléctricas en el proceso de adaptación con el uso de prótesis utilizando la característica de la envolvente de la señal mioeléctrica.

Mientras más fuerza de contracción aplique el paciente, mayor movimiento produce en el servomotor que controlaría la prótesis, y a su vez, en la aplicación móvil, se registra su nivel de fuerza, actuando así también como reeducador muscular.

Una ventaja del equipo aquí propuesto es la portabilidad característica útil cuando es necesario que el paciente se esté ejercitando de forma continua. Esta es una diferencia importante respecto a otros sistemas similares presentados en la literatura científica [Barraza, 2010], [Ramírez, 2006] donde se propone como interfaz gráfica, en un caso una prótesis virtual que se presenta en un monitor de computadora personal y en el otro un conjunto de figuras geométricas cuya superficie se rellena de acuerdo al nivel de actividad muscular, estas también presentadas en un monitor de computadora personal. Debido a esto, limitando el sistema de terapia a un espacio físico fijo.

Para una persona con una extremidad amputada, en este caso el antebrazo, la terapia resulta de gran ayuda. Si la ayuda que puede otorgársele es mediante un

medidor de fuerza que garantice un seguimiento de su mejoría, esta tiene un impacto incluso emocional en el paciente. Si a la vez, mediante contracciones controladas puede abrir una prótesis de pinza tanto como él desee, considerando las restricciones mecánicas, significa un gran apoyo tanto en terapia como funcional para el paciente.

Para no provocar esfuerzos extenuantes como fatiga muscular en los pacientes, la auto calibración presenta una gran ventaja. Es decir, el paciente no se ve forzado a realizar grandes esfuerzos físicos para abrir o cerrar la pinza, o para medir su fuerza de contracción. Sino que, de acuerdo a su máxima contracción muscular, se establece un límite, que asegura que los parámetros de apertura de la prótesis como los niveles de fuerza del reeducador muscular en la aplicación móvil, siempre estarán en un rango aceptable del paciente, dado que él mismo fija el valor máximo.

El uso de la envolvente de la señal EMG ayuda para su manejo como variable de control ya que suaviza su comportamiento aleatorio e incluso, omite ruido que pudiera existir dentro de los rangos de interés de la señal EMG. Y, además, como punto principal en toda esta investigación, permite el correcto control, mediante regulación PWM del servomotor. Es decir, la señal envolvente de la señal EMG determina la apertura de la prótesis de pinza. Así, no importa en realidad si la contracción generada por el paciente genera valores anormales, como, por ejemplo, momentos en los que el paciente genera una contracción con valores picos que sobresalen de la señal promedio. El comportamiento de la envolvente garantiza un buen control para el servomotor, y con esto, de la prótesis de mano.

5. Conclusiones

Se desarrolló un sistema de entrenamiento mioeléctrico basado en una aplicación móvil orientado al entrenamiento de pacientes potenciales usuarios de prótesis mioeléctricas de mano de un grado de libertad. El objetivo de este tipo de sistemas es ayudar a la simbiosis entre amputado y prótesis. El sistema aquí desarrollado incluye la instrumentación electrónica necesaria para captar y

registrar la señal EMG y una interfaz gráfica de usuario sobre una aplicación móvil cuya funcionalidad permite al usuario:

- Monitorear su actividad muscular de forma amigable a través de una barra de fuerza.
- Controlar un dispositivo protésico de mano virtual de un grado de libertad.

A diferencia de otros sistemas de entrenamiento que requieren de monitores o un equipo de mayor complejidad, esta propuesta de menor tamaño y complejidad y además portabilidad facilita a los próximos usuarios el poder realizar el entrenamiento de una manera más sencilla y con una interfaz bastante amigable. Además, el sistema cuenta con una función de autocalibración que está en función de una máxima contracción muscular voluntaria que el usuario realiza al inicio del entrenamiento. Esto presenta la ventaja de evitar la fatiga muscular durante el proceso del entrenamiento.

Como una siguiente etapa a futuro del trabajo aquí desarrollado se planea valorar el funcionamiento del sistema con un grupo de usuarios con la intención de obtener una retroalimentación desde el punto de vista de funcionalidad clínica para hacer modificaciones en caso de ser necesarias y posteriormente promover el sistema para su implantación en uso clínico.

6. Bibliografía y Referencias

- [1] Atmel ATmega328/P, datasheet: http://www.atmel.com/Images/Atmel-42735-8-bit-AVR-Microcontroller-ATmega328-328P_Datasheet.pdf, 12/05/2017.
- [2] Barraza-Madriral, J. A., Ramírez-García, A., y Muñoz-Guerrero, R. (2010, September). A virtual upper limb prosthesis as a training system. In *Electrical Engineering Computing Science and Automatic Control (CCE)*, 2010 7th International Conference on IEEE, pp. 210-215, 2010.
- [3] Ramírez, A., R Muñoz, L Leija, y A Vera, Sistema de Entrenamiento Muscular con Retroalimentación Visual, Pan American Health Care Exchanges (PAHCE 2006), Long Beach, California, pp. 36-39, 30 de enero al 3 de febrero 2006.

- [4] Bundhoo, V., Design and evaluation of a shape memory alloy-based tendon-driven actuation system for biomimetic artificial fingers, Doctoral dissertation, 2009.
- [5] Casallas, E. C., Toro, J. D. R., & Castrillón, I. F. T., Virtual coaching system for transradial myoelectric prosthesis using bioelectrical signals, *Revista politécnica*, 11(21), pp. 97-106, 2016.
- [6] Campo, O., Rovetta, A., & Caicedo, E., Uso de vibraciones musculares para identificar la intención de movimiento en amputados transfemorales. In 8 Congreso Iberoamericano de Ingeniería Mecánica, 2007.
- [7] Clingman, R., & Pidcoe, P., A novel myoelectric training device for upper limb prostheses. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(4), pp. 879-885, 2014.
- [8] Dupont, A. C., & Morin, E. L., A myoelectric control evaluation and trainer system. *IEEE Transactions on Rehabilitation Engineering*, 2(2), pp. 100-107, 1994.
- [9] Guyton, A. C., Hall, J. E., & Guyton, A. C., *Tratado de fisiología médica*. Elsevier Brasil, pp. 109, 2006.
- [10] JL, R. P., & Roca, O., Uso de la isoestación B-200® y electromiografía de superficie en la valoración del dolor lumbar. *Mapfre Medicina*, 12(4), pp. 241-249, 2001.
- [11] Loreto Vergara, B., *Desarrollo de la medicina física y rehabilitación como especialidad médica*, 2014.
- [12] Morant-Arilla, D., Martín-Ruiz, J., Gallego-Cerveró, C., Tamarit-Grancha, I., & Pérez-Pérez, J., Comparación de la electromiografía superficial en el ejercicio de press de banca mediante el uso de Electroestimulación Eléctrica en el test de una Repetición Máxima: estudio piloto. *Revista Andaluza de Medicina del Deporte*, 8(4), pp.182-183, 2015.
- [13] Ramírez-García. A., *Desarrollo de un equipo electrónico de entrenamiento muscular con retroalimentación visual*. Tesis, Maestría, Patente 275763, 2005.

- [14] Ramírez, J. C., & Peláez, A., Conceptos básicos para el análisis electromiográfico. *Revista CES Odontolntfu VoL*, 15(1), 2002.
- [15] San Mauro Martín, I., González Fernández, M., & Collado Yurrita, L., Aplicaciones móviles en nutrición, dietética y hábitos saludables: análisis y consecuencia de una tendencia a la alza. *Nutrición Hospitalaria*, 30(1), pp. 15-24, 2014.
- [16] Trujillo, J. C. G., Muñoz, J. E., & Villada, J. F., Exergames: una herramienta tecnológica para la actividad física. *Revista Médica de Risaralda*, 19(2), 2013.
- [17] Texas Instruments Ina114AP: <http://www.ti.com/lit/ds/symlink/ina114.pdf>, 20/02/2017.

FUSIÓN MORFOLÓGICA DE IMÁGENES IR Y VISUALES UTILIZANDO EL MODELO LIP

Oscar Ricardo Delfín Santiesteban

Cidesi

oscar.delfin.santiesteban@gmail.com

Iván Ramón Teról Villalobos

Cideteq

iterol@cideteq.mx

Resumen

La fusión de imágenes es el proceso de combinar la información de una escena que proviene de dos o más imágenes fuente en una sola con una mejor percepción visual y espacial que puede proporcionar detalles que en su conjunto, no pueden ser observados en las imágenes por separado. En este estudio, se presenta una metodología que permite realizar este procedimiento combinando el modelo de procesamiento logarítmico de imágenes (LIP Model) y las transformaciones morfológicas por reconstrucciones.

Palabras Claves: Imagen Visual, imagen IR, modelo LIP, morfología matemática.

Abstract

Image fusion is the process of combining information from a scene that comes from two or more source images into a single one with better visual and spatial perception that can provide details that as a whole cannot be seen in separate images. In this study, a methodology is presented that allows performing this procedure combining the logarithmic image processing model (LIP Model) and the morphological transformations by reconstructions.

Keywords: LIP Model, IR Image, mathematical morphology, visual Image.

1. Introducción

Capturar y registrar nuestro entorno, a través de imágenes o videos, ha sido una de las actividades de las que se ha valido el hombre para resolver algunas problemáticas o tener evidencia de que ciertos eventos ocurrieron. Para llevar a cabo la captura de las escenas de interés, hacemos uso de determinados sensores que se ajustan a las condiciones del entorno de manera tal, que nos provean de información visual suficiente para tomar alguna decisión. El gran avance tecnológico nos permite hoy en día contar con diferentes tipos de sensores de los cuales, podemos obtener imágenes bajo condiciones muy particulares. Por ejemplo, cámaras visuales, cámaras de visión nocturna, cámaras infrarrojas, cámaras de longitud de onda milimétrica, cámaras de rayos X o cámaras pancromáticas [Omar, 2014]. Las imágenes que se generan de estas tecnologías, aportan diferente tipo de información de una misma escena. Por ejemplo, en la figura 1a, se presenta una imagen visual tomada de noche en la que no se puede apreciar que hay una persona cruzando. Esta información es revelada en la toma 1b, que corresponde a la misma escena, pero la imagen es infra-roja. Observe como, una misma escena con dos sensores que registran cosas diferentes, proveen información complementaria.

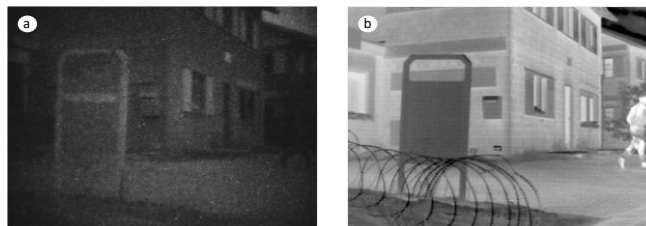


Figura 1 Imagen cortesía de TNO Human Factors Institute, The Netherlands

El tener la posibilidad de unir dos o más fuentes de información provenientes de dos o más tecnologías diferentes y obtener en una sola toma lo mejor de ambas permite observar detalles y características de una misma escena que tal vez de forma separada no es posible percibir. Al proceso de combinar imágenes provenientes de diferentes sensores y generar otra con las mejores características de estas es lo que se conoce como fusión de imágenes. Los propósitos de la

fusión de imágenes son: a) obtener una imagen con una mejor percepción visual y espacial de las estructuras que componen las imágenes por separado, b) minimizar la información redundante c) maximizar la información la relevante.

Para lograr fusionar las imágenes provenientes de varias tecnologías, se han desarrollado una gran cantidad de metodologías que se ajustan a alguna aplicación o problemática en particular. Las aplicaciones que se han valido de estas metodologías son diversas:

- *Geo-ciencia*. El objetivo es la detección, clasificación y seguimiento de fenómenos terrestres. En este tipo de aplicaciones, los sensores ofrecen imágenes de baja y alta resolución que por lo general incluyen dos imágenes: una pancromática de alta resolución espacial (PAN) y una multi-espectral con baja resolución espacial (MS). El fusionar estos dos tipos de imágenes permite que los investigadores obtengan imágenes donde se aprecia una gran extensión de tierra con muchos detalles [Yang, 2012].
- *Diagnóstico médico*. Se tienen diferentes tipos de imágenes como resultado de la aplicación de ciertas técnicas no invasivas para detectar tumores, tejidos blandos, tejidos duros y diversas patologías. Resonancia magnética (MRI), Tomografía Computarizada (TM), Tomografía por emisión de positrones (PET), y ultra sonido son algunas de las imágenes utilizadas en este campo [Li, 2014]. La fusión de este tipo de imágenes permite a los médicos realizar un diagnóstico de mayor calidad y con mayor asertividad.
- *Imágenes multi-foco*. Una misma escena capturada con diferentes sensores presentan zonas en la que aparece el primer plan bien definido en una, pero en la otra se presenta como un segundo plano [Li, 2013].

En este trabajo de investigación, se abordará la fusión de imágenes IR y visuales en escalas de grises. Las imágenes infra-rojas (IR), son ampliamente utilizadas, debido a que, cómo se sabe, cualquier objeto que tenga una temperatura por encima del cero absoluto emite una radiación en la zona infrarroja. Esta radiación es captada por sensores diseñados para recibir estas longitudes de onda que son no perceptibles por el ojo humano y las presenta en un formato que podemos

entender. Las imágenes IR son una herramienta muy versátil que es aplicada en la resolución de innumerables problemas. En aplicaciones petroquímicas es utilizada para la detección de pérdida de aislamiento en los procesos de refinería, inspección de soldaduras, inspección de tuberías, evaluación de la eficiencia y calidad de intercambiadores de calor, entre otros. En aplicaciones aeroespaciales, son utilizadas para el diagnóstico de llantas y frenos, diagnóstico de sistemas de descongelación en alas, detección de desgaste de materiales compuestos, inspección de líneas eléctricas de alta tensión, inspección de tuberías y áreas con grietas por corrosión. En sistemas de seguridad, se han reportado aplicaciones para detectar portación de armas ocultas [Xue, 2003], identificación y rastreo de objetivos, detección de minas y autenticación de personal [Mayet, 1996], reconocimiento de rostros [Piella, 2003].

El proceso de fusionar imágenes puede ser clasificada en tres niveles de acuerdo al tipo de procesamiento que se efectúe [Pohl, 1998]:

- A nivel de pixel: se genera una imagen fusionada en la que la información asociada a cada pixel se determina a partir de los pixeles de las imágenes originales.
- A nivel de características: se extraen características que requieran fusionar con mayor interés y después se identifican tamaños, formas, contrastes y texturas previo al paso de fusión.
- A nivel de decisión: la fusión se realiza procesando las imágenes por separado determinando no solo algún objeto de interés, sino también algunos otros parámetros como contraste, ruido, vecindad entre pixeles.

Al escoger cual es el esquema más adecuado de procesamiento dependerá en gran medida de factores como el tipo de datos de entrada, la aplicación y las herramientas con las que se dispone [Piella, 2003]. En particular, las imágenes IR, ofrecen muy baja resolución espacial en comparación con las imágenes visuales, pero ofrecen detalles que no son perceptibles a la vista humana y que, en conjunto permiten entender una escena. Para llevar a cabo la fusión de estos dos tipos de imágenes se han desarrollado innumerables métodos. Por ejemplo,

se han desarrollado métodos de procesamiento multi-resolución, en el que las imágenes de entrada son descompuestas a través de diversas transformaciones wavelet y curvelet, de manera tal, que se van va extrayendo la información sobresaliente para, posteriormente, fusionarlas a través de reglas de fusión [Pohl, 1998], [Pinoli, 1997].

Otros métodos de descomposición multi-escala propuestos en [Toet, 1989], [Toet, 1990], [Toet, 1992], [Toet, 1989] hacen uso de filtros pasa baja que permiten preservar el contraste de las imágenes de entrada. Multi-resolución con filtros morfológicos. Independientemente del método de fusión que se desee emplear, es requisito que las imágenes de entrada estén alineadas para garantizar que la información de cada toma corresponde con las estructuras físicas de la realidad. Una de las grandes aportaciones en el procesamiento de imágenes, ha sido el desarrollado en [Pinoli, 1997], en el que se propone un conjunto de operadores que permiten manipular de forma no lineal los tonos de una imagen formada en escala de grises.

En este estudio, se presenta un método de fusión en el que se incluyen filtros morfológicos en el modelo LIP utilizando un modelo de procesamiento de imágenes conocido como *LIP MODEL*.

2. Métodos

Modelo de Procesamiento Logarítmico de Imágenes

Una imagen está formada por luz reflejada que puede ser representada de manera vectorial. Si se intenta adicionar dos valores de intensidad o bien, multiplicar un valor por un factor constante, el resultado pudiera no estar dentro del intervalo en el que se acotan las imágenes de entrada [Mayet, 1996]. Bajo esta premisa, el modelo LIP, define un conjunto de operadores que permiten manipular las intensidades de grises de una o más imágenes acotando su resultado al espacio de trabajo de cada una de ellas. Este modelo fue introducido por [Pinoli, 1992] en 1980 y se desarrolló para el procesamiento de imágenes hasta 1997 y ha sido utilizado en aplicaciones como remoción de fondos de una imagen, corrección de iluminación, interpolación de imágenes, reconstrucción 3D, restauración de

imágenes, estimación del contraste, segmentación de imágenes, descomposición multi-escala y compresión de imágenes [Mayet, 1996]. Pese a la gran variedad de aplicaciones de este modelo, no se han reportado su aplicación para el problema de fusión. De acuerdo con el modelo LIP, la adición de dos imágenes está definido a través de un operador a un intervalo definido para la cantidad máxima de tonos posibles $(0, M]$. En particular, la adición de una imagen $f(x, y)$ con otra $g(x, y)$ queda definida como en la ecuación 1.

$$f(x, y) \dagger g(x, y) = f(x, y) + g(x, y) - (f(x, y) * g(x, y) / M) \quad (1)$$

Con $M = 2^n, \forall x, y \in \mathbb{Z} \rightarrow (0, M]$ donde $n = 8$.

Este operador \dagger llamado adición LIP, permite la adición cerrada en el intervalo $(0, M]$. También es posible amplificar el contenido de una imagen a través del operador de producto escalar, cuya definición es mostrada en la ecuación 2.

$$\alpha * f(x, y) = M - M \left(1 - \frac{f(x, y)}{M} \right)^\alpha \quad (2)$$

Con $M = 2^n$. En general, cuando $\alpha > 1$, se eleva a tonos claros el contenido de la imagen. Cuando $\alpha < 1$, se oscurece la imagen. En $\alpha = 1$, el contenido de la imagen queda sin cambios. En la siguiente figura se muestra como son afectadas las tonalidades de unas imágenes para cierto rango de valores de α .

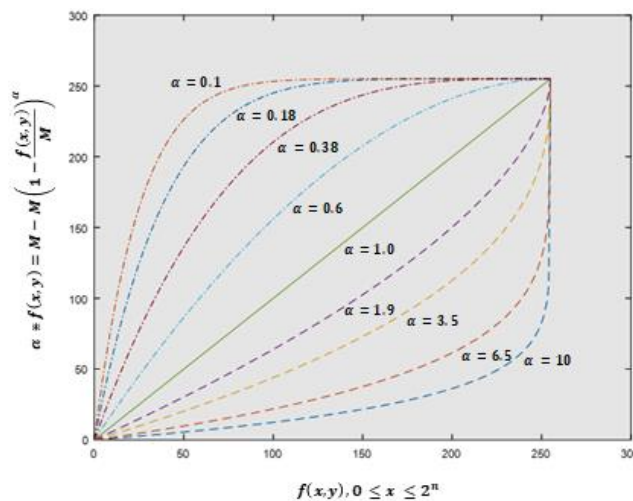


Figura 2 Modificación del tono por el producto escalar LIP.

Una manera de conocer el valor de alfa [Michoud, 1997] y que permite maximizar la dinámica de la imagen está dada por ecuación 13.

$$\alpha(f) = \frac{\ln\left(\frac{\ln\left(\frac{M - f_{max}}{M}\right)}{\ln\left(\frac{M - f_{min}}{M}\right)}\right)}{\ln\left(\frac{M - f_{min}}{M - f_{max}}\right)} \quad (3)$$

Donde f_{min} y f_{max} representan respectivamente los niveles de gris mínimo y máximo de una imagen.

Morfología Matemática

La Morfología Matemática descansa en la teoría de conjuntos desarrollada por el matemático ruso Hermann Minkoski y el alemán Hugo Hadwiger en los primeros años del siglo XX. La reformulación de ésta teoría por los franceses Jean Serra y Georges Matheron dieron como consecuencia una teoría para el análisis y procesamiento de estructuras geométricas cuya aplicación más popular ha sido en el análisis de imágenes. En general, la morfología matemática permite realizar transformaciones no lineales sobre estructuras reticulares en espacios bidimensionales mediante el uso de un conjunto de forma conocida llamado elemento estructural cuyo tamaño y forma se escoge, *a priori*, de acuerdo con la morfología del conjunto sobre el que va a operar y en función de lo que se desea obtener. Las operaciones básicas de la teoría morfológica son la erosión y la dilatación y son denotadas en las ecuaciones 4 y 5.

$$\delta_{\lambda\beta}(f)(x) = (f \oplus \lambda\hat{\beta})(x) = \vee \{f(y) : y \in \lambda\hat{\beta}_x\} \quad (4)$$

$$\varepsilon_{\lambda\beta}(f)(x) = (f \ominus \lambda\hat{\beta})(x) = \wedge \{f(y) : y \in \lambda\hat{\beta}_x\} \quad (5)$$

Estas operaciones no admiten inversa, por tanto, no se puede determinar el origen x desde las imágenes $\delta_{\lambda\beta}(f)(x)$ o $\varepsilon_{\lambda\beta}(f)(x)$. Sin embargo, dada la dualidad de estas transformaciones, es posible aproximarse al elemento original realizando una dilatación dada la erosión o bien, erosionar dada una dilatación. Estas operaciones son llamadas Apertura y Cerradura morfológica y son denotados en ecuaciones 6 y 7.

$$\gamma_{\lambda\beta}(x) = (x \ominus \lambda\beta) \oplus \lambda\beta = \delta_{\lambda\tilde{\beta}}(\varepsilon_{\lambda\beta}(x)) \quad (6)$$

$$\varphi_{\lambda\beta}(x) = (x \oplus \lambda\beta) \ominus \lambda\beta = \varepsilon_{\lambda\tilde{\beta}}(\delta_{\lambda\beta}(x)) \quad (7)$$

La apertura morfológica es de utilidad para eliminar detalles luminosos o claros en relación con elemento estructural quedando el resto de la imagen relativamente sin modificaciones. El cierre, en cambio, elimina detalles oscuros en relación con el elemento estructural. Con estas definiciones formamos una transformación que permite descubrir la información que la apertura y el cierre remueven: la transformada Top-Hat, que se define como la diferencia entre la imagen original y la apertura, llamada Top-Hat sobre blancos y, la diferencia entre el cierre y la imagen original, llamado Top-Hat sobre negros. Su notación es mostrada en las ecuaciones 8 y 9.

$$ThW_{\lambda\beta}(X) = X - \gamma_{\lambda\beta}(x) \quad (8)$$

$$ThB_{\lambda\beta}(X) = \varphi_{\lambda\beta}(x) - X \quad (9)$$

La principal desventaja de la aplicación de las operaciones morfológicas es la distorsión producida por el elemento estructural sobre las estructuras originales de la imagen. A diferencia de las transformadas morfológicas, en las que, a través de aperturas y cerraduras se eliminan ciertas regiones para resaltar algunos otros detalles, las transformadas por reconstrucción corresponden a un grupo de transformadas que permiten preservar la información de origen permitiendo eliminar ciertas zonas que previamente han sido marcadas de acuerdo algún propósito específico. En general, las transformaciones por reconstrucción se forman a partir de una imagen de referencia X y una imagen Y llamada “marcador” que crece al interior de la primera. Para construir estas transformaciones se hace uso del concepto de erosión y dilatación geodésica.

La dilatación geodésica de tamaño 1, es denotada en ecuación 10.

$$\delta_X^1(Y) = X \cap \delta_B(Y) \quad (10)$$

Cuando esta dilatación es iterada hasta la estabilidad, obtenemos la transformación por reconstrucción dada por ecuación 11.

$$\rho_X(Y) = R(X, Y) = \lim_{n \rightarrow \infty} \delta_X^n(Y) = \delta_X^1 \delta_X^1 \dots \delta_X^1(Y) \quad (11)$$

Dado que las operaciones morfológicas son duales, entonces, para el caso de la erosión geodésica, tenemos la notación para tamaño 1 en la ecuación 12.

$$\varepsilon_X^1(Y) = X \cap \varepsilon_\beta(Y) \quad (12)$$

Iterando hasta la estabilidad, obtenemos la transformación dual por reconstrucción dada por la ecuación 13.

$$\rho_X^*(Y) = R^*(X, Y) = \lim_{n \rightarrow \infty} \varepsilon_X^n(Y) = \varepsilon_X^1 \varepsilon_X^1 \dots \varepsilon_X^1(Y) \quad (13)$$

Cuando el conjunto marcador Y es igual a la erosión X, obtenemos la apertura y cierre por reconstrucción, ecuaciones 14 y 15.

$$\tilde{Y}_{\lambda B}(X) = \lim_{n \rightarrow \infty} \delta_X^n(\varepsilon_{\lambda B}(X)) = R(X, \varepsilon_{\lambda B}(X)) \quad (14)$$

$$\tilde{\varphi}_{\lambda B}(X) = \lim_{n \rightarrow \infty} \varepsilon_X^n(\delta_{\lambda B}(X)) = R(X, \delta_{\lambda B}(X)) \quad (15)$$

De la misma forma en que fueron establecidas las operaciones Top-Hat morfológica sobre blancos y Top-Hat sobre oscuros, para el caso de la reconstrucción tenemos las ecuaciones 16 y 17.

$$ThW(X) = X - \tilde{Y}_{\lambda B}(X) \quad (16)$$

$$ThB(X) = \tilde{\varphi}_{\lambda B}(X) - X \quad (17)$$

Con los elementos mostrados, a continuación, describiremos como es utilizada esta última transformada para poder intercambiar la información de dos imágenes.

Modelo LIP como Operador Primitivo de Fusión

Una de las premisas para llevar a cabo la fusión de imágenes, es encontrar un operador que, a nivel de pixel, adicione sus valores y cuyo resultado este acotado a un rango que, para efectos de imágenes codificadas en 8 bits, sea 255. Dado que el operador de adición LIP cumple con esta premisa, se verá a continuación que es lo que ocurre cuando aplicamos este operador sobre 2 imágenes de entrada. En la figura 3, se presenta una imagen visual y una de tipo IR de la misma escena. El resultado de aplicar el operador de adición LIP sobre estas entradas se puede apreciar en figura 4.



Figura 3 Imágenes de entrada.



Figura 4 Resultado de la adición LIP al conjunto de imágenes de entrada.

Como se aprecia en la figura 5, si bien la imagen resultante muestra el contenido de las dos de entrada, aún no es posible apreciar con detalle la información complementaria que la imagen IR aporta. Con la intención de amplificar algunos detalles, a través de la ecuación 3, determinamos el valor de $\alpha = 0.8956$ para la imagen resultante y aplicamos el producto LIP por este escalar de la ecuación 2. El resultado se muestra en la figura 5.

Observe cómo, efectivamente algunas zonas se ven más claras en comparación con el resultado obtenido en la figura 4, sin embargo, ésta imagen y al igual en el primer resultado, no es posible apreciar la información complementaria que aporta la imagen IR a toda la escena. Es necesario, entonces, encontrar un mecanismo

que, a partir de la adición y multiplicación LIP sobre imágenes de entrada, incorpore los detalles complementarios para terminar el proceso de fusión.



Figura 5 Producto LIP por escalar de la adición de dos imágenes de entrada.

Modelo Fusión Morfológica usando LIP Model (MLiF)

Como se expuso en la sección anterior, el operador de adición LIP en conjunto con el operador de multiplicación LIP por un escalar, nos ofrece una primera aproximación a un modelo de fusión en el que la información contenida en la imagen fusionada represente en mayor medida la información de las imágenes de entrada.

Para incluir en el resultado final de la fusión la información complementaria de la imagen IR, se propone entonces la extracción de características de la siguiente manera. Sea $F(x, y)$ y $G(x, y) \forall (x, y) \in \mathbb{N}$ dos imágenes definidas en 2^n niveles de grises con $n = 8$. Aplicando las ecuaciones 16, 17 a las imágenes de entrada, obtenemos las ecuaciones 18, 19, 20 y 21.

$$thW(F) = F - \tilde{\gamma}_{\lambda\beta}(F) \quad (18)$$

$$thB(F) = \tilde{\varphi}_{\lambda\beta}(F) - F \quad (19)$$

$$thW(G) = F - \tilde{\gamma}_{\lambda\beta}(G) \quad (20)$$

$$thB(G) = \tilde{\varphi}_{\lambda\beta}(G) - G \quad (21)$$

Donde β es un elemento estructural plano de 3×3 y $\lambda \in \mathbb{N}$. Con estas transformaciones por reconstrucción, obtenemos dos nuevos conjuntos que los denotamos en las ecuaciones 22 y 23.

$$D_n = V(thW(F), thB(F)) \quad (22)$$

$$D_m = V(thW(G), thB(G)) \quad (23)$$

Y representan las zonas más claras que sobresalieron en función del tamaño del elemento estructural. A continuación, se determinan las estructuras brillantes de ambas imágenes de entrada a través de una función de composición denotada en las ecuaciones 24 y 25.

$$C_{m,n} = D_n \circ D_m = \begin{cases} 2^n & \text{si } D_n > D_m \\ 0 & \text{en otro caso} \end{cases} \quad (24)$$

$$\bar{C}_{m,n} = C_{m,n} \quad (25)$$

Esta función de composición permite unir las zonas brillantes de ambas imágenes de entrada que fueron obtenidas a través de las transformaciones morfológicas Top – Hat. Esta función de composición es filtrada para homogenizar las zonas brillantes. Finalmente, la imagen fusionada se obtiene mediante la ecuación 26.

$$Fm(F, G) = \begin{cases} \alpha * (F + G) + D_n & \text{si } \bar{C}_{m,n} = 2^n \\ \alpha * (F + G) & \end{cases} \quad (26)$$

En la figura 6, se presenta de manera gráfica una representación del método que acabamos de describir y, en la figura 7, se muestra el pseudocódigo de este.

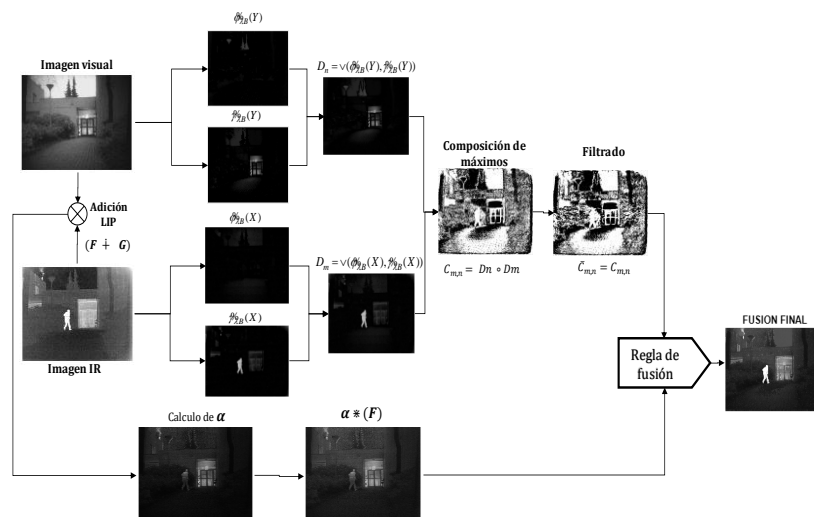


Figura 6 Diagrama de bloques del modelo MLiF.

Algoritmo propuesto

```

inputs : Imagen visual ( $I_{vis}$ ), Imagen IR ( $I_{ir}$ )
output : imagen fusionada ( $ImFus$ )

lipAdd  $\leftarrow$  adición LIP( $I_{vis}, I_{ir}$ )
alphaLIP  $\leftarrow$  cálculo del parámetro  $\alpha$ 
preFusLIP  $\leftarrow$  producto escalar LIP(lipAdd, alphaLIP)

IvisTNW  $\leftarrow$  TopHat( $I_{vis}$ )
IvisTnB  $\leftarrow$  BottonHat( $I_{vis}$ )
Dn  $\leftarrow$  max(IvisTNW, IvisTnB)

IirTNW  $\leftarrow$  TopHat( $I_{ir}$ )
IirTnB  $\leftarrow$  BottonHat( $I_{ir}$ )
Dm  $\leftarrow$  max(IirTNW, IirTnB)

Para todo pixel (i, j) de Dn hacer
    Si Dn > Dm entonces
        Rn(i, j)  $\leftarrow$  2n
    sino
        Rn(i, j)  $\leftarrow$  0
fin para todo

medianRn  $\leftarrow$  filtro de mediana(Rn)

Para todo pixel(i, j) de Rn hacer
    Si medianRn(i, j) == 2n
        ImFus(i, j)  $\leftarrow$  preFus(i, j) + Dn(i, j)
    sino
        ImFus(i, j)  $\leftarrow$  preFus(i, j)
fin para todo
    
```

Figura 7 Pseudocódigo del método propuesto.

3. Resultados

Conjuntos de Datos Experimentales

Para comprobar la eficiencia del método propuesto, se utilizó un conjunto de 20 pares de imágenes como entrada, previamente registradas y que fueron tomadas de TNO Humans Factors¹, que contienen tomas militares en diferentes escenarios y que fueron tomadas con diferentes tipos de cámaras. En la figura 8 se muestran los 20 pares de imágenes de entrada que se utilizaron para validar nuestro modelo de fusión; a la izquierda de cada par tenemos la imagen IR y a la derecha se muestra la misma escena, pero en imagen visual. La eficiencia de un método puede valorarse desde dos puntos de vista: a partir de la aplicación que lo requiere a través del observador que hará uso de la fusión y, a través de métricas que cuantifiquen el grado en que la información de entrada fue intercambiada para obtener un resultado final. En cuanto a la primera, pudiera resultar un poco complicado determinar qué método de fusión fue el mejor ya que, visualmente la diferencia entre un resultado y otro pudiera no ser claramente perceptible. Quizá para alguna aplicación en particular, para un observador, un determinado método es mejor frente a otro. Bajo esta idea, la valoración de eficiencia se dará en términos cualitativos y será subjetiva ya que estará sujeta a la apreciación del observador y estará en función de lo que pudiera buscar o esperar.

¹ El conjunto de datos de entrada está disponible en http://figshare.com/articles/TNO_Image_Fusion_Dataset/1008029

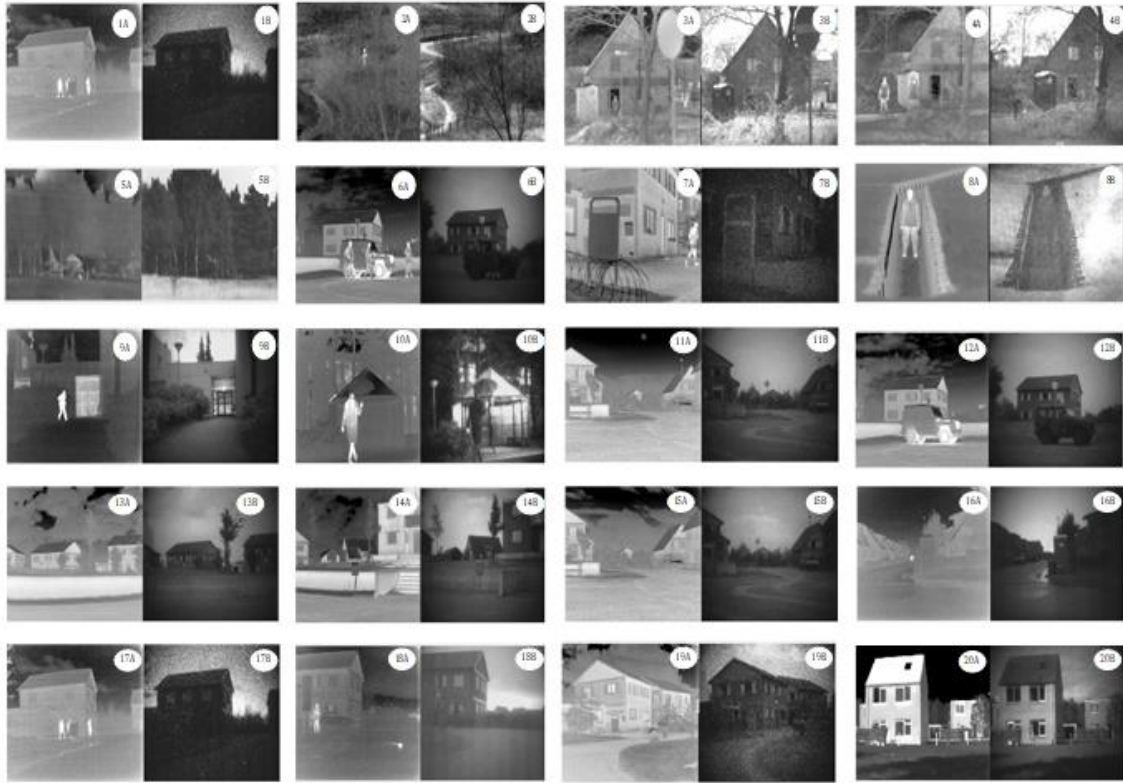


Figura 8 Conjunto de datos experimentales.

Por otro lado, para valorar de manera cuantitativa la posible ventaja de un modelo sobre otro, se utilizaron algunos indicadores que comparan la similitud, los contornos, el contenido de información, el contraste, entre otras, de las imágenes de entrada con la imagen fusionada. Esto no quiere decir, a priori, que un indicador sea mejor que otro, simplemente miden cosas diferentes. Para este estudio, fueron utilizados cuatro indicadores, i.e., Correlación Cruzada (Corr2D), Error Cuadrático Medio (RMSE), el Índice de la Calidad de la Fusión (IoQ) y el Índice de Similitud Estructural (SSIM). La definición de estos cuatro indicadores se presenta a continuación:

- Correlación Cruzada (Corr2D): Es una medida estadística de similitud entre dos variables aleatorias. Para determinar la correlación de la imagen fusionada contra las originales, decimos que nuestro valor de correlación es el producto de la correlación de las originales contra la fusionada. Es decir, si $r_{AF} = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(F_{mn} - \bar{F})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (F_{mn} - \bar{F})^2)}}$ es la correlación de la imagen A

(una del conjunto de entrada) con la fusionada (F) y $r_{BF} = \frac{\sum_m \sum_n (B_{mn} - \bar{B})(F_{mn} - \bar{F})}{\sqrt{(\sum_m \sum_n (B_{mn} - \bar{B})^2)(\sum_m \sum_n (F_{mn} - \bar{F})^2)}}$ es la correlación de la imagen B (la segunda del conjunto de entrada) con la fusionada, entonces nuestro indicador de fusión será: $r_{AF} * r_{BF}$.

- Error Cuadrático Medio (RMSE), permite cuantifica el error medio entre cada una de las imágenes de entrada contra la imagen fusionada.
- El Índice de Calidad de la Fusión (IoQ), es una métrica que combina el coeficiente de correlación, la distorsión lumínica y la distorsión del contraste e indica el grado con el que se integraron las imágenes de entrada en la fusionada. El máximo de este valor es 1 y es alcanzado cuando la totalidad de las entradas están integradas en la fusionada.
- El Índice de Similitud Estructural (SSIM) es un indicador que permite valorar calidad de la fusión con base a la degradación de estructuras de información contenidas en la imagen fusionada contra otra de referencia comparando patrones locales de intensidad que han sido normalizados en luminancia y contraste.
- Para contrastar los resultados y poder valorar si el método propuesto es eficiente, se utilizó el mismo conjunto de imágenes con dos algoritmos de fusión que también hacen uso de transformaciones morfológicas para lograr la fusión y que han sido reportados en [Toet, 1989] [Mukhopadhyay, 2001]. Todos los experimentos se realizaron en una laptop con Intel core i5 a 3.3 GHz y 8 GB de memoria y los algoritmos fueron codificados de Matlab.

Resultados Obtenidos

Los resultados de nuestra experimentación son mostrados en la figura 9. En esta gráfica se presenta el comparativo de la fusión realizada por 3 métodos diferentes. En la imagen de la izquierda, se aprecia el resultado del método propuesto, al centro se presenta el método reportado [Toet, 1989] y a la derecha la fusión realizada por el método reportado en [Mukhopadhyay, 2001]. Visualmente se puede apreciar que los tres métodos de fusión cumplen el objetivo de integrar las imágenes de entrada.

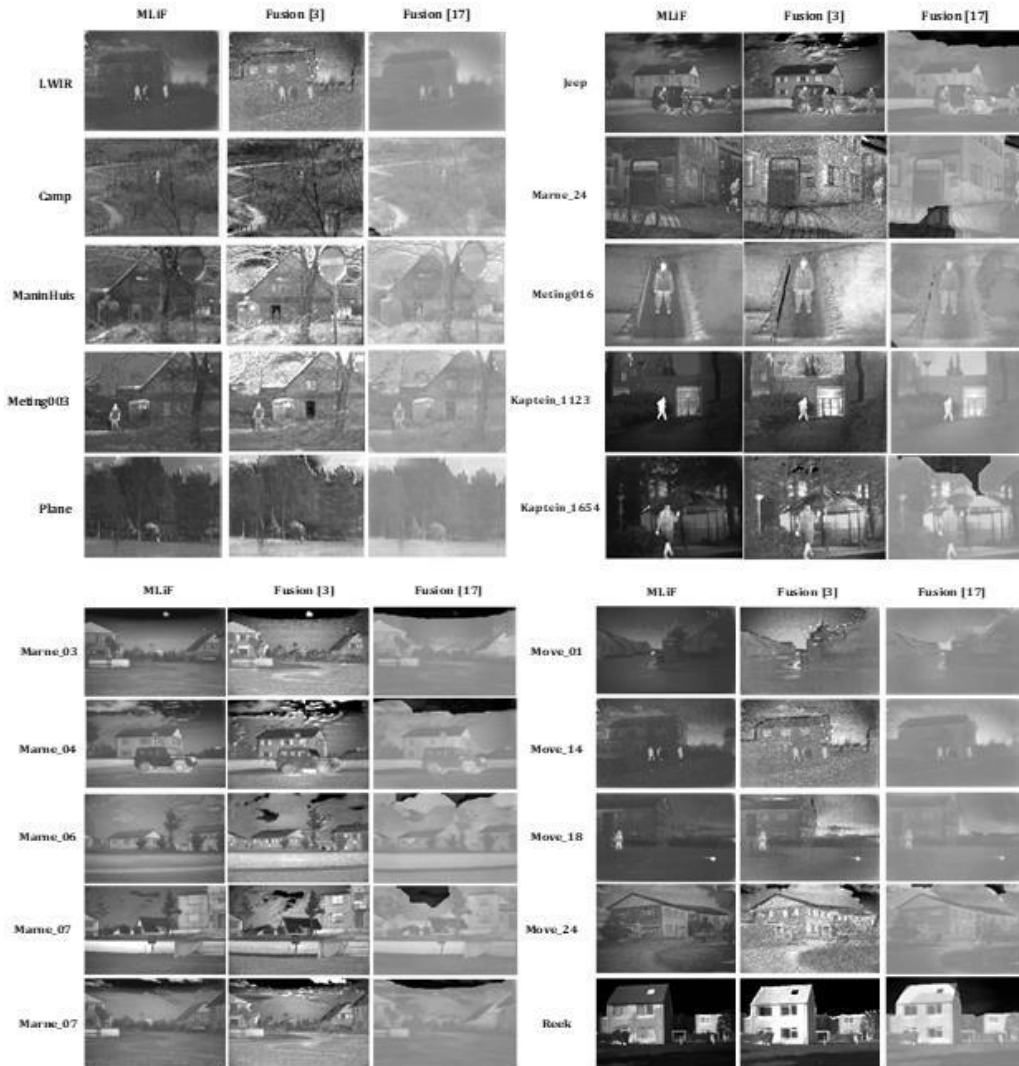


Figura 9 Resultados de la fusión de imágenes de entrada y su comparación con dos métodos de fusión.

Sin embargo, también es posible observar que, en algunos pares fusionados por los métodos reportados en [Toet, 1989], [Mukhopadhyay, 2001] no fusiona de forma eficiente. Se aprecian zonas que no fueron fusionadas o bien, aparecieron algunos contornos que ninguna de las originales tiene y esto es debido a la cantidad de escalamientos del elemento estructural que fueron usados para llevar a cabo la fusión. La ventaja del modelo propuesto es que el elemento estructural no es escalado, su tamaño permanece fijo y solo es usado para extraer la información complementaria de la toma IR e integrarla a la adición LIP previamente realizada.

En la tabla 1 se presentan los resultados obtenidos de aplicar el modelo de fusión propuesto y la fusión realizada con los dos modelos mencionados. Las celdas marcadas indican el mejor método de fusión de cada una de las imágenes de entrada. Se puede apreciar, que el indicador IoQ resultó el mejor método en la gran mayoría. Sin embargo, para el indicador SSIM el método propuesto fue el mejor en 13 de los 20 pares de entrada. En cuanto a la correlación cruzada 9 de los 20 pares de entrada y para el error cuadrático medio 10 de los 20 pares el método propuesto resulto mejor.

Tabla 1 Tabla comparativa de modelos de fusión con diferentes indicadores.

	IoQu			PEARSON			RMSE			SSIM		
	MLiF	Fusion[3]	Fusion[17]	MLiF	Fusion[3]	Fusion[17]	MLiF	Fusion[3]	Fusion[17]	MLiF	Fusion[3]	Fusion[17]
Lwir	-0.0041	0.1408	0.0972	-0.0041	0.0075	-0.0849	8.2542	10.5531	22.5769	0.3873	0.1611	0.2648
Camp	0.0472	0.2148	0.0608	0.0472	0.1137	-0.0577	0.7724	0.9917	55.9463	0.3760	0.3550	0.2407
MainHuis	0.1122	0.2513	0.1171	0.1122	0.1816	-0.0129	25.8455	7.4859	9.8593	0.3140	0.3654	0.4332
Meting003	0.0204	0.2339	0.1136	0.0204	0.0105	-0.0546	17.8241	5.6631	8.4309	0.4068	0.4194	0.5034
Plane	0.2032	0.2194	0.1017	0.2032	0.1998	-0.0064	3.4090	2.0385	19.3434	0.5934	0.6223	0.5894
Jeep	-0.0727	0.2218	0.1325	-0.0727	-0.0307	-0.1971	6.3579	9.3890	35.1170	0.4139	0.3039	0.3473
Marne_24	0.1563	0.1932	0.1641	0.1563	0.2156	0.0451	8.1696	7.7910	15.4476	0.2873	0.1441	0.2261
Meting016	0.1879	0.1735	0.0748	0.1879	0.0483	0.0671	2.9577	2.4209	3.7885	0.5958	0.4812	0.5388
Kaptein_1123	0.3097	0.2960	0.0829	0.3097	0.3129	0.2435	8.5048	4.5340	50.7185	0.3869	0.4018	0.3377
Kaptein_1654	0.2459	0.1383	0.0614	0.2459	0.2171	0.1065	16.8994	4.4315	30.0477	0.2775	0.3146	0.3455
Marne_03	-0.0646	0.1916	0.1236	-0.0646	-0.1132	-0.1467	14.2445	18.0274	25.3263	0.4042	0.2421	0.3723
Marne_04	-0.0691	0.2293	0.1303	-0.0691	-0.0159	-0.1776	8.3654	15.6860	24.6565	0.4104	0.2612	0.3759
Marne_06	0.0516	0.2025	0.1410	0.0516	-0.0664	0.0128	10.9190	6.2471	15.6696	0.4471	0.3662	0.4611
Marne_07	-0.0160	0.2594	0.1385	-0.0160	0.0340	-0.1257	10.9692	12.7138	28.6898	0.4341	0.3228	0.3845
Marne_11	0.0956	0.2024	0.1377	0.0956	-0.0324	-0.0954	10.8614	15.4343	17.8831	0.4354	0.2619	0.4046
Move_01	-0.0848	0.2054	0.0942	-0.0848	0.0357	-0.0552	7.9024	9.2538	17.6585	0.5392	0.3592	0.4753
Move_14	-0.0584	0.1350	0.0911	-0.0584	0.0724	-0.1953	10.3856	15.6774	24.6844	0.3756	0.1427	0.2516
Move_18	0.0823	0.3532	0.0904	0.0823	-0.0381	0.0144	5.1085	1.9798	5.9776	0.5601	0.5142	0.5580
Move_24	0.0388	0.1432	0.1295	0.0388	0.1314	0.0452	11.9498	16.8183	15.1465	0.3192	0.1206	0.2626
Reek	0.3668	0.2306	0.1199	0.3668	0.3519	0.0596	18.5196	4.0143	25.8226	0.2614	0.3782	0.2314

4. Discusión

Por los resultados obtenidos, se concluye qué si bien el uso de indicadores permite cuantificar el grado de fusión de un método sobre otro, el resultado también dependerá de qué es lo que mida un indicador en particular y cómo lo hace. Mientras un indicador pudiera concluir que un método de fusión resultó

mejor que otro, para otro indicador que mida algo distinto, tal vez otro método de fusión resulte más eficiente. Por tanto, no habrá un indicador a través del cual podamos concluir que un método es mejor que otro y por esta razón los métodos de fusión deberán de estar en función de la aplicación para la que se quiera y de lo que el observador este buscando y/o esperando encontrar.

5. Conclusiones

En este estudio se mostró el uso de los operadores básicos del modelo LIP para fusionar imágenes visuales e imágenes IR. Obtenemos una imagen “pre-fusionada”, tomando como base la adición LIP de las imágenes de entrada. Posteriormente, aplicando el operador de producto LIP por un escalar α de esta “pre-fusión” logramos acentuar las zonas más brillantes. Sin embargo, también se mencionó que la sola aplicación de éstas operación no es suficiente para lograr una fusión de calidad, ya que estos operadores operan a nivel de pixel y no permite extraer la información complementaria que ofrecen las tomas IR de una escena en particular. Se usaron las transformaciones por reconstrucción Top Hat para extraer dicha información y desarrollamos un modelo de fusión que permite unir el resultado de este procesamiento al resultado que ofrecen los operadores de adición y multiplicación por un escalar LIP.

Dado que el método propuesto hace uso de transformaciones morfológicas, hemos comparado nuestra propuesta contra dos métodos que se han reportado en [Toet, 1989] [Mukhopadhyay, 2001] que también hace uso de estas transformaciones y que, además, hacen uso del procesamiento multi-escala, haciendo que el proceso de fusión resulte más complejo. Los resultados revelan que, en términos visuales, nuestra propuesta resultó mejor ya que solo empleamos las transformaciones morfológicas para extraer la información complementaria y la agregamos a la pre-fusión realizada a través del modelo LIP, evitando de esta manera, la aparición de contornos no deseados como fue el caso con los otros dos modelos.

Se ha mostrado, cómo cuatro indicadores que actúan de diferente manera reportan que, de acuerdo a como realiza la medición, un método de fusión resulta

efectivo para ciertas imágenes mientras que para otras el mismo indicador pudiera ser diferente. Esto nos habla de la complejidad del proceso de fusión de señales y que su efectividad estará acompañada por dos factores: de la aplicación para la cual será utilizada y de lo que un observador espera encontrar.

6. Bibliografía y Referencias

- [1] Gyaourova A., G. Bebis, I., Fusion of infrared and visible images for face recognition. *Computer Vision–ECCV*, 2004.
- [2] Lewis J. J., R.J. O’Callaghan, S.G. Nikolov, D.R. Bull. Pixel–and region-based image fusion with complex wavelets. *Information Fusion*, Volume 8, Issue 2, pp. 119–130, 2007.
- [3] Li S., X. Kang, J. Hu, B. Yang. Image matting for fusion of multifocus in dynamic scenes. *Information Fusion* 14, pp. 147–162, 2013.
- [4] Li T., Y. Wang, C. Chang, N. Hu, Y. Zheng, Color-Appearance based fusion of gray and pseudo-color images for medical applications. *Information Fusion* 19, pp. 103–114, 2014.
- [5] Ma J., C. Chen, C. Li, J. Huang, Infrared visible image fusion via gradient transfer and total variation minimization. *Information Fusion* 31, 2016.
- [6] Matsopoulos G. K., S. Marshall, Application of morphological pyramids: fusion of MR and CT phantoms. *Journal of Visual Communications and Image Representation* 6 (2), PP. 167 –207, 1995.
- [7] Mayet F., J.C. Pinoli, M. Jourlin, Physical Justification and applications of the LIP Model for the Processing of transmitted light images. *Traitement du Signal*, Volume 13 No. 3, 1996.
- [8] Michoud P., J.C. Pinoli, M. Jourlin, Les applications industrielle et biomédicales du modèle LIP, *Seizème Colloque Gretsi*, pp. 15–19, Septembre 1997.
- [9] Mukhopadhyay S., B. Chanda, Fusion of 2D grayscale images using multiscale morphology. *Pattern Recognition* 34, pp. 1939–1949, 2001.
- [10] Omar A., T. Stathaki. Image Fusion: an overview. *Fifth International Conference on Intelligent System, modelling and Simulation*, 2014.

- [11] Piella G., A general framework for multiresolution image fusion: from pixels to regions. *Information Fusion* 4, pp. 259-280, 2003.
- [12] Pinoli J. C., The Logarithmic Image Processing Model: Connection with Human Brightness Perception and Contrast Estimator. *Journal of Mathematical Imaging and Vision* 7, pp. 341–158, 1997.
- [13] Pohl C., J.L. Genderen, Multisensor image fusion in remote sensing. *International Journal of Remote Sensing*. 19 (5), pp. 823–854, 1998.
- [14] Singh R., A. Khare, Fusion of multimodal medical using Daubechies complex wavelet transform–A multiresoution approach. *Information Fusion* 19, pp. 49–60, 2014.
- [15] Toet A., Image Fusion by ratio of low–pass pyramind. *Pattern Recognition Letters* 9 (4), pp. 245–253, 1989.
- [16] Toet A., Hierarchical Image Fusion, *Machine Vision and Applications*, 1990.
- [17] Toet A., Multiscale contrast enhancement with applications to image fusion. *Optical Engineering* 31(5), pp. 1026-1031, may 1992.
- [18] Toet A., Merging thermal and visual images by a constrast pyramids. *Optical Engineering* 28(7), pp. 789 -792, July 1989.
- [19] Xue Z., R.S.Blum, Concealed Weapon Detection Using Color Image Fusion. Electrical and Computer Engineering Deparment. Lehigh University, 2003.
- [20] Yang S., M. Wang, L. Jiao. Fusion of multispectral and panchromatic images based on support value transform and adaptive principal component analysis. *Information Fusion* 13, pp. 177–184, 2012.

MÓDULO DE CONTROL DE CARGA PARA EVALUAR CELDAS DE COMBUSTIBLE -HARDWARE-

Shirley Yahaira Echánove Gómez

Universidad Autónoma del Carmen

mrodriguez@pampano.unacar.mx

Marco Antonio Rodríguez Blanco

Universidad Autónoma del Carmen

mrodriguez@pampano.unacar.mx

Juan Manuel Tadeo Sierra Grajeda

Universidad Autónoma del Carmen

mrodriguez@pampano.unacar.mx

Luis Enrique Vidal Burelo

Universidad Autónoma del Carmen

mrodriguez@pampano.unacar.mx

Resumen

En este trabajo se presenta el acondicionamiento eléctrico-electrónico de un prototipo de control de cargas eléctricas para evaluar celdas de combustible utilizando componentes mínimos como una tarjeta Raspberry Pi para el control secuencial, interfaz hombre máquina HMI y adquisición de datos, así como la utilización de un opto acoplador lineal como sensor de voltaje y relevadores mecánicos para la etapa de potencia. El objetivo en términos generales de este prototipo es por un lado optimizar los costos de desarrollo y por otro proporcionar una HMI amigable y adecuado de protocolo abierto para obtener curvas de polarización en celdas de combustible con capacidad de 1.5 kW.

Palabras Claves: Cargas eléctricas, celdas de combustible, interfaz hombre máquina, instrumentación.

Abstract

This paper presents the electrical-electronic conditioning of a prototype electric charge control to evaluate fuel cells using minimal components such as a Raspberry Pi card for sequential control, human machine interface HMI and data acquisition, As well as a linear optocoupler voltage sensor and mechanical relays for the power stage. The main objective is to optimize development costs as well as to have a friendly and adequate open protocol HMI

Keywords: *Electrical loads, fuel cells, Human Machine Interface, instrumentation.*

1. Introducción

Actualmente las principales fuentes de energía a nivel mundial provienen de los combustibles fósiles y debido al alto consumo de estos, es conveniente cuestionarse cuanto tiempo durarán las reservas, además hay que tomar en cuenta que la producción excesiva de CO₂ está afectando las condiciones atmosféricas. Existen diversas fuentes de energías limpias alternativas como la solar y eólica, el problema es que la producción está restringida a la falta de energía solar y altas velocidades de viento o intermitencia respectivamente, en tanto que la energía geotérmica, hidráulica y bioenergía se limitan a la disponibilidad localizada de ciertos puntos energéticos y que difícilmente puedan ser portables en un medio móvil.

Por otro lado, la celda de combustible es dispositivo que convierte la reacción química de oxidación de un combustible y reducción de un oxidante (frecuentemente hidrógeno y oxígeno) en energía eléctrica. Entonces, las celdas de combustible en conjunto con los sistemas fotovoltaicos y aerogeneradores pueden acoplarse para almacenar energía en forma de hidrógeno y producir electricidad en periodos en los que no se disponga de energía solar o eólica. La utilización de las celdas de combustible para sistemas aún sigue en desarrollo y comúnmente está siendo investigada en laboratorios que poseen bastos recursos económicos dado el alto costo de los equipos instrumentales tales como el emulador de cargas eléctricas entre otros.

Una de las principales barreras para la investigación científica en los países en desarrollo es la falta de instrumentación e instalaciones. En este caso de celdas de combustible, los instrumentos de monitoreo comercial son caros y generalmente están fuera del límite del presupuesto de los científicos locales. El hardware de código abierto puede proporcionar una solución a las necesidades de equipamiento para los investigadores nacionales y otros países en desarrollo. La filosofía de código abierto [Coley, 2017] es dotar a los usuarios de esquemas, componentes, costos de material y disposición de circuitos, el hardware de código abierto se está empleando para muchos elementos que van desde microcontroladores [Davison, 2004], hasta redes de sensores inalámbricos [Bermann, 2010], y una variedad de diferentes equipos de laboratorio [Harnett, 2011]. Por otra parte, en cuanto a medición de corriente y voltaje para caracterizar el desempeño celdas solares, en el mercado existen trazadores de curvas I-V como se propone en [Tatiana, 2014] que emulan una corriente alta mediante un circuito de conexión y desconexión de cargas capacitivas para generar un impulso de corriente durante un tiempo transitorio como se muestra en la figura 1. En el esquema general trazador de curvas I-V la carga se desconecta temporalmente para activar una carga capacitiva durante un tiempo suficiente en donde es posible alcanzar corrientes transitorias muy altas al mismo tiempo que la medición de I-V son realizadas ($S1=B$ y $S2=1$) al término de la medición $S2$ se libera y $S3$ se activa durante un cierto tiempo para descargar el capacitor ($S2=0$ y $S3=1$) y poder comenzar en cualquier momento una nueva caracterización.

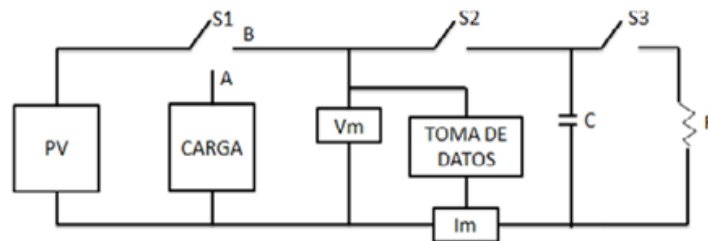


Figura 1 Esquema general Trazador de curvas I-V.

El esquema anterior resulta ser exclusivamente para evaluar celdas solares dado que la caracterización es durante un tiempo transitorio (micro o milisegundos)

determinado por la resistencia interna R_s de la celda PV y el capacitor C. El problema que presenta en el anterior esquema es que no puede ser utilizado para evaluar celdas de combustible dado que la reacción química del hidrogeno y oxigeno esta dado en el orden de los segundos. Por otro lado, el instituto de energías renovables IER del grupo de materiales solares de la UNAM con sede en Temixco Morelos México [Romero, 2015] ha desarrollado un equipo en proceso de patente el cual instrumenta sensores de corriente y voltaje dedicados a medir parámetros de celdas solares y de combustible por lo que su desarrollo se orienta enfáticamente al software. Sin embargo, en muchas ocasiones la carga tanto para celdas solares como de combustible no puede ser llevada al laboratorio por lo que la emulación de la carga es muy conveniente tal como lo proponemos en este trabajo. Por otro lado, [Ocampo, 2014] no solo diseñó la interfaz gráfica sino también diseñó el control del flujo y el control de cargas para caracterizador celdas de combustible, el problema principal es que la carga resistiva es transitoria lo cual es adecuada para celdas solares, pero no para celdas de combustible debido al tiempo largo de estabilización de los gases de la celda.

En este trabajo se propone el desarrollo e implementación de un prototipo experimental para evaluar celdas de combustible de 1.5 kW con elementos tecnológicos mínimos y suficientes capaces de proporcionar al usuario la gráfica de polaridad y la densidad de potencia. La intención de este desarrollo es minimizar el costo de fabricación ya que los equipos actualmente comerciables son extremadamente caros

Celdas de Combustible

Una celda de combustible es un dispositivo electroquímico que convierte la energía química de las reacciones de oxidación de un combustible y reducción de un oxidante, en energía eléctrica y calor [Morales, 2005]. Su concepto es similar al de una batería, consiste en la producción de electricidad mediante el uso de sustancias químicas, frecuentemente hidrógeno y oxígeno, donde el hidrógeno actúa como elemento combustible, y el oxígeno es obtenido directamente del aire. Hidrógeno + Oxígeno (del aire), es decir Electricidad + Calor + Agua.

Las celdas de combustible ofrecen muchas ventajas en comparación con los motores de combustión interna y las baterías, por ejemplo:

- No producen ninguna contaminación durante su operación y los únicos desechos son agua y calor.
- El agua producida tiene suficiente pureza para usarla como agua potable.
- Las celdas de combustibles son eficientes en un 40 – 50%, silenciosas y limpias.
- Reducción de peligro medioambiental inherente de las industrias extractivas.

Existen diversas aplicaciones de las celdas de combustible, tanto en C.C. estacionarias para aplicaciones de servicios públicos, redes de comunicación, como en la industria automotriz, algunos tipos de celdas de combustible existentes son:

- *Ácido fosfórico (PAFC)*: Su temperatura de operación es alrededor de los 220 °C. Es el tipo de C.C. más desarrollada a nivel comercial, su uso se encuentra en clínicas, hospitales, hoteles, edificios, escuelas, plantas eléctricas y terminales aeroportuarias. Con una eficiencia de más del 40% y cerca del 85% si el vapor es usado en cogeneración.
- *Polímero sólido (PEMFC)*: Con un rango de temperatura en operación de 50 a 100 °C, operan con una eficiencia de 40 a 60% y son capaces de manejar los grandes y repentinos cambios de potencia de salida, por eso son adecuadas para aplicaciones en automóviles y otros vehículos espaciales.
- *Carbonato fundido (MCFC)*: Con una temperatura de operación es alrededor de 600 °C, puede alcanzar de 50 a 60% de eficiencia y de 70 a 80% en aplicaciones de cogeneración, se implementan en aplicaciones estacionarias en servicios públicos y empresas, proporcionando energía primaria de alta calidad y energía de respaldo.
- *Oxido sólido (SOFC)*: Su temperatura de funcionamiento oscila alrededor de los 900 °C, estas C.C. tienen eficiencias eléctricas de 50 a 60% y de 70 a 80% en aplicaciones de cogeneración. Las SOFC son utilizadas en una

amplia gama de aplicaciones, desde pequeñas unidades de potencia auxiliar residencial que suministran calor y energía a los hogares hasta empresas grandes.

- *Metanol directo (DMFC)*: Operan en un rango de 50 a 60 °C, sus aplicaciones van desde pequeños aparatos electrónicos, como cargadores de baterías y portátiles, al igual que como energía estacionaria para copia de seguridad de las telecomunicaciones.
- *Alcalinas (AFC)*: Su temperatura de operación va desde 50 a 250 °C, con un potencial de eficiencia de 60% y de 80 a 90% en aplicaciones de cogeneración, son mejor conocidas por su participación en la NASA. Utilizan el hidrógeno como fuente de combustible y pueden fallar cuando se exponen a CO₂, es por ello por lo que se utilizan principalmente en la industria aeroespacial controlada y aplicaciones bajo el agua.

Curva de Polarización

Comúnmente la curva de polarización de una celda de combustible permite conocer el desempeño una pila o celda de combustible. En una curva de polarización, el eje de abscisas representa la densidad de corriente por unidad de área activa de la pila en A/cm² mientras que en el eje de ordenadas se representa la caída de tensión de la pila en Volts. En la figura 2 se representa la curva de polarización típica de una monocelda. Se observa que la curva está dividida en 3 regiones:

- Región 1 (Pérdidas por activación): Proviene de la energía de activación de las reacciones electroquímicas en los electrodos. Depende del material y la microestructura del electro catalizador y de la actividad química de los reactantes.
- Región 2 (Pérdidas Óhmicas): Se deben a la resistencia iónica en el electrolito y los electrodos, a la resistencia electrónica en los electrodos, colectores y a la resistencia de contacto. Son proporcionales a la densidad de corriente y dependen del tipo de material utilizado, la geometría de la celda y la temperatura.

- Región 3 (Pérdidas por concentración): Son el resultado de las limitaciones, debido a las tasas finitas de transferencia de masa de los reactantes y dependen fuertemente de la densidad de corriente y la estructura de los electrodos y las placas.

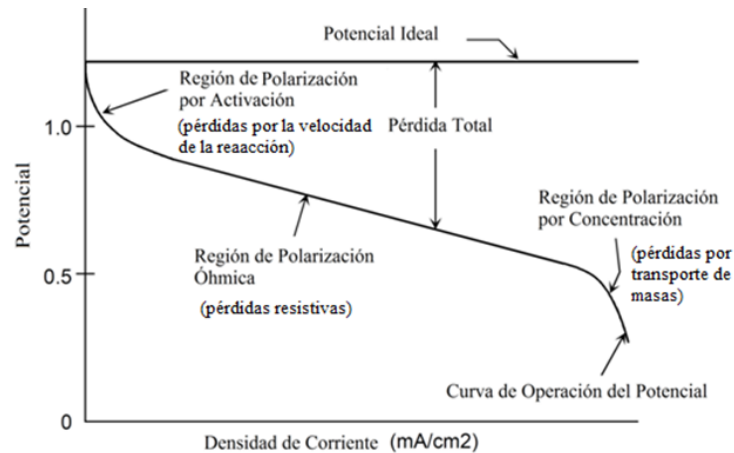


Figura 2 Curva de polarización típica de una pila de combustible tipo PEM.

El voltaje de la celda representa la máxima tensión que se puede extraer de la pila de combustible, siempre y cuando no existan pérdidas térmicas. La ecuación que describe la curva de polarización se muestra en la ecuación 1.

$$V_{cell} = E^{\circ}(T, P) - V_{act_ánodo} - V_{act_cátodo} - V_{ohmico} - V_{con_ánodo} - V_{con_cátodo} \quad (1)$$

Donde

V_{cell}	Tensión en la monocelda de combustible (V)
$E^{\circ}(T, P)$	Potencial de Nerst (V)
V_{act_x}	Pérdidas por activación ya sea en ánodo o cátodo (V)
V_{con_x}	Pérdidas por concentración (V)
V_{ohmico}	Pérdida óhmica (V)

En la ecuación 2 se define el potencial de Nerst siempre y cuando los gases se comporten como gases ideales.

$$E^{\circ}(T, P) = \frac{-\Delta G^{\circ}(T)}{nF} + \frac{RT}{nF} \cdot \ln \left[\frac{\left(\frac{y_{H_2} P_{ánodo}}{P^{\circ}} \right) \left(\frac{y_{O_2} P_{cátodo}}{P^{\circ}} \right)^{1/2}}{\frac{y_{H_2O} P_{cátodo}}{P_{sat}(T)}}} \right] \quad (2)$$

Donde:

$\Delta G^\circ(T)$	Variación de la función de Gibbs (J)
n	Número de electrones transferidos (2 para una PEM)
F	Constante de Faraday (96486 C/eq)
R	Constante de los gases ideales (8.314 J/(mol K))
Y_i	Fracción molar de la especie i
P_i	Presión en ánodo o cátodo (Pa)
P°	Presión de referencia (101325 Pa)
$P_{sat}(T)$	Presión de saturación del agua a la temperatura T (Pa)

En las ecuaciones 3 y 4 se muestra el modelo simplificado de pérdidas por activación en el ánodo y el cátodo respectivamente, lo anterior siempre y cuando la densidad de corriente sea de un valor bajo.

$$V_{act_ánodo} = \frac{RT}{\alpha_{ánodo} F} \cdot \ln\left(\frac{i}{i_{o,ánodo}}\right) \quad (3)$$

$$V_{act_ánodo} = \frac{RT}{\alpha_{ánodo} F} \cdot \ln\left(\frac{i}{i_{o,ánodo}}\right) \quad (4)$$

Donde

i_o	Densidad de corriente de referencia (A/cm ²)
α	Coefficiente de transferencia de carga (Adimensional)
i_{cell}	Densidad de corriente de la celda (A/cm ²)

2. Métodos

La configuración experimental del prototipo experimental para trazar curvas V-I se muestra en la figura 3. La innovación de este desarrollo se enfoca en hacer uso de recursos tecnológicos mínimos y suficientes para disminuir los costos de fabricación. En este sentido, el prototipo propuesto tiene como unidad central de proceso una tarjeta de bajo costo Raspberry Pi, con periféricos de bajo costo como teclado y mouse inalámbricos y como elemento visualizador un monitor. Además, los sensores de voltaje y actuadores de relevadores son totalmente aislados para garantizar una medición y actuación correcta.

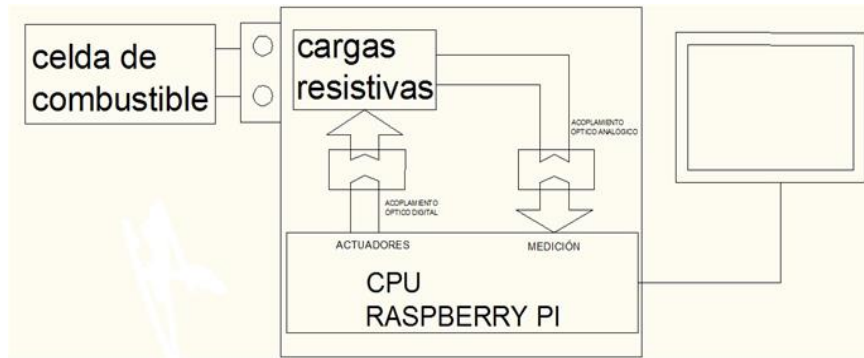


Figura 3 Configuración experimental propuesto para evaluación de celdas de combustible.

En la figura 4 se muestra de manera general un diagrama de tubería e instrumentación DTI normalizado el cual es utilizado para explicar el funcionamiento del módulo de control de carga propuesto.

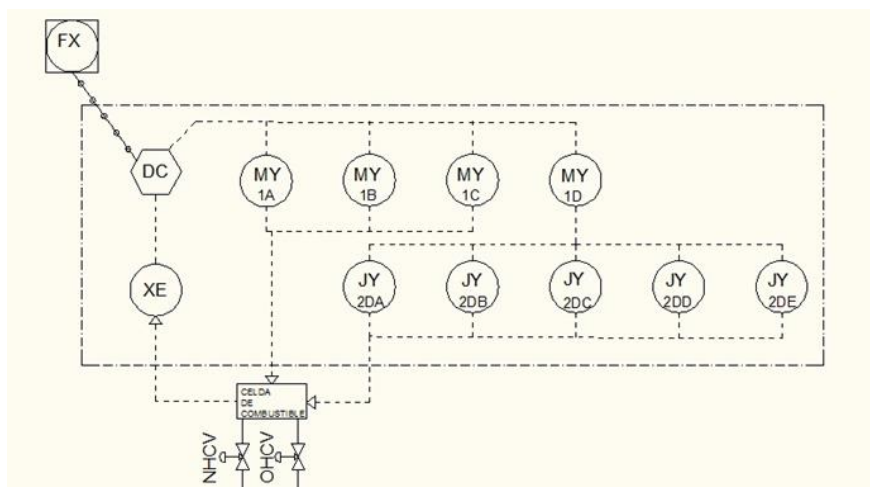


Figura 4 Diagrama de tubería e instrumentación del módulo de control de cargas.

De la figura anterior el DC (dispositivo de control Raspberry) comanda una señal de activación a los MY (módulos de relevadores) 1A, 1B, 1C, 1D, el MY 1D a su vez activa secuencialmente los JY (relevadores de potencia) 2DA, 2DB, 2DC, 2DD, 2DE. La carga activada es aplicada a la celda de combustible, el resultado de la prueba se puede visualizar en el FX (monitor), a su vez la celda de combustible es alimentada normalmente por hidrógeno y oxígeno a través de dos válvulas manuales NHCV y OHCV respectivamente.

Análisis de las Cargas Resistivas Empleadas

Las cargas del módulo propuesto se basan en la conexión en paralelo de cargas totalmente resistivas por lo que la resistencia equivalente R_t esta dada por ecuación 5.

$$R_t = \frac{1}{\sum_{m=1}^n \left(\frac{1}{R_n} \right)} \quad (5)$$

Considerando al conjunto de resistencias iguales, entonces se ecuación 6.

$$\sum_{m=1}^n \left(\frac{1}{R_n} \right) = \frac{n}{R} \quad (6)$$

Y simplificando ecuación 1, se obtiene ecuación 7.

$$R_t = \frac{R}{n} \quad (7)$$

Para un valor de n taps o eventos de carga cualquiera, ecuaciones 8 a la 11.

$$R_{Unidades_de_n} = \frac{100\Omega}{unidades_de_n} \quad (8)$$

$$R_{Decenas_de_n} = \frac{10\Omega}{decenas_de_n} \quad (9)$$

$$R_{Centenas_de_n} = \frac{1\Omega}{centenas_de_n} \quad (10)$$

$$R_{Millares_de_n} = \frac{0.1\Omega}{millares_de_n} \quad (11)$$

Entonces la resistencia equivalente para un valor de n taps R_{t_n} es

$$R_{t_n} = \frac{1}{\frac{1}{R_{unidades}} + \frac{1}{R_{decenas}} + \frac{1}{R_{centenas}} + \frac{1}{R_{millares}}} \quad (12)$$

Por ejemplo, para $n = 125$ taps o muestras:

$$R_{Unidades}=5, R_{Decenas}=20, R_{Centenas}=1, R_{Millares}=a \text{ entonces } R_{t_n}=0.8 \text{ ohms}$$

Sensores y Actuadores

Comúnmente para medición de corriente y voltaje de baja potencia no es necesario utilizar un aislamiento eléctrico de medición, cuando se utiliza la misma tierra. Sin embargo, cuando se manejan potencias elevadas el dv/dt suele incrementarse de manera significativa y provocar daños en la tarjeta de control. Para evitar problemas asociados a lo anterior, se diseñó un sensor de voltaje con aislamiento óptico utilizando un optoacoplador lineal IL300 como se muestra en la figura 5.

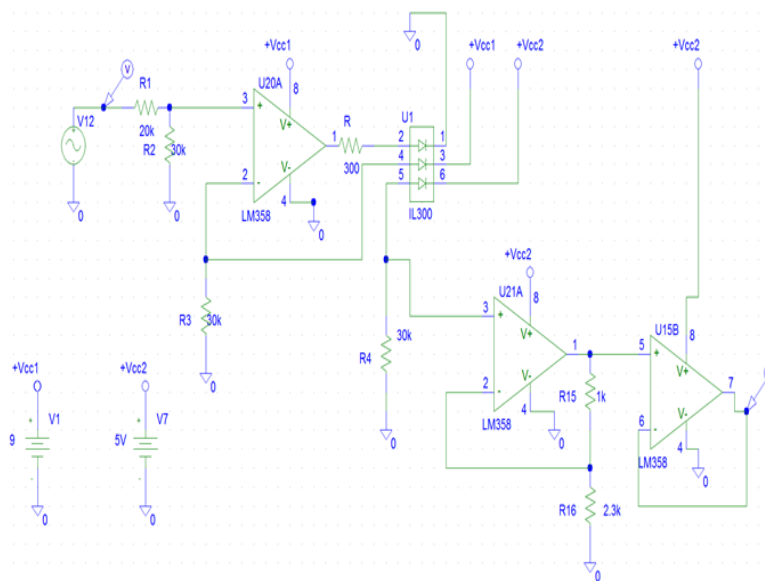


Figura 5 Circuito de medición de voltaje con el IC IL300.

Con respecto a los actuadores se utilizaron módulos de relevadores integrados con entradas optoacopladas lo cual no solo facilita la implementación, sino que también aíslan la etapa de potencia con el circuito de control, brindando seguridad al equipo.

3. Resultados

En la figura 6 se muestra el módulo de control de cargas desarrollado e implementado en la Universidad Autónoma del Carmen diseñado para evaluar celdas de combustible.

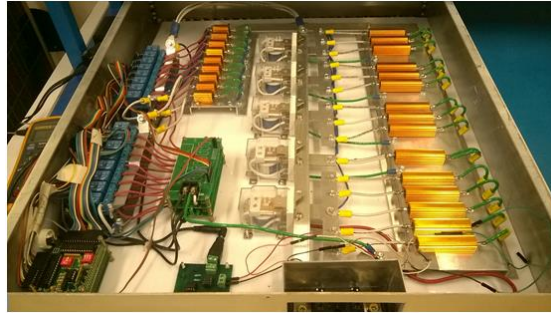


Figura 6 Módulo de control de cargas resistivas.

En la figura 7 se muestra la configuración experimental durante la evaluación de una celda de combustible de 4 ensambles con un área activa de 9 cm².

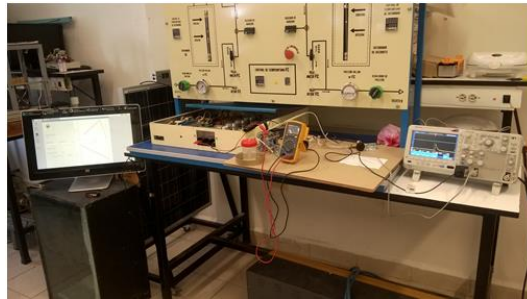


Figura 7 Configuración experimental durante la evaluación de una celda de combustible.

En la figura 8 se muestra el comportamiento de la activación de cargas resistivas en función del número de taps en donde es posible observar que el comportamiento no es lineal.

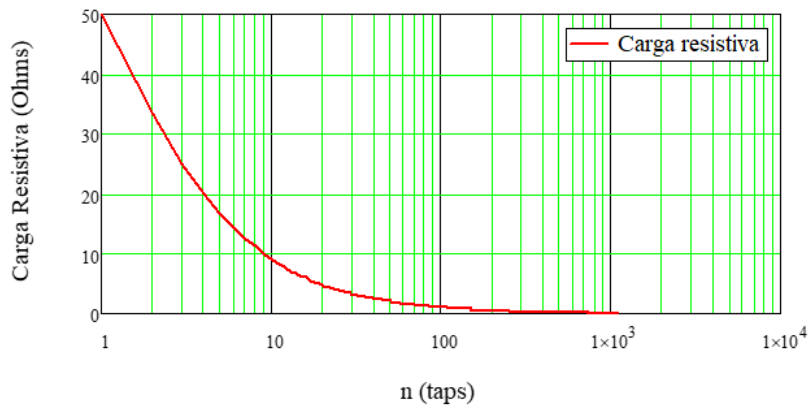


Figura 8 Carga resistiva Vs taps.

La figura 9 muestra la curva de polarización obtenida con la HMI desarrollada con la tarjeta de control Raspberry Pi en donde se observa el comportamiento característico.

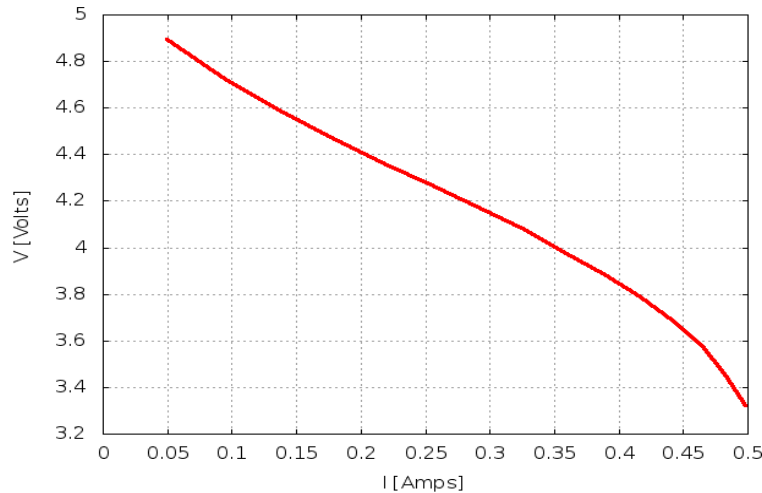


Figura 9 Curva de polarización obtenida con la HMI.

En la figura 10 se muestra la respuesta de una celda con insuficiente hidrógeno lo cual deja observar que el voltaje de inicio es de 2.5 volts en lugar de 4.8 volts (4 celdas de 1.2 volts) además se observa que cada vez que se aplica una carga, para este caso cada 3 segundos, la celda logra estabilizarse y el gradiente de voltaje cada vez es menor por lo que la corriente llegará a un valor máximo y de ahí en adelante disminuirá la corriente, tal como se muestra en la figura 11.

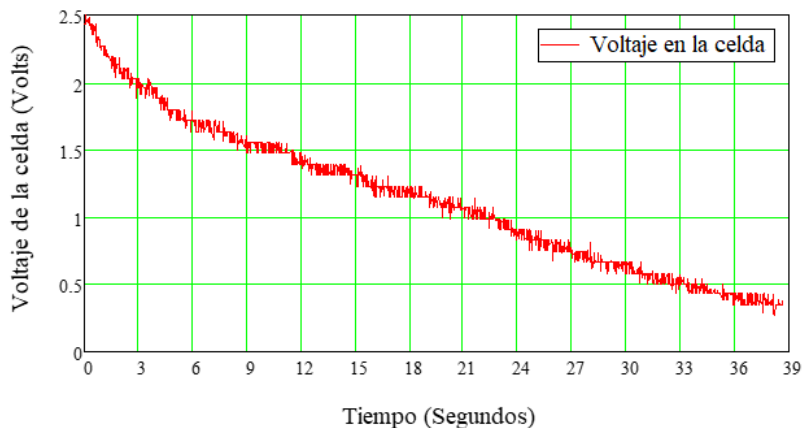


Figura 10 Estabilización del voltaje en la celda cada 3 segundos de conmutación.

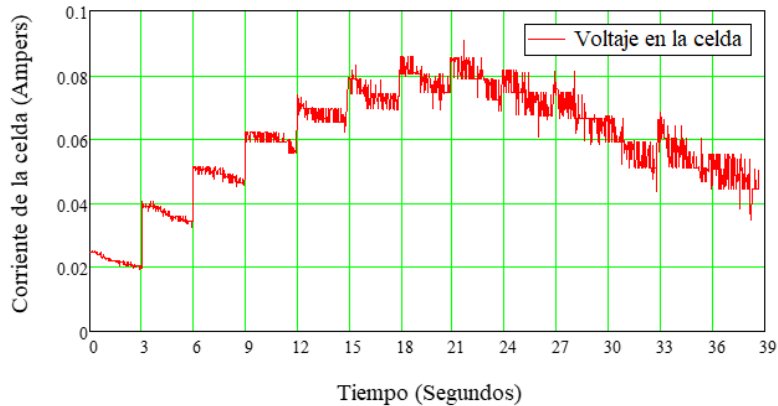


Figura 11 Estabilización de la corriente en la celda cada 3 segundos de conmutación.

La carga resistiva aplicada cada 3 segundos depende de la ecuación 12 para un valor de $n = 13$ es decir 100, 50, 33.3, 25, 20, 166, 14.3, 12.5, 11.1, 10, 9.9, 8.33 y 7.69 Ω , respectivamente cada 3 segundos.

4. Discusión

Cargas Eléctricas

A partir de la ecuación 12, el valor de Rt_n es posible obtenerlo en función de n . Sin embargo, la función inversa solo puede ser obtenida de manera precisa utilizando métodos numéricos o utilizando la tabla de identificación y de manera aproximada utilizando la gráfica representada en la figura 8. El decremento resistivo no lineal sobre esta misma gráfica podría parecer una desventaja, sin embargo, no lo es porque con pocos taps es posible alcanzar corrientes muy grandes y muchos taps el sistema tiene un mejor control del manejo de corriente.

Sensores y Actuadores Aislados

Uno de los elementos esenciales de este prototipo es la correcta medición del voltaje en el bus+ de la celda de combustible. Es decir, la interferencia inducida en circuito a tierra se disminuyó utilizando el punto de medición lo más cercano físicamente a los bordes del bus+ de la celda, La interferencia capacitiva y electromagnética se reduce utilizando un gabinete metálico con conexión a tierra física y las inductancias parásitas de cableado en la etapa de potencia se reducen utilizando un diseño robusto con rieles de conexión y cableado estructurado.

Control

La interfaz HMI desarrollada en la tarjeta Raspberry pi tiene como opciones principales:

- Controlar el tiempo de cada tap: esto es para lograr la estabilización de la reacción química y mantener un voltaje constante de la celda
- Controlar el número de tap: esto es alcanzar una corriente máxima dependiendo el voltaje de la celda bajo prueba
- Controlar el voltaje de referencia: esto es tener mejor resolución en la medición.
- Disponibilidad de la curva de polarización y densidad de potencia.
- Grabado de curvas en formato digital.

5. Conclusiones

- El costo total de la implementación no sobrepasa a los \$13000 pesos mexicanos.
- El diseño robusto garantiza una confiable y precisa medición, Por otro lado, la exactitud de la medición seguirá siendo ajeno al desarrollo, dado que es problema principal de una certificación.
- La posible utilización de relevadores de estado sólido en lugar de mecánicos pudiera ser de gran provecho si el equipo estuviera trabajando de manera de manera continua.
- La corriente no es medida dado que es posible estimarla utilizando el voltaje medido y la carga resistiva por cada tap. Lo anterior podría parecer desventaja, pero no lo es, dado que se disminuyen aún más los costos, sin afectar el desempeño.

6. Bibliografía y referencias

- [1] G. Coley. Take advantage of open source hardware internet: <http://www.edn.com/design/systems-design/4313253/Take-advantage-of-open-source-hardware>, 25 mayo 2017.

- [2] C. Harnett. Open-source hardware for instrumentation and measurement. Instrumentation and measurement, Magazine, IEEE, June 2011.
- [3] Laura Romero., Open-source hardware. Gaceta digital UNAM, no. 4742, pp. 12, Lunes 23 de noviembre de 2015.
- [4] Morales Coutiño M., Desarrollo de un emulador de celdas de combustible. Tesis de maestría, Centro Nacional de Investigación y Desarrollo Tecnológico. Cuernavaca, Morelos, 2005.
- [5] N. Bergmann. Low cost prototyping system for sensor networks, Sensor, Sensor network, pp.19-24, 2010.
- [6] Rafael Ocampo Martínez, Diseño y Construcción de un Sistema de Caracterización de Celdas de Combustible y De Adquisición de Datos con Labview. Tesis de Maestría en Energías Renovables por parte del Centro de Investigación en Materiales Avanzados CIMAV México, Nov. 2014.
- [7] S. Davinson, Open-source hardware. IEEE Desing and Test of Computers, vol. 21, no. 5, pp. 446-456, 2004.
- [8] Tatiana Vargas y Augusta Abrahamse, An open-source hardware I-V Curve Trace for Monitoring PV Output in Bolivia. Investigación y Desarrollo, vol. 1, no. 14, pp. 100-116, 2014.

EFFECTO DE LA LONGITUD DEL DIÁMETRO EN LA ESTABILIDAD TÉRMICA DE LA CAPA LIBRE DE LAS MEMORIAS RAM MAGNÉTICAS

Marco A. Escobar

Universidad de La Salle Bajío
maescobar@delasalle.edu.mx

Rafael Guzmán Cabrera

Universidad de Guanajuato
maescobar@delasalle.edu.mx

Miguel Torres Cisneros

Universidad de Guanajuato
maescobar@delasalle.edu.mx

Jorge Ramón Parra Michel

Universidad de La Salle Bajío
maescobar@delasalle.edu.mx

Rafael Martínez Peláez

Universidad de La Salle Bajío
maescobar@delasalle.edu.mx

Resumen

Una de las nuevas aplicaciones del magnetismo en medios de almacenamiento de datos son las memorias RAM magnéticas (MRAM, por sus siglas en inglés). El periodo de tiempo que se puede mantener un bit en una MRAM está íntimamente relacionado con la estabilidad térmica del dispositivo, la cual depende de las propiedades de los materiales utilizados y de la geometría. En el presente trabajo presentamos un estudio de cómo afecta el diámetro del dispositivo a la estabilidad térmica de una MRAM. A partir de los resultados obtenidos es posible explicar que

al incrementar del diámetro de una MRAM, en algún punto el proceso de inversión de la magnetización deja de ser una rotación coherente y se convierte un movimiento de pared de dominio, lo cual a su vez ocasiona que la barrera de energía no sea proporcional al volumen, presentándose una disminución en el valor de la barrera de energía.

Palabras Claves: Almacenamiento, Estabilidad, MEP, MRAM.

Abstract

One of the novel applications of magnetism in data storage is the use of Magnetic Random Access Memories (MRAM). The period that a bit can be stored in such a device is closely related to the thermal stability, which in turn depends on the materials used, and on the geometry. In the present work, we performed a study on the effect of the junction diameter on the thermal stability of an MRAM. From our results it is possible to explain the reason why when the diameter of an MRAM is increased the reversal process goes from coherent rotation to domain wall movement, leading to a decrease in the energy barrier.

Keywords: MEP, MRAM, Storage, Stability.

1. Introducción

En la última década se especuló que la nueva tecnología de memorias RAM magnéticas (MRAM) tiene el potencial de desplazar a algunas soluciones actuales [Slaughter, 2016]. Esto por su bajo consumo de energía, gran rapidez de operación, por tratarse de una memoria no volátil y que además, en principio, posee un cierto grado de compatibilidad con la tecnología CMOS [Khvalkovskiy, 2013].

Una de las propuestas para desarrollar MRAMs es la STT-MTJ, la cual basa su funcionamiento de escritura en el fenómeno conocido como transferencia de spin-torque (STT, spin transfer torque) en una *Magnetic Tunnel Junction* (MTJ), y para la lectura en el fenómeno *Tunneling Magnetoresistance* (TMR). La estructura de un STT-MTJ consiste de al menos dos películas magnéticas delgadas separadas por una película aislante. La información está guardada en el estado de

magnetización de una de las películas magnéticas a la cual se le conoce como capa libre (FL). La segunda película recibe el nombre de capa de referencia (RL) y provee un marco de referencia para la lectura y escritura de la información.

En el caso de escritura, el fenómeno STT permite que los electrones que pasan a través de la unión MTJ transfieran el momento angular del spin entre las capas magnéticas, lo cual resulta en un torque de la magnetización en la FL. Por lo tanto, el estado de la magnetización en dicha capa puede ser modificado si un torque suficientemente fuerte es aplicado.

Por otro lado, la lectura se da a partir del fenómeno de TMR, en el cual la resistencia de la MTJ depende fuertemente de la orientación relativa de la magnetización entre las FL y RL. Una representación de una célula STT-MRAM, del llamado tipo perpendicular al plano, se presenta en la figura 1. En esta, se representa en amarillo la capa FL, en negro una capa aislante, las regiones azules y verde en su conjunto representan la capa RL, la capa verde representa un material que proporciona acoplamiento anti-ferromagnético, y las flechas representan los estados de magnetización estables posibles para cada una de las capas. Típicamente la FL presenta una anisotropía uni-axial grande, la cual impide que el estado de la magnetización cambie debido a fluctuaciones térmicas. Un sistema de almacenamiento magnético debe ser estable a las fluctuaciones térmicas, la estimación de dicha estabilidad es una parte fundamental de la teoría magnética. En presencia de una perturbación ocasionada por efectos térmicos se puede ocasionar la llamada activación térmica, la cual puede ocasionar una transición de un sistema magnético de un estado estable a otro, pudiendo así causar pérdida de información. En caso de análisis presentado, la transición puede ocasionar que un "1" lógico pase a un "0" lógico. La obtención de resultados computacionales para sobre la estabilidad térmica nos permite explicar a detalle los procesos que intervienen en la inversión de la magnetización y proveen una pauta para el diseño de dispositivos estables [Uhlir, 2013].

En el presente trabajo realizamos un estudio sobre el efecto que tiene la longitud del diámetro de la MTJ en la estabilidad térmica. Para calcular la estabilidad térmica generalmente se estima la barrera de energía entre dos estados estables,

esto se estudia utilizando el *Nugded Elastic Band Method* (NEB), el cual se describe en la siguiente sección. En la sección de resultados se describe la geometría y los materiales de las MRAMs bajo estudio, y con los cuales se obtiene una estimación bastante cercana a los resultados experimentales existentes en la literatura [Sato, 2011]. En la sección de discusión se presenta un análisis de los resultados obtenidos y finalmente se concluye que al incrementar el diámetro de la MRAM más allá de un punto el camino de mínima energía (MEP, *minimum energy path*) pasa de una rotación coherente, al movimiento de la pared entre dominios (DWM, *domain wall motion*), lo cual implica que la barrera de energía no sigue a la curva $E_b=K_{eff}V$ donde K_{eff} [$J\ m^{-3}$] es la anisotropía efectiva y V [m^3] es el volumen de la capa FL.

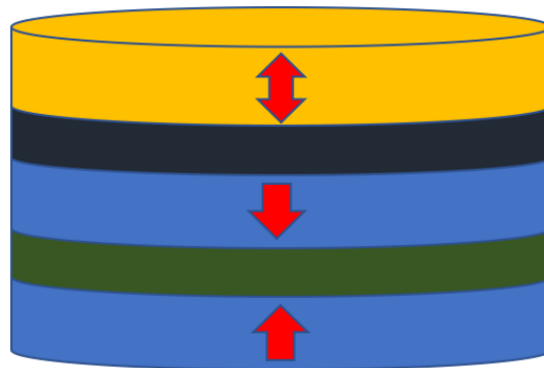


Figura 1 Representación gráfica de una MRAM perpendicular.

2. Métodos

Los simuladores micromagnéticos son ampliamente usados para investigar nuevas nanotecnologías como las memorias *racetrack*, o la lógica magnética [Lubarda, 2012]. Dada la tendencia de miniaturización existente para alcanzar una alta densidad de componentes por centímetro cuadrado, la estabilidad térmica de los componentes magnéticos se ha visto comprometida. Cuando algún componente magnético posee dimensiones nanométricas las fluctuaciones térmicas existentes, incluso a temperatura ambiente, ocasionan una transición entre dos estados estables. La mayoría de los componentes comerciales requieren una estabilidad de al menos 10 años.

Una metodología para estudiar la estabilidad térmica de medios magnéticos es el NEB, el cual fue adaptado de la química al micromagnetismo [Dittrich, 2002]. El método busca la transición más probable entre un estado inicial y un final, usando el principio de mínima energía mediante el cual se obtiene el MEP, y por consiguiente la llamada energía de barrera, E_B [J]. A partir del resultado obtenido se puede estimar el tiempo de relajación sobre una barrera de energía usando la ley de Arrhenius-Néel [Tudosa, 2012], ecuación 1.

$$\tau = \frac{1}{f_0} e^{\frac{E_B}{k_B T}} \quad (1)$$

Donde f_0 [1/s] es una constante que depende de la geometría, $k_b = 1.38 \times 10^{-23}$ J/K es la constante de Boltzmann, y $T = 300$ K es la temperatura ambiente.

En este método se propone una transición solución, la cual se discretiza en una secuencia de imágenes. Por medio de un proceso de relajación, dichas imágenes tienen la libertad de moverse en la dirección que minimiza la energía requerida para realizar la transición, este proceso matemáticamente toma la forma de ecuación 2.

$$\frac{dE}{du} = -\nabla E - (\nabla E \cdot \hat{t})\hat{t} \quad (2)$$

Donde E [J] es la energía total de una de las imágenes, ∇E es el gradiente de la energía con respecto a la magnetización, \hat{t} es un vector tangente a las imágenes, que en principio puede conectar una imagen determinada con la imagen siguiente y la anterior, y u es una variable que carece de significado físico real y que se introduce únicamente con el propósito de relajar el sistema.

Dentro de la llamada aproximación micromagnética se considera el vector de magnetización como una variable continua y al no considerar variaciones de la amplitud causadas por la temperatura toma la forma de $\mathbf{M} = M_s \hat{\mathbf{m}}(\mathbf{r}, t)$, donde M_s [A m⁻¹] es conocida como la polarización de saturación, \mathbf{r} [m] es la posición y t [s] es el tiempo. Dicho de otra forma, en esta aproximación no consideramos ni a los átomos ni a los electrones de forma individual. Una consecuencia inmediata de esta aproximación es que los fenómenos físicos se representan dentro de un

material perfectamente bien usando diferencias finitas (FDM, *finite differences method*) o elemento finito (FEM, *finite element method*).

En el caso de cálculos micromagnéticos, la densidad de energía total del sistema se relaciona con el campo efectivo H_{eff} [A m⁻¹] por medio de una derivada funcional con respecto al vector de magnetización \hat{m} [Schrefl, 2007], ecuación 3.

$$H_{eff} = -\frac{1}{\mu_0 M_s} \frac{\partial \mathcal{E}}{\partial m \partial V} \quad (3)$$

Con $\mu_0 = 4\pi \times 10^{-7}$ NA⁻² la permeabilidad del vacío.

El campo efectivo a su vez contiene contribuciones de diversos campos, ecuación 4.

$$H_{eff} = H_d + H_a + H_{ext} + H_{exc} \quad (4)$$

En donde podemos identificar al campo magnetostático H_d , ecuación 5.

$$H_d = \frac{1}{4\pi} \iiint_V \frac{\nabla' \cdot M_s \hat{m}}{\|r - r'\|^3} dr'^3 - \frac{1}{4\pi} \iint_S \frac{M_s \hat{m} \cdot \hat{n}'}{\|r - r'\|^2} dr'^2 \quad (5)$$

El campo de anisotropía uniaxial H_a , ecuación 6.

$$H_a = \frac{2K_u}{\mu_0 M_s} (\hat{m} \cdot \hat{k}) \hat{k} \quad (6)$$

Donde K_u [Jm⁻³] es una constante para el caso de anisotropía uniaxial, y el vector unitario \hat{k} representa el eje de dicha anisotropía.

Existe contribución del campo originado por la interacción de intercambio H_{exc}

$$H_{exc} = \frac{2A}{\mu_0 M_s} (\nabla^2 \hat{m}) \quad (7)$$

Donde a A [Jm⁻¹] se le conoce como constante de intercambio. También puede existir contribución de algún campo externo H_{ex} .

Una exposición detallada sobre como calcular los elementos que conforman el vector tangente \hat{t} es demasiado extensa para ser incluida aquí, pero se puede consultar la referencia [Escobar, 2016].

3. Resultados

Para realizar los cálculos utilizamos un código basado en Fortran [Chang, 2011] y que utiliza la librería SUNDIALS [Hindmarsh, 2005] para la solución de las ecuaciones diferenciales no lineales. Además, las propiedades de los materiales y la geometría están inspiradas en los experimentos presentados en la referencia [Sato, 2011] y la validez del estudio se toma a partir de la congruencia obtenida con dichos resultados experimentales.

El dispositivo se simplifica tomando en consideración únicamente la FL, la cual consiste en un cilindro de altura 1 nm y el diámetro D toma los valores de 45, 50, 60, 70 y 75 nm. En todos los casos la transición fue dividida en 10 imágenes y en todos los casos la estimación inicial de solución fue una rotación coherente del vector de magnetización. Las propiedades del material se resumen en la tabla 1. La anisotropía uniaxial está orientada en la dirección de la altura de la celda MRAM.

Tabla 1 Propiedades del material utilizado en los cálculos

Propiedad	Valor
M_s	$1.27 \times 10^7 \text{ A m}^{-1}$
A	12 pJm^{-1}
K_u	1 MJm^{-3}

Para ejemplificar el procedimiento, en la figura 2 se presenta la energía de barrera, E_b [J], de una estimación inicial a la solución para una MRAM de diámetro $D = 70$ nm, es común normalizar los resultados usando $k_b T$, con k_b la constante de Boltzman y $T = 300$ K, obteniendo en este caso un resultado de $E_b = 70 k_b T$.

Si comparamos con la barrera de energía ($E_b = 45 k_b T$) del MEP que se presenta en la figura 3, podemos observar una gran diferencia entre la estimación inicial y el resultado. Esto se debe a que el MEP es en realidad un movimiento de la pared entre dominios un ejemplo se muestra en la secuencia de imágenes en la figura 4, en donde un estado estable de magnetización se representa por vectores apuntando hacia abajo (azul) y el otro estado estable de magnetización por vectores apuntando hacia arriba (rojo).

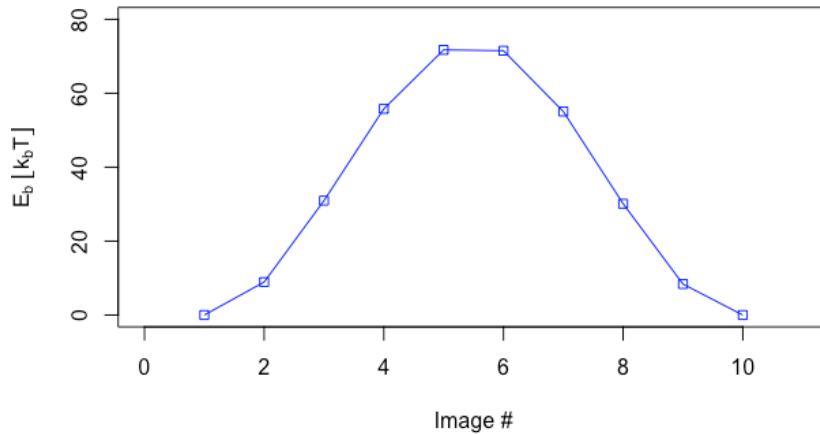


Figura 2 Barrera de energía de una MRAM de 70 nm de diámetro en una transición de rotación coherente del estado de magnetización.

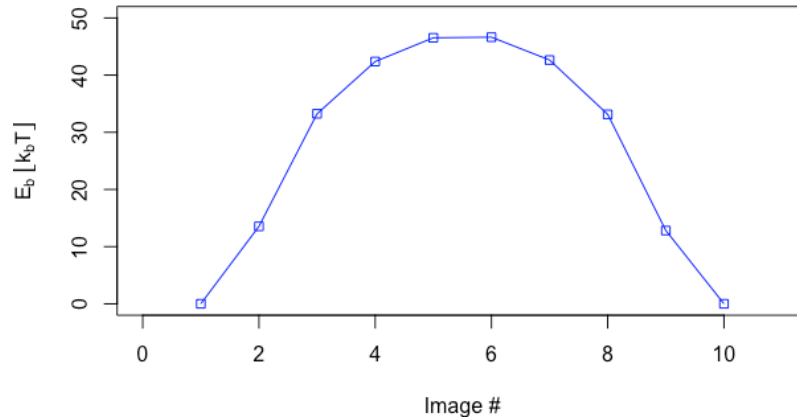


Figura 3 Camino de mínima energía de una MRAM de 70 nm de diámetro.

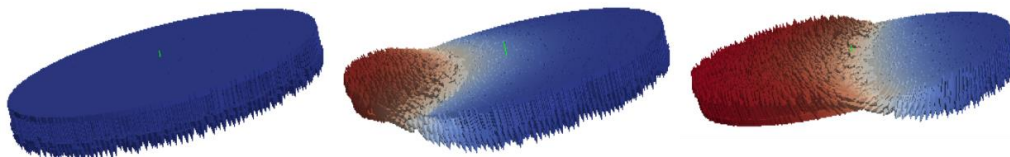


Figura 4 Representación de la secuencia de transición entre dos estados estables.

Finalmente, en la figura 5 se muestra los valores de la barrera de energía obtenidas al encontrar el MEP para cada una de las MRAMs bajo estudio. Es posible observar que al incrementar el diámetro más allá de los 60 nm, la barrera de energía de la MRAM se reduce en vez de aumentar.

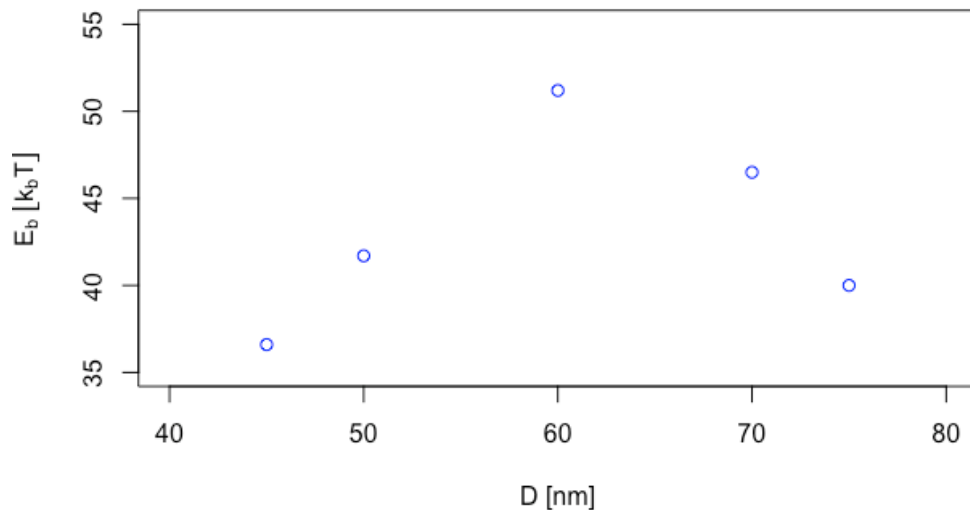


Figura 5 Barrera de energía para MRAMs de 45, 50, 60, 70 y 75 nm de diámetro.

4. Discusión

Para entidades de un solo dominio, la barrera de energía se calcula por medio de $E_b = K_{eff}V$, de tal suerte que al aumentar el diámetro de un cilindro la barrera de energía debiera aumentar de forma cuadrática. Es fácil observar a partir de la figura 5 que esta tendencia no se sigue y esto es porque al incrementar el diámetro llegamos a un punto en que la formación de paredes de dominio es energéticamente favorable y por tanto el MEP ya no consiste en una rotación coherente. Esto concuerda bien la discusión presentada en [Dittrich, 2002], en la cual se muestra que al incrementar la longitud de un dispositivo el movimiento de paredes de dominio es favorecido por el principio de mínima energía. Una consecuencia lógica de esto es que el incremento del volumen de una MRAM disminuye la estabilidad térmica de esta, por tanto las herramientas computacionales como el NEB cobran importancia, ya que nos permiten hacer estudios de estabilidad y potencialmente nos permitirán optimizar los diseños. Es importante recalcar que a pesar de las simplificaciones hechas en la geometría los resultados tienen validez puesto que el principal efecto de agregar las capas que conforman el RL es del de introducir un *offset* en la barrera de energía, no así el considerar un arreglo de MRAMs ya que al hacerlo se puede obtener barreras asimétricas, este efecto se discute en la referencia [Escobar, 2016].

5. Conclusiones

Por medio del NEB se realizaron cálculos del MEP para estimar la barrera de energía de la capa libre de MRAMs perpendiculares. Los cálculos realizados concuerdan bien con los resultados experimentales de la referencia [Sato, 2011]. Se puede apreciar que al aumentar el diámetro de la capa libre de una MRAM en algún punto la barrera de energía empieza a decrecer. Esto es atribuido a un cambio en el MEP que pasa de una rotación coherente de la magnetización a consistir en un movimiento de la pared entre dominios.

Agradecimientos

El autor MAE desea agradecer al Prof. Vitaliy Lomakin de la Universidad de California San Diego por fructíferas discusiones respecto al tema. Los autores MAE, JRPM, RMP agradecen la Universidad de La Salle Bajío el apoyo en la realización del presente trabajo.

6. Bibliografía y Referencias

- [1] Chang, R., Li, S., Lubarda, M. V., Livshitz, B., & Lomakin, V. FastMag: Fast micromagnetic simulator for complex magnetic structures. *Journal of Applied Physics*, Vol. 109, No. 7, 07D358, 2011.
- [2] Dittrich, R., Schrefl, T., Suess, D., Scholz, W., Forster, H., & Fidler, J. A path method for finding energy barriers and minimum energy paths in complex micromagnetic systems. *Journal of Magnetism and Magnetic Materials*, Vol. 250, 12–19, 2012.
- [3] Escobar, M. A. Efficient Micromagnetics for Magnetic Storage Devices. Tesis de doctorado. Universidad de California San Diego. San Diego, USA, 2016.
- [4] Khvalkovskiy, V., Apalkov, D., Watts, S., Chepulskaa, R., Beach, R. S., Ong, A., Tang, X., Driskill-Smith, A., Butler, W. H., Visscher, P. B., Lottis, D., Chen, E., Nikitin, V., & Krounbi, M. Basic principles of STT-MRAM cell operation in memory arrays. *Journal of Physics D: Applied Physics*, Vol. 46, No. 7, 74001, 2013.

- [5] Hindmarsh, A. C., Brown, P. N., Grant, K. E., Lee, S. L., Serban, R., Shumaker, D. E., & Woodward, C. S. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software*, Vol. 31, No. 3, 363–396, 2005.
- [6] Lubarda, M. V., Escobar, M. A., Li, S., Chang, R., Fullerton, E. E., & Lomakin V. Domain wall motion in magnetically frustrated nanorings. *Physical Review B - Condensed Matter and Materials Physics*, Vol. 85, No. 21, 214428, 2012.
- [7] Slaughter, J. M., Nagel, K., Whig, R., Deshpande, S., Aggarwal, S., DeHerrera, M., Janesky, J., Lin, M., Chia, H.-J., Hossain, M., Ikegawa, S., Mancoff, F. B., Shimon, G., Sun, J. J., Tran, M., Andre, T., Alam, S. M., Poh, F., Lee, J. H., Chow, Y. T., Jiang, Y., Liu, H. X., Wang, C. C., Noh, S. M., Tahmasebi, T., Ye, S. K., & Shum, D. Technology for reliable spin-torque MRAM products. *IEEE International Electron Devices Meeting (IEDM)*. San Francisco, USA, Diciembre 2016.
- [8] Sato, H., Yamanouchi, M., Miura, K., Ikeda, S., Gan, H. D., Mizunuma, K., Koizumi, R., Matsukura, F., & Ohno, H. Junction size effect on switching current and thermal stability in CoFeB/MgO perpendicular magnetic tunnel junctions. *Applied Physics Letters*, Vol. 99, No. 4, 1–3, 2011.
- [9] Schrefl, T., Hrkac, G., Bance, S., Suess, D., Ertl, O., & Fidler, J. Numerical Methods in Micromagnetics (Finite Element Method). *Handbook of Magnetism and Advanced Magnetic Materials*. Wiley. USA. 1–30, 2007.
- [10] Tudosa, I., Lubarda, M. V., Chan, K. T., Escobar, M. A., Lomakin, V. & Fullerton, E. E. Thermal stability of patterned Co/Pd nanodot arrays. *Applied Physics Letters*, Vol. 100, No. 10, 102401, 2012.
- [11] Uhlir, V., Urbánek, M., Hladík, L., Spousta, J., Im, M.-Y., Fischer, P., Eibagi, N., Kan, J. J., Fullerton, E. E., & Sikola, T. Dynamic switching of the spin circulation in tapered magnetic nanodisks. *Nature Nanotechnology*, Vol. 8, 341–6, 2013.

DISEÑO Y CONSTRUCCIÓN DE PROTOTIPO DE ENTRENAMIENTO PARA PRÁCTICAS EN INSTRUMENTACIÓN Y CONTROL

Eladio Flores Martínez

Universidad Tecnológica del Sureste de Veracruz
layo06@live.com.mx

Pablo Reyna Guerra

Universidad Tecnológica del Sureste de Veracruz
pabloreynag@hotmail.com

José Luis Jiménez Reyes

Universidad Tecnológica del Sureste de Veracruz
jluis@201085@hotmail.com

Javier Garrido Meléndez

Universidad Veracruzana0
vgarrido@uv.mx

Quetzalcoatl Cruz Hernández Escobedo

Universidad Veracruzana
qhernandez@uv.mx

Resumen

En base a las encuestas que realiza la Universidad con los alumnos egresados se detectó que tienen problemas al conectar, calibrar y sintonizar equipos de tipo industrial, esto se debe a que en la Universidad existen prototipos didácticos para el control de variables de procesos, pero los alumnos cuando egresan y se emplean en alguna industria, los sensores, actuadores y controladores de los equipos didácticos no se parecen a los equipos industriales. En el presente trabajo se diseñó y construyó un prototipo de entrenamiento para prácticas en instrumentación y control de la variable de proceso Nivel, con la finalidad que los

alumnos desarrollen los conocimientos, habilidades y destrezas en el área de instrumentación y control. La variable podrá ser manipulada a través de controladores, PLC y tarjetas de adquisición de datos que permitirán la implementación de Interfaces para comunicarse con software Matlab y LabView.

Palabras Claves: Control, entrenamiento, instrumentación, monitoreo, variables.

Abstract

Based on the surveys carried out by the University with its graduates, it was detected that students have problems connecting, calibrating and tuning industrial equipment, this is because in the university there are didactic prototypes for the control of process variables, But the students when they graduate and are employed in some industry, the sensors, actuators and controllers of the didactic equipment do not look like the industrial equipment. In the present work designed and built a prototype of training for practices in instrumentation and control of the Process variables: Level, with the aim that the students develop the knowledge, skills and abilities in the area of instrumentation and control. The variables can be manipulated through controllers, PLCs and data acquisition cards that will allow the implementation of Interfaces to communicate with Matlab and Labview software.

Keywords: Control, instrumentation, monitoring, training, variables.

1. Introducción

La universidad Tecnológica del sureste de Veracruz se encuentra ubicada en una zona industrial, donde se localizan los complejos petroquímicos de la transformación de energía que emplean la instrumentación para el control de sus procesos, generalmente son procesos no regulatorios, debido a que para el llenado y vaciado de los tanques se utilizan bombas. Los sistemas no regulatorios son los que no tienen un cambio limitado en la salida para un cambio sostenido en la entrada [LeBlanc, 2008], aunado a esto el control que predomina en la industria petroquímica es el control clásico PID, el cual se ha comprobado no es totalmente conveniente ya que al ser un parámetro de control fijo, la no linealidad de los sistemas hace que ocurran cambios o perturbaciones, que no se pueden corregir

[Wang, 2011]. Por lo anterior es necesario que los estudiantes cuenten con los conocimientos, habilidades y destrezas en estas áreas. Esta es una de las razones por las que se decide diseñar y construir un prototipo de entrenamiento para la instrumentación y control de la variable de proceso nivel.

Existen trabajos con propósitos didácticos relacionados con la Supervisión, Control y Adquisición de Datos (por sus siglas en inglés, SCADA), Interfaces Hombre Maquina (por sus siglas en inglés, HMI), todos estos interconectados con el software de LabView [Adamo, 2007]. Aplicaciones de control de nivel y sistemas SCADA utilizando controladores avanzados sin módulos de lógica difusa, donde el controlador se programa con las instrucciones básicas del controlador lógico programable (por sus siglas en inglés PLC) [Aydogmus, 2009]. Sistemas con controladores Proporcional + Integral + Derivativo (PID), discretos y algunos aspectos prácticos para su sintonización [Lopez, 2007]. Además del análisis para obtener los valores de sintonización PID para minimizar los criterios de desempeño cuadrático de la integral del error [Zhuang, 1993], las investigaciones anteriores en algunos casos utilizan prototipos didácticos, la ventaja de tener equipo de tipo industrial, además de utilizarlo para realizar prácticas por parte de los alumnos, se podrá desarrollar trabajos de investigación por parte de los docentes. Sobre técnicas de control moderno que utilizan controles difusos para el control de nivel: de uno o más tanques[Meng, 2013].

Para el diseño del prototipo de entrenamiento se utilizaron en algunos casos materiales y equipos reciclados o donados por empresas de la zona por lo cual la metodología fue la siguiente:

- Diseño del DTI.
- Construcción del Prototipo.
- Instrumentación del prototipo.
- Prueba y puesta en marcha del Prototipo.

Este prototipo experimental permite a los estudiantes realizar prácticas de: Instalación, Cableado, Configuración, Calibración de los Instrumentos, implementación de interfaces en los software: LabView, Matlab y HMI [Ritter,

2002] para la variable de proceso nivel, además podrán realizar la sintonización de los controladores, interpretación de diagramas de instrumentación y tuberías (DTI) y finalmente la puesta en marcha del sistema. Siguiendo esta metodología posteriormente se implementará en este prototipo las variables Flujo, Temperatura y Presión.

2. Métodos

La metodología empleada para la realización del prototipo de entrenamiento para el control y monitoreo de las principales variables de proceso fue la siguiente: en una primera etapa se realizó el diagrama de tuberías e instrumentos (DTI_NIVEL) como se muestra en la figura 1, está compuesto por dos Tanques (T1 y T2) con capacidad de 300 lts, y los instrumentos utilizados para el control de nivel son: Controlador lógico programable (PLC) de la marca SIEMENS Modelo S7-1200, una computadora personal en la cual está instalado la interface Hombre Maquina (HMI), dos bombas de $\frac{1}{2}$ hp (B1 y B2), un transmisor de nivel (LT2) de 2 hilos con protocolo Hart modelo: 3051 marca Rosemount con alimentación de 24 vcd, una válvula manual (LV1) y una válvula de control (V2) marca Neles-Jamesbury modelo: VF613 tipo globo con posicionador electroneumático, 24 vcd entrada de 4-20 ma. La cual estará calibrada con offset para protección de las líneas y bombas.

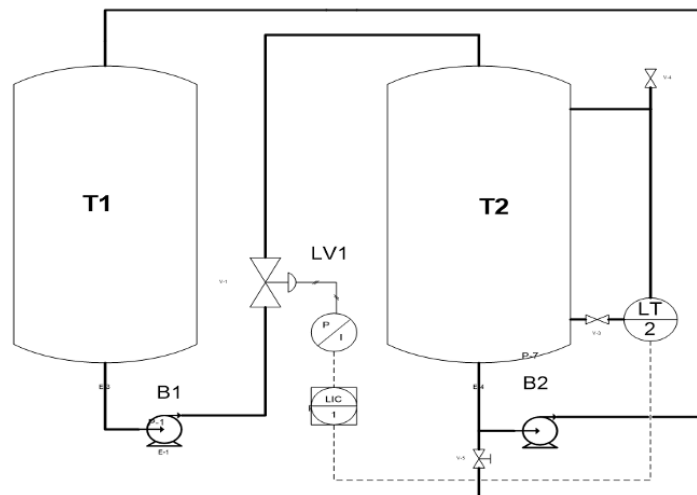


Figura 1 Diagrama de tuberías e instrumentos de la variable nivel.

El sistema cuenta con un lazo de control de la variable Nivel lo que permite controlar y monitorear la variable de proceso, además de esto se pretende que el sistema sea abierto debido a que se podrán obtener las señales de los transmisores de campo (4-20 mA) para que estas puedan ser manipuladas con diferentes tipos de controladores como: tarjetas de adquisición de datos, controladores de campo, plataformas abiertas como Arduino o Microcontroladores, estos últimos ejecutarán su algoritmo de control sobre los elementos finales como: Válvulas de control, Bombas, intercambiador de calor, para corregir la desviación de las variables.

Construcción del Prototipo

Para la construcción se utilizaron ángulos de 6" de acero al carbón, los cuales se soldaron para formar la base que tiene una medida de 1 m de ancho por 3 m de largo. El siguiente paso consistió en el montaje de los tanques de capacidad de 300 L de acero al carbón, dos bombas de agua de 1/2 de hp, así como también la instalación de la válvula de control neumática, el transmisor de presión diferencial, y la conexión de sus cámaras de alta y baja presión al proceso por medio de tubing de acero inoxidable de 3/8", y la conexión del PLC. En el caso de la válvula de control se elaboró un tubo bridado de 1" para acoplarla a la tubería del proceso donde circulará el líquido de un tanque a otro; estas tuberías son de material PVC de 1". En la figura 2 se muestra la construcción de la base e instalación de los tanques en el prototipo.



Figura 2 Base del prototipo de entrenamiento.

En la figura 3 se muestran los equipos e instrumentación instalados del prototipo de entrenamiento para el control y monitoreo de la variable Nivel.



Figura 3 Construcción final del Prototipo.

Siguiendo con la construcción se fabricó un gabinete de madera triplay de $\frac{1}{2}$ " de espesor, y sus medidas fueron 80x100x30 cm y soportado sobre la base del proceso. Como se muestra en la figura 4, dentro del gabinete se instaló los contactores eléctricos, relevadores, circuitos de protección para los equipos y finalmente un controlador lógico programable (PLC) SIEMENS S7-1200 para el control, monitoreo de la variable nivel y creación de las HMI en software como LabView.



Figura 4 Gabinete de control.

La programación del PLC dependerá de las condiciones iniciales de operación del proceso y de acuerdo a los tipos de entradas y salidas. Para el lazo de control de nivel se utilizaron: dos entradas digitales para el sistema de arranque y paro general, dos salidas digitales una para cada bomba, una entrada analógica que recibe señal del transmisor de presión diferencial montado en campo y finalmente una salida analógica para la señal a la válvula de control.

Para la conexión eléctrica de las bombas de agua, se utilizó cable calibre 12, para el manejo de voltaje de 110 Vac. Con respecto a los instrumentos del lazo de control se utilizó cable calibre 16 para el manejo de voltaje de 24 Vcd, el cual fue cableado hasta el tablero de control y conectado al PLC.

Pruebas y Puesta en Marcha del Prototipo

Teniendo todos los instrumentos instalados y conectados al panel de control se procede a la configuración y verificación de estos, para ello se utilizó el calibrador de procesos Fluke 744 con protocolo HART (por sus siglas en inglés de highway addressable remote transducer) compatible con diferentes Instrumentos. El primer instrumento a configurar es el transmisor de nivel (LT2). La figura 5 muestra la conexión al transmisor con el configurador para establecer su rango de medición (0 – 30 “ de H₂O) de acuerdo a la pierna de nivel [Solé, 2008].

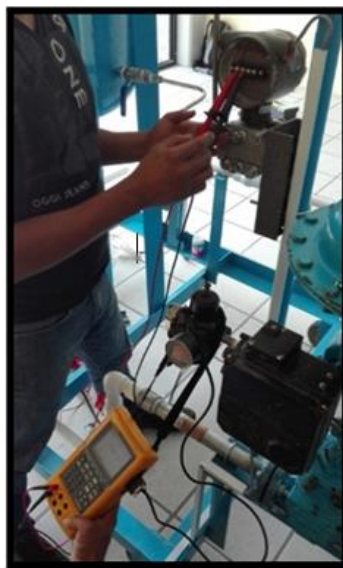


Figura 5 Calibración del trasmisor de presión diferencial.

El segundo instrumento calibrado fue la válvula neumática de control, como se observa en la figura 6, esto se realizó con el calibrador suministrándole una corriente en forma de rampa al posicionador electroneumático de la válvula.



Figura 6 Calibración de la válvula neumática de control.

En la tabla 1 se muestra los valores de la corriente suministrada al posicionador y el recorrido del vástago. En caso de no ser los valores correctos hacer el ajuste necesario.

Tabla 1 Calibración de Válvula Neumática de Control.

Señal de Entrada al Posicionador en mA	Señal de Salida del Posicionador en PSI	Recorrido del Vástago de la Válvula en %
4	3	0
8	6	25
12	9	50
16	12	75
20	15	100

Prueba de Funcionalidad de Bombas

Instaladas las bombas, con su conexión eléctrica y conectadas las tuberías del proceso, fue necesario probar su funcionalidad. Con el tanque 1 previamente llenado al 50%.de nivel, se energizo la bomba 1 para transferir el agua de este al

tanque 2 y de ahí se recircula agua del tanque 2 al tanque 1 por medio de la bomba 2, detectándose y corrigiendo al mismo tiempo posibles fugas en las líneas de tuberías de PVC y tubing del transmisor.

Prueba de Lazo

Finalmente habiendo probado la funcionalidad de los equipos y de los instrumentos, fue necesario realizar una prueba final al lazo de control la cual consistió en enviar una señal del calibrador en modo transmisor desde la posición del transmisor hasta el panel de control y del panel de control a la válvula de control. En la figura 7 se muestra la prueba del lazo de control del prototipo.



Figura 7 Prueba de lazo de control.

3. Resultados

Para verificar el correcto funcionamiento del sistema se implementó un controlador PI. Para la selección de los parámetros de las ganancias se procedió a calcular el modelo del sistema como un proceso de nivel de líquido con salida de flujo constante [Coughanowr, 2009] como se muestra en la figura 8.

El modelo matemático que representa la función de transferencia del proceso se representa con ecuación 1.

$$q_o(t) - q_o = A \frac{dh}{dt} \quad (1)$$

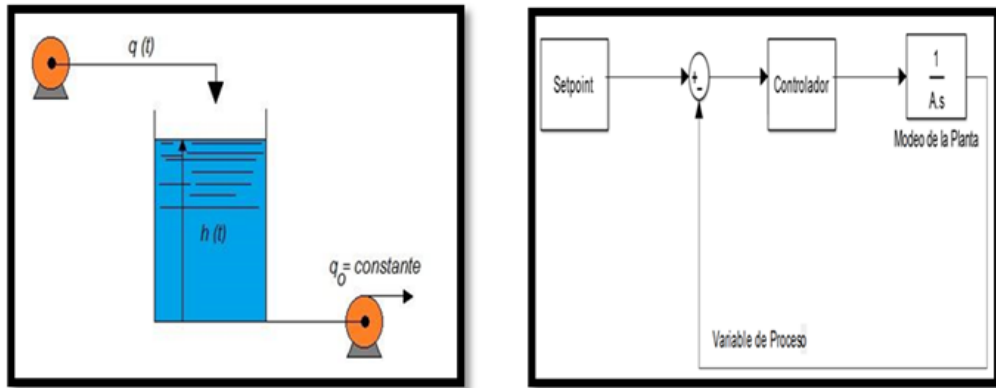


Figura 8 Sistema de nivel de flujo no regulatorio y sistema en Lazo cerrado.

Donde $q(t)$ = Razón de flujo de entrada y q_o = Razón de flujo de salida constante.

En el instante inicial, se tiene que la razón de flujo de entrada está dado por ecuación 2.

$$q_o = q_s \quad (2)$$

Sustituyendo ecuación 2 en ecuación 1, se tiene ecuación 3.

$$q_o(t) - q_s = A \frac{dh}{dt} \quad (3)$$

Introduciendo las variables de desviación $Q = q_o(t) - q_s$ y $H = h - h_s$, se obtiene ecuación 4.

$$Q = A \frac{dH}{dt} \quad (4)$$

Tomando la transformada de Laplace de cada lado de ecuación 4 y resolviendo para H/Q se obtiene ecuación 5.

$$\frac{H(s)}{Q(s)} = \frac{e^{-\tau s}}{As} = \frac{e^{-0.8s}}{0.2933 * s} \quad (5)$$

Donde $H(s)$ es el nivel del tanque, $Q(s)$ es el flujo de entrada, es el retardo de tiempo, A representa el área del tanque. En la figura 9 se muestra la respuesta en lazo abierto del sistema.

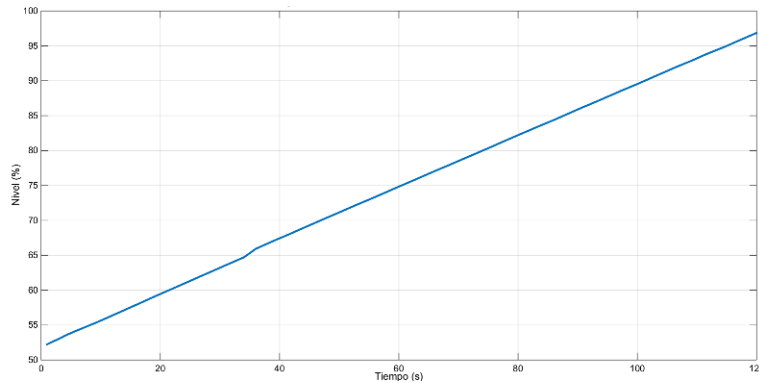


Figura 9 Respuesta del sistema en Lazo abierto.

Para comprobar analíticamente que el error en estado estacionario (e_{ss}) utilizando un controlador PI del sistema que se muestra en la figura 10, se calcula el sistema en lazo abierto $G(s)$, ecuaciones 6 y 7.

$$E(s) = \frac{1}{1 + G(s)} R(s) = \frac{As^2}{As^2 + e^{-ts}k_p(s+k_i)} \left(\frac{1}{s}\right) \quad (7)$$

Donde $E(s)$ es el error en estado estacionario, $R(s)$ es la entrada de referencia y utilizando el teorema del valor final para calcular el error en estado estacionario, ecuación 8.

$$\lim_{t \rightarrow \infty} e_{ss}(t) = \lim_{s \rightarrow 0} s E(s) = s \left(\frac{As}{As^2 + e^{-ts}k_p(s+k_i)} \right) = 0 \quad (8)$$

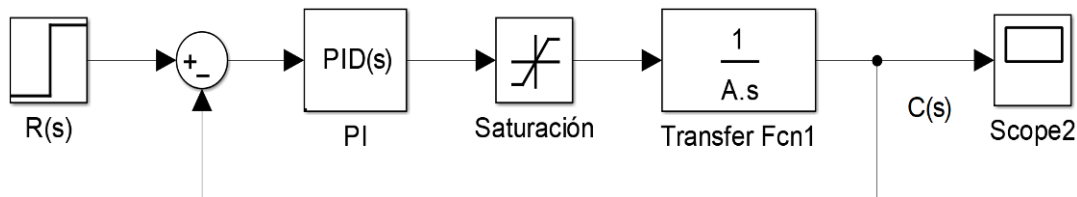


Figura 10 Sistema en lazo cerrado.

Se implementó un controlador PI, para sintonizar las ganancias, se usó el segundo método de Ziegler-Nichols [Ogata, 2003], con las siguientes ganancias:

$$K_{cr} = 10, \quad P_{cr} = 90 \quad K_p = 0.45K_{cr} = 4.5, \quad T_i = \frac{1}{0.2} P_{cr} = 75 \text{ seg.}, \quad K_i = \frac{K_p}{T} = 0.06$$

Donde K_{cr} es valor crítico proporcional y P_{cr} es oscilación sostenida con periodo en segundos, K_p ganancia proporcional, K_i es la constante integrativa y T_i tiempo de integración. Con estos valores se obtuvo la respuesta del sistema en lazo cerrado, figura 11.

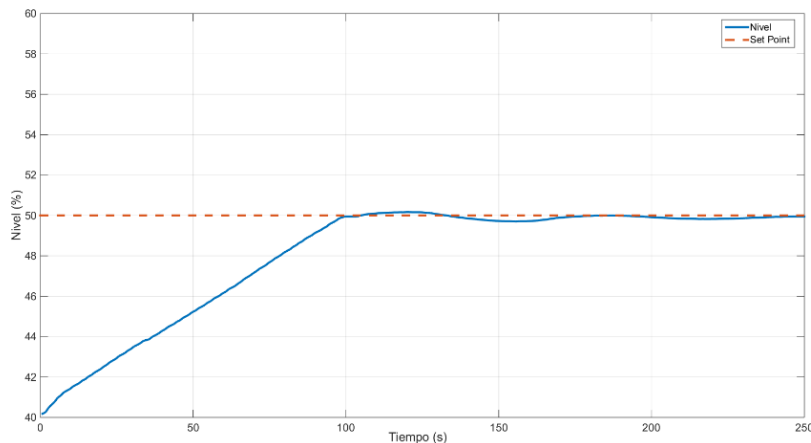


Figura 11 Respuesta del sistema con controlador PI.

La universidad tecnológica del sureste de Veracruz cuenta con un prototipo experimental que permitirá a los estudiantes realizar prácticas de las materias como: Instrumentación industrial, Integración de sistemas automáticos, Instrumentación virtual, Sistemas de control automático, Control lógico avanzado, Sistemas lineales para la automatización, Programación Visual. y Diseño de interfaces electrónicas, por mencionar algunas. Con el lazo de control de la variable Nivel instrumentada permitirá configurar y calibrar los instrumentos, así como sintonizar el lazo con sus modos de control: P, P+I, P+I+D.

Cabe mencionar que el prototipo fue diseñado con la intención de crecer agregando lazos de control de las variables Presión, Flujo y Temperatura. Estos lazos de control tendrán bornes de conexión rápida en el tablero para las señales de entrada y salida del proceso, y con ello realizar el control con diferentes tipos de controladores y medir la eficiencia de cada uno de ellos en los lazos de control. También podrán realizar interfaces electrónicas, programación de PLC, Interpretación de diagramas de tuberías e instrumentos (DTI) y la creación de interfaces Humano Maquinas (HMI).

4. Discusión

Actualmente en nuestra región hay pocas instituciones de educación superior que imparten la especialidad de instrumentación industrial y que cuenten con la infraestructura adecuada para este propósito. El prototipo tendrá la capacidad de realizar el control y monitoreo utilizando la instrumentación que se utiliza actualmente en la industria, como son: válvulas de control con posicionadores inteligentes, transmisores, etc. El modulo podrá comunicarse a través de tarjetas de adquisición de datos con software como Matlab y Labview, para el análisis y monitoreo de las variables de proceso y poder aplicar las leyes de control como: Controladores PID, Controladores con lógica Difusa, Controladores con redes neuronales, Controladores de modos deslizantes para la variable temperatura, Control Robusto, lo que significa que la plataforma de entrenamiento será un sistema abierto y modular.

5. Conclusiones

Con este prototipo industrial el alumno de la UTSV podrá egresar con los conocimientos teóricos y prácticos en el área de instrumentación industrial. Debido a que podrá realizar prácticas en las materias relacionadas como control automático, Integración de Sistemas Automáticos, Controladores Lógicos Programables, Control Lógico Avanzado, Instrumentación Virtual, Hidráulica y Neumática. Este prototipo al contar con instrumentación industrial también se podrá utilizar para capacitar a trabajadores de las empresas de nuestra región. Por otra parte debido a que en el tablero de control se adecuaran bornes de conexión rápida para la entrada de la variable de proceso, y salidas a válvulas, bombas etc., el módulo podrá comunicarse a través de tarjetas de adquisición de datos y PLC's, para crear interfaces con software como Matlab o Labview, para el análisis y monitoreo de las variables de proceso y poder aplicar las leyes de control como:

- Controladores PID
- Controladores con lógica Difusa
- Controladores con redes neuronales
- Controladores de modos deslizantes para la variable temperatura

- Control Robusto.

Cabe mencionar que el prototipo cuenta con una variable Nivel y que un futuro cercano siguiendo la metodología planteada, se agregarán lazos de control para las variables: flujo, presión y temperatura.

6. Bibliografía y Referencias

- [1] Adamo, F., Attivissimo, F., Cavone, G., & Giaquinto, N., SCADA/HMI Systems in Advanced Educational Courses. *IEEE Transactions on Instrumentation and Measurement*, pp. 4-10, 2007.
- [2] Aydogmus, Z., *Implementation of a fuzzy-based level control using SCADA. Expert Systems with Applications*, 2009.
- [3] LeBlanc, S. and Coughanowr, D., *Process Systems Analysis and*. McGraw-Hill Higher Education, 2008.
- [4] Lopez, I., & Cerezo, Y., Some practical aspects about performance and tuning of a multirate discrete PID controller. Paper presented at the 2007 Mediterranean Conference on Control & Automation, 2007.
- [5] Meng, Q. ,Wang, Q. and Wei, H., "Design of fuzzy controller for liquid level control system based on MATLAB/RTW," in *Proceedings of 2013 2nd International Conference on Measurement, Information and Control*, 2013, pp. 1090-1094, 2013.
- [6] Ogata, K., *Ingeniería de control moderna*: Pearson Educación, 2003.
- [7] Ogata, K., *Ingeniería de control moderna*: Pearson Educación, 2003.
- [8] Ritter, D. J., *LabVIEW GUI: Essential Techniques*: McGraw-Hill, 2002.
- [9] Wang, Z. and Wang Q., "Application of Fuzzy Controller in Drum water level Control," *IEEE International Conference on Mechatronic Science, Electric Engineering and Computer*,2011.
- [10] Solé, A. C., *Instrumentos industriales: su ajuste y calibración*, Marcombo, 2008.
- [11] Zhuang, M., & Atherton, D. P., Automatic tuning of optimum PID controllers. *IEE Proceedings D - Control Theory and Applications*, pp. 216-224, 1993.

SISTEMAS PARA LA EXTRACCIÓN DE FRASES CLAVE EN DOCUMENTOS CIENTÍFICOS

Gerardo Flores Petlascalco

Benemérita Universidad Autónoma de Puebla

gerardo.florespe@alumno.buap.mx

Mireya Tovar Vidal

Benemérita Universidad Autónoma de Puebla

mtovar@cs.buap.mx

Hilda Castillo Zacatelco

Benemérita Universidad Autónoma de Puebla

hilda@cs.buap.mx

José A. Reyes-Ortiz

Universidad Autónoma Metropolitana

jaro@correo.azc.uam.mx

Resumen

En este documento se describen dos sistemas para la extracción de frases clave en textos científicos. El primer sistema usa la generación de *n-gramas* y posteriormente se realiza la discriminación de términos candidatos usando reglas empíricas. El segundo sistema se basa en la construcción de patrones para la eliminación de frases candidatas. Además, se hace una comparación de estos sistemas con otros propuestos que realizan la misma tarea y se muestran los resultados obtenidos en la evaluación.

Palabras Claves: Frases clave, *n-gramas*, patrones.

Abstract

*In this document, we describe two systems for keyphrase extraction on scientific texts. The first system use *n-gram* generation and candidate term discrimination*

using empirical rules. The second system is based in the patterns construction for candidate phrases elimination. Further, we do a systems comparison with other approaches that perform the same task and we show the evaluation results.

Keywords: *keyphrases, n-grams, patterns.*

1. Introducción

Las palabras clave tienen como funcionalidad capturar la información importante del contenido de un texto con el objetivo de ayudar a los lectores al momento de estudiar o resumir dando una idea general del tema que aborda [Siqueira, 2015]. Por otro lado, en los sistemas computacionales, su funcionalidad radica en los sistemas de indexación. Estos recurren muchas veces a las palabras para agruparlas en tópicos y hacer sencilla una recuperación de información en caso de alguna consulta.

La selección de palabras clave no es una tarea sencilla, se tienen que evitar términos muy generales puesto que se corre el riesgo de que sean intrascendentes o demasiado objetivos, que provoquen que los lectores no los encuentren por desconocimiento de ellos. Una adecuada combinación de ambas características y técnicas de selección como pueden ser la frecuencia de ciertos términos en el documento o la identificación de conceptos fundamentales que describan nombres, acciones o características del trabajo nos darán métodos para identificar las palabras clave adecuadas que contengan lo que el autor quiere explicar en su trabajo. Este proceso se hace de forma manual por parte del creador del documento o expertos, sin embargo, es una tarea compleja y pesada que requiere mucho tiempo, además de que no está exenta de fallos. Por esta razón es que áreas del Procesamiento de Lenguaje Natural ven este problema como una oportunidad para implementar modelos que extraigan de forma automática palabras clave que ayuden a realizar la selección correspondiente.

SemEval es una serie de evaluaciones para sistemas computacionales de análisis semántico, desde el 2012 publican problemas donde se involucra el área de Procesamiento de Lenguaje Natural (PLN) y todos sus campos de estudio. En el año 2017 se propusieron un total de 12 tareas, cada una con sus respectivos

objetivos, evaluaciones y recursos que implican áreas de estudio específicas. Para objeto de estudio en este trabajo se dará solución a la tarea número diez que tiene como objetivo crear un Sistema de extracción de palabras clave y relaciones semánticas aplicado a textos científicos [Augenstein, 2017]. La tarea consta de tres subtareas:

- Extracción de palabras clave en textos.
- Clasificación de palabras clave identificadas.
- Extracción de relaciones de sinonimia o hiperonimia en las frases clave identificadas.

En esta investigación se aborda la subtarea 1, que tiene como objetivo la creación de un sistema que realice de forma automática una extracción de frases clave en publicaciones científicas, los sistemas propuestos deben recibir como entrada un extracto del texto y devolver las frases clave del mismo.

Con el objetivo de resolver esta subtarea, nuestro sistema es un enfoque basado en la extracción y discriminación de *n-gramas* sobre los textos procesados, la discriminación se realiza siguiendo reglas empíricas y posteriormente se adiciona una regla más al método para obtener mejores resultados.

Durante el desarrollo de esta investigación se realizó un estudio del estado del arte en materia de extracción de frases clave, entre los trabajos consultados se encuentran los siguientes:

- [Matsuo, 2010] presentaron una propuesta de algoritmo para extraer palabras clave pertenecientes a un corpus, usan un modelo probabilístico que evalúa la ocurrencia de los términos dentro de los documentos y después miden su rendimiento con base a una métrica llamada χ^2 -medida.
- [Park, 2010] usan un método basado en Naïve Bayes donde extraen las palabras clave en documentos científicos bien estructurados, consideran la posición de un candidato en el título, encabezado o cuerpo del texto para contar sus ocurrencias dentro de las distintas secciones y asignarle un puntaje, los más altos son seleccionados como palabras clave.

- [Thuy Dung, 2010] presentan un enfoque basado en la estructura lógica del texto para determinar si un candidato podría ser o no frase clave basado en su ocurrencia dentro de las secciones que lo conforman, esto limita mucho la cantidad de candidatos que son considerados y hace más sencilla la evaluación. Además, se puede combinar con otras técnicas de extracción para mejorar el rendimiento.
- [Stuart, 2010] describen a RAKE, un sistema para la extracción de palabras clave en textos individuales sin necesidad de que pertenezcan a un corpus. El enfoque usa la premisa de que las palabras clave raramente contienen puntuación y palabras vacías, bajo este enfoque se extraen candidatos que son divididos en términos individuales, luego a cada término se le asigna un puntaje basado en la frecuencia dentro del texto y a cada candidato se le calcula una evaluación usando la suma de sus términos. Se ordenan los resultados y los mejores puntajes son los elegidos.
- [Ouyang, 2010] proponen PolyU, un sistema para la obtención de frases usando la identificación y posterior expansión de palabras núcleo. Éstas serán conseguidas por medio de la frecuencia dentro del cuerpo siguiendo la premisa de que palabras muy repetidas serán consideradas como importantes, después de la identificación cada palabra será expandida por medio de una combinación de la frecuencia de su contexto y un patrón *PoS tagger*.
- [Ortiz, 2010] crean un sistema que combina dos técnicas para el descubrimiento de las palabras claves en textos científicos. Dichas técnicas son las secuencias de frecuencia máxima y el algoritmo de *PageRanking*. La secuencia de frecuencia máxima se realiza usando *n-gramas*, seleccionando aquellas que tengan alta ocurrencia para hacer un ordenamiento por *PageRanking* y determinar las palabras clave.

En la investigación proponemos un enfoque basado en el uso de *n-gramas* que se extraen desde el cuerpo del texto y forman nuestra colección de términos candidatos, usamos reglas de discriminación empíricas para disminuir la cantidad

y finalmente obtenemos los más relevantes por medio de patrones morfológicos obtenidos sobre los términos clave de un conjunto de datos de entrenamiento proporcionado por SemEval 2017 Tarea 10. El documento está organizado de la siguiente manera, iniciamos en la sección 2 donde presentamos los sistemas para la extracción de frases clave, en la sección 3 se presentan los resultados obtenidos y finalmente una discusión del trabajo en la sección 4.

2. Métodos

En esta sección mostramos los sistemas propuestos para dar solución a la subtarea 1 de SemEval 2017 Tarea 10 que tiene como objetivo la extracción de frases clave en textos científicos. Se explican dos sistemas, el primero se basa en un enfoque de extracción de *n-gramas* y discriminación de términos usando reglas empíricas obtenidas por observación. El segundo, continúa con el enfoque del primero, pero se le añade una mejora al pre-procesado y una nueva regla de discriminación usando patrones obtenidos después de un etiquetado *PoS Tagger*, es decir su etiqueta gramatical, sobre un conjunto de entrenamiento.

Primer Sistema GMBUAP

El primer sistema, nombrado como GMBUAP, es un enfoque que inicia con una extracción de *n-gramas* de un texto, después se hace una discriminación usando reglas empíricas conseguidas por observación sobre el conjunto de datos de entrenamiento proporcionado por SemEval 2017 Tarea 10 [Augenstein, 2017].

Este sistema consta de las siguientes fases, que se ilustran en el algoritmo de la figura 1:

- Pre-procesamiento de los documentos. Al texto de los documentos se eliminaron caracteres extraños como corchetes o no imprimibles. Dejando únicamente los símbolos: paréntesis, comas, puntos, guion medio, guion bajo, comillas, punto y coma, diagonal, llaves. Posteriormente, el documento fue dividido en oraciones.
- Formación de *n-gramas*. Cada oración del texto fue dividida en términos y se encontraron todas las posibles combinaciones de *n* palabras. En esta

aproximación decidimos formar gramas que van de 1 hasta una longitud máxima de 5 (*1-grama*, ... ,*5-gramas*).

```
Entrada: Texto_cientifico
Salida: Conjunto de palabras_clave
Inicio
  Gramas <- []
  TextoPreprocesado <- Preprocesamiento(Texto_cientifico)
  Oraciones <- PartirEnOraciones(TextoPreprocesado)
  Para cada oracion en Oraciones hacer:
    GramasDeLaOracion <- CrearGramas(oracion)
    Gramas <- Gramas + GramasDeLaOracion
  Finpara
  //Parte de la discriminación de candidatos
  NuevosCandidatos <- Regla_A(Gramas)
  NuevosCandidatos <- Regla_B(NuevosCandidatos)
  PalabrasClaveDelTexto <- Regla_C(NuevosCandidatos)
  Return PalabrasClaveDelTexto
Fin
```

Figura 1 Algoritmo del Primer sistema (GMBUAP).

- Discriminación de candidatos. Usando la observación sobre los datos de la tarea, fueron propuestas tres reglas empíricas para la reducción de candidatos:
 - a. Eliminar candidatos con palabras vacías (*stopwords*) al inicio y al final. Esta regla se creó bajo la premisa de que las frases clave identificadas en las anotaciones no contenían estas palabras al inicio. Por lo tanto, se decidió eliminar las que estaban al final puesto que dan por entendido que la idea sigue, pero se vio truncada en el momento de generar los gramas (ver resultado de regla A en la tabla 1). Además, en este paso eliminamos aquellas frases candidatas de longitud un carácter.
 - b. Eliminar candidatos que no formen parte del texto. El pre-procesado ocasiona que algunos caracteres considerados como no imprimibles fueran eliminados y con esas deficiencias se formaron los *n-gramas*. Al tener este problema las frases clave candidatas formadas pueden no estar completas en el texto al hacer el mapeo y por consiguiente son eliminadas (ver resultado regla B en la tabla 1).
 - c. Eliminar candidatos que no tengan ambos paréntesis. Las palabras clave pueden tener paréntesis dentro de ellas pero es necesario que posean ambos, el paréntesis de inicio y el paréntesis de cierre (ver resultado regla C en la tabla 1).

Tabla 1 Ejemplo del funcionamiento de las reglas.

Texto	Frases candidatas	Regla	Resultado
...such as X-ray absorption spectroscopy (XAS) and X-ray emission spectroscopy (XES) at...	and X-ray emission spectroscopy	A	Rechazada
...such as X-ray absorption spectroscopy (XAS) and X-ray emission spectroscopy (XES) at...	X-ray emission spectroscopy (XES)	B	Aceptada
<i>Frases clave:</i> X-ray emission spectroscopy (XES)	emission spectroscopy (XES)	C	Rechazada

En la tabla 1 se muestra un ejemplo de la aplicación de las reglas mencionadas anteriormente, en la columna uno se muestra un extracto de un documento, en la columna dos un ejemplo de frase candidata, es decir, un *5-grama*, en la columna 3 la regla aplicada y el resultado de esta en la columna 4.

Segundo Sistema

El segundo sistema continua con el enfoque de creación de candidatos con base a *n-gramas* y siguiendo las técnicas de discriminación antes descritas. Sin embargo, en esta aproximación se hacen operaciones adicionales al pre-procesado y se añade una regla de discriminación que utiliza patrones.

La finalidad de añadir otras operaciones al pre-procesado responde a una deficiencia localizada en el paso de la división de un texto en oraciones debido a abreviaciones encontradas en el cuerpo. Estas abreviaciones interfieren al momento de realizar el corte del texto puesto que nuestro criterio es hacerlo al encontrar el carácter punto '.'. Para evitar estos conflictos reemplazamos las abreviaciones con el carácter punto en su cuerpo de tal forma que el proceso pudiera ser revertido y no comprometieran la integridad de los candidatos. En la tabla 2, se muestran las abreviaciones con su respectivo reemplazo que fueron aplicadas a los textos.

Con el fin de encontrar un complemento a las tres primeras reglas, se recurrió al etiquetado gramatical o *Part-of-Speech Tagging (PoS)* en inglés [Toutanova, 2003]. Su objetivo es agregar una etiqueta a cada término de la frase clave. El etiquetado gramatical, en base a su categoría léxica, brinda información sobre una

palabra y su contexto [Rodríguez, 2013]. En este caso, lo usamos para formar patrones.

Tabla 2 Abreviaciones localizadas y su reemplazo.

Abreviación	Reemplazo
e.g.	e-g
Fig.	Fig-
Eq.	Eq-
Ref.	Ref-
al.	al-

Para ello, las palabras clave encontradas en los archivos de anotaciones fueron etiquetadas mediante *PoS tagger*, después de realizar este paso se obtuvieron un total de 1420 patrones que fueron ordenados de acuerdo a la cantidad de repeticiones que tenían dejando solamente aquellos que tuvieran diez o más para motivos prácticos. Los patrones seleccionados se muestran en la tabla 3, junto con sus frecuencias. La finalidad de esta regla fue validar y discriminar candidatos respetando aquellos que sean lo más consistentes con una estructura de frase clave.

Por lo tanto, el funcionamiento de la segunda aproximación se describe a continuación y se ejemplifica en el algoritmo de la figura 2:

- Obtención de patrones de palabras clave para la validación. Se emplea un etiquetado *PoS Tagger* sobre las frases clave de los archivos de anotaciones del corpus de entrenamiento y se realiza la extracción de patrones considerando aquellos con frecuencia de aparición mayor a diez (ver tabla 3).
- Pre-Procesado de los documentos. Se realiza un nuevo pre-procesado sobre el texto, en esta ocasión se usó la nueva técnica de reemplazo en abreviaciones conflictivas (ver tabla 2).
- Formación de *1-gramas*, ... ,*5-gramas*.
- Discriminación de candidatos. Las tres primeras reglas se describen en el algoritmo de la figura 1.
 - a. Eliminar candidatos con palabras vacías al inicio y al final.

- b. Eliminar candidatos que no formen parte del texto.
- c. Eliminar candidatos que no tengan ambos paréntesis.

Tabla 3 Patrones para el segundo sistema y sus frecuencias.

Patrón	Frecuencia	Patrón	Frecuencia
NN	990	NNP NNP NNP	17
NN NN	492	NN IN DT NN	17
JJ NN	390	JJ NN NN NNS	16
NN NNS	265	RB NN	15
JJ NNS	252	NNP NN NNS	15
NNS	231	NNP	15
JJ NN NN	191	NN VBG	15
JJ NN NNS	135	NN JJ NN	15
NNP NN	129	NNP NNP	14
NN NN NN	89	NN JJ NNS	14
NNP NNS	82	JJ NNP NNS	13
VBG	51	JJ CC JJ NNS	13
NN NN NNS	51	DT NN	13
JJ JJ NN	50	VBN JJ NN	12
JJ	43	VBG NN NNS	12
VBG NN	39	VBG DT JJ NN	12
JJ JJ NNS	31	NN IN NNS	12
VBN NN	30	NN CC NN NNS	12
VBN	24	JJ JJ NN NN	12
VBG NNS	23	CD NNS	12
JJ NN NN NN	23	CD NN	12
NNP NN NN	21	NNP JJ NN	11
NN IN NN	21	NN IN NN NNS	11
VBN NNS	20	VB DT JJ NN	11
VBG NN NN	18	NNP NNP NN	10
JJ NNP	18	NN VBG NN	10
NNS NN	17	NN IN NN NN	10

```

Entrada: Texto_científico
Salida: Conjunto de palabras_clave
Inicio
Gramas <- []
Patrones <- CargaPatrones()
TextoPreprocesado <- NuevoPreprocesamiento(Texto_científico)
Oraciones <- PartirEnOraciones(TextoPreprocesado)
Para cada oracion en Oraciones hacer:
    GramasDeLaOracion <- CrearGramas(oracion)
    Gramas <- Gramas + GramasDeLaOracion
Finpara
//Parte de la discriminación de candidatos
NuevosCandidatos <- Regla_A(Gramas)
NuevosCandidatos <- Regla_B(NuevosCandidatos)
NuevosCandidatos <- Regla_C(NuevosCandidatos)
PalabrasClave <- ValidacionPorPatrones(NuevosCandidatos,Patrones)
Return PalabrasClave
Fin
    
```

Figura 2 Algoritmo del Segundo sistema.

- Validación de candidatos usando patrones. Los candidatos se recibieron, etiquetaron y posteriormente mapearon con los patrones encontrados en la tabla 3 para validarlos mediante su estructura y definir si es una frase clave candidata.

3. Resultados

Los sistemas fueron evaluados usando las medidas clásicas de *Precisión* (ecuación 1), *Exhaustividad* (ecuación 2) y *Medida-F₁* (ecuación 3) que dan como resultado el rendimiento general del sistema [Tolosa, 2008]. Ambos se calificaron usando un script proporcionado por los organizadores de la Tarea 10 de SemEval 2017, junto con un *gold* estándar [Augenstein, 2017].

$$Precisión(S) = \frac{Cantidad\ de\ términos\ relevantes\ recuperados}{Cantidad\ de\ términos\ recuperados} \quad (1)$$

$$Exhaustividad(S) = \frac{Cantidad\ de\ términos\ relevantes\ recuperados}{Cantidad\ de\ términos\ relevantes} \quad (2)$$

$$Medida - F_1(S) = \frac{2}{\frac{1}{Precisión(S)} + \frac{1}{Exhaustividad(S)}} \quad (3)$$

Conjunto de Datos

Los datos proporcionados por los organizadores de SemEval 2017 tarea 10, constan de 500 artículos científicos del área de Ciencias de la Computación, Ciencias de Materiales y Física. Estos fueron divididos en tres grupos, 350 como conjunto de entrenamiento, 50 como conjunto de desarrollo y 100 para realizar las pruebas de nuestros sistemas. El total de frases clave del *gold* estándar es de 2051.

Cada uno de los 500 artículos fue extraído de la página de *ScienceDirect* y consta de tres partes, un archivo XML con todo el texto, TXT con un párrafo del texto y un archivo de anotaciones, el archivo contiene las palabras clave con un identificador, su clasificación, la posición dentro del texto y la frase clave. Todo el conjunto de datos se encuentra en el idioma inglés.

Evaluación

El primer sistema (GMBUAP) es un algoritmo reportado en la Tarea 10 de SemEval [Augenstein, 2017], el número de frases claves candidatas se presentan en la tabla 4 y su disminución a nivel de renglones al aplicar cada regla del algoritmo, ver columna 2 de la tabla 4. El número de frases claves candidatas del segundo sistema se presenta en la tercera columna de la tabla 4. Como puede observarse la aplicación de la regla correspondiente a patrones del segundo sistema consiguió un decremento notable de términos candidatos para la evaluación.

Tabla 4 Decremento de frases clave candidatas en el primer y segundo sistema.

Filtro	Frases clave candidatas	
	Primer sistema	Segundo sistema
Original	71804	77339
Primera regla	27994	27254
Segunda regla	21553	24611
Tercera regla	20871	22398
Patrones	----	11909

En la tabla 5 se muestran los resultados experimentales de los dos sistemas comparándolos con los resultados de otros equipos que participaron en SemEval 2017 Tarea 10. Se observa que el primer sistema GMBUAP obtiene bajos resultados, mientras que el segundo sistema logra superar los resultados del primero, consiguiendo un 0.22 de *Medida-F₁*.

Tabla 5 Evaluación general de las aproximaciones.

Equipo	<i>Medida-F₁</i>
SciX	0.42
IHS-RD-BELARUS	0.41
Know-Center	0.39
LIPN	0.38
SZTE-NLP	0.35
Segundo sistema	0.22
GMBUAP	0.08

La tabla 6 presenta los resultados experimentales de los dos sistemas. Se observa un incremento en los resultados de *Precisión*, *Exhaustividad* y *Medida-F₁* del

Segundo sistema con respecto a GMBUAP. Sin embargo, la gran cantidad de candidatos generados por el procedimiento de *n-gramas* comprometieron la precisión y con ello la calificación general. El siguiente objetivo de nuestra investigación consiste en disminuir la cantidad de frases candidatas con la intención de mejorar los resultados.

Tabla 6 Resultados de precisión, recuerdo y medida- F_1 de ambos sistemas.

	Precisión	Exhaustividad	Medida-F_1
Segunda aproximación	0.18	0.63	0.22
Primera aproximación	0.04	0.53	0.08

4. Discusión

Se presentan dos sistemas para la extracción de frases clave en textos científicos, el primer sistema utiliza *n-gramas* y realiza discriminación usando reglas empíricas. Posteriormente, presentamos una ligera mejora como segundo sistema que utiliza patrones obtenidos a través de un etiquetado *PoS tagger*.

El segundo sistema logra un 0.22 de *Medida- F_1* superando al primer sistema que consiguió un 0.08 de *Medida- F_1* . Asimismo, la *Exhaustividad* y la *Precisión* del segundo sistema también aumento de manera evidente con respecto al primer sistema, lo cual indica que el segundo sistema recupera más de la mitad de las frases clave en el conjunto de prueba y disminuye la cantidad de términos que no son consideradas frases clave. Partimos de la hipótesis de que las frases clave mantiene una estructura morfológica en el texto, es decir, están formadas normalmente por adjetivos y/o sustantivos, lo cual permitió disminuir la cantidad de frases clave candidatas y validar la hipótesis.

La puntuación en *Medida- F_1* de ambos sistemas es competitivo con respecto a los otros equipos participantes en la tarea de SemEval, según lo reportado por [Augenstein, 2017] en el documento de descripción de la tarea. En ambos sistemas solo se utilizaron los datos proporcionados por los organizadores, sin embargo, los otros equipos utilizaron recursos externos y enfoques supervisados que les permitió mejorar los resultados de sus propuestas.

5. Conclusiones

En esta investigación se presentan dos sistemas para la extracción de frases clave en textos científicos, el primer sistema usa una extracción de *n-gramas* para la extracción de frases clave candidatas y después se discriminan usando reglas empíricas que se obtuvieron por observación sobre el conjunto de datos de entrenamiento de la tarea. Posteriormente, se realizó una mejora al primer sistema y a las reglas empíricas a través del uso de patrones obtenidos después de un etiquetado *Part-of-Speech (PoS)* sobre frases clave identificadas del conjunto de entrenamiento y usando aquellos patrones con una frecuencia de aparición mayor a 10.

Los resultados obtenidos en la evaluación de los sistemas en *Medida-F₁* son de 0.08 para el primer sistema y 0.22 para el segundo sistema. Lo anterior, muestra que el enfoque propuesto basado en la hipótesis de que una frase clave tiene una estructura morfológica obtiene resultados competitivos comparado a otros enfoques que realizan su detección de frases clave usando aprendizaje supervisado y recursos externos para mejorar sus resultados.

Para futuras investigaciones, se pretende disminuir el número de términos extraídos y mejorar la puntuación de nuestros sistemas. Por lo cual, se están estudiando otras formas de extracción que puedan usarse de forma independiente o combinarlas con lo propuesto, entre ellas está la intención de aplicar similitud semántica, recurrir a técnicas más tradicionales como es el pesado de términos, es decir, *TF-IDF* o incluso modelos de aprendizaje supervisado, por ejemplo, Naïve Bayes, Máquinas de Soporte Vectorial y Árboles de decisión.

6. Bibliografía y Referencias

- [1] Augenstein, I., Riedel, S., Vikraman, L., McCallum, A., & Das, M., SemEval-2017 task 10: Extracting keyphrases and relations from scientific publications. The 11th International Workshop on Semantic Evaluation (SemEval-2017). Vancouver, Canada: Association for Computational Linguistics, 2017.

- [2] Matsuo, Y., & Ishizuka, Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information. *FLAIRS*, pp. 392-396, 2003.
- [3] Ortiz, R., David, P., Tovar, M., & Jiménez-Salazar, H., BUAP: An Unsupervised Approach to Automatic Keyphrase Extraction from Scientific Articles. *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 174-177, 2003.
- [4] Ouyang, Y., Li, W., & Zhang, R., 273. Task 5. Keyphrase Extraction Based on Core Word Identification and Word Expansion. *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 142–145, 2010.
- [5] Park, J., Gun Lee, J., & Daille, B., UNPMC: Naïve Approach to Extract Keyphrases from Scientific Articles. *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 178–181, 2010.
- [6] Rodriguez , F. J., *Nuevas fuentes de información para entrenamiento de etiquetados gramaticales*. Buenos Aires: Universidad de Buenos Aires, 2013.
- [7] Siqueira, C., ¿Cómo encontrar las palabras clave en un texto?, 22 de Diciembre de 2005. Obtenido de Universia.net: <https://goo.gl/q1JgPy>
- [8] Stuart, R., Dave, E., Nick Cramer, & Wendy Cowley, Automatic keyword extraction from individual documents. *Text Mining: Applications and Theory*, pp. 1-20, 2010.
- [9] Thuy Dung, N., & Minh-Thang, L., WINGNUS: Keyphrase Extraction Utilizing Document Logical Structure. *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 166-169, 2010.
- [10] Tolosa, G. H., & Bordignon, F. R., *Introducción a la Recuperación de Información*. Buenos Aires : Tolosa y Bordignon, 2008.
- [11] Toutanova, K., Klein, D., Manning, C., & Singer, Y., Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of HLT-NAACL 2003*, pp. 252-259, 2003.

SIMULACIÓN Y CONTROL DEL PROCESO DE MACERACIÓN DE UNA CERVECERÍA ARTESANAL

Jesús Antonio Flores Tovar

Universidad Autónoma Metropolitana Azcapotzalco
ing.antonio.flores.tovar@gmail.com

Miguel Magos Rivera

Universidad Autónoma Metropolitana Azcapotzalco
mrm@correo.azc.uam.mx

José Antonio Lara Chávez

Universidad Autónoma Metropolitana Azcapotzalco
jalch@correo.azc.uam.mx

José Manuel Domínguez Martínez

Universidad Autónoma Metropolitana Azcapotzalco
jmdm-legolas@hotmail.com

Juan Alberto Godínez Viveros

Universidad Autónoma Metropolitana Azcapotzalco
beetoo819@hotmail.com

Resumen

Uno de los procesos involucrados en la fabricación de cerveza es la maceración. En éste, se mezcla el grano con agua caliente para obtener un líquido denominado mosto. La calidad y sabor de la cerveza depende directamente de la proporción de grano y agua empleados, así como de la temperatura de ésta última. En este artículo se describe la elaboración de un simulador del proceso de maceración de una cervecería artesanal. El objetivo de este sistema es el de facilitar la realización de pruebas a los algoritmos que se emplearán para controlar las variables del sistema real. A partir de las ecuaciones de balance de materia y energía del proceso, el simulador entrega a un Controlador Lógico Programable

valores de temperatura, flujo y nivel. El controlador a partir de la información recibida, y empleando algoritmos PID, genera las señales de control para las válvulas virtuales del simulador. El intercambio de datos entre el equipo de control y el simulador se realiza por medio de un servidor OPC. El sistema desarrollado ha permitido validar y ajustar los parámetros de control que se emplearán en el sistema real.

Palabras Claves: Cervecería Artesanal, Control de Procesos, Controlador Lógico Programable, Interface Hombre Máquina, Simulación de Procesos.

Abstract

One of the phases involved in brewing is the mashing process. In this, the grain is mixed with hot water to obtain a liquid called wort. The quality and taste of the beer depends directly on the proportion of grain and water used, as well as its temperature. In this paper a simulator of the mashing process of a craft brewery is presented. The objective of this system is to facilitate the testing of the algorithms that will be used to control the variables of the real system. From a mass and energy balance of the process, the simulator delivers to a Programmable Logic Controller (PLC) temperature, flow and level values. Based on the information received, and using PID algorithms, the PLC generates the control signals for the virtual valves of the simulator. Data exchange between the control unit and the simulator is via an OPC Server. The developed system has allowed to validate and adjust the control parameters that will be used in the real system

Keywords: *Craft Brewery, Human Machine Interface, Process Control, Process Simulation, Programmable Logic Controller.*

1. Introducción

La producción de cerveza se encuentra dentro del 5% de las actividades manufactureras más importantes en México. Actualmente el país ocupa el primer lugar mundial como exportador y el cuarto como productor de cerveza con 105 millones de hectolitros al año. Esta área productiva representa cerca del 30% de la producción total de la rama industrial de bebidas en el país [INEGI, 2016]. Al ser

México uno de los principales consumidores de cerveza, sexto a nivel mundial, la búsqueda de sabores distintos por parte de un sector del mercado ha ido en incremento. Es de esta forma que las pequeñas cervecerías artesanales, que proponen productos distintos a los ofrecen las grandes empresas, han presentado un ritmo sostenido promedio del 35% anual desde el año 2010 [Deloitte, 2017]. A finales del 2016, se contaban cerca de 400 cervecerías artesanales formalmente registradas, las cuales producían casi 105 mil hectolitros [Acermex, 2017].

El proceso de fabricación de la cerveza ya sea en grandes volúmenes o a nivel artesanal, consta de un conjunto de etapas. En estas, se extraen algunos componentes del grano original, el líquido resultante es fermentado y posteriormente acondicionado para ser envasado.

Una de las primeras etapas en la fabricación de la cerveza es la denominada maceración. En este proceso el grano triturado en la etapa anterior se mezcla con agua caliente en un tanque agitado. La figura 1 muestra el diagrama de flujo de las operaciones involucradas en la producción de cerveza, haciendo énfasis en el lugar que ocupa el proceso de maceración.

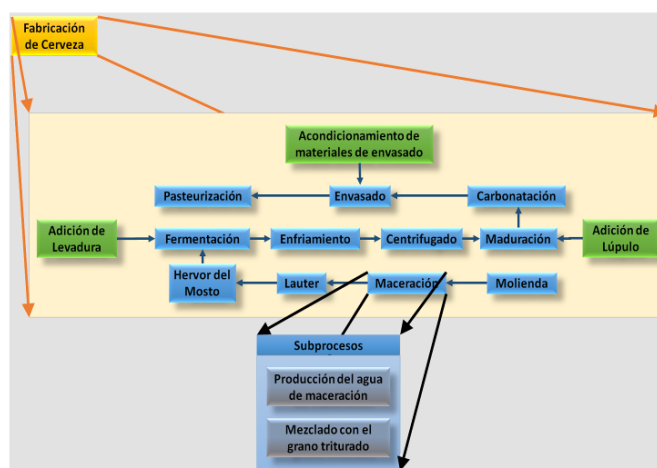


Figura 1 Diagrama de flujo del proceso de fabricación de la cerveza.

La operación de maceración permite extraer del grano determinados componentes, los cuales quedan contenidos en el líquido final denominado mosto. La calidad y sabor de la cerveza depende directamente de la proporción de grano y agua empleados, así como de la temperatura de ésta última. Debido a lo

anterior, el proceso de maceración debe llevarse a cabo bajo condiciones controladas de volumen, temperatura y tiempo [Kunze, 2006].

Al igual que muchas de las cervecerías artesanales del país, la empresa en la cual se realizó el proyecto que se presenta en este artículo inició sus operaciones con equipos poco automatizados. Lo anterior hace depender del factor humano muchos de los parámetros asociados a la producción de la bebida. Como parte de la modernización de sus equipos, la empresa inicio un proyecto para la automatización de la etapa de maceración. Se trata de controlar, de forma automática, la temperatura y el flujo del agua que entra al tanque donde se lleva a cabo el proceso. Además, se debe controlar tanto la temperatura de la mezcla agua-grano, como el nivel del tanque donde se lleva a cabo la maceración.

Un problema al cual se enfrenta el equipo que trabaja en este proyecto de automatización, es el hecho de tener que detener la línea de producción para poder realizar pruebas a los algoritmos y secuencias de control diseñados. También se debe considerar que las pruebas que se efectúan implican un cierto desperdicio de energía y materia prima. Además de existir el riesgo, tanto para los equipos como para las personas, por estar manipulando sistemas por los cuales circula agua a temperaturas cercanas a 80 °C.

Tomando en cuenta lo anterior, se propuso la elaboración de un simulador del proceso de maceración empleado en la empresa. A partir de condiciones de operación iniciales, y con base en ecuaciones de balance de materia y energía, un programa de cómputo determina los valores de las principales variables físicas involucradas en el proceso. El simulador desarrollado tiene la capacidad de enviar, mediante un servidor OPC (OLE for Process Control), la información obtenida a un Controlador Lógico Programable.

Para el proceso de producción del agua de maceración, el controlador determina el grado de apertura de las válvulas de agua caliente y agua fría. Lo anterior basado en los valores enviados por sensores virtuales. Además, en el proceso de mezclado del agua con el grano, a partir de la señal de un sensor de nivel también virtual, el controlador regula la cantidad de agua que se introduce al tanque, así como el momento en el cual inicia la alimentación del grano. Adicionalmente, se

cuenta con tres válvulas de vapor las cuales, por medio de intercambiadores de calor, permiten ajustar la temperatura de la mezcla al interior del tanque.

El paquete de cómputo empleado para realizar la simulación fue Intouch de la compañía Wonderware. Además de poder programarse las ecuaciones que describen el comportamiento del proceso, en este paquete es posible realizar el monitoreo del sistema mediante interfaces gráficos animados. La razón de emplearlo en esta aplicación es debido a que cuenta con la capacidad de comunicarse con controladores industriales, lo que permite simular únicamente el proceso. De esta forma, la acción de control puede ser realizada por los equipos de control que se emplearían en la realidad.

En la literatura existen diversos trabajos en los cuales se plantea el control y la automatización de alguna parte del proceso de fabricación de la cerveza. En [Buttrick, 2007], el autor hace una revisión de las principales aportaciones que la automatización aportó a la industria cervecera en las últimas décadas del siglo pasado y primera del presente. Mediante estudios de costo/beneficio, el autor presenta justificaciones para automatizar los procesos involucrados en la fabricación de la cerveza. Finalmente, una visión del futuro de la automatización en esta rama industrial también es presentada. Por su parte en [Birle *et al.*, 2013] los autores hacen una revisión de trabajos relacionados con la implementación de lógica difusa en el control de procesos de fabricación de bebidas y alimentos. Específicamente, en [Lujan *et al.*, 2010] se presenta el uso de lógica difusa para el control de las etapas de maceración y cocción en una planta artesanal de cerveza. En este trabajo, el control y monitoreo del proceso se realizó con ayuda de Labview. De igual manera, en [O'Connor *et al.*, 2002] se desarrolla un modelo basado en lógica difusa que considera la relación entre las diversas variables involucradas en el proceso de fermentación de la cerveza. El modelo obtenido es, finalmente, integrado al algoritmo de control del proceso. Por otro lado, en [De Andres *et al.*, 1997] los autores plantearon la optimización dinámica de procesos de fermentación de cerveza. El objetivo fue encontrar, mediante algoritmos genéticos, el mejor perfil de temperatura bajo el cual realizar la fermentación para acelerar el proceso.

La simulación de procesos ha sido de gran utilidad ya sea para estudiar o predecir el comportamiento de algunos procesos, así como para tareas de capacitación. Diversas herramientas de cómputo han sido empleadas para la elaboración de simuladores. En [Rodman *et al.*, 2016] se presenta la simulación en Matlab del proceso de fermentación de la cerveza. En este trabajo, los autores muestran el efecto que tiene la selección del perfil de temperatura a utilizar, al igual que el ajuste correcto de las condiciones iniciales, sobre la calidad de la cerveza. En [Andrés-Iglesias, *et al.*, 2015] el paquete de simulación Aspen es empleado para la simulación de un proceso de producción de cerveza sin alcohol. También ChemCAD ha sido empleado para la simulación de procesos de la industria de bebidas. En [Valderrama, *et al.*, 2012], los autores emplean este paquete de cómputo para simular tres distintos procesos de producción de bebidas alcohólicas y así demostrar la utilidad de la herramienta en esta área de la industria.

El paquete de cómputo empleado en el trabajo que aquí se expone, Intouch de Wonderware, es empleado en múltiples plantas industriales, siendo su principal aplicación el monitoreo y supervisión de procesos [Invensys Systems Inc., 2005]. En la literatura se encuentran algunos trabajos en los cuales se emplea esta herramienta. Una aplicación en la simulación de procesos, aunque no dentro de la industria alimenticia, es la que se presenta en [Wijaksono, *et al.*, 2015]. En este trabajo se simula el funcionamiento de un reactor nuclear del tipo agua a presión programando el modelo matemático en Intouch. Este paquete de cómputo ha sido empleado en la elaboración de interfaces de usuario. En [Xiaodong, *et al.*, 2015], se presenta una propuesta de control mediante PLC de un proceso de fermentación de cerveza. La interface de monitoreo en este trabajo se desarrolló con Intouch. Por último, podemos mencionar el trabajo que se presenta en [Wang, *et al.*, 2013], se emplea Intouch para la elaboración de la interface de monitoreo de un conjunto de máquinas CNC. Combinando este paquete de cómputo con Matlab y SQLServer, se desarrolló un sistema que alerta a los operadores cuando alguna de las máquinas fabrica piezas fuera de especificación, con lo que se logra elevar la eficiencia de la producción.

2. Métodos

Como ya se mencionó, el proceso de maceración puede ser dividido en dos grandes subprocesos: la producción del agua de maceración y el mezclado de ésta con el grano. Se puede considerar que el sistema desarrollado cuenta con dos simuladores, uno para cada subproceso. La parte central en cada uno de ellos es el modelo matemático que describe el comportamiento de cada etapa. Adicionalmente, cada bloque tiene una interface que permite establecer los parámetros de la simulación, así como pantallas para monitorear su operación. El servidor OPC y el Controlador Lógico Programable completan el sistema. En el diagrama de la figura 2, se muestra el diagrama de bloques del simulador completo, cada una de las partes que lo componen se describe en el resto del artículo.



Figura 2 Diagrama de bloques del sistema desarrollado.

Producción del agua de maceración

En la maceración se mezcla agua caliente con el grano con el fin de convertir los almidones en azúcares. Como ya se mencionó, la calidad y sabor de la cerveza depende directamente de la temperatura del agua empleada. En la planta, el agua de maceración se obtiene mezclando un flujo de agua caliente con otro de agua fría. En el proyecto de automatización, se plantea que un Controlador Lógico Programable lea el valor de flujo y temperatura de la mezcla de agua y a partir de esta información regule automáticamente la abertura de dos válvulas

proporcionales. El objetivo en esta parte del proyecto, es obtener un flujo de agua constante con un valor determinado de temperatura. Cabe mencionar que ni la temperatura ni el flujo del agua caliente y fría son constantes.

Modelo matemático

Las ecuaciones que describen el comportamiento de esta parte del proceso corresponden a un balance de materia y energía. El flujo de la mezcla de agua caliente y fría (flujo de maceración) (\dot{V}_{AM}) es igual a la suma de los flujos de entrada, multiplicados cada uno por la abertura de la válvula de agua fría y de agua caliente (A_f y A_c), ecuación 1.

$$\dot{V}_{AM} = \dot{V}_{Af} A_f + \dot{V}_{Ac} A_c \quad (1)$$

Por su parte, la temperatura del agua de maceración (T_{AM}), está dada por un promedio ponderado de la temperatura de ambos caudales, ecuación 2.

$$T_{AM} = \frac{\dot{V}_{Af} A_f T_{Af} + \dot{V}_{Ac} A_c T_{Ac}}{\dot{V}_{Af} A_f + \dot{V}_{Ac} A_c} \quad (2)$$

Las ecuaciones anteriores se programan en el paquete de computo Intouch. La sintaxis que maneja esta aplicación es muy similar a la del lenguaje C.

Interface de usuario

Se elaboraron dos las pantallas que permiten la interacción del usuario con la sección del simulador asociada al proceso de producción del agua de maceración. La primera de ellas permite configurar la simulación y monitorear su funcionamiento, mientras que la segunda permite visualizar un gráfico animado de esta parte del proceso:

- **Parámetros iniciales de simulación.** Esta ventana permite configurar los parámetros de la simulación, así como monitorear cuantitativamente el comportamiento de las variables de interés. Los parámetros se encuentran agrupados en seis grupos funcionales. Los primeros tres bloques permiten ajustar los parámetros de la simulación, mientras que los tres últimos permiten monitorear el comportamiento de esta.

En el primer grupo se especifican los valores de flujo y temperatura para las dos líneas de alimentación de agua. Para cada entrada de agua se cuenta con un campo en el cual se puede especificar el tiempo de respuesta de las válvulas. Este parámetro se refiere al tiempo que tardan las válvulas en pasar de totalmente cerradas a totalmente abiertas. Las válvulas que se proponen para la aplicación real son del tipo electroválvulas proporcionales, las cuales mediante un motor, desplazan el vástago de las mismas. Los tiempos de recorrido son del orden de 2 minutos, esta característica es contemplada por el simulador. El siguiente grupo corresponde a los parámetros de consigna de flujo y temperatura del agua de salida, es decir, los valores a los cuales se desean mantener estas dos variables. La propuesta de control del sistema se basa en dos algoritmos PID independientes, es en la tercera ventana de la interface que el usuario proporciona los parámetros de ambos controladores. Estos algoritmos están implementados en el Controlador Lógico Programable externo, el cual por medio del servidor OPC lee los valores correspondientes. El quinto bloque de variables despliega los valores de temperatura y flujo a la salida de las válvulas, así como el valor real de apertura de cada una de estas. Este valor se determina, como ya se mencionó, tomando en cuenta el retardo propio de los dispositivos. El último grupo despliega los valores finales de temperatura y flujo del agua de maceración.

Adicionalmente se cuenta con botones que permiten saltar a otras ventanas del simulador.

- **Monitoreo gráfico de la simulación.** En esta ventana se tiene un gráfico en el cual se observan las dos líneas de alimentación de agua. Parejas de indicadores de temperatura y flujo se ubican antes y después de cada una de las válvulas, así como en la tubería de salida.

Mezcla con grano

El siguiente bloque del simulador corresponde al tanque en el cual se realiza la mezcla del grano molido y el agua de maceración resultante de la etapa anterior.

Modelo matemático

En este bloque del simulador, se determina el nivel en el mezclador, así como la temperatura de la mezcla. El valor del nivel depende del volumen de agua (\dot{V}_A) y grano vertidos en el macerador (\dot{V}_G) por unidad de tiempo, así como de la geometría del tanque. En forma simplificada el volumen total (V_T) está dado por la ecuación 3.

$$V_T = (\dot{V}_A + \dot{V}_G) \cdot t \quad (3)$$

Por su parte, la temperatura de la mezcla depende de la cantidad y temperatura de los materiales introducidos al tanque, así como del estado de las tres válvulas que controlan el paso de vapor hacia igual número de intercambiadores de calor colocados alrededor del mezclador. En forma simplificada la temperatura del producto al interior del tanque está dada por la ecuación 4.

$$T_m = T_{A_{mat}} + T_{A_{val1}} + T_{A_{val12}} + T_{A_{val123}} \quad (4)$$

Como puede observarse, la ecuación de la temperatura de mezcla (T_m) está compuesta por cuatro términos. El primero depende únicamente del volumen y temperatura del agua de maceración y del grano ($T_{A_{mat}}$). Los tres términos restantes dependen de estado en que se encuentren cada una de las válvulas de vapor (temperaturas de aporte válvulas: $T_{A_{val1}}$, $T_{A_{val12}}$ y $T_{A_{val123}}$). Cabe mencionar que la segunda válvula de vapor solo puede accionarse si la primera se ha abierto. Una condición similar se tiene para el tercer dispositivo, solo se abre si los dos primeros están permitiendo el paso de vapor.

La decisión respecto a cuantas válvulas activar, la tiene el Controlador Lógico Programable. Este equipo, con base en el valor de la temperatura y nivel de la mezcla, decide cuantas válvulas abrir. Lo anterior hasta llevar la temperatura del líquido en el mezclador al valor correcto.

Interface de usuario

Monitoreo gráfico de la simulación. En esta ventana se tiene la imagen animada del tanque de mezclado. En este gráfico el estado de las válvulas de

vapor se refleja con un cambio en la orientación de las palancas correspondientes, así como con el color de los dispositivos. El nivel del líquido al interior del tanque se muestra en forma animada en esta interface. El motor que envía el grano al macerador, así como el agitador de este, cuentan con efectos que simulan su movimiento. Por último, se tienen indicadores numéricos que despliegan los valores del nivel y la temperatura de la mezcla.

Servidor OPC

Este bloque permite el intercambio de información entre el paquete de cómputo en el cual se realiza la simulación y el Controlador Lógico Programable. Se eligió un Servidor OPC para esta tarea. OPC (OLE for Process Control) es un estándar para el intercambio de datos entre dispositivos empleados en aplicaciones de automatización a nivel industrial. Es una herramienta abierta que la mayoría de fabricantes de equipos de automatización ha incluido en sus dispositivos. Para esta aplicación, se seleccionó el servidor OPC Top Server de la compañía Software Toolbox [Software Toolbox, 2013].

De esta forma, vía el Servidor OPC, el simulador en Intouch envía los datos de temperatura y flujo del agua de maceración, así como la temperatura de la mezcla en el macerador, al PLC. Por el mismo medio, el controlador transmite al simulador las señales de control para las dos válvulas proporcionales (agua caliente y agua fría), las tres válvulas de vapor y los dos motores (grano y agitador).

Controlador Lógico Programable

El PLC seleccionado para esta aplicación es el modelo S7-1200 CPU 214 de la marca Siemens [Siemens AG, 2009]. Se trata de un controlador con 14 entradas y 10 salidas digitales. A partir de la información recibida vía el Servidor OPC, el equipo controla la operación de las válvulas de agua y de vapor, así como los motores del grano y del agitador. En la figura 3, se muestra el diagrama de flujo del programa realizado para el controlador.

Al ejecutarse el programa, el controlador queda en espera de que en la interface se active la perilla de Inicio de Simulación. Una vez que esta condición se cumple,

el equipo lee los valores de temperatura y flujo del agua de maceración y ejecuta el algoritmo PID, a partir del resultado se ajustan las válvulas proporcionales. Posteriormente, si ya se tiene un nivel mínimo de agua en el tanque de maceración, se enciende el motor que lo alimenta de grano. Al mismo tiempo se activa el motor de agitación. Enseguida se verifica si la temperatura del líquido al interior del tanque es la adecuada, en caso contrario, se inicia el encendido de las válvulas de vapor para ajustar este parámetro. Finalmente, se verifica si se ha alcanzado el nivel final de producto en el tanque, en caso afirmativo, la simulación concluye.

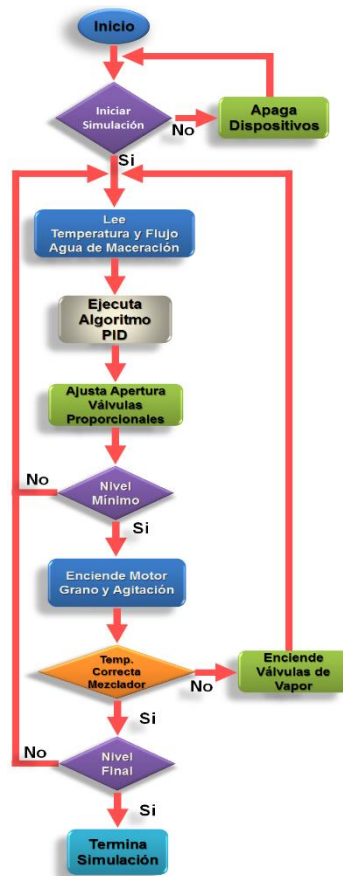


Figura 3 Diagrama de flujo del programa elaborado para el PLC.

3. Resultados

El resultado obtenido es un programa de cómputo que simula el comportamiento de las secciones de generación de agua de maceración y del

mezclado de esta con el grano molido. Mediante diversas ventanas el usuario puede configurar la simulación, monitorear cuantitativamente el proceso o bien, visualizar mediante animaciones el estado del mismo. También se incluyeron gráficas que permiten observar el comportamiento de las variables de interés en función del tiempo. La figura 4 muestra la ventana de bienvenida del simulador. En ésta se tienen 5 botones que permiten el acceso a las otras ventanas de la aplicación.

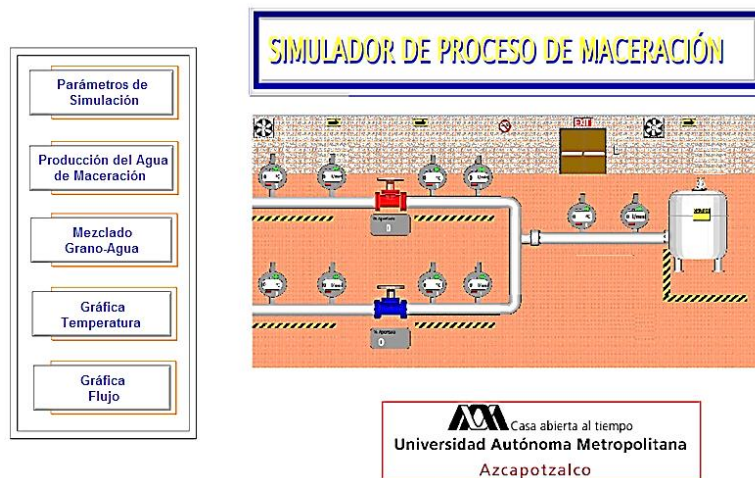


Figura 4 Ventana de inicio del simulador desarrollado.

La ventana de configuración de parámetros de la simulación se muestra en la figura 5. En el ejemplo se especifica una alimentación de agua caliente de 20 l/min a una temperatura de 85 °C. En cuanto al agua fría se estableció un caudal de 25 l/min a 14 °C. Las válvulas consideradas tienen un tiempo de respuesta de 120 s. Se desea tener un flujo de agua de maceración de 30 l/min a 75 °C. Los parámetros del algoritmo PID el control de temperatura del agua de maceración se establecieron en: $K_p=5$, $T_i=100$ y $T_d=0$. Mientras que para el flujo se propusieron en: $K_p=7$, $T_i=120$ y $T_d=0$. Al realizar la captura de pantalla del ejemplo, el algoritmo de control establecía aberturas del 100% y del 61.9% para las válvulas de agua caliente y fría respectivamente. Dado el retardo especificado para las válvulas, los valores reales de abertura al momento de la captura de pantalla eran: 90% y 30%. Con estas aberturas, los valores de la temperatura y del flujo a la salida de la

válvula de agua caliente eran: 85 °C y 18 l/min. Mientras que a la salida de la válvula de agua fría eran: 14 °C y 7.5 l/min. Los valores instantáneos de estas variables una vez mezclados los flujos eran: 64.1 °C y 25.5 l/min.

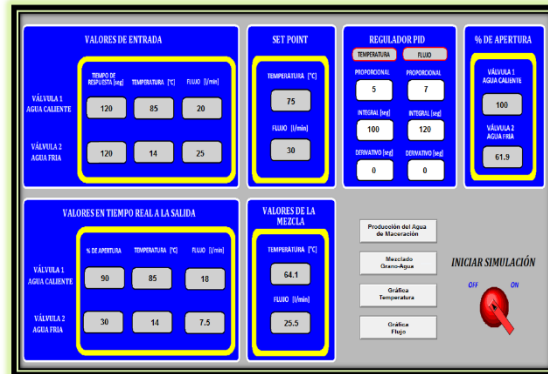


Figura 5 Ventana de configuración de parámetros de simulación.

Adicionalmente se tienen los botones que permiten cambiar de ventana a desplegar, así como la perilla que inicia la simulación con los parámetros establecidos.

En la imagen que se muestra en la figura 6 se observa la pantalla de monitoreo gráfico del proceso de producción del agua de maceración. Los indicadores señalan los valores instantáneos de las variables.

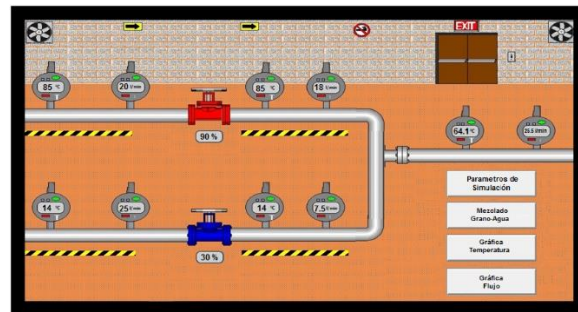


Figura 6 Monitoreo gráfico del proceso de producción del agua de maceración.

Finalmente, la figura 7 muestra la sección de la interface donde se realiza el monitoreo gráfico del mezclado del grano con el agua de maceración.

En este caso se tiene animación sobre algunos de los elementos que conforman la interface. El nivel en el gráfico aumenta de acuerdo con la variable simulada. Así

mismo, la posición de las manivelas de las válvulas de vapor cambia a la vez que el color del cuerpo del dispositivo pasa de rojo a verde dependiendo de su estado. En el ejemplo, se tienen abiertas las dos primeras válvulas de vapor. Por último, los motores muestran una vibración cuando están encendidos y se observa la caída del grano al interior del tanque.

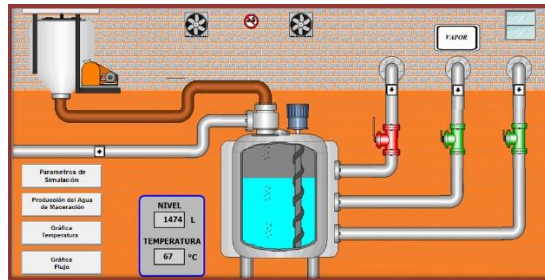


Figura 7 Monitoreo gráfico del proceso de mezclado grano-agua.

Las últimas ventanas de la interface corresponden a las gráficas de comportamiento del flujo y de la temperatura del agua de maceración contra tiempo. La figura 8 muestra la gráfica de flujo vs tiempo, para una simulación realizada. El trazo en verde corresponde al valor de consigna especificado para el flujo, mientras que en rojo se observa el comportamiento en el tiempo de la variable real. Podemos observar como después de iniciada la simulación el valor real alcanza al de consigna y se mantiene en ese punto. Cabe mencionar que, manualmente se produjo una perturbación en alguno de los flujos de alimentación para observar el comportamiento de los algoritmos de control elaborados para el PLC. El resultado se observa en la gráfica, después de algunos segundos, el valor real vuelve a ajustarse. Para la temperatura se tiene una ventana de graficación similar.

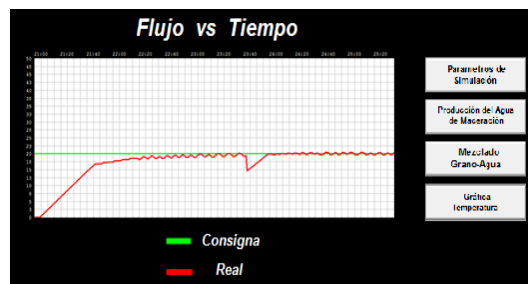


Figura 8 Gráfica del comportamiento en el tiempo del flujo del agua de maceración.

4. Discusión

Antes de desarrollar el proyecto que aquí se describe, las de pruebas y ajustes a los algoritmos de control que se plantean emplear para la automatización del sistema real, eran complicadas de realizar. No obstante que la empresa está interesada en alcanzar un cierto grado de automatización de sus procesos, el detener la producción, así como el desperdiciar materia prima y energía, era algo que no les satisfacía. Adicionalmente se tenía el riesgo hacia el personal y los equipos al trabajar con agua a alta temperatura.

La automatización de un equipo involucra 4 elementos básicos: el proceso, los sensores, los actuadores y el controlador. El sistema desarrollado considera la simulación de los tres primeros elementos.

Respecto a la simulación del proceso se puede comentar que las ecuaciones que describen el comportamiento de los dos subprocesos considerados son sencillas y se obtienen de un balance de materia y energía, el cual es un cálculo perfectamente documentado en la bibliografía. Lo anterior permite considerar que esta parte del simulador tiene un comportamiento bastante cercano al del equipo real. Respecto a los sensores, la simulación parte de un comportamiento ideal de los mismos, no se tiene en cuenta ruido electromagnético, caídas de voltaje o algún tipo de perturbación en la señal. Para la implementación real del control se considera el uso de sensores que emplean estándares de transmisión industrial, tales como: 4 a 20 mA o transmisión digital. Así mismo, estos dispositivos se encontrarán a no más de 10 metros de distancia del controlador. Lo anterior, aunque no elimina totalmente, si reduce bastante la probabilidad de perturbaciones en las señales, esto valida la simulación. Por último tenemos las válvulas proporcionales que regulan el flujo de agua caliente y fría para producir el agua de maceración. En este caso, uno de los parámetros que podría afectar el comportamiento del sistema real es el tiempo de respuesta, sobre todo considerando que el valor de esta variable es del orden de varias decenas de segundos. Como se describió en la sección correspondiente, el simulador realiza la abertura de las válvulas considerando este retardo. Lo anterior se ve reflejado

en las variables que se envían al controlador por lo que se tiene la posibilidad de que el algoritmo de control a implementar tome en cuenta ese retardo.

5. Conclusiones

Se considera que los objetivos planteados al inicio de la elaboración del simulador que se presenta en este artículo fueron alcanzados. Se logró la elaboración de un sistema compuesto por varias pantallas que sirven de interface para configurar y monitorear experimentos en forma sencilla. El uso del simulador descrito está facilitando las pruebas y ajustes de los algoritmos de control que se implementarán en un futuro cercano en el equipo real.

Al momento de redactar este trabajo, el simulador ha permitido hacer ajustes a los algoritmos originalmente propuestos. Las respuestas obtenidas son satisfactorias por lo que se tiene la certeza de que la implementación física de la automatización tendrá los resultados esperados.

Como trabajo a futuro se pretende iniciar en las próximas semanas las adecuaciones físicas al sistema real para instalar los equipos que permitirán automatizar el proceso. Así mismo, se está considerando incluir en la simulación realizada, la siguiente etapa de la fabricación de la cerveza. Lo anterior permitiría facilitar la automatización de ésta en caso de que la empresa así lo requiera.

6. Bibliografía y Referencias

- [1] Birle S., Hussein M.A., Becker T, Fuzzy logic control and soft sensing applications in food and beverage processes, *Food Control*, Vol. 29, No. 1, pp. 254-269, Enero 2013.
- [2] Acermex, Cerveceros de México, Copa Cerveza, Estado de la industria cervecera artesanal 2016-2017, México, 2017.
- [3] Acermex, Etherpower, SUN 2232, Manual de usuario, Buenos Aires, Argentina, 2015.
- [4] Andrés-Iglesias C., Garcia-Srna J., Montero O., Blanco C., Simulation and flavor compound analysis of dealcoholized beer via one-step vacuum distillation, *Food Research International*, Vol. 73, No. 3, pp. 751-760, 2015.

- [5] Buttrick P., Towards the lights-out brewery, A brewer's view of automation, *The brewer & distiller international*, Vol. 3, No. 4, pp. 1-5, August 2007.
- [6] De Andres Toro B., Girón-Sierra J.M., Lopez-Orozco J.A., Fernandez-Conde C., Optimization of a batch fermentation process by genetic algorithms, *IFAC Proceedings Volumes*, Vol. 30, No. 9, pp. 179-184, June 1997.
- [7] Deloitte, *La cerveza artesanal, Una experiencia multisensorial*, México, 2017.
- [8] INEGI, *La actividad de elaboración de cerveza*, México, 2016.
- [9] Invensys Systems Inc, *Wonderware Factory Suite, Intouch User's Guide*, Lake Forest, California, USA, 2005.
- [10] Kunze W, *Tecnología para Cerveceros y Malteros*, Ed. VLB Berlin, 2006.
- [11] Lujan M., Vásquez V., Control automático con lógica difusa de la producción de cerveza artesanal en las etapas de maceración y cocción, *Scientia Agropecuaria*, Vol. 1, pp. 125-137, 2010.
- [12] O'Connor B., Riverol C., Kelleher P., Bevan R., Hinchy E., D'Arcy J., Integration of fuzzy logic based control procedures in brewing, *Food Control*, Vol. 13, No. 1, pp. 23-31, January 2002.
- [13] Rodman A., Gerogiorgis D., Dynamic simulation and visualisation of fermentation: effect of process conditions on beer quality, *IFAC Papers on line*, Vol. 49, No. 7, pp. 615-620, 2016.
- [14] Siemens AG, *Controlador Programable S7-1200, Manual del Sistema*, Nuremberg, Alemania, 2009.
- [15] Software Toolbox Inc., *OPC Quick client user guide*, Matthews, North Caroline, USA, 2013.
- [16] Valderrama J., Toselli L., Faúndez C., Advances on modeling and simulation of alcoholic distillation, Part 2: Process simulation, *Food and Bioproducts Processing*, Vol. 90, No. 4, pp. 832-840, October 2012.
- [17] Wang J., Zhang Y., Monitoring system of machine tools based on the intouch, *Proceedings of the International Conference on Mechanical and Automation Engineering*, MAEE 2013, pp. 70-72, Jiujiang, China, 2013.

- [18] Wijaksono U., Abdullah A.G., Hakim D.L., Design of virtual SCADA simulation system for pressurized water reactor, Proceedings of international seminar on mathematics, science, and computer science education, MSCEIS 2015, Bandung, Indonesia, October 2015.
- [19] Xiaodong Z., Jie Z., Ke L., Design and implementation of control system for beer fermentation process based on SIMATIC PLC, Proceedings of the 27th Chinese Control and Decision Conference, CCDC 2015, pp. 5653-5656, Qingdao, China, 2015.

DISEÑO E IMPLEMENTACIÓN DE UN MULTÍMETRO DIGITAL CON FUNCIONES AMPLIADAS DE BAJO COSTO

Rafael García Arredondo

Tecnológico Nacional de México en Celaya

m1703043@itcelaya.edu.mx

Juan Carlos Gómez Cortez

Tecnológico Nacional de México en Celaya

m1703042@itcelaya.edu.mx

Diego de Jesús Padierna Arvizu

Tecnológico Nacional de México en Celaya

m1703018@itcelaya.edu.mx

José Eleazar Peralta López

Tecnológico Nacional de México en Celaya

m1703016@itcelaya.edu.mx

Francisco Javier Pérez Pinal

Tecnológico Nacional de México en Celaya

francisco.perez@itcelaya.edu.mx

Luis Antonio Ramírez Arredondo

Tecnológico Nacional de México en Celaya

m1703017@itcelaya.edu.mx

Julio Cesar Regalado Sánchez

Tecnológico Nacional de México en Celaya

m1703015@itcelaya.edu.mx

Resumen

Este documento describe el diseño e implementación de un multímetro digital de bajo costo para la medición de voltaje y corriente (directa y alterna), impedancia y temperatura. Para su implementación, se utilizó un divisor de voltaje, un sensor

de corriente ACS712, el medidor de impedancias AD5933, un LM35 para la medición de temperatura. Después del acondicionamiento respectivo de cada sensor, las señales se adquirieron por un microcontrolador PIC16F877A para finalmente realizar el despliegado de datos en una pantalla de cristal líquido (LCD, por sus siglas en inglés) de 16x2. El multímetro diseñado realiza mediciones de 0 a 500 V, de 0 a 2 A, impedancias desde 1 k Ω hasta 10 M Ω y temperaturas de 0 a 150 °C. El multímetro posee rangos aceptables de medición, con la capacidad de realizar tales mediciones sin la necesidad de hacer un cambio de escala. De igual manera, su costo es menor comparado con equipos comerciales con capacidades similares.

Palabras Claves: Corriente, impedancia, multímetro digital, temperatura, voltaje.

Abstract

This document describes the design and implementation of a low-cost digital multimeter for measuring temperature, impedance, voltage and current. For its implementation, an ACS 712 current sensor, the AD5933 impedance meter, an LM35 for temperature measurement and a voltage divider for voltage measurement were used. After the respective conditioning of each sensor, the signals were acquired by a PIC16F877A microcontroller to finely perform the data display on a 16x2 LCD (liquid crystal display). The designed multimeter supports 0 to 500 V measurements, 0 to 2 A, 0 to 150 °C and impedance measurements from 1 k Ω to 10 M Ω .

Keywords: Current, digital multimeter, impedance, temperature, voltage.

1. Introducción

El multímetro Digital

El multímetro digital (MMD) es un instrumento electrónico de medición que en su forma más básica calcula voltaje, resistencia y corriente. Los MMD a base de baterías y en producción en masa datan de principios de los años setenta [Green, 1974]. Gracias al MMD podemos comprobar el correcto funcionamiento de componentes y circuitos electrónicos [Circuito, 2017]. Hoy día existen MMD que

utilizan técnicas estroboscópicas para su funcionamiento [Petrovic, 2004]. Sin embargo, en general todo MMD, se conforma de cuatro etapas básicas, las cuales son: a) sensado, b) acondicionamiento, c) control y d) desplegado [Petrovic, 2004]. Adicionalmente a estas etapas básicas se les puede agregar una etapa de comunicación [Kumar, 2002]; y a algunos MMD comerciales se les ha evaluado su relación señal a ruido en modo muestreado [Lapuh, 2017]. Actualmente en el mercado existen MMD de diferentes tamaños y precios, donde las características que los diferencian son: desviación, exactitud, precisión, repetibilidad, resolución y sensibilidad. En particular, este documento describe el diseño e implementación de un multímetro digital de bajo costo para la medición de voltaje y corriente (directa y alterna), impedancia y temperatura. Se dará una breve descripción de los diferentes dispositivos seleccionados y se presentaran resultados experimentales de cada sección del MMD.

Divisor de Voltaje como Sensor de Voltaje

Un divisor de voltaje es un circuito simple que reparte el voltaje de una fuente entre una o más impedancias conectadas. Con sólo dos resistencias en serie y un voltaje de entrada, se puede obtener un voltaje de salida equivalente a una fracción del voltaje de entrada, figura 1.

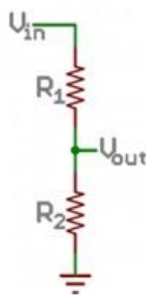


Figura 1 Divisor de voltaje.

La ecuación 1 muestra la relación entre las resistencias y los voltajes de entrada/salida.

$$V_{out} = V_{in} \left(\frac{R_2}{R_1 + R_2} \right) \quad (1)$$

Donde, V_{out} es el voltaje de salida, V_{in} es el voltaje de entrada, y R_1 , R_2 son las resistencias que forman el divisor. La ventaja más importante de un divisor de voltaje es el precio de implementación, en comparación al uso de un amplificador operacional con una ganancia menor a 1 para reducir el valor de salida.

Sensor de Corriente ASC712

El sensor ASC712 ofrece una solución económica y precisa para la detección de corriente alterna o corriente continua en sistemas industriales, comerciales y sistemas de comunicaciones. Las aplicaciones típicas de este dispositivo incluyen control de motores, detección y administración de cargas, fuentes conmutadas y protección contra fallos de sobre corriente [Allegro, 2017]. El sensor trabaja internamente con un sensor de efecto Hall que detecta el campo magnético que se produce por inducción de la corriente que circula por la línea que se está midiendo. Es decir, el sensor entrega una salida de voltaje proporcional a la corriente [Allegro, 2017]. El rango de corriente que se puede medir y sensibilidad varían dependiendo del modelo del circuito integrado, tabla 1.

Tabla 1 Parámetros de sensores de corriente ACS712.

Modelo	Rango (A)	Sensibilidad (mV/A)
ACS712ELCTR-05B-T	-5 a 5	185
ACS712ELCTR-20B-T	-20 a 20	100
ACS712ELCTR-30A-T	-30 a 30	66

El sensor entrega un valor de 2.5 V para una corriente de 0 A y, a partir de allí, incrementa proporcionalmente de acuerdo a la sensibilidad, teniendo una relación lineal entre la salida de voltaje del sensor y la corriente. Dicha relación es una línea recta regida por la ecuación 2.

$$V = ml + 2.5 \quad (2)$$

Donde, “m” es la pendiente (equivale a la sensibilidad), V es el voltaje de salida e I es la corriente medida. Despejando “I”, se obtiene la ecuación 3 para hallar la corriente a partir de la lectura del sensor.

$$I = \frac{V - 2.5}{\text{Sensibilidad}} \quad (3)$$

Sensor de Impedancia AD5933

El AD5933 es un sensor que combina un generador de frecuencia en conjunto con un convertidor analógico digital (ADC, por sus siglas en inglés) de 12 bits y 250 kilo muestras por segundo (ksps). El generador de frecuencia permite que una impedancia compleja externa sea excitada con una frecuencia conocida. La señal de respuesta de la impedancia es muestreada por el ADC y mediante un procesador de señales digitales (DSP) integrado, se efectúa una transformada discreta de Fourier para después devolver una palabra de datos real e imaginaria en cada frecuencia de salida, figura 2 [Medidor, 2017].

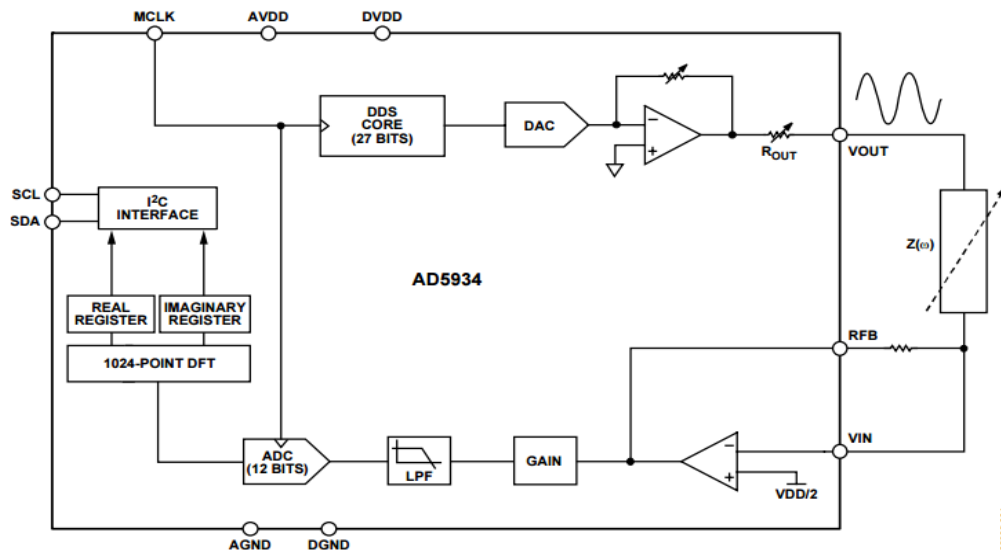


Figura 2 Diagrama a bloques del sensor AD5934, similar al AD5933 [Medidor, 2017].

Este sensor tiene un rango de medición de impedancia entre 1 kΩ y 100 MΩ, con un rango de frecuencia de 1 kHz hasta 100 kHz.

Sensor de Temperatura LM35

La serie LM35 pertenece a la familia de circuitos integrados que entregan una salida de voltaje proporcional a la temperatura registrada. El dispositivo LM35 no requiere obtener un voltaje constante grande de la salida para obtener un

escalamiento conveniente en grados centígrados. No requiere de calibración externa para proporcionar precisiones típicas de $\pm \frac{1}{4} \text{ }^\circ\text{C}$ a temperatura ambiente y $\pm \frac{3}{4} \text{ }^\circ\text{C}$ en un rango completo de temperaturas de $-55 \text{ }^\circ\text{C}$ a $150 \text{ }^\circ\text{C}$, teniendo un factor de escala de $10\text{mV}/^\circ\text{C}$. De acuerdo al fabricante, la baja impedancia de salida, salida lineal y la calibración inherente precisa del dispositivo LM35 hace que la interconexión con los circuitos de lectura o control sea relativamente sencillo [Sensor, 2017].

2. Métodos

El diseño del multímetro se divide principalmente en cuatro secciones: sensado de parámetros, acondicionamiento de señales, adquisición, control y procesamiento de señales y finalmente el despliegado de mediciones, figura 3.

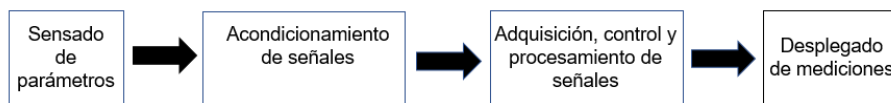


Figura 3 Diagrama de bloques del multímetro digital diseñado.

Divisor de Voltaje Corriente Directa

Para la medición en voltaje de corriente directa, se consideraron los siguientes parámetros para el diseño del divisor: una entrada de 500 V (V_{in}) y un voltaje de salida de 5 V (V_{out}), que será procesado después por el PIC16F877A. La relación de resistencias se muestra en la ecuación 4.

$$\frac{V_{out}}{V_{in}} = \frac{5}{500} = 0.01 = \frac{R_2}{R_1 + R_2} \quad (4)$$

Utilizando valores comerciales de resistencias para la implementación, con los siguientes valores $R_1 = 1 \text{ M}\Omega$ y $R_2 = 10 \text{ k}\Omega$ obtenemos una relación de 0.0099 aproximado al valor deseado.

Divisor de Voltaje Corriente Alterna

Para el diseño del divisor en corriente alterna, se usó un puente rectificador de onda completa. Dentro de las consideraciones se estableció un voltaje de entrada

(V_{in}) de 500 Vrms (equivalente a 707 Volts pico), y un voltaje de salida de 5 V. Sustituyendo en la ecuación 1 se obtuvo la relación de resistencias, ecuación 5.

$$\frac{V_{out}}{V_{in}} = \frac{5}{707} = \frac{R_2}{R_1 + R_2} = .007 \quad (5)$$

Utilizando valores comerciales de resistencias para la implementación, con los siguientes valores $R_1= 1 \text{ M}\Omega$ y $R_2= 10 \text{ k}\Omega$ obtenemos una relación de 0.0099 aproximado al valor deseado. Se propusieron resistencias de valores comerciales $R_2= 10 \text{ k}\Omega$ y $R_1= 1.5 \text{ M}\Omega$, por lo que se obtuvo una relación de 0.0066 muy próxima a la deseada, figura 4.

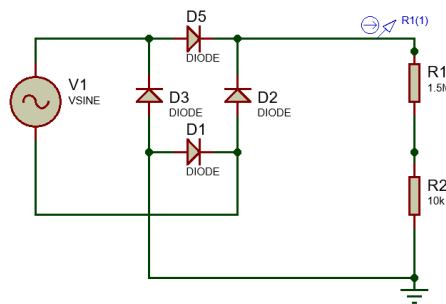


Figura 4 Circuito implementado.

Corriente

Para la implementación del sensor de corriente ACS712, se utilizó una resistencia de 33Ω a 5 W y una fuente de voltaje, figura 5. En dicho arreglo, el ACS712 provee un valor de voltaje, el cual se debe de acondicionar para obtener el valor correspondiente a la corriente. Dicho acondicionamiento se efectuó con la ecuación 2, en la cual se consideró una sensibilidad de 185 mA. Dicha sensibilidad cubre un rango de medición de -5 hasta 5 A.

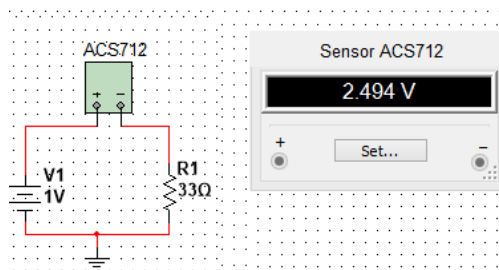


Figura 5 Esquema de medición del ACS712.

Impedancia

El sensor AD5934 está conectado a una tarjeta Arduino Uno en donde se lleva a cabo el registro de las mediciones. Estas mediciones se mandarán al PIC16F877A, donde se procesarán los datos para posteriormente sean mostradas en el LCD. La implementación del sensor se lleva a cabo por los siguientes pasos figura 6. Primero se utilizó el LM555 para generar el tren de pulsos de entrada.

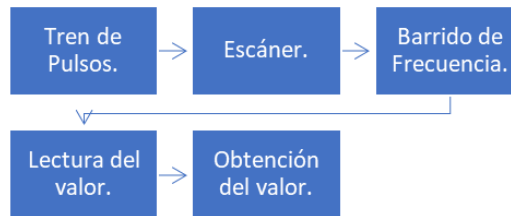


Figura 6 Diagrama de implementación del sensor AD5934.

La ecuación 6 determina la frecuencia en el circuito.

$$T = 0.694(R_1 + 2R_2)C_1 \quad (6)$$

Para generar una frecuencia cercana a 16 MHz, que es la que soporta el reloj maestro (MCLK), se usó el valor de $R_1= 10 \text{ k}\Omega$, $R_2= 33 \text{ }\Omega$ y $C_1= 10 \text{ pF}$. Con estos valores se obtiene una frecuencia de 14.31 MHz, la señal del tren de pulsos debe entrar al MCLK del sensor figura 7. Posteriormente se realiza un escaneo del sensor para determinar que reconoce cada una de las direcciones del sensor, y la comunicación I2C. Este escaneo de igual manera se realiza con un programa de Arduino, figura 8. Para realizar el barrido de frecuencia en el Arduino Uno, se inicia con 1 hasta 100 kHz, con aumentos de 1 kHz. Estos valores son definidos dependiendo de la tarjeta de Arduino usada y con estas especificaciones de frecuencia, se determina la lectura de punto por punto, en este caso se realizó una medición de 100 lecturas.

Temperatura

Para la implementación del LM35, se realizó un acondicionamiento de señal para aprovechar el máximo rango de voltaje del ADC del PIC (0-5V). Para este

diseño, se propuso trabajar en un rango de temperatura de 0 a 150 °C por lo que, al tener una temperatura de 150 °C, en la entrada del ADC del PIC se debe tener un voltaje de 5 V mientras que al tener 0 °C se tendrá 0 V. Para el acondicionamiento, se utilizó un amplificador operacional en configuración de amplificador no inversor, figura 9. De lo anteriormente expuesto, puede notarse que la salida es lineal y cada grado centígrado equivale a 10 mV. Una aproximación para el diseño del acondicionamiento es utilizar un amplificador no inversor con una ganancia que entregue una salida de 5 V máximo.

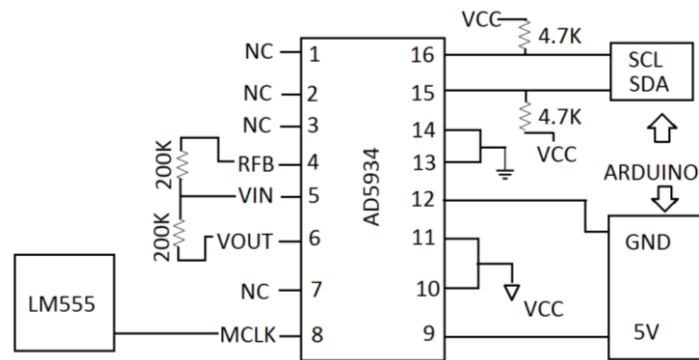


Figura 7 Diagrama de conexión del sensor.

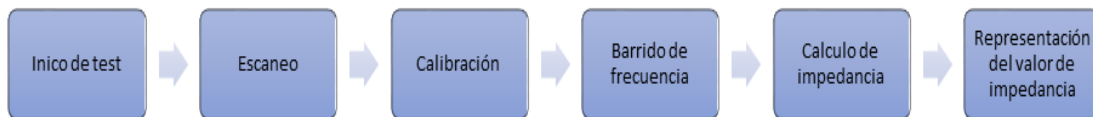


Figura 8 Algoritmo implementado en la tarjeta de Arduino Uno.

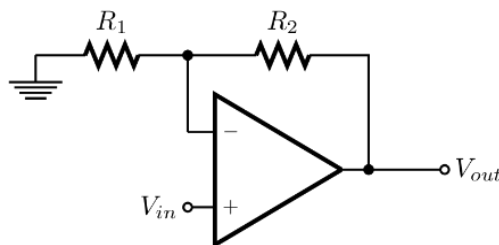


Figura 9 Amplificador operacional no inversor.

La ecuación 7 presenta la relación para la obtención de la ganancia.

$$V_{out} = V_{in}\left(1 + \frac{R_1}{R_2}\right) \quad (7)$$

Considerando una medición máxima de 150 °C, entonces $V_{in} = 1.5$ V y sabiendo que la entrada máxima permisible del PIC es de 5 V, entonces $V_{out} = 5$ V. En la ecuación 8 se presenta el despeje para obtener R_1/R_2 .

$$\frac{V_{out}}{V_{in}} - 1 = \frac{R_1}{R_2} \quad (8)$$

Sustituyendo los valores ya conocidos en la ecuación 8, obtenemos la ecuación 9.

$$\frac{5}{1.5} - 1 = \frac{R_1}{R_2} = 2.33 \quad (9)$$

Se utilizaron valores de resistencias comerciales para R_1 y R_2 por lo que se obtuvo una relación de 2.2. En la ecuación 8, se calculó el nuevo valor de V_{out} con los valores de resistencias propuestos.

Adquisición y Procesamiento de Señales

Para la adquisición de datos hacia el microcontrolador PIC16F877A, se utilizó un multiplexor analógico MUX508IPWR para introducir una medición a la vez. Para la selección de la medición a efectuar, se ocupó un selector giratorio de nueve posiciones conectado al puerto B del PIC. El algoritmo implementado en el microcontrolador consiste en detectar la medición seleccionada por el usuario, controlar al multiplexor para recibir la señal de medición respectiva, realizar el escalamiento de señales y finalmente el control del LCD para el despliegado de mediciones.

3. Resultados

Medición de Voltaje

Para comprobar que el rectificador y divisor de voltaje trabajan correctamente se alimentó el circuito con el voltaje de línea mostrado en la figura 10a. Posteriormente se midió la señal de salida con ayuda del osciloscopio y se obtuvo

una señal mostrada en la figura 10b, la cual presenta una señal rectificada con un voltaje pico de 1.3 V. Para calcular el voltaje RMS se utiliza ecuación 10.

$$V_{rms} = \frac{V_{out} * 707}{5 * \sqrt{2}} \quad (10)$$

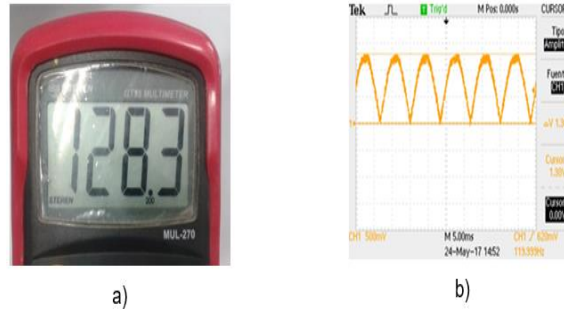


Figura 10 Señal de voltaje de salida rectificada.

Al sustituir el voltaje medido en la ecuación se obtiene un voltaje de 130 V lo que nos indica un valor similar al medido con el multímetro. Para la evaluación de la confiabilidad del instrumento en su medición de voltaje en CA, se realizaron diez mediciones y se calculó el error, ecuación 11, entre el valor medido con el osciloscopio y el valor entregado por el multímetro diseñado.

$$Error = \left\| \left(\frac{RMS_{real} - RMS_{medido}}{RMS_{real}} \right) \right\| \times 100 \quad (11)$$

Donde:

- RMSreal: El valor en Volts (RMS) entregado por la fuente.
- RMSmedido: El valor en Volts (RMS) entregado por el multímetro.

De igual manera para la evaluación de la confiabilidad del instrumento se realizaron diez mediciones y se calculó el error, ecuación 12, entre el valor real y valor entregado por el multímetro.

$$Error = \left(\frac{V_{real} - V_{medido}}{V_{real}} \right) \times 100 \quad (12)$$

Donde:

- Vreal: El valor en Volts entregado por la fuente.
- Vmedido: El valor en Volts entregado por el multímetro.

Se realizaron las mediciones para ambos casos, tabla 2. De igual manera se realizó el promedio del error con el número de mediciones realizadas. Dando un promedio de error es 3.43% para la medición en CA y de 1.77% para CD.

Tabla 2 Mediciones de voltaje del multímetro diseñado.

Voltaje de CD			Voltaje de CA		
RMSreal	RMSmedido	% Error	Vreal	Vmedido	% Error
33	34	3.030	16.3	16.16	0.859
60	62	3.333	18.6	18.18	2.258
73	74	1.370	20.6	20.20	1.942
90	86.4	4.000	24.5	24.24	1.061
101	98	2.970	31.4	31.31	0.287
108	104	3.704	39.6	38.38	3.081
116	110	5.172	44.5	43.43	2.404
121	116	4.132	49.5	48.48	2.061
125	120	4.000	54.5	53.53	1.780
130	126	3.077	59.6	58.58	1.711
132	128	3.030	62.9	61.61	2.051
Promedio		3.438	Promedio		1.77

Corriente

Se realizaron mediciones de un voltaje de 1 a 20 volts de CD, utilizando la resistencia de 33Ω como referencia, figura 11a. Donde se observa el comportamiento del sensor figura 11b, en comparación con un multímetro digital. Al igual que la medición realizada para corriente en corriente directa, la configuración del circuito para la prueba en corriente alterna es la misma, con una resistencia de 300Ω a 25 W. Se efectuaron mediciones de una fuente de CA variable, en la cual se varió el voltaje de 35 a 80 volts.

La figura 12a muestra el comportamiento del sensor, y la figura 12b reporta la comparación con un multímetro digital.

Temperatura

Para comprobar que el sistema trabaja de forma correcta se realizó una medición con el LM35, un termómetro laser y el sistema diseñado, figura 13. Para calcular la temperatura medida por el sensor solo se tiene que multiplicar el valor del voltaje obtenido del operacional por 32 para obtener el valor de temperatura

real. Por lo que al multiplicar el valor obtenido de 0.91 V por 32 °C se obtiene un valor de 29.12 °C que es el valor de la temperatura real.

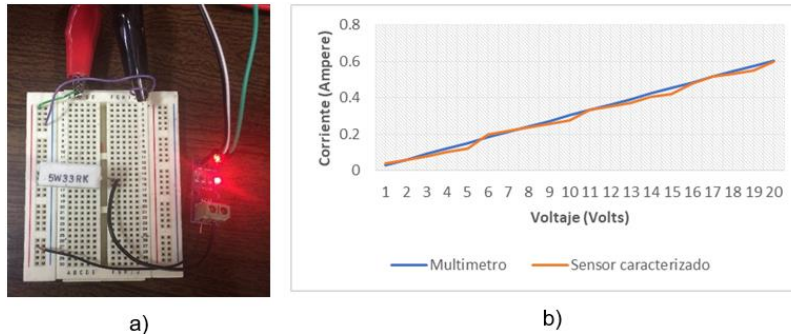


Figura 11 Gráfica comparativa de respuesta.

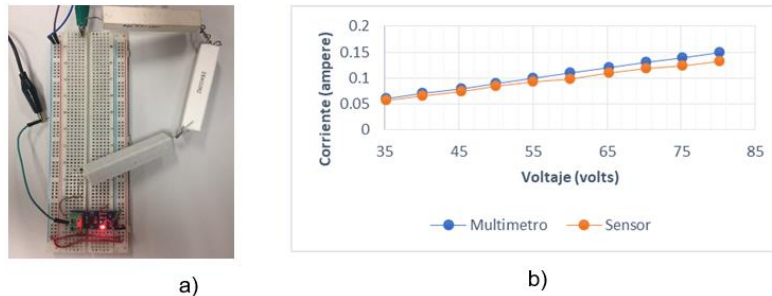
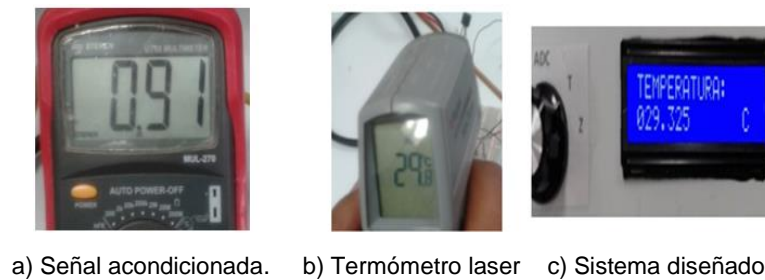


Figura 12 Gráfica comparativa de respuesta en CA.



a) Señal acondicionada. b) Termómetro laser c) Sistema diseñado.

Figura 13 Comparativa de respuesta.

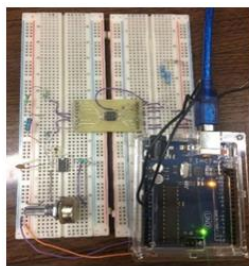
Impedancia

El fabricante indica que el sensor, al conectarlo con los parámetros calculados para la medición de impedancia, calcula su magnitud en cada punto de frecuencia. Con esto, se obtiene el valor real (R) e imaginario (I), dirección de registro 0x94 y 0x95 (R); y las direcciones de registro 0y96 y 0x97 (I). Para obtener los valores R

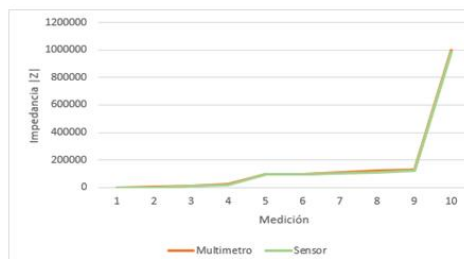
e / se usa la fórmula: $\sqrt{R^2 + I^2}$, donde el valor obtenido se debe multiplicar por el factor de ganancia. Este factor de ganancia se obtiene colocando una resistencia entre los pines Vin y Vout del integrado, colocando una resistencia de 200kΩ se obtiene un factor de 1 (caso utilizado); finalmente, para obtener el valor de la impedancia se usa la ecuación 13.

$$\text{Impedancia} = \frac{1}{\text{Factor de ganancia} * \text{Magnitud}} \quad (13)$$

La frecuencia es para el barrido de lecturas que indicará cuantas muestras realizará, este barrido se efectúa con tres parámetros: frecuencia de inicio, incremento de frecuencia y el número de incrementos. Se implementó un programa, utilizando una tarjeta Arduino, figura 14a. En el programa se realizan las operaciones para determinar la parte real e imaginaria de la impedancia, así como su magnitud figura 14b. Al tener estas mediciones se procesan en un microcontrolador y se envía el dato en forma de voltaje digital, que a su vez lo lee el microcontrolador maestro.



a)



b)

Figura 14 Sensor de impedancia, circuito, multímetro comercial y sistema diseñado.

Sistema Integrado y Lista de Componentes

La figura 15 muestra el sistema diseñado. En la tabla 3 se reporta el costo de cada uno de los elementos requeridos para el sistema presentado. Como se puede observar el costo total de los componentes no supera los \$1150, por lo que se puede concluir que se trata de un sistema de bajo costo con características competitivas.

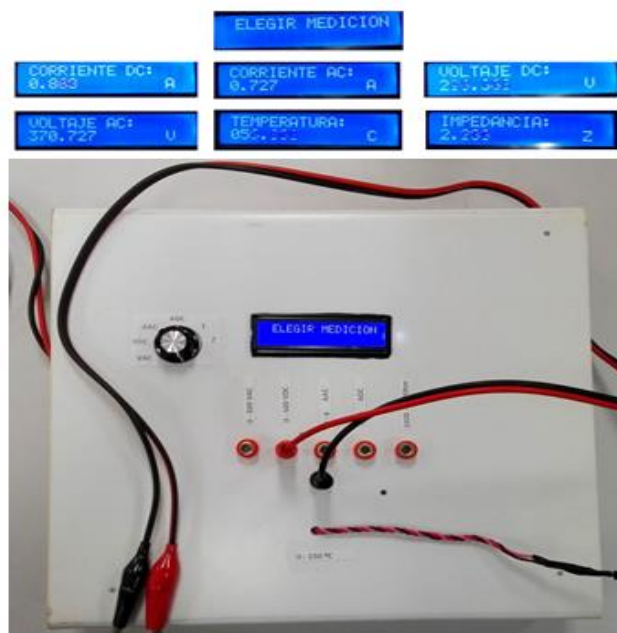


Figura 15 Adquisición, procesamiento y despliegado mediciones multímetro.

Tabla 3 Precio de componentes utilizados.

Elemento	Costo Unitario	Cantidad	Subtotal
ASC712	\$120.00	1	\$120.00
AD5934	\$400.00	1	\$400.00
LM555	\$10.00	1	\$10.00
Arduino	\$180.00	1	\$180.00
LM35	\$37.00	1	\$37.00
Amplificador Operacional LM358	\$13.00	1	\$13.00
PIC16F877A	\$95.00	1	\$95.00
Multiplexor CIM14067	\$151.00	1	\$151.00
LCD 2x16	\$30.00	1	\$30.00
Selector giratorio 9 posiciones	\$9.00	1	\$9.00
Resistencias	\$18.00	1	\$18.00
Capacitores	\$5.00	1	\$5.00
Cristal de cuarzo 4 MHz	\$8.00	1	\$8.00
Zócalo 10 pines	\$10.00	1	\$10.00
Zócalo 40 pines	\$10.00	1	\$10.00
KBPC3510	\$50	1	\$50
TOTAL			\$1,146.00

4. Discusión

Este documento describe el diseño e implementación de un multímetro digital de bajo costo para la medición de temperatura, impedancia, voltaje y corriente.

Para su implementación, se utilizaron sensores comerciales y se aplicaron los conocimientos adquiridos durante los estudios de la licenciatura en Ing. Electrónica. El prototipo final servirá como una muestra de las habilidades del Ing. Electrónico y será utilizado en los diferentes foros de orientación vocacional de nivel medio superior.

5. Conclusiones

Este multímetro digital cumple con las funciones primordiales de medición de corriente y voltaje, tanto para señales de corriente directa tanto como alterna. Además, se añadieron las funciones de medición de temperatura e impedancia, resaltando esta última función debido a la complejidad de la implementación a través del sensor AD5933, obteniendo resultados satisfactorios. El multímetro posee rangos aceptables de medición, con la capacidad de realizar dichos procesos sin la necesidad de hacer un cambio de escala como se hace regularmente con los equipos comerciales. Este trabajo se planea presentar como un proyecto didáctico para jóvenes en el que se pueda apreciar la integración de cada uno de los componentes en una herramienta que todas las personas que se desempeñan en el área de la electricidad y electrónica conocen. Como trabajo a futuro se pretende implementar la medición para valores negativos, así como complementar el multímetro con algunas otras funciones que poseen los multímetros comerciales.

6. Bibliografía y Referencias

- [1] Allegro ASC712: <http://www.alldatasheet.es/datasheet-pdf/pdf/168326/ALLEGRO/ACS712.html>, consultada el 21/05/2017.
- [2] B. R. Kumar, K. Sridharan and K. Srinivasan, "The design and development of a web-based data acquisition system," in *IEEE Trans. Instrum. Meas.*, vol. 51, no. 3, pp. 427-432, June 2002.
- [3] J. U. Green, "A battery-operated digital multimeter-instrument design using large-scale integration," in *Electronics and Power*, vol. 20, no. 14, pp. 573-575, Aug. 1974.

- [4] Circuito: http://www.circuitoselectronicos.org/2007/11/el-multmetro-digital-tester-digital-o_10.html.
- [5] Medidor, <http://www.analog.com/media/en/technical-documentation/data-sheets/AD5933.pdf>, consultada el 24/05/2017.
- [6] Multímetro: <http://tecnoedu.com/F1000/Multimetro.php>.
- [7] P. Petrovic, "New digital multimeter for accurate measurement of synchronously sampled AC signals," in *IEEE Trans. Instrum. Meas.*, vol. 53, no. 3, pp. 716-725, June 2004.
- [8] R. Lapuh, B. Voljč, M. Lindič and O. F. O. Kieler, "Keysight 3458A noise performance in DCV sampling mode," in *IEEE Trans. Instrum. Meas.*, vol. 66, no. 6, pp. 1089-1094, June 2017.
- [9] Sensor: <http://www.ti.com/lit/ds/symlink/lm35.pdf>, consultada el 25/05/2017.

SISTEMA DE MONITOREO DE TEMPERATURA EN RED

Ricardo Godínez Bravo

Universidad Autónoma Metropolitana Azcapotzalco

rgb@correo.azc.uam.mx

Miguel Magos Rivera

Universidad Autónoma Metropolitana Azcapotzalco

mrm@correo.azc.uam.mx

Resumen

Algunas de las materias primas empleadas en la industria del plástico son almacenadas en recintos con temperatura controlada ya que pueden reaccionar exotérmicamente si se presenta un incremento térmico. Con el fin de proporcionar una solución a la medida para una empresa productora de polímeros, se realizó el diseño y la construcción de un sistema de monitoreo de temperatura para cuartos térmicos funcionando en red. Los equipos desarrollados cuentan con elementos que visualizan y alertan, de forma local y remota, cuando la temperatura en algunos de los puntos de monitoreo sale del intervalo correcto. Una interface de computadora desarrollada, además de desplegar los valores de las variables, se encarga de enviar correos electrónicos a usuarios designados para señalarles una situación de peligro. La comunicación entre los dispositivos se realiza mediante una red Ethernet. Los equipos construidos tienen más de 18 meses operando adecuadamente, lo que ha permitido incrementar la seguridad al proporcionar un monitoreo permanente de los cuartos fríos de almacenamiento.

Palabras Claves: Interface de usuario, Microcontrolador, Monitoreo de temperatura, Red Ethernet.

Abstract

Some of the raw materials used in the plastics industry are stored in rooms with controlled temperature. This is because they can react exothermically if a thermal

increase occurs. In order to provide a customized solution for a polymer company, the design and construction of a temperature monitoring system for thermal rooms operating in a network was carried out. The developed equipment has elements that visualize and alert, locally and remotely, when the temperature in some of the monitoring points leaves the correct interval. A developed computer interface, in addition to displaying the values of the variables, is responsible for sending emails to designated users to indicate a dangerous situation. The communication between the devices is made through an Ethernet network. The equipment built has more than 18 months operating properly. This has increased safety by providing permanent monitoring of cold storage rooms.

Keywords: *Ethernet network, Microcontroller, Temperature monitoring, User interface.*

1. Introducción

Empresas dedicadas a la fabricación de polímeros emplean en sus procesos diversas resinas y catalizadores; estos materiales, debido a propiedades químicas, pueden reaccionar exotérmicamente bajo diversas situaciones: agitación, iluminación, calentamiento, etc. Los materiales empleados en esta rama industrial son altamente combustibles por lo que su almacenamiento requiere de normas de seguridad muy estrictas. Las materias primas, en las industrias dedicadas a la fabricación de materiales plásticos, se encuentran generalmente almacenadas en cuartos con temperatura controlada. Lo anterior para evitar una posible reacción química que degrade sus propiedades o que provoque su combustión [García, 2013].

La empresa para la cual se desarrolló el proyecto que se describe en este trabajo, se dedica a la fabricación de diversos termoplásticos moldeables que son la base para la elaboración de componentes utilizados en la industria automotriz, eléctrica y de productos electrodomésticos, por mencionar algunas. La fabricación de este material se realiza mezclando diversos materiales tales como: resinas, catalizadores y fibras. Una de las plantas de esta empresa, ubicada en una de las zonas industriales de la Ciudad de México, cuenta con diversos almacenes tanto

de materia prima como de producto terminado. En estos espacios se encuentran instalados sistemas cuyo único objetivo es mantener la temperatura dentro de rangos seguros, esto es, no cuentan con mecanismos de despliegue de información, ni de alerta en caso de alarma.

Con la finalidad de aumentar la seguridad, tanto de su personal como de sus instalaciones, la empresa decidió diseñar y construir, en colaboración con el Departamento de Electrónica de la Universidad Autónoma Metropolitana-Azcapotzalco, un sistema de monitoreo y alarma para los cuartos fríos de la empresa. El aparato a desarrollar debería trabajar en forma complementaria al sistema de control existente, esto es, operaría como un equipo que permitiría el monitoreo, de forma local y remota, de la temperatura controlada. Así mismo, debería alertar, también en forma local y remota, cuando la variable se encuentre fuera de los rangos adecuados.

En forma local, el aparato cuyo diseño y construcción se presenta en este documento, permite la visualización de la temperatura mediante indicadores numéricos que pueden ser leídos a una distancia mínima de 10 metros. Así mismo, los equipos cuentan con indicadores luminosos que permiten alertar al personal cuando el valor de la temperatura ha salido de los intervalos correctos de operación. La figura 1 muestra una imagen con uno de los aparatos instalados al exterior de uno de las bodegas de producto en la planta.



Figura 1 Vista de uno de los equipos instalado en la planta.

En forma remota, el equipo cuenta con dispositivos que le permiten comunicarse, vía red Ethernet, con computadoras en la planta. Es de esta manera que, con un programa previamente instalado, es posible observar en la pantalla de la computadora las temperaturas de los aparatos conectados en la red de monitoreo. La figura 2 presenta la imagen de una sección de la pantalla de la computadora con los valores de las temperaturas leídas en tiempo real por los equipos distribuidos en la planta.



Figura 2 Pantalla de monitoreo de temperaturas del programa de cómputo.

El programa de cómputo desarrollado permite, no sólo monitorear las temperaturas, sino que también tiene la función de enviar un correo electrónico señalando al personal involucrado con la seguridad de la planta respecto a una situación de alarma debido a un valor fuera de rango.

En el mercado existen equipos con características similares a las del sistema diseñado. La compañía Etherpower ofrece el modelo SUN-2232 [Etherpower, 2015], mientras que Xytronix Research & Design, Inc., tiene el módulo de monitoreo de temperatura X-DAQ [Xytronix Research & Design, Inc., 2016]. Ambos equipos permiten el monitoreo remoto de temperatura mediante algún navegador web. El acceso a la información se encuentra protegido mediante claves de seguridad. Al igual que el equipo que se describe en este artículo, estos aparatos envían mensajes de correo electrónico y permiten, mediante un programa de las mismas empresas, desplegar y almacenar el comportamiento de la temperatura. Sin embargo, estos dispositivos no cuentan con elementos de visualización y alertas locales.

Existen opciones que proporcionan indicación local de la temperatura, la empresa Omega cuenta con el modelo ILD44-UTP [Omega Engineering Inc., 2014],

mientras que Laurel Electronics Inc., propone la serie M24 [Laurel Electronics Inc., 2016]. Ambos equipos tienen indicadores numéricos de 100 mm de altura. En forma estándar, ninguno de los dos aparatos cuenta con funciones de comunicación, esta es una opción adicional. Lo anterior implica que el monitoreo y alerta remoto debe ser implementado con herramientas adicionales.

Así mismo, en la literatura se encuentran diversos desarrollos relacionados con el monitoreo en red de temperatura. En [Silveira et. al., 2016], se presenta un sistema de monitoreo de temperatura empleando una red de tipo WLAN. El dispositivo que se presenta, se basa en el principio del Internet de las Cosas (IoT, por sus siglas en inglés) para proporcionar una lectura de temperatura en forma remota. El aparato no está diseñado para alertar, ni local ni remotamente, sobre funcionamiento fuera de rango. En [Wen-Tsai et. al., 2011], se presenta un sistema de monitoreo remoto de diversas variables existentes en una industria. En este caso la comunicación es mediante una red inalámbrica tipo ZigBee. De la misma manera que en el caso anterior, el dispositivo no cuenta con elementos para alertar que algún valor de las variables consideradas se encuentra fuera de rango.

2. Métodos

Con el objetivo de facilitar la explicación relacionada al diseño y construcción del sistema, este puede dividirse en varios bloques funcionales, mismos que se muestran en el diagrama de la figura 3 y que se describen en esta sección del artículo.



Figura 3 Diagrama de bloques del sistema desarrollado.

Proceso

- **Controlador Red Lion.** En el recinto a monitorear se colocó un termopar tipo k, el cual está conectado a un controlador de temperatura modelo T48 de la marca Red Lion [Red Lion Controls, 2007]. Se trata de un controlador de temperatura en encapsulado 1/16 DIN que acepta distintos tipos de termopares y RTD. El funcionamiento del controlador se configura mediante un conjunto de botones y dos juegos de indicadores de 7 segmentos. El modelo empleado en esta aplicación cuenta con un puerto serial RS-485 que le permite transmitir información sobre su operación, en particular, la temperatura monitoreada. La figura 4 muestra una imagen del controlador empleado.



Figura 4 Controlador modelo T48 de la compañía RedLion empleado.

En esta aplicación, el controlador está configurado para enviar en forma permanente una cadena de caracteres dentro de la cual se encuentra el valor de la temperatura. Los parámetros de comunicación empleados en esta aplicación son: 9600 Baudios, sin paridad, 8 bits de datos y 1 bit de parada; se asignó la dirección “1” para el puerto de red 1 del dispositivo.

- **Recepción de Datos.** El formato de la cadena de datos que el controlador envía por su puerto está definido por el fabricante, la estructura del mismo se muestra en la figura 5:
 - ✓ **Dirección Nodo:** aquí se señala la dirección del equipo al cual se le solicita la información, se trata de un número que varía entre 00 y 99.
 - ✓ **Espacio:** corresponde al carácter ASCII que representa un espacio.

- ✓ **Identificador de Parámetro:** este conjunto de tres caracteres indica el parámetro que se está enviando, la opción para temperatura es “TMP”.
- ✓ **Signo:** este carácter indica el signo del valor numérico que se está enviando.
- ✓ **Dato, Parte Entera:** por medio de tres caracteres, el controlador envía la parte entera del valor numérico del parámetro solicitado.
- ✓ **Punto:** este carácter corresponde al punto decimal del valor numérico enviado.
- ✓ **Dato, Parte Decimal:** este carácter corresponde al único dígito después del punto decimal que el controlador envía del valor numérico del parámetro solicitado.
- ✓ **Unidades:** Corresponde a las unidades de ingeniería del valor numérico del parámetro solicitado.
- ✓ **Retorno de Carro:** corresponde al carácter ASCII que representa un retorno de carro.
- ✓ **Salto de Línea:** corresponde al carácter ASCII que representa un salto de línea.

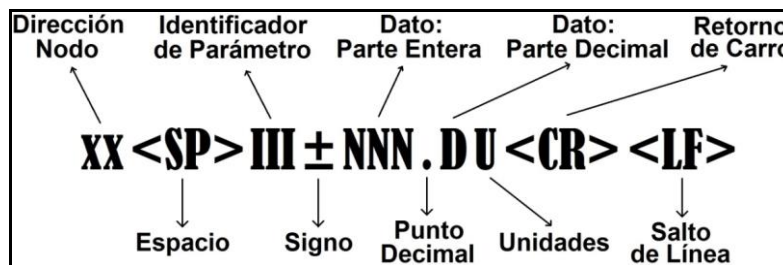


Figura 5 Formato de la cadena de recepción de datos.

Sistema Electrónico

- **Convertidor RS-485 a TTL.** Este bloque, basado en el circuito MAX485, tiene como función convertir los niveles de voltaje que envía el puerto serial del controlador a niveles TTL con los cuales trabaja el sistema digital.
- **Convertidor TTL a Ethernet.** Esta etapa del sistema permite la conexión Ethernet del sistema electrónico. El dispositivo empleado es el modelo NE-

4100T de la compañía Moxa Inc., se trata de un módulo electrónico servidor serie a Ethernet 10/100 Mbps embebido [Moxa Inc., 2008]. Se pueden seleccionar modos de operación tales como: Real COM, TCP Cliente, TCP Servidor y UDP, lo que facilita su integración a una red ya establecida. Su tamaño es de 45 x 36 mm y se alimenta con 5 vdc con un consumo de 1.5 W. Las señales de operación y comunicación están disponibles en un par de tiras de terminales de 13 contactos cada una. La figura 6, muestra una imagen de este dispositivo.



Figura 6 Módulo convertidor NE4100T de Moxa Inc.

El modo de operación, los parámetros de comunicación con el sistema digital y la dirección IP, entre otros valores, son configurados mediante una serie de ventanas almacenadas en la memoria del dispositivo y a las cuales se puede tener acceso mediante algún navegador. Para esta aplicación, el módulo se configuró para trabajar en modo Real COM. La información de la temperatura es enviada a este dispositivo por el microcontrolador a 9600 Baudios, sin paridad, 8 bits de datos y 1 bit de parada. Los parámetros de red también son configurados para este elemento: IP, máscara y puerto de enlace. El valor de la temperatura es enviado al módulo mediante una cadena de 6 bytes en formato ASCII:

[Decenas][Unidades][.][Decimas][CR][LF]

- **Sistema Digital.** El bloque central del sistema electrónico es el sistema digital. Este bloque es el encargado de recibir y decodificar la cadena de caracteres enviada por el controlador de temperatura y convertida a formato TTL por el convertidor RS-485-TTL. A su vez, conforma la cadena de

información que debe transmitirse vía el convertidor TTL-Ethernet. Este mismo bloque genera las señales hacia el conjunto de indicadores numéricos de 7 segmentos, así como a los indicadores de estado, que señalan si la variable de interés se encuentra dentro o fuera del rango correcto. La base de esta etapa es el microcontrolador AT89S52 de ATMEL. Este dispositivo contiene una CPU de 8 bits, 8K Bytes de memoria tipo Flash, 256 Bytes de memoria RAM, 4 puertos de 8 bits bidireccionales para entrada y salida en paralelo, los cuales pueden ser direccionados por bit; contiene señales de control para comunicación serial, 3 temporizadores/Contadores de 16 bits, terminales para detectar interrupciones externas y puede trabajar como procesador con una capacidad de direccionamiento de hasta 64K de memoria externa ROM y RAM, además de contar con un puerto serie de tipo "FULL DUPLEX", el cual se emplea en esta aplicación para comunicarse con el Controlador de Temperatura y la red local LAN, lo anterior multiplexando sus líneas de comunicación. En la figura 7 se muestra en forma simplificada el diagrama de flujo del programa desarrollado para el microcontrolador, mismo que se encuentra residente en memoria.

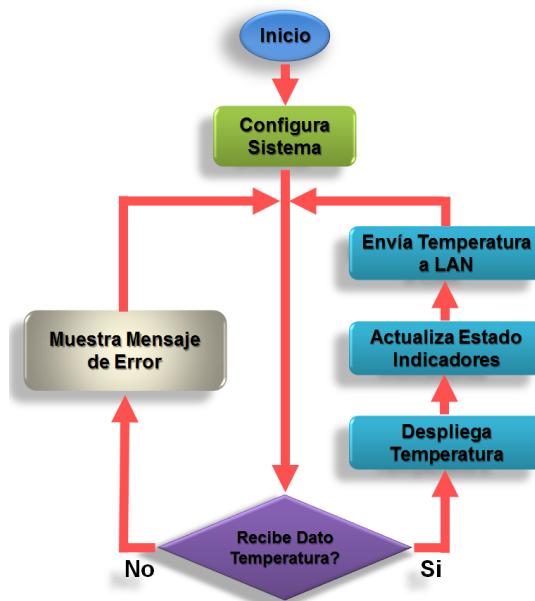


Figura 7 Diagrama de flujo del programa del microcontrolador.

Se observa como el microcontrolador se encuentra en un ciclo infinito leyendo el dato que el controlador le envía. La condición de error se basa en un intervalo de tiempo, si un dato no llega pasado el período establecido, se despliega un código de error en los indicadores numéricos. Si el dato es recibido dentro del tiempo programado, se procede a desplegarlo en los indicadores numéricos, se actualiza el estado de los dos indicadores luminosos y se envía el valor a la red de área local.

Indicadores

- **Numéricos.** El equipo cuenta con tres indicadores numéricos de 7 segmentos de 3 pulgadas de alto, los cuales permiten la visualización de la temperatura para un monitoreo local de la misma. Este conjunto de indicadores permite desplegar el valor de la temperatura ambiente al interior del espacio a monitorear con la precisión de una décima de grado centígrado y a una distancia de por lo menos 10 metros.
- **Luminosos.** Adicionalmente se incluyó un indicador luminoso de color verde y otro amarillo. En forma local, el primero señala si la temperatura se encuentra dentro del rango seguro, mientras que el segundo indica una situación de precaución debido a una temperatura ligeramente fuera del intervalo correcto. Cabe mencionar que el equipo cuenta con un relevador el cual es activado cuando la temperatura se encuentra en un valor de peligro. Los contactos de este elemento pueden ser empleados para activar una alarma sonora y/o luminosa que alerte de la situación a zonas remotas de la planta.

Programa de Cómputo

Como ya se mencionó, uno de los requisitos planteados para este equipo, era la posibilidad de monitorear en forma remota la temperatura en los recintos bajo supervisión. Para esto se desarrolló una interface para computadora en Visual Basic. En la figura 8 se muestra el diagrama de flujo del programa elaborado.

El programa de cómputo está diseñado para poder leer la información de varios monitores de temperatura, siempre y cuando estos se encuentren conectados a la

misma red. Por lo anterior, la primera acción que se realiza al arrancar la aplicación, es seleccionar a cuál de los equipos conectados se le solicitará la información. Posteriormente, una vez que se ha establecido comunicación con el servidor de red, se lee el valor de la temperatura, la cual es desplegada en los indicadores de texto de la interface. Si el valor de la variable se encuentra fuera de rango, se ejecuta la subrutina encargada de enviar correos electrónicos señalando la situación a las direcciones previamente configuradas. El programa se encuentra en un ciclo infinito realizando las acciones descritas, lo anterior, hasta que el usuario decide cerrar la aplicación.

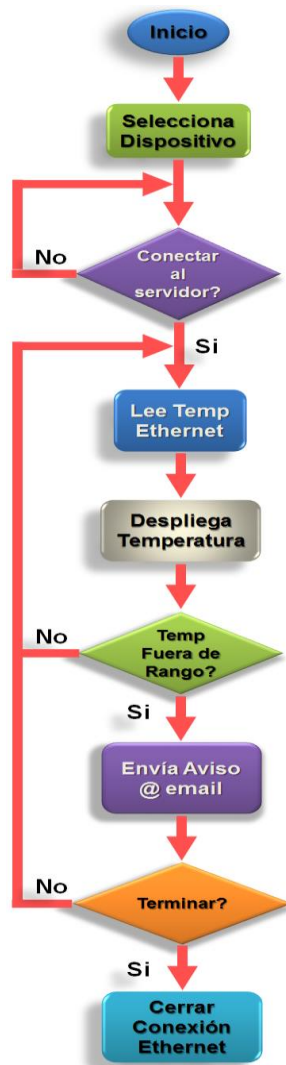


Figura 8 Diagrama de flujo del programa en Visual Basic.

3. Resultados

El resultado que se presenta de este proyecto es el equipo físico y la interface desarrollada. Una vez que el diseño de las distintas partes del equipo fue realizado, se procedió a construir el módulo, el cual se encuentra dentro de un gabinete, así como a elaborar el programa de cómputo. En esta sección del artículo, se presentan los resultados obtenidos.

Monitor de Temperatura

Se diseñaron y construyeron dos tarjetas electrónicas, la primera de ellas corresponde al sistema electrónico. En esta se encuentran los elementos asociados al sistema digital, así mismo, en esta tablilla se ubicaron los convertidores RS-485 a TTL y TTL a Ethernet. La segunda tarjeta soporta los tres indicadores numéricos de 7 segmentos. La primera de las tarjetas se encuentra instalada en una platina al interior del gabinete, mientras que la correspondiente a los indicadores se colocó al frente del mismo. La conexión entre las dos tarjetas se realiza mediante un conector de cable plano.

Adicionalmente se tiene una fuente de 24 VDC, la cual alimenta los distintos elementos del equipo, así como un interruptor de perilla que permite el encendido del equipo. El gabinete empleado es el modelo CRN-3025/150 de la marca Himel, se trata de un gabinete metálico con puerta y platina interior, sus dimensiones externas son: 300 x 250 x 150 mm. La figura 9 muestra una vista del interior del gabinete con los elementos instalados.



Figura 9 Vista del interior del monitor de temperatura construido.

En la imagen se observa al lado izquierdo el interior del gabinete. Se tiene en la parte superior la fuente de voltaje, así como un conjunto de clemas que facilitan las conexiones y el mantenimiento del equipo. Del mismo lado, pero en la parte inferior se encuentra la tarjeta electrónica correspondiente al sistema electrónico. En el lado derecho de la imagen se tiene la parte trasera de la puerta del gabinete. En esta se observa en la parte superior la tarjeta electrónica con los indicadores numéricos, mientras que en la parte inferior se tienen los indicadores luminosos que señalan localmente el estado de la variable monitoreada. Se observan también la parte posterior del controlador de temperatura, así como del interruptor del equipo. En la figura 10 se tienen dos vistas frontales del monitor de temperatura encendido, mientras que la figura 11 se muestran imágenes de dos monitores instalados en la planta.



Figura 10 Vistas frontales del monitor de temperatura construido.



Figura 11 Vistas de dos monitores instalados en la planta.

Interface de Monitoreo

La interface que se elaboró contempla la conexión de tres monitores de temperatura. La figura 12 muestra la vista de la pantalla de monitoreo por computadora del sistema desarrollado.



Figura 12 Pantalla de monitoreo de la interface desarrollada.

La ventana de la interface desarrollada es bastante sencilla. En la parte superior se despliega el logotipo de la empresa, en este documento se omitió por razones de propiedad industrial. En la parte inferior se despliegan tres campos en los cuales se muestra el valor de la temperatura en igual número de recintos. Se incluyó un indicador que señala cuando el equipo ha logrado conectarse a la red en la cual se encuentran los equipos.

4. Discusión

Anterior al desarrollo de este proyecto, la única forma de saber si había un problema era cuando el personal detectaba esta situación y verbalmente lo hacía saber a los responsables. En caso de alguna situación de peligro, una alarma era activada manualmente por el personal. Respecto al monitoreo de la temperatura, un empleado estaba encargado de leer cada dos horas termómetros instalados al interior de cada cuarto frío. La forma establecida de supervisar la temperatura, al ser dependiente de personas, implicaba demasiados riesgos, además de no ser funcional por las noches o en los días que la planta no tenía labores.

Los equipos construidos y descritos en este artículo han permitido a la empresa contar con un sistema de monitoreo y de alarma, tanto de forma local como remota, confiable. Lo anterior redundó en un incremento de la seguridad tanto para las instalaciones como para el personal. Debido a que desde el inicio de operaciones de la empresa nunca se ha presentado algún problema mayor con respecto al control de la temperatura en los cuartos fríos, no se puede cuantificar la ventaja de contar con los equipos construidos. El beneficio radica en contar con un sistema confiable que deja de ser dependiente del personal.

5. Conclusiones

Tomando en cuenta los requisitos propuestos por la empresa al momento de plantear el proyecto, se considera que los objetivos fueron alcanzados en forma adecuada. Al momento de redactar este artículo, se tienen instalados tres monitores de temperatura en igual número de puntos de la planta. Los equipos tienen cerca de 18 meses operando. Durante este período los aparatos han facilitado las tareas relacionadas con el monitoreo de la temperatura de forma local y remota. Respecto al sistema de alertamiento remoto se tiene el problema que éste es funcional solo si una persona se encuentra frente a la computadora y tiene abierta la aplicación. Así mismo, la opción de enviar correos electrónicos fue desactivada al poco tiempo de empezar a operar los equipos, lo anterior debido a que los rangos de operación para las temperaturas establecidos en los manuales de seguridad estaban demasiado reducidos generando constantemente envíos de mensajes.

Como trabajo a futuro se están elaborando nuevas versiones de los programas, tanto para el sistema electrónico, como para la interface, que contemplan comentarios realizados por el personal de la planta. Así mismo, se considera el manejo de la información de forma remota con ayuda de una página web, lo anterior permitiría a los usuarios el tener acceso a la misma, desde cualquier dispositivo capaz de conectarse a internet.

6. Bibliografía y Referencias

- [1] Etherpower, SUN 2232, Manual de usuario, Buenos Aires, Argentina, 2015.
- [2] García S., Recomendaciones, salud y seguridad en la manipulación de las resinas, Revista Iberoamericana de Polímeros, ISSN 1988-4206, Vol.14, No.5, Septiembre 2013.
- [3] Laurel Electronics Inc., Magna Process 4-Large Digit Process Indicator, California, USA, 2016.
- [4] Moxa Inc., NE-4100 Series, User's Manual, USA, 2008.
- [5] Omega Engineering Inc., IDL-44 UTP Big Display Universal Temperature & Process Controller Manual, Connecticut, USA, 2014.

- [6] Red Lion Controls, The 1/16 DIN Controllers Temperature/Process, Instruction Manual, USA, 2007.
- [7] Silveira E., Bonho S., Temperature Monitoring Through Wireless Sensor Network Using an 802.15.4/802.11 Gateway, IFAC Papers on Line, Vol. 49, No. 30, pp. 120-125, 2016.
- [8] Wen-Tsai S., Yao-Chi H., Designing an industrial real-time measurement and monitoring system based on embedded system and ZigBee, Expert Systems with Applications, Vol. 38, No. 4, pp. 4522-4529, April 2011.
- [9] Xytronix Research & Design Inc., X-DAQ. Temperature Module, User's Manual, Utha, USA, 2016.

IMPLEMENTACIÓN DE EFECTOS DE SONIDO PARA GUITARRA ELÉCTRICA EN LA TARJETA C6713 DSK

Claudia Gómez Borrás

Universidad Autónoma Metropolitana-Azcapotzalco

cgb@correo.azc.uam.mx

Javier Alducin Castillo

Universidad Autónoma Metropolitana-Azcapotzalco

jac@correo.azc.uam.mx

Resumen

La implementación de efectos de sonido en forma digital mediante procesamiento de audio es una opción económica y flexible de modificar el audio para crear un tono o efecto particular acorde al gusto del músico. En la actualidad existen sólo unos cuantos pedales de guitarra programables comerciales, así como pedales multi-efectos digitales que brindan al usuario algunas opciones para poder elaborar sus propios efectos. Teniendo en cuenta lo anterior, en este trabajo se presenta el desarrollo de un sistema de procesamiento de audio en tiempo real, a través del DSP TI's C6713 DSK, que tiene la ventaja de ser completamente modificable al tener acceso al código fuente. Se implementan exitosamente efectos de sonido como el efecto wah-wah y reverberación, así como dos filtros digitales básicos, los cuales se diseñaron previamente en el software Octave y se trasladaron para su realización en el DSP. El sistema desarrollado brinda la posibilidad de elegir el efecto deseado, sin reprogramar el DSP, para su inclusión en una señal de audio generada en tiempo real, a través de un instrumento como una guitarra. Por otro lado, basta modificar algunos parámetros en el programa elaborado para obtener variantes de tales efectos evitando así la necesidad de adquirir otro equipo como sucede en el caso de los dispositivos analógicos.

Palabras Clave: DSP C6713, efectos de audio, Octave, procesamiento de audio.

Abstract

The implementation of digital sound effects through audio processing, is an economical and flexible option to modify sound to create a special tone according to the taste of the musician. Currently there are just a few commercial programmable guitar pedals as well as digital multi-effects pedals that provide options to the users to make their own effects. Considering the above, this work shows the development of a real-time audio processing system, through DSP TI C6713 DSK, which has the advantage of being completely modifiable by having access to the source code. Sound effects such as the wah-wah effect and reverberation, as well as two basic digital filters, which were previously designed in the Octave, were successfully implemented in the DSP. The developed system offers the possibility to choose the desired effect without reprogramming the DSP for inclusion in a real-time audio signal through an instrument such as a guitar. On the other hand, it is enough to modify some parameters in the program elaborated to obtain variants of the effects and thus avoids the need to acquire another equipment as it happens in the case of analog devices.

Keywords: *Audio effects, audio processing, DSP C6713, Octave.*

1. Introducción

Hoy en día, las aplicaciones principales de audio con DSP's son: codificación de alta calidad de audio, así como la generación y manipulación digital de señales de música. En el ámbito musical, un DSP puede resultar útil para realizar varias funciones, entre ellas, la generación de efectos de sonido mediante la aplicación de diferentes filtros digitales. Algunos efectos de sonido como eco y reverberación pueden ser generados a través de un sistema lineal; otros efectos como la distorsión requieren sistemas no lineales.

En las etapas de acondicionamiento y procesamiento de una señal de audio es posible, mediante sistemas discretos y digitales, sustituir a sistemas analógicos como amplificadores, moduladores, filtros, entre otros. Esto se debe a que los DSP's han permitido que sea más factible implementar determinadas funciones mediante técnicas de procesamiento digital de señales. Además, de forma

comparativa, los componentes analógicos requieren mayor espacio; deben ser ajustados en sus valores adecuados por única vez al construirse, por lo que los ajustes disponibles son aquellos con los que fueron configurados de fábrica y presentan alteraciones tras el cambio de temperatura, así como con el transcurso del tiempo de uso y posibles fallas en sus componentes. Un sistema de procesamiento de audio implementado con un DSP es mucho más sencillo de modificar, corregir en caso de error y actualizar su funcionamiento, pues basta con reprogramarlo. Particularmente la producción de efectos de guitarra, mediante procesamiento digital de señales, tiene como ventajas: poder tener múltiples efectos en un solo sistema y contar con características y parámetros personalizables que los sistemas analógicos no pueden lograr, sin embargo, es necesario tener en cuenta que se tienen limitaciones debido a las características del DSP, del procesamiento y de programación.

En este trabajo se propone realizar la implementación de dos efectos de sonido: *wah-wah* y *reverberación* usando un DSP TI C6713 DSK y también la implementación de dos filtros básicos: *filtro pasa-bajas* y *pasa-altas*. Si bien, el DSP es un modelo con algunos años en el mercado, posee características básicas para poder realizar el procesamiento de audio en tiempo real. Una de las motivaciones para realizar este trabajo es que se plantea que el sistema tenga la capacidad de ser modificable en los parámetros de cada efecto, de tal manera que el usuario logre la configuración que desee y pueda seleccionar entre cada uno.

Las señales de audio se pueden modificar de acuerdo con una serie de parámetros de un sistema digital, de acuerdo con el criterio del usuario; la configuración de los parámetros produce el efecto de sonido deseado. En la literatura, existen algunos trabajos de revisión de diversos efectos de sonido desde el punto de vista analógico y digital; respecto a este último se pueden encontrar trabajos como [Dattoro, 1997] y [Zölzer, 2003], donde se describen una gran cantidad de efectos de sonido. En [Ballou, 2013], no sólo describen efectos de sonido, sino también realiza una interesante descripción física del sonido, así como los diversos fenómenos físicos que pueden alterarlo. En [Ballou, 2013], se utiliza el DSP modelo C6713 DSK de TI para efectos de guitarra, los filtros

digitales fueron diseñados en Matlab generando los siguientes efectos: eco, coro, reverberación, flanger, wah-wah, ecualizador de 8 bandas y distorsión. Una de las desventajas de ese trabajo es que no se integran los efectos en un solo programa, es decir, para realizar alguno de los efectos de sonido, se reprograma cada vez el DSP. Además, requiere el uso de un software comercial de alto costo (Matlab) para el diseño de los filtros digitales, además de requerir diversos *toolbox* los cuales también tienen un costo significativo.

Por otro lado, en [Digital Guitar Effects, 2017] se implementaron varios efectos de sonido, con fines didácticos usando Matlab junto con un DSP cuyo modelo no se especifica. Los efectos implementados fueron flanger y wah-wah, sin embargo, los autores mencionan que debido a errores en el desarrollo del código no se obtuvo la respuesta deseada, ya que el audio de salida del proceso no mostraba los efectos esperados. Otro enfoque de algunos trabajos es únicamente la simulación de efectos usando Matlab, por ejemplo, en [Vargas, 2016], realizan la simulación del efecto *booster* (incremento de magnitud de la señal) y del efecto *delay*, sin embargo, no implementan los efectos en tiempo real. En [Adil, 2015], también simulan los efectos *pitch*, *echo*, *tremolo* haciendo uso de Matlab. En [Glover, 2011], proponen el uso de Python junto con la librería *Modal*, para procesar audio e implementar efectos de sonido, si bien hacen uso de software libre, no privativo, no se ejecuta en un DSP.

El procesamiento de audio es un nicho de mercado, por esta razón, muchos de los sistemas comerciales tienen patentes específicas, ya que la configuración de parámetros, así como el diseño de los sistemas que producen los efectos son protegidos celosamente, por ello existen diversas patentes, por ejemplo, en [Daniel, 2009] se plantea un sistema multiefectos para guitarra mediante un DSP y una unidad de control dinámico sensible al tacto, lo que ayuda a controlar y procesar las señales producidas por la guitarra. En [Ryle, 2010], se propone un sistema basado en DSP con un módulo extraíble, el cual contendrá los efectos de sonido a realizar, es decir, hay una protección adicional al sistema desarrollado, ya que el DSP no tendrá las funciones a realizar, si no un módulo adicional proporcionará al DSP las funciones a realizar. En [Pennock, 2006], se patentó un

sistema basado en DSP que realiza efectos de sonido, la descripción que realizan menciona modelos matemáticos complejos, sin embargo, al leer detenidamente la patente, realizan filtros digitales básicos, donde los coeficientes son importantes, ya que suelen ser éstos los que diferencian a un sistema de otro.

En el mercado también existen los *pedales programables* [Programmable Guitar Pedals, 2017]. Hay que considerar que algunos de estos sistemas son de código abierto y es posible personalizarlos, aunque ciertos pedales requieren programarse en lenguaje ensamblador en vez de C, o bien utilizar su software bajo un sistema operativo de paga lo cual resulta desventajoso para el usuario. Otros inconvenientes son que algunos de estos dispositivos no cuentan con un preamplificador que permita tratar directamente la señal proveniente de la guitarra, no tienen suficientes perillas o tienen una fidelidad muy pobre.

El uso de Matlab para el diseño de filtros digitales a implementarse en el DSP o para simulación, es extendido. Sin embargo, es un software comercial costoso, por ello, en este trabajo se propone realizar el diseño de los sistemas que realizan los efectos de sonido usando Octave, un software de licencia *GNU*, muy similar a Matlab.

Características de los Efectos

Los efectos digitales de audio se catalogan de acuerdo con la forma en que se realiza su procesamiento. Clasificación de los efectos implementados:

- Filtrado Básico: Pasa bajas, Pasa altas.
- Filtros de tiempo variable: Wah-wah.
- Efectos espaciales: Reverberación.

Reverberación

La reverberación natural es un sonido que resulta del reflejo atenuado y en distintos ángulos de las ondas sonoras en las superficies de un espacio cerrado. Este sonido varía según las características del lugar como el tamaño y hasta el material de las superficies que lo componen, debido a la forma en como es absorbido y reflejado.

Es posible reproducir este efecto de forma artificial usando un sistema digital, sin embargo, tiene dos principales limitantes: memoria disponible y velocidad de procesamiento. Esto es debido a que se requiere almacenar datos para poder crear la sensación de escuchar el sonido reflejado tras un breve retardo.

Reverberación Digital

Existen diversos tratamientos, entre ellos los digitales, para conseguir un efecto de reverberación. Parte de estas técnicas se encuentran limitadas por los requerimientos de cómputo: por ejemplo, el modelado físico tridimensional de un cuarto reverberante, el cual precisa de la aplicación de distintas leyes físicas, así como de diversos parámetros [Ballou, 2013]. La mayoría de los trabajos para reproducir la reverberación se basan en el empleo de sistemas cuyas variables sean controlables, pues resulta más simple aun cuando no simulen con fidelidad las características de un determinado espacio reverberante.

Para producir un sistema básico de reverberación se usa generalmente un *filtro peine*, ver figura 1, resulta sencillo modificar heurísticamente el número de retrasos (m) y atenuación (g) para producir el efecto de reverberación. Sin embargo, en la práctica no es fácil, por ello los equipos digitales que lo generan conllevan un desarrollo y patentes por parte de las compañías que los producen y desarrollan. Un sistema ampliamente utilizado es el realizado en [Moorer, 1979], conocido como *reverberador de Moorer*. Se considera un sistema complejo de reverberaciones, ya que actúan con n número de etapas en paralelo; es capaz de recrear reflexiones tempranas y tardías de sonido (60ms a 100ms). Cada filtro de peine actúa como un reverberador y permite simular el efecto de reflexión del sonido en las paredes de un cuarto, así como el tiempo de trayecto de una pared a otra. Además, al añadir la señal de entrada a la salida permite generar la sensación de estar cerca de la fuente.

En el mismo trabajo se sugiere que los retardos sean números primos, ya que de esta forma se evita la acumulación de picos sobre la misma muestra, lo que a su vez permite que el decaimiento sea más denso y uniforme.

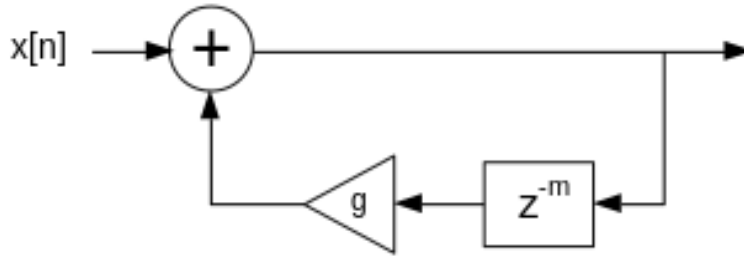


Figura 1 Filtro de peine.

Wah-wah

Este efecto de sonido recibe dicho nombre debido a que imita la articulación de la expresión *wah* mediante la voz humana. Comúnmente se utiliza en guitarras, donde el sonido se hace pasar por filtros que lo modifican al añadir y suprimir alternadamente sus componentes en distintas frecuencias. Esto además de percibirse como la imitación de la pronunciación de un *wah*, crea la sensación de que el tiempo de duración de la nota ejecutada mediante el instrumento se prolonga. Para su implementación de forma digital se utiliza un filtro de estado variable con una frecuencia central f_c variable, estos filtros fueron tratados inicialmente en [Oppenheim, 1976] y ayudan en la generación del efecto wah-wah. La descripción del filtro implementado para este efecto se amplía en la sección Métodos.

Características del TMS20C6713 DSK

La tarjeta TMS20C6713 DSK pertenece a la familia TMS320 C67x de DSP's de punto flotante de Texas Instruments, tiene arquitectura VLIW (Very Long Instruction Word) con 8 unidades de ejecución en su CPU y velocidades de reloj que alcanzan hasta los 350MHz. Tiene palabras de 32 bits y se programa en lenguaje C. Dispone de un Códec Estéreo AIC23, el cual tiene una tasa de muestreo en un rango de 8-96kHz, de 16 a 32 bits. Este códec lleva a cabo las funciones de un ADC y un DAC. Cuenta con líneas de entrada, de salida, así como para micrófono y audífonos. La tarjeta tiene también LED's y DIP switch que facilitan algunas aplicaciones que el usuario desee implementar.

2. Métodos

Se implementaron los efectos descritos anteriormente siguiendo el diagrama a bloques de la figura 2.

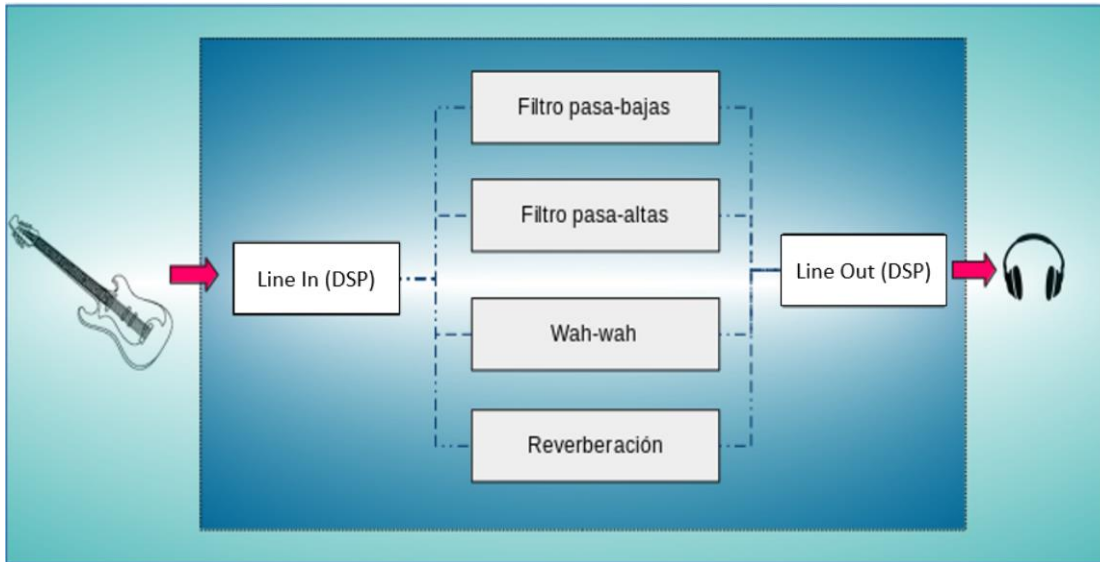


Figura 2 Diagrama a bloques del sistema desarrollado.

DSP

El bloque de DSP está constituido por la tarjeta C6713. Se utilizaron las conexiones de Line In y Line Out, para conectar la guitarra (señal de entrada) y los audífonos (señal de salida). Se configuró el códec con una frecuencia de muestreo de 24 kHz. La estructura del código principal se muestra en el siguiente pseudocódigo:

```
#include <stdio.h>
#include <dsk6713.h>
#include <dsk6713_aic23.h>
#include <access.h>
Set DSK6713_AIC23 Default
Set CodecHandle
Initialize Board Support Library
Initialize DIP switch
Initialize LED
Open the codec in eI_handle0
Set Freq of AIC23 Codec
while(1){
```

```
while(read AIC23){
  conversion data "typecasting"
  if (switch(1) On)
    low pass filter      endif
  if (switch(2) On)
    high pass filter     endif
  if (switch(3) On)
    wah-wah effect       endif
  if (switch(4) On)
    reverberation filter  endif
  conversion data "typecasting"
  while(write AIC23)}}}
```

Inicialmente, se carga la configuración de puertos y memoria predeterminada, se configura el códec de audio para poder utilizar el ADC y DAC adecuadamente. Se establece la frecuencia de muestreo; se inicializa el DIP switch, ya que a través de éste se podrá elegir el efecto deseado; al mismo tiempo se inicializan los led's disponibles, únicamente como indicador del efecto en funcionamiento. Este programa se ejecuta en tiempo real, por ello se usa una sentencia while, para asegurarse que las líneas de código se ejecuten cuando el canal sea leído. A continuación, se realiza una conversión de datos, a punto flotante unitario, por lo que se divide por 32767 (valor máximo de 15 bits) y se realiza un *typecasting* a short (16 bits) y después a flotante, de tal manera que los valores de la señal se encontrarán en el rango 0-1. Con el DIP switch se selecciona el efecto o sistema a ejecutar, más adelante se detallará el funcionamiento de cada uno. Una vez realizado el efecto es necesario escribir los datos de salida para que se pueda transferir la información al puerto Line out, previo acondicionamiento de datos (multiplicando por 32767 y realizando un *typecasting* a short), de tal forma se pueda escuchar adecuadamente la señal de salida, la escritura también se realiza mediante una sentencia while, para asegurar que los datos se escriben correctamente antes de pasar a la siguiente línea. Con el pseudocódigo anterior, basta para poder configurar adecuadamente el DSP para la lectura y escritura de datos usando el códec de audio. Cada uno de los efectos fue realizado en código C para su implementación en el DSP.

Filtro Pasa Bajas

Se implementó mediante un filtro FIR de fase lineal de orden 40 con una ventana de Blackman, ésta permite un rizado constante en la banda de paso y en la banda de rechazo, la zona de transición es estrecha. La frecuencia de corte superior se fijó en 3 kHz. Los coeficientes utilizados, fueron obtenidos a través Octave y trasladados para su uso en el DSP. Al seleccionar el DIP switch 0 del DSP se implementa este filtro, atenuando componentes de frecuencia altos, que corresponde a tonos agudos.

Filtro Pasa Altas

Se implementó mediante un filtro FIR de fase lineal de orden 40 con una ventana de Blackman. La frecuencia de corte inferior se fijó en 9 kHz. Los coeficientes utilizados, fueron obtenidos a través Octave y trasladados para su uso en el DSP. Al seleccionar el DIP switch 1 del DSP se implementa este filtro, atenuando componentes de frecuencia bajos, es decir, eliminando sonidos graves. El uso de un orden reducido en los filtros pasa bajas y pasa altas de tipo FIR, obedece a un compromiso de diseño en tiempo real, entre mayor sea el orden del filtro, mayor memoria (muestras a almacenar) es requerida además de incrementarse el retardo de grupo.

Efecto wah-wah

En Figura 3, se muestra el diagrama de un filtro de estado variable propuesto por [Oppenheim, 1976] y que se tomó como base para este trabajo. Este filtro se modela a través de las ecuaciones 1, 2 y 3, donde el parámetro F_1 (ecuación 4) se recalcula para cada iteración del filtro, dependiente del valor de la frecuencia F_c (modificada en cada iteración), éste último parámetro es la frecuencia central variable que produce el efecto wah-wah. La ecuación 6 muestra la función de transferencia discreta que modela el comportamiento del filtro de estado variable propuesto.

$$y_1[n] = F_1 y_b[n] + y_1[n - 1] \quad (1)$$

$$y_b[n] = F_1 y_h[n] + y_b[n - 1] \quad (2)$$

$$y_h[n] = (x[n] - y_1[n - 1]) - Q_1 y_b[n - 1] \quad (3)$$

Donde:

$$F_1 = 2 * \text{sen} \left(\pi \frac{f_c}{f_s} \right); \quad (4)$$

$Q_1 = 2\xi$, siendo ξ el factor de amortiguamiento

$$r = F_1, \quad q = 1 - F_1 Q_1 \quad (5)$$

$$H(z) = \frac{r^2}{1 + (r^2 - q - 1)z^{-1} + qz^{-2}} \quad (6)$$

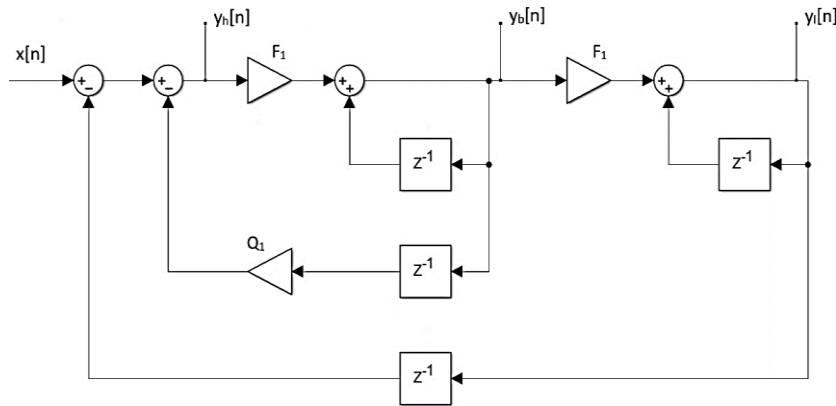


Figura 3 Diagrama a bloques para el efecto wah-wah.

Para realizar el efecto wah-wah se implementaron en el DSP las ecuaciones de diferencias mostradas en las ecuaciones 4, 5 y 6. La frecuencia de muestreo (f_s), requerida para la ecuación 4, es la misma implementada por el códec AI123 (24kHz), la frecuencia f_c corresponde a la frecuencia central de la implementación wah-wah, para este trabajo se definió una frecuencia central mínima de 500 Hz y máxima de 3 kHz, con incrementos de 0.125 Hz por iteración. Las ecuaciones de diferencias se realizan en cada iteración recalculando el valor de F_1 a partir de la variación de f_c , es decir, en cada iteración se logra un filtro digital de frecuencia central variable incrementando y decrementando la frecuencia central. Lo anterior produce el efecto wah-wah. Al seleccionar el DIP switch 2 del DSP, se implementa este efecto.

Durante la implementación de este efecto de sonido en el DSP, se realizaron dos tipos de modalidades. En la primera, el efecto wah-wah afecta de manera continua

a la entrada del DSP al dejar presionado el DIP switch 2. La segunda modalidad consiste en que al accionar el DIP switch 2 se realiza el efecto wah-wah similar a como si se presionará un pedal (por tiempo limitado), al dejar de presionar el pedal (DIP switch 2) se inhibe el efecto.

Reverberación

Se elaboró el sistema de reflexiones tardías, para lo cual se requirió diseñar los filtros de peine (que actúan como reverberadores individuales), eligiendo sus atenuaciones y retardos, la variación de estos parámetros simulan la superficie de reflexión del sonido, así como el tiempo de recorrido de la fuente a las superficies y su retorno, de tal forma que se simula una gran sala de concierto. Se utilizaron en total cuatro filtros de peine cuyas características se muestran en la tabla 1. Debido a que la tasa de muestreo (f_s), seleccionada fue de 24 KHz, se tiene que el periodo de muestreo (T_s), es de 41.667 μ s. Con esta información es posible calcular el tiempo de retardo entre cada reflexión correspondiente a los reverberadores tras haber establecido el número de muestras de retardo, ver tabla 1.

Tabla 1 Especificaciones de los reverberadores.

Reverberador	Atenuación	Muestras de retardo	Tiempo de retardo [ms]
1	0.8	2311	96.29
2	0.31	4800	200
3	0.15	6307	262.79
4	0.03	8101	337.54

Los valores mostrados en la tabla 1, fueron elegidos de manera heurística, a criterio de los autores, debido a que la percepción de los efectos de sonido es subjetiva, y con ellos se lograr percibir el efecto adecuadamente, cabe destacar, que en los dispositivos comerciales la configuración de retardos y atenuaciones es resguardada mediante patentes o la no divulgación de dichos parámetros. Los filtros peine diseñados se implementaron en paralelo, es decir, la señal de entrada de audio ($x[n]$) es la entrada a cada uno de los reverberadores, los cuales realizan

el retardo y atenuación de la señal como se indica en la tabla 1, la salida de los cuatro filtros peine se suman, éste resultado es afectado por un filtro pasa todo, con el objetivo de obtener una difusión del sonido de los reverberadores, tratando de obtener un efecto lo más natural posible.

El diagrama a bloques del modelo diseñado se observa en la figura 4, y que constituye el sistema de reflexiones tardías basado en cuatro etapas de reverberación en cascada con un filtro pasa todos. En él, se observan los parámetros elegidos tanto para los retardos como las ganancias involucradas, mostradas en la tabla 1.

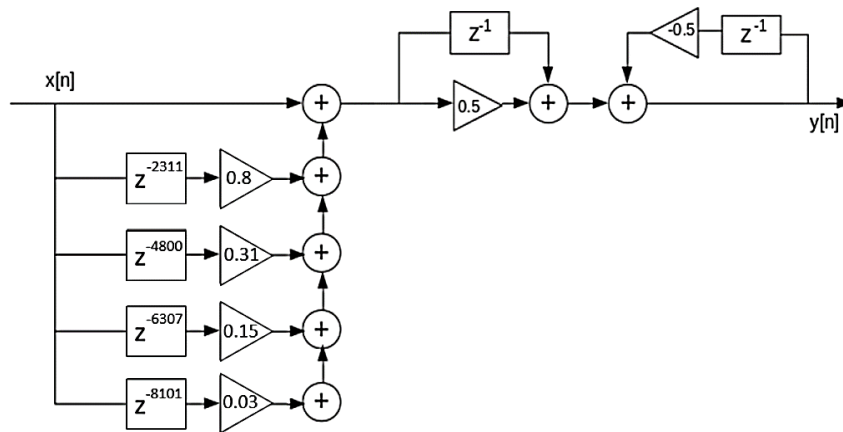


Figura 4 Diagrama del reverberador implementado.

Del análisis del diagrama anterior se obtiene que la función de transferencia (FT) de la etapa de los filtros peine en paralelo, ecuación 7.

$$H_1(z) = 1 + 0.8z^{-2311} + 0.31z^{-4800} + 0.15z^{-6307} + 0.03z^{-8101} \quad (7)$$

La FT de la segunda etapa (filtro pasa todos), ecuación 8.

$$H_2(z) = \frac{z^{-1}-0.5}{1-0.5z^{-1}} \quad (8)$$

Una vez obtenidas las FT's de ambas etapas, se puede obtener la FT de todo el sistema en el dominio de z mostrada en la ecuación 9.

$$H(z) = \frac{-0.5z^{-1}-0.4z^{-2311}+0.8z^{-2312}-0.155z^{-4800}+0.31z^{-4801}-0.075z^{-6307}}{1-0.5z^{-1}} + \dots \frac{0.15z^{-6308}ah-0.015z^{-8101}+0.03z^{-8102}}{1-0.5z^{-1}} \quad (9)$$

3. Resultados

En la figura 5, se muestran los componentes físicos del trabajo elaborado y su conexión, así como la ventana de ejecución de CCS con el código del proyecto en modo *run* (ejecución) en el monitor de la computadora. El equipo de cómputo utilizado fue una *workstation* Dell T5400, con 4 Gb de memoria RAM, y un procesador Intel Xeon a 2.0 Ghz. Los resultados de este trabajo se corroboraron de forma audible mediante la ejecución de diversas notas y acordes en guitarra, además de probarlo con la reproducción de canciones.



Figura 5 Conexión de los elementos del proyecto.

En el caso de los filtros pasa bajas y pasa altas no existió mayor complicación en identificar su efecto sobre el sonido, por lo que se confirma que la respuesta del DSP tuvo el comportamiento esperado. El uso este DSP brindó la posibilidad de realizar un procesamiento de señal de audio suficiente para poder generar adecuadamente los efectos propuestos. Al oprimir el switch 0 se mantuvieron los sonidos graves al tener una frecuencia de corte superior de 3 kHz ($\pi/4$). En el caso de seleccionar el switch 1 se escucharon los sonidos agudos debido a las configuraciones del filtro pasa altas. Cabe recordar que el diseño de ambos filtros está elaborado bajo la consideración de que la frecuencia de muestreo configurada en el códec AIC23 fue de 24 kHz, por lo que, si se eligiera una frecuencia de muestreo diferente, tendrían que ser rediseñados los filtros para cumplir las expectativas del usuario.

En el caso del efecto wah-wah existen muchas variantes dependiendo los ajustes realizados ya que puede ser un sonido corto o prolongado; sin embargo, para los

parámetros establecidos en este desarrollo se comprobó la audición de la articulación “wah” con una duración corta, al tocar diferentes notas en la guitarra. Cabe señalar que no se generó distorsión alguna de la señal de salida.

Finalmente, para el reverberador fue posible percibir las reflexiones de sonido para lo cual se probaron distintos ajustes en los coeficientes de retardo y ganancia, obteniendo así distintos resultados. Con la variación de estos parámetros fue posible recrear una sensación de habitación con reflexión o bien sonidos para los que podía distinguirse una reverberación que creaba la sensación de que las notas ejecutadas en la guitarra tuvieran una duración mayor a la original.

4. Discusión

Si bien se lograron implementar los efectos de audio propuestos, se notó que la tarjeta C6713 DSK tiene ciertas limitaciones como lo es la atenuación que sufre la señal de audio debido al códec, ya que el circuito de entrada contiene un divisor de voltaje que disminuye la amplitud en un factor de 2 como se indica en [DSP, 2017]. Esto se compensó añadiendo ganancia a la señal de salida, pero resulta insuficiente el volumen obtenido; además, si se excede en el valor de ganancia el audio comienza a presentar saturación. Por otro lado, el efecto wah-wah fue generado con éxito en sus dos modalidades, duración corta o prolongada, sin distorsión o saturación de sonido. Finalmente, los parámetros elegidos en el caso del reverberador no generan distorsión en el audio con guitarra, aunque en la música sí debido a que contiene un mayor rango de frecuencias, lo que provoca una saturación del sonido debido a las reflexiones generadas.

5. Conclusiones

La apreciación de efectos de sonido es subjetiva y depende del “oído” de cada usuario, por ello existe una gran variedad de pedales. Una alternativa al uso de pedales analógicos de efectos, es la implementación de filtros y efectos de sonido en señales de audio en tiempo real digitalmente, gracias a los DSP's. Las características propias de los DSP (arquitectura y recursos) son determinantes en el posible acondicionamiento a realizar en las señales de audio. En este trabajo,

se muestra que para el efecto de reverberación, se requieren diferentes retardos de tiempo aplicados a la señal de entrada. Lo anterior afecta la obtención de la señal de salida, provocando en algunos casos una distorsión, por lo que según el tipo de audio, será conveniente reajustar los parámetros involucrados en su diseño.

El efecto wah-wah y los filtros digitales básicos se realizan adecuadamente. El efecto wah-wah requiere tan sólo de almacenar una muestra anterior de las secuencias $y_1[n]$ ver ecuación 4, y $y_b[n]$ ver ecuación 5, de esta manera, las ecuaciones de diferencias que modelan dicho efecto, son ejecutadas sin la necesidad de aplicar grandes retardos de tiempo.

Existen una gran cantidad de efectos de audio que se pueden realizar a través de un DSP, la implementación de un sistema multiefecto puede brindar la posibilidad de satisfacer las preferencias de configuración de cada usuario con el objetivo de generar un efecto sonoro único y personalizado. Se propone como trabajo futuro complementar el sistema realizado, con la adición de otros efectos de sonido, en un solo script de ejecución, buscando desarrollar un sistema multiefecto.

6. Bibliografía y Referencias

- [1] A. Oppenheim, W. Mecklenbrauker & R. Mersereau, Variable cutoff linear phase digital filters, *IEEE Transactions on Circuits and Systems*. Vol. 23. No. 4, pp. 199-203, 1976.
- [2] Adil, Duaa. Nehad, Yazan, "Digital Signal Processing and Sound Effects". Department of Computer Engineering. Baghdad University, 2015.
- [3] Anuj Dharia & Rosham Gummattira, Signal Processing Examples Using the TMS320C67x Digital Signal Processing Library (DSPLIB): <http://www.ti.com/lit/an/spra947a/spra947a.pdf>. 2009, Abril 2017.
- [4] Ballou, Glen, Handbook for sound engineers, 4ta Edición, Taylor & Francis. Hoboken, N. J., pp. 1737, 2013.
- [5] Daniel, S., Multi-sound effect system including dynamic controller for an amplified guitar. US Patent 7,541,536: <https://www.google.com/patents/US7541536>, 2009.

- [6] Dattorro, Jon, Effect design, part 2: Delay line modulation and chorus. *Journal of the Audio Engineering Society*. Vol.45. No. 10, pp. 764-788, 1997.
- [7] Digital Guitar Effects Unit and Amplifier: <http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1190&context=eesp>, Marzo 2017.
- [8] DSP Audio Effects: https://courses.engr.illinois.edu/phys406/Student_Projects/Spring01/PPoongbunkor/Piya_Poongbunkor_DSP.pdf, Marzo 2017.
- [9] Glover, John C. et. al. Python for audio signal processing, 2011: <http://eprints.maynoothuniversity.ie/4115/1/40.pdf>, último acceso: Julio, 2017.
- [10] Moorer, James A., About this Reverberation Business, *Computer Music Journal*. Vol. 3. No. 2, pp.13-28, Junio 1979.
- [11] Pennock, J. D. Urry, R. M. et. al., Musical effect customization system. US Patent 7,026,539, 2006: <https://www.google.com/patents/US7026539>.
- [12] Programmable Guitar Pedals: <http://diydsp.com/livesite/pages/GuitarPedals>, Abril 2017.
- [13] Ryle, M. Doidic, M.A. Audio signal processor with modular user interface and processing functionality. US Patent 7,711,442, 2010: <https://www.google.com/patents/US7711442>.
- [14] Steven A. Tretter, Communication System Design Using DSP Algorithms with Laboratory Experiments for the TMS320C6713™ DSK. University of Maryland. Springer, pp. 1-25, 2008.
- [15] Siddhesh N. Upasani et. al. Review on Implementation of Digital Music Equalization (Echo & Reverberation) Model Using Simulink and TMS320C6713 DSK, 2014: www.ijcstjournal.org/volume-2/issue-2/IJCST-V212P26.pdf, Abril 2017.
- [16] Thad B. Welch, Cameron H. G. Wright, et. al., Real-Time Digital Signal Processing from MATLAB to C with the TMS320C67x DSPs. CRC Press. 28, 2012.

- [17] Vargas, Juan S. Burgos, Jaime A., Análisis de sistemas procesadores de señales de guitarras eléctricas Estudio de las señales en efectores de guitarra eléctrica orientado a tecnologías de la actuación, *Ingenium Revista de la facultad de ingeniería*, Vol. 17, No. 34, pp. 76-89, 2016.
- [18] Zölzer, Udo. Smith III, Julius O., DAFX-digital audio effects, *The Journal of the Acoustical Society of America*. Vol. 114. No. 5, pp. 2527-2528, 2003.

DISEÑO DE UN CIRCUITO DE CONTROL DE ILUMINACIÓN PARA UN SISTEMA FORMADOR DE IMÁGENES DE PURKINJE

Armando Gómez Vieyra

Universidad Autónoma Metropolitana
agvte@correo.azc.uam.mx

Ezequiel Martínez Solís

Universidad Autónoma Metropolitana
ezequiel.martinez.s@hotmail.com

Juan Jesús Ocampo Hidalgo

Universidad Autónoma Metropolitana
jjoh@correo.azc.uam.mx

Karla Beatriz Vergara Vázquez

Universidad Autónoma Metropolitana
karlavergara_b@outlook.com

Uriel Calderón Uribe

Universidad de Guanajuato
urielcal91@gmail.com

Geovanni Hernández Gómez

Universidad de Guanajuato
geov.hernandez@ugto.mx

Resumen

Las imágenes de Purkinje son generadas por reflexión de la luz en las diferentes interfaces oculares (córnea anterior y posterior, y cristalino anterior y posterior). El estudio de estas imágenes es de suma importancia tanto en la

optometría como en la oftalmología. En la UAM Azcapotzalco se ha desarrollado un sistema formador de imágenes de Purkinje, el cual permite generar y detectar dichas imágenes de individuos in-vivo. El sistema optoelectrónico presentado, consta de una iluminación infrarroja a 840nm, alimentada por un convertidor digital/analógico (current-steering DAC) con salida de corriente, el cual fue diseñado con transistores y multiplexores analógicos CMOS. El DAC entrega de 6.5 a 7.5 mA de corriente por LED en pasos de 0.25 mA (n bits de resolución), en las diferentes matrices. De este modo, las matrices funcionan apropiadamente sin saturarse, y por ende es posible generar las Imágenes de Purkinje sin reflexión y sin ruido de fondo, con lo cual se ha reducido el intervalo de la prueba (con tiempos de 10 a 15 minutos por sujeto).

Palabras Claves: DAC de corriente, iluminación LED, imágenes de Purkinje, sistema ocular humano, tecnología CMOS.

Abstract

Purkinje images are generated by the light reflection at different ocular interfaces (anterior and posterior cornea, anterior and posterior lens). The study of these images is of paramount importance in both optometry and ophthalmology. In UAM-Azcapotzalco, a Purkinje image-forming system has been developed, which allows the generation and detection of such images of in-vivo. The optoelectronic system presented, consists of an infrared illumination at 840nm, powered by a current-steering DAC with current output, which was designed with analogue CMOS transistors and multiplexers. The DAC delivers from 6.5 to 7.5 mA of current by LED in steps of 0.25 mA (n bits of resolution), in the different matrices. In this way, the matrices function properly without saturation, and therefore it is possible to generate the Purkinje Images without reflection and without background noise, which has reduced the interval of the test (with times of 10 to 15 minutes per subject).

Keywords: Current DAC, human eye system CMOS technology, LED lighting, Purkinje images.

1. Introducción

Las Imágenes de Purkinje fueron observadas por vez primera por Johannes Evangelista Purkinje (Fisiólogo) en 1821, con la ayuda de una linterna y más tarde confirmadas por Luis Joseph Sanson en 1837 [Marquez, 1926]. Las imágenes son cuatro, ver figura 1, dos corneales y otras dos del cristalino.

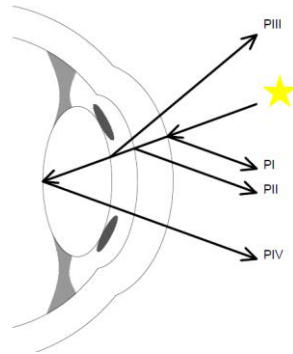


Figura 1 Esquema de la formación de las imágenes de Purkinje.

La primera imagen (PI) se produce por reflexión sobre la superficie anterior de la córnea que actúa como un espejo convexo y da lugar a una imagen virtual. Esta imagen es más intensa debido a la gran diferencia de índices de refracción entre el aire y las células de la córnea. Dicha imagen está situada a nivel del plano pupilar y tiene un tamaño intermedio entre las imágenes que producen las 2 superficies del cristalino. La imagen en cuestión se conoce también como reflejo luminoso corneal. La segunda imagen (PII) se produce sobre la superficie posterior de la córnea. Es de escasa intensidad debido a que la diferencia entre los índices de refracción de la córnea y el humor acuoso es muy pequeña. Además, está localizada muy cerca de la PI, ya que el radio de curvatura de la cara posterior es ligeramente inferior al de la cara anterior. Todo ello hace que esta imagen pase desapercibida ya que su tamaño también es ligeramente menor. La tercera imagen de Purkinje (PIII) se forma sobre la superficie anterior del cristalino que también actúa como un espejo convexo. Es la mayor de todas en cuanto a tamaño debido a que el radio de curvatura de la cara anterior del cristalino es mayor que los de la córnea, pero su intensidad luminosa es la más débil debido a varios factores como son: el mayor tamaño de la imagen, la escasa

diferencia entre el índice de refracción del humor acuoso y el del cristalino, una superficie menos lisa que la de la córnea y la conformación del cristalino con diferentes índices de refracción. Durante el estímulo visual conocido como acomodación, esta imagen se hace más pequeña, ya que disminuye el radio de curvatura, es decir, aumenta la curvatura de la cara anterior del cristalino. La cuarta imagen (PIV), se forma por reflexión sobre la superficie posterior del cristalino, que a diferencia de las otras actúa como un espejo cóncavo, por lo que se produce una imagen real e invertida. Su intensidad es mayor que la de la tercera imagen, pero su tamaño es menor que el de la primera (superficie anterior de la córnea) [Kaschke, 2014].

Recientemente, Taberero y coautores [Taberero, 2006] desarrollaron un sistema para estudiar la alineación de las estructuras oculares a partir de las Imágenes de Purkinje. Basado en esta investigación, se ha desarrollado un instrumento que permite obtener las imágenes de Purkinje, en una sola toma [Escamilla, 2014], [Cosme, 2016]. Este instrumento consta de una cámara CCD conectada a una PC, con lo cual realiza la captura y almacenamiento de las imágenes, empleando un módulo desarrollado bajo la plataforma LabView. Como se mencionó anteriormente, las Imágenes de Purkinje son cuatro (PI, PII, PIII y PIV), que se generan cuando se hace incidir luz infrarroja (840 nm) desde la superficie corneal, hasta el cristalino, del ojo humano. El uso de diodos emisores de luz (LEDs) infrarrojos conectados en serie a un transistor BJT que estabiliza la corriente consumida presenta muchas desventajas, que incrementan el grado de dificultad para generar las imágenes de interés. Las desventajas que tiene este arreglo son: El arreglo de diodos no irradia la intensidad luminosa apropiada: Dadas las características I-V no-lineales de los LEDs y los transistores BJT, ligeros aumentos en el voltaje provocan aumentos muy drásticos en la corriente, lo cual vuelve muy crítico el ajuste. En consecuencia, la generación de las imágenes de Purkinje se vuelve muy tardada.

Las imágenes captadas con dicho sistema de control de la corriente suministrada al LED, se observan con mucha reflexión y ruido, esto hace que se pierda nitidez en los reflejos de interés. No tener uniformidad en la iluminación hace que la

intensidad del haz de luz incidente en la superficie corneal se aprecie entre sí en algunas partes con más potencia que en otras, como consecuencia de esto, al localizar las coordenadas del centro de cada imagen, en la mayoría de los casos están fuera del eje céntrico, esto se observa en la figura 2 [Cosme, 2016].

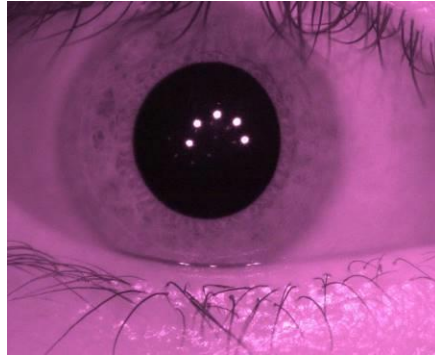


Figura 2 Reflejos de Purkinje en la superficie de interés.

En este trabajo se presenta el diseño y la implementación de un sistema electrónico, que permite lograr una iluminación uniforme y controlada para este instrumento.

2. Métodos

Se procedió a caracterizar 10 LEDs infrarrojos, tanto óptica como electrónicamente, para definir los parámetros necesarios del sistema electrónico. Empleamos una esfera integradora de 6" Edmund Optics, un Power Meter Marca Newport Modelo 1916-R, una fuente de alimentación controlada digitalmente, un multímetro Keithley DMM5.5 y un multímetro AMPROBE Compact 15XP-B. Las curvas características encontradas se pueden observar en la figura 3, donde se corroboró que tenían propiedades similares y que son acordes a la literatura.

El sistema completo a bloques que se desarrolló se muestra en la figura 4. Donde se observa un bloque del sistema de alimentación general, que es una fuente de voltaje de corriente directa variable de los 3 a los 15 V, basada en un regulador LM317 y monitoreada por un microcontrolador ATmega328 que muestra el voltaje de salida en una pantalla de cristal líquido. Un bloque de fuente de corriente, el cual basa su construcción en un espejo de corriente simple con transistores

MOSFET. El bloque de control digital será construido a partir de la implementación de multiplexores analógicos (con tecnología MOSFET), que podrán ser controlados manualmente o mediante una computadora personal o un microcontrolador. Este sistema permite trabajar de manera manual o implementar un sistema automatizado para análisis de las estructuras oculares.

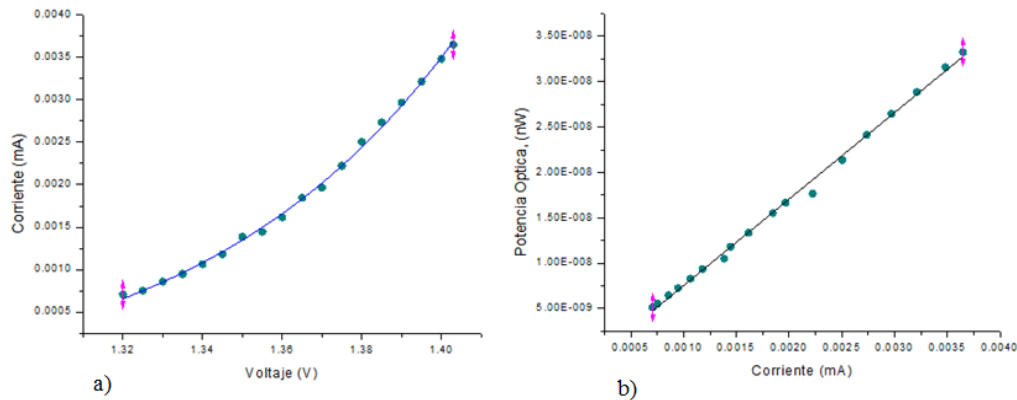


Figura 3 a) Curva de corriente contra el voltaje. b) Potencia óptica contra la corriente.

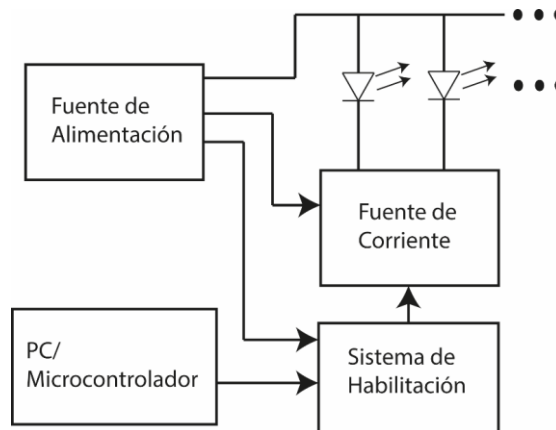


Figura 4 Sistema a bloques para el control de iluminación.

Tomando en cuenta a los dispositivos disponibles en el mercado y a las características de éstos, se decidió basar todo el diseño en tecnología CMOS ya que son más estables a la temperatura que la tecnología BJT, presentan una elevada impedancia de entrada y no son sensibles a la radiación externa. Sin embargo, el manejo y ensamblaje se complica debido a que se dañan con la electricidad estática.

La Fuente de Corriente

Existen diversas configuraciones para crear este subsistema, por lo que se eligió como punto de partida, una fuente de corriente de dos transistores, llamada también espejo de corriente, que se ilustra en la figura 5 [Neamen, 2010]. Donde se observa una corriente de referencia (I_{REF}), que entra al primer transistor que se conecta como un diodo. El voltaje en este transistor activa el segundo transistor. Lo que provoca que las corrientes de ambas terminales fuente (I_{S1} e I_{S2}) son prácticamente iguales, debido a que las compuertas están en paralelo. Con lo anterior se logra que la corriente en la terminal de drenaje (I_{D2}) sea igual a la corriente de referencia (I_{REF}).

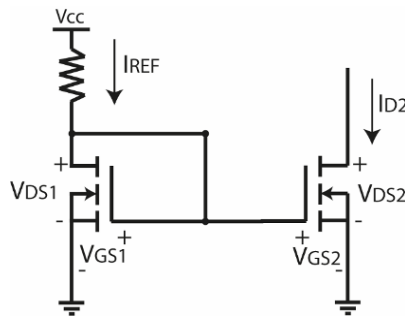


Figura 5 Fuente básica de corriente básica MOSFET de dos transistores.

Por lo que la resistencia, ajustaría la corriente de referencia, y podría proporcionar los pasos adecuados para incrementar la corriente mediante la incorporación de salidas múltiples, dejando una corriente base (con dos o más salidas activas) y el resto controlados para ir incrementando la corriente en pasos constantes, ver figura 6.

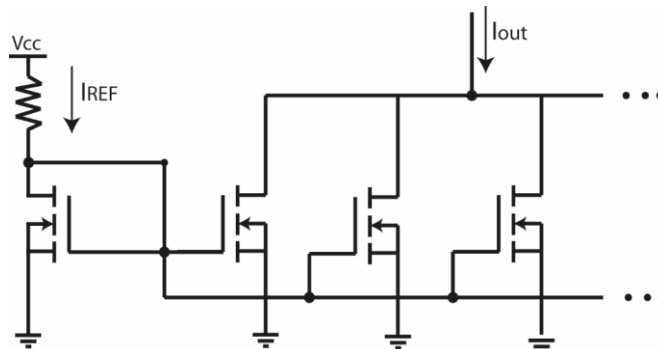


Figura 6 Fuente de corriente de transistores de salidas múltiples.

El Sistema de Habilitación

Este sistema es diseñado empleando multiplexores analógicos, que pueden construirse a partir de interruptores analógicos CMOS. Idealmente, los interruptores, presentan una resistencia nula cuando están cerrados y una impedancia infinita cuando están abiertos. Al implementarse el multiplexor analógico, se emplean interruptores bilaterales CD4016BE. El diagrama a bloques del sistema a implementar se muestra en la figura 7.

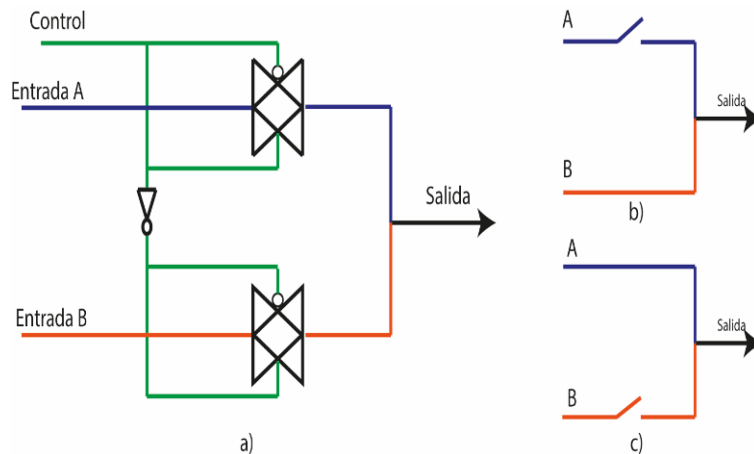


Figura 7 Formación de un Multiplexor Analógico a partir de interruptores bilaterales.

En el bloque SWA y SWB están en posibles salidas lógicas: 0 y 1, la señal de corriente que reciben las compuertas lógicas están presentes en los pines de control A y control B; al presentarse un estado cero lógico en el SWA la señal circundante deja de pasar, dirigiéndose solamente en el bloque SWB con estado uno lógico, al final del recorrido se tiene para éste una salida S igual a B, lo anterior puede ilustrarse por medio del circuito equivalente de la figura 7b.

Integración del Sistema Completo

Para la integración final del sistema, se decidió implementar de forma manual el control del sistema de habilitación, sin embargo, es claro que puede habilitarse fácilmente desde un microcontrolador o una computadora personal. Para la etapa de la fuente de corriente se emplearon transistores MOSFET 2N7000 y un potenciómetro de 20 vueltas de valor nominal de 100 k Ω .

Por la complejidad del sistema, se muestra únicamente una célula acotada del sistema, figura 8, que consta de un DAC de corriente de 2 bits.

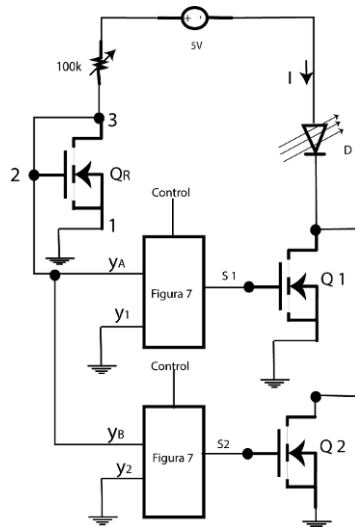


Figura 8 DAC de corriente de 2 bits implementado para controlar “I” en “D1”.

La figura 9 nos muestra el ensamble final para el control de un solo LED. El cual nos permite variar la intensidad de este, a partir de un interruptor miniatura presente en el arreglo, con intensidades bien definidas y bajo muestreos definidos. En este caso se realizó un DAC de corriente de 4 bits.

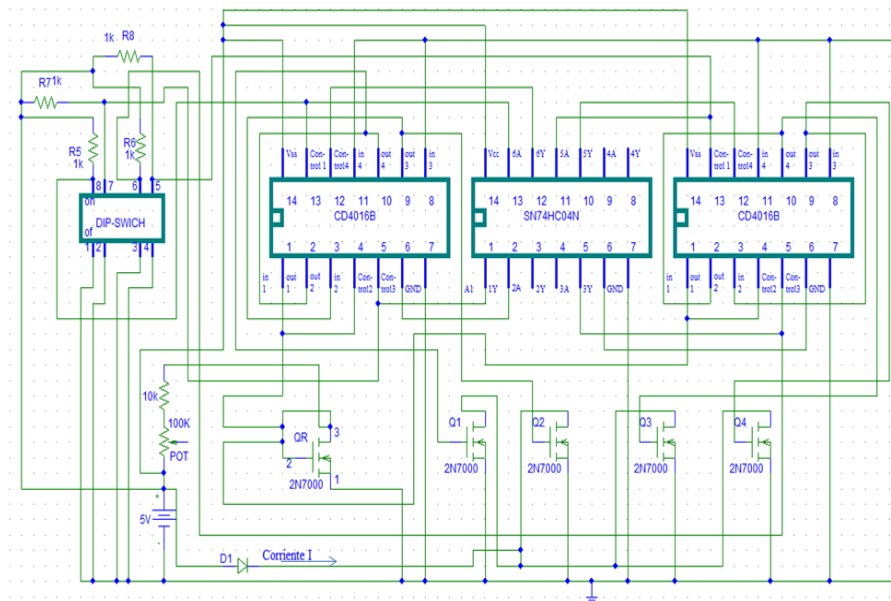


Figura 9 Ensamble final para el control de un solo LED infrarrojo.

Para completar un arreglo de tres, cuatro o más sistemas de iluminación LED, únicamente se tienen que reproducir los ensambles necesarios, mostrados en la figura 9. Cada ensamble comparte el mismo transistor de referencia para gobernar la corriente en el drenaje (salida) de cada uno de los transistores. Y con eso garantizar la misma emisión. Las compuertas lógicas empleadas fueron de tecnología CMOS, que en este caso fueron las SN-74HC04N. La fuente de alimentación estuvo regulada a 5 V.

Acondicionamiento de los LED

Los diodos infrarrojos para el sistema de iluminación que brindó el fabricante, son como el que se observa en la figura 10a. Este diodo por la forma geométrica que tiene el lente (esférico) presentaba problemas, pues el encapsulado epóxido (lente del LED) hace que la luz de éste se dirija aleatoriamente. Este hecho fue un inconveniente, ya que para este trabajo es necesario que la luz del LED esté lo menos dispersa posible (haz de luz puntual). Para resolver este inconveniente, cada LED se sometió a la eliminación del lente; para hacer esto, se sometió cada pieza a un esmerilado abrasivo, para posteriormente pulirlos.

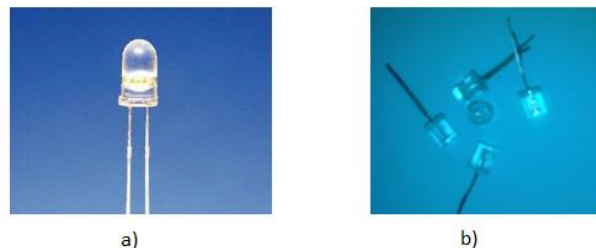


Figura 10 Diodo LED: adquirido con el fabricante y después del proceso de pulido.

El esmerilado abrasivo consistió en devastar el lente del LED a través de una lija de agua del número 1600. El número de la lija se refiere a la porosidad (rugosidad) o a la capacidad para esmerilar. La número 1600 es colocada sobre una superficie plana y sobre de ésta se hace verticalmente el desbaste del LED. La superficie a esmerilar debe quedar horizontalmente de forma plana, sin ningún ángulo de inclinación. El proceso finaliza cuando el esmerilado esta 1.5 mm de la parte interna del LED, ver figura 10b.

El material que se usó en el proceso de pulido fue: Óxido de Silicio de $0.5 \mu\text{m}$ de granulo, una charola de plástico (25 por 40 cm), un despachador de agua (una botella de refresco con un agujero) y un cuadro de Poliuretano para pulido (10 cm de lado). El proceso de pulido comienza en donde finalizó la etapa de esmerilado. El pulido finaliza cuando la superficie ha adquirido la misma apariencia que tiene el resto del LED.

3. Resultados

El arreglo final donde se integró el sistema desarrollado se muestra esquemáticamente en la figura 11. Antes de la realización de este trabajo se obtenían imágenes, como las mostradas en la figura 12.

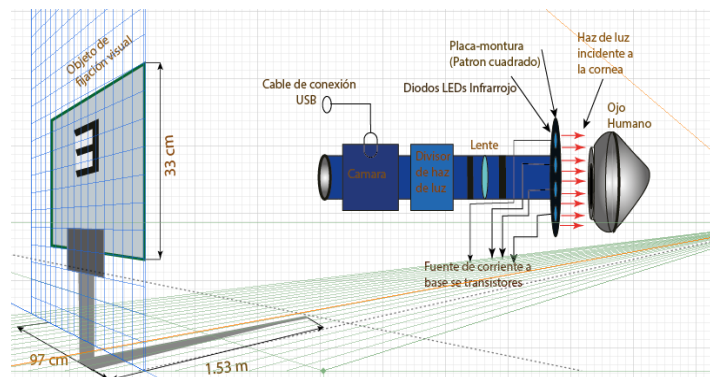


Figura 11 Configuración del Sistema Formador de Reflejos de Purkinje.

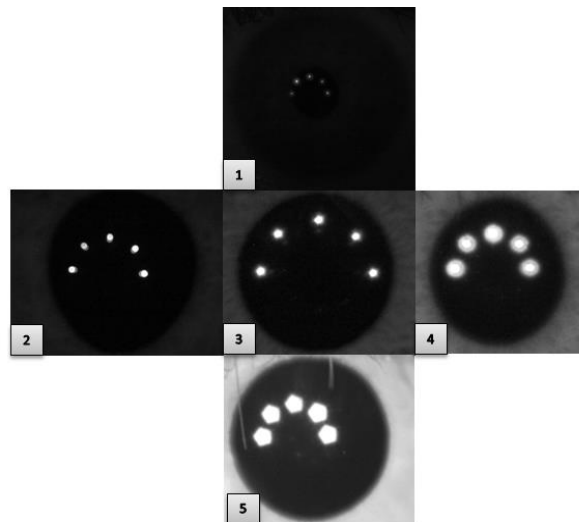


Figura 12 Imágenes típicas obtenidas sin el desarrollo del control de iluminación.

En la figura 13, se muestra las imágenes obtenidas a partir de la implementación de las fuentes de corriente. Debe tenerse presente que se ha cambiado la geometría del patrón de iluminación. Pero los avances cualitativos de este trabajo no pueden evidenciarse cuantitativamente, por ejemplo, el tiempo que tardamos en la toma de datos se reduce a 15 minutos, aun con sujetos que no han usado el instrumento, la comodidad para el individuo de no sufrir un cansancio, etc.

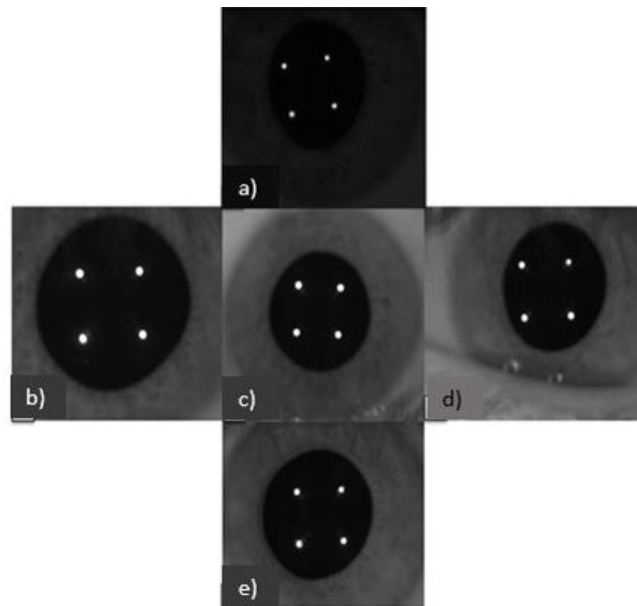


Figura 13 Imágenes para cinco sujetos, implementación arreglo de fuentes de corriente.

La figura 13a: fue captada a 1.92 mA. La figura 13b corresponde al sistema de iluminación alimentado a 2.23 mA. La figura. 13c, generada a 3.5 mA. La figura 13d tomada con una alimentación con 5.5 mA. La figura 13e generada con el sistema de captación de reflejos de Purkinje funcionando con 6.5 mA. Todos los sujetos de prueba son distintos, con edades que oscilan de los 22 años a los 30 años. Se respetan las normas ANSI para la máxima irradiación permitida a un individuo, de hecho, se trabaja con al menos 100 veces por debajo de la norma.

4. Discusión

La integración en este trabajo del Sistema de Formador de Imágenes de Purkinje ha beneficiado en varios aspectos tanto subjetivos como objetivos. Es

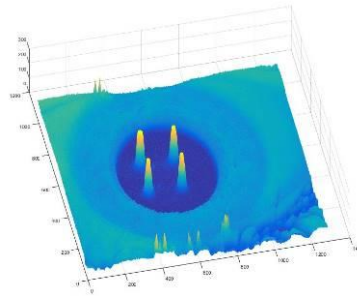
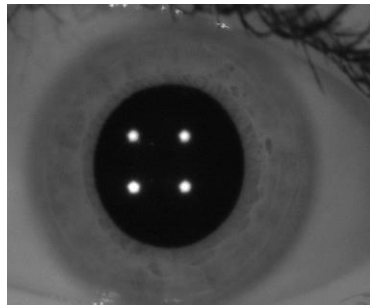
claro que la disminución del tiempo requerido para obtener imágenes bien definidas de las reflexiones intraoculares beneficia tanto al paciente como al usuario del equipo. El control de la emisión de luz a partir de la fuente de corriente, demostró que es posible obtener resultados más homogéneos, con menores reflexiones, aun en condiciones de maquillaje, facilitando el análisis de resultados. La modificación de la forma esférica de los LEDs, permite que los diodos emitan de forma casi puntual. La idea central de la electrónica empleada, fue considerar el umbral de polarización de los diodos LEDs (entre los 5 a los 40 mA). El umbral apropiado se obtuvo con la construcción y diseño de la fuente de corriente. Una muestra de los resultados tanto empíricos como los niveles de detección se presentan en la figura 14, donde se evaluaron sujetos en un rango de 22 a 33 años, hombres y mujeres.

Es recomendable hacer la adquisición de las Imágenes de Purkinje entre las 8:00 y 10:00 horas de la mañana, ya que, si los sujetos de prueba se someten a horas más avanzadas del día, es muy probable que las imágenes de interés no sean adecuadas, porque se estaría experimentando con sujetos con la vista cansada, en consecuencia, adelgazamiento de la lágrima y su actividad neuronal limitaría su capacidad de enfoque visual.

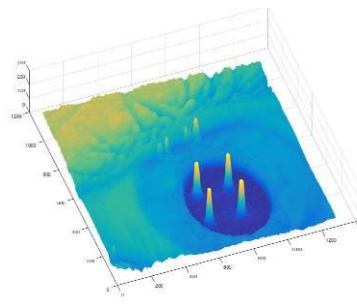
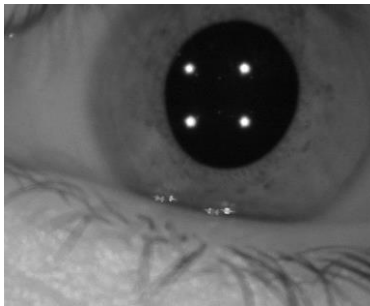
5. Conclusiones

En este trabajo se diseñó e implementó un sistema electrónico que permite controlar la iluminación infrarroja de una matriz de LEDs. Esto a partir de una fuente de corriente en cascada, cuyo sistema de habilitación fue construido con multiplexores analógicos y que tiene la versatilidad de poder ser controlado mediante un sistema automático o de forma manual. Con este sistema se tiene una iluminación adecuada para realizar diversos experimentos en el Sistema Generador de Imágenes de Purkinje. Los resultados se ven reflejados en la no saturación y la homogeneidad de las Imágenes de Purkinje (principalmente la PI y PIV), logrando una confiabilidad y eficiencia en nuestros experimentos (ver figura 14). Los alcances de este sistema pueden aplicarse en el diseño y construcción de los sistemas de iluminación de cámaras de fondo de ojo, donde se está

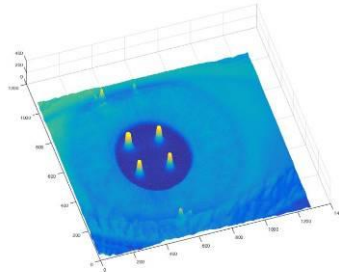
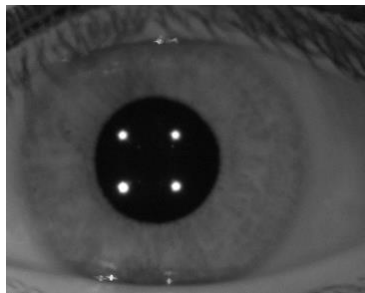
experimentando con diodos de 1 y 3 W, que tienen diferentes rangos en sus curvas características.



a) Sujeto masculino 33 años.



b) Sujeto masculino 23 años.



c) Sujeto femenino 26 años con rímel.

Figura 14 Imágenes de Purkinje (principalmente la PI y PIV).

6. Bibliografía y Referencias

- [1] G. A. Escamilla Ruiz. A. Gómez Vieyra. Construcción de un medidor de desalineación de superficies intraoculares basado en imágenes de Purkinje. VII Congreso Internacional de Ingeniería Física, México, pp. 391-392, 28 de noviembre del 2014.

- [2] D. A. Neamen, *Microelectronics Circuit Analysis and Design*, 4th edition, Mc Graw-Hill. Singapore. pp. 687-706, 2010.
- [3] I.U. Cosme-Cisneros. G. A. Escamilla-Ruiz. D. Flores-Montoya. G. Hernández-Gómez. A. Gómez-Vieyra. Instrument for recording Purkinje images. Conference Proceedings of the Society for Experimental Mechanics Series: 5th International Symposium on Experimental Mechanics and the 9th Symposium on Optics in Industry. Springer International Publishing AG, Agosto del 2016.
- [4] J. Tabernero. A. Benito. V. Nourrit. P. Artal. Instrument for measuring the misalignments of ocular surfaces. *Optics Express*. Vol 14. N. 22, pp. 10945-10956, 30 de octubre del 2006.
- [5] M. Kaschke, K. H. Donnerhacke, M. S. Rill, *Optical Devices in Ophthalmology and Optometry*. 1st edition, WILEY-VCH. Germany, pp. 3-14 y pp. 81-87, 2014.
- [6] M. Marquez. Interpretación y valor clínico de las imágenes de Purkinje-Sanson. *Archivos de la Sociedad Oftalmológica Hispano-Americana* Vol. 26. N. 308, pp. 421-435, agosto 1926.

ANÁLISIS COMPARATIVO DE LOS TIEMPOS DE EJECUCIÓN SOBRE SBC PARA DOS SISTEMAS OPERATIVOS DE TIEMPO REAL

Diana Lizet González Baldovinos

Instituto Politécnico Nacional ESIME Culhuacan

glez_lizet@hotmail.com

Jose Luis Cano Rosas

Instituto Politécnico Nacional ESIME Culhuacan

lucskyr@gmail.com

Pedro Guevara López

Instituto Politécnico Nacional ESIME Culhuacan

pguevara@real-time.com.mx

Resumen

El análisis del desempeño de los Sistemas Operativos de Tiempo Real (SOTR) permite aplicarlos adecuadamente en diferentes ámbitos y desarrollar posibles aplicaciones de tiempo real de manera efectiva. En este sentido, el presente trabajo considera los tiempos de ejecución como parte del desempeño total del sistema operativo; para ello se plantea la comparativa de respuesta de los tiempos de ejecución generados al realizar una tarea de alta complejidad temporal, implementada en dos extensiones de Tiempo Real basadas en Linux, particularmente PREEMPT_RT y Xenomai. La tarea consiste en un algoritmo basado en inversión de matrices, mediante el método de Gauss-Jordan con dimensión de 2^n , en el rango desde 2×2 hasta 128×128 . La finalidad de este trabajo, es mostrar que los tiempos de ejecución, son parte del desempeño total de los Sistemas Operativos de Tiempo Real y que un mismo hardware puede tener diferente desempeño con diferentes extensiones de tiempo real, al manejar prioridad y políticas de planificación. Para los experimentos se utilizó una SBC (Single Board Computer) Raspberry Pi 3.

Palabras Claves: Complejidad Temporal, inversión de matrices, SBC, SOTR, tiempos de ejecución.

Abstract

The analysis of Real-time Operating Systems (RTOS) performance allows to be properly apply them in different environments and development real-time applications. In this sense, this paper consider the response of computing times as part of total performance of the operating system. So, we propose the comparison of execution times response generated by a task of high temporal complexity, implemented in two Real-time extensions based on Linux, PREEMPT_RT and Xenomai. The task is created by an algorithm based on Gauss-Jordan of matrix inversion method, with dimension of 2^n , in the range of 2×2 to 128×128 . The goal of this paper is to demonstrate that execution times, are part of the total performance of Real-time operating systems and the same hardware could have different performance with two real-time operating system extensions, modifying priority level and scheduling policies. For the experiments we used an SBC (single board computer) Raspberry Pi 3.

Keywords: Execution times, temporal complexity, matrix inversion, RTOS, SBC.

1. Introducción

Las computadoras embebidas o SBC como Raspberry Pi, pueden emplearse para el desarrollo de sistemas en tiempo real; Sin embargo, generalmente cuentan con un sistema operativo de tiempo compartido, donde su núcleo (kernel) puede adecuarse para dar soporte de tiempo real. Los Sistemas Operativos de Tiempo Real dan preferencia a los procesos del mundo físico sobre el usuario, brindan soporte para el manejo de tareas y sincronización, manejo de memoria, comunicación entre procesos, manejo de dispositivos, así como del reloj de la computadora. Existen algunas extensiones de Tiempo Real basadas en Linux, tales como: PREEMPT_RT y Xenomai. PREEMPT_RT es una extensión de tiempo real, que ejecuta al kernel de Linux como un thread (hilo de ejecución) de menos prioridad que las tareas de tiempo real. Con este diseño, las tareas de

tiempo real y los manejadores de interrupciones nunca se ven retrasados por operaciones que no son de tiempo real. Estas tareas y los manejadores, ejecutan cuando se necesitan en detrimento de lo que estuviera ejecutando Linux.

Xenomai es un proyecto de software libre desarrollado en 2001, que implementa un framework de tiempo real en plataformas Linux [Xenomai, 2017]. Por lo anterior, se puede afirmar que Xenomai no es un sistema operativo de tiempo real, en cambio, es una extensión que permite realizar tareas en tiempo real y cuya criticidad dependerá de la integridad del sistema en su totalidad y no únicamente de la extensión, es decir, que podrá trabajar con limitantes. Xenomai utiliza un kernel dual mediante una extensión al kernel nativo de Linux, mismo que se encarga de atender interrupciones y planificar tareas de tiempo real [Xenomai, 2017].

Los trabajos que han servido como punto de referencia para esta investigación, se mencionan a continuación: Del trabajo Doctoral [Valdez, 2015], se toma el estudio sobre los tiempos de ejecución, donde se menciona que c_k es el tiempo en que se procesa la información en el intervalo k hasta completarse el procesamiento, sin considerar los bloqueos por lectura o escritura en los canales de comunicación, desalojos del procesador u otro tipo de suspensiones. Se menciona, que el tiempo de respuesta del sistema en tiempo real, depende en gran parte del comportamiento de los tiempos de ejecución $c_{i,k}$ de cada instancia $j_{i,k}$ de cada tarea en tiempo real J_i , el cual varía debido a diversos factores. Éste trabajo también hace referencia a [Stappert, 2000] y [Bernat, 2003], donde los autores exponen que el comportamiento fluctuante de los tiempos de ejecución $c_{i,k}$, se debe a los factores computacionales como son: el caching, pipeline, la búsqueda de la ruta de ejecución considerando a la exclusión mutua, número de lazos, predicciones y otras interacciones.

De acuerdo al trabajo de [Brown, 2010], se presentó un sistema de prueba para evaluar el rendimiento de tres núcleos; el kernel base de Linux, el mismo kernel de Linux con parche PREEMPT_RT y el mismo kernel de Linux pero con el parche de Xenomai aplicado. Se midieron los tiempos para realizar dos tareas; La primera tarea consistió en alternar una salida de E/S de propósito general (GPIO) en un

período fijo. La segunda tarea consistió en responder a un pin GPIO de entrada cambiante haciendo que el valor de un pin de salida GPIO lo siga. Para esta tarea, en lugar de realizar sondeos, utilizaron una interrupción para notificar cuándo cambia la entrada GPIO. Todas estas pruebas fueron realizadas en el sistema embebido BeagleBoard, los autores exponen que el kernel con parche PREEMPT_RT, es el que tiene un mejor tiempo de respuesta cuando se alterna una salida de GPIO.

En el trabajo de [Parikh, 2013] se analiza un panorama fundamental de los parámetros que determinan el rendimiento de algunos Sistemas Operativos de Tiempo Real. Con esta investigación, se observó que la planificación es el parámetro que más influye en el desempeño de los SOTR. Además, el documento proporciona una descripción intuitiva de tres Sistemas operativos de Fuente abierta, los cuales son: RT-Linux, FreeRTOS y eCos. Los autores afirman que el planificador, es el parámetro más importante que rige el funcionamiento de todo el sistema, incluido el hardware. Los sistemas de tiempo real crítico, requieren interrupciones y latencias extremadamente bajas y para ello se necesita un sistema de prevención con un mecanismo eficiente de manejo de interrupciones. RT-Linux es el SOTR más prometedor de los tres, ya que proporciona amplias funcionalidades sin comprometer lo esencial de un SOTR.

2. Métodos

La parte principal en el desarrollo de este trabajo, es experimentar el efecto de los tiempos de ejecución en una SBC Raspberry Pi 3, con dos extensiones de tiempo real basadas en Linux. Para ello, se ha desarrollado e implementado un algoritmo programado en ANSI C denominado *matrices.h*, descrito como una biblioteca de funciones para la inversión de matrices basada en los fundamentos teórico-prácticos que caracterizan a la matriz y su inversión, empleando principalmente el método de Gauss-Jordan.

El algoritmo está elaborado con el propósito de generar carga computacional y se emplea para hacer la medición de los tiempos de ejecución. Dicho algoritmo, fue implementado en los dos sistemas operativos de tiempo real, en cada instancia se

realiza la inversión de la matriz con dimensión de 2^n , abarcando desde 2×2 hasta 128×128 . La medición de los tiempos de ejecución se realiza dentro del algoritmo, para efectuar las mediciones, existe una serie de funciones dentro de la biblioteca `time h`.

La función `clock()`, es una herramienta de los sistemas POSIX, también se encuentra la función `time()`, `ClockTime()` y `clock_gettime()`, siendo esta última la que se usa en esta investigación. La diferencia de esta última con la función `clock()`, es que la anterior trabaja con múltiplos enteros de “ticks” de reloj y registra valores en el orden de los microsegundos, en cambio con la función `clock_gettime()`, la resolución puede ser ajustada permitiendo obtener valores más precisos de los tiempos de ejecución, medidos en el orden de los nanosegundos [Cano, 2015]. En la figura 1 se muestra el fragmento de código donde se realiza la medición del tiempo de ejecución.

```
for (k=0; k<ITERACIONES; k++){
clock_gettime(CLOCK_REALTIME, &inicio);
matriz_inversa=inversa(matriz, M);
clock_gettime(CLOCK_REALTIME, &fin);
milisegundos_i=(inicio.tv_sec*1000.0)+((double)inicio.tv_nsec/1000000.0);
milisegundos_f= (fin.tv_sec*1000.0) + ((double)fin.tv_nsec/1000000.0);
ejecucion= milisegundos_f-milisegundos_i;
printf ("c(%d)= %f milisegundos\n", k, ejecucion);
sprintf(buffer, "\n%f\n", ejecucion);
write(des, buffer, sizeof(buffer));
}
```

Figura 1 Medición de los tiempos de ejecución para la inversión de una matriz.

Los tiempos de ejecución medidos se guardan en un archivo `.txt` para posteriormente ser graficados fuera de línea en el software Octave. En el análisis se consideran los primeros dos momentos de probabilidad, el primer momento de probabilidad es la media o valor esperado de una variable aleatoria y es denotada por μ . La función con la cual se obtiene, se expresa en la ecuación 1.

$$E\{x_i\} = \sum_{i=1}^n \frac{x_i}{n} = \mu \quad (1)$$

El segundo momento de probabilidad o varianza de una variable aleatoria, es una medida de la dispersión de sus valores alrededor de la media μ y se denota por σ^2 [Bernat, 2003]. La función con la cual se obtiene, se define mediante ecuación 2.

$$E\{x_i\}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left[\frac{1}{n} \sum_{i=1}^n x_i \right]^2 = \sigma^2 \quad (2)$$

Las métricas que se emplean para realizar la comparativa de los tiempos de ejecución en cada SOTR, consisten en el manejo de prioridades y manejo de política de planificación. Se hace uso del manejo de prioridad por amabilidad, empleando el comando nice, la cual consiste en establecer la preferencia de ejecución de un proceso, la escala de prioridad está estandarizada con los valores de -20 a 19 donde el negativo representa la más alta prioridad y el positivo indica baja prioridad. Para establecer la máxima preferencia de ejecución al proceso, se ingresa la siguiente línea de comando desde la terminal: `sudo nice -n -20 ./programa`. Al utilizar el comando nice se observa que los tiempos de ejecución disminuyen un poco, sin embargo pueden disminuir aún más. Por lo tanto se hace uso de la función `setpriority()` que se encuentra dentro de la librería `sys/resource.h`, la función `setpriority()` requiere tres argumentos `int setpriority(int which, id_t who, int value)` los procesos de destino se especifican mediante los valores de los argumentos `which` y `who`. El primer argumento puede ser uno de los siguientes valores: `PRIO_PROCESS`, `PRIO_PGRP` o `PRIO_USER`, el segundo argumento `who` debe interpretarse como un ID de proceso, un ID de grupo de proceso o un ID de usuario efectivo, respectivamente, finalmente el tercer argumento indica el valor de prioridad [Linux, 2017], en la figura 2 se muestra el uso de la función.

```
#include <sys/resource.h>
...
int which = PRIO_PROCESS;
id_t pid;
int priority = -20;
int ret;

who = getpid();
ret = setpriority(which, who, priority);
```

Figura 2 Uso de la Función `setpriority()`.

Por otro lado se emplea también la llamada al sistema `sched_setscheduler()` la cual se encuentra en la librería `sched.h`, ésta llamada establece tanto la política de planificación como los parámetros para el hilo cuyo ID se especifica en el `pid`, en la figura 3 se muestra el uso de la función.

```
#include <sched.h>

int sched_setscheduler(pid_t pid, int
policy, const struct sched_param *param);
```

Figura 3 Uso de la Función `sched_setscheduler()`.

En este trabajo se utilizaron dos políticas de planificación de tiempo real, éstas políticas son `SCHED_FIFO` una política de primero en entrar, primero en salir; y `SCHED_RR` una política round-robin, los valores de prioridad que manejan respectivamente van desde 1 a 99 donde 99 es la más alta prioridad [SCHED, 2017].

3. Resultados

En el siguiente apartado, se presenta un compendio de figuras donde se muestran los resultados experimentales de la medición de los tiempos de ejecución, al realizar inversión de matrices con manejo de prioridades en `PREEMPT RT` y `Xenomai`. Se expone la diferencia de tiempo de ejecución de la misma tarea, para 1000 instancias efectuadas en la misma SBC.

PREEMPT RT

Figuras 4 y 5 muestran el resultado de los tiempos de ejecución con prioridad *nice* para la inversión de matrices de diferentes dimensiones, en este caso se introdujo desde la terminal la línea de comando: `sudo nice -n -20 ./ejecución`. Puesto que el propósito de este trabajo es mostrar una comparativa de los tiempos de ejecución cuando se ejerce alta carga temporal, se han elegido las gráficas que muestran los tiempos al invertir matrices de dimensiones 64x64 y 128x128, con el

respectivo manejo de prioridades. En la figura 6 se observan los tiempos de ejecución al invertir matrices de 64×64 , el valor de prioridad asignado para `setpriority()` y `nice` es `-20`, y el valor para `sched_setscheduler()` de acuerdo a la política de planificación FIFO y Round Robin se estableció en `95`.

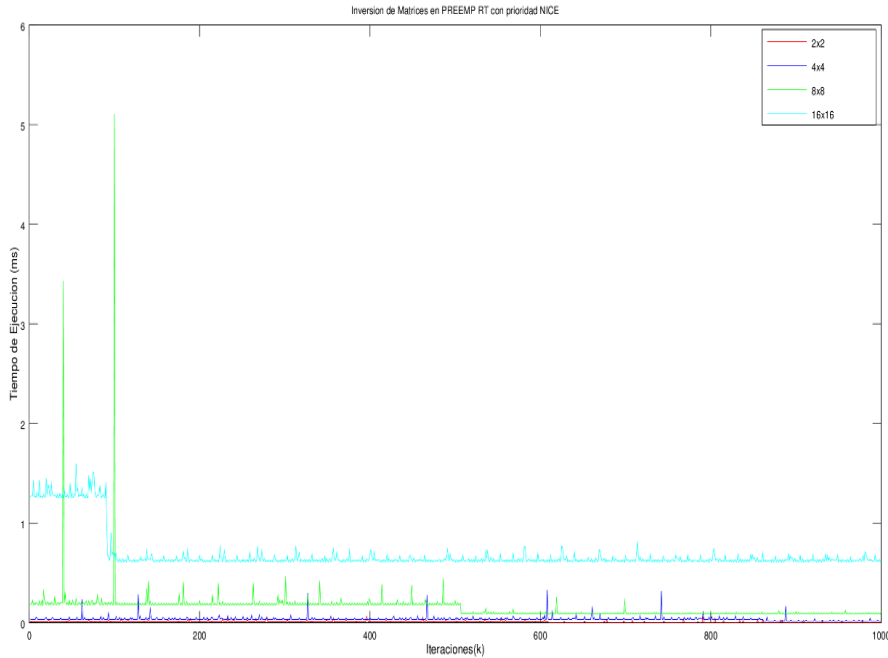


Figura 4 Inversión de matrices desde 2×2 hasta 16×16 , PREEMP_RT prioridad *nice*.

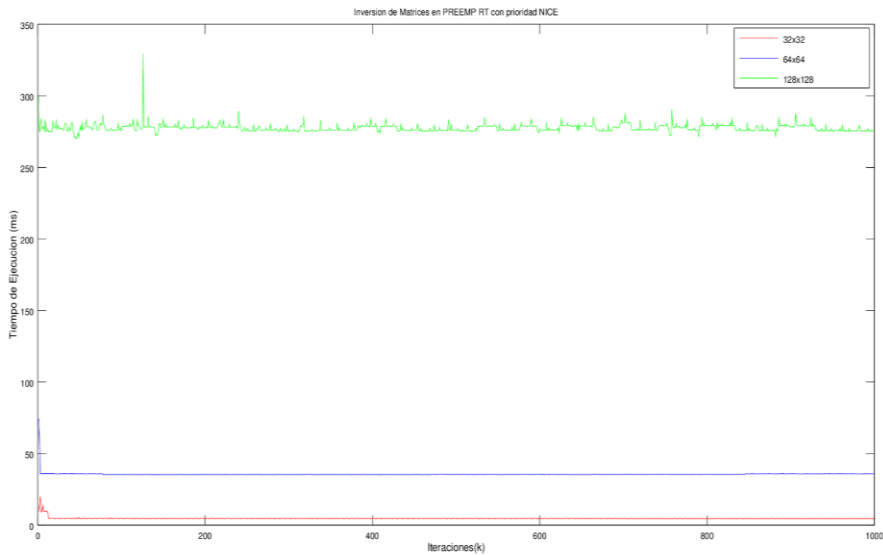


Figura 5 Inversión de matrices desde 32×32 hasta 128×128 , PREEMP RT prioridad *nice*.

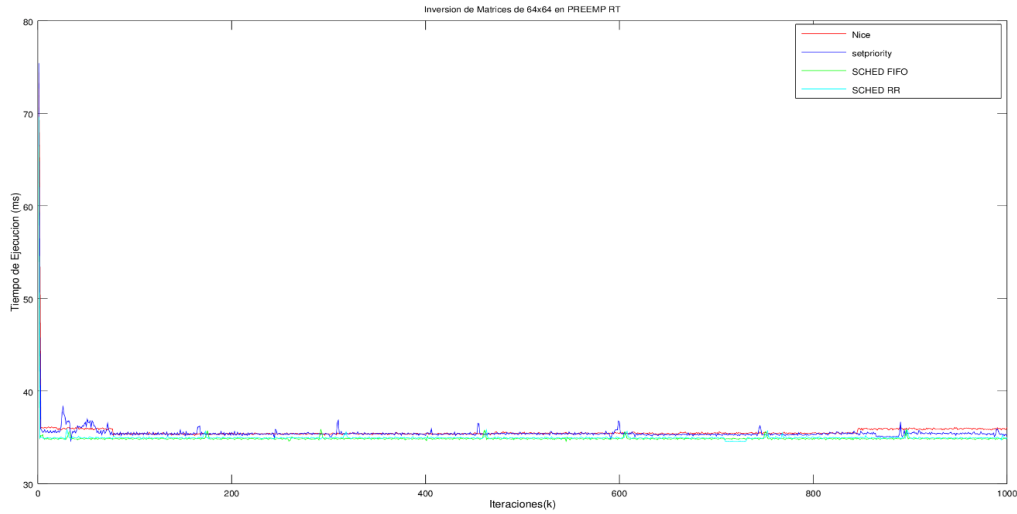


Figura 6 Tiempos de ejecución inversión de matrices de 64x64, manejo de prioridades.

En la figura 7 se muestran los tiempos de ejecución al invertir matrices de 128x128, con los valores de prioridad mencionados anteriormente.

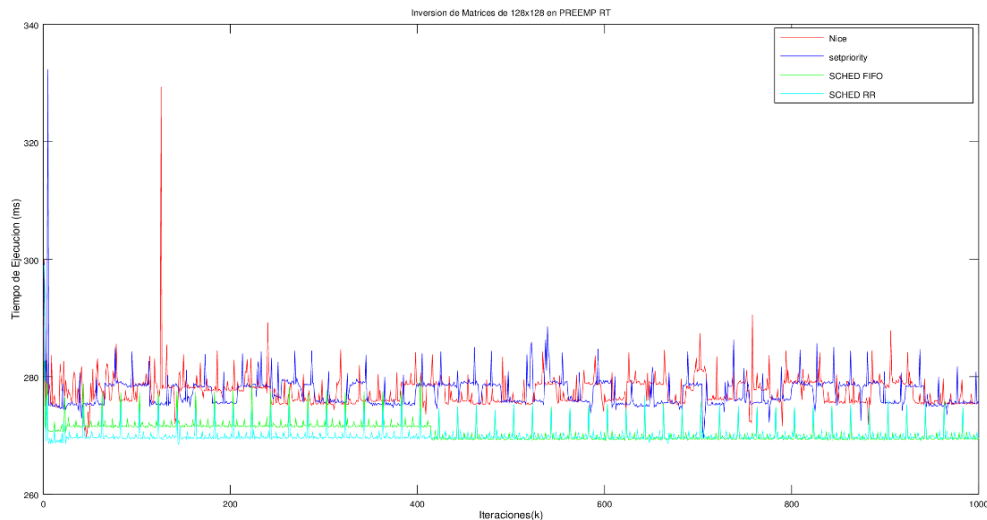


Figura 7 Tiempos de ejecución inversión de matrices de 128x128, manejo de prioridades.

Con esta extensión de tiempo real, se observó que el manejo de prioridad por planificación Round Robin presenta tiempos de ejecución menores a diferencia de los demás. En la figura 8 se presentan la media y varianza de los tiempos de ejecución de la inversión de matrices de 128x128 con prioridad por planificador Round Robin.

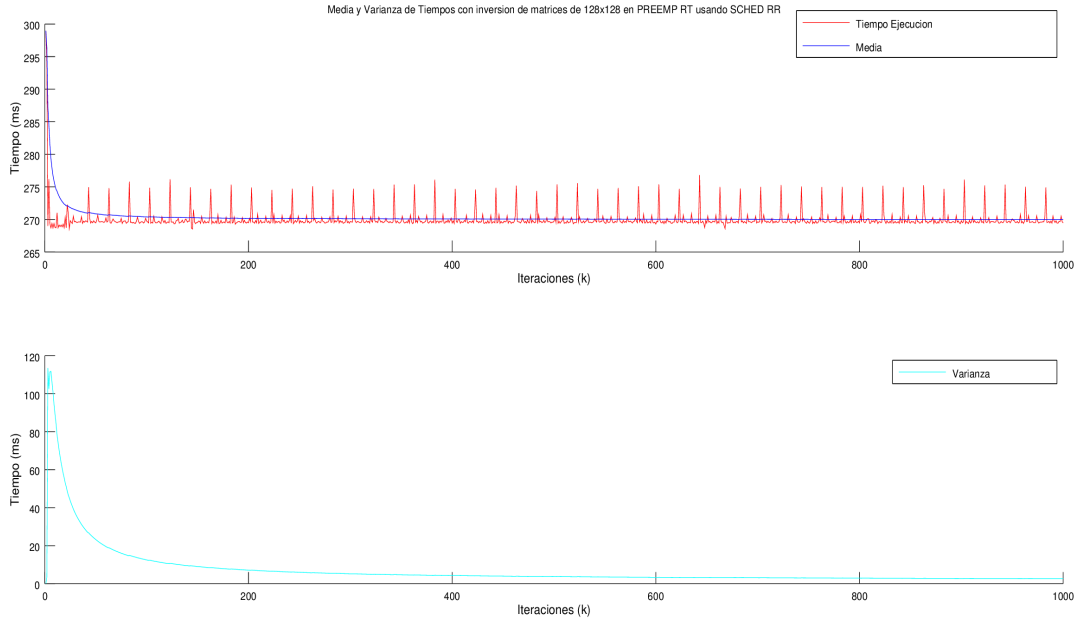


Figura 8 Media y Varianza de Tiempos e ejecución de inversión de matrices de 128x128, con prioridad Round Robin.

Xenomai Los tiempos de ejecución obtenidos, son considerablemente mayores a los que se obtuvieron en PREEMPT_RT, a continuación se presentan las gráficas resultantes al invertir matrices de dimensión 64x64 y 128x128. En figuras 9 y 10 se muestran los tiempos de ejecución, con los valores de prioridad nombrados inicialmente.

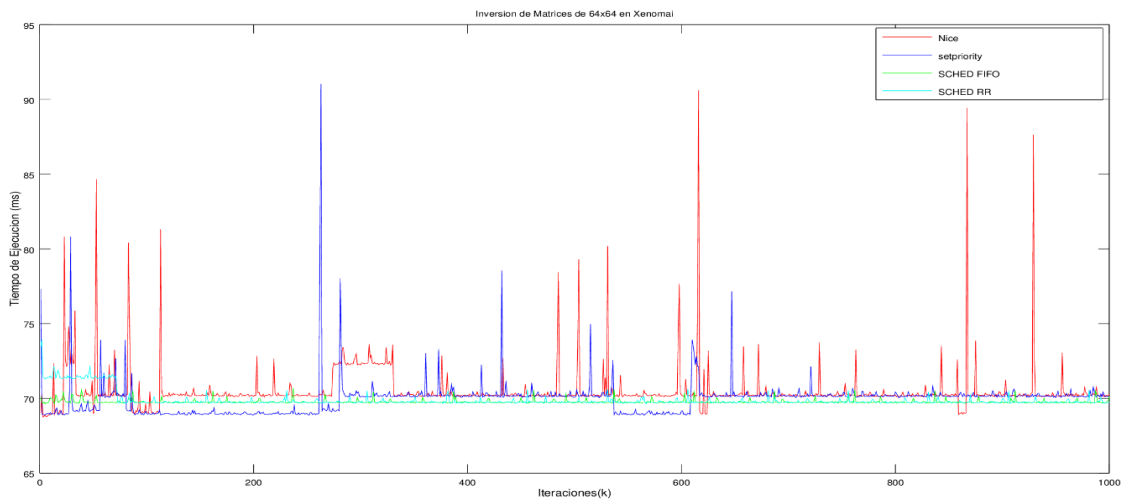


Figura 9 Tiempos de ejecución inversión de matrices de 64x64, manejo de prioridades.

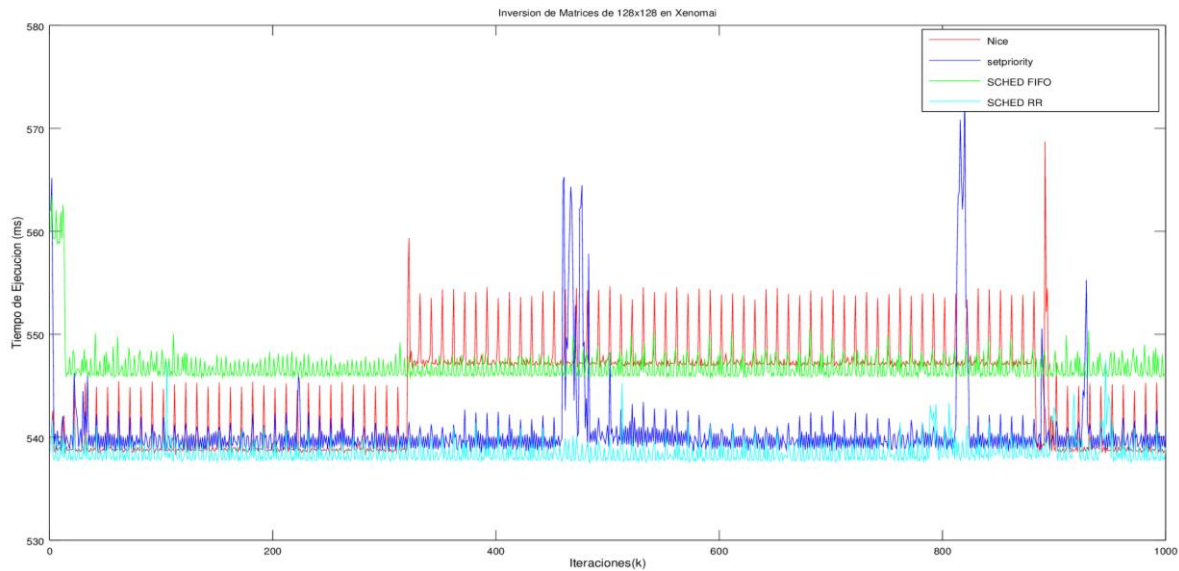


Figura 10 Tiempos de ejecución inversión de matrices de 128x128, manejo de prioridades.

De igual forma que en PREEMPT_RT el manejo de prioridad por planificación Round Robin presenta tiempos de ejecución menores a diferencia de los demás. En la figura 11 se presentan la media y varianza de los tiempos de ejecución de la inversión de matrices de 128x128 con prioridad por planificador Round Robin.

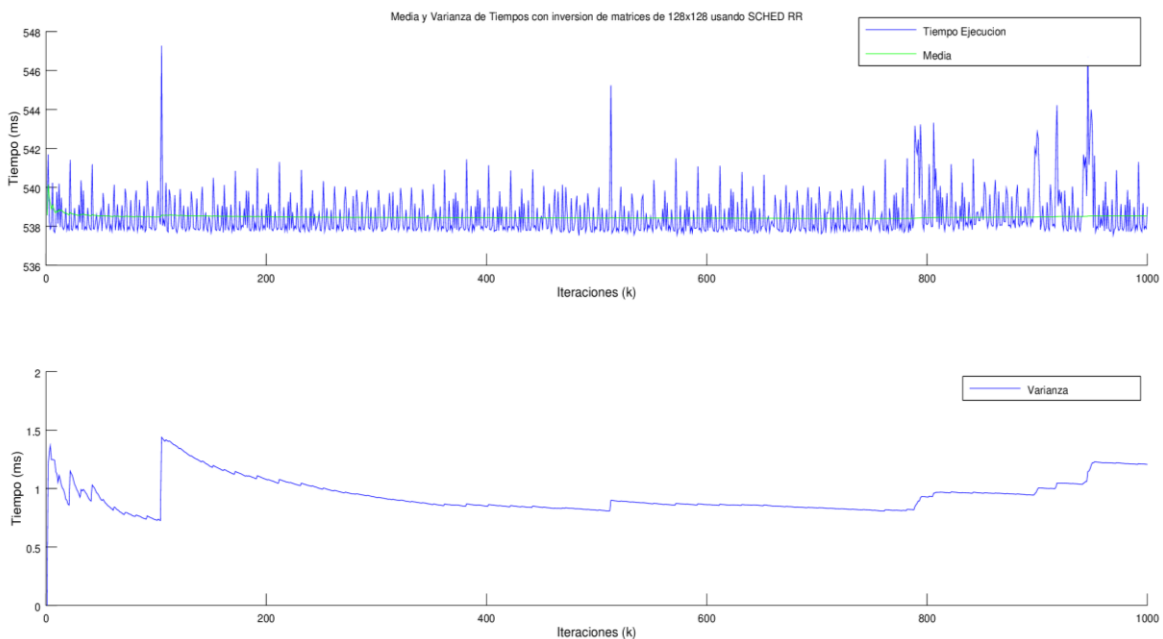


Figura 11 Media y Varianza de Tiempos de ejecución de inversión de matrices de 128x128, con prioridad por planificador Round Robin.

4. Discusión

El comportamiento de los tiempos de ejecución, de acuerdo al algoritmo de inversión de matrices con alta complejidad computacional temporal, realizado en los sistemas operativos de tiempo real basados en el kernel de Linux, tiene una diferencia sustancial en cuanto a tiempo de procesamiento, como se ilustra en la figura 12. Es posible comprobar que en PREEMPT_RT los tiempos de ejecución son menores, debido a que esta extensión trata al kernel de Linux como una tarea de menor prioridad y en contraste Xenomai trabaja con kernel dual. Por lo tanto, hace que los recursos se dividan para ambos y el tiempo de procesamiento sea mayor. También se observa que los tiempos de ejecución en PREEMPT_RT no son tan variables, esto otorga mayor estabilidad y fiabilidad al sistema cuando se aplica alta carga computacional temporal, a diferencia de Xenomai que muestra gran variabilidad en los tiempos.

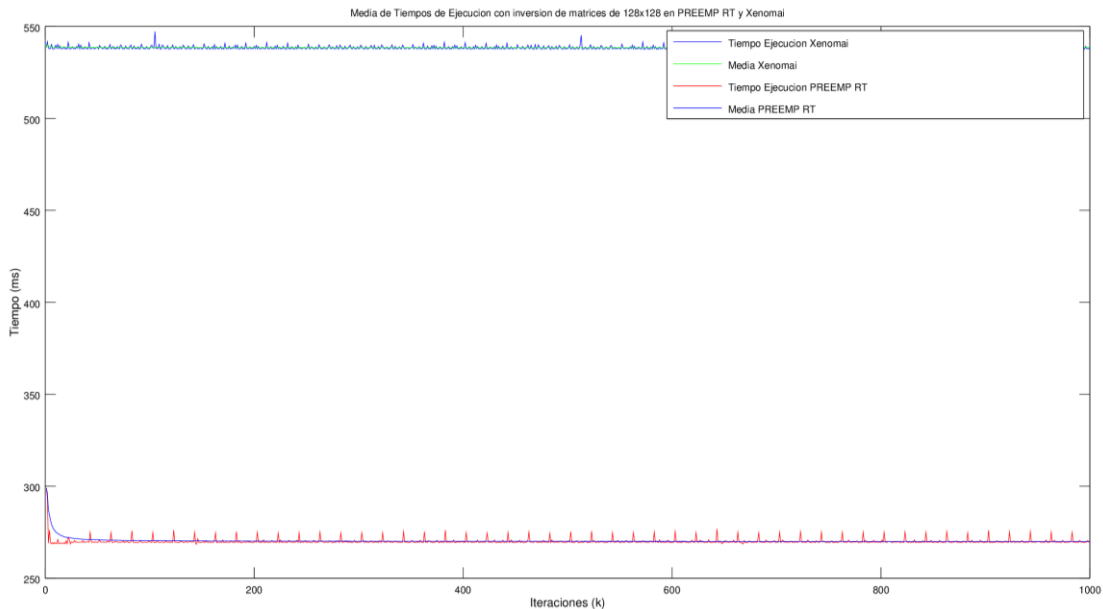


Figura 12 Media y Varianza de Tiempos de ejecución inversión de matrices de 128x128, en PREEMPT_RT y Xenomai con prioridad por planificador Round Robin.

5. Conclusiones

Las computadoras basadas en placa reducida (SBC) como Raspberry Pi, son una excelente alternativa para el desarrollo de sistemas de tiempo real embebidos;

algunas razones son su pequeño tamaño, bajo consumo energético y sobretodo el soporte para software libre; todo esto hace que el resultado final sea modular y de muy bajo costo. En este proyecto, se instaló en la SBC Raspberry Pi 3 el sistema operativo Raspbian, que es un derivado del sistema operativo GNU/Linux Debian con características de tiempo compartido. Para dar soporte de tiempo real se utilizaron PREEMPT_RT y XENOMAI, ambas extensiones se basan en arquitecturas y funcionamiento diferente; para analizar su comportamiento sobre los tiempos de ejecución se llevó a cabo este estudio experimental donde se manipularon prioridades, se midieron tiempos de ejecución y se realizó un análisis cuantitativo basado en los primeros momentos de probabilidad de las mediciones, utilizando como banco de pruebas un algoritmo de complejidad NP, como la inversión de matrices por el método de Gauss-Jordan. Los resultados obtenidos reflejaron que en PREEMPT_RT los tiempos de ejecución fueron menores debido a que esta extensión trata al kernel de Linux estándar como una tarea de menor prioridad, en contraste con Xenomai cuya ejecución fue de mayor demanda temporal. Pudo observarse que en PREEMPT_RT los tiempos tuvieron mayor estabilidad, es decir, su varianza no fue demasiado alta, lo cual determina mayor confiabilidad en ésta extensión, cuando se genera alta carga temporal.

6. Bibliografía y Referencias

- [1] Bernat G, Colin A, Petters SM, PWCET: A tool for probabilistic Worst-Case Execution Time Analysis of Real-Time Systems. Technical Report YCS-2003-353. Department of Computer Science, University of York, Reino Unido, York, pp. 1-18, 2003.
- [2] Brown, J. H., & Martin, B., How fast is fast enough Choosing between Xenomai and Linux for real-time applications. In proc. Of the 12th Real-Time Linux, 2010.
- [3] Cano Rosas J. L., Efecto Del Overclocking Sobre Los Tiempos De Ejecución Generados Por Inversión De Matrices En Una Computadora Embebida. Sección de Estudios de Posgrado e Investigación, ESIME Culhuacan IPN, 2015.

- [4] Linux, Linux Setpriority Man Page, 2017: <https://linux.die.net/man/3/setpriority>, May 2017.
- [5] Parikh, H., Shah, R., Shah, U., & Deshmukh, S., Performance parameters of RTOSs; comparison of open source RTOSs and benchmarking techniques. In *Advances in Technology and Engineering (ICATE)*, 2013 International Conference on IEEE, pp. 1-6, 2013.
- [6] Stappert F and Altenbernd P., Complete Worst-Case Execution Time Analysis of Straight-line Hard Real-Time Programs. *Journal of Systems Architecture* V. 46, I. 4, pp. 339–355, 2000.
- [7] SCHED SETSCHEDULER, Linux Programmer's Manual: http://man7.org/linux/manpages/man2/sched_setscheduler.2.html, May 2017.
- [8] Valdez, J. S. (2015). *Medición, Caracterización y Reconstrucción de los Tiempos de Ejecución y Tiempos de Transporte para Sistemas de Telecontrol en Tiempo Real*. Tesis Doctoral. Sección de Estudios de Posgrado e Investigación. ESIME Culhuacan. IPN.
- [9] Xenomai, About Xenomai, Xenomai.org, 2017: <https://xenomai.org/about-xenomai/>, May 2017.
- [10] Xenomai, Start Here, Xenomai.org, 2017: https://xenomai.org//start-here/#How_does_Xenomai_deliver_real-time, May 2017.

GUÍAS DE DISEÑO WEB PARA FACILITAR EL ACCESO A LA INFORMACIÓN DESDE TELÉFONOS INTELIGENTES

Beatriz A. González Beltrán

Universidad Autónoma Metropolitana

bgonzalez@correo.azc.uam.mx

Araceli Granados García

Universidad Nacional Autónoma de México

aragranadosg@comunidad.unam.mx

Resumen

Actualmente, los usuarios tienen mayor acceso a contenidos web desde sus teléfonos inteligentes y tabletas que desde sus computadoras de escritorio. Este hecho les permite estar comunicados y disponibles en cualquier momento y en cualquier lugar. Sin embargo, cuando el usuario navega en la web móvil, se enfrenta a problemas de visualización debido a que muchos de los contenidos web no están adaptados para consultarse en este tipo de dispositivos y aquellos contenidos que si lo están, inciden en problemas de usabilidad.

Tomando como base esta problemática, en el presente trabajo se plantea la necesidad de generar guías de diseño web que orienten la adaptación de contenidos usables para los usuarios de teléfonos inteligentes. Se propone la integración de la metodología del diseño centrado en el usuario así como el uso de las mejores prácticas de diseño web móvil propuestas por la W3C.

Palabras Claves: Diseño centrado en el usuario, diseño web móvil, guías de diseño web.

Abstract

Users have greater access to web content from smartphones and tablets, than from desktop computers. This allows them to be communicated and available at any time and at any moment. However, when the user browses the mobile web, he

experiences visualization problems because many of the web contents are not adapted to be accessed on this type of devices and even those mobile-adapted have usability problems.

To address this issue, this work aims to generate web design guidelines that support the adaptation of mobile web contents for the mobile phone user. We propose the integration of the user-centered design methodology as well as the use of the mobile web best practices proposed by the W3C.

Keywords: *Mobile web design, user-centered design, web design guidelines.*

1. Introducción

Actualmente las personas hacen mayor uso de los teléfonos inteligentes que de las computadoras de escritorio, por lo que el acceso a contenidos web desde estos dispositivos también se ha incrementado. La ventaja que el usuario encuentra al tener acceso a contenidos web desde teléfonos inteligentes es que la web móvil está disponible en cualquier momento y en cualquier situación que lo requiera. Sin embargo, el usuario aún se enfrenta a problemas de usabilidad debido a que muchos de los contenidos web no están adaptados para visualizarse en dispositivos móviles y aquellos que si lo están, inciden en problemas de navegación, de presentación, de sobrecarga de información y de legibilidad [Lynch, 2016].

Los desarrolladores de contenidos web pueden utilizar un documento llamado guías de diseño que les sirve de soporte en la construcción de sitios web usables. Apple posee una guía de diseño para el navegador Safari [Apple, 2016] cuyo objetivo es optimizar el diseño de sitios web para mejorar la experiencia de uso desde dispositivos móviles. En [Cremin et al., 2007] consideran varias etapas para la creación de sitios web móviles, desde la creación de una estrategia web móvil considerando sus ventajas y desventajas, hasta el uso de estándares para la web móvil. En [Häkkinen, 2006] se proponen guías de diseño en el desarrollo de aplicaciones móviles sensibles al contexto. En este trabajo abordan los problemas de usabilidad y su impacto en los diferentes aspectos del desarrollo de este tipo de aplicaciones. Por su parte, en [Baharuddin et al., 2013] proponen un modelo para

evaluar la usabilidad de las aplicaciones móviles en el cual consideran cuatro factores contextuales: el usuario, el ambiente, la tecnología y la actividad o tarea. En [Park et al., 2011] desarrollan unas guías de diseño para teléfonos móviles donde consideran tres factores: principios generales de usabilidad, los componentes de interfaz de usuario de los teléfonos móviles y los atributos de un elemento de la interfaz. En [Shitkova et al., 2015] proponen unas guías de usabilidad para el desarrollo de aplicaciones o sitios web móviles. Este trabajo divide las guías en cinco tópicos: disposición (*layout*), navegación, diseño, contenido y desempeño. En [Nayebi et al., 2012] se realiza una revisión de los métodos de evaluación de la usabilidad en interfaces de usuario móviles. El estudio reveló que aún no existen medidas estandarizadas ni métodos de evaluación de usabilidad estandarizados.

En esta investigación se propone que la metodología del diseño centrado en el usuario (DCU) se integre a las guías de diseño web móvil. La metodología DCU ha sido aplicada exitosamente en el diseño de nuevos productos, incluyendo el diseño de páginas web [Williams, 2009].

2. Métodos

En esta sección se describe el procedimiento que se llevó a cabo para construir las guías diseño web móvil. Los pasos que se llevaron a cabo fueron:

- Caracterización del contexto móvil (usuarios, dispositivos móviles y la web móvil).
- Análisis del diseño web móvil.
- Caracterización de la metodología del diseño centrado en el usuario.
- Generación de guías de diseño web móvil.
- Prueba de factibilidad de las guías de diseño propuestas.

Caracterización del Contexto Móvil

En primer lugar, se caracterizó el contexto móvil. En [Abowd et al., 1999] se define al contexto como “cualquier tipo de información que puede ser utilizada para caracterizar la situación de una entidad. Una entidad es una persona, lugar, u

objeto que es considerado importante en la interacción entre un usuario y una aplicación, incluyendo al usuario y a la aplicación”. A partir de esta definición de contexto, se definió al entorno móvil como el ambiente cambiante que se crea en torno al usuario de dispositivos móviles cuando accede a la web móvil en cualquier momento y en cualquier lugar. Los elementos que conformaron nuestro contexto móvil son los **usuarios móviles**, los **dispositivos móviles** y la **web móvil**.

Caracterización de los Usuarios Móviles

Consideramos a los usuarios móviles a las personas que hacen uso de dispositivos de mano tales como teléfonos celulares o teléfonos inteligentes, para estar comunicados en cualquier momento y en cualquier lugar. Los usuarios móviles utilizan sus dispositivos en diferentes lugares, cuando están caminando, cuando están viajando de un lugar a otro, etc., lo que ocasiona que tengan necesidades de información específicas y requieran de recursos específicos como conexiones inalámbricas para una comunicación móvil efectiva.

En [Love, 2005] se consideran cinco características esenciales de los usuarios móviles: la capacidad espacial, la personalidad, la memoria, la capacidad verbal y las experiencias previas:

- La **capacidad espacial** se refiere a la habilidad del usuario para lidiar con el espacio físico en el que se encuentra, así como para visualizar mentalmente las tareas a realizar. Esta característica la aplica el usuario móvil al hacer uso de su dispositivo en un espacio físico lleno de distracciones como objetos u otras personas que se cruzan en su camino.
- La **personalidad** define los patrones de conducta y los modos de pensamiento que determinan la adaptación de un individuo a su entorno. En el usuario móvil, la personalidad tiene un efecto en la percepción que tiene del sistema con el que está interactuando y le atribuye rasgos de personalidad con los que él mismo se identifica.
- La **memoria** se divide en memoria a largo plazo y memoria a corto plazo. El usuario móvil utiliza más la memoria a corto plazo, por lo que es importante reducir la complejidad de la información que se le presenta.

- La **capacidad verbal** se refiere a la habilidad para comprender palabras habladas o escritas. Esta habilidad se relaciona con el mejor desempeño del usuario móvil en sus tareas a realizar.
- La **experiencia previa** se refiere a la habilidad del usuario en el manejo de su dispositivo, por el uso cotidiano que le da.

Debido a que el entorno móvil es muy diferente al entorno que se da con el uso de las computadoras de escritorio, es importante investigar el efecto del entorno en la experiencia del usuario móvil. Cuando se hace uso de computadoras de escritorio se puede tener un cierto control de los aspectos del entorno como la ubicación de la computadora, la luz ambiental, el aislamiento del ruido, entre otros; sin embargo, en el entorno al que se enfrenta el usuario de un dispositivo móvil existen muchas distracciones que están fuera de su control, por ejemplo, el propio movimiento del usuario si es que va caminando o en algún transporte, la luz en exteriores, el ruido del lugar en que hace uso del dispositivo, etcétera. Esto determina la facilidad o complejidad que se tiene al usar estos dispositivos móviles para hablar por teléfono, para conectarse a la web, para tomar y ver fotografías, para escribir, para leer, etcétera.

En la tabla 1 se presentan algunas diferencias entre usuarios de computadoras de escritorio y usuarios de dispositivos móviles.

Las características y necesidades específicas de los usuarios móviles deben conocerse muy bien para poder diseñar contenidos web que sean usables y así mejorar su experiencia en la web móvil; por ejemplo, si se sabe que los usuarios móviles hacen mayor uso de la memoria a corto plazo, se deberán diseñar contenidos concretos, sólo con la información necesaria y con una navegación más simple.

En la presente investigación, se consideraron sólo algunas de las características del usuario móvil mostradas en la tabla 1, tales como: que usan un dispositivo móvil, que usan conexiones inalámbricas a Internet, que hacen un mayor uso de la memoria a corto plazo, que manejan contenidos web breves y que realizan muchas consultas de poco tiempo en su dispositivo móvil.

Tabla 1 Diferencias entre usuario fijo y usuario móvil.

Usuario fijo	Usuario móvil
Utiliza una computadora de escritorio	Utiliza un dispositivo móvil
Tiene una ubicación fija	Va de un lugar a otro
Utiliza Internet alámbrico	Utiliza Internet inalámbrico
Utiliza una computadora de escritorio	Utiliza un dispositivo móvil
Tiene una ubicación fija	Va de un lugar a otro
Utiliza Internet alámbrico	Utiliza Internet inalámbrico
Las condiciones físicas están controladas	Existen muchos distractores en el ambiente físico tales como el ruido, la luz, el movimiento, otras personas, etc.
Mayor uso de la memoria a largo plazo	Mayor uso de la memoria a corto plazo
Maneja contenidos web extensos	Maneja contenidos web breves
Realiza pocas consultas de largos periodos de tiempo	Realiza muchas consultas de poco tiempo

Caracterización de los Dispositivos Móviles

Los dispositivos móviles también conocidos como dispositivos de mano, computadoras de mano, *palmtop* o *handheld* son dispositivos de bolsillo que tienen una pantalla pequeña y un teclado miniatura o un sistema táctil para ingresar datos; ejemplo de ellos son los teléfonos inteligentes, los relojes inteligentes y las tabletas.

Tabla 2 Características de los teléfonos inteligentes.

Característica
Sistema operativo
Tamaño de pantalla
Resolución de pantalla
Profundidad de color
Procesador
Memoria interna
Almacenamiento
Batería
Tipo de pantalla
Orientación de pantalla
Modalidades de interacción (ingreso de datos y navegación)
Navegadores

Los primeros móviles se caracterizaron por el diseño y optimización de dispositivos móviles para la comunicación por voz; después se centraron además en la transmisión de datos a través de la conexión inalámbrica a Internet; posteriormente, se combinaron la comunicación por voz, datos y servicios

multimedia. Sin embargo, aunque los dispositivos móviles son cada vez más ubicuos, pequeños y un poco más inteligentes [Edmondson et al., 2014], podemos considerar que la generación de estos dispositivos se encuentra presente y está caracterizada por la combinación de los accesorios móviles tales como los sensores y otros dispositivos como los lentes de realidad virtual.

Esta investigación se centró en los teléfonos inteligentes. La tabla 2 nos muestra las características que se consideran en la evaluación de los teléfonos móviles. Cabe señalar que, aunque todas las particularidades enlistadas en esta tabla dependen de sus propiedades físicas, para el diseño de interfaces las características de diseño que nos interesó resaltar fueron: el tipo de conexión inalámbrica, la resolución de pantalla, la profundidad de color, el tipo de pantalla y los navegadores. Cada una de estas características servirá para definir las medidas de diseño a implementar.

Caracterización de la Web Móvil

El término web móvil se refiere a la conexión a Internet de manera inalámbrica, a través de un dispositivo móvil, lo que facilita el acceso a contenidos web actualizados en cualquier momento [O'Reilly, 2005]. Sin embargo, la web móvil también es la interacción entre aplicaciones web y dispositivos móviles. Esta definición coincide con los principios de la web móvil, ver tabla 3.

Tabla 3 Principios de la web móvil.

Principios
La web como plataforma
El aprovechamiento de la inteligencia colectiva
Los datos en bases de datos especializadas
El final del ciclo de la liberación del software
Los modelos de programación ligera
El software no está limitado a un solo dispositivo
Experiencias de usuario enriquecedoras

La W3C tiene una recomendación sobre las mejores prácticas para el desarrollo de la web móvil cuyo objetivo es ayudar a los desarrolladores a diseñar y publicar contenido web que funcione adecuadamente en dispositivos móviles [WWW,

2008]. Estas pautas de desarrollo se pueden resumir en diez puntos clave que se muestran en la tabla 4.

Tabla 4 Mejores prácticas de la web móvil.

Puntos clave
Diseñar para una web única
Confiar en los estándares web
Evitar los riesgos conocidos
Ser prudente con las limitaciones de los dispositivos móviles
Simplificar la navegación
Comprobar los gráficos y los colores
Diseñar sitios web concisos
Economizar el uso de la red
Facilitar la entrada de datos
Pensar en los usuarios móviles

Aunque los principios de la web móvil están acordes al desarrollo de aplicaciones web, desde el punto de vista del diseño, solo puede aplicarse el principio “experiencias de usuario enriquecedoras”, por lo que nos pareció fundamental aplicar solo las pautas que propone la W3C para estandarizar los procesos de diseño y desarrollo eficientes y puntuales.

Diseño Web Móvil

En [Ballard, 2007] se da una definición de diseño móvil separando estos dos conceptos. Por un lado, considera al diseño como el medio para comunicar un mensaje y, por otra parte, afirma que lo móvil hace referencia al usuario, no al dispositivo o a la aplicación que se desarrolle; por lo que sugiere una experiencia de usuario completamente diferente a la que se tiene en una computadora de escritorio. Considerando esta definición se puede decir que diseñar para la web móvil, no sólo es “miniaturizar” el contenido y mostrarlo en pantallas pequeñas; es adaptar los contenidos a un contexto móvil, para lo cual se deben de considerar las características de los usuarios, las capacidades de los dispositivos y las limitaciones técnicas de la web móvil.

Aunque [Ballard, 2007] menciona que para el desarrollo de una aplicación móvil y una de escritorio existen diferencias en la plataforma, el contexto de uso, el

dispositivo y la tecnología; sin embargo, los fundamentos de diseño web siguen siendo los mismos. Por lo tanto, consideramos que, de la misma manera que el lenguaje básico del diseño se adaptó al diseño web, ahora se adapta al diseño web móvil. La tabla 5 muestra las adaptaciones que consideramos para el diseño web móvil.

Tabla 5 Características del diseño web móvil.

Elemento	Descripción
Retícula	El uso de una retícula debe considerar el tamaño de pantalla del dispositivo. Evitar el uso excesivo de la barra de desplazamiento vertical.
Color	La profundidad de color mínima es de 16 bits. Se pueden utilizar los mismos colores para el diseño web móvil que para el diseño web convencional.
Tipografía	Considerar las tipografías sans-serif. Considerar las tipografías por omisión de los dispositivos móviles.
Gráficos	Recomendar el uso de gráficos de manera adecuada.
Animaciones	Minimizar el uso de animaciones.

El diseño Centrado en el Usuario

La metodología de diseño centrado en el usuario (DCU) es una metodología de diseño donde tanto los diseñadores como los usuarios colaboran activamente. Los usuarios dirigen el diseño y son la clave para determinar la usabilidad de un producto (e.g. sitio web). La metodología DCU está dividida en cuatro fases que funcionan de manera iterativa: análisis, diseño, construcción y evaluación (ver figura 1).

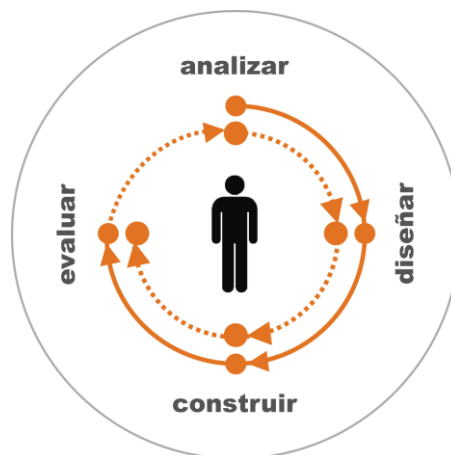


Figura 1 Metodología de diseño centrado en el usuario.

La fase de análisis permite estudiar el contexto de uso del usuario objetivo. La fase de diseño permite transformar los requisitos en propuestas de diseño. La fase de construcción permite la construcción del producto que coincide con los requerimientos. Por último, la fase de evaluación tiene como principal objetivo asegurar que el sistema desarrollado cumpla con las necesidades del usuario. La evaluación proporciona información que permite detectar deficiencias que haya que corregir [Williams, 2009].

Generación de las Guías de Diseño Web Móvil

La generación de las guías de diseño web móvil que se presentan en este trabajo se realizó con el apoyo de la caracterización del contexto móvil (usuarios, dispositivos móviles y la web móvil), del análisis del diseño web móvil y de la metodología DCU. En primer lugar, se incluyeron dentro de las guías de diseño los pasos necesarios para caracterizar a los usuarios, a los dispositivos y a la web. Esta caracterización forma parte del contexto de uso del usuario objetivo. Estos pasos forman parte de la etapa de análisis del DCU y permiten recoger la información sobre el usuario y sus circunstancias individuales como tareas, objetivos, entorno, etcétera. En segundo lugar, se integraron las mejores prácticas de la web móvil y las recomendaciones del diseño web móvil en la fase de diseño del DCU. En tercer lugar, se consideraron también las mejores prácticas de la web móvil para integrarse en la etapa de construcción del DCU. Por último, se propuso realizar pruebas de usabilidad, mismas que forman parte de la última etapa del diseño centrado en el usuario: la fase de evaluación.

Prueba de Factibilidad de las Guías de Diseño Propuestas

Se realizó una prueba de factibilidad de las guías de diseño propuestas a través de la realización, por parte de un diseñador de aplicaciones, de un caso de estudio de diseño de una aplicación móvil.

3. Resultados

Esta sección presenta en primer lugar las guías de diseño web móvil propuestas. En segundo lugar muestra una prueba de factibilidad de dichas guías.

Guías de Diseño Web Móvil

En la tabla 6 se muestran las guías de diseño web móvil propuestas e incluyen la etapa del diseño centrado en el usuario. Cabe señalar que las guías propuestas en las etapas de análisis del contexto de uso y de evaluación de la aplicación no existen en otras guías de diseño. Sin embargo, si existen las guías propuestas para la etapa del diseño y de construcción.

Tabla 6 Guías de diseño web móvil propuestas.

Etapa del diseño centrado en el usuario	Guía de diseño
Analizar el contexto de uso	Identificar las necesidades del usuario
	Conocer las características del dispositivo móvil
	Detectar el tipo de información a adaptar
Diseñar propuesta gráfica	Analizar la factibilidad tecnológica
	Evitar la sobrecarga de información
	Buscar simplicidad en el diseño de interfaces
	Favorecer el uso de gráficos
	Utilizar tipografía diseñada para pantalla
	Utilizar contrastes de color
	Evitar el uso de la barra de desplazamiento horizontal y vertical
Construir la aplicación	Simplificar la navegación
	Dar opciones al usuario
	Aplicar los estándares web
Evaluar la aplicación	Evitar los riesgos en el diseño
	Evitar los tamaños fijos
	Realizar pruebas de usabilidad

Analizar el Contexto de Uso

- Identificar las necesidades del usuario. Para empezar cualquier proyecto de adaptación de contenido web es necesario conocer primero las características de los usuarios móviles, los usos que le dan a la web y sus necesidades a cubrir.
- Conocer las características del dispositivo móvil. Es importante conocer las características del dispositivo móvil donde se presentará el contenido adaptado, pues dependiendo del tipo de conexión inalámbrica, del tipo de pantalla, resolución y profundidad de color que soporta; y de los navegadores en que se visualizan los contenidos, se considerará el tipo de

material (gráficos, videos, animaciones, etc.) que se puede adaptar y presentar.

- Detectar el tipo de información a adaptar. De acuerdo a la información obtenida de la guía #1, se puede detectar el tipo de información a adaptar que sea relevante para el usuario.

Diseñar la propuesta gráfica

- Analizar la factibilidad tecnológica. Una vez conociendo la información que se va a adaptar, y conociendo las características del dispositivo, se puede definir el tipo de aplicación a desarrollar, por lo que es importante seleccionar la tecnología a utilizar, por ejemplo, el lenguaje de programación, las características del servidor, cómo se va a solicitar la información, si se reconocen hojas de estilo, entre otros.
- Evitar la sobrecarga de información. La información que se presenta en una pantalla pequeña debe ser breve, clara y concisa, y deberá permitir al usuario identificar fácilmente las opciones de navegación que tiene.
- Buscar simplicidad en el diseño de interfaces. La simplicidad en el diseño de interfaces consiste en emplear elementos claros, sencillos y lógicos que le sean familiares al usuario para que pueda interactuar con ellos sin ninguna complicación.
- Favorecer el uso de gráficos. Considerando que los dispositivos móviles se utilizan para realizar consultas rápidas y que el usuario está haciendo otras actividades al mismo tiempo, se deberá favorecer el uso de gráficos e imágenes sobre el texto, pues esto permite identificar más rápidamente el tipo de información que se está presentando.
- Utilizar tipografía diseñada para pantalla. En cuanto al texto, es importante utilizar tipografías diseñadas para visualizarse en pantalla como la *Arial*, *Verdana* o *Tahoma*, que se caracterizan por ser tipografías *sans-serif*, con ojos más abiertos y con rasgos bien definidos para facilitar la lectura. Es importante mencionar que debido a que el usuario hace consultas rápidas y

breves en su dispositivo móvil debe limitarse al mínimo posible el número de caracteres que se presentan en cada pantalla.

- Utilizar contrastes de color. Para facilitar la visualización de los contenidos en entornos donde las condiciones de luz son variables (muy poca o muy intensa) es recomendable utilizar contrastes de color muy marcados para crear un énfasis visual. Esto se logra utilizando los colores complementarios: rojo-verde, amarillo-morado y azul-naranja, así como el contraste máximo blanco-negro.
- Evitar el uso de la barra de desplazamiento horizontal y vertical. Para facilitar la navegación dentro de la aplicación adaptada, se deberá evitar en lo posible el uso de la barra de desplazamiento vertical y horizontal, por lo que se deberá presentar en la primera pantalla la información más relevante y dar la opción de pasar a las siguientes pantallas sin usar la barra de desplazamiento vertical para ver más información.
- Simplificar la navegación. Se debe de considerar que se está haciendo uso de pantallas pequeñas y de un ancho de banda limitado, por lo tanto, se deberá organizar la información del sitio de tal manera que la navegación sea sencilla para que el usuario encuentre la información rápidamente pulsando no más de tres enlaces para que pueda cambiar de pantalla sin complicaciones.
- Dar opciones al usuario. A pesar de que se esté presentando al usuario una adaptación de la información, siempre se le deberá dar la opción de seleccionar la versión no adaptada, o de seleccionar algunas características de cómo quiere ver la información; por ejemplo, sin imágenes y sólo el texto o viceversa.

Construir la aplicación

- Aplicar los estándares web. Es importante aplicar los estándares web como el lenguaje extensible de marcado de hipertexto para el contenido, así como las hojas de estilo en cascada para la apariencia, ya que esto permite que se use menos código, se utilice menos ancho de banda y se muestre más

rápido el contenido, mejorando así la experiencia del usuario. Además, las aplicaciones desarrolladas son más fáciles de actualizar, son más accesibles y compatibles con diferentes navegadores.

- Evitar los riesgos en el diseño. Un riesgo en el diseño es el uso de elementos que no son estándares web y que dificultan el acceso a la información como son las ventanas emergentes, el uso de tablas, marcos, mapas de imagen, entre otros. Evitar estos riesgos en el diseño ayudará a reducir los problemas de usabilidad causados por pantallas más pequeñas.
- Evitar los tamaños fijos. El poner tamaños fijos en la tipografía puede limitar al usuario si quiere hacer más grande o más chica la letra de la información presentada. De igual manera, si se ponen tamaños fijos para el ancho de las columnas se pueden causar problemas en la presentación de la información cuando se tiene acceso a esta desde una pantalla de diferente tamaño.

Evaluar la aplicación

- Realizar pruebas de usabilidad. Utilizar métodos de evaluación de la usabilidad como evaluaciones al final de cada iteración. Por ejemplo, evaluaciones de desempeño, pruebas en condiciones controladas en un laboratorio de usabilidad, cuestionarios o entrevistas para comprobar que la propuesta de interfaz gráfica de usuario presentada cubre las necesidades de los usuarios, y si no es así, para que los mismos usuarios den las pautas sobre lo que se adapta mejor a sus necesidades.

Prueba de factibilidad de las guías de diseño propuestas

La prueba de factibilidad de las guías propuestas fue realizada de la siguiente manera. Se le pidió a un diseñador de aplicaciones que empleara las guías en una propuesta para la modificación de los resultados del buscador de Google para móviles. Para conocer el modo de utilización de las guías, pedimos al diseñador que describiera cada una de ellas, de acuerdo a los criterios utilizados en cada guía, mismas que se indican a continuación:

- Necesidades del usuario. Los usuarios objetivo son personas en el rango de 25 a 34 años, con licenciatura y pertenecen al nivel socioeconómico C+ (clase media alta); hacen uso de la web móvil varias veces al día para enviar y recibir correos electrónicos, así como para buscar información y leer noticias.
- Características del dispositivo móvil. Teléfono inteligente con sistema operativo Android, con acceso a la web móvil a través de una conexión inalámbrica de cuarta generación; cuenta con una pantalla táctil de alta resolución; soporta más de 65,000 colores.
- Tipo de información a adaptar. La modificación de los resultados de búsqueda del buscador de Google incluye al título, resumen y URL.
- Factibilidad tecnológica. Se propone utilizar la *API GoogleSearch* para desarrollar la aplicación web que recuperará los resultados de la búsqueda de *Google*. Esta aplicación se puede desarrollar en lenguaje *PHP* y se puede colocar en un servidor intermediario.
- Simplicidad en el diseño de interfaces. Se respeta el diseño actual de *Google* en colores, tipografía y estilo de gráficos. Se agrega la palabra Móvil al logotipo de *Google* para confirmar al usuario que está en la versión móvil. En la página de inicio del buscador, además del campo para la búsqueda se agregó la opción de seleccionar el país y el idioma. Se agregaron los enlaces a los servicios más solicitados de *Google* (figura 2).



Figura 2 Página de inicio de la propuesta para el buscador de Google.


En las páginas de los resultados, se cambió el logotipo completo de *Google* por su ícono, para dar mayor espacio a los resultados. Se utiliza una retícula de máximo dos columnas para dar más orden a los resultados obtenidos y para poder ver más resultados en la primera vista (figura 3). Se utiliza otra vista de resultados para dar opciones al usuario (figura 4).



Figura 3 Vista 1 para mostrar los resultados de búsqueda.



Figura 4 Vista 2 para mostrar los resultados de búsqueda.

- Sobrecarga de información. Se limita a 40 el número de caracteres para el título de la fuente de información. Se limita a 80 el número de caracteres para el resumen de la fuente de información. Se muestra solo el nombre del servidor de la fuente de información. Esto con el objetivo de mostrar el mayor número de resultados posibles en una pantalla con la información mínima necesaria que le ayude a decidir al usuario si entra a un sitio web o no, figura 3.
- Uso de gráficos.  Se propone que en cada resultado se presente el logotipo de la empresa o institución que proporciona la información, el cual se obtiene del ícono denominado favicon que despliega cada sitio web, figura 3. Este apoyo visual ayudará al usuario a decidir si le interesa obtener información de esa fuente o no, y agilizará su búsqueda de información.
- Tipografía. Se utilizará una tipografía Arial a 12 pixeles, adecuada para visualización en pantalla.
- Contrastes de color. Los colores básicos del diseño de *Google* ofrecen un contraste de color efectivo, ya que utiliza los colores primarios como el rojo y el azul, y los secundarios como el verde y el naranja para el logotipo; azul para los títulos de la fuente de información, verde para los *URL* y texto negro sobre fondo blanco para la descripción de los sitios.
- Uso de la barra de desplazamiento vertical y horizontal. En cada pantalla se despliega la mayor cantidad de resultados posibles: ocho para la vista condensada y cuatro para la vista extendida y en la parte inferior de cada pantalla se encuentran los enlaces para pasar de una pantalla a otra sin tener que hacer uso excesivo de la barra de desplazamiento vertical.
- Navegación simplificada. Cada pantalla presenta el campo de búsqueda, para que el usuario pueda iniciar una nueva búsqueda, además cada pantalla tiene una barra de navegación superior que muestra los enlaces a los cinco servicios más solicitados de *Google* de acuerdo a los resultados de la encuesta. En la parte inferior se presentan los enlaces para cambiar de vista o pasar a la página siguiente y a la anterior.

- Opciones al usuario. Se diseñan dos vistas de los resultados: una vista condensada con menos información textual para simplificar la interfaz y agilizar la búsqueda de información, y una vista extendida con más información textual como un apoyo extra a la búsqueda de información. El usuario podrá seleccionar entre la vista condensada y la vista extendida de los resultados. En ambas vistas se tiene la opción de pasar de una a otra. También tiene la opción de cambiar de la versión móvil del buscador a la versión clásica.
- Estándares web. Para el armado de la estructura del buscador se empleará *XHTML Mobile* y *PHP*, así como hojas de estilo en cascada *CSS Mobile* para la apariencia. Se debe utilizar la etiqueta *handheldfriendly* en el código de cada página para identificarlas como páginas que puedan ser navegadas usando dispositivos móviles.
- No riesgos conocidos en el diseño de las páginas. No se utilizan tablas, ni marcos, ni mapas de imágenes, ni ventanas emergentes.
- No a los tamaños fijos. Para el armado de las pantallas se utilizan porcentajes en lugar de tamaños fijos, para que el diseño se pueda adaptar a las variaciones de tamaño de pantalla de los diferentes modelos de teléfonos.
- Pruebas de usabilidad. Para las pruebas de usabilidad se realizarán búsquedas de información tanto en la versión actual del buscador de Google como en la nueva.

4. Discusión

Se han propuesto diversas guías de diseño web con mucho detalle como aquellas proporcionadas por las grandes compañías de software como Apple. Además, existe un estándar que propone la W3C para el desarrollo web móvil. Sin embargo, en este trabajo propusimos integrar la metodología de diseño centrado en el usuario con las guías de diseño web que propone la W3C. Este trabajo es un esfuerzo para el diseño y evaluación de sitios web usables para los teléfonos inteligentes.

Las guías de diseño propuestas han sido aplicadas de manera exitosa en una prueba de factibilidad por parte de un diseñador de aplicaciones mismo que reconoce que las guías de diseño le fueron útiles para diseñar la propuesta de modificación del buscador de Google y comenta que no tuvo tiempo de diseñar de manera completa la prueba de usabilidad. Reconocemos que falta solicitar a más diseñadores el uso de las guías de diseño. Sin embargo, el éxito de esta prueba de factibilidad nos permite extender el dominio de aplicación de las pruebas.

5. Conclusiones

Este trabajo presentó un conjunto de guías de diseño web móvil. Los elementos considerados en la guía fueron el resultado de la caracterización del usuario móvil, de los dispositivos móviles y de la web móvil. Además se consideró la integración de la metodología de diseño centrado en el usuario y de las mejores prácticas de diseño de la W3C. En primer lugar, se incluyeron dentro de las guías de diseño los pasos necesarios para caracterizar a los usuarios, a los dispositivos y a la web, mismos que forman parte de la etapa de análisis del DCU. En segundo lugar, se integraron las mejores prácticas de la web móvil y las recomendaciones del diseño web móvil en la fase de diseño del DCU. En tercer lugar, se consideraron también las mejores prácticas de la web móvil para integrarse en la etapa de construcción del DCU. Por último, se propuso integrar fase evaluación del DCU al proponer la realización de pruebas de usabilidad.

La prueba de factibilidad fue exitosa y nos permite recomendar que las guías pueden ser utilizadas tanto para evaluar las guías de diseño en casos de estudio como para el desarrollo de aplicaciones web móviles.

Como trabajo futuro se considera solicitar al diseñador construir la aplicación propuesta para la modificación de los resultados del buscador de Google para móviles, terminar el diseño de las pruebas de usabilidad y aplicar las pruebas. También se considera solicitar el uso de las guías de diseño a más diseñadores.

6. Bibliografía y Referencias

- [1] Apple Inc., Safari Web Content Guide, Apple Inc, 2016.

- [2] Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., and Steggles, P., Towards a Better Understanding of Context and Context-Awareness, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 304-307, 1999.
- [3] Ballard, B., Designing the Mobile User Experience. West Sussex: Wiley, 2007.
- [4] Baharuddin, R., Singh, D., & Razali, R., Usability dimensions for mobile applications-a review. *Research Journal of Applied Sciences, Engineering and Technology*, 5(6), pp. 2225-2231, 2013.
- [5] Cremin, R., Rabin, J., Fling, B. y Robinson, K., *DotMobi Mobile Web Developer Guide*, mTLD, Ltd. Dublin, Ireland, 2007.
- [6] Edmondson, J., Anderson, W., Gray, J., Loyall, J. P., Schmid, K., and White, J., Next-generation mobile computing. *IEEE Software*, 31(2), pp. 44–47, 2014.
- [7] Häkkinen, J. and Mäntyjärvi, J., Developing design guidelines for context-aware mobile applications. In *Proceedings of the 3rd International Conference on Mobile Technology, Applications & Systems, Mobility 06*, New York, NY, USA. ACM, 2006.
- [8] Love, S., *Understanding Mobile Human-Computer Interaction*. Oxford: Elsevier, 2005.
- [9] Lynch, P. J. y Horton, S., *Web Style Guide, 4th Edition: Foundations of User Experience Design*, Yale University Press, New Haven, CT, USA, 2016.
- [10] Nayebi, F., Desharnais, J. M., y Abran, A., The state of the art of mobile application usability evaluation. In *2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp. 1–4, 2012.
- [11] O'Reilly, T., *What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software*. O'Reilly, 2005.
- [12] Park, W., Han, S. H., Kang, S., Park, Y. S., and Chun, J., A factor combination approach to developing style guides for mobile phone user interface. *International Journal of Industrial Ergonomics*, 41(5), pp. 536–545, 2011.

- [13] Shitkova, M., Holler, J., Heide, T., Clever, N., Becker, J., Towards Usability Guidelines for Mobile Websites and Applications, *Wirtschaftsinformatik Proceedings 2015*. Osnabrück, Germany, pp.1603-1617, 2015.
- [14] Williams, A., User-centered Design, Activity-centered Design, and Goal-directed Design: A Review of Three Methods for Designing Web Application. In *Proceedings of the 27th ACM International Conference on Design of Communication*. ACM, New York, USA, pp.1-8, 2009.
- [15] World Wide Web Consortium, *Mobile Web Best Practices 1.0*. W3C, 2008.

ESTUDIO DE UNA ANTENA DE MICROCINTA FRACTAL TIPO E PARA LA BANDA DE LOS 2.4 GHZ

Iván R. González Rangel

Universidad Autónoma Metropolitana Unidad Azcapotzalco

Ivan.gonzalez.rangel@gmail.com

Javier Vargas Sánchez

Universidad Autónoma Metropolitana Unidad Azcapotzalco

Genaro Hernández Valdez

Universidad Autónoma Metropolitana Unidad Azcapotzalco

ghv@correo.azc.uam.mx

Mario Reyes Ayala

Universidad Autónoma Metropolitana Unidad Azcapotzalco

J. R. Miranda Tello

Universidad Autónoma Metropolitana Unidad Azcapotzalco

Edgar A. Andrade González

Universidad Autónoma Metropolitana Unidad Azcapotzalco

Resumen

En este trabajo, se presenta el estudio experimental y por simulación de una antena de parche basada en la geometría fractal tipo E para operar en la banda libre de los 2.4 GHz. La estructura electromagnética de la antena diseñada se analiza utilizando el simulador HFSS de ANSI, mientras que el prototipo experimental se caracteriza utilizando un analizador de redes de 6.5 GHz acoplado a un equipo de medición de campo cercano para medir los parámetros importantes de esta antena. Considerando la geometría fractal tipo E de dos iteraciones, los resultados en las simulaciones muestran que la antena diseñada

presenta una frecuencia de resonancia de 2.463 GHz con pérdidas por retorno cercanas a -50 dB y un ancho de banda de acoplamiento de 74 MHz. Los parámetros correspondientes en el prototipo experimental resultaron ser de 2.624 GHz, -17 dB y 100 MHz, respectivamente. Estos resultados demuestran la factibilidad de utilizar la antena de microcinta fractal tipo E para aplicaciones en las comunicaciones inalámbricas para la banda de los 2.4 GHz.

Palabras Claves: ancho de banda de acoplamiento, antenas de banda ancha, antenas de microcinta, bandas sin licencia, geometría fractal tipo E.

Abstract

In this paper, the simulation and experimental study of an E-shape fractal patch antenna for the unlicensed band of 2.4 GHz is presented. The electromagnetic structure of this antenna is analyzed using the High Frequency Structure Simulator HFSS of ANSY, while the experimental antenna prototype is characterized using a network analyzer of 6.5 GHz and a (near-field based) antenna pattern measurement equipment. Simulation results show that the designed antenna presents a resonant frequency of 2.463 GHz with return losses equal to -50 dB, and a coupling bandwidth of 74 MHz. On the other hand, the corresponding experimental parameters are 2.56 GHz, -17 dB, and 100 MHz, respectively. The study performed in this paper demonstrates the suitability of using the E-shape fractal patch antenna for wireless applications at the 2.4 GHz unlicensed band.

Keywords: broadband antennas, coupling bandwidth, ISM frequency bands, patch antennas, type-E fractal geometry.

1. Introducción

El diseño actual de antenas para aplicaciones de radio móvil requiere que éstas sean de tamaño compacto, bajo perfil, fáciles de diseñar, de fabricar, de integrar a dispositivos electrónicos, y que posean características de banda ancha o multibanda. Diversas soluciones se han propuesto para lograr uno o más de estos requerimientos. Por ejemplo, los cinco primeros requerimientos mencionados se han logrado al desarrollar antenas de parche o microcinta [Balanais, 1982], [Khan,

2015]. Recientemente, la tecnología de microcinta se ha combinado con las propiedades de la geometría fractal para mejorar las características de respuesta en frecuencia de las antenas para dispositivos portátiles [Gianvittorio, 2002]. En particular, las antenas fractales de microcinta poseen una estructura geométrica con propiedades auto similares que les permite tener características multi-resonantes o de banda ancha, las cuales resultan idóneas para aplicaciones de comunicaciones móviles [Werner, 2003]. Entre las formas geométricas que han resultado de mayor interés en el desarrollo de antenas fractales se encuentran el triángulo y carpeta de Sierpinsky, copo de nieve de Koch, árbol fractal, curva de Hilbert, estructura hexagonal, formas “E”, “T” y “U”, entre otras [Werner, 2003]. Entre estas, la geometría basada en el triángulo de Sierpinsky es ampliamente utilizada en la literatura para proponer antenas de microcinta multi-resonantes [González, 2016]. Por otra parte, aunque las antenas de parche basadas en la geometría fractal tipo E han sido poco estudiadas en la literatura, el interés por mejorar sus características multi-resonantes y de banda ancha ha aumentado en los últimos años [Bayatmaku, 2011], [Asghar, 2013], [Navukarasu, 2016], [Zakir, 2011]. En este trabajo, se estudia el desempeño (experimental y por simulación) de una antena de parche basada en la geometría fractal tipo E para aplicaciones inalámbricas localizadas en la banda libre de los 2.4 GHz. La idea es que este tipo de antena pueda ser empleada en dispositivos portátiles que trabajen bajo el estándar para redes inalámbricas de área local IEEE 802.11b/g.

La antena de microcinta fractal tipo E se diseña tomando como base la metodología de fabricación de la antena de parche rectangular [Balanais, 1982]. La antena diseñada se analiza utilizando el simulador de estructuras electromagnéticas de alta frecuencia (HFSS, por sus siglas en inglés) de ANSI. Los parámetros considerados en el diseño se programan en el simulador HFSS para mejorar las características de desempeño de la antena propuesta. Una vez que los resultados de simulación son satisfactorios, se procede a construir un prototipo y se caracteriza utilizando un analizador de redes “FieldFox” de 6.5 GHz de “Agilent Technologies” acoplado a un equipo de medición de parámetros de antenas “RFxpert”. Los resultados, tanto de la simulación como de la

experimentación, demuestran que es viable utilizar la antena de microcinta fractal tipo E para aplicaciones inalámbricas en la banda de los 2.4 GHz.

En la sección 2, se presenta la metodología de diseño en la que se basa la antena de parche de geometría fractal tipo E. La sección 3 presenta los resultados de la simulación y de la experimentación. En la sección 4 se presenta la discusión. Finalmente, la sección 5 muestra las conclusiones del trabajo.

2. Métodos

El diseño de la antena fractal tipo E de microcinta se basa en la metodología de diseño de la antena de parche rectangular. La antena de parche rectangular es una de las antenas básicas que más se han estudiado en la literatura, para su diseño se requiere conocer el material sobre el cual se fabrica ya que se considera el valor de la constante dieléctrica para calcular sus dimensiones. Se utilizan las ecuaciones 1, ecuación 4, las cuales describen el modelo simplificado de la antena [Balanis, 1982], [Gianvittorio, 2002], [Werner, 2003], [González, 2016]. El ancho del parche (w) se calcula mediante ecuación 1 [Balanis, 1982].

$$w = \frac{c}{2fr} \sqrt{\frac{2}{\epsilon_r + 1}} \quad (1)$$

Donde c representa el valor de la velocidad de la luz en el vacío, ϵ_r representa el valor de la constante dieléctrica del sustrato y f_r representa la frecuencia de resonancia en la que se desea operar. Antes de calcular la longitud del parche, es necesario evaluar el valor efectivo de la constante dieléctrica (denotado por ϵ_{reff}), el cual está dado por ecuación 2 [Balanis, 1982].

$$\epsilon_{reff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left[1 + 12 \frac{h}{w} \right]^{-\frac{1}{2}} \quad (2)$$

Donde h representa el espesor del sustrato sobre el cual se fabrica la antena. De esta forma, la longitud incremental de la antena se calcula mediante ecuación 3 [Balanis, 1982].

$$\Delta L = 0.412h \frac{(\epsilon_{\text{reff}} + 0.3) \left(\frac{w}{h} + 0.264 \right)}{(\epsilon_{\text{reff}} - 0.258) \left(\frac{w}{h} + 0.8 \right)} \quad (3)$$

Finalmente, la longitud del parche (denotada por L) se calcula con ecuación 4 [Balanais, 1982].

$$L = \frac{c}{2fr\sqrt{\epsilon_{\text{reff}}}} - 2\Delta L \quad (4)$$

Para este tipo de antena utilizamos placas de circuito impreso con doble capa de cobre y el sustrato es el FR4 ya que tiene una constante dieléctrica con un valor relativamente bajo ($\epsilon_r = 4.4$), una de las caras de cobre nos permite dibujar el parche mientras que la otra cara funciona como plano de tierra.

Las dimensiones calculadas para la antena de parche rectangular tienen los valores $w=38.03$ mm y $L=28.44$ mm. La ubicación del punto de alimentación de la antena es algo crucial, ya que esta debe ser de 50 ohms para acoplar la impedancia con los dispositivos a los que se conecta. Desafortunadamente, cada que se realiza un cambio en las dimensiones de la antena, la impedancia característica del parche cambia [Werner, 2003]. Para tener el mejor acoplamiento de la antena, mediante simulación se encontró que en el centro de un extremo de la dimensión L del parche, la impedancia es muy aproximada a los 50 ohms.

Como se mencionó anteriormente, la estructura de la antena de parche fractal tipo E tiene como base la antena de microcinta rectangular, la cual se utiliza típicamente en una sola frecuencia. El efecto multi-resonante de la antena tipo E se debe al flujo de corriente alrededor de la antena al incorporar las ranuras, esto permite que también funcione a una frecuencia más baja. El ancho de la longitud del parche permite que la impedancia de entrada de la antena se mantenga acoplada [Gianvittorio, 2002]. Para obtener la antena de parche tipo E, la forma del parche rectangular se altera para operar en las frecuencias deseadas, esto se logra haciendo defectos en el parche tales como las ranuras y modificando la línea de alimentación [Werner, 2003]. La antena tipo E con diversos defectos permite

controlar o modificar algunos de sus parámetros, uno de estos defectos se logra aplicando la geometría fractal. En este trabajo, la optimización del funcionamiento de la antena fractal tipo E se realiza utilizando simulación por computadora.

3. Resultados

Resultados de la Simulación de la Antena Tipo E

Se utiliza el paquete computacional HFSS de ANSY para la simulación de la estructura de la antena fractal tipo E. HFSS es un programa que permite elegir entre varios tipos de métodos para realizar un análisis de gran precisión. Se simula la antena con las dimensiones obtenidas por el método simplificado descrito en la sección anterior. Posteriormente, se ajustan los parámetros de diseño para optimizar su respuesta en la banda de frecuencia deseada [Skrivervik, 2001]. En todas las simulaciones, la impedancia de entrada de las antenas se encuentra acoplada en la frecuencia de resonancia (donde se obtienen las menores pérdidas por retorno) a un valor aproximado de $R=50$ ohms para la parte real y cero ohms para la parte imaginaria. La estructura de la antena de parche fractal tipo E de una iteración simulada se ilustra en la figura 1.

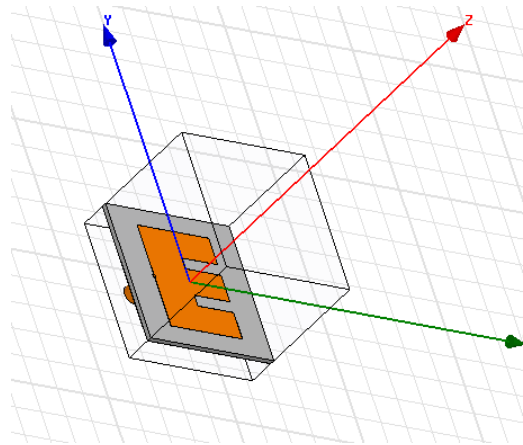


Figura 1 Antena tipo E simulada.

En la figura 2 se presenta la respuesta en frecuencia de la antena fractal de una iteración, en donde se puede apreciar el ancho de banda de acoplamiento (el cual es de 80 MHz), el cual nos garantiza que la antena trabaje bien para la banda de

2.4 GHz con el estándar IEEE 802.11b para los 11 canales asignados en el Continente Americano. La frecuencia de resonancia resultó ser de 2.439 GHz con pérdidas por retorno de -43.19 dB.

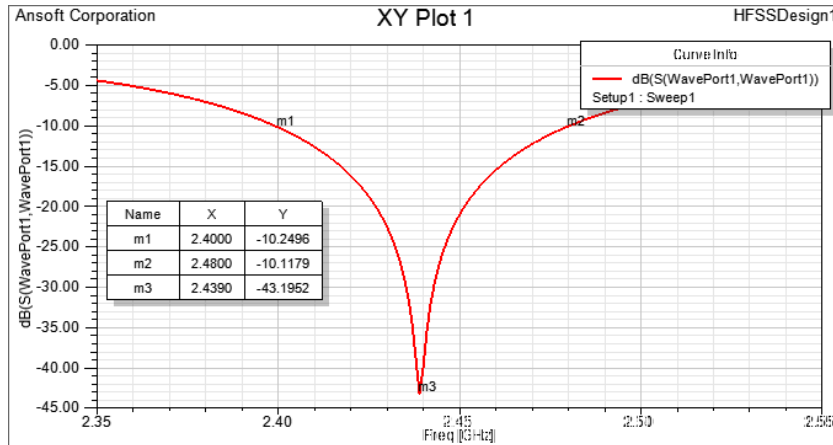


Figura 2 Simulación de la respuesta en frecuencia de la antena tipo E.

La figura 3 muestra el patrón de radiación de la simulación de la antena de parche tipo E de una iteración. En esta figura se puede observar que el patrón de radiación de la antena, tiene un ancho del haz amplio que le permite operar en la banda libre de 2.4 GHz y presenta una ganancia de 2.65 dB en la dirección de máxima radiación.

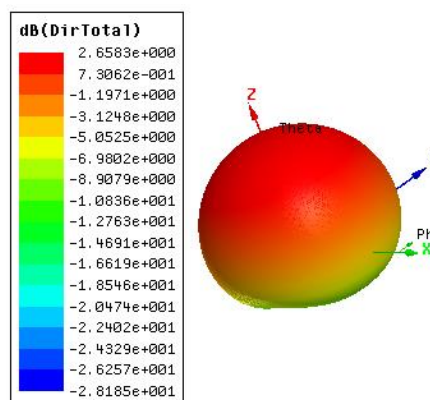


Figura 3 Patrón de radiación simulado de la antena fractal tipo E de una iteración.

A continuación, se presentan los resultados de la simulación en HFSS de la antena fractal tipo E de dos iteraciones para determinar las ventajas que se

consiguen. La estructura electromagnética de la antena de parche basada en la geometría fractal tipo E de dos iteraciones se muestra en la figura 4.

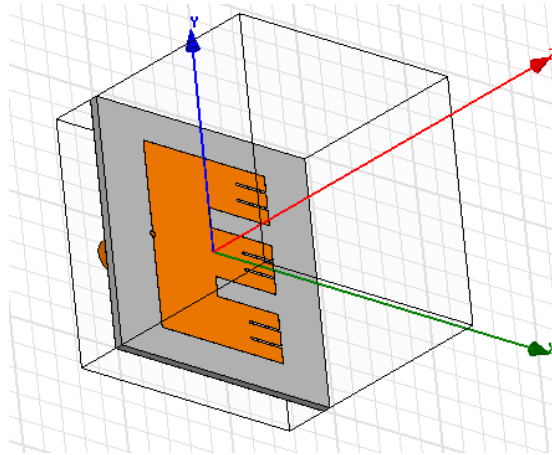


Figura 4 Antena de parche fractal tipo E de dos iteraciones simulada.

En la figura 5 se presenta la respuesta en frecuencia de la antena fractal tipo E de dos iteraciones. En ésta figura se observa un ancho de acoplamiento de 74 MHz, el cual comprende los 11 canales utilizados en el Continente Americano definidos en el estándar IEEE 802.11b. También se observa que la antena presenta una frecuencia de resonancia de 2.436 GHz con pérdidas por retorno de -49.23 dB.

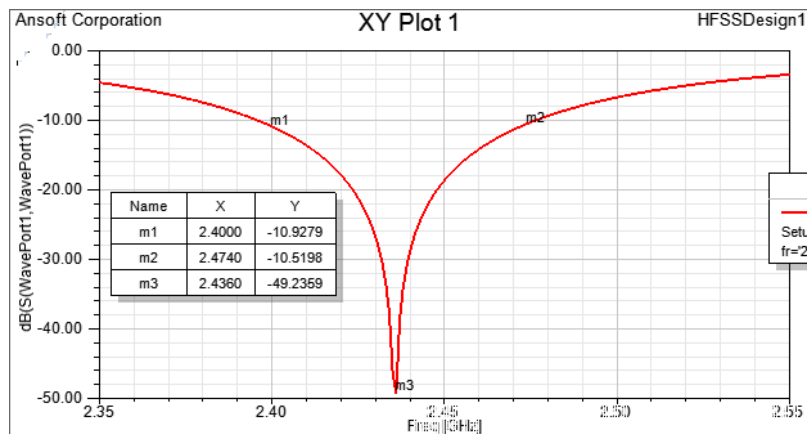


Figura 5 Respuesta en frecuencia de la antena fractal tipo E de dos iteraciones.

La figura 6 muestra el patrón de radiación, obtenido por simulación, de la antena de parche fractal tipo E de dos iteraciones. En esta figura se muestra que el haz

se encuentra en la parte superior del parche, también se observa un patrón de radiación con ancho de haz amplio, idóneo para aplicaciones inalámbricas. Un inconveniente de esta antena es su ganancia, la cual tiene un valor de 0.345 dB en la dirección de máxima radiación, por lo que se deben de realizar acciones para mejorar su desempeño en este aspecto. Por ejemplo, se pueden introducir defectos en el plano de tierra o utilizar un doble sustrato para mejorar el desempeño de la antena en su ganancia directiva [Werner, 2003].

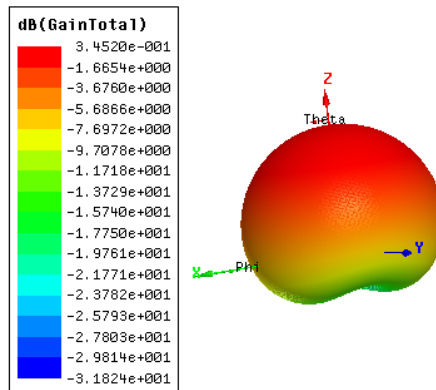


Figura 6 Patrón de radiación simulado de antena de parche fractal tipo E dos iteraciones.

En la tabla 1 se comparan los resultados de las simulaciones de las antenas de parche rectangular y de las antenas fractales tipo E, se observa que éstas últimas tienen un mejor desempeño en cuanto a pérdidas por retorno.

Tabla 1 Resultados de simulación de las antenas de parche analizadas.

Tipo de antena	Intervalo de acoplamiento (GHz)	Frecuencia de resonancia (GHz)	Pérdidas por retorno (dB)	Impedancia característica (Ohms)	Ancho de banda (MHz)
Rectangular	2.401 – 2.489	2.445	-27.5	$0.969R + j0.07$	88
Tipo E	2.400 – 2.48	2.439	-43.1	$1.00R - j0.12$	80
Tipo E fractal	2.400 – 2.474	2.436	-49.2	$0.942R - j0.008$	74

Resultados Experimentales de la Antena Fractal Tipo E

Para la caracterización de la antena construida (antena de parche fractal tipo E de dos iteraciones) se utilizó el analizador de redes “FieldFox”, con el cual se

midió la frecuencia de resonancia, las pérdidas por retorno y su impedancia característica. Además, se utilizó el equipo para medir parámetros de antenas en conjunto con el software “RFXpert” y el analizador de redes para medir el patrón de radiación, ancho del haz y ganancia en potencia de la antena bajo estudio. En la figura 7 se presenta el sistema de medición, el cual incluye los equipos utilizados y la antena bajo estudio.

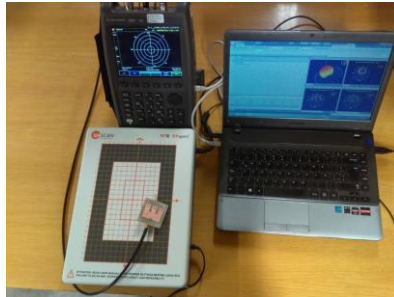


Figura 7 Sistema experimental para caracterizar antena de parche fractal dos iteraciones.

En la figura 8 se observa la respuesta en frecuencia (pérdidas por retorno) de la antena fractal tipo E de dos iteraciones. El ancho de banda de acoplamiento experimental resultó ser de 100 MHz, el cual comprende el intervalo de frecuencias de 2.56 GHz hasta 2.66 GHz con un acoplamiento máximo en 2.62 GHz, donde se obtuvieron pérdidas por retorno de -16.52 dB.



Figura 8 Pérdidas por retorno de la antena de parche fractal tipo E de dos iteraciones.

En la figura 9 se muestra la impedancia de entrada (compleja) en función de la frecuencia de la antena fractal tipo E de dos iteraciones. Observe que, en la

frecuencia con mayor acoplamiento, la impedancia de entrada resultó ser de $(21.1 + j10.0)$ ohms. Esto indica que es necesario construir un acoplador de impedancia para mejorar la respuesta de la antena (trabajo de investigación a futuro).



Figura 9 Impedancia de entrada de antena de parche fractal tipo E dos iteraciones.

La figura 10 muestra el patrón de radiación experimental de la antena de parche fractal tipo E de dos iteraciones. En ésta figura se observa que el haz se concentra sobre el parche y tiene un ancho de haz amplio, además, se observa que la ganancia en potencia experimental es mejor (9.66 dBi). Estas características indican que la antena de parche fractal tipo E es una opción para ser utilizada en aplicaciones inalámbricas y de radio móvil localizadas en la banda de los 2.4 GHz.

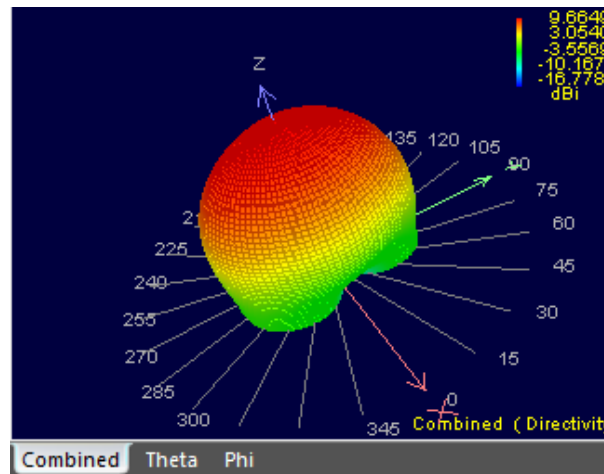


Figura 10 Patrón de radiación experimental de antena fractal tipo E de iteraciones.

4. Discusión

A partir del análisis y la comparación entre las antenas de parche rectangular, su modificación como antena fractal tipo E de una y dos iteraciones, se decidió construir y analizar experimentalmente las características de la antena fractal tipo E de dos iteraciones. Los resultados de simulación permitieron adaptar el funcionamiento de ésta antena al estándar IEEE 802.11b, el cual es uno de los más utilizados en la actualidad. Mediante simulación por computadora se encontró que, comparada con la antena de parche rectangular, posee mayores pérdidas por retorno y un ancho de banda de acoplamiento ligeramente menor; no obstante, cumple con el ancho de banda requerido para redes inalámbricas de área local que operan en la banda de 2.4 GHz. Una vez que los resultados de simulación fueron satisfactorios se procedió a construir esta antena. Mediante las ecuaciones del modelo simplificado se obtuvieron dimensiones para $w = 37.33 \text{ mm}$ y $L = 28.90 \text{ mm}$, mientras que en la simulación para que la antena cumpla con el estándar IEEE 802.11b, las dimensiones obtenidas tienen los siguientes valores: $w=35.79 \text{ mm}$ y $L = 27.68 \text{ mm}$. Los resultados experimentales muestran que la antena opera en un intervalo de frecuencias mayor al de la simulación, esta diferencia se debe a errores en las dimensiones físicas de la antena al momento de fabricar el prototipo. Además, el modelo experimental presentó una impedancia de entrada diferente de la esperada, por lo cual las pérdidas por retorno del prototipo experimental resultaron menores al del modelo simulado, sin embargo, el ancho de banda de su frecuencia de operación resultó ser mayor. Como trabajo a futuro se pretende utilizar técnicas de acoplamiento de impedancias, introducir defectos al plano de tierra y utilizar doble dieléctrico para mejorar el desempeño de la antena de parche con geometría fractal tipo E. El uso de estas técnicas permite que la antena opere en múltiples bandas de frecuencia (2.4 GHz y 5.8 GHz para aplicaciones basadas en el estándar 802.11 a/b/g).

5. Conclusiones

Tomando como base la metodología de análisis matemático para diseñar antenas de parche rectangular y el uso del simulador para estructuras

electromagnéticas de alta frecuencia (HFSS), se determinaron las dimensiones físicas para construir una antena de parche basada en la geometría fractal tipo E de dos iteraciones para operar en la banda sin licencia de los 2.4 GHz. La caracterización del prototipo de la antena construida se realizó utilizando el analizador de redes "FieldFox" en conjunto con el equipo para medir parámetros de antenas "EMSCAN" y el software "RFXpert". Se observó que, comparada con la antena rectangular, la antena fractal tipo E de dos iteraciones tiene un ancho de banda ligeramente menor, el cual se compensa con pérdidas por retorno mayores. En particular, el ancho de banda de acoplamiento experimental resultó ser de 100 MHz, con un acoplamiento máximo en 2.62 GHz, donde se obtuvieron pérdidas por retorno de -16.52 dB. El prototipo construido presenta un patrón de radiación con ancho de haz amplio, idóneo para aplicaciones inalámbricas, y una ganancia en potencia de 9.66 dBi. Los resultados tanto de simulación como experimentales dan evidencia de que es factible utilizar la antena de microcinta fractal tipo E para aplicaciones inalámbricas en la banda de los 2.4 GHz.

6. Bibliografía y Referencias

- [1] Asghar Abbas Razzaqi. Wideband E-Shaped Antenna Design for WLAN Applications, IEEE 9th International Conference on Emerging Technologies (ICET). pp. 1-6, 2013.
- [2] Balanis A. Constantine, Antenna theory analysis and design. Wiley. New York, 1982.
- [3] Bayatmaku N., Lotfi P., Azarmanesh M., and Soltani S., Design of simple multiband patch antenna for mobile communication applications using new E-shape fractal, IEEE Antennas and Wireless Propagation Letters, Vol. 10, 2011.
- [4] Gianvittorio J. P., and R.-Samii Y., Fractal Antennas: A novel antenna miniaturization technique, and applications, IEEE Antennas and Propagation Magazine, vol. 44, no. 1, pp. 20-36, February 2002.
- [5] González Rangel I. R., Hernandez-Valdez G., Andrade-Gonzalez E. A., Reyes Ayala M., Miranda-Tello J. R., and Serrano Chávez J., Relationship

- among resonant frequencies of Sierpinski multiband fractal antennas, The 2016 International Conference Applied Mathematics, Computational Science and Systems Engineering (AMCSE 2016), November de 2016.
- [6] Khan M. U., Sharawi M. S., Mittra R., Microstrip patch antenna miniaturization techniques: a review, *IET Microwave, Antennas & Propagation*, Vol. 9, No. 9, pp. 913-922, 2015.
- [7] Navukarasu G. J., Design of an E Shaped Patch Antenna for GPS and IRNSS Application, *International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, pp. 179-183, 2016.
- [8] Skrivervik A. K., Zurcher J.-F., Staub O., and Mosig J. R., PCS Antenna Design: The Challenge of Miniaturization, *IEEE Antennas and Propagation Magazine*, vol. 43, no. 4, pp. 12-26, August 2001.
- [9] Werner D. H., and Ganguly S., An overview of fractal antenna engineering research, *IEEE Antennas and Propagation Magazine*, vol. 45, no. 1, pp. 12-26, pp. 38-57, February 2003.
- [10] Zakir A., E-Shaped Microstrip Antenna on Rogers Substrate for WLAN Applications, *International Conference on Computational Intelligence and Communication Networks*, pp. 342-345, 2011.

CONDITIONING AND SIGNAL AMPLIFICATION STAGES FOR A SMART GAS MICROSENSOR MEMS

J. L. González Vidal

UAEH

jvidal@uaeh.edu.mx

M. A. Reyes Barranca

CINVESTAV-IPN *mreyes@cinvestav.mx*

E. N. Vázquez Acosta

CINVESTAV-IPN

neo_wolfx@hotmail.com

Resumen

En este trabajo, el objetivo principal es el diseño de una etapa nueva de amplificación y acondicionamiento de señal utilizando opamps, la cual será utilizada con un microsensor de gas inteligente. Las dimensiones del diseño fueron calculadas en base al modelo de pequeña señal de los MOSFET. Las configuraciones del opamp se eligieron de acuerdo a los requerimientos del sensor. El sensor está hecho de una película delgada de ZnO, depositada sobre la capa superior de una microplaca caliente sobre un microfoso micromaquinado. Gracias a la integración de las etapas de acondicionamiento de señal dentro del diseño, se dice que es un sensor inteligente. Se respaldó la operación de la etapa de amplificación y de acondicionamiento de señal por medio de técnicas de análisis de estabilidad, tales como el método del lugar geométrico de raíces. Se realizaron simulaciones y graficaron diagramas de Bode del sistema.

Palabras clave: Sensor de gas, MEMS, Opamp, sensores inteligentes.

Abstract

In this work, the principal aim is the design of novel signal amplification and conditioning stages using opamps, for an intelligent gas microsensor. The design dimensions were computed based on a small signal model of a MOSFET. The necessary configurations for the opamp were chosen according to the requirements of the sensor, which means, that the input signals are taken from this one. The sensor is based on a ZnO thin film deposited on the top side of a micro hot plate located within a micromachined pit. Thanks to the integration of signal conditioning stages within the design, it is said that it is an intelligent sensor. The right operation of the amplification stage and signal conditioning is supported by means of techniques of analysis of stability like the roots locus method, and computed Bode diagrams and simulations have been developed.

Keywords: Gas sensors, MEMS, opamp, smart sensors.

1. Introduction

Today, MEMS microsensors and smart sensors have very important and multiple applications in industry, automotive sector, aeronautics, biomedicine, consumer staples, etc. The most important advantages about MEMS are: their size (they are small devices), their low power consumption, batch fabrication and very low costs. MEMS sensors make a rich design space of networked sensors viable [González, 2006], [Howe, 1996].

Intelligent microsensors have a sensing stage and a conditioning stage due to the fact that the sensing stage provides a very low current in the nanoampere range [González, 2006]. Such current is a nonlinear signal and, since it is extremely low, it is hard to be manipulated by any ordinary electronic circuitry. For this reason, it is necessary to add linearization and amplification stages within the same monolithic integrated circuit. With these stages, undesired effects in the integrated circuit can be eliminated, such as parasitic charges, interface problems, noise, among others [Vázquez, 2008].

The signal provided by the intelligent microsensor can then be converted to digital mode and then it be manipulated by a digital system or a digital computer.

2. Methods

Consider a gas microsensor which provides a small and nonlinear current signal. That signal must be amplified and linearized. With this in mind, an opamp was designed using CMOS technology.

Three-Stage Opamp

A three-stage opamp was proposed as is shown in figure 1, where the first stage, A1, is a differential amplifier. The later has an input labeled V_{di} . The second stage, A2, is the gain stage, which decreases the gain at high frequencies; and the last stage, X1, provides gain in current mode.

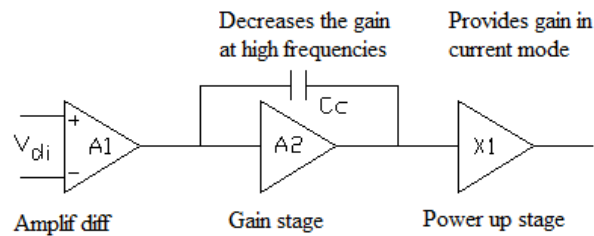


Figure 1 Three-stage opamp.

The diagram of the designed opamp is shown in figure 2. This opamp diagram was performed using Orcad® software. Subsequently, the opamp shall be explained in detail.

The Opamp design was developed according to AMIS 0.5 foundry technology and the layout was developed by using the L-Edit® software by Tanner-EDA. Values of $L = 2\lambda_D$, $W = 3\lambda_D$ and $\lambda_D = 1.5 \mu\text{m}/2$ were proposed.

Differential Pair (Inputs)

To better understand how opamp works, it was divided in several parts, the first one is the differential pair input.

The differential pair is formed by M_1 and M_2 n type transistors. It is well known that the current across the differential pair, comes from a current mirror (M_3 and M_4 , p type transistors). The dimensions of the transistors were calculated under the same conditions. That way, I_{SS} is given by equation 1.

$$I_{SS} = 10\mu A = \frac{KP_n}{2} \frac{W_{1,2}}{L_{1,2}} (V_{GS5,6} - V_{THN})^2 \quad (1)$$

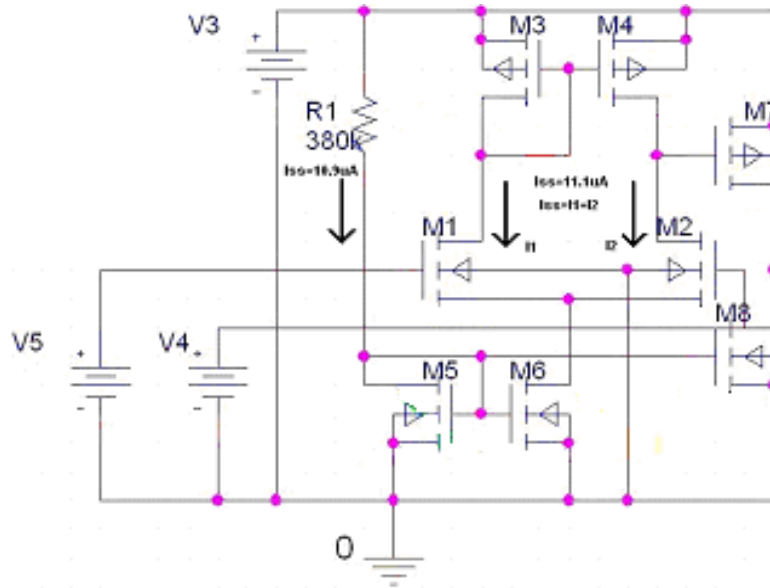


Figure 2 Opamp schematic circuit.

From equation 1 $W_1=W_2=15.61 \mu m \approx 21\lambda_D=15.75 \mu m$ and the current $I_{SS}=10.09 \mu A$ were calculated, this current flows across M_1 and M_2 . For frequency compensation, these dimensions may be increased, even when the voltage applied affects too much adversely the current value which should be a constant value. [Vázquez, 2008], [Tsviois, 1999], [Baker, 1997], [Gray, 2001].

The same design rules are now applied for p type transistors. The current source load was calculated to assure a $I_{SS}=10.09\mu A$ current and a voltage of $V_{GS3,4}=V_{THP}+0.3V=1.48V$, $V_{DD}=5V$, $V_{SS}=0V$, then W_3 can be calculated using equation 2.

$$I_{SS} = 10.09\mu A = \frac{KP_p}{2} \frac{W_{3,4}}{L_3} (V_{GS3,4} - V_{THN})^2 \quad (2)$$

$W_3=W_4=52.93 \mu m \approx 71\lambda_D=53.25 \mu m$ and $I_{SS}=10.15 \mu A$ were calculated. This current parameter increases when the load potential of the current source load needs to decrease.

Current Mirror

To better understand how the differential amplifier stages work, it is necessary to understand how the current mirror works. This configuration is frequently used in circuit designing.

The current source was designed assuming that $V_{DD}=5V$, $V_{SS}=0V$, $I_{SS}=10\mu A$, $V_{GS5,6}=0.85V$, with these voltage and current values both saturations of M_1 and M_2 are assured. The value of R is calculated assuming that $I_{D5}=I_{D6}=10\mu A$, if Solving for R , equation 3.

$$R = \frac{V_{DD} - V_{GS5,6} - V_{SS}}{I_{SS}} \quad (3)$$

Therefore, M_5 and M_6 currents have the same values; both dimensions of M_5 and M_6 were calculated by using equation 4.

$$I_{SS} = \frac{KP_n}{2} \frac{W_{5,6}}{L_{5,6}} (V_{GS5,6} - V_{THN})^2 \quad (4)$$

Although there is a small increase in the power consumption and the gain decreases, this will not affect the designed circuit.

Output impedance is denoted by equation 5.

$$r_o = \frac{\lambda_c}{I_D} = \frac{1}{\lambda I_D} \quad (5)$$

λ is calculated based on experimental data (0.06 V⁻¹ typical), equation 6.

$$r_o = \frac{\lambda_c}{I_D} = \frac{1}{\lambda I_D} \quad (6)$$

Common Source Stage (Output)

This stage is formed with transistors M_7 and M_8 . Transistor M_8 is biased as current source and the input of the amplification stage is connected to the gate of M_7 .

The current through M_8 and the current mirror (M_5 and M_6) must be the same. Therefore, M_8 , M_5 and M_6 have the same dimensions $W_5=W_6= W_8=15.61\mu m \approx 21\lambda_D=15.75\mu m$, whereas M_7 has the same dimensions than M_8 , therefore, the value of the current mirror and the differential pair is the same current.

Once the simulation is carried out, the transistor dimensions should be adjusted to fulfill the polarization requirements, signal conditioning, power consumption, stability and gain, among others [Vázquez, 2008], [Razavi, 2001], [Giurgiutiu, 2010], [Schilling, 1999].

3. Results

With the objective of determining the characteristics of the opamp, the circuit performance was simulated, 2.5 V were applied to the non-inverting input (V5) with two voltage sweeps. The first sweep is from 0 to 5 V and the second one from 2.47 V to 2.53 with 0.001 V steps in the non-inverting terminal (V4).

The diagram of opamp was simulated with Orcad® software, according to the parameters of V3.1, level 7 of AMIS 0.5µm technology libraries, equations have a complexity degree of level 3. The results from simulations are shown in figure 3, where the I_{SS} in M_5 is raised to 10.9 µA due to channel modulation effects (figure 3a), and I_{SS} in M_6 is raised to 11.1 µA (figure 3b). On the other hand, voltage V_{GS} in M_5 is 858 mV (figure 3c). From figure 3d it is shown that the equation $V_{GS5} \geq 0.3V + V_{THN}$ is satisfied which means that the transistors are in saturation [Vázquez, 2008].

Gain

Gain is expressed by the equation 7.

$$A_{OL} = A_1 \cdot A_2 = g_{m1} (r_{o2} \parallel r_{o4}) \cdot \left[-g_{m7} (r_{o7} \parallel r_{o8}) \right] \quad (7)$$

From figure 4, it can be determined that the gain is given by equation 8.

$$|A_{OL}| = \frac{\Delta V_{Output}}{\Delta V_{Input}} = \frac{2.93V}{6815 \mu V} = 4,305 \quad (8)$$

Gain is modified due to body effect. Gain results in the λ variation and, in consequence, the variation of equation 8. The output current of the second stage will be the same current of the design, this means $I_{SS} = 10.9 \mu A$.

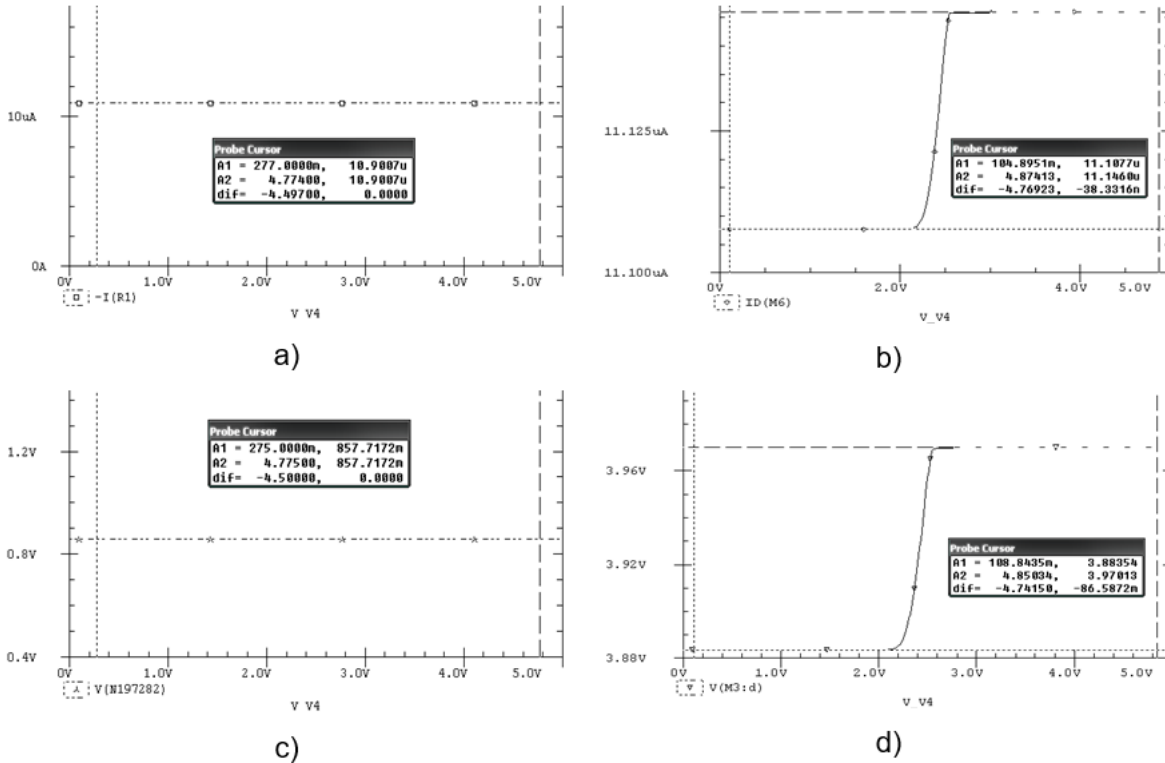


Figure 3 Simulation plots of schematic of figure 2.

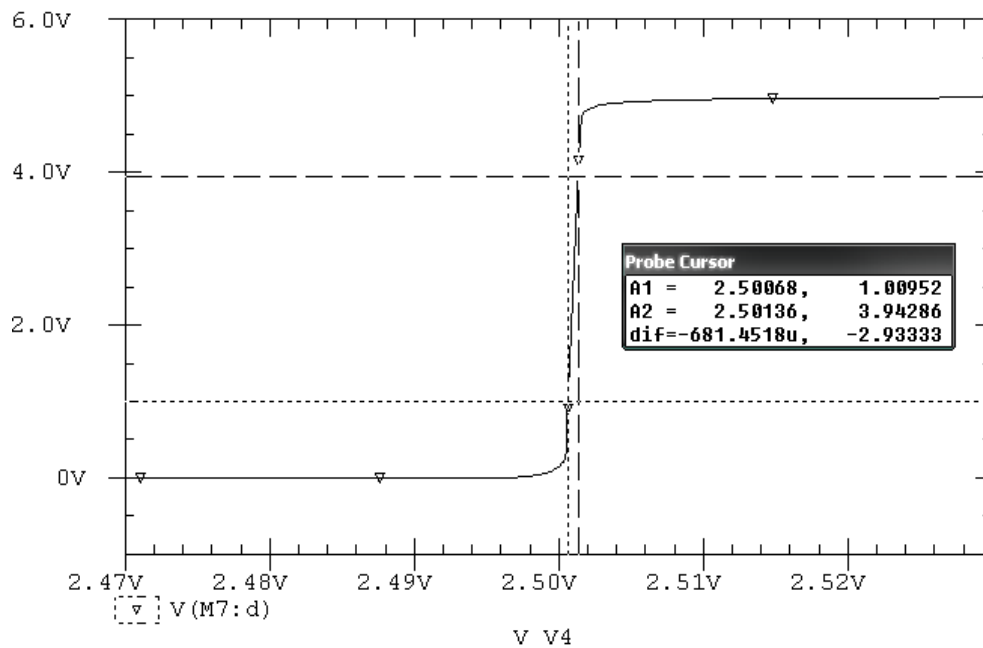


Figure 4 Opamp response due two voltage sweeps applied.

Small Signal Analysis

In a two-stage amplifier, figure 5, and considering high impedance nodes, which determine the dominant poles, the output resistance R_1 of the opamp is connected to ground, is given by equation 9.

$$R_1 = r_2 \parallel r_4 \quad (9)$$

Where both r_2 and r_4 are the output impedances of M_2 and M_4 respectively.

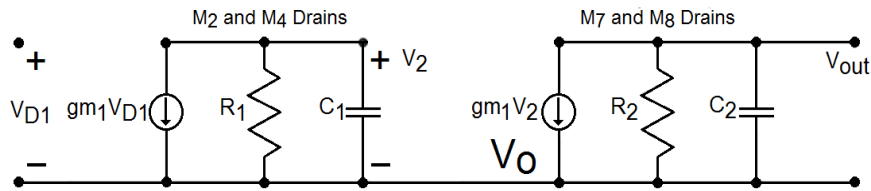


Figure 5 Small signal model of a two stages amplifier.

Capacitance C_1 , can be calculated using equation 10.

$$C_{TOT} = C_L + C_{db4} + C_{gd4} + C_{db2} + C_{gd2} \quad (10)$$

Where C_{TOT} is the total capacitance of the circuit, the load capacitance C_L is $C_{dg7} + C_{gs7}$, which are drain-gate and gate-source capacitances of M_7 , C_{db4} and C_{gd4} are drain-bulk and gate-drain capacitances of M_4 , C_{db2} and C_{gd2} are drain-bulk and gate-drain capacitances of M_2 . Since C_{dg7} is connected between the input and the output of the amplifier, the Miller theorem can be applied to divide the capacitance in two parts. The first capacitance is connected to the input and the second one to the output of the amplification stage. Miller capacitances are supposed to be connected to the gate and to the physical ground and between the drain and the physical ground; then, the load capacitance is now by equation 11.

$$C_L = C_{gs7} + C_{MI} \quad (11)$$

Where C_{MI} is known as Miller input capacitance, C_{MI} is denoted by equation 12.

$$C_{MI} = (1 + |A_2|) \quad (12)$$

And C_{MO} , the output miller capacitance, is denoted by equation 13.

$$C_{MO} = \left(1 + \frac{1}{|A_2|} \right) \quad (13)$$

On the other hand, the drain node of M_7 is characterized by R_2 and C_2 , where $R_2 = R_1$, because the current of polarization is the same.

C_2 is given by equation 14.

$$C_2 = C_{gd7} \left(1 + \frac{1}{|A_2|} C_1 = 1.14 \text{ pF} \right) + C_{db7} + C_{db8} + C_{gd8} \quad (14)$$

Finally, the calculated values are $R_1 = R_2 = 1502 \text{ k}\Omega$, $C_1 = 1.14 \text{ pF}$ and $C_2 = 1.84 \text{ fF}$ [Vázquez, 2008], [Giurgiutiu, 2003].

Frequency Response Analysis

The closed-loop gain of the opamp is described in terms of equation 15.

$$A_{CL} = \frac{A_{OL}}{1 + A_{OL}\beta} \quad (15)$$

Where A_{OL} is the open loop gain of the opamp.

The amplifier reaches stability, equation 16.

$$|A_{OL}\beta| = 1 \quad \text{and} \quad \angle A_{OL}\beta = \pm 180^\circ \quad (16)$$

Where β represents the amount of output signal that could be feedback and subtracted to the input of the amplifier; the highest possible value of β with amplification is given when $\beta=1$ and this condition is achieved in the voltage follower configuration.

Figure 6 shows the Bode plots resulting from the simulations. These were performed using the Orcad® software, applying an input signal of 0.01 Hz up to 10 GHz, with 10-decade steps. In figure 6a, the system shows a unit gain (0dB), which corresponds to a phase angle of 45° , whereas figure 6b shows that the gain is -25.997dB, which means that the system does not require external compensation [Vázquez, 2008], [Razavi, 2001], [Giurgiutiu, 2010], [Schilling, 1999].

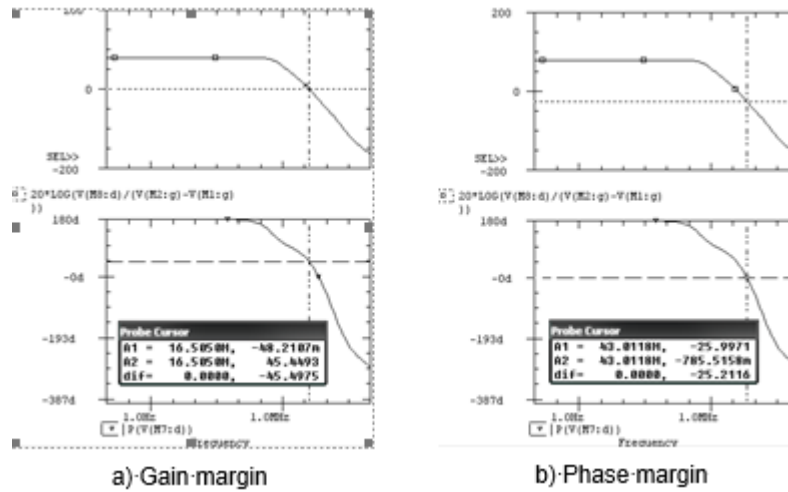


Figure 6 Bode plots.

Opamp Characterization

One quality of the opamp is to reject a common signal (CMRR) applied to its two inputs and it is represented by the gain in common mode given by equation 17.

$$CMRR = 20 \log \left| \frac{A_v}{A_c} \right| = 20 \log \left| g_{m1} (r_{o2} \parallel r_{o4}) \cdot 2g_{m4} r_{o6} \right| \quad (17)$$

In figure 7 it can be noticed that $CMRR = -42.92dB$, which is a convenient value for a good performance of the amplifier. On the other hand, the power supply rejection ratio (PSRR) is used to describe the amount of noise that a voltage power supply of a particular device can reject (equation 18).

$$PSRR = A_{OL} / (v_{out} / v_{sin}) \quad (18)$$

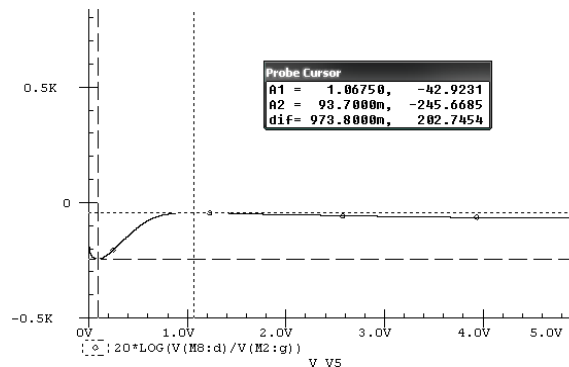


Figure 7 Common mode gain (CMRR) for an input signal applied (0 through 5 V, with 0.0001 V steps).

Where A_{OL} is the open-loop gain, v_{out} is output voltage and v_{sin} is input sinusoidal voltage; for this opamp PSRR = -3.704dB. Figure 8 shows the maximum cut frequency, where a f_{max} = 195 KHz can be noticed [Vázquez, 2008].

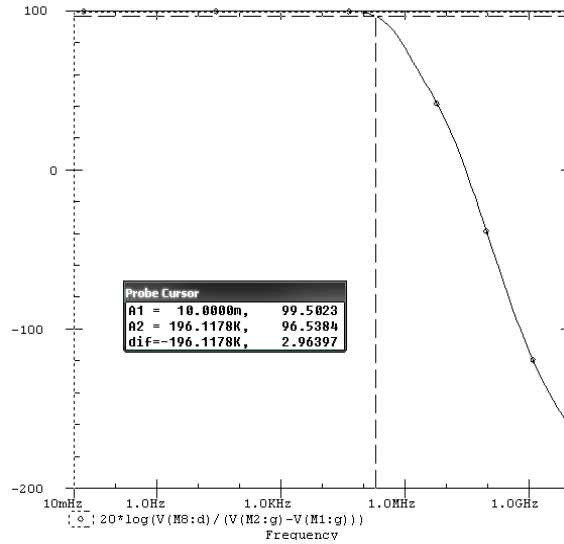


Figure 8 PSRR simulation.

Gas sensor-Opamp Connection

It is well known that electrical properties of semiconductors have a tight dependence on temperature. In the case of semiconductor oxides, a ZnO thin film must reach temperatures close to 300 °C so the right adsorption and reduction processes can be correctly carried out in presence of oxidizing or reducing species. The main characteristic of a ZnO thin film is that the variation of its resistance is not a linear in presence of a reducing species such as CO. For this reason, the most of gas sensors are based on thin films of semiconductor oxides, this process is explained in detail in [González, 2005, 2006, 2013].

In figure 9, the behavior of a gas sensor can be observed. The gas sensor measurements were compared to simulations that were carried out using Matlab® and Orcad® softwares. The use of a 100MΩ resistance in series (voltage divider) with the gas sensor, allowed getting voltage values for different CO concentrations. Due to sensor resistance variation, it cannot be measured directly. For this reason, an opamp in voltage follower configuration is connected to the voltage divisor

output; this allows the interaction between the gas sensor and the load. In this case, the load is a variable gain amplifier (VGA).

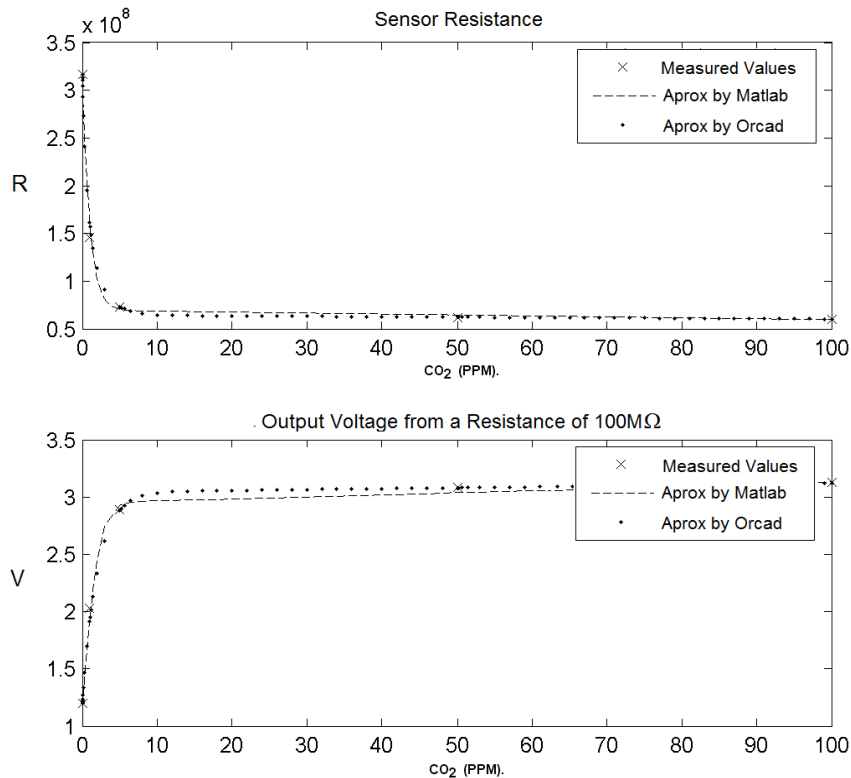


Figure 9 Variation of sensor resistance and voltage divider output potential for several CO_2 concentrations.

Since the output in voltage mode of a voltage divisor has a logarithmic behavior, an output in a proportional voltage mode of CO concentration is desired. The VGA provides a linear variation of output voltage of the signal conditioning stage, with respect to a CO concentration variation. After analyzing the slope of the logarithmic plot, there are four regions where such variation of the plot is minimum. Therefore, the VGA has four possible changes in its gain (although there might be fifteen changes). Each of the four regions already mentioned have different slopes that must be interpreted. VGA lets every region to have the same slope; nevertheless, the slope increases or decreases the voltage value in each region, which means that it moves the whole region upwards or downwards. That is why compensation voltages are required. Compensation voltages are different in each region and they

depend on the gain. Voltages are added or subtracted to a whole region depending on whether the signal is amplified or attenuated, respectively [Vásquez, 2008].

The current follower circuit is connected to the VGA output with the purpose of avoiding any interaction with the load. Figure 10 shows the schematic circuit. It was designed with connections for external resistances, because external resistances can modify the gain. First and third terminals are designed for power supply connection. The second terminal is designed for connection of the 100 MΩ resistance, V_{div} is the output voltage of A_{s1} ; an input resistance R_1 is connected to V_{div} and R_1 terminals; V_c is the voltage that compensates the gain change of A_{gv} ; $C_{v1,2,3,4}$ are control signals which determine feedback resistance. Feedback resistances R_f are connected between $R_{f1,2,3,4}$ and V_{fb} . The output signal of A_{gv} is connected to the non-inverting output of the current follower A_{s2} . A_{s2} avoids the load to drain current to A_{gv} , and finally V_{out} is the output voltage of the circuit [Vásquez, 2008].

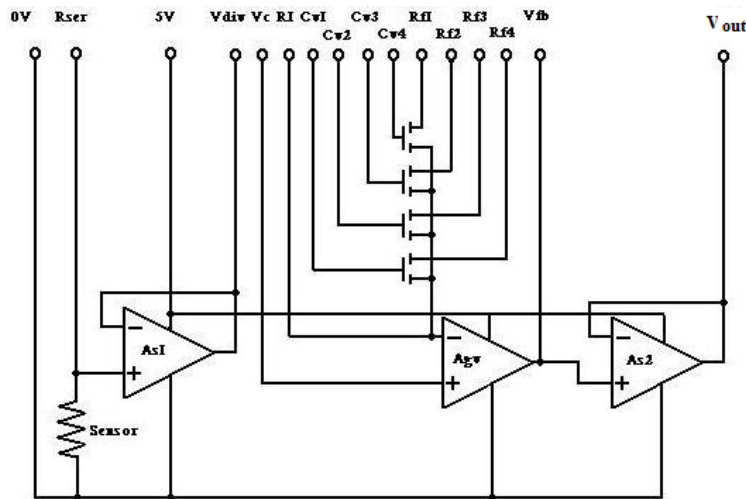


Figure 10 Schematic circuit developed.

4. Discussion

The signal conditioning stage is very important, because sensor signal linearization is needed. The sensor's linear signal is taken to a standard range for right measurements. Signal conditioning stages must have a right approximation of the sensor's behavior. The approximations were developed by Matlab® using

recursive gradient method. However, Orcad® allows designing more accurate functions for the electrical behavior of the device, because resistive, capacitive and inductive elements are used. Nevertheless, the functions developed in Matlab® had problems when they were translated to Orcad®.

On the other hand, the use of PMOS instead of NMOS makes all the transistors not to work in a saturation mode, and it was observed that their behavior was unstable. The use of a logarithmic conditioning stage does not work because the constants defined in this stage are different from the sensor's response parameters; therefore, they do not eliminate each other. Finally, using a resistance of 100MΩ in series with the gas sensor allows obtaining voltage values for different CO concentrations. Since the sensor resistance variation cannot be measured directly, an opamp in current follower configuration is connected to the voltage divisor output, and generating the possibility of interaction between the gas sensor and the load. In this case, the load is a VGA. The realization of this on-chip device will be developed.

5. Conclusions

A three-stage opamp was designed, such stages are three opamps called A1, A2 and X1; since this circuit will have a special application, a lot of values were computed.

In realizing of the opamp design, it was divided in a differential pair (formed by M_1 and M_2 transistors), current load source (formed by M_3 and M_4 transistors), current mirror (formed by M_5 and M_6 transistors) and common source stage (formed by M_7 and M_8 transistors). Drain currents I_D and transistor dimensions as W and L were calculated. Opamp design was developed according to 0.5μm AMIS technology and the layout was developed by using L-Edit software by Tanner EDA. Simulations were developed using Matlab® and Orcad® softwares. Several characteristics of the opamp were analyzed, such as gain, small signal analysis, frequency response analysis; in addition, CMRR and PSRR were calculated.

Opamp design and calculations were based on basis of semiconductors, electrical physics, magnetism concepts, modeling and transistors performance. In addition,

frequency stability was applied. Computational techniques gave us a very reliable approximation of the performance and behavior of the opamp. The designed Opamp was developed according to a particular gas sensor whose surface resistance varies two orders of magnitude.

Acknowledgments

This work was sponsored by UAEH.

6. Bibliography and References

- [1] Behzad Razavi, Design of Analog CMOS Integrated Circuits, McGraw-Hill Higher Education, 2001.
- [2] Donald L. Schilling, Charles Belove, Tuvia Apelewicz, Raymond J. Saccardi, Electronic Circuits, Discrete and Integrated, Third Edition, McGraw-Hill Book Company, 1999.
- [3] Edgar Norman Vázquez Acosta, Diseño de etapa de amplificación y conversión A/D en un circuito integrado para un microsensord de gases (MEMS) inteligente. Master degree Thesis, UAEH, Mineral de la Reforma, Hidalgo, Mexico, 2008.
- [4] J. L. González Vidal, Alfredo Reyes Barranca y Wilfrido Calleja Arriaga, Technological Processes for Micro-Heater and Micro-Hot-Plate in the Implementation of a MEM Gas Sensor, 2nd International Conference on Electrical and Electronics Engineering (ICEEE) and XI Conference on Electrical Engineering, Mexico City, Mexico, pp. 440-443, 2005.
- [5] J. L. González-Vidal, Alfredo Reyes Barranca, Wilfrido Calleja Arriaga, Juan Silva F e I. Juárez, Caracterización de la interfase de Polisilicio-ZnO, para un microsensord de gases micromaquinado, XXV Congreso Nacional Sociedad Mexicana de Ciencia y Tecnología de Superficies y Materiales, Zacatecas, Zacatecas, Mexico. September 2005.
- [6] José Luís González Vidal, Aplicación de Estructuras Micro-Electro-Mecánicas (MEMS) con Tecnología CMOS para Sensores de Parámetros Físicos. Ph D. Thesis, CINVESTAV-IPN, Mexico City, Mexico, 2006.

- [7] Paul R. Gray, Robert G. Meyer, *Analysis and Design of Analog Integrated Circuits*. Wiley and Sons, 2001.
- [8] R Jacob Baker, Harry W. Li and David E. Boyce, *CMOS Circuit Design, Layout, and Simulation*, Willey-IEEE PRESS, 1997.
- [9] Roger T. Howe, Charles G. Sodini, *Microelectronics, an Integrated Approach*, Prentice Hall, pp 176-179, 1996.
- [10] Victor Giurgiutiu, Sergey Edgard Lyshevsky, *Micromechatronics. Modelling, Analysis, and Design with MATLAB®*, CRC PRESS, 2003.
- [11] Yannks P. Tsviois., *Operation and Modeling of the MOS Transistor*, 2nd ed, Oxford University press, 1999.

TÉCNICA DE CONMUTACIÓN SUAVE PARA UN CONVERTIDOR REDUCTOR-ELEVADOR DOBLE CON APLICACIONES EN ILUMINACIÓN

Pablo Israel Guzmán Tafoya

Tecnológico Nacional de México en Celaya

M1503104@itcelaya.edu.mx

Nimrod Vázquez Nava

Tecnológico Nacional de México en Celaya

n.vazquez@ieee.org

René Osorio Sánchez

Universidad de Guadalajara

rene.osorio@profesores.valles.udg.mx

Resumen

En este artículo se presenta una técnica de conmutación suave para el convertidor Reductor-Elevador Doble (CRED) con aplicación en lámparas LED. El convertidor consta de un convertidor Reductor-Elevador de dos etapas que presenta un interruptor principal y salida positiva. La primera etapa, conectada a la red eléctrica, contiene un inductor que opera en modo de conducción discontinua (MCD), permitiendo alcanzar naturalmente un factor de potencia elevado. La segunda etapa busca mejorar la respuesta dinámica de la corriente en la lámpara LED de 69 W trabajando en modo de conducción continua (MCC), obteniendo así, un rizo de corriente pequeño a la salida del convertidor. La topología, al ser una integración de dos convertidores, presenta una concentración de esfuerzos en el interruptor cuando se activa con conmutación dura. La técnica de conmutación suave propuesta consiste en un snubber activo de voltaje y un snubber pasivo de corriente que busca modificar las señales en el transistor, de voltaje en el apagado

y de corriente en el encendido. Simulaciones en el software PSim son utilizadas para demostrar el funcionamiento de los circuitos auxiliares.

Palabras Claves: Corriente, interruptor, reductor-elevador, snubber, voltaje.

Abstract

In this paper, a soft switching technique for a LED driver based on an Integrated Double Buck-Boost (IDBB) converter is presented. The converter is a two-stage cascade Buck-Boost topology featuring a single main switch and positive voltage output. The first stage is connected to the grid and it is operated in discontinuous conduction mode (DCM); working in this way a high power factor is naturally achieved. The second stage is used to improve the dynamic performance of the LED current and works in continuous conduction mode (CCM), which allows a low output current ripple in a 69W LED lamp. As an integrated converter, the power losses of the two stages are concentrated in the single transistor, when it is in hard switching. The proposed soft switching technique consists on a lossless voltage active snubber and a current passive snubber, which allows that the converter switches at zero voltage at the turn off, and in zero current at during the turn on. A comparison between hard switching and soft switching techniques is made based on a PSim simulation.

Keywords: Buck-boost, current, snubber, switch, voltage.

1. Introducción

La iluminación LED se ha convertido en una fuente confiable de luz por sus características como el tiempo de vida que presentan sus componentes, el empaquetado pequeño y libre de mercurio y las eficiencias elevadas [Zhou, 2008], [Bo, 2009]. A pesar de los rasgos positivos que las lámparas LED presentan, deben contar con un convertidor que asegure las especificaciones correctas como un factor de potencia elevado y una corriente constante para evitar cambios bruscos en la iluminación.

Se han propuesto diferentes soluciones para mantener un factor de potencia elevado. Circuitos pasivos y activos se han encontrado como respuestas. De igual

manera, para reducir la intermitencia de la luz, se aplican circuitos que entregan rizados de corriente pequeños y sin grandes variaciones [Gobbato, 2016], [Kim, 2017].

El CRED logra obtener el factor de potencia elevado con un componente pasivo operado en modo de conducción discontinua. La salida del convertidor ofrece un rizo de corriente pequeño debido a la integración de un segundo convertidor Reductor-Elevador alimentado por la primera etapa. El convertidor consta de un interruptor, tres diodos, dos inductores y dos capacitores (figura 1) [Alonso, 2012]. La topología presenta la limitación causada por el interruptor principal. Este elemento une ambas partes del circuito acumulando las pérdidas de potencia en el mismo debido a que el voltaje y la corriente de ambos subcircuitos concurren en este punto en particular.

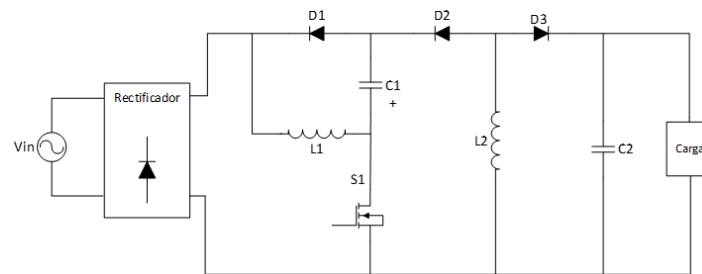


Figura 1 Esquemático del CRED.

Para reducir el esfuerzo en el transistor, circuitos auxiliares como snubbers o clamps pueden ser utilizados. Este tipo de arreglos modifican o limitan las señales de voltaje y corriente. Redes de este tipo pueden llegar a ser activas o pasivas ayudan al interruptor a conseguir una conmutación suave sin alterar las características que presenta el sistema.

Existen, de igual manera, circuitos snubber disipativos y sin pérdidas. Los primeros usan elementos como resistencias en donde se disipa la energía, causando una disminución en la eficiencia del sistema. Las redes sin pérdidas utilizan elementos almacenadores de energía que reciclan la misma dentro del circuito. Sin embargo, existen topologías que pueden causar problemas, como los que utilizan inductancias acopladas. Si éstas no están correctamente acopladas la

inductancia de dispersión puede generar problemas con sobretiros de voltajes [Tseng, 1998], [Marshall, 2006], [Zhao, 2011]; [Mohammadi, 2015], [Konishi, 2007], [Yao, 2002], [Chu, 2010], [Fujiwara, 1999], [Choe, 2014]. Los circuitos sin pérdidas son preferidos por la eficiencia que permiten alcanzar.

En este artículo se presenta un snubber activo de voltaje y un snubber pasivo de corriente para el CRED que permitirá disminuir las pérdidas en el interruptor al modificar el encendido y el apagado del mismo. Es importante recalcar que, a pesar de que se mejora el rendimiento del interruptor al reducir las pérdidas por conmutación, el precio por pagar es que se agregan más elementos al sistema.

La red snubber propuesta, al igual que el CRED son explicados en detalle en la sección 2, mientras que en la sección 3 se presentan resultados de simulación, en la sección 4 resultados experimentales son mostrados y finalmente es redactada una conclusión.

2. Métodos

Convertidor Reductor Elevador Doble

En la figura 1 se muestra el convertidor CRED, conectado a la línea eléctrica, que se compone de dos etapas Reductor-Elevadoras integradas compartiendo el interruptor principal.

La primera etapa consiste de un rectificador de onda completa y un convertidor Reductor-Elevador flotado. Esta primera parte reducirá el rizo de voltaje para la alimentación del segundo subcircuito mientras asegura un factor de potencia elevado debido al inductor que opera en modo de conducción discontinua. El inductor L_1 , el capacitor C_1 y el diodo D_1 conforman la primera etapa. El segundo segmento y la salida de la topología está conformada por el inductor L_2 , el capacitor C_2 y el diodo D_3 . El interruptor S_1 y el diodo D_2 sirven como eslabones para la unión de ambos subcircuitos.

El convertidor consta de tres modos de operación. Los subcircuitos se muestran en la figura 2 mientras que las formas de onda esperadas se observan en la figura 3. Al encender el interruptor S_1 en t_0 se genera el primer modo de operación. El inductor L_1 se carga con el voltaje rectificado de entrada mientras que el capacitor

C_1 descarga la energía almacenada en L_2 y el capacitor C_2 entrega voltaje a la carga. El segundo modo de operación entra en funcionamiento cuando S_1 se apaga en t_2 . En ese instante, el diodo D_1 se polariza directamente permitiendo que el inductor L_1 se descargue sobre el capacitor C_1 y el inductor L_2 se descarga igualmente sobre el capacitor C_2 con la polarización directa del diodo D_3 y el apagado del diodo D_2 . El tiempo D_2T es el tiempo de descarga del inductor L_1 .

El circuito entra en el tercer modo de operación cuando el inductor L_1 se descarga completamente en t_2 , terminando en t_3 . El diodo D_3 se polariza directamente y el voltaje sobre el interruptor S_1 es la suma de los voltajes en C_1 y C_2 .

Las ecuaciones de diseño se pueden encontrar en [Alonso, et. al., 2012].

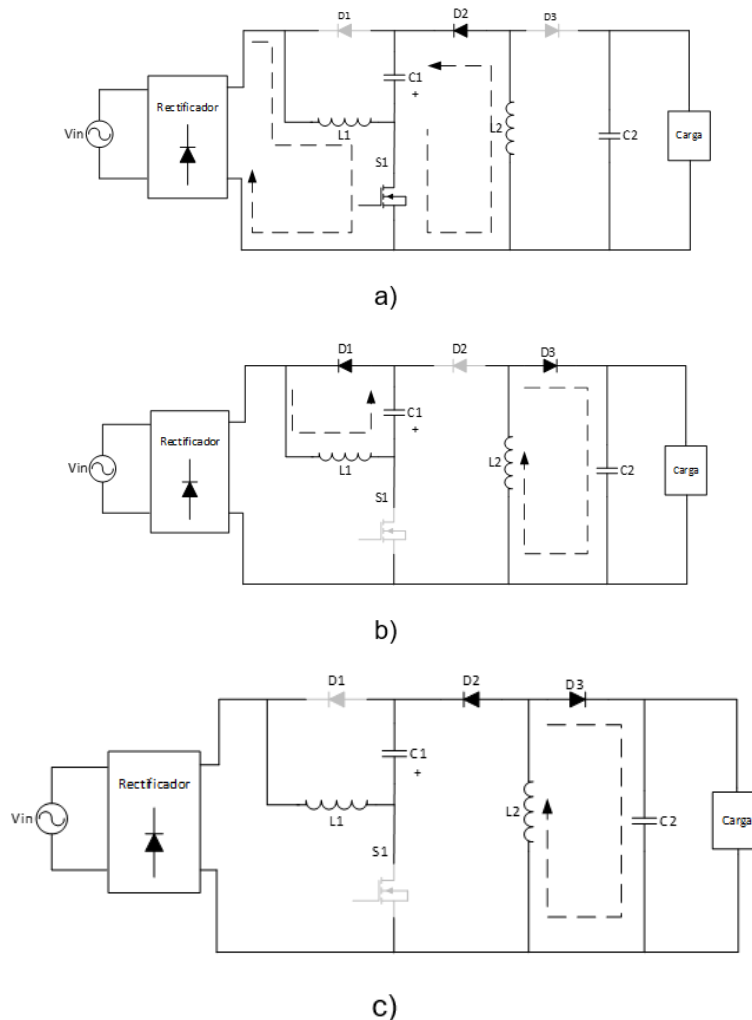


Figura 2 Modos de operación.

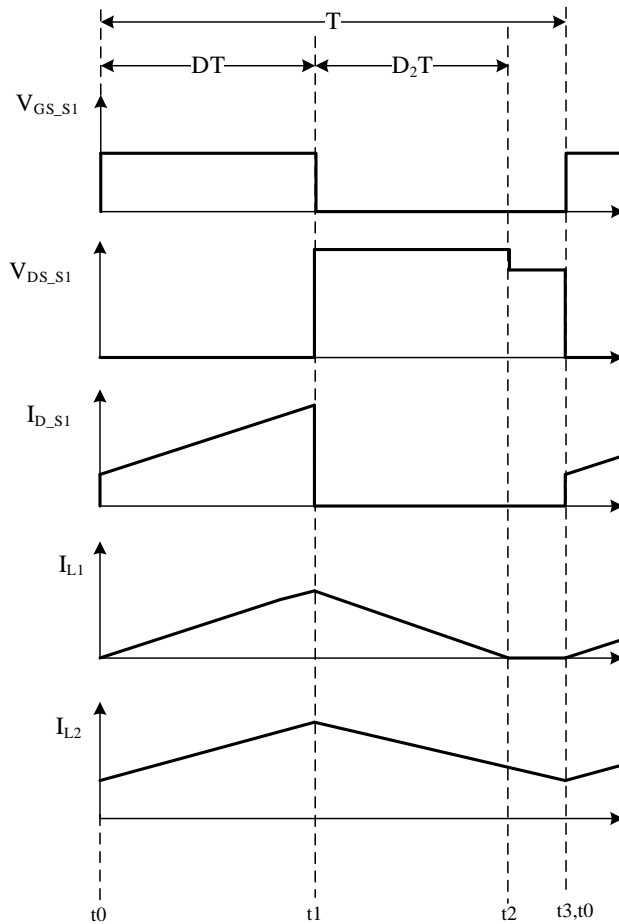


Figura 3 Formas de onda de convertidor.

Con el objetivo de aliviar las pérdidas por conmutación en el interruptor S_1 , se presenta el diseño de redes snubber pasiva y activa, que se muestra en figura 4 y las formas de onda esperadas, que se pueden apreciar en la figura 5.

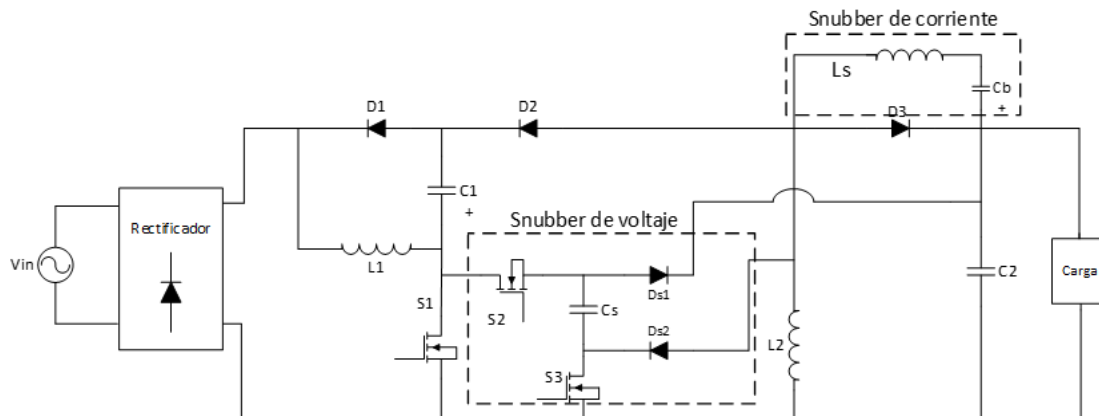


Figura 4 Convertidor Reductor Elevador Doble con redes snubber.

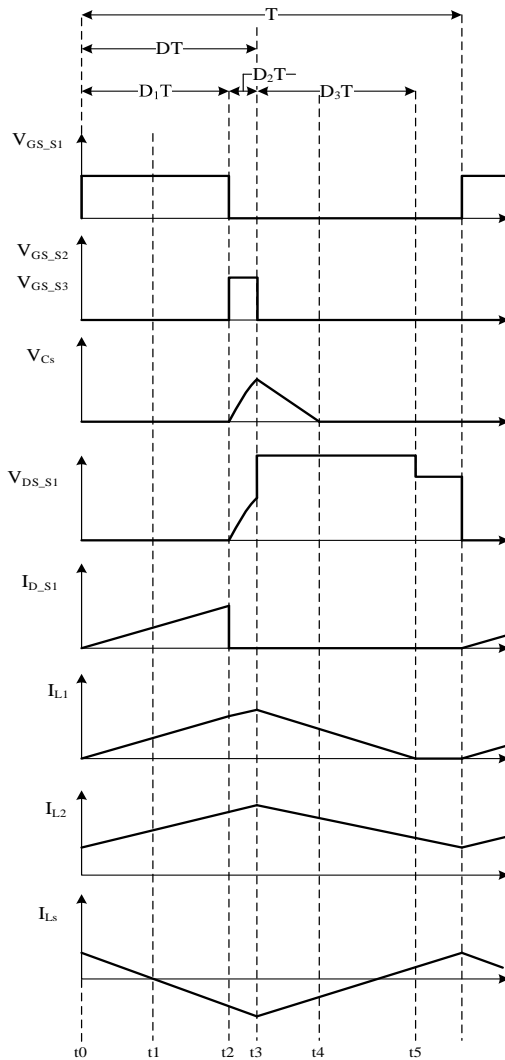


Figura 5 Formas de onda esperadas del convertidor con redes snubber.

Convertidor Reductor-Elevador Doble con snubbers

El snubber activo de voltaje está compuesto de dos interruptores auxiliares, S_2 y S_3 , dos diodos, D_{s1} y D_{s2} , y el capacitor C_s .

El circuito de snubber de voltaje comienza a funcionar en el momento en que S_1 se apaga. En ese instante los interruptores S_2 y S_3 se encienden permitiendo un flujo de corriente a través del capacitor C_s , como se muestra en la figura 6b. El voltaje en el capacitor C_s se incrementa con respecto a las corrientes en L_1 , L_2 , L_s y el ciclo de trabajo D_2 , ecuación 1.

$$C_s = \frac{(i_{L1} + i_{L2} - i_{Ls})D_2T}{\Delta V_{C_s}} \quad (1)$$

Donde ΔV_{Cs} es el voltaje máximo al que se cargará dicho capacitor y D_2T es el tiempo en el que los interruptores S_2 y S_3 permanecen encendidos.

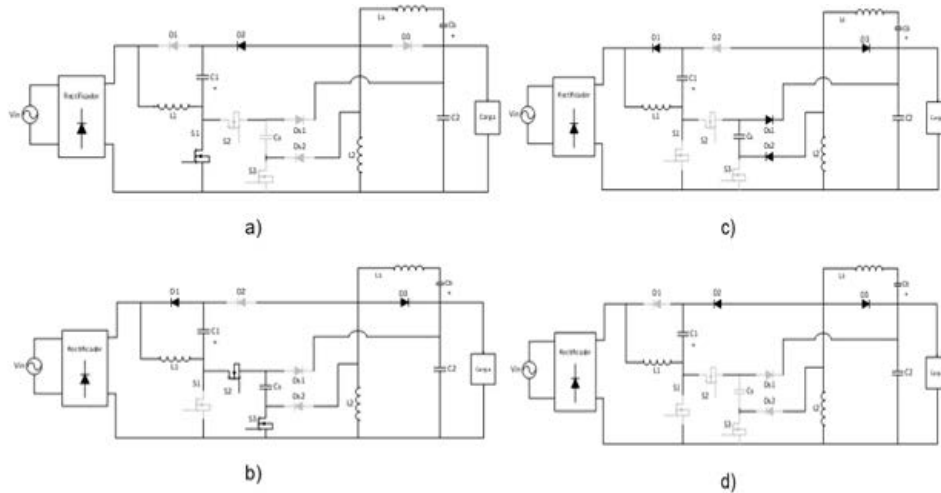


Figura 6 Modos de operación del circuito con snubbers.

Definiendo el valor de la capacitancia de C_s , se encuentra el tiempo necesario de encendido para los interruptores S_2 y S_3 como se muestra en ecuación 2.

$$D_2 = \frac{C_s \Delta V_{Cs}}{(i_{L1} + i_{L2} - i_{Ls})T} \quad (2)$$

El capacitor C_s se cargará al valor del voltaje de salida V_{C2} debido a la activación de D_{s1} a menos que los interruptores sean apagados antes de alcanzar este voltaje debido al cambio en baja frecuencia de la corriente en el inductor L_1 . Los interruptores auxiliares deben sincronizar su activación con el interruptor principal. Cuando el interruptor S_1 es apagado los interruptores auxiliares S_2 y S_3 son encendidos y el inductor L_1 permanece en el estado de carga. Para el ciclo de trabajo actual del interruptor se deben considerar ambas señales de control, del interruptor principal y de los interruptores auxiliares, ecuación 3.

$$D = D_1 + D_2 \quad (3)$$

Una vez apagados los tres interruptores el diodo D_3 se enciende y la corriente del inductor L_2 fluye a través de los diodos D_{s1} y D_{s2} descargando el capacitor C_s figura 6c.

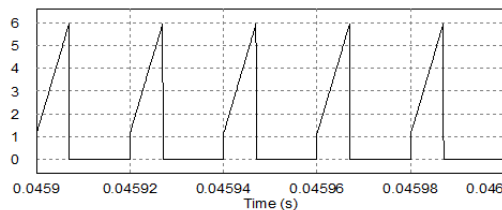
El snubber de corriente se utiliza de manera que se anule la corriente del inductor L_2 , que opera en MCC, sobre el interruptor S_1 . El circuito, conformado por L_s y C_b , permite la carga del inductor L_2 mientras el inductor L_s se descarga, figura 6a, figura 6d. Debido a que los voltajes promedios de C_b y C_2 son iguales, el inductor L_s se ve afectado únicamente por el voltaje de capacitor C_1 , ecuación 4.

$$L_s = \frac{V_{C1}DT}{\Delta I_{L_s}} \quad (4)$$

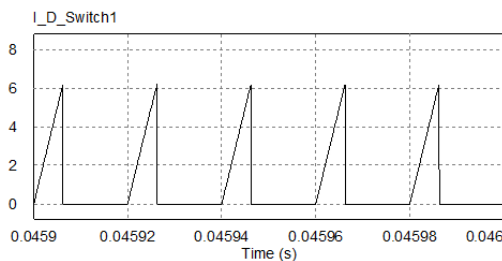
En el tiempo t_1 el flujo de corriente carga al inductor L_s nuevamente pero con una corriente que fluye en sentido contrario.

3. Resultados

En esta sección se presentan los resultados de simulación en PSim. Cabe mencionar que la acción de las resistencias en serie equivalentes de los elementos inductivos y capacitivos tienen un efecto despreciable en la simulación. En la figura 7 se muestra la corriente a través del interruptor S_1 . En a) se observa la corriente en el modo conmutación dura que es el resultado de la suma de las corrientes de los inductores L_1 y L_2 , mientras que en b) se presenta la técnica de conmutación suave en la que se minimiza el efecto de la corriente proveniente del inductor L_2 .



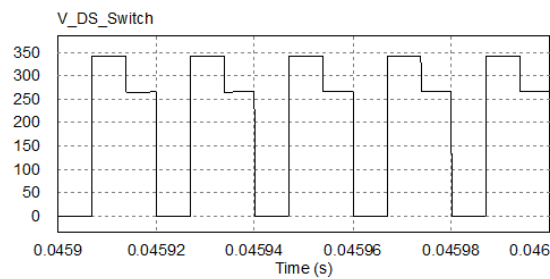
a) Conmutación dura



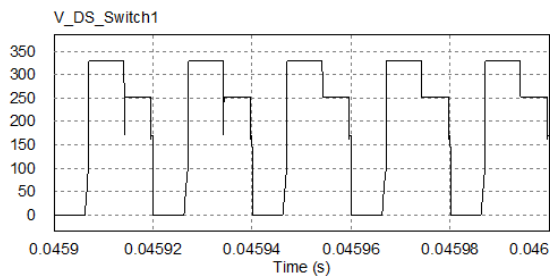
b) Conmutación suave

Figura 7 Corriente en Drain de interruptor S_1 .

La figura 8 muestra el voltaje Drain-Source del interruptor S_1 . En figura 8a, se observa el voltaje en conmutación dura, cuando el voltaje inicia en un valor elevado, por su parte, la figura 8b despliega el voltaje en conmutación suave en donde éste tiene una pendiente menos pronunciada debido a la carga del capacitor C_s y el hecho de que el mismo capacitor se encuentra en paralelo con el interruptor S_1 .



a)



b)

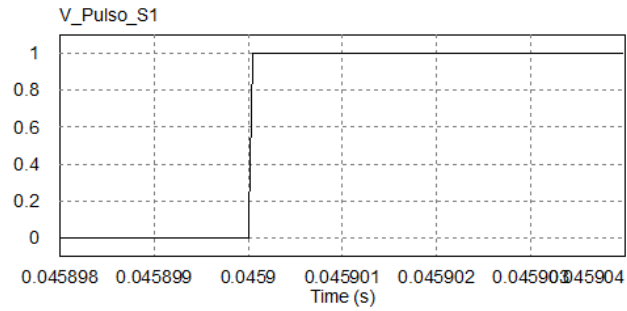
Figura 8 Voltaje Drain-Source de interruptor S_1 .

En la figura 9 se muestra un acercamiento en la señal de activación del interruptor S_1 en su flanco positivo en a) y en b) el voltaje y la corriente en el encendido del mismo interruptor. En la figura 10 se puede apreciar la señal de activación del interruptor S_1 en su flanco descendente en a) y en b) el voltaje y la corriente en el apagado de dicho interruptor.

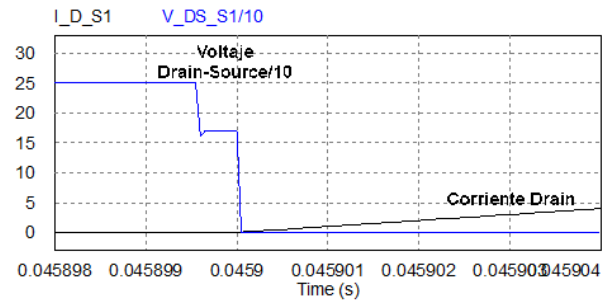
En la figura 11 se observan las formas de la corriente en los inductores L_2 y L_s en la topología con la red snubber.

La figura 12 muestra el voltaje del capacitor C_s mientras que en la figura 13 se observan las señales de activación de los interruptores S_1 , S_2 y S_3 .

En la tabla 1 se aprecian los valores utilizados para la simulación del convertidor.

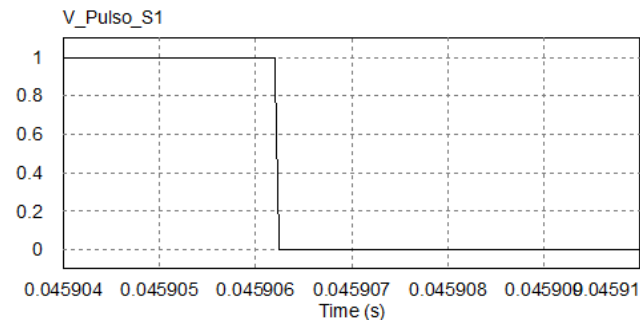


a)

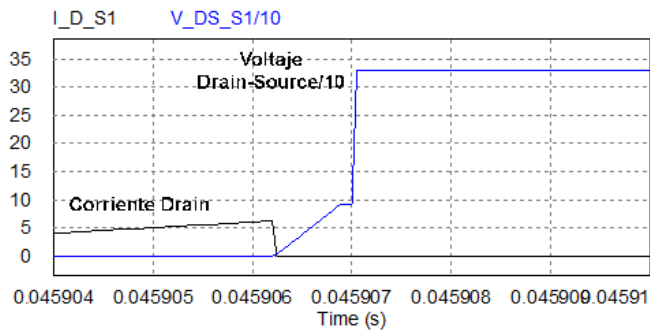


b)

Figura 9 Flanco ascendente de señal de activación interruptor S_1 y voltaje drain-source.

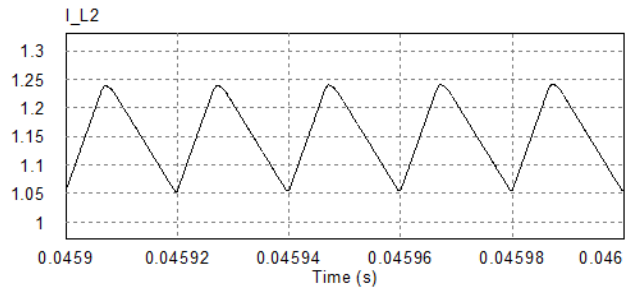


a)

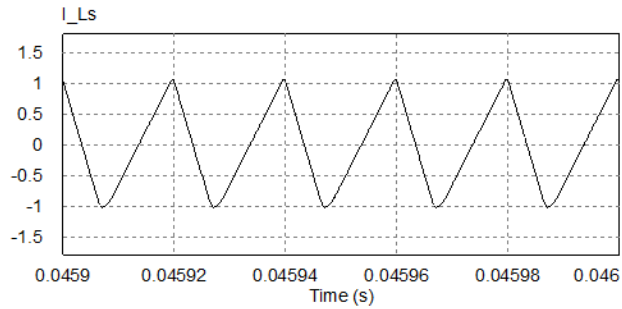


b)

Figura 10 Flanco descendente de señal de activación interruptor S_1 y voltaje drain-source.



a)



b)

Figura 11 Corrientes de los inductores.

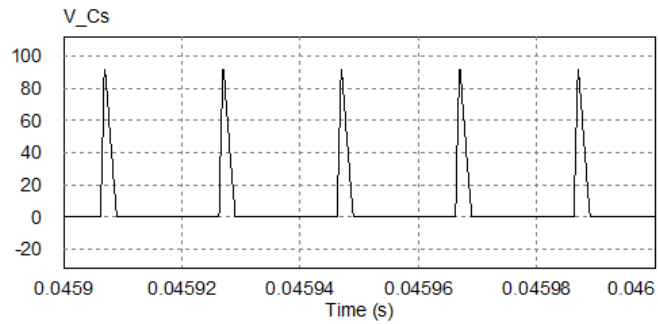


Figura 12 Carga y descarga del capacitor Cs.

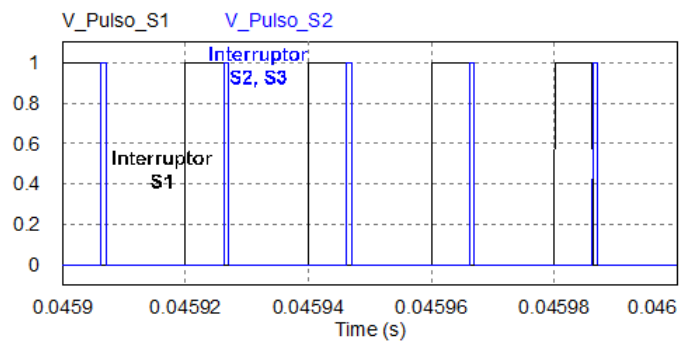


Figura 13 Señales de activación del interruptor.

Tabla 1 Valores de los elementos utilizados en la simulación del convertidor.

Elementos	Valor	Unidades
L1	256	μH
L2	5.6	mH
Ls	500	μH
C1	100	μF
C2	10	μF
Cs	47	nF
Cb	10	μF
Carga	122	Ω
Vin	120, 60	Vrms, Hz
F_interrupt	50	kHz

4. Discusión

La presente técnica de reducción de esfuerzos permite disminuir las pérdidas por conmutación en el CRED debido a la modificación de las señales de voltaje y corriente, permitiendo alcanzar la conmutación con corriente cero y voltaje cero.

5. Conclusiones

Los convertidores de potencia con aplicaciones para iluminación LED deben satisfacer ciertas necesidades como un factor de potencia elevado o un rizo de corriente pequeño en la salida. El convertidor que aquí se presenta cumple ambas características antes mencionadas, con la limitación de la concentración de esfuerzos debido a su naturaleza como topología integrada. La solución propuesta es la implementación de una red snubber sin pérdidas que permita la conmutación suave del interruptor y, con esto, disminuir las pérdidas por conmutación sin afectar el rendimiento del convertidor.

Los resultados de simulación corroboran el funcionamiento del circuito al modificar las señales de conmutación dura a conmutación suave. Ésta también muestra que al agregar elementos como las resistencias en serie equivalentes de los componentes utilizados no se modifica la operación de la topología.

Es importante recalcar que la conmutación suave permite mejorar el rendimiento de la topología al disminuir las pérdidas por conmutación. Sin embargo, al agregar más elementos se espera una disminución en la eficiencia y un incremento en el costo de un prototipo.

6. Bibliografía y Referencias

- [1] C. Gobatto, S. V. Kohler, I. H. de Souza, G. W. Denardin, J. de Pelegrini Lopes, Integrated topology of DC converter for street lighting system based on LED modular drivers, 2016 12th IEEE International Conference on Industry Applications (INDUSCON), Curitiba, pp. 1-6, 2016.
- [2] Ching-Jung Tseng, Chern-Lin Chen. A passive lossless snubber cell for nonisolated PWM DC/DC converters, *IEEE Transactions on Industrial Electronics*, vol. 45, no. 4, pp. 593-601, 1998.
- [3] Enhui Chu, Weiyu Hu, Jinxing Gong, Rui Hou, Mutsuo Nakaoka, A novel high frequency ZVS-PWM boost DC-DC converter with auxiliary resonant snubber, 2010 International Conference on Communications, Circuits and Systems (ICCCAS), Chengdu, pp. 576-580, 2010.
- [4] H. C. Kim, M. C. Choi, S. Kim, D. K. Jeong., An AC–DC LED Driver With a Two-Parallel Inverted Buck Topology for Reducing the Light Flicker in Lighting Applications to Low-Risk Levels, *IEEE Transactions on Power Electronics*, vol. 32, no. 5, pp. 3879-3891, 2017.
- [5] H. J. Choe, Y. C. Chung, C. H. Sung, J. J. Yun, B. Kang, Passive Snubber for Reducing Switching-Power Losses of an IGBT in a DC–DC Boost Converter, *IEEE Transactions on Power Electronics*, vol. 29, no. 12, pp. 6332-6341, 2014.
- [6] J. M. Alonso, J. Vina, D. G., Vaquero, G. Martinez, R. Osorio. Analysis and Design of the Integrated Double Buck–Boost Converter as a High-Power-Factor Driver for Power-LED Lamps. *IEEE Transactions on Industrial Electronics*, vol. 59, no. 4, pp. 1689-1697, 2012.
- [7] J. Marshall, M. Kazerani, A Novel Lossless Snubber for Boost Converters, 2006 IEEE International Symposium on Industrial Electronics, Montreal, Que., pp. 1030-1035, 2006.
- [8] Kaiwei Yao, F. C. Lee, Yu Meng, Jia Wei, Tapped-inductor buck converters with a lossless clamp circuit, APEC. Seventeenth Annual IEEE Applied Power Electronics Conference and Exposition (Cat. No.02CH37335), Dallas, TX, pp. 693-698, 2002.

- [9] K. Fujiwara, H. Nomura, A novel lossless passive snubber for soft-switching boost-type converters, in *IEEE Transactions on Power Electronics*, vol. 14, no. 6, Noviembre 1999, pp. 1065-1069.
- [10] K. Zhou, J. G. Zhang, S. Yuvarajan, D. F. Weng. Quasi-Active Power Factor Correction Circuit for HB LED Driver. *IEEE Transactions on Power Electronics*, vol. 23, no. 3, pp. 1410-1415, 2008.
- [11] M. Mohammadi, E. Adib, M. R. Yazdani, Family of Soft-Switching Single-Switch PWM Converters With Lossless Passive Snubber, *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3473-3481, 2015.
- [12] Y. Zhao, W. Li, Y. Deng, X. He., High step-up boost converter with passive lossless clamp circuit for non-isolated high step-up applications, *IET Power Electronics*, vol. 4, no. 8, pp. 851-859, 2011.
- [13] Yoshihiro Konishi, Yung-Fu Huang, Soft-switching buck boost converter using passive snubber composed of pulse current regenerative resonant circuit, *INTELEC 07 - 29th International Telecommunications Energy Conference*, Roma, pp. 886-890, 2007.
- [14] Zhang Bo, Yang Xu, Xu Ming, Chen Qiaoliang, Wang Zhaoan. Design of Boost-Flyback Single-Stage PFC converter for LED power supply without electrolytic capacitor for energy-storage. 2009 *IEEE 6th International Power Electronics and Motion Control Conference*, Wuhan, pp. 1668-1671, 2009.

DISEÑO Y SIMULACIÓN DE RODILLA MECÁNICA MONOCÉNTRICA

José Alexis Hernández Aguilar

Universidad Veracruzana, Campus Xalapa

alexis_9319@hotmail.com

Ervin Jesús Álvarez Sánchez

Universidad Veracruzana, Campus Xalapa

eralvarez@uv.mx

Andrés López Velázquez

Universidad Veracruzana, Campus Xalapa

andlopez@uv.mx

Resumen

En este trabajo se presenta una propuesta de diseño mecánico de una rodilla monocéntrica, con la finalidad de que posteriormente sea implementada en una prótesis transfemoral que pueda ser utilizada por personas de bajos recursos. El diseño mecánico de la articulación de rodilla se realiza mediante el software Autodesk Inventor ® y tiene como requisito alcanzar como mínimo un ángulo de flexión de 93°, que es uno de los requerimientos básicos de una prótesis para su adecuado funcionamiento. Mediante simulaciones numéricas, se verifica el ángulo máximo que puede flexionarse la articulación de rodilla, así como las fuerzas que actúan sobre la rodilla si se construye de plástico ABS utilizando una impresora 3D. Los resultados obtenidos muestran las ventajas y desventajas de utilizar este material en una prótesis de rodilla.

Palabras Claves: Articulación, diseño mecánico, prótesis, rodilla, transfemoral.

Abstract

In this paper a mechanical design for a monocentric prosthetic knee that will be implemented in a transfemoral prosthesis for low-income people is presented. The

knee joint mechanical design is made by Autodesk Inventor ® and has a requirement to reach at least 93° bending angle, which is one of the basic requirements of a prosthesis for their proper functioning. By means of numerical simulations, the maximum angle reached by the knee joint is verified. Also the forces acting on the prosthetic knee if is build in ABS plastic, using a 3D printer, are simulated. The results show the advantages and disadvantages of employing this material in a prosthetic knee.

Keywords: *Articulation, knee, mechanical design, prosthesis, transfemoral.*

1. Introducción

La prótesis de articulación de rodilla es uno de los componentes protésicos más complejos que existen, ya que imitar el comportamiento de una rodilla humana se dificulta por su estructura anatómica, además de que es de vital importancia para que el paciente vuelva a caminar. Es por esto, que es necesario el seleccionar la más adecuada, ya que existen distintas clasificaciones acordes a las necesidades del paciente.

La principal clasificación de las prótesis de rodilla radica en su nivel de control, ya que pueden ser mioeléctricas o mecánicas, ambas con ventajas y desventajas. Las prótesis mioeléctricas se basan en el concepto de que los músculos del cuerpo al contraerse o flexionarse, producen una pequeña señal eléctrica de entre 5 a 20 V, que es creada por la interacción química en el cuerpo. Sin embargo, debido al nivel de instrumentación que se requiere [Torrealba et al, 2010], el costo que llegan a tener este tipo de prótesis es considerable [Murthy, 2015] y requieren de un plan de mantenimiento regular [González, 2005].

Las prótesis mecánicas son las más accesibles para los pacientes, debido a que necesitan menor mantenimiento que las prótesis mioeléctricas y tienen un menor costo, sin embargo, se requiere ejercer una fuerza mayor para realizar movimientos. Además, en la fase de apoyo del talón la rodilla no se debe flexionar y en la fase final, el usuario de la prótesis debe ser capaz de flexionarla.

Otras formas de clasificar las prótesis son a través del movimiento cinemático que presenta el eje de rotación, ya que puede ser monocéntrico o policéntrico [Näder,

2003]. Las rodillas policéntricas son las que realiza un movimiento combinado de giro y traslación por su multiaxialidad, en el cual el punto de giro cambia la posición dependiendo en donde se encuentre en la posición de flexión. [Näder, 2003]. Este tipo de rodillas tienen usualmente un mecanismo de 4 barras para proporcionar mayor estabilidad al paciente [Menghini, 2015], sin embargo la mayoría están contruidos de materiales metálicos, lo cual se ve reflejado en el incremento de peso de las prótesis, lo que hace primordial realizar simulaciones numéricas para evaluar la resistencia a cargas utilizando el método de elementos finitos [Vega, 2015], aunque cabe aclarar que estos análisis de esfuerzos también deben realizarse aún cuando los materiales utilizados sean plásticos [Rodríguez et al, 2016].

En el caso de las rodillas monocéntricas, estas presentan un movimiento similar al de una bisagra, es decir, la rotación se presenta solo en el plano sagital y su eje articular está ubicado atrás del eje de carga, con la finalidad de evitar que la rodilla se flexione cuando el pie toque el suelo por primera vez [García, 2007]. Además, este tipo de rodilla contiene un elemento que ayuda a absorber la energía durante la fase de apoyo y extender la prótesis durante la fase de balanceo [Torrealba et al, 2010]. La geometría básica de este mecanismo es un triángulo con dos lados constantes, mientras que el tercero, que es el que generalmente realiza la tarea de extender la pierna, varía de forma proporcional al ángulo de flexo-extensión [Nájera, 2013]. Lo anterior, se puede reforzar mediante un análisis cinemático para determinar las trayectorias del mecanismo, los ángulos y desplazamiento del actuador utilizado en la rodilla [Valencia et al, 2017]. Este tipo de clasificaciones cinemáticas se encuentra relacionado directamente con el ciclo de la marcha, el cual es un proceso cíclico en donde el centro de gravedad se mueve hacia adelante por efecto del movimiento armónico de las extremidades inferiores e indirectamente con ayuda de las extremidades superiores al balancear los brazos. El ciclo de la marcha puede dividirse en dos fases: apoyo y balanceo, o en tres intervalos sobre el plano sagital, en los cuales se da a conocer el ángulo que logra la rodilla [Sánchez et al, 2006]. Estos intervalos representan el movimiento de las articulaciones y son:

- Entre el contacto del talón con el suelo y el punto medio de apoyo. Este intervalo se subdivide en 4 etapas, las cuales se explican en la tabla 1 y se muestran en la figura 1.
- Entre el apoyo medio y el despegue del pie del suelo. Este intervalo se subdivide en 3 fases, dadas en la tabla 2 y mostradas en la figura 2.
- En la etapa de balanceo. Se subdivide en dos etapas, dadas en la tabla 3 y mostradas en la figura 3.

Tabla 1 Movimientos de la rodilla en el intervalo 1.

Etapa	Movimiento de la rodilla
Inmediatamente antes del contacto del talón con el suelo	La articulación de la rodilla se encuentra en completa extensión.
Simultáneamente con el contacto del talón con el suelo.	La articulación de la rodilla comienza a flexionarse y continúa hasta que la planta del pie esté plana en el suelo.
Inmediatamente después de haber alcanzado la posición plana del pie.	La rodilla tiene aproximadamente un ángulo de 20° de flexión y comienza a extenderse.
En el apoyo medio	La rodilla tiene aproximadamente un ángulo de 10° de flexión y continúa extendiéndose.

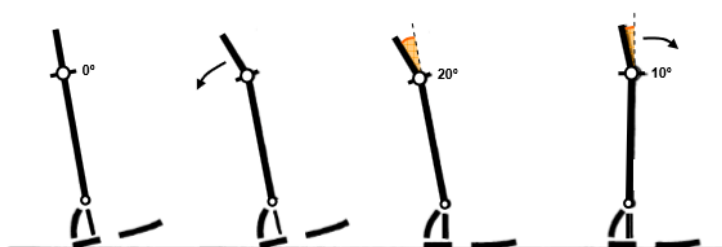


Figura 1 Primer intervalo de movimiento de la rodilla [Hernández, 2008].

Tabla 2 Movimientos de la rodilla en el intervalo 2.

Etapa	Movimiento de la rodilla
En el apoyo medio	La rodilla tiene aproximadamente un ángulo de 10° de flexión y continúa extendiéndose.
Inmediatamente antes de que el talón pierda contacto con el suelo.	La rodilla está a 4° de la extensión completa.
Entre el despegue del talón y el de los dedos.	La articulación de la rodilla se mueve de una extensión casi completa a 40° de flexión.

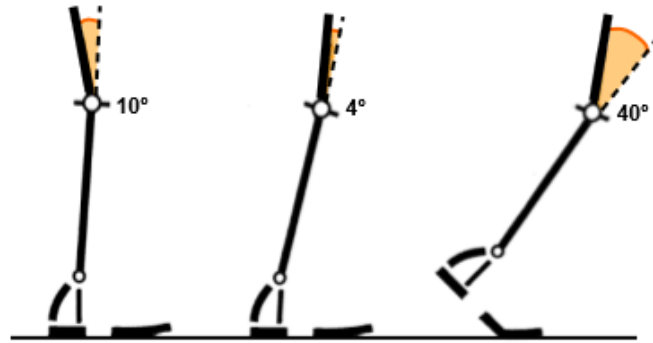


Figura 2 Segundo intervalo de movimiento de la rodilla [Hernández, 2008]

Tabla 3 Movimientos de la rodilla en el intervalo 3.

Etapa	Movimiento de la rodilla
Entre el despegue del pie y la parte media de la etapa de balanceo.	La rodilla se flexiona de una posición inicial de aproximadamente 40° a un ángulo de máxima flexión de aproximadamente 65°
Entre la parte media de la etapa de balanceo y el contacto del talón.	La rodilla se extiende casi completamente hasta el último instante de la etapa de balanceo.

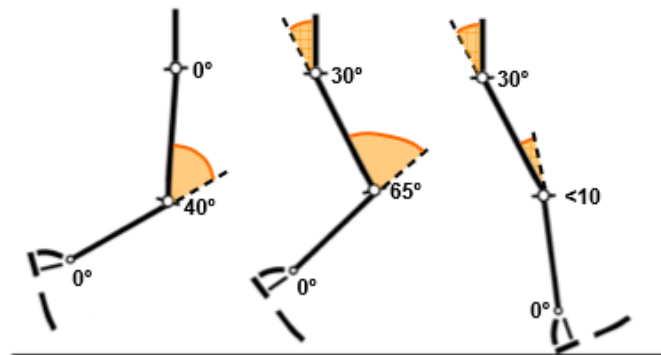


Figura 3 Tercer intervalo de movimiento de la rodilla [Hernández, 2008].

Los movimientos realizados por la rodilla durante la marcha requieren de una flexión-extensión que abarca un rango de 0° a 65° aproximadamente, sin embargo existen otros movimientos relacionados con las actividades cotidianas, los cuales requieren de otros ángulos de flexión para llevarse a cabo. El rango angular y la actividad que lo requiere se muestran en la tabla 4.

Tabla 4 Rangos angulares para la rodilla dependiendo de la actividad [Nordin, 2001].

Actividad	Rango angular
Caminar	0 - 67°
Subir escalones	0 - 83°
Bajar escalones	0 - 90°
Sentarse	0 - 93°
Probarse un zapato	0 - 106°
Subir un obstáculo	0 - 117°

Utilizando la información expuesta y estos rangos lo siguiente es proponer un diseño mecánico para la prótesis de rodilla.

2. Métodos

En esta sección se presenta la propuesta de diseño de una rodilla monocéntrica, la cual cumple con los requerimientos para los intervalos de marcha y rangos angulares, incluyendo las características mecánicas que deben cumplirse para un adecuado funcionamiento de la marcha. Con la finalidad de llevar a cabo un adecuado diseño mecánico, se utilizó el software Inventor®, mediante el cual se puede realizar el modelado de sólidos en 3D, basándose las características que se requerían que tuviera la prótesis, así como el funcionamiento que realiza cada componente.

Para iniciar con la propuesta de diseño de la rodilla protésica, primero se tomó en cuenta la funcionalidad de la rodilla humana para lograr los movimientos tradicionales de caminar, bajar y subir escaleras y sentarse, por lo que el ángulo de flexión elegido fue de 0 a 105°. Por otra parte, por cuestiones de seguridad para el futuro usuario de la prótesis, se limitó al movimiento en el plano sagital, proponiéndose además que la rodilla estuviera conformada por tres componentes: el flexor-extensor, el superior y el inferior, los cuales se describen a continuación.

Componente Flexor-Extensor

El primer elemento que se diseñó es el componente flexor-extensor, de forma que pudiera proporcionar un amortiguamiento a los impactos y que estos no se transmitieran hacia el socket-muñón del paciente. Dimensionar este componente fue primordial, ya que mediante el mismo se establecieron los espacios que

requieren los componentes inferior y superior. Se seleccionó el modo de amortiguación considerado como el más sencillo, ya que no depende de la velocidad de movimiento, además de que no contiene ningún fluido, y se propuso el uso de un resorte como elemento de restitución para la posición. En la figura 4 se muestra el diseño propuesto para el elemento flexor-extensor. Este diseño comprende una carcasa (figura 4a) que protege el vástago (figura 4b) a la vez que le permite deslizarse en su interior, ambas partes tiene soportes para alojar al resorte (figura 4c) y limitarlo por ambos lados, así como pestañas con perforaciones para fijación a los componentes superior e inferior de la rodilla.

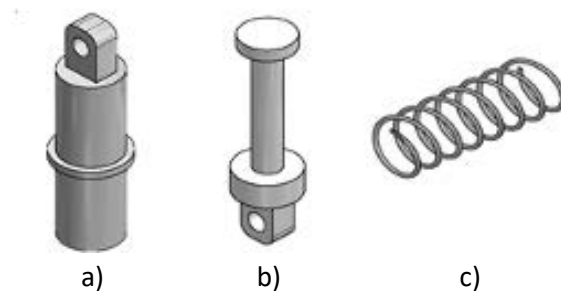


Figura 4 Elementos del componente flexor-extensor.

Componente Superior

Este componente se ubica en la parte superior de la rodilla y tiene como finalidad emular la acción de flexión del fémur en la rodilla. En la figura 5 se muestra la vista en isométrico del diseño propuesto.

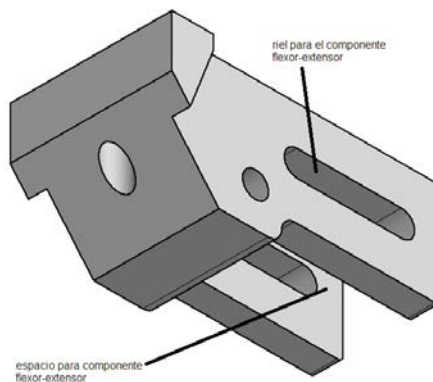


Figura 5 Vista en isométrico del elemento superior.

El elemento superior se diseñó para cumplir con las características requerida para la adaptación y funcionamiento de la rodilla, las cuales son pestañas para seguridad en el bloqueo, soporte para el componente flexor-extensor, eje articular y aditamento superior. A continuación, se describen estas características:

- Pestañas para seguridad en el bloqueo. En el diseño se estableció una pestaña de seguridad de cada lado, de tal manera que se bloquea el regreso del componente superior, manteniéndolo a un ángulo de 0° respecto del eje horizontal. La forma dos pestañas en cada lado, que funcionan para bloquear el regreso del componente superior cuando este se encuentra en flexión. Manteniendo el componente superior a 0° . Son de forma geométricamente poligonal, donde el lado más largo es el que se acopla con el componente inferior, en la figura 6 se muestra una vista lateral del componente superior, en donde se pueden apreciar las pestanas de seguridad.

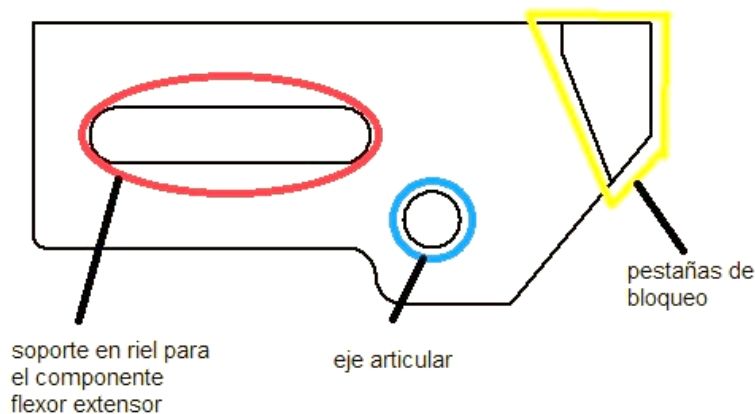


Figura 6 Vista lateral del componente superior.

- Soporte para el componente flexor-extensor. Este soporte se diseñó de tal manera que se pudiera modificar la posición angular del componente flexor-extensor, por lo que se tiene una ranura alarga que permite acercarlo o alejarlo del eje articular. Las pestañas con perforaciones, del elemento flexor-extensor, se unen con el elemento superior por medio de un perno pasado.

- Eje articular. El eje articular, es el punto donde la rodilla realiza el movimiento de flexión y no cambia de lugar sin importar el movimiento de flexión que se realice. Este eje se ubicó detrás del eje de carga, tal y como se requiere en los diseños de prótesis. Por otra parte, el eje articular, también tiene como función unir el componente superior con el componente inferior, mediante un perno pasado.
- Aditamento superior. Aunque este aditamento no forma parte del elemento superior, es importante su ubicación superior, ya que es una pirámide que sirve para alinear la parte del socket (componente de la prótesis que recubre al muñón) con la parte inferior de la pierna, la forma de unir la pirámide con el componente superior es mediante tornillos, por ese motivo en el componente superior se ubicaron ranura para acople de tornillos.

Componente Inferior

El componente inferior fue diseñado con la finalidad de emular la parte superior de la tibia, permitiendo uniones con el elemento superior y el elemento flexor-extensor. En la figura 7 se muestran las vistas lateral y frontal de este componente.

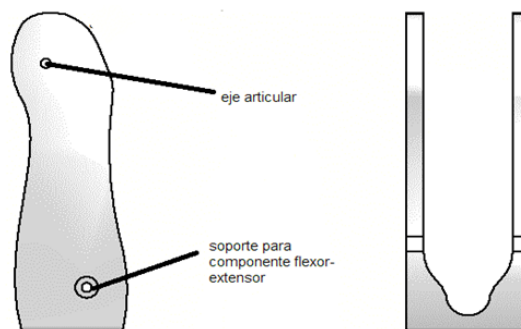


Figura 7 Vistas del elemento inferior.

Este componente también presenta orificios para el eje articular y para el componente flexor-extensor. La forma en que se diseñó fue para darle mayor estética y reducir el peso total, además se incluyeron ranuras en la parte posterior para evitar que los pernos no obstruyan la flexión total alcanzable. En la figura 8

se muestra una vista isométrica del elemento inferior, en donde se observa la ubicación de las ranuras.

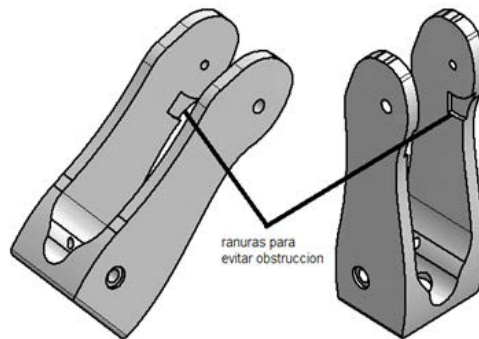


Figura 8 Vista isométrica del componente inferior.

Ensamble de la Rodilla

Después de haber definido los elementos que estarán en la rodilla mecánica, se inicia el ensamble con los tres componentes, donde se utilizan pernos para las uniones, en la figura 9 se muestran la vista lateral y frontal del ensamble, donde con una línea azul se indica el eje de carga, observando que en la vista lateral esta adelante del eje articular y en la vista frontal se encuentra centrado.

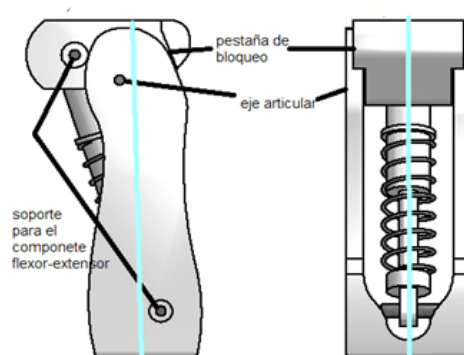


Figura 9 Ensamble de la rodilla.

Una vez que ya se tiene el diseño de los componentes de la rodilla y su respectivo ensamble, lo siguiente es realizar simulaciones numéricas para corroborar que se logra el movimiento de flexión, para posteriormente verificar los resultados que se obtienen bajo una carga.

3. Resultados

En esta sección se muestran los resultados obtenidos mediante el software Autodesk Inventor ® para la simulación de flexión y para la aplicación de carga sobre la rodilla diseñada.

Movimiento de Flexión

El movimiento en el plano sagital, ya que es donde realiza el movimiento de flexión y extensión. En la figura 10 se muestra una vista lateral del prototipo, en donde se muestran la posición de reposo a 0° y la posición flexionada a 105.5°.

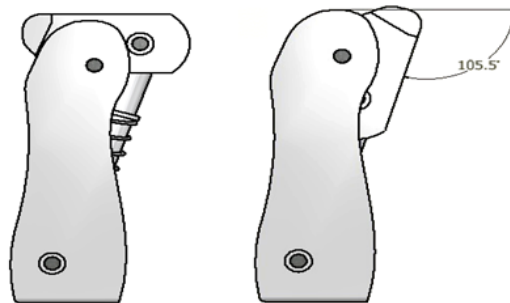


Figura 10 Vista lateral de la rodilla.

Simulación de Tensión

Para llevar a cabo esta simulación, la rodilla es sometida a una fuerza aproximada de 734 N, lo que sería equivalente a una persona cercana a los 75 kg. Se realiza el análisis de tensión aplicando la fuerza en el componente superior de la rodilla mientras se mantiene fijo el componente inferior. Cabe mencionar que el material utilizado para simular los pernos es acero.

Los resultados que arroja la simulación son: tensión de Von Mises, 1a tensión principal, 3a tensión principal, desplazamiento y factor de seguridad. A continuación, se muestran las gráficas que arroja la simulación:

- Tensión de Von Mises. En la figura 11 se muestran la vista posterior y lateral de la rodilla, en donde se observa que el rango de tensión simulada está entre 0 MPa (azul fuerte) y 135.6 MPa (Rojo).

La simulación muestra que desde una vista lateral toda la superficie tiene una tensión de 0 MPa, mientras que en la vista posterior se aprecia que las

zonas que se encuentran a mayor tensión son los pernos, lo cuales conectan el componente flexor-extensor con los componentes superior e inferior. Estas zonas en color azul claro se encuentran indicadas mediante círculos rojos.

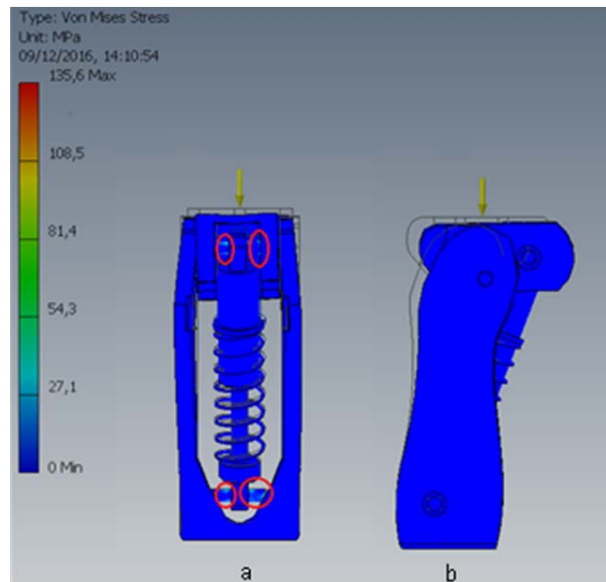


Figura 11 Vista a. posterior y b. frontal.

- 1a tensión principal. Los resultados obtenidos se muestran en la figura 12, donde se observa que en la parte izquierda la escala de colores el valor mínimo de tensión está en color azul fuerte y equivale -24.4 MPa, mientras que el máximo valor de tensión se indica en color rojo y equivale 135.6 MPa, en la figura se muestra la rodilla en una vista posterior en donde la mayor parte de la rodilla se encuentra con tensión negativa (compresión), en las zonas marcadas en una circunferencia roja existe una mayor tensión ya que se encuentra de color azul claro.
- 3a tensión principal. En la figura 13 se observa el resultado que arroja la simulación para la 3a tensión principal. En la parte izquierda de la imagen, la escala de colores muestra que el rango mínimo es de -134 MPa, color azul fuerte, mientras que el rango máximo es de 134.5 MPa, color rojo. Las zonas marcadas con un círculo y un rectángulo azul es donde se concentra la mayor compresión.

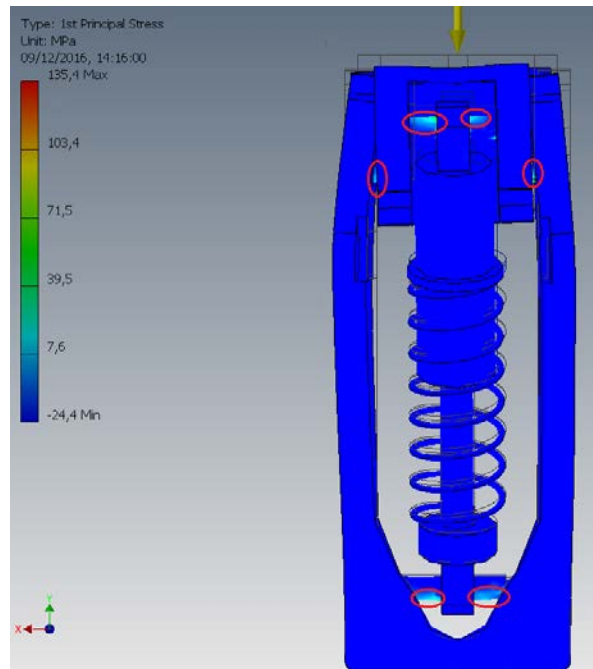


Figura 12 Primera tensión principal.

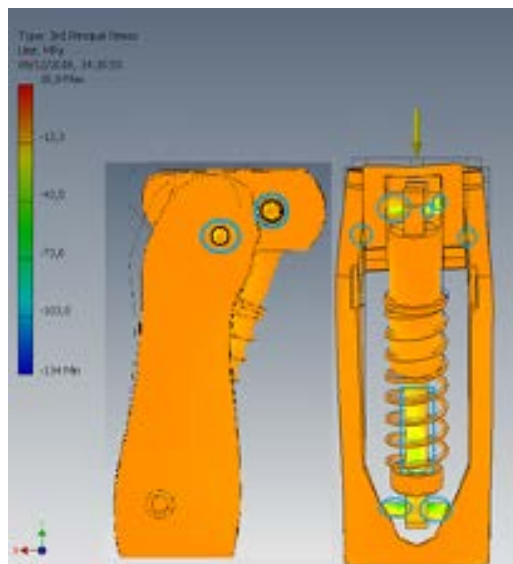


Figura 13 Tercera tensión principal, vistas lateral y posterior.

- Desplazamiento. El desplazamiento es mostrado en la figura 14, en donde en la parte izquierda de la figura la escala de colores muestra un rango de 0 a 0.2894 mm, en color azul fuerte y rojo, respectivamente. Se puede observar que el mayor desplazamiento sucede en la parte superior de la rodilla, siendo los puntos más afectados las pestañas de bloqueo y la parte

más elevada del componente inferior, mientras que en la parte inferior de la rodilla no se aprecia que exista desplazamiento.

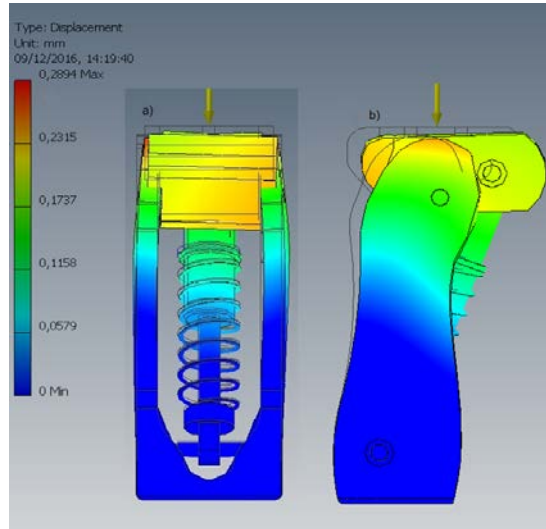


Figura 14 Desplazamiento, vistas frontal y lateral.

- Factor de seguridad. El factor de seguridad aplicado se muestra en la figura 15, donde la escala de colores tiene un rango en el cual el mínimo es 0, en color rojo, y el máximo es de 15 en color azul fuerte. Se observa que en las zonas marcadas por circunferencia naranja se tiene menor factor de seguridad que el resto de la rodilla.

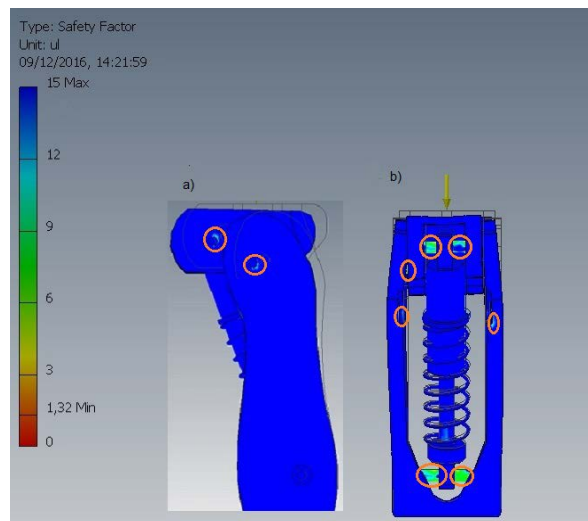


Figura 15 Factor de seguridad, vistas lateral y posterior.

Una vez que se han llevado a cabo las simulaciones de la rodilla diseñada, bajo la acción de una fuerza constante, lo siguiente es discutir los resultados obtenidos.

4. Discusión

Mediante la primera simulación para la flexión, se pudo comprobar que se logra el ángulo sobre el plano sagital que permitirá a una persona realizar los distintos movimientos, además de que se verificó que las pestañas de bloqueo actual adecuadamente para limitar la rotación dentro de los límites de diseño.

En lo que respecta a las pruebas de tensión, el prototipo diseñado presenta puntos de tensión de Von Mises solamente en los pernos, lo cual puede ocasionar una posible falla al momento de que se esté utilizando. Por otra parte, la 1a y 3a tensión principal muestra que la mayor compresión se da en la parte superior debido a que es la ubicación en donde se aplica el peso directamente. Aunado a esto se puede tener un ligero desplazamiento en la rodilla, lo cual debido a la magnitud no representaría un peligro al momento del montaje y uso de la rodilla diseñada. Finalmente, a pesar de que se tiene un número considerable de puntos con factor de seguridad menor al resto de la rodilla, tampoco es significativo, por lo que se considera que en general la rodilla diseñada puede ser construida, sin riesgo de fallo.

5. Conclusiones

El diseño mediante software especializado permite corroborar el funcionamiento de los mecanismos, por lo que resulta un paso que debe ser recomendado previo a la construcción de cualquier prototipo mecánico. Por otra parte, el uso de los nuevos materiales que han ido emergiendo recientemente, como el plástico ABS, permiten disminuir los tiempos de construcción, el peso final del ensamble y el costo de manufactura.

Sin embargo, debe tomarse en cuenta que al ser un material plástico, este no siempre tendrá las mismas propiedades de resistencia a la tensión y a la deformación como un material metálico. Esto puede corroborarse si en un trabajo a futuro, la rodilla se simula utilizando por ejemplo aluminio y acero inoxidable,

que, si bien aumentarán el peso y el costo del prototipo, también se tendrá un aumento en el tiempo de vida.

6. Bibliografía y Referencias

- [1] Dorador González, J. M., Robótica y prótesis inteligentes, Revista digital universitaria, vol. 6, No. 1, 2005.
- [2] García Vargas, L. A. y Vargas Duque, S. A., Diseño y simulación de un sistema controlado de amortiguación para la rodilla de la prótesis transfemoral. Tesis de Licenciatura. Universidad de la Salle, Bogotá, Colombia, 2007.
- [3] Hernández Stengele, F., Diseño y construcción de prototipo neumático de prótesis de pierna humana. Tesis Licenciatura. Universidad de las Américas Puebla, México, 2008.
- [4] Menghini, M., Diseño y construcción de prototipo funcional de prótesis total de rodilla policéntrica y pie. III Jornadas de Investigación, Transferencia y Extensión de la Facultad de Ingeniería. p. 25-30, La Llata, Argentina, abril 2015.
- [5] Murthy Arelekatti, V. N. and Winter, A. G., Design of a Fully Passive Prosthetic Knee Mechanism for Transfemoral Amputees in India. IEEE International Conference on Rehabilitation Robotics (ICORR). 11-14 Aug, Singapore, Singapore, 2015.
- [6] Näder, M. and Näder, H. G., Compendio de prótesis. Prótesis para miembro inferior. Berlin, Alemania: schiele & schon, 2003.
- [7] Nájera Castrejón, J. A., Diseño del control para una prótesis de rodilla tipo policéntrica. Tesis de Maestría. Universidad Nacional Autónoma de México, 2013.
- [8] Nordin, M. and Frankel, V. (Eds), Basic biomechanics of the musculoskeletal system. Lippincott Williams & Wilkins, Baltimore-USA, 2001.
- [9] Rodríguez-Sánchez, A. E, Méndez Aguirre, J. S. A. y Robles Sánchez, R., Análisis de esfuerzos en eslabones impresos en 3D para un mecanismo de prótesis policéntrica de rodilla. UTCJ Theorema, No. 3, pp. 34-41, 2016.

- [10] Sánchez Lacuesta, J. Hoyos, J. V. Viosca, E. Soler Gracia, C. Comín, M. Lafuente, R. Cortés, A. Vera, P., *Biomecánica de la marcha humana normal y patológica*, Valencia España, Instituto de Biomecánica de Valencia, 2006.
- [11] Torrealba, R. R., Pérez-D'Arpino, C, Cappelletto, J, Fermín-León, L, Fernández-López, G, and Grieco, J. C., *Through the Development of a Biomechatronic Knee Prosthesis for Transfemoral Amputees: Mechanical Design and Manufacture, Human Gait Characterization, Intelligent Control Strategies and Tests*. IEEE International Conference on Robotics and Automation. May 3-8, Anchorage, Alaska, USA, 2010.
- [12] Valencia, F, Mejía, C. y Erazo, V., *Desarrollo de una prótesis de rodilla para amputaciones transfemorales usando herramientas computacionales*. *UIS Ingenierías*, vol. 16, no. 2, pp. 21-32, 2017.
- [13] Vega, D. Y Escobar, E. *Diseño de rodilla policéntrica, simulación y evaluación de la resistencia a la fatiga*. *RIDTEC* Vol. 11, No. 2, pp. 29-34, 2015.

DISPOSITIVO DE ILUMINACIÓN LED CON INCORPORACIÓN DE ELECTRÓNICA DIGITAL Y CONTROL DESDE ANDROID POR BLUETOOTH

Mario Alberto Hernández Alves

Universidad Autónoma Metropolitana- Azcapotzalco
marbeto_lvs@hotmail.com

Leonardo Sánchez

Universidad Autónoma Metropolitana- Azcapotzalco
marbeto_lvs@hotmail.com

José A. Reyes Ortiz

Universidad Autónoma Metropolitana- Azcapotzalco
jaro@correo.azc.uam.mx

Resumen

El control de factores como la humedad, temperatura e iluminación en un entorno específico resulta de gran importancia ya que pueden interferir en la realización de actividades. Este artículo presenta un dispositivo de control de iluminación para un entorno o espacio. El dispositivo presentado utiliza tecnología LED y es controlado mediante un dispositivo móvil con Bluetooth. Se diseñó e implementó el dispositivo utilizando un circuito basado en electrónica digital sobre una placa Arduino con interfaz Bluetooth. Además, se definió un protocolo de aplicación para manipular el dispositivo a distancia mediante un dispositivo móvil. Los experimentos realizados sobre el dispositivo presentado muestran un funcionamiento estable y resultados prometedores.

Palabras Claves: Comunicación Bluetooth, control de un espacio, dispositivo de iluminación LED.

Abstract

The control of factors such as humidity, temperature and lighting in a specific environment is absolutely important because they can interfere with the

performance of several activities. This paper presents a lighting control device for an environment or space. The presented device uses LED technology and it is controlled by a wireless mobile device by means of Bluetooth. A circuit based on digital electronics has been designed and it has been implemented on an Arduino board with a Bluetooth interface. Also, an application protocol was defined for manipulating the proposed device by means of a mobile device. The experimentation performed on the presented device shows a stable operation of it and promising results.

Keywords: *Bluetooth communication, LED lighting device, space control.*

1. Introducción

Las diversas actividades realizadas por las personas de nuestra sociedad se ven afectadas por distintos parámetros o circunstancias [Aaronson, 1943], [Chandrinós, 1998]. Variables como la temperatura, humedad e iluminación están relacionadas con la correcta realización de una tarea en particular. Esto significa que los valores para estas variables cambian de acuerdo al tipo de actividad. En particular, la iluminación es un tema que ha tomado bastante fuerza en los últimos años, gracias al uso de tecnología LED (*Light Emitting Diode*) y al uso de plataformas de hardware libre como Arduino [Becky, 2015].

La tecnología LED es un invento de la década de los 60 del siglo pasado, pero su uso para fines de iluminación doméstica es reciente. Ésta tiene enormes ventajas respecto a la iluminación incandescente e iluminación halógena. Una bombilla LED, con un promedio de vida de 50000 horas, es 30 veces más duradera que una bombilla incandescente y 25 veces más que una bombilla halógena. En cuanto a eficiencia y ahorro, las bombillas que incorporan tecnología LED son capaces de brindar la misma cantidad de luz (en lúmenes¹) que una bombilla incandescente, pero con un ahorro del 80% de energía. La principal razón para ello es que las bombillas incandescentes y halógenas convierten hasta el 90% de la energía que consumen en calor y sólo el 10% se transforma en iluminación [Illuminet.com, 2016]. Por otro lado, la principal desventaja de esta tecnología es

¹ Lúmen: unidad del Sistema Internacional de Medidas para medir el flujo luminoso, una medida de la potencia luminosa emitida por la fuente.

su precio, ya que una bombilla de tecnología LED es aproximadamente 30% más cara que una bombilla incandescente y 10% más cara que una bombilla halógena [Illuminet.com, 2016].

Por otra parte, desarrolladores independientes han logrado crear versiones modificables de muchos dispositivos innovadores, como impresoras 3D, cuyas especificaciones, diagramas esquemáticos y códigos son de acceso público ya sea de forma gratuita o de pago. Este tipo de dispositivos electrónicos libres y modificables permiten que la comunidad realice adaptaciones de nuevas funciones para éstos. Por ello es de gran importancia la creación de este tipo de dispositivos que hacen más inclusivo el desarrollo de tecnología en el mundo.

Para acercar a la gente común a la tecnología LED, empresas como Apple inc. y Philips han implementado dispositivos que aprovechan la corriente directa e incorporan electrónica digital para crear iluminadores novedosos. Algunos de éstos incorporan funcionalidades extras tales como: cambio de color de la luz emitida, control de intensidad, encendido y apagado desde un Smartphone. Si bien es cierto que estas características hacen más atractivo el uso de este tipo de bombillas; también lo es que elevan notoriamente el precio de éstas.

De esta manera, es evidente que el número de dispositivos de electrónica libre es muy reducido comparado con la enorme cantidad de dispositivos privativos que se lanzan al mercado cada año. Esto provoca que muchos de los productos electrónicos de nuevo lanzamiento terminen siendo “cajas negras” para los usuarios, aún para los que tienen conocimientos en computación y/o electrónica, limitando el posible potencial de estos.

Así, en este trabajo se diseña e implementa un dispositivo de control de iluminación habitacional que integra tecnología LED y funcionalidades que pueden ser ejecutadas desde un dispositivo Android. Dicho dispositivo se basa en el uso de electrónica libre para su fácil fabricación e incluye funciones que extiendan el paradigma de la iluminación actual. La idea es que cualquier persona con conocimientos básicos en electrónica lo puede manufacturar a bajo costo y las personas con conocimientos intermedios en electrónica pueden realizar modificaciones de hardware y software al dispositivo.

El trabajo previo relacionado con el diseño de dispositivos configurables y accesibles a distancia es vasto. Esto se debe a que algunos trabajos incluyen características configurables y otros involucran tecnologías de comunicación inalámbricas diferentes.

En [González, 2015] los autores presentan un control de iluminación con tecnología LED. El objetivo principal de dicho trabajo es identificar y evaluar los diferentes tipos de control para tiras LED utilizando los circuitos de control existentes en el mercado. Los autores en [Hernández, 2005] presentan un sistema que controla las luces LED de un escenario a través de un programa y una red cableada. En [Pérez, 20015] se presenta un sistema que permite controlar la iluminación LED de forma inalámbrica utilizando la tecnología infrarroja.

Por otro lado, empresas como Philips han desarrollado bombillas que permiten ser controladas a distancia mediante WiFi. Las denominadas Bombillas Hue [Philips, 2016], son capaces de emular cualquier color de una paleta de colores y pueden ser controladas a través de una aplicación vía WiFi. Otra empresa que se dedica a la comercialización de bombillas que se pueden controlar a distancia, es Luz Wi-Fi-Bombillas Inteligentes [Luzwifi.com, 2016]. Al igual que Philips, esta empresa diseñó una bombilla que cambia de colores y que se controla vía WiFi, pero además puede variar la intensidad de la luz emitida. Por su parte, JUNG [Jung.de, 2016] es una empresa alemana que se ha enfocado a la domótica en general. Cuentan un con un control de iluminación para múltiples casos de uso con diferentes formas de manipulación incluyendo manipulación automática y aprendizaje.

2. Métodos

El desarrollo de este trabajo involucra el diseño e implementación de dos dispositivos: el dispositivo de control a distancia y el dispositivo de iluminación. El primero, es un dispositivo móvil con sistema operativo Android que permite la manipulación del dispositivo de iluminación. Para ello, este envía señales inalámbricas a través de una interfaz Bluetooth en un formato particular. El segundo, integra un dispositivo de control fijo, una placa Arduino UNO, una

interfaz de comunicación Bluetooth y tres tiras de led. Este recibe comandos mediante su interfaz Bluetooth o a través de un botón asociado al dispositivo de control fijo. Cada uno de los comandos tiene un formato específico para indicar un modo de funcionamiento distinto.

La figura 1 muestra la interacción entre los dispositivos con el usuario. Se puede observar que el dispositivo de iluminación se puede manipular tanto manual como de manera remota.

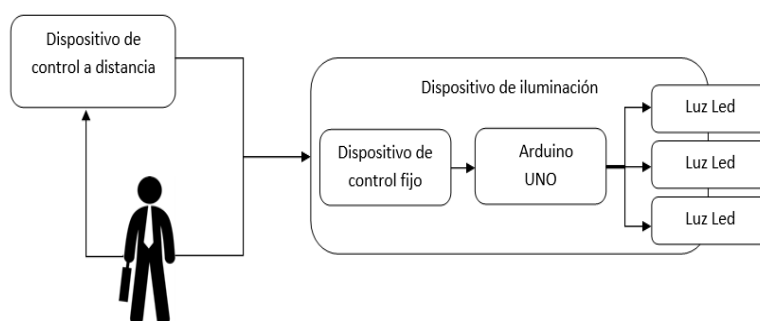


Figura 1 Interacción entre los dispositivos y el usuario final.

Se proponen cuatro modos de funcionamiento:

- Modo 1: El sistema mantendrá las iluminaciones apagadas en espera de instrucciones desde el dispositivo de control a distancia.
- Modo 2: El sistema encenderá la iluminación a un 75%
- Modo 3: El sistema encenderá la iluminación a un 40%
- Modo 4: El sistema encenderá por la acción el sensor de movimiento.

Todos los modos de funcionamiento son accesibles desde el dispositivo de control fijo y desde el dispositivo de control a distancia y en conjunto brindan las siguientes funciones al dispositivo de iluminación:

- Encendido/apagado progresivo (manual y mediante aplicación Android).
- Almacenamiento de intensidad preferida.
- Encendido completo o encendido parcial.
- Control de intensidad mediante aplicación Android.
- Programación de apagado automático mediante aplicación Android.
- Efectos de iluminación activables desde la aplicación Android.

- Encendido mediante sensor de movimiento.
- Control de estados mediante aplicación Android y mediante el control fijo.
- Simulación de estancia.

La implementación de estas funcionalidades implica el diseño e implementación de diversos módulos que describiremos a continuación.

Módulo de control a distancia

Este módulo consiste de una aplicación que permite la manipulación del módulo de iluminación de manera inalámbrica. El objetivo de este módulo es transformar las acciones del usuario sobre la aplicación en comandos que serán enviados al módulo de iluminación, tal y como se muestra en la figura 2.

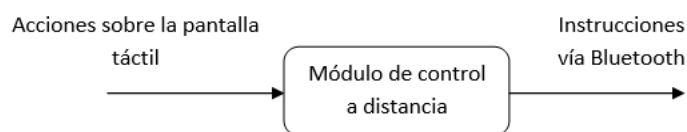


Figura 2 Módulo de control a distancia.

Para la implementación se utilizó el entorno de desarrollo integrado App Inventor 2 con programación orientada a eventos. Se generó un archivo APK instalable en cualquier sistema operativo Android (Tablet o Smartphone) versión 2.4 o superior. Se utilizaron los elementos estándar del sistema operativo Android como botones, listas, textos, etc. con la finalidad de facilitar su comprensión para los usuarios. Las imágenes utilizadas se descargaron del repositorio libre “pixabay.com” y se tomaron las siguientes consideraciones de programación visual:

- Se evitó en la medida de lo posible el uso del color blanco, ya que es color más brillante en la pantalla y puede llegar a cansar la vista.
- Se evitó colocar elementos que puedan confundir al usuario como imágenes no relacionadas con la acción del botón o mensajes confusos.
- Se incorporó prevención de errores mediante mensajes de texto en partes estratégicas de la aplicación para evitar que el usuario pierda el flujo de ejecución.

La figura 3a muestra la pantalla inicial de la aplicación en la que se indica que se debe seleccionar uno de posibles sistemas de iluminación. La figura 3b muestra la pantalla en la que se obtiene la lista de los dispositivos Bluetooth disponibles.

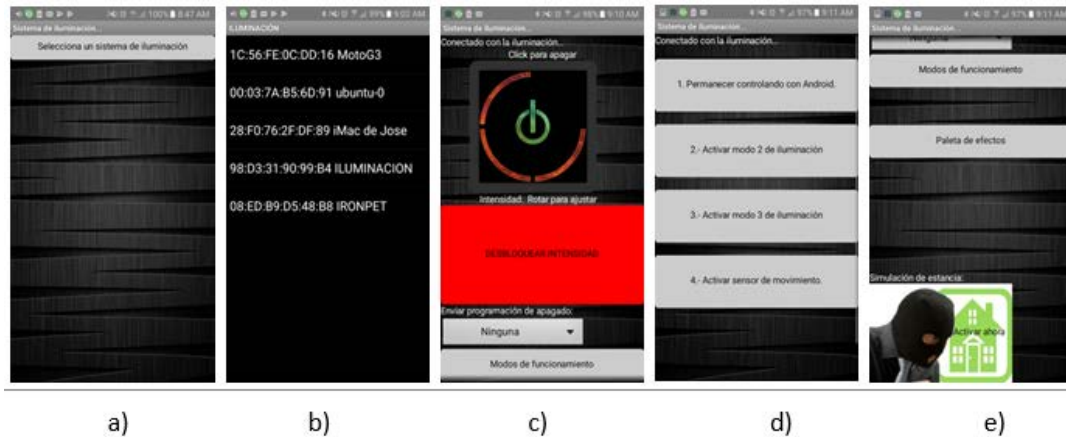


Figura 3 Pantallas iniciales de la aplicación.

La figura 3c muestra la pantalla inicial de la aplicación una vez que se ha encendido el dispositivo de control de iluminación de manera remota. El botón de desbloqueo de intensidad permite modificar la intensidad luminosa de cada una de las salidas del módulo de iluminación. Se utilizó el acelerómetro del teléfono para subir y bajar la intensidad para un uso más intuitivo de la aplicación y así reducir la curva de aprendizaje. La figura 3d muestra los posibles modos de funcionamiento del dispositivo de control de iluminación, los cuales corresponden a los descritos en la sección anterior. La figura 3e muestra la funcionalidad de “simulación de estancia”, en la que el dispositivo de control de iluminación se mantiene en un ciclo continuo generando tiempos aleatorios para el encendido y apagado del módulo de iluminación.

Módulo de la recepción de instrucciones por Bluetooth

Este módulo valida la conexión con el módulo de control a distancia y además captura y retransmite los comandos recibidos de forma inalámbrica por un medio de comunicación físico al módulo de procesamiento de instrucciones, como se muestra en la figura 4.

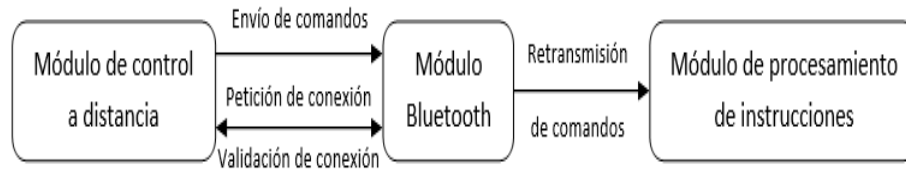


Figura 4 Módulo de recepción y retransmisión de comandos.

Para la implementación de este módulo se utilizó el módulo HC-06 junto con una placa Arduino UNO. Se utilizó el entorno de desarrollo de Arduino para configurar el módulo Bluetooth como servidor y también los parámetros necesarios para su identificación y funcionamiento: velocidad de ciclo de reloj, nombre del dispositivo Bluetooth y contraseña de acceso, tal y como se muestra en la figura 5.

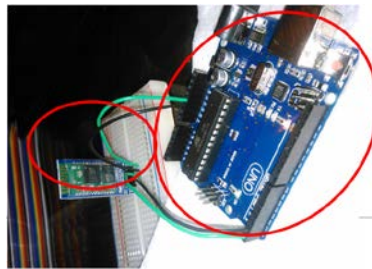


Figura 5 Módulo de recepción de instrucciones Bluetooth a través de placa Arduino UNO.

Módulo de Recepción de Señales de Presencia y Botón de Control

La función de este módulo es detectar presencia en la habitación o pulsaciones en el botón de control y transformarlas en señales que se envían por un medio físico al módulo de procesamiento de instrucciones. La figura 6 muestra la interacción entre los módulos antes mencionados.

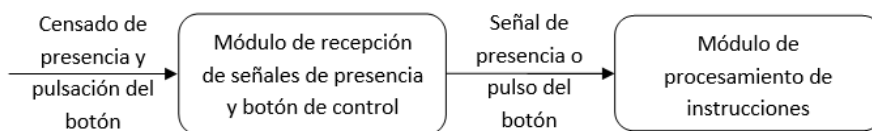


Figura 6 Detección de presencia o actividad del botón del control.

Este módulo es un circuito que integra el módulo de recepción de instrucciones por Bluetooth, un pulsador (botón) y un sensor PIR HC-sr501 (sensor de

presencia). Los últimos dos permiten generar una interrupción en el sistema, el botón de forma manual y el sensor a partir de la detección de una presencia.

La figura 7 muestra el diagrama del circuito. Como se puede observar, el sensor PIR HC-Sr501 es un transistor que se satura con cambios de temperatura. Por otro lado, la resistencia de 1K utilizada en este circuito podría ser de cualquier valor ya que sus únicas funciones son: mantener un bit bajo y evitar que la corriente fluya directamente a tierra cuando el pulsador está cerrado. La construcción del circuito se realizó sobre una placa fenólica de 1.5 milímetros de base aislante, y 0.105 milímetros de lámina de cobre (conductor) tal como lo recomienda la norma UNE 20-621-84/3 [ftp.ehu.es, 2016].

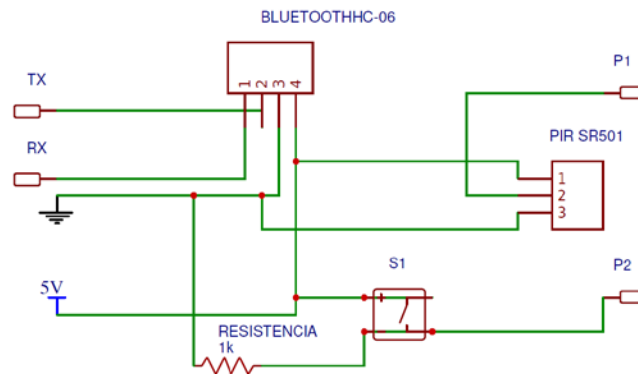


Figura 7 Circuito electrónico para el módulo de la recepción de señales de presencia y botón de control.

Módulo de Potencia

La función de este módulo es la de amplificar las señales PWM², provenientes del módulo de procesamiento de instrucciones, a una señal PWM de mayor voltaje y enviarla al módulo de iluminación como se ve en la figura 8.

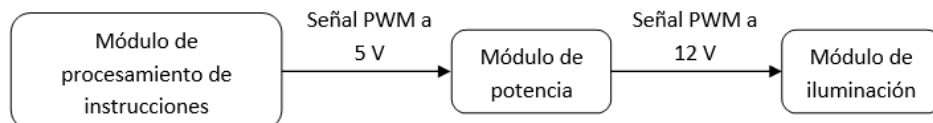


Figura 8 Amplificación de la señal por el módulo de potencia.

² La modulación por ancho de pulsos de una señal o fuente de energía es una técnica en la que se modifica el ciclo de trabajo de una señal periódica (una senoidal o una cuadrada), de forma digital sin necesidad de usar potenciómetros.

El circuito integra los siguientes componentes electrónicos: transistor TIP120 Darlington, diodo rectificador 1N4001 A y resistencia de 1K. El primero nos permite transformar las señales PWM de 5 V a 12 V. El segundo permite proteger las salidas PWM de algún error humano al conectar el circuito o bien, de algún componente con mal funcionamiento. El tercero se utiliza para que la corriente proveniente de las salidas PWM mantengan al transistor TIP120 en su área de saturación.

Al igual que el circuito anterior, la construcción de este circuito se realizó sobre una placa impresa fenólica de 1.5 milímetros de base aislante y 0.105 milímetros de lámina de cobre siguiendo la norma UNE 20-621-84/3 [ftp.ehu.es, 2016]. La figura 9 muestra el diagrama del circuito.

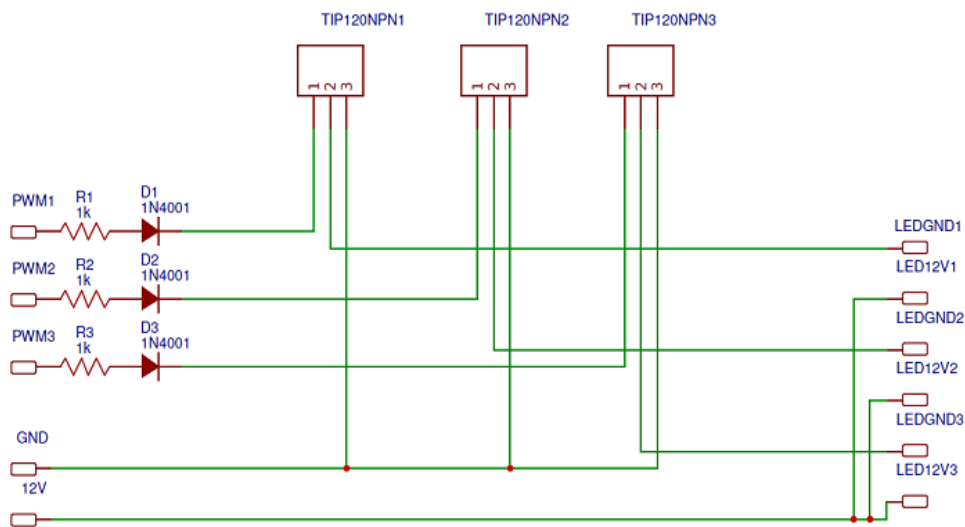


Figura 9 Circuito electrónico para el módulo de potencia.

Módulo de Iluminación

Es el módulo encargado de brindar una respuesta de iluminación visible al usuario a partir de las instrucciones que se están ejecutando. El módulo de iluminación es un componente de hardware visible que influye en la presentación del dispositivo general.

Para el diseño del módulo de iluminación, se seleccionó la tira LED SMD 5050 [Datasheet SMD-LED-5050, 2017] que trabaja a 12 V. Esta tiene 3 LED por cada 5

cm y cada LED brinda 15 lúmenes. Si se considera que una lámpara LED convencional emite 1000 lúmenes, entonces se necesitan 120 cm de tira LED aproximadamente. La disposición de dicha tira se consideró de forma que se genere un ambiente envolvente dentro de un espacio, tal y como se muestra en la figura 10.

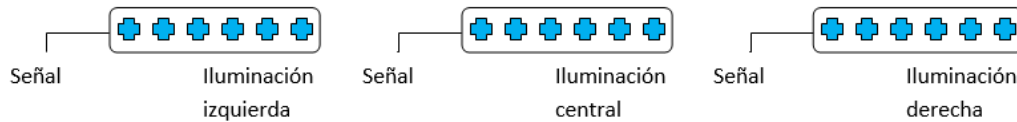


Figura 10 Diseño del módulo de iluminación.

Para la implementación del módulo de iluminación se consideró la siguiente disposición:

- 45 cm para el módulo de iluminación izquierdo
- 45 cm para el módulo de iluminación derecho
- 30 cm para el módulo de iluminación central

Para cada módulo se diseñó una base de 18 por 480 o 330 mm según el caso. A ésta se le agregó un desnivel de 13 mm y en los extremos se colocaron dos orificios de 2 mm para el cableado; todo sobre un material de madera pino de 18 mm de espesor.

Módulo de Procesamiento de Instrucciones

Su función es la de recibir, interpretar, ejecutar y administrar las instrucciones de iluminación inalámbricas del módulo de la recepción de instrucciones por Bluetooth, así como procesar las señales del módulo de recepción de señales de presencia y botón de control para generar las señales PWM como se observa en la figura 11.

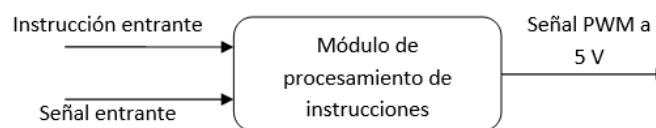


Figura 11 Funcionamiento del módulo de procesamiento de instrucciones.

Este módulo es el más complejo de todos, ya que se encarga de administrar los recursos de hardware y brindar los servicios necesarios para ejecutar las diferentes funciones del dispositivo de control de iluminación. Está constituido por dos partes: hardware y software.

Para el Hardware se decidió utilizar la arquitectura de Arduino Uno R3, principalmente debido a que es una de las arquitecturas más económicas del mercado.

Dicha arquitectura tiene las siguientes características: una velocidad de procesamiento de 16 MHz, una memoria de almacenamiento permanente de 32 KB, para sistema y datos permanentes, una memoria principal de 2 KB para procesos y datos temporales, dos puertos seriales RX y TX, 14 puertos digitales y seis puertos analógicos.

Para el software, al contar con una arquitectura limitada, el sistema desarrollado se adaptó a esta, considerando las siguientes restricciones:

- Cada proceso con sus correspondientes variables locales y variables globales no puede equivaler a más de 2 KB cargado en memoria principal.
- No debe generar uso de memoria principal para datos que pueda saturarla por completo en tiempo de ejecución y provocar el fallo total del sistema.
- No puede exceder los 32 KB de almacenamiento permanente para su funcionamiento.
- El sistema debe administrar de forma correcta los tiempos de ejecución de las interrupciones y procesos.
- No debe contener procedimientos que monopolicen la CPU, cada proceso debe contener al menos una interrupción de salida esto evita que el sistema pueda quedar pasmado en un estado.

Así, la distribución de puertos de la arquitectura Arduino es la siguiente:

- El puerto 3 analógico corresponde a la salida de iluminación central.
- El puerto 5 analógico corresponde a la salida de iluminación izquierda.
- El puerto 6 analógico corresponde a la salida de iluminación derecha.
- El puerto 8 digital corresponde a la entrada de bits del sensor de presencia.

- El puerto 12 digital corresponde a la entrada de bits del pulsador.
- Los puertos seriales tx y rx trabajan en conjunto para mantener la comunicación con el módulo de la recepción de instrucciones por Bluetooth.
- Se utiliza una variable para almacenar la intensidad establecida por el usuario y utilizarla al inicio del sistema.

Para la implementación del Software se utilizó el entorno de desarrollo integrado (IDE) de Arduino que utiliza su propio lenguaje de programación.

Integración de módulos y estructura de presentación

La integración de los módulos corresponde a la interconexión de estos. Para ello se considera el diseño de una estructura/carcasa para el almacenamiento y presentación de todos los componentes. Se tomaron en cuenta las siguientes consideraciones para el diseño de dicha estructura:

- No debe generar interferencias sobre alguno de los componentes y debe ser de un material ligero, maleable y fácil de trabajar.
- Debe permitir al sensor de presencia un rango de mínimo 120 grados para sensor.
- Debe tener expuestas la salida USB del Arduino así como la de alimentación, además debe tener orificios para las salidas de los módulos de iluminación.
- Debe permitir que los indicadores de los componentes estén a la vista.
- Debe permitir el flujo de aire frío hacia los transistores para evitar sobre calentamiento.

La carcasa se diseñó en dos partes: la parte frontal y la parte posterior. Para la parte posterior se seleccionó el material MDF por ser económico y tener propiedades aptas para los requerimientos: ligero y maleable. Para la parte frontal se seleccionó el material acrílico por ser fácil de conseguir y tener propiedades aptas para los requerimientos: ligero, translucido, maleable y pueden obtenerse buenos acabados. Para la implementación se seleccionó el proceso de fresado

por control numérico computarizado. La figura 12 muestra la integración de los módulos en la carcasa ya terminada.

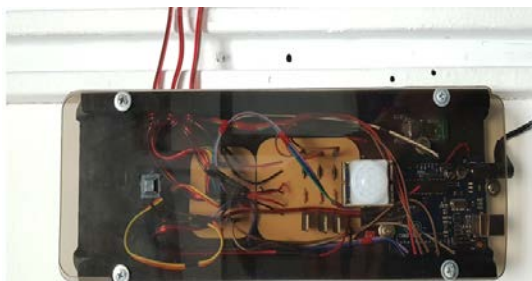


Figura 12 Integración de módulos en dispositivo de control de iluminación armado.

3. Resultados

El resultado de este proyecto corresponde al correcto funcionamiento del dispositivo de control de iluminación diseñado, respecto a las siguientes funcionalidades:

- El sistema se conecta a la alimentación e inicia en modo 1. Esto implica que el módulo de iluminación está pagado de forma inmediata.
- Al presionar el botón de control el sistema cambia de modo 1 a modo 2 y encienden las iluminaciones izquierda y derecha a una intensidad de 80%.
- Al presionar el botón de control nuevamente el sistema de iluminación cambia de modo 2 a modo 3 y enciende la iluminación central a un 60%.
- Al presionar el botón de control nuevamente el sistema de iluminación cambia de modo 3 a modo 4 y se activa el sensor de movimiento. Este enciende sólo ante la presencia de una persona y el tiempo de encendido es de 15 segundos.
- Al presionar el botón de control nuevamente el sistema regresa al modo 1.
- Al conectar el sistema de iluminación con un dispositivo Android mediante la aplicación, el sistema indica la conexión apagándose y respondiendo a las instrucciones enviadas desde el celular.
- El encendido desde el dispositivo Android se realiza de forma progresiva y se respeta la intensidad guardada.

- El apagado desde el dispositivo Android se realiza de forma progresiva como se tenía pensado.
- Al rotar el dispositivo Android hacia el lado derecho la intensidad del sistema de iluminación aumenta y al rotarlo hacia el lado izquierdo la intensidad disminuye.
- Al enviar una programación de apagado desde un dispositivo Android el sistema se apaga en el tiempo previsto tal y como debería de suceder. Para validar esto se utilizó un cronometro.
- Al enviar una funcionalidad de prueba desde un dispositivo Android el sistema de iluminación reacciona de forma esperada de acuerdo con cada funcionalidad
- Al enviar un modo de funcionamiento desde un dispositivo Android el sistema de iluminación se coloca en el modo correspondiente.
- Al activar la simulación de estancia desde un dispositivo Android el sistema de iluminación realiza el encendido y apagado dentro de los rangos previstos.

El dispositivo de control de iluminación no presenta ninguna anomalía común en circuitos electrónicos como calentamientos o reinicios repentinos: Se probó el dispositivo de control de iluminación a su máxima intensidad durante 24 horas continuas y no presentó ningún tipo de calentamiento o anomalía. El dispositivo de control de iluminación no presenta problemas de programación como ejecución de instrucciones no solicitadas o retrasos notables en la ejecución de los procesos que impidan su correcto funcionamiento: El dispositivo de control de iluminación ejecuta todas las instrucciones de forma inmediata. El sistema de iluminación se puso en funcionamiento durante una semana ininterrumpida y no presento ningún problema o anomalía de hardware o software.

4. Discusión

La proliferación de las aplicaciones tanto de hardware como de software libre han dado lugar a diversas implementaciones de dispositivos que permiten facilitar

diversas tareas o actividades en el mundo real. Una de las aplicaciones que ha llamado mucho la atención en los últimos años es la iluminación de espacios. En este trabajo se presenta el diseño e implementación de un dispositivo de control de iluminación de hardware libre con una aplicación Android. El diseño del dispositivo de control de iluminación se pensó para ser económico y de fácil fabricación para que cualquier persona con conocimientos en electrónica lo pueda reproducir para su uso o venta. Por otro lado, tanto el diseño del sistema, como el de la aplicación, se pensaron para ser lo más abiertos e intuitivos posibles para el usuario. Esto quiere decir que una persona con conocimientos en programación puede modificar fácilmente el funcionamiento de la aplicación y puede agregar funcionalidades nuevas al dispositivo de control de iluminación.

5. Conclusiones

Con la aparición de plataformas de desarrollo de hardware como Android, se ha facilitado el diseño y desarrollo de dispositivos dedicados y de propósito específico.

Las aplicaciones de domótica son cada vez más comunes en nuestra sociedad. En particular, el control de la iluminación dentro de una habitación mediante un dispositivo Android brinda una experiencia distinta a la forma habitual de utilizar la iluminación y es posible que pueda convertirse en una necesidad a corto plazo.

Este artículo ha presentado un dispositivo de control de iluminación para un espacio habitacional utilizando tecnología LED y controlado mediante un dispositivo móvil por medio de comunicación Bluetooth. El dispositivo de control de iluminación se constituye de un circuito basado en electrónica digital diseñado sobre una placa Arduino, con un funcionamiento estable.

Para proyectos posteriores se puede pensar en diseñar e implementar un dispositivo similar pero que funcione mediante la tecnología Wifi. También podemos pensar en modificaciones que integren inteligencia artificial al diseño actual o un software para programar efectos mediante una interfaz gráfica sin necesidad de utilizar código.

6. Bibliografía y Referencias

- [1] Aaronson, S. A. Theory of Human Motivation. *Psychological Review*, No. 50, pp. 370-396, 1943.
- [2] Becky Stewart, *Adventures in Arduino*, Wiley, 2015.
- [3] Circuitos Eléctricos, ftp.ehu.es, 2016: http://ftp.ehu.es/cidira/dptos/depjt/Tecnologia/BK-ANGEL/Presentaciones/02_Circuitos%20Impresos.pdf/,.
- [4] Chandrinos, K. V., and Trahanias, P. E., *Web-based Information Systems. ERCIM Workshop Proceedings*. Toronto, Canada, October, 1998.
- [5] Datasheet SMD-LED-5050, 2017: <https://www.tweaking4all.com/wp-content/uploads/2014/01/5050LED.pdf>.
- [6] González Ramírez R., *Control de iluminación con tecnología LED proyecto terminal*, División de Ciencias Básicas e Ingeniería, Universidad Autónoma Metropolitana Azcapotzalco, México, 2015.
- [7] Hernández Borja C., *Iluminación inteligente de escenarios proyecto terminal*, División de Ciencias Básicas e Ingeniería, Universidad Autónoma Metropolitana Azcapotzalco, México, 2005.
- [8] Iluminet.com, *Revista de iluminación*, 2016: <http://www.iluminet.com/>.
- [9] Jung.de, JUNG- eNet Control de la iluminación Técnica: <http://www.jung.de/es/925/productos/tecnica/control-de-la-iluminacion/enet/>, 2016.
- [10] Luzwifi.com, *Bombillas Led Wifi controladas con el móvil*, 2016: <http://www.luzwifi.com/>.
- [11] Pérez Carbajal C., *Sistema de iluminación con control inalámbrico infrarrojo basado en tecnología leds, proyecto terminal*, División de Ciencias Básicas e Ingeniería, Universidad Autónoma Metropolitana Azcapotzalco, México, 2015.
- [12] Philips Hue, Phillips Inc: 2016: <http://www2.meethue.com/es-mx/>, 2016.

DESIGN AND FABRICATION OF A 64-QAM MODULATOR FOR ANALYSIS OF SIGNALS BETWEEN STAGES

Jorge Andrés Hernández Carrillo

Tecnológico Nacional de México/Instituto Tecnológico de Tijuana
jorgeandreshernandezcarrillo@gmail.com

José Ricardo Cárdenas Valdez

Tecnológico Nacional de México/Instituto Tecnológico de Tijuana
jose.cardenas@tectijuana.edu.mx

Virgilio Rosendo Pérez

Universidad Tecnológica de Tijuana (UTT)
virgilio.perez@uttijuana.edu.mx

Manuel de Jesús García Ortega

Tecnológico Nacional de México/Instituto Tecnológico de Tijuana
manuel.garcia@tectijuana.mx

Andrés Calvillo Téllez

Instituto Politécnico Nacional – CITEDI
calvillo@citedi.mx

Resumen

Este trabajo propone el diseño de un modulador digital 64-QAM en tarjeta impresa basado en tecnología en chip, la tarjeta desarrollada consta de dos fases de retardo y dos fases de amplificación, el modulador fue desarrollado usando el software PCB Wizard y fueron establecidos puntos de prueba para cada etapa. El diseño desarrollado fue construido a través de una máquina de control numérico por computadora. Los resultados experimentales muestran una mejora significativa en la precisión alcanzada de NMSE de -51 dB. La tarjeta desarrollada ofrece una herramienta de diseño para desarrolladores de hardware donde los

productos de intermodulación puedan ser evaluados, este modulador digital evita inductancias parásitas y capacitancias entre líneas. Los resultados del sistema desarrollado son efectivos para agregar ruido blanco Gaussiano y comprobar los los productos de intermodulación de hasta 3er orden que se pueden agregar, así como el efecto de recrecimiento espectral. El sistema es una herramienta de diseño de hardware capaz de mostrar por etapas los cambios en amplitud y fase que involucra una modulación de tipo QAM.

Palabras Claves: CNC, Diseño, Intermodulación, 64-QAM.

Abstract

This paper proposes the design of a 64-QAM digital modulator on board based on chip technology, the developed board comprises two phase delays and two amplification stages, the modulator was developed using the software PCB Wizard and were established test points for each stage. The developed layout was built through a computer numerical control machine. Experimental results show a significant improvement of the accuracy based on a reached NMSE= -51 dB. The developed board offer a design tool for hardware developers where the intermodulation products can be evaluated, this digital modulator avoid parasitic inductances and capacitances between the lines. The results of the developed system are effective to add white Gaussian noise and probe the intermodulation products up to 3rd order that can be added, as well as the effect of spectral regrowth. The system is a hardware design tool able to show by stages the amplitude and phase changes that involves a 64-QAM modulation.

Keywords: CNC, Intermodulation, Layout, 64-QAM.

1. Introduction

With the explosive increase of mobile and portable communications, digital transmission with wide baud rate, the spectral availability has become so scarce. To make possible this global condition is required the use of digital modulation with big transfer schemes such as quadrature amplitude modulation (QAM). The design of the modulation, interleaving, coding, testing and characterization of cable

systems in North America is an important necessity [ANSI/SCTE, 2006], nowadays the main universities in Mexico related to Telecommunications are teaching the main digital concepts solely to explain the interpretation of digital constellations and the obtained performance in a digital link. Unfortunately, the students do not understand the behavior between stages even in the field there is no hardware related to QAM devices where the researchers and students can obtain the signals previous to the antenna in the transmitter.

However, such techniques are more susceptible to noise since a greater number of combinations means that these combinations are closer to each other and therefore the noise signal can be switched more easily. The probability of error in the transmitter chain depends primarily on the noise that is added to the modulated signal over the communication channel. In addition, the designers related to digital schemes requires platforms to properly address and correct the non-desirable effects that QAM system can achieve. For this reason, hardware and system level designers have made special efforts, not only through software [Cárdenas, 2012], but also in hardware implementation [Yan, 2013], [Gnauck, 2011], [Shin, 2003] testing QAM links [Kameda, 2011], [Besnoff, 2015], [Oguma, 2009] or creating sequences for multicarrier applications [Chang, 2010], [Lee, 2016] and QAM structures implemented in FPGA [Vu, 2010].

The authors agree that in order to carry out the feasibility studies of high modulation schemes is required the description in hardware of each stage. In this paper is developed a 64-QAM modulator in board controlled by an open-source electronic prototyping platform programmed in C language. It means that each amplification and phase delay stage is represented into board with the goal of provide a hardware option of QAM analysis for designers, each stage is built through the use of operational amplifiers and digital multiplexers.

This digital modulation technique is mainly used to send data on the downstream channel coaxial cable networks. It is a very efficient technique, supports transmission speeds up to 28 Mbps over a single 6 MHz channel. Although it is susceptible to interference signals, which makes it is not used in the upstream channel because it is very sensitive to the noise.

The main objective of this work is to provide an analysis tool for researchers and students related to telecommunications issues. A further work involves the performance analysis based on bit error rate (BER) compared with the energy per bit to noise power spectral density ratio (EbNo) and a study regarding to antenna coupling, in this case the performance study is not possible because the developed hardware just involves the circuitry and construction between stages.

Additionally, the purpose of this work aims to focus on the development of a 64-QAM modulator where all the test points are perfectly established for signal analysis.

The organization of this correspondence consists of three parts. In the section 2, we state the methodology and the fabrication of the 64-QAM modulator. In the section 3 are described the board performance and the accuracy in terms of normalized mean squared error (NMSE), the section 4 has the main discussion main results and accuracy obtained results the construction of the 64-QAM modulator controlled by FPGA. Finally, in conclusion are summarize the results obtained of the proposed tool.

2. Methods

Employed multilevel schemes as QAM requires a higher signal-to-noise ratio (SNR) than others binary ones under the same bit error rate (BER). Hence the importance of a proper capability for hardware design for these kind of schemes. M-ary QAM is a non-binary memoryless modulation technique in which one of M different symbols is transmitted per time using two orthogonal carriers (in quadrature). In this words are call I and Q signals that are divided by two channels. Each symbol represents a bit stream pack, the M-ary QAM can be represented by equations 1 and 2 [Correa, 2003].

$$s(t) = A_1\gamma_1(t) + A_2\gamma_2(t) \quad (1)$$

$$q = \log_2 M \quad (2)$$

Where $\gamma_1(t) = \sqrt{2/T_s} \cos(2\pi f_c t)$ for $0 \leq t \leq T_s$, otherwise $\gamma_1(t) = 0$ and $\gamma_2(t) = -\sqrt{2/T_s} \sin(2\pi f_c t)$ for $0 \leq t \leq T_s$, otherwise $\gamma_2(t) = 0$, in which T_s

represents one symbol's transmission time interval, f_c is the carrier frequency and A_1 as well as A_2 are the orthogonal carrier coefficients.

Considering that n is bigger than q . The n -bits in N sets of q -bits can be grouped, where q is given by equation 1 and N is an integer number. Based on the 64-QAM overview (figure 1) the q bits will travel simultaneously through the channel forming a 64-QAM symbol.

The used signals for the 64-QAM modulator have the same band width reaching a higher efficiency with the signals $I(t)$ and $Q(t)$ that are modulated by two carriers with the same frequency but with a phase delay of 90° , the equation 2 is shows the resulting expression.

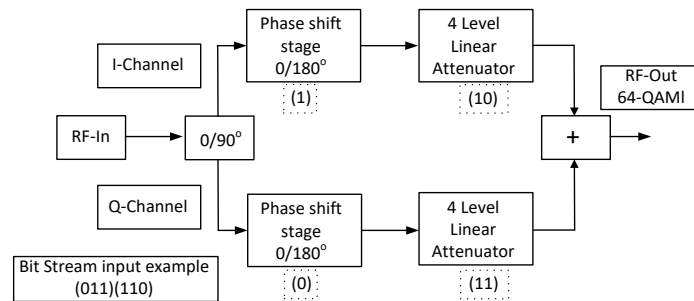


Figure 1 Overview of a typical 64-QAM modulator.

The 64-QAM architecture is based on the direct conversion principle to the RF transmission frequency. Considering the 64-QAM modulator the output signal is based on the RF input frequency [ANSI/SCTE, 2006]. The required frame for 64-QAM is represented in figure 2.

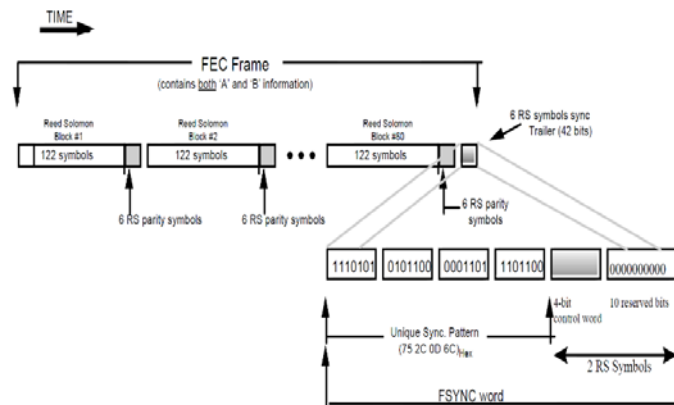


Figure 2 Frame packet format for 64-QAM modulator.

The RF output signal of the 64-QAM modulator can be denoted as $s(t)$ as it is showed in the equation 3.

$$s(t) = I(t) \cos(2\pi f_0 t) + Q(t) \cos(2\pi f_0 t - 90^\circ) = I(t) \cos(2\pi f_0 t) + Q(t) \sin(2\pi f_0 t) \quad (3)$$

Where f_0 represents the carrier frequency of the system. If the digital modulator is properly designed, a 64-QAM receptor should be able to demodulate the $s(t)$ signal, adding a local oscillator. In the receptor side, the recovered signal $r_i(t)$ is proportional to $s(t)$, an ideal form of $I(t)$ is recovered and expressed by the equations 4.

$$\begin{aligned} I(t) &= r_i(t) \cos(2\pi f_0 t) \\ &= I(t) \cos(2\pi f_0 t) + Q(t) \sin(2\pi f_0 t) \cos(2\pi f_0 t) \\ &= I(t) \cos(2\pi f_0 t) \cos(2\pi f_0 t) + \sin(2\pi f_0 t) \cos(2\pi f_0 t) \\ &= I(t) \cos(2\pi f_0 t)^2 + Q(t) \sin(2\pi f_0 t) \cos(2\pi f_0 t) \end{aligned} \quad (4)$$

using trigonometric functions, the equation 4 can be expressed as equations 5.

$$\begin{aligned} I(t) &= \frac{1}{2} I(t) [1 + \cos(4\pi f_0 t)] - \frac{1}{2} Q(t) [\sin(4\pi f_0 t)] \\ &= \frac{1}{2} I(t) + [\cos(4\pi f_0 t)] - Q(t) [\sin(4\pi f_0 t)] \end{aligned} \quad (5)$$

Based on the equations 4 y 5 is designed the circuitry with dual 4-line to 1-line multiplexer that digitally separate the signal between stages, The multiplexers can select 2 bits of data from up to four sources selected by common the bit stream information, in this case the bit 0 and 1 of the 64-QAM symbol represent the address of the circuit, by other hand the bit 3 and 4 are controlling the second amplification stage. This circuit is used to strobe the outputs independently. The design between stages consist of high gain operational amplifiers.

Was used an open-source software for the design, the circuitry in figure 3 was developed in PCB Wizard. Once that the design was completed, and the code was generated using the tool CooperCAM is exported to the CNC machine. In the figure 4 can be seen the design to be exported to the CNC machine.

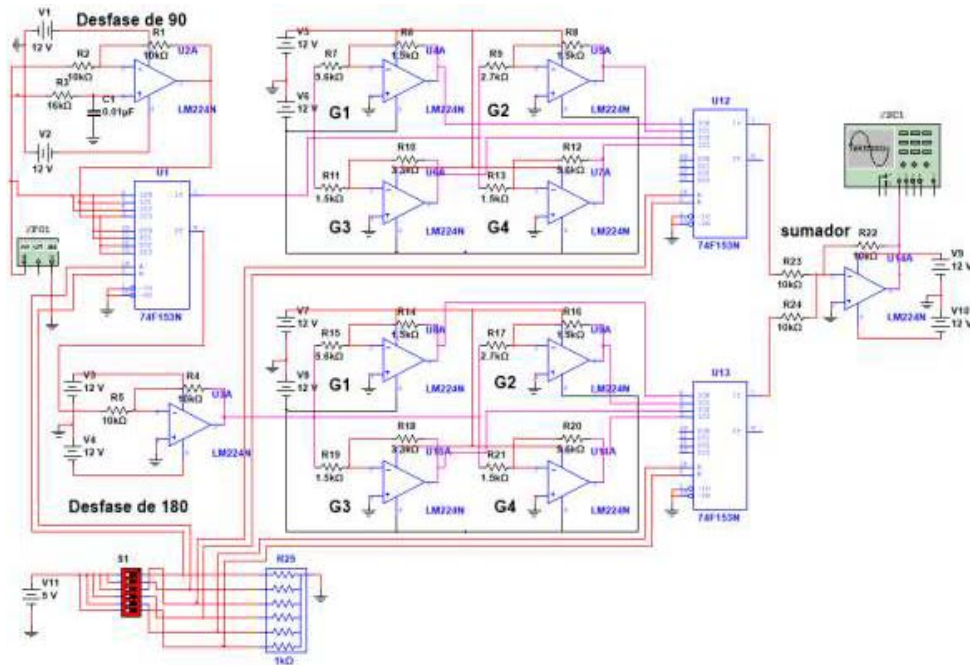


Figure 3 Circuitry of the 64-QAM digital modulator.

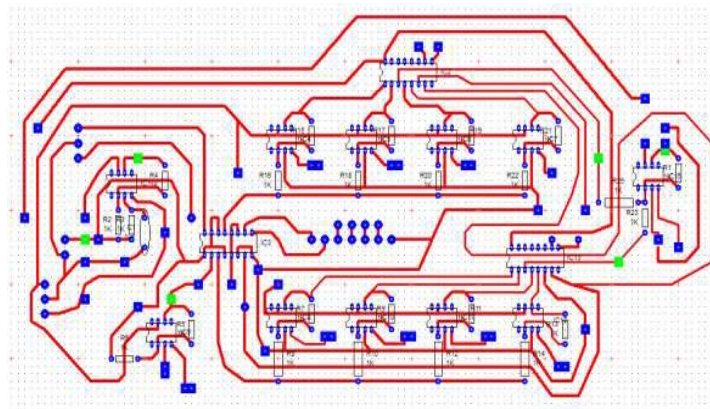


Figure 4 Overview of the developed layout.

The spaces between lines are optimized previous to generate the code for the CNC machine, the figure 5 depicts the general distribution of the devices taking into account the operational circuits and digital multiplexers. Should be noted that right angles were avoided in the corners in order to avoid induced antennas included by a poor design.

In the figure 6 is depicted the simulation and the tools that must be used for the drilling and cutting the modulator, the finals details related to time and length are corrected in this stage.

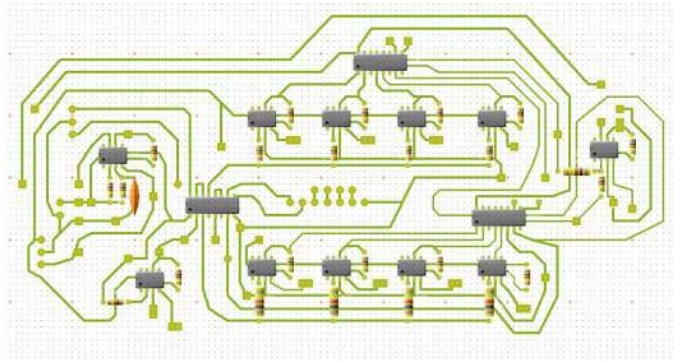


Figure 5 Optimization of the developed board.

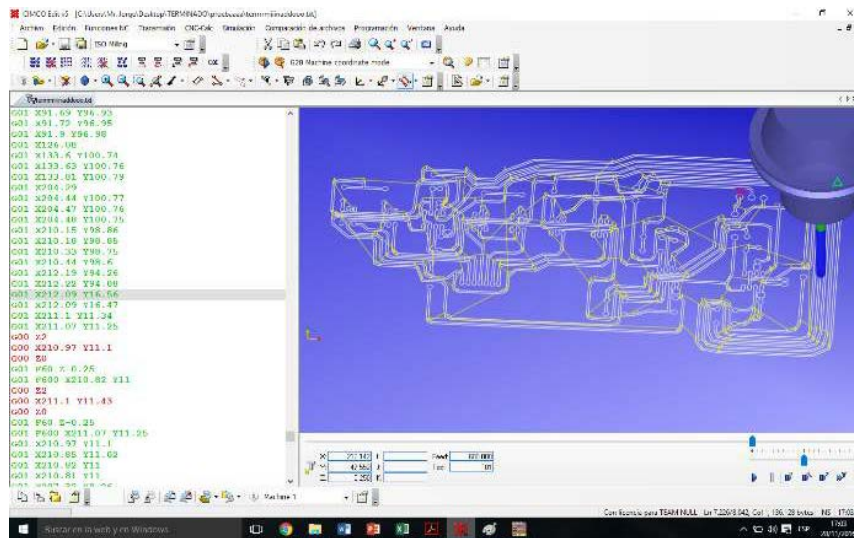


Figure 6 CIMCO Software of the CNC machine.

The figure 7 shows the generated 64-QAM digital modulator, where the test points were established properly and the parasitic inductances and capacitances between the lines were improved.

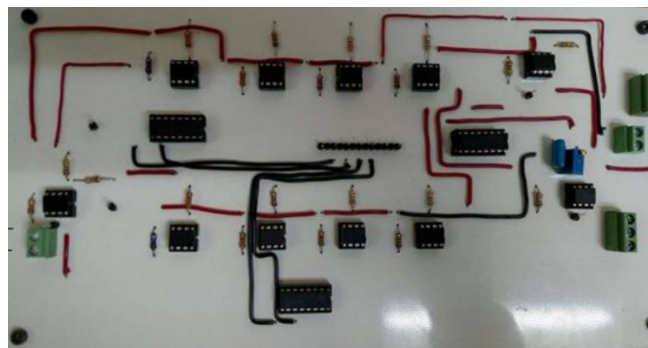


Figure 7 Overview of the developed 64-QAM board.

The figure 8 shows working the 64-QAM modulator, in this case a randomized bit stream was send as information and was packed in symbols (6 bits for 64-QAM), this information was implemented thought the board Arduino Uno taking advantage of the open source platform but other controllers can be adopted to the input port.



Figure 8 Photo of the measurement setup of the 64-QAM modulator.

3. Results

The predicted results based on the equations 4 and 5 were compared with the obtained results of the test bench setup showed in the figure 9. The input data is sent from host computer using Matlab for the result and by other way the data for the board is sent from the Arduino Uno platform, this study integrates a complete high performance for a digital serial stream. The figure 9 shows the signal with a phase offset of 90° in this case: a) represent the model in Matlab and b) the result in the scope for the first stage.

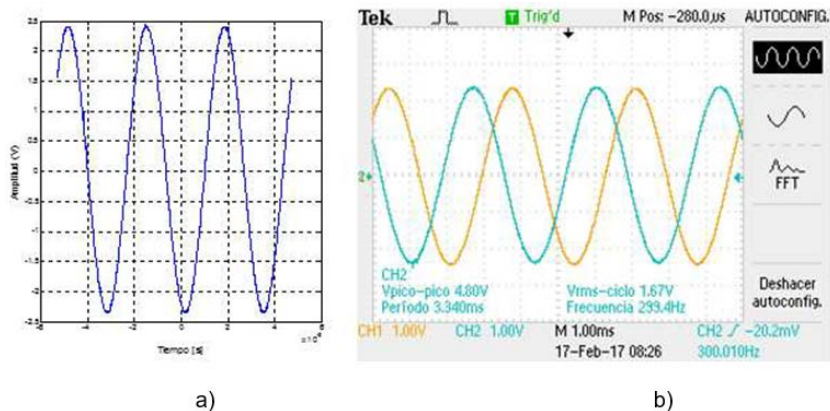


Figure 9 Obtained signal in the first stage with a phase delay of 90° in Matlab and Obtained signal in the scope measured in the Laboratory.

The figure 10 represents the signal with a phase offset of 90° taking into account the bit stream in the input port, in this case the bits numbered as 3 and 6 are controlling the delay stages, a) represent the model in Matlab and b) the result in the scope for the phase delay stages.

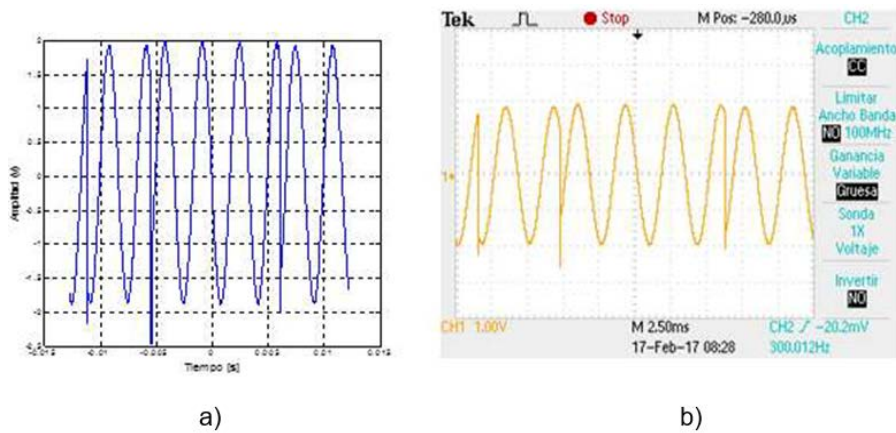


Figure 10 Obtained signal in the first stage with a phase delay of 90° in the two stages in Matlab and Obtained signal in the scope measured in the Laboratory.

The figures 11 and 12 show the amplitude changes after that the phase delay process was implemented, in this case are tested the two amplification stage controlled by four bits, a) represent the obtained model in Matlab and b) the result in the scope after the amplification process involving four amplifiers for each stage.

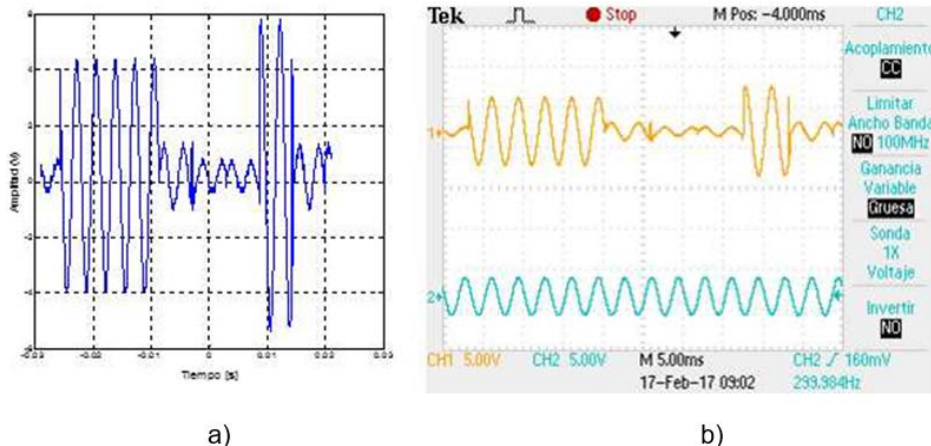


Figure 11 Amplified signal previous to the final stage controlled by two bits after the phase delay process for a) Matlab result and Obtained signal in the scope.

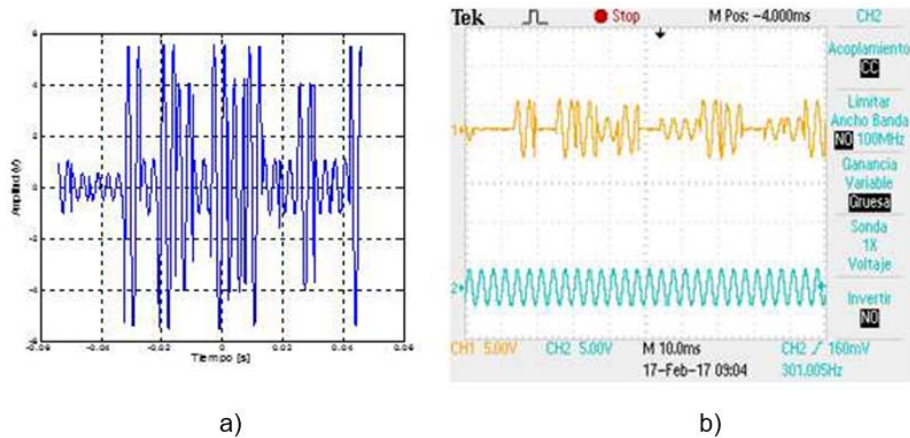


Figure 12 Amplified signal of the second stage previous controlled by two bits after the phase delay process for Matlab result and Obtained signal in the scope.

The used devices allow to work up to 1.3 MHz, the figure 13 shows a general waveform of a 64-QAM signal, a further works that include this methodology allow to emigrate to a higher band in the spectrum.

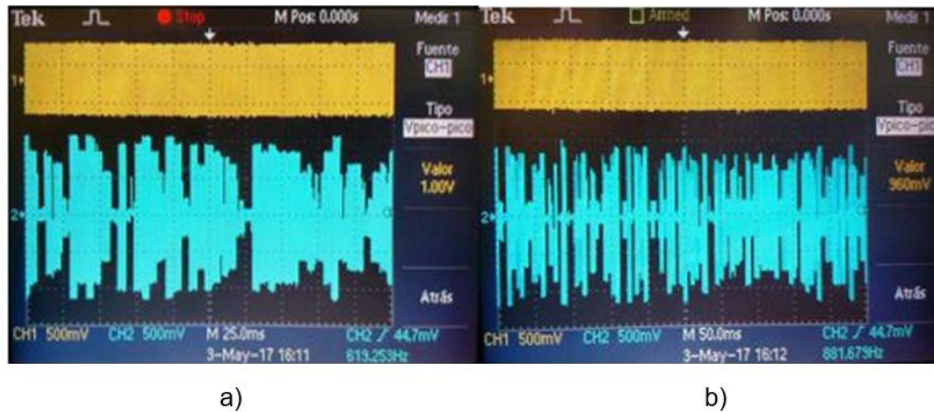


Figure 13 Two general waveforms of a 64-QAM signal up to 1.3 MHz.

4. Discussion

The hardware developers require the use of flexible platform in order to understand properly the performance of a digital signal, as in this case the 64-QAM. Experimental results show firstly in protoboard a result of NMSE= -19.5 dB that was strongly improved to NMSE= -51 dB with the use of this methodology. According with the obtained results the objective was reached and is showing with a high accuracy the signals between stages.

The results modeled in Matlab compared with the general model of a N-QAM system and the measurements done in the laboratory tell us that the board fabricated through CNC machine was properly designed. A lot of knowledge about developing impressed board was gain by the authors.

5. Conclusions

The conclusions are summarized as follows:

- The simulated model and the general performance of the fabricated board had an error or NMSE=-51 dB improving the NMSE=-19.5 dB of a schematic develop in protoboard.
- This work derives in a hardware design tool for analysis of communication links that use 64-QAM.
- The developed tool can be used for academic and research purpose due to the details that comprise each stage.
- The next stage in this study is to measure the spectrum and verify if the spectral regrowth was reduced
- Further work require the use of mounting board an increase the frequency in order to emigrate to VHF and UHF used for ATSC in North America.
- This work contributes with a low cost solution for the national problem of connectivity in Mexico in the analysis stage.

6. Bibliography and References

- [1] ANSI/SCTE, Society of Cable Telecommunications Engineers, ANSI/SCTE 07 2006 Digital Transmission Standard For Cable Television, Engineering Committeee, 2006.
- [2] Yan, S. et al., Generation of 64-QAM signals using a single dual-drive IQ modulator driven by 4-level and binary electrical signals, Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference, pp. 1-3, Anaheim, USA, March 2013.
- [3] Cárdenas-Valdez, J. R., et al., Amplification of 4-, 8-, 16-, 32- and 64-QAM through the Memory Polynomial-Model as Special Case of the Volterra Series

- Implemented in a RF Satellite Link, in IEEE Ninth Electronics, Robotics and Automotive Mechanics Conference, pp. 349-352, Cuernavaca, Nov. 2012.
- [4] Besnoff, J. and Ricketts, D. Quadrature Amplitude Modulated (QAM) Communication Link for Near and Mid-Range RFID Systems, pp. 151-157, San Diego, USA, April 2015.
- [5] Chang, C., Li, Y. and Hirata, J. New 64-QAM Golay Complementary Sequences, In IEEE Transactions on Information Theory vol. 56, no. 5, pp. 2479-2485, May 2010.
- [6] Correa, R. Performance Analysis Of M-QAM with Viterbi Soft-Decision Decoding, Master of Science in electrical engineering thesis, Naval Postgraduate School, March 2003.
- [7] Gnauck, A. H. et al., Generation and Transmission of 21.4-Gbaud PDM 64-QAM Using a Novel High-Power DAC Driving a Single I/Q Modulator, Journal of Lightwave Technology, vol. 30, no. 4, December 2011.
- [8] Kameda, S. et. al., Coverage estimation of uplink 64 QAM signal up to 20 MHz bandwidth based on field trial results: coverage issue of broadband uplink signal, Wireless Personal Multimedia Communications (WPMC), 14th International Symposium on Wireless Personal Multimedia Communications, pp. 1-5, Brest, France, October 2011.
- [9] Lee, H. and Golomb, S. W. A new construction of 64-QAM golay complementary sequences, IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1663-1670, Melbourne, Australia, April 2016.
- [10] Oguma, H. et al., Feasibility Study of Uplink Transmission with 64 QAM Based on Results of MBWA System Field Trial, IEEE 5th Broadband Wireless Access Workshop, Hawaii, U.S.A., 2009.
- [11] Shin, J. et al., The Implementation of 256 QAM CDMA Modulator, Chapter High-Speed Networks and Multimedia Communications, Lecture Notes in Computer Science, vol. 2720, pp 326-332, 2003.
- [12] Vu, X., Duc, N. A. and Vu, T. A. 16-QAM Transmitter and Receiver Design Based on FPGA, Fifth IEEE International Symposium on Electronic Design, Test and Application, pp. 95 – 98, Ho Chi Minh, Vietnam, January 2010.

INTEGRACIÓN DE UN SISTEMA CEREBRO COMPUTADORA EMPLEANDO SOFTWARE LIBRE

Irving Ulises Hernández Miguel

Universidad de la Sierra Sur

irving.u.h.m@gmail.com

Alejandro Jarillo Silva

Universidad de la Sierra Sur

ajarillo@unsis.edu.mx

Víctor Alberto Gómez Pérez

Universidad de la Sierra Sur

vgomez@unsis.edu.mx

Resumen

La generación de sistemas de interacción entre la computadora y el cerebro humano ha crecido en los últimos años gracias al avance tecnológico. En este artículo se muestra y propone la integración de un sistema cerebro computadora, que permite monitorear la actividad eléctrica del cerebro durante la interacción con otros sistemas (e. g. aplicaciones móviles, web, etc.). El objetivo de la integración es emplear tecnología accesible y software libre, además de proporcionar al laboratorio de usabilidad de la Universidad de la Sierra Sur un sistema capaz de medir la actividad eléctrica aplicando diferentes pruebas de usabilidad, de tal manera que se obtenga información cuantitativa durante dicha interacción. En el desarrollo se hicieron diferentes pruebas con diversas tecnologías. Una de las primeras fue construir la etapa de adquisición y procesamiento de las señales, otra fue realizar pruebas con tecnología de bajo costo que integran la etapa de adquisición y filtrado, y la última etapa fue proponer una arquitectura para la integración del sistema empleando software libre. Para demostrar la funcionalidad del sistema, se llevaron a cabo dos experimentos, donde los usuarios realizaron tareas específicas con diferentes grados de dificultad durante la interacción con un

sitio web, y de manera paralela se registró la actividad eléctrica de cada uno de ellos.

Palabras Claves: EPOC+, interacción cerebro computadora, software libre.

Abstract

The generation of systems of interaction between the computer and the human brain has grown in recent years thanks to technological advances. This article shows and proposes the integration of a computer brain system, which allows monitoring the electrical activity of the brain during interactions with other systems (e. g. mobile applications, web, etc.). The objective of the integration is to use accessible technology and free software, in addition to providing the software usability lab of the Universidad de la Sierra Sur with a system capable of measuring the electrical activity by applying different usability tests, in order to obtain quantitative information during the interaction. In the development different tests were made with diverse technologies. One of the first steps was to build the stage of acquisition and processing of signals, another stage was to perform tests with low-cost technology that integrate the stage of acquisition and filtering, and the last stage was to propose an architecture for system integration using free software. To demonstrate the functionality of the system, two experiments carried out, in which the users performed specific tasks of different degrees of difficulty during the interaction with a website, and simultaneously the electrical activity of each one was recorded.

Keywords: Brain computer Interaction, EPOC+, free software.

1. Introducción

Nuestro cerebro produce pequeños impulsos eléctricos (potenciales de acción) que viajan a través de las neuronas. Estos impulsos eléctricos forman ritmos que son conocidos como señales u ondas cerebrales [Psicología de la percepción visual, 2017]. Las señales cerebrales muestran la actividad cerebral y pueden observarse en un electroencefalograma mediante el uso de un electroencefalógrafo, figura 1.



Figura 1 Electroencefalograma.

La llegada del electroencefalógrafo produjo numerosas investigaciones de las ondas cerebrales y estados de conciencia. Esto dio lugar a una clasificación de las señales cerebrales; Beta, Alfa, Theta, Delta y Gamma [Rojas et al., 2012], figura 1c. Las ondas Beta con frecuencias de 14 a 40 Hz, se producen cuando el cerebro está despierto o se encuentra en actividades mentales intensas, las ondas Alfa con 7.5 a 14 Hz se manifiestan cuando hay una escasa actividad cerebral o relajación, las ondas Theta con 4 a 8 Hz se alcanzan en un estado de calma profunda, las ondas Delta con frecuencias de 0.5 a 4 Hz se generan cuando hay un estado de “sueño profundo” y finalmente las ondas Gamma con frecuencias mayores a los 40 Hz que se asocian a una repentina introspección [Rojas et al., 2012].

El estudio de patrones de señales cerebrales se había limitado a que sólo pudieran hacerlo instituciones (eg institutos neurológicos privados, centros de investigación de biomédica, etc.) que contaban con dispositivos biomédicos especializados. Hoy en día con el nacimiento de nuevas tecnologías se abre una puerta al campo científico para abundar más en la investigación, la cual involucra las señales eléctricas que ocurren en el cerebro, para ello es necesario integrar diferentes módulos, mismos que con la ayuda de la informática y de ingeniería es posible desarrollar una tecnología, la cual tiene como objetivo principal ser flexible y factible para estudiantes, profesores e investigadores. Es decir, contar con una tecnología de fácil implementación o utilización que obtenga la información de los impulsos eléctricos del cerebro.

La forma de obtener dicha información cerebral es hacer una lectura del voltaje de los impulsos eléctricos que generan los grandes conjuntos de neuronas (redes de

neuronas) [Erp, 2012]. Para obtener estos voltajes se podría emplear una tecnología invasiva, es decir sensores en contacto directo con el cerebro colocados a través de cirugía [Sepúlveda, 2011] (figura 2a), esta tecnología es la ideal, debido a que se hace una lectura directa del cerebro, pero su realización requiere especialistas en el área médica, lo cual resulta difícil, costoso y sobre todo invasivo. La alternativa sería utilizar tecnologías no invasivas [Sepúlveda, 2011], las cuales consisten en colocar electrodos en contacto directo con el cuero cabelludo, figura 2b.

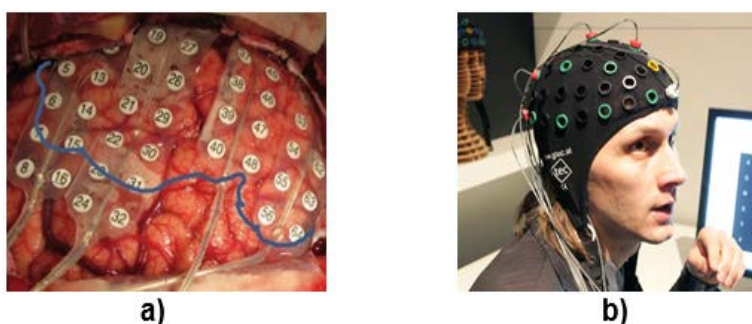


Figura 2 Aplicación de electrodos con tecnología invasiva, tecnología no invasiva

Actualmente las tecnologías más accesibles que podrían ayudarnos a la investigación de las señales cerebrales son las denominadas tecnologías BCI (Interfaz Cerebro Computadora “Brain Computer Interface”). Una Interfaz Cerebro Computadora es un sistema de ingeniería capaz de traducir nuestras intenciones en interacción real con un mundo físico o virtual [Sepúlveda, 2011]. Existen tecnologías comerciales que entran en la clasificación de las tecnologías BCI no invasivas [Erp, 2012]. Algunos ejemplos de las más populares son; la diadema MindWave de la empresa NeuroSky, los cascos EPOC+, EPOC, EPOC Insight de la empresa Emotiv, la banda BrainBand de la empresa MyndPlay y la diadema XWave headset de la empresa PLX devices. Con ellas se podría crear un sistema BCI que obtenga las señales cerebrales que serían estudiadas o utilizadas para algún propósito.

La cantidad de aplicaciones de un sistema BCI es inmenso, desde aplicaciones médicas en las que los pacientes puedan ser tratados para rehabilitación motriz, investigaciones en electroencefalografía, la industria de los videojuegos o en la

interacción con ambientes virtuales hasta ambientes virtuales de realidad aumentada. También es posible llevar a cabo implementaciones en cómputo ubicuo [Santiago, 2015], como la activación de dispositivos con la mente, por ejemplo, encender la luz de un cuarto o un televisor con tan solo pensarlo [Román, 2012]. En este artículo se propone la integración de una arquitectura basada en el uso de tecnología BCI, el objetivo es determinar los niveles actividad eléctrica que manifiestan los usuarios al interactuar con otro tipo de sistemas, y de esta manera el evaluador tendría una herramienta más para determinar la usabilidad de un sistema en particular. Por otra parte, el laboratorio de Usabilidad de la Universidad de la Sierra Sur carece de herramientas para llevar a cabo evaluaciones de usabilidad basándose en la actividad eléctrica del cerebro, es por ello que dicha integración puede ser aplicada en dicho laboratorio.

Este artículo se encuentra estructurado de la siguiente manera: en la sección 2 se encuentra la parte de métodos, en esta sección se describe la metodología y todo el proceso de integración, en las secciones 3 y 4 se presentan resultados y discusiones, y finalmente en la sección 5 se dan las conclusiones y trabajos futuros.

2. Métodos

En este apartado se describen los materiales y métodos que se utilizaron en las cuatro fases de experimentación.

Materiales

- Técnica de aplicación de electrodos para EEG Electro-Cap (figura 3a) de la empresa Electro-Cap-International.
- Diadema MindWave (figura 3b) de la empresa NeuroSky.
- Casco EPOC+ (figura 4) de la empresa Emotiv.
- Software Libre.
 - ✓ Arduino.
 - ✓ Linux Ubuntu 14.04
 - ✓ Software de graficación GNUPlot.

- ✓ Software para automatización de código CMake
- ✓ Sistema Gestor de Base de Datos MySQL
- ✓ Software Emokit
- ✓ Lenguajes de programación; Java y C.
- Hardware
 - ✓ Equipo de cómputo
 - ✓ Arduino Hardware Open Source
 - ✓ Componentes electrónicos (amplificadores de instrumentación y operacionales)



Figura 3 Técnica EEG Electro-Cap., colocación de la diadema en usuario.

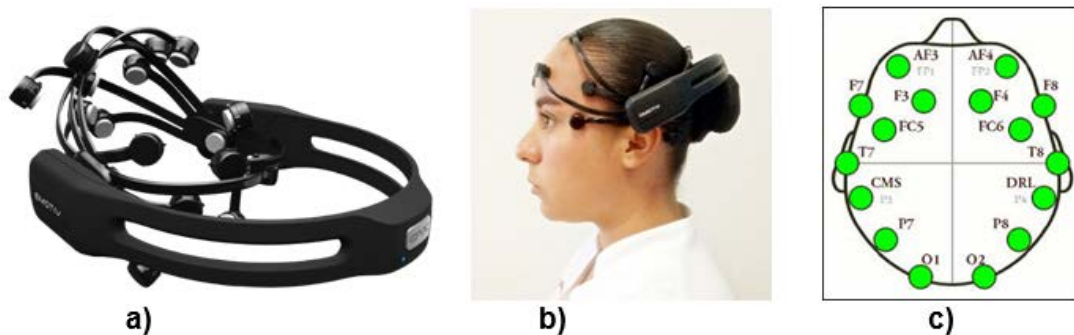


Figura 4 Casco EPOC+, Colocación del casco en usuario, posición de electrodos.

La técnica de colocación de electrodos para EEG Electro-Cap, figura 3a, es una especie de malla para la cabeza que están hechas de un tejido de tipo spandex elástico con electrodos de estaño empotradas a la tela. Los electrodos en las

tapas estándar están posicionados para el método internacional de colocación de los electrodos 10-20 [Caps, 2017].

La diadema MindWave, figura 3b, es también denominada como un auricular inalámbrico EEG, que cuenta con características muy particulares, tabla 1, y es la culminación de décadas de investigación en tecnología de ondas cerebrales con biosensores EEG de laboratorio. Esta diadema se adapta cómodamente en la cabeza, figura 3b, ya que tiene aplicaciones en juegos, en la educación, y la investigación [Mindwave, 2017].

Tabla 1 Características de la diadema NeuroSky MindWave.

Características MindWave	
Principales	<ul style="list-style-type: none">• Conexión directa (electro seco).• Un canal EEG + Referencia + Tierra.• Detección de señal de nivel extremadamente bajo.• Filtro avanzado con alta inmunidad al ruido.• RAW EEG a 512 Hz.• Frecuencia de muestreo de 512 Hz.• Rango de frecuencia de 3-100 Hz.• Protección ESD: Contacto 4 kV.• Descarga: Aire 8 kV.• Consumo máximo de energía: 15 mA 3.3 V.• Voltaje de operación: 2.97 ~3.63 V.• UART (Serial): 1200, 9600, 57600 baudios.
Datos de salida	<ul style="list-style-type: none">• Señales RAW EEG.• Atención.• Meditación.• Ondas Delta, Theta, low alpha, high alpha, low beta, high beta y gamma.

El casco EPOC+ figura 4a es una tecnología BCI comercial de la empresa australiana Emotiv [Epoc, 2017] denominada también como una neuroheadset EEG inalámbrico, que ofrece alta resolución de 14 canales asociados a una posición, tabla 2 y figura 4c diseñado para aplicaciones avanzadas de interfaz cerebro computadora (BCI) y la investigación contextualizada. El EPOC+ proporciona acceso de alta calidad de datos EEG y una gran comodidad en su colocación, figura 4b.

Tabla 2 Características del casco Emotiv EPOC+.

Características EPOC+	
Señales	<ul style="list-style-type: none">• 14 canales: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF42• Referencias: P3/P4 para configuración de cancelación de ruido CMS/DRL.
Resolución de las señales	<ul style="list-style-type: none">• Método de muestreo: Muestreo secuencial con ADC individual.• Velocidad de muestreo: 128 SPS o SPS* (2048 Hz Internos)• Resolución: 14 bits o 16 bits• Ancho de banda: 0.2–43 Hz, filtro digital Notch digital a 50 Hz y 60 Hz• Filtración: Filtro Sinc digital de 5to orden• Rango dinámico (entrada referida): 8400 μV(pp)• Modo de acoplamiento: AC acoplado
Conectividad	<ul style="list-style-type: none">• Inalámbrico: Bluetooth® Smart• Propietario inalámbrico: Banda de 2.4 GHz
Alimentación	<ul style="list-style-type: none">• Batería: Batería interna de polímero de litio de 640mAh• Duración de la batería: Hasta 12 horas

Desarrollo y Experimentación

En este apartado se describe cada fase de experimentación. Éstas se basan en el uso de una técnica EEG o tecnología BCI. La primera fase se enfoca en la construcción de una tecnología BCI empleando la técnica EEG Electro-Cap. La segunda fase es la modificación a nivel hardware y software de la tecnología MindWave. La tercera fase de experimentación se basa en utilizar el casco EPOC+ empleando el software libre Emokit, con el fin de extraer y procesar los datos de los sensores de dicho casco para la implementación de un graficador de señales. Por último, en la cuarta fase de experimentación se realizan pruebas a dos usuarios para medir la actividad eléctrica de su cerebro durante la interacción con un sitio web.

Primera Fase de Experimentación

Se construyó tecnología BCI no invasiva utilizando componentes electrónicos; circuitos integrados, amplificadores de instrumentación, amplificadores operacionales y software libre (Linux, Java y C). Se implementaron circuitos electrónicos analógicos y digitales especializados en conjunto con la técnica de EEG Electro-Cap. Además, se elaboraron PCBs [Cifuentes, 2010] empleando

software de simulación y diseño; Proteus, ISIS y National Instruments Multisim. En la parte de la amplificación y filtrado de las señales se utilizaron los amplificadores de instrumentación INA128 y AD623 [Cifuentes, 2010]. Cabe señalar que debido a la presencia de ruido eléctrico los resultados en las señales fueron poco legibles.

Segunda Fase de Experimentación

Se realiza una búsqueda de tecnología BCI ya existente, la cual podría ser la solución a la obtención de las señales. La diadema MindWave se adquirió y modificó [MindWave, 2017], de tal manera que las señales entregadas ahora son procesadas por una tarjeta Arduino.

Al realizar pruebas de contacto con esta tecnología se logró obtener, procesar y graficar las señales cerebrales con una velocidad de muestreo de 1 Hz. Sin embargo, esta lectura es muy lenta, ya que en el intervalo de un segundo se manifiestan, mismas que no son obtenidas por la diadema. Además de contar con un electrodo, que se coloca en la frente, el cual no alcanza a cubrir las diferentes zonas del cuero cabelludo, por lo tanto, no se tiene una representación válida de la actividad eléctrica del cerebro. A partir de los resultados del muestreo de 1 Hz y la limitante de utilizar un solo electrodo, la tecnología MindWave es descartada.

Tercera Fase de Experimentación

Con la finalidad de resolver el problema del muestreo y hacer un barrido de la mayor cantidad de señales del cerebro se toma la decisión de emplear la tecnología EPOC+ en la integración del sistema que se encuentra constituido por la siguiente arquitectura, figura 5:

1. *Módulo de actividad cerebral*: para obtener las señales del cerebro se necesita un usuario, el cual realizará diferentes actividades según cada caso.
2. *Módulo de obtención de señales*: la tecnología EPOC+ se va encargar de obtener, filtrar y proporcionar cada una de las señales de los 14 sensores con los que cuenta. El muestreo de trabajo es de aproximadamente 128 Hz [EPOC, 2017].

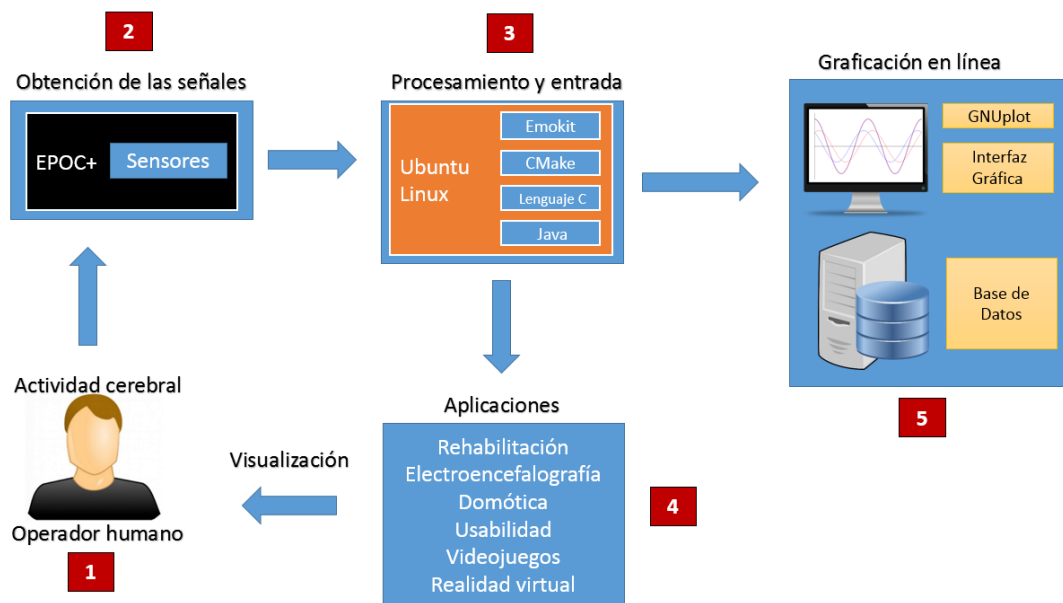


Figura 5 Arquitectura general para la visualización de las señales cerebrales.

3. *Módulo de procesamiento y entrada*: los datos son procesados y almacenados por una computadora con software libre; Linux Ubuntu 14.04, CMake, C y Java.
4. *Módulo de aplicaciones*: para esta aplicación en especial, el evaluador tendrá la opción de visualizar en tiempo real la activación eléctrica del cerebro, además de guardar en un archivo CSV todos los datos para su posterior análisis de ser necesario.
5. *Módulo de graficación*: este módulo conlleva la generación de la representación gráfica de las señales eléctricas, para ello se emplearon herramientas como Java, C y GNUPlot. Las lecturas hechas son guardadas en una base datos en MySQL que podrían servir para estudios posteriores.

La tecnología EPOC+ es el módulo clave para hacer la integración del sistema, esta tecnología comercial es de las más avanzadas a nivel mundial. Este casco requiere de software comercial, el cual se limita en función del costo de este. Lo que quiere decir que cuenta con distintos tipos de licencia y costos para su uso [Epic, 2017]. Al hacer la inversión de adquirir el dispositivo se toma el camino de crear y utilizar el software libre Emokit para obtener la información del casco

EPOC+ y lograr implementar el módulo de Obtención de señales cerebrales. Con esto se evita el uso del software de la empresa que tiene un costo significativo.

Para los experimentos se utilizó una computadora con las siguientes características; Procesador Intel Core i3-3220 a 3.30 Ghz, Memoria RAM 6 GB, Ubuntu 14.04 y Tarjeta de video AMD Radeon Graphics.

Se obtienen los datos guardados en archivos CSV de las señales cerebrales del casco EPOC+ utilizando software libre. Este software denominado Emokit [Open Emotiv, 2017] se modificó para que los datos del casco EPOC+ pudieran ser obtenidos, procesados y visualizados numéricamente en la computadora, figura 6(a). Los datos del archivo CSV describen la calidad de contacto, lectura del voltaje de los sensores, la posición del giroscopio, la cantidad de ciclos de lectura y el estado de la batería, tabla 3 y figura 6a.

Tabla 3 Datos obtenidos con software libre Emokit modificado.

Datos en archivo CSV	
Calidad de contacto	<ul style="list-style-type: none">Datos etiquetados en calidad Bueno, Regular y Malo.
Sensores	<ul style="list-style-type: none">Sensores: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF42.Valores con rango de 0 a 16383.
Giroscopio	<ul style="list-style-type: none">Valores en eje X e Y con rango de -128 a 128.
Ciclos de lectura	<ul style="list-style-type: none">Valores con rango de 0 a 128.
Batería	<ul style="list-style-type: none">Valores con rango de 0 a 100.



Figura 6 Captura de pantalla de datos crudos del EPOC+ y el graficador de señales.

Al tener los datos de manera numérica, es necesario poder visualizarlos de alguna manera gráfica que sea lo más intuitiva posible para una fácil interpretación. Una manera de hacerlo es desarrollar una interfaz gráfica que muestre las señales

cerebrales en la computadora, y con esto se desarrolla el módulo de Graficación en línea. Los datos numéricos se grafican como señales cerebrales con el software GNUPlot y lenguaje C, figura 6a.

Para validar que el procesamiento de señales realmente nos proporciona información de lo que sucede en el cerebro, se realizaron pruebas de contacto con un usuario, figura 6b. Para estas pruebas el usuario recibió las siguientes indicaciones: cerrar y abrir los ojos rápidamente, relajarse, hablar en voz alta, reír, mover los ojos y pensar algo profundamente. Los resultados demuestran que cuando el usuario realiza cada una de las tareas mencionadas existe la presencia de diferentes señales, con diferente forma de onda y periodo. Lo cual significa que la adquisición, codificación y filtrado de las señales se está llevando de la manera correcta, figura 6.

Cuarta fase de Experimentación

Este experimento consiste en evaluar la actividad eléctrica cerebral de un usuario al interactuar con un sitio web. Para ello el usuario debe realizar una serie de tareas, mismas que se le estarán informando cuando culmine una de ellas. Con esto se mide su actividad cerebral durante el transcurso de su navegación hasta que termine con éxito la tarea encomendada o de lo contrario a que hayan pasado 5 minutos.

Se eligen dos usuarios, un usuario que está familiarizado con el sitio web de la UNSIS y otro que no lo está. A los usuarios figura 7 se les asigna diversos tipos de tareas, las cuales tienen su respectiva complejidad, tabla 4.



Figura 7 Usuario familiarizado, usuario no familiarizado.

Tabla 4 Tareas para navegación en el sitio web de la UNSIS.

No	Tarea	Nivel de dificultad
1	<ul style="list-style-type: none">Encontrar el plan de estudios de la carrera en informática.	Sencillo
2	<ul style="list-style-type: none">Encontrar el nivel académico de la profesora Teresita de J. Mijangos Martínez.	Normal
3	<ul style="list-style-type: none">Encontrar a los editores de la revista "Salud y Administración" que publica la universidad.	Moderado
4	<ul style="list-style-type: none">Encontrar el número de teléfono de la UNSIS para pedir informes de inscripción.	Difícil
5	<ul style="list-style-type: none">Encontrar el reglamento de la biblioteca de la universidad.	Difícil

3. Resultados

El primer resultado es la obtención de los datos crudos del casco EPOC+ a través de la modificación de la librería Emokit. También la implementación de un visualizador de señales cerebrales empleando software libre (Linux, Java y C).

Las pruebas de medición de la actividad cerebral de los dos tipos de usuario al navegar en un sitio web arrojaron los siguientes resultados:

- **Usuario familiarizado:** en el sitio web con la tarea 3 manifestó las siguientes señales cerebrales tardando un tiempo de 45 segundos en realizar dicha tarea, figura 8a, durante la realización de la tarea el usuario habló en voz alta diciendo hacia dónde se dirigía y mostró una actividad ocular rápida. Al realizar la tarea 4 con complejidad "difícil" el usuario tardó 22 segundos, figura 8b, el cual es un tiempo menor al de la tarea 3 de complejidad "moderada". Esta diferencia está relacionada a que el usuario ya había buscado esa sección de la página del sitio en otras ocasiones. Por lo que recordó fácilmente la ruta y no dudó hacia dónde dirigirse en la navegación.
- **Usuario no familiarizado:** en el sitio web realizó la tarea 2 de complejidad "normal" con un tiempo de 4 minutos y 38 segundos, figura 9a. La tarea 3 de complejidad "moderada" tardó un tiempo de 47 segundos figura 9b. En ambas tareas el usuario mostró cambios de posición corporal, respiraciones

profundas, movimientos faciales y de cabeza notorios, debido a que no encontraba la sección pedida por la tarea.

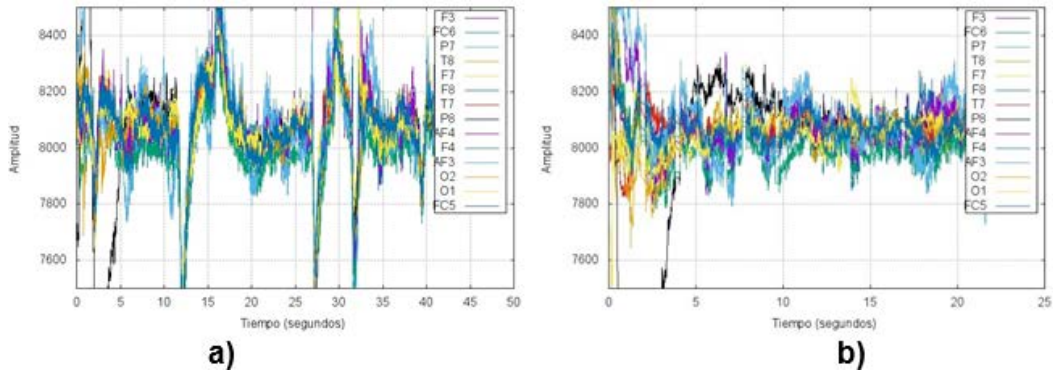


Figura 8 Usuarios familiarizados realizando las tareas 3 y 4.

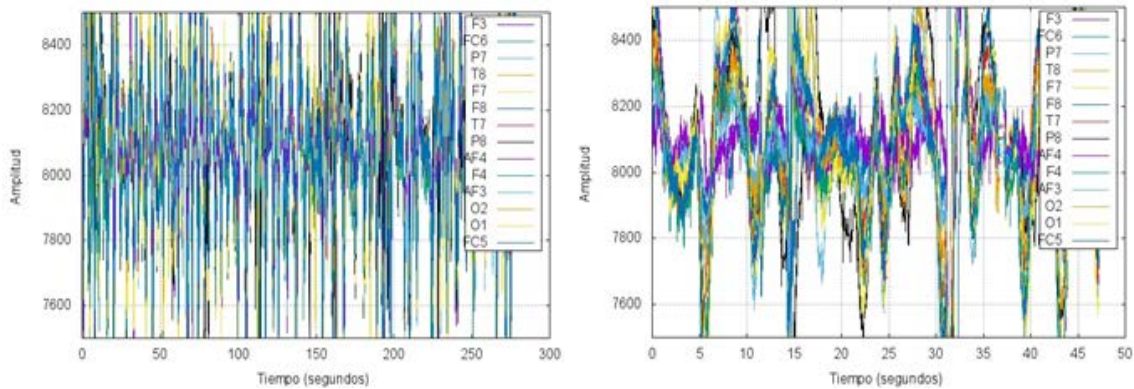


Figura 9 Usuarios no familiarizados realizando las tareas 2 y 3.

4. Discusión

El software Emokit proporciona los datos más importantes (datos crudos de los sensores) como lo hace la API comercial de la empresa Emotiv en su función EEG [Epo, 2017].

El graficador de señales cumple la funcionalidad de mostrar las señales que proporciona cada sensor como lo hace la aplicación EMOTIVPureEEG RAW EEG comercial con la particularidad de que no existe aún una clasificación de ondas [EPOC, 2017] por parte del software Emokit.

Un usuario que realice una actividad que le resulte estresante se verá reflejada en su actividad cerebral en el graficador de señales. Esto se observa al comparar la

realización de la tarea 3 por ambos usuarios (figura 8a y figura 9b), el usuario no familiarizado presenta más oscilaciones y más presencia de actividad eléctrica en sus señales que el usuario familiarizado.

Cuando un usuario no se concentra mucho o no le cuesta trabajo realizar alguna actividad, su actividad cerebral es mínima. Esto se logra observar en la tarea 4 realizada por el usuario familiarizado, ya que su tiempo para realizar la tarea fue poco y la actividad cerebral no presentó oscilaciones notorias.

Es notable la presencia del aumento de la actividad cerebral en el usuario no familiarizado en comparación al familiarizado, basta con observar la amplitud de los picos en todas las tareas del no familiarizado. De esta manera se demuestra que es posible determinar niveles de estrés generados durante la interacción cuando un usuario no es capaz de acceder a una sección de un sitio.

5. Conclusiones

Al poder integrar el casco EPOC+ al sistema y obtener los datos crudos, se puede iniciar desde cero la investigación de búsqueda de patrones de señales, y esto da pauta a poder hacer aplicaciones de las señales cerebrales, es decir que ahora se podría desarrollar un sistema BCI con múltiples aplicaciones en las diferentes áreas científicas. En este artículo mostramos la implementación de un graficador de señales cerebrales, el cual es capaz de proporcionar información verídica de lo que sucede en nuestro cerebro, ya que se comprobó que realmente muestra la presencia de actividad cerebral cuando el usuario realiza una tarea y cuando no la realiza.

Como trabajo a futuro se pretende desarrollar una nueva interfaz visual en lenguaje Java que sirva para tener una mejor manipulación y clasificación de las ondas cerebrales para la medición de usabilidad de aplicaciones en el laboratorio de Interacción Humano Computadora de la Universidad de la Sierra Sur. También se liberará el código para facilitar la extracción de datos del casco EPOC+ y eliminar la limitante del software privativo en este dispositivo, esto se hace con el fin de que más investigadores aborden estos temas con esta tecnología. Por último, se pretende convertir en una API (Interfaz de Programación de

Aplicaciones) el software Emokit añadiéndole nuevas funcionalidades para interactuar fácilmente con lenguajes como Python y Java.

6. Bibliografía y Referencias

- [1] Caps: <http://electro-cap.com/>, 12 de abril de 2017.
- [2] Cifuentes González, I. A., Diseño y construcción de un sistema para la detección de señales electromiográficas. Tesis de licenciatura. México, Universidad Autónoma de Yucatán; 2010.
- [3] Epoc: <https://www.emotiv.com/epoc/>, 12 de abril de 2017.
- [4] EPOC Technical Specifications: <https://www.emotiv.com/epoc/>, marzo 2017.
- [5] Erp, J. Lotte, F. Tangermann, M. Brain-Computer Interfaces: Beyond Medical Applications. IEEE Computer Society. Vol 45. No 4, pp. 26-34, 2012.
- [6] MindWave and Arduino: http://developer.neurosky.com/docs/doku.php?id=Arduino_tutorial, 12 de abril de 2017.
- [7] MindWave Technical Specs: <http://neurosky.com/>, 20 de marzo de 2017.
- [8] Open Emotiv EPOC EEG RAW: <https://github.com/openyou/emokit>, 11 de abril de 2017.
- [9] Psicología de la percepción visual, Las ondas cerebrales: www.ub.edu/pa1/node/130, 14 de marzo de 2017.
- [10] Rojas, S. Garzón, J. Martínez, D. Escobar, M. Robayo, C. Montenegro D., Lector de ondas cerebrales para implementar un sistema alternativo y aumentativo de comunicación. Latin American and Caribbean Conference – International Competition of Student Posters and Papers (LACCEI). No.10, 2012.
- [11] Román Pérez, M. A., Control de un robot manipulador mediante la interpretación de ondas cerebrales. Tesis de maestría. México, CIDETEC Instituto Politécnico Nacional; 2012.
- [12] Santiago López, J. L. Gómez Pérez, V. A. Ramírez Díaz, A. J. Jarillo Silva, A. Santiago López, J. C., Arquitectura de descubrimiento de servicios para

entornos hospitalarios (módulo tiny application). *Pistas Educativas*, No 112, 2015.

[13] Sepúlveda Cervantes, G. Montaña Martínez, N. Román Pérez, M. A., Control de un robot manipulador virtual, utilizando una interfaz cerebro-computadora, Congreso Nacional de Ingenierías Mecánica, Eléctrica, Electrónica y Mecatrónica (CIMEEM), 2011.

[14] Sepúlveda Cervantes, G. Montaña Martínez, N. Román Pérez, M. A., Interfaz Cerebro Computadora para el posicionamiento de un Robot Virtual. XII Simposio Mexicano en Cirugía Asistida por Computadora y Procesamiento de Imágenes Médicas (MEXCAS), 2011.

ELEMENTOS DE LOS PARQUES EÓLICOS QUE DEBEN SER CONTROLADOS PARA SU INTERCONEXIÓN CON REDES ELÉCTRICAS

Jorge Eduardo Hernández Miranda

Universidad Autónoma Metropolitana-Azcapotzalco

jorge11752@gmail.com

Irvin López García

Universidad Autónoma Metropolitana-Azcapotzalco

ilg@azc.uam.mx

Eduardo Campero Littlewood

Universidad Autónoma Metropolitana-Azcapotzalco

ecl@correo.azc.uam.mx

Francisco Beltrán Carbajal

Universidad Autónoma Metropolitana-Azcapotzalco

fran_belt29@hotmail.com

Victor Manuel Jiménez Mondragón

Universidad Autónoma Metropolitana-Azcapotzalco

vmjm1986@gmail.com

Resumen

En este trabajo se presentan los requisitos que los parques eólicos deben cumplir para su interconexión a una red eléctrica en México. El trabajo se enfoca al análisis de los elementos técnicos que requieren de control, que están incluidos en el Código de Red nacional: parámetros de potencia activa y reactiva, límites de operación de voltaje y frecuencia, así como de la metodología de operación del parque eólico durante una contingencia en la red eléctrica o en el propio parque eólico. También se presenta un estudio de algunas alternativas tecnológicas que

permiten satisfacer los requisitos establecidos en el Código de Red para una interconexión segura, desde el punto de vista de estabilidad transitoria y aprovechamiento de la energía eólica, entre los parques eólicos y la red nacional.

Palabras Claves: Código de red, oscilaciones de voltaje y frecuencia, parque eólico, potencia activa y reactiva, red eléctrica nacional.

Abstract

This paper presents the requirements that the wind farms are fulfilled for their interconnection to an electric network in Mexico. The work focuses on the analysis of the technical elements that require control, which are included in the National Grid Code: parameters of active and reactive power, voltage and frequency operation limits, as well as the methodology of operation of the wind farm during a contingency in the electric network or in the wind farm itself. It also presents a study of some technological alternatives that allow to satisfy the requirements established in the Grid Code for a secure interconnection, from the point of view of transient stability and use of wind energy, between wind farms and the national grid.

Keywords: *Active and reactive power, frequency, grid code, national electric network, voltage sags, wind farm.*

1. Introducción

Las fuentes de energías renovables están tomando un papel cada vez más importante en el mundo por el fenómeno del calentamiento global provocado por los gases de efecto invernadero [INECC, 2017], [IPCC, 2017]. La generación de energía eléctrica convencional, basada en combustibles fósiles, tiene un efecto importante en este fenómeno. Es por ello que el uso de fuentes de energía limpia, como lo es el viento en la generación de la energía eléctrica que según el Consejo Global de Energía Eólica (GWEC, por sus siglas en inglés) tiene un crecimiento del 20% anual. De acuerdo con la Asociación Mexicana de Energía Eólica (AMDEE), la Secretaría de Energía (SENER) y la Comisión Reguladora de Energía (CRE), México es el segundo país en Latinoamérica más importante en el

uso de la generación eólica se ha planteado como reto incrementar la potencia de generación de energía eléctrica con esta fuente a 15,000 MW para los años 2020-2022 [CRE, 2016], [DOF, 2016], [GWEC, 2015].

Para el cumplimiento de este reto, se requiere llevar a cabo una evaluación de los elementos que integran los parques eólicos y el impacto que estos tienen en su interacción con la red eléctrica por los problemas inherentes a la variabilidad del viento. El análisis incluye las variaciones de tensión, frecuencia, potencia activa y reactiva, e intensificación de carga [Cialdea, 2012], [Jadhav, 2011]. Existe información en la literatura básica y especializada al respecto de los problemas que pueden afectar la seguridad y calidad de la energía de la red eléctrica [Chen, 2005], [Chompoo, 2005], [CRE, 2016], [DEFU Comittee Reports, 1998], [El Moursi, 2008], [Ekraft systems, 2004], [Heier, 2014], [IEC-61400-12], [IEC-61300-21, 2001].

El objetivo de este trabajo es presentar un análisis general de las características que presenta un parque eólico cuando se interconecta con una red eléctrica. Se abordan temas como: la planeación del parque, problemas debido a la naturaleza del viento, distribución de aerogeneradores, cuestiones de desempeño y generación, así como opciones con las que se pueda mejorar su eficiencia en la generación y suministro de la de energía eléctrica. La intención del trabajo es señalar los problemas que se presentan al momento de la interconexión de estos sistemas con la red eléctrica y con ello estar en la posibilidad de hacer recomendaciones que hagan que se respeten los límites y valores establecidos por el Código de Red Nacional [CRE, 2016].

El trabajo está organizado de la siguiente manera. En la sección 2 se describe el concepto de parque eólico, los componentes que lo conforman y las problemáticas que este presenta en su funcionamiento tanto en condiciones de operación generales, como en su interconexión a la red eléctrica. En la sección 3 se presentan los parámetros establecidos por el código de red nacional para la interconexión del parque eólico a la red eléctrica, en la sección 4 se analizan estos resultados, por último, en la sección 5 se presentan las conclusiones del trabajo.

2. Métodos

Un parque eólico es una agrupación de aerogeneradores que utiliza la energía cinética del viento para generar energía eléctrica a través de turbinas eólicas, generalmente de eje horizontal. Los componentes básicos de un aerogenerador son: la torre, turbina eólica (rotor con tres álabes unidas al cubo), eje con engranaje mecánico (multiplicador), generador eléctrico, mecanismo de guiñada (paleta de cola, sensores) y algoritmos de control [Patel, 2005], [Retana, 2016]. Aunque la capacidad de generación de un parque eólico depende sobre todo del recurso eólico en la zona donde se encuentra ubicado, la potencia del parque eólico también puede estar condicionada por las especificaciones de operación de la red eléctrica a la cual está interconectado [Patel, 2005], [Rajib, 2002] [Muller, 2002].

El constante incremento de la capacidad eólica incluida en las redes eléctricas hace necesario que la generación eólica tenga un funcionamiento que colabore con la estabilidad de la red [Erlich, 2006], [Pearmine, 2007]. Es necesario conocer los efectos claves causados por la integración de la energía eólica a gran escala en el sistema de energía. Por esta razón, los códigos de red expuestos actualmente exigen que los grandes parques eólicos soporten variaciones de tensión, que se especifican en función del porcentaje de cambio y la duración de la variación. Tales requisitos se conocen como Fault Ride Through (FRT), que significa, requerimientos para resistir una falla sin sufrir daños o requisitos de transición de baja tensión (LVRT) Low Voltage Ride Through y se describen por una característica de voltaje contra tiempo, que denotan la resistencia mínima requerida de la central de energía eólica a la baja tensión del sistema [Tsil, 2009], [Patel, 2005], [Ackerman, 2005]. Los requisitos de FRT incluyen una restauración de potencia de salida activa y reactiva rápida a los valores previos a la falla, después de que la tensión del sistema vuelva a su funcionamiento normal, con el fin de soportar la tensión del sistema. Las centrales eólicas pueden participar activamente en la operación de la red y en el control mediante la regulación de su potencia de salida. Todos los códigos de red actualmente imponen requisitos sobre las capacidades de regulación de la potencia activa de los parques eólicos.

Dentro de la potencia activa disponible (condiciones de viento predominantes), la potencia de salida puede regularse a un valor específico [Ackerman, 2005], [Ekraft systems, 2004], o tener una relación fija con la potencia disponible de manera que se mantenga una reserva especificada, ya sea en MW o como porcentaje de la potencia disponible [Ackerman, 2005], [Ekraft systems, 2004]. Los requisitos adicionales incluyen la limitación de la velocidad de cambio de la potencia de salida [Ackerman, 2005], [Ekraft systems, 2004]. Las velocidades de rampa son posibles para los aumentos de potencia, pero la operación con una reserva de marcha es necesaria para ser efectiva cuando la potencia de salida disminuye.

Las capacidades de regulación de potencia reactiva son requeridas por muchos códigos (en algunos países se le conoce como “de cuadrícula”). Esto se efectúa ya sea proporcionando externamente un valor de potencia reactiva específico o mediante un factor de potencia específico. Además, la capacidad de regulación de la potencia reactiva puede explotarse para el control de la tensión en el punto de conexión del parque eólico a la red eléctrica, o en un nodo más distante. Las referencias [Ackerman, 2005], [Ekraft systems, 2004] muestran los requisitos típicos para el rango de regulación del factor de potencia, en función del voltaje de los terminales y de la potencia de salida activa del parque eólico, respectivamente, los cuales van de 1.0 a 0.95 tanto en adelanto como en atraso.

En relación con los problemas que presenta el parque eólico, existe literatura especializada [Abad, 2011], [Ackerman, 2005], [Eisa, 2017], [Ferdous, 2016] [Muljadi, 1998], [Muller, 2002]. [Patel, 2005], [Petru, 2002], [Rajib, 2002], [Rosmin, 2012], [Thiringer, 2004], [Xing, 2016], [Zhang, 2011] donde se presentan elementos específicos como metodologías de “pitch control” (control del ángulo de ataque) [Eisa, 2017], [Ferdous, 2016], [Xing, 2016], [Zhang, 2011], “stall regulated” (control aerodinámico de pérdida de velocidad) [Muljadi, 1998], [Petru, 2002], [Rosmin, 2012], [Thiringer, 2004] y selección del generador eléctrico [Abad, 2011], [Ackerman, 2005], [Muller, 2002], [Patel, 2005], [Rajib, 2002]. Lo que presenta el método de “pitch control” es un control activo que hace variar el ángulo de ataque, es decir, gira los álabes alrededor de su eje, para disminuir el par producido en una turbina de velocidad fija y para disminuir la velocidad de rotación en turbinas

de velocidad variable. Este tipo de control se emplea normalmente para evitar que las altas velocidades de viento (generalmente por encima de la velocidad nominal), provoquen altas velocidades de rotación que pudieran dañar el equipo. Cuando las velocidades del viento llegan a ser muy altas, los álabes se hacen girar de manera que haya menos elevación y más resistencia debido al aumento de la separación del flujo a lo largo de la longitud de la hoja. Esto reducirá la velocidad de rotación de la turbina o el par transferido al eje de modo que la velocidad de rotación o el par se mantenga constante. Por otro lado, los aerogeneradores por “stall regulated” tienen sus álabes diseñados para que cuando la velocidad del viento sea alta (arriba de cierto valor), se disminuya la producción de energía. La disminución de potencia con velocidades de viento altas se debe a los efectos aerodinámicos en los álabes del aerogenerador. El beneficio de “stall regulated” sobre el “pitch control” es el costo de la turbina, así como un menor mantenimiento asociado con menos partes móviles. Al igual que el aerogenerador controlado por “pitch control”, el aerogenerador controlado por “stall regulated” también tiene frenos para detener la turbina a velocidades extremas del viento.

La diferencia entre el aerogenerador por “pitch control” y “stall regulated” radica en que los sistemas de regulación por “stall regulated” dependen del diseño aerodinámico de los álabes para controlar la velocidad de rotación del aerogenerador en altas velocidades del viento, los sistemas regulados por “pitch control” utilizan un control de paso activo para las cuchillas, que permite que los sistemas regulados tengan una potencia de salida constante, mientras que los sistemas con “stall regulated” no son capaces de mantener una potencia constante en vientos fuertes.

Asimismo, dependiendo del ángulo y dirección con la que el viento impacta los aerogeneradores del parque eólico tenemos algunos problemas como son, vibraciones y perturbaciones en los aerogeneradores del parque eólico. Se encuentra el fenómeno denominado “wake effect” [Akdag, 2013], [Rebecca, 2012], [Rolán, 2010], [Sun, 2009], también conocido como efecto sombra, que provoca la disminución de la potencia de giro del aerogenerador colocado atrás (de acuerdo

con la dirección del viento), es decir, la turbulencia provocada en ese aerogenerador representa pérdidas de generación. Por ello para minimizar las pérdidas por este efecto se proponen métodos de distribución de los aerogeneradores a lo largo de la zona del parque. La separación adecuada entre aerogeneradores es de 3 a 4 veces el diámetro del rotor en los costados y de 6 a 8 veces el diámetro del rotor en la dirección del viento, figura 1.

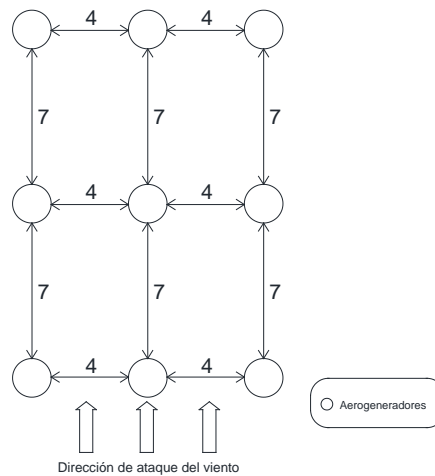


Figura 1 Distribución de aerogeneradores.

De la misma manera para atender la dirección de impacto del viento al parque se propone un sistema de desconexión de aerogeneradores para mitigar el "efecto sombra" en el parque y reducir la turbulencia entre los elementos del mismo. Ver figura 2.

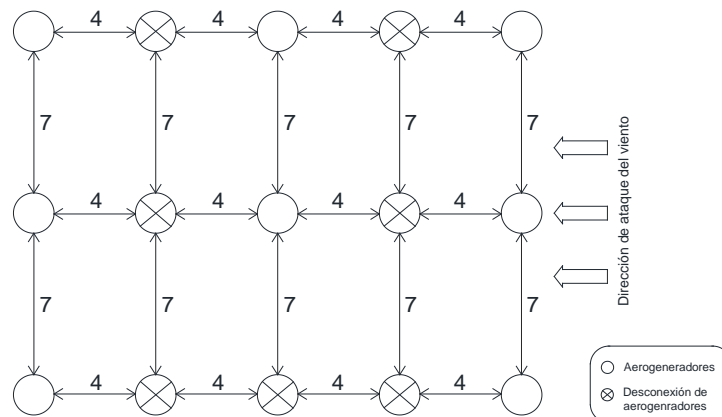


Figura 2 Distribución y desconexión de aerogeneradores debido a la dirección del viento.

Como se pudo observar en los párrafos anteriores, la velocidad del viento toma un papel fundamental para la generación de energía eléctrica. Es por ello que la selección del tipo de generador eléctrico debe hacerse con el estudio de los factores antes mencionados, ya que se cuentan con generadores que son capaces de funcionar a velocidad fija y a velocidad variable, en las cuales podremos encontrar generadores tipo síncrono o asíncrono, comúnmente llamado de inducción [Abad, 2011], [Muller, 2002], [Patel, 2005], [Rajib, 2002], [Rolán, 2010]. Con base en el estudio hecho al perfil de viento en la zona (por lo menos de un año) se puede planear la disposición del diseño del parque.

Para poder interconectar un parque eólico a la red eléctrica, es necesario tener en cuenta cuales son los parámetros que cumplir en la conexión. Al respecto cada país cuenta con un código de red [CRE, 2016]. El cual tiene como objetivo definir los requerimientos técnicos para la interconexión de las centrales eléctricas al Sistema Eléctrico Nacional, manteniendo en todo momento la confiabilidad y seguridad de la red. Para ello clasifica las centrales eléctricas en diferentes tipos, como se puede observar en la tabla 1.

Tabla 1 Clasificación de las Centrales Eléctricas [6].

Áreas Síncronas	Central Tipo A	Central Tipo B	Central Tipo C	Central Tipo D
Sistema Interconectado Nacional	$P < 500$ kW	$0.5 \leq P < 10$ MW	$10 \leq P < 30$ MW	$P \geq 30$ MW

La central eléctrica debe mantenerse operando dentro de los rangos de frecuencia y tiempo definidos en la tabla 2.

Las centrales eléctricas deberán mantenerse interconectados a la red y operando ante cambios de la frecuencia. La central eléctrica debe controlar la potencia activa en respuesta a los aumentos de frecuencia en un tiempo menor a 2 segundos. Debe activarse a partir de 60 Hz, con una característica de regulación entre 3% y 8%. Estos son los factores principales por atender en el suministro de energía de un parque eólico.

Tabla 2 Tiempos máximos en los que la Central Eléctrica puede operar en frecuencias diferentes del valor nominal, sin desconectarse de la red [6].

Área Síncrona	Rango de Frecuencias	Tiempos máximos de Operación
Sistema Interconectado Nacional	$61.8 < f < 62.4$ Hz	15 min
	$61.2 < f < 61.8$ Hz	30 min
	$58.8 < f < 61.2$ Hz	Ilimitado
	$58.2 < f < 58.8$ Hz	30 min
	$57.0 < f < 58.2$ Hz	15 min

3. Resultados

El Código de Red nacional muestra los parámetros de potencia activa y reactiva, límites de operación de voltaje y frecuencia. Los cuales deben cumplir los parques eólicos para su interconexión a una red eléctrica en México. Como se observó la distribución de los aerogeneradores toma un papel muy importante para mitigar los problemas de vibraciones y perturbaciones que presenta el efecto sombra, el cual provoca la disminución de la potencia de giro del aerogenerador colocado atrás (de acuerdo con la dirección del viento), es decir, pérdidas de generación. Aún teniendo métodos de distribución de los aerogeneradores a lo largo del área disponible del parque, como se mostró en el artículo, para cumplir las demandas de operación establecidas por el Código de Red se presentan elementos complementarios para satisfacer con los requerimientos para resistir una falla sin sufrir daños. Estos elementos pitch control y stall regulated son las alternativas necesarias para mantener al parque eólico dentro de una interconexión segura, debido a que colaboran para mantener la generación de energía eléctrica dentro del margen solicitado para su distribución, al igual que mantienen la curva de potencia entregada por los aerogeneradores en límites de operación aceptables desde el punto de vista de estabilidad transitoria y aprovechamiento de la energía eólica.

4. Discusión

A lo largo del artículo se presentaron diferentes factores que afectan de manera considerable el funcionamiento del parque eólico en materia de generación y calidad de la energía entregada y que deben ser controlados. Se expusieron

algunos puntos fundamentales para la evaluación de la calidad de la energía, estos fueron presentados con base al código de red [CRE, 2016] (México) y en él se exponen los valores aceptables por la red, si la central eléctrica generadora, en este caso eólica, está interconectada a la red eléctrica nacional [Chen, 2005], [CRE, 2016], [Jadhav, 2011], [IEC-61400-12, 2017], [Tsili, 2009]. Si bien es vital cumplir con los valores solicitados por el código de red [CRE, 2016], es fundamental atender las fallas que presente el parque eólico ya que de ello dependerá el correcto y eficiente suministro de energía al sistema eléctrico de potencia. Como pudimos observar en los párrafos anteriores, el parque eólico depende totalmente del viento, la fuente de energía renovable, pero en ella encontramos que debemos contar con estrategias de control y tecnologías para mitigar las perturbaciones posibles, y cumplir con los valores estipulados por la red. La intermitencia del viento resulta ser la variable [Patel, 2005], [Rajib, 2002] de mayor interés, debido a que puede provocar problemas de flujo de potencia activa y reactiva [IEC-61400-21, 2001], [IEC-61400-12, 2017], [Muljadi, 1998] [Ackerman, 2005], variaciones de tensión y frecuencia [Heier, 2014], [Retana, 2016], [Thiringer, 2004]. Los métodos de control de velocidad por “pitch control” [Eisa, 2017], [Ferdous, 2016], [Xing, 2016], [Zhang, 2011], o “stall regulated” [Muljadi, 1998], [Petru, 2002], [Rosmin, 2012], [Thiringer, 2004] son muy utilizados y opciones viables puestos en práctica en condiciones reales. También se observa un avance tecnológico importante de los generadores eléctricos de velocidad variable [Muller, 2002], [Patel, 2005], [Rajib, 2002], [Rolán, 2010] que pueden ser controlados para mitigar las variaciones de potencia y frecuencia que presenta el aerogenerador.

5. Conclusiones

En este trabajo se presentó un análisis general de las características que presenta un parque eólico cuando se interconecta con la red eléctrica. El código de red nacional nos permite mantener el suministro de energía eléctrica dentro de los parámetros convenidos con el Sistema Eléctrico Nacional (SEN) ante cualquier probabilidad de contingencia como lo son, la estabilidad de factor de potencia,

voltaje, frecuencia, potencia activa y reactiva. En el código de red se expresan cuáles son los límites de operación de la red eléctrica nacional cuando se interconecta un parque eólico, es decir que mantenga un nivel adecuado de confiabilidad en la generación de energía eléctrica. Para lograr esto se realizó un estudio de los factores que intervienen en el desempeño del parque eólico como lo es el efecto sombra, “wake effect”, la intermitencia de la velocidad del viento y la transmisión de la energía eléctrica. Por ello son necesarios elementos que ayuden en el control de estos factores. Por otra parte, resulta importante la selección del tipo de generador eléctrico, la distribución de los aerogeneradores en el parque eólico y el número de aerogeneradores con que este cuente. Lo importante es ayudar a mantener la estabilidad en el suministro de energía eléctrica y un nivel adecuado de confiabilidad en el Sistema Eléctrico Nacional.

6. Bibliografía y Referencias

- [1] Abad, G., López, J., Rodríguez, M., Marroyo, L., & Iwanski, G., Doubly fed induction machine: modeling and control for wind energy generation (Vol. 85). John Wiley & Sons, 2011.
- [2] Ackermann, T. (Ed.), Wind power in power systems. John Wiley & Sons, 2005.
- [3] Akdag, S. A., Guler, O., & Yagci, E., Wind speed extrapolation methods and their effect on energy generation estimation. In Renewable Energy Research and Applications (ICRERA), 2013 International Conference on IEEE, pp. 428-430, 2013.
- [4] Chen, Z., Issues of connecting wind farms into power systems. In Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES, pp. 1-6, 2005.
- [5] Diario Oficial de la Federación, Código de Red. CDMX. PDF, pp. 179, 2016.
- [6] Eisa, S. A., Stone, W., & Wedeward, K. (2017, March). Mathematical Modeling, Stability, Bifurcation Analysis, and Simulations of a Type-3 DFIG Wind Turbine's Dynamics with Pitch Control. In Green Technologies Conference (GreenTech), 2017 Ninth Annual IEEE, pp. 334-341, 2017.

- [7] Chompoo-Inwai, C., Lee, W. J., Fuangfoo, P., Williams, M., & Liao, J. R., System impact study for the interconnection of wind generation and utility system. *IEEE transactions on Industry Applications*, 41(1), pp, 163-168, 2005.
- [8] Comisión Reguladora de Energía, Código de Red. CDMX: PDF, pp.144-164, 2016.
- [9] DEFU Committee reports 111-E (2nd edition): Connection of wind turbines to low and medium voltage networks, 1998.
- [10] El Moursi, M., Joos, G., & Abbey, C., A secondary voltage control strategy for transmission level interconnection of wind generation. *IEEE Transactions on Power Electronics*, 23(3), pp. 1178-1190, 2008.
- [11] English version of Technical Regulations TF 3.2.6, Wind turbines connected to grids with voltage below 100 kV –Technical regulations for the properties and the control of wind turbines, Eltra and Ekraft systems, 2004.
- [12] Erlich I., Shewarega F., Interaction of large wind power generation plants with the power system, *Proc. IEEE Int. Power and Energy Conf.*, Kuala Lumpur, 2006.
- [13] Ferdous, A. M. I., Sheikh, M. R. I., & Shobug, M. A. (2016, December). Controlling of frequency fluctuation of wind turbine generator using wind speed controlled pitch controller. In *Electrical, Computer & Telecommunication Engineering (ICECTE)*, International Conference on IEEE, pp. 1-4, 2016.
- [14] Global Wind Energy Council, Global Wind Statistics, 2015: http://www.gwec.net/wp-content/uploads/vip/GWEC-PRstats-2015_LR.pdf.
- [15] H. T. Jadhav and Ranjit Roy, A Critical Review on the Grid Integration Issues of DFIG based Wind Farms, National Institute of Technology, Surat, India, 2011.
- [16] Instituto Nacional de Ecología y Cambio Climático: <http://www.gob.mx/inecc/acciones-y-programas/gases-y-compuestos-de-efecto-invernadero>.
- [17] Intergovernmental Panel on Climate Change (ipcc), https://www.ipcc.ch/publications_and_data/ar4/wg1/es/faq-10-3.html.

- [18] Heier, S., *Wind Energy Conversion Systems*, in *Grid Integration of Wind Energy: Onshore and Offshore Conversion Systems*, John Wiley & Sons, Ltd, Chichester, UK, 2014.
- [19] IEC 61400-12: Wind turbine generator systems. Power performance measurement techniques.
- [20] IEC 61400-21: Power quality requirements for wind whines, 2001.
- [21] Muljadi, E., Pierce, K., & Migliore, P., Control strategy for variable-speed, stall-regulated wind turbines. In *American Control Conference*, 1998. *Proceedings of the 1998*, Vol. 3, pp. 1710-1714, IEEE, 1998.
- [22] Muller, S., Deicke, M., & De Doncker, R. W., Doubly fed induction generator systems for wind turbines. *IEEE Industry applications magazine*, 8(3), PP. 26-33, 2003.
- [23] Patel, M. R., *Wind and solar power systems: design, analysis, and operation*. CRC press, 2005.
- [24] Pearmine R., Song Y.H., Chebbo A., Influence of wind turbine behaviour on the primary frequency control of the British transmission grid, *IET Renew. Power Energy*, 1, (2), pp. 142– 150, 2007.
- [25] Petru, T., & Thiringer, T. (2002). Modeling of wind turbines for power system studies. *IEEE transactions on Power Systems*, 17(4), 1132-1139.
- [26] Rajib Datta and V. T. Ranganathan, Senior Member, IEEE, Variable Speed Wind Power Generation Using Doubly Fed Wound Rotor Induction Machine—A Comparison With Alternative Schemes, *IEEE Transactions On Energy Conversion*, Vol. 17, NO. 3, 2002.
- [27] Rebecca L. Busby, *Wind Power: The Industry Grows Up PemWell*, 2012.
- [28] Retana Mora Francisco Alejandro, Rodríguez García Bruno, *Análisis de estabilidad dinámica ante pequeños disturbios aplicado en aerogeneradores (Tesis de Pregrado)*. Instituto Politécnico Nacional, México. 2016.
- [29] Rolán, A., Luna, Á., Rocabert, J., Aguilar, D., & Vázquez, G. (2010, July). An approach to the performance-oriented model of variable-speed wind turbines. In *Industrial Electronics (ISIE)*, 2010 IEEE International Symposium on IEEE, pp. 3853-3858, 2010.

- [30] Rosmin, N., Samsuri, S., Hassan, M. Y., & Rahman, H. A., Power optimization for a small-sized stall-regulated variable-speed wind turbine. In Power Engineering and Optimization Conference (PEDCO) Melaka, Malaysia, 2012 Ieee International, pp. 373-378, 2012.
- [31] S. Cialdea, M. Peart, and W. Walton, Analysis and mitigation of harmonics in wind turbine transformers Worcester Polytechnic Institute, B.Sc Thesis, 2012.
- [32] Sun, Y. Z., Lin, J., Li, G. J., & Li, X., A review on the integration of wind farms with variable speed wind turbine systems into power systems. In Sustainable Power Generation and Supply, 2009. SUPERGEN'09. International Conference on IEEE, pp. 1-6, 2009.
- [33] Thiringer, T., Petru, T., & Lundberg, S., Flicker contribution from wind turbine installations. IEEE transactions on Energy Conversion, 19(1), pp. 157-163, 2004.
- [34] Tsili, M., & Papathanassiou, S., A review of grid code technical requirements for wind farms. IET Renewable Power Generation, 3(3), pp. 308-332, 2009.
- [35] Xing, H., & Yao, C. (2016, November). Coordinated pitch and generator control for wind turbine flexible power tracking. In Electrical Machines and Systems (ICEMS), 2016 19th International Conference on IEEE, pp. 1-5, 2016.
- [36] Zhang, L., Sun, Y., & Lv, X., Design on pitch-control wind turbine system based on Bladed. In Control and Decision Conference (CCDC), 2011 Chinese, pp. 3921-3923, IEEE, 2011.

DETECCIÓN ACTIVA DE FALTAS EN SISTEMAS DE EVENTOS DISCRETOS

Karen Hernández Rueda

Universidad de Guadalajara-CUCSUR

karenhr@cucsur.udg.mx

María E. Meda Campaña

Universidad de Guadalajara-CUCEA

emed@cucea.udg.mx

Bernardo Haro Martínez

Universidad Autónoma de Guadalajara

bernardo.haro@gmail.com

Resumen

El objetivo de este trabajo es presentar una propuesta de solución de Diagnóstico Activo en Sistemas de Eventos Discretos modelado por redes de Petri. La propuesta se basa en un controlador llamado Circuito de Regulación Inteligente que reduce la distancia relativa entre las transiciones que modelan faltas y las del resto de la red de Petri, permitiendo la detección y diagnóstico de faltas mientras se mantiene la vivacidad del sistema y se reduce la flexibilidad del sistema sólo en los estados requeridos. Finalmente, los resultados presentados se ilustran en un ejemplo.

Palabras Claves: Detección activa, diagnosticabilidad, redes de Petri, sistemas de eventos discretos.

Abstract

The aim of this work is to present a proposal of Active Diagnosis in Discrete Event Systems modeled by Petri nets. This approach is based on a controller named Intelligent Regulation Circuit which reduces the relative distance among the system transition allow in the detection and diagnosis of faults while maintaining

the liveness of the system. Finally, the results presented are illustrated by an example.

Keywords: *Active detection, diagnosability, discrete event system, Petri nets.*

1. Introducción

En la actualidad los sistemas industriales cada vez más se vuelven grandes y complejos por lo que no están exentos de sufrir cualquier tipo de desviación de su comportamiento especificado (normal), comprometiendo la seguridad tanto de los sistemas como de los operadores humanos. Por lo tanto, las tareas de detección y localización de faltas deben incluirse en los controladores modernos para incrementar la confiabilidad de los sistemas.

La detección y localización de faltas han sido estudiadas extensamente en la literatura desde el punto de vista de los autómatas finitos (FA) y de las redes de Petri (RP). Existen varios enfoques que usan FA; en [Sampath et al., 1995] y [Sampath et al., 1998] se caracteriza la propiedad de diagnosticabilidad y se resuelven problemas de detección y localización de faltas en línea. Después de estos trabajos seminales, estos conceptos han sido extendidos y aplicados a diferentes áreas y herramientas formales. Por ejemplo, en [Lafortune, 2007] se aborda la diagnosticabilidad en sistemas distribuidos. En [Seatzu, 2005] y [Wu, 2005] se trata el problema de diagnóstico usando redes de Petri. Posteriormente, en [Dotoli et al., 2009] se usa un problema de programación entera para determinar cuál secuencia de transiciones fue disparada y así determinar la ocurrencia de una falta. En [Lefebvre, 2011] se usa una función probabilista para dar una medida a la ocurrencia de una falta; en [Ramírez et al., 2012] la propiedad de diagnosticabilidad se caracteriza usando RP Interpretadas (RPI); en [Ruiz et al., 2014] se presentan algoritmos para construir diagnosticadores y probar la diagnosticabilidad del sistema.

Un problema relacionado con la diagnosticabilidad es forzar la diagnosticabilidad en Sistemas de Eventos Discretos (SED). Esto significa, encontrar formas de hacer que un SED sea diagnosticable agregándole elementos al sistema, tales como sensores y/o controladores. Este problema ha sido abordado desde

diferentes puntos de vista. En [Ziqiand et al., 2014] la diagnosticabilidad es forzada seleccionando las palabras apropiadas para conseguir la detección de faltas y su aislamiento (diagnosticabilidad activa) para un caso específico y no se puede generalizar. En [Cabasino et al., 2013] se resuelve un problema de localización de sensor para garantizar la diagnosticabilidad. Por otro lado, en [Hernández et al., 2015] se presenta un enfoque para forzar la diagnosticabilidad en una clase de RP utilizando un controlador nombrado como un Circuito de Regulación [Densel, 1995], la solución es estructural y consiste en añadir nuevos lugares para restringir el disparo de las transiciones en la RPI. Sin embargo, la inclusión de estos lugares reduce el número y variedad de palabras que el sistema puede realizar y, si no se realiza adecuadamente, la inclusión de estos lugares puede bloquear a la red.

Este trabajo presenta una propuesta de diagnosticabilidad activa a través de un Circuito de Regulación Inteligente (CRI) en una clase de redes de RP que no es diagnosticable pero sí acotada y viva. La solución es estructural, considera un marcado de k marcas en el CRI para asegurar que la ocurrencia de cualquier falta pueda ser detectada y aislada.

En la siguiente sección se presentan la propuesta y los sustentos teóricos necesarios para su comprensión, así como la caracterización de una clase de RP donde se puede realizar un diagnóstico activo.

2. Métodos

Esta sección introduce los conceptos básicos de RP y diagnóstico que serán necesarios para la explicación del material presentado en el trabajo. Un lector interesado puede consultar las referencias [Densel, 1995] y [Murata, 1989] para más información.

Redes de Petri

- Una estructura de una red de Petri es un dígrafo bipartito definido por la 4-tupla $Q=(P,T,I,O)$, donde $P = \{p_1,p_2,\dots,p_n\}$, $T = \{t_1,t_2,\dots,t_m\}$ son conjuntos finitos de lugares y transiciones respectivamente. $P \cap T = \emptyset$ y $P \cup T \neq \emptyset$. $I: P \times T \rightarrow \{0,1\}$ y $O: P \times T \rightarrow \{0,1\}$ son las funciones de entrada y salida que

describen los arcos que van de los lugares a las transiciones y de las transiciones a los lugares respectivamente.

- Un marcado es una función $M: P \rightarrow \{0, 1, 2, 3, \dots\}$ que asigna a cada lugar un número entero no negativo, nombrado como el número de marcas que residen dentro de cada lugar. M_0 es la distribución inicial de marcado.
- Una red de Petri N es una estructura Q junto con un marcado inicial, esto se denota como $N=(Q, M_0)$.
- La matriz de incidencia C de $n \times m$ de N está definida por $C_{\{i,j\}} = O(t_j, p_i) - I(p_i, t_j)$. La notación $\bullet t = \{p \mid I(p, t) \neq 0\}$, $t \bullet = \{p \mid O(p, t) \neq 0\}$, $\bullet p = \{t \mid O(p, t) \neq 0\}$ y $p \bullet = \{t \mid I(p, t) \neq 0\}$ representa los lugares de entrada y de salida de t , y las transiciones de entrada y de salida de p respectivamente.
- Sea (Q, M_0) una RP. Los vectores X_i tal que $CX_i = 0$, $X_i \geq 0$ son conocidos como T-semiflujos. El soporte de un T-semiflujo X_i , denotado por $\langle X_i \rangle$, es el conjunto de transiciones $T_i = \{t_j \mid X_i(j) > 0\}$. La subred $T_i = \{(P_i, T_i, I, O), M_{0i}\}$ de N generada por el T-semiflujo X_i es una T-componente si $P_i = (\bullet \langle X_i \rangle \cap \langle X_i \bullet)$, $T_i = \langle X_i \rangle$; I_i , O_i y M_{0i} son las funciones de entrada y salida, y el marcado inicial restringido a P_i y T_i respectivamente.

Una transición t_j se dice que está habilitada en el marcado M_k si este tiene $M_k(p_i) \geq I(p_i, t_j)$ marcas en cada lugar p_i de entrada a t_j . Una transición habilitada t_j se puede disparar, remueve $I(p_i, t_j)$ marcas de p_i y añade $O(t_j, p_k)$ marcas a p_k produciendo un nuevo marcado M_{k+1} (representado por $M_k \xrightarrow{t_j} M_{k+1}$) que puede ser calculado usando la ecuación de estado $M_{k+1} = M_k + C \vec{t}_j$ donde C es la matriz de incidencia y $\vec{t}_j(i) = 1$ si $i=j$ y $\vec{t}_j(i) = 0$ en cualquier otro caso.

Observe que $M_0 \xrightarrow{t_j} M_1$ puede ser extendido a una secuencia de transiciones $M_0 \xrightarrow{\sigma} M_q$ donde $\sigma = t_a t_b \dots t_r$. En este caso M_q se dice que es alcanzable desde M_0 . El conjunto de alcanzabilidad de N , denotado por $R(Q, M_0)$, es el conjunto de todos los posibles marcados alcanzables desde M_0 , disparando solamente las transiciones habilitadas:

Definición 1. Una RP (Q, M_0) es viva (o equivalentemente M_0 es un marcado de N vivo) si, no importa cual marcado ha sido alcanzado desde M_0 , es posible disparar de última instancia cualquier transición de N al progresar a través de alguna secuencia de disparo adicional.

Definición 2. Una RP (Q, M_0) es k-segura si $\forall M \in R(Q, M_0)$ y $\forall p \in P, M(p) \leq k$. Si se cumple que $\forall M \in R(Q, M_0)$ y $\forall p \in P, M(p) \leq 1$, la red es llamada 1-segura (segura o binaria).

Definición 3. Una RP (Q, M_0) es fuertemente-conexa para cualesquiera dos nodos de la red X, Y (lugares o transiciones) hay un camino de X a Y y de Y a X.

Definición 4. Un sifón (o cerrojo) es un subconjunto de lugares $S = \{p_1, \dots, p_s\} \subseteq P$ de una RP tal que $\bullet S \subset S \bullet$. Las siguientes definiciones están relacionadas con la secuencia de transiciones de disparo con los vectores de observación de salida.

Definición 5. Una secuencia de transiciones de disparo en una RP (Q, M_0) es una secuencia $\sigma = t_i t_j \dots t_k \dots$ tal que $M_0 \xrightarrow{t_i} M_1 \xrightarrow{t_j} M_2 \dots M_w \xrightarrow{t_k} \dots$

En este trabajo se asume que la RP es evento-detectable, es decir, que el disparo de cualquier transición siempre es detectado. En [Ramírez et al., 2012] y [Rivera et al., 2005] se presenta esta propiedad. Gráficamente una RP se puede ver como en la figura 1a.

Diagnosticabilidad

En este trabajo sólo se consideran las faltas permanentes f_i . En la figura 1b se representan dos faltas permanentes f_1 y f_2 en una RP, estas son subredes. En la figura 1c se muestra los subconjuntos de lugares P y transiciones T considerados en la RP en estado normal y de falta.

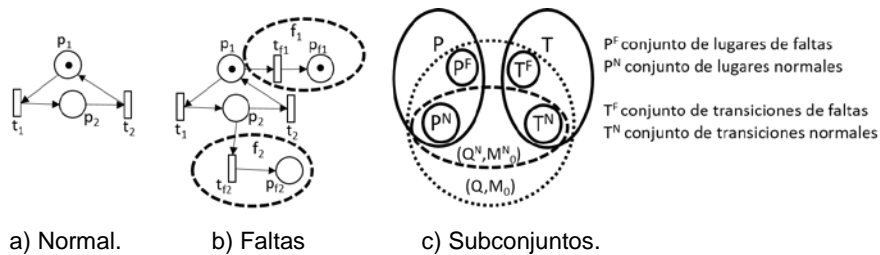


Figura 1 Conjuntos de P y T.

En este trabajo se supone que el disparo de las transiciones de falta no es evento-detectable. En el caso de que lo sea el problema de detección de faltas estaría resuelto.

La siguiente definición es tomada de [Ramírez et al., 2012].

Definición 6. Sea (Q, M_0) una RP y $t_{fi} \in T^F$. El conjunto de lugares de pre-riesgo de t_{fi} es $P_i^R = \{p_k | p_k \bullet t_{fi}\}$. El conjunto de lugares de post-riesgo de t_{fi} es $P_i^{PR} = \{p_k | p_k \in (\bullet t_{fi})^{**} \cap P^N\}$. El conjunto de transiciones de pre-riesgo de t_{fi} es $T_i^R = \{t_k | t_k \in \bullet P_i^R \cap T^N\}$ y el conjunto de transiciones de post-riesgo de t_{fi} es $T_i^{PR} = \{t_k | t_k \in \bullet P_i^{PR} \cap T^N\}$.

La propiedad de diagnosticabilidad entrada-salida de un SED basada en los modelos RP se define a continuación.

Definición 7. Una RP viva dada por (Q, M_0) es diagnosticable en $k < \infty$ pasos si usando cualquier secuencia de disparo de transiciones de longitud igual o mayor a k y la estructura de (Q, M_0) son suficientes para distinguir la ocurrencia de una falta en el SED. Esta definición es equivalente a la presentada en [Sampath et al., 1996] desde el punto de vista de las RP. Como se muestra en la figura 2, si un ciclo a) que contiene una falta f_i cuya salida RP es igual a otro ciclo b) que no contiene la falta f_i entonces la RP es no diagnosticable entrada-salida.

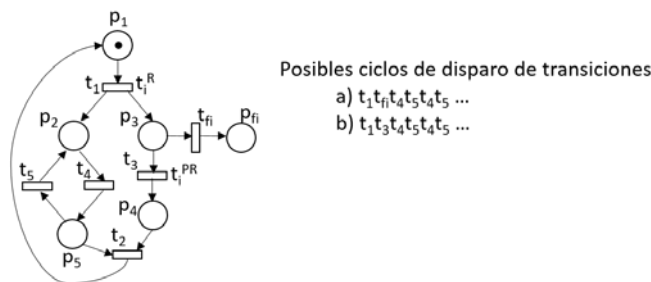


Figura 2 RP con ciclos de disparo indistinguibles.

Si $t_3 \in T^{PR}$ es evento-detectable (es decir, el disparo de esta transición se detecta), entonces los ciclos pueden distinguirse si t_3 pertenece a cualquier secuencia de disparo de transiciones finita. El intentar disparar t_3 y no poder hacerlo indica que su lugar de entrada no tiene marcas, y esto ocurre porque la marca está retenida en el lugar de falta, es decir ocurrió la falta. Por lo tanto, si la transición de post-riesgo está en cualquier secuencia finita en el comportamiento normal de la red,

entonces en un número finito de pasos se intentará disparar dicha transición, y su disparo o no disparo permite determinar si la falta existió. Este hecho se estudia a través de la distancia relativa [Ruiz et al., 2007] entre dos transiciones y los sifones [Densel, 1995] de la red.

Definición 8. Sea (Q, M_0) una RP limitada, la distancia relativa $D_R(t_i, t_j)$ entre cualquier par de transiciones $t_i, t_j \in T$, es el número máximo de veces que t_j puede ser disparado sin que se dispare t_i cuando una marca se retiene en el lugar $\bullet t_i$, esto es, el marcado que habilita a t_i no puede usarse para disparar la transición t_j . La distancia máxima relativa $D_H(t_i, t_j)$, entre cualquier par de transiciones $t_i, t_j \in T$ es $D_H(t_i, t_j) = \max\{D_R(t_i, t_j), D_R(t_j, t_i)\}$.

El problema de caracterizar la diagnosticabilidad de las faltas permanentes necesita el cálculo de las distancias máximas relativas. Este cálculo parece ser un problema complejo. Sin embargo, existen condiciones estructurales de la RP que pueden ser explotadas para determinar polinómicamente la distancia máxima relativa entre las transiciones en una clase de RP.

La siguiente proposición presentada en [Ruiz et al., 2014] caracteriza la diagnosticabilidad en términos de la distancia relativa máxima (si los sifones se desmarcan todas las transiciones no son vivas).

Proposición 1. Sea (Q, M_0) una RP limitada, donde (Q^N, M_0^N) es viva, acotada y fuertemente-conexa. Sea t_{fi} una falta permanente, p_k un lugar de riesgo y S_{ti} el sifón que se desmarcará cuando t_{fi} se dispare. Se asume que $|p_k \bullet| = 1$ y la transición post-riesgo $t_a \in p_k \bullet$ y las transiciones pre-riesgo son evento-detectable. (Q, M_0) es diagnosticable respecto a t_{fi} si todos los T - semiflujos de la red contienen transiciones en $\bullet S_{ti} \cap S_{ti} \bullet$.

En la proposición anterior, la notación $\bullet S_{ti} \cap S_{ti} \bullet$ indica las transiciones de entrada y de salida a los lugares que forman el sifón S_{ti} .

Diagnosticabilidad Activa

Las RP que tienen ciclos indeterminados no son diagnosticables. Sin embargo, como se indica en [Hernández et al., 2015], es posible remover estos ciclos modificando la estructura de la RP. La modificación de la RP se realiza a través de

la adición de un Circuito de Regulación (CR) [Densel y Esparza, 1995]. Esto es, si \exists un conjunto $T_r = \{t_i, t_j, \dots, t_q\} \subseteq T$ tales que $\bullet t_i = \bullet t_j = \dots = \bullet t_q$ entonces se añade un conjunto $C_r = \{p_i', p_j', \dots, p_q'\}$ conocido como un CR para T_r y arcos tales que $\bullet p_i' = t_i, p_i' \bullet = t_j, \bullet p_j' = t_j, \dots, \bullet p_q' = t_q$ con una marca inicial en uno de los lugares de C_r . Considere, por ejemplo, un estacionamiento automatizado que tiene tres entradas (Entrada1, Entrada2, Entrada3), una salida y cuatro cajones (Lugar1, Lugar2, Lugar3, Lugar4) para estacionarse. La figura 3 muestra de lado izquierdo un esquema del estacionamiento y de lado derecho su modelo en RP.

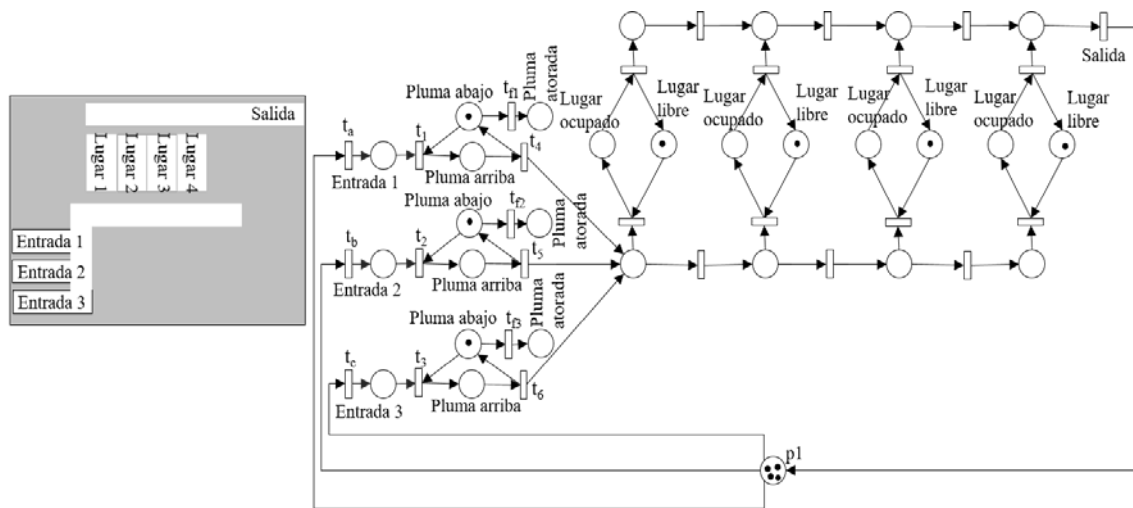


Figura 3 Estacionamiento de capacidad cuatro y su modelo en RP.

Imagine como faltas que las plumas fallen al levantarse ante la llegada de un coche. Hay muchas otras faltas, sin embargo, con la explicación de éstas bastará para ejemplificar al CRI y las demás faltas se pueden tratar exactamente igual. Las partes que destacar en el modelo son las secciones de entrada, las secciones de cada pluma de entrada y las posibles faltas donde las plumas se puede quedar atoradas. En este caso $t_4, t_5, t_6 \in T^R$ son transiciones de pre-riesgo a las fallas de la pluma de la Entrada1 atorada, pluma de la Entrada2 atorada y pluma de la Entrada3 atorada respectivamente, y $t_1, t_2, t_3 \in T^{PR}$ son transiciones de post-riesgo de las faltas de la Entrada 1, 2 y 3 respectivamente. Las transiciones $t_{r1}, t_{r2}, t_{r3} \in T^F$ son las faltas Pluma1 atorada, Pluma2 atorada y Pluma3 atorada, respectivamente. Haciendo el análisis de diagnosticabilidad se obtiene que $D_H(t_b,$

$t_1)=\infty$, $D_H(t_c, t_1)=\infty$, $D_H(t_a, t_2)=\infty$, $D_H(t_c, t_2)=\infty$, $D_H(t_a, t_3)=\infty$, $D_H(t_b, t_3)=\infty$, por lo que ninguna de las faltas es diagnosticable.

Según [Hernández et al., 2015] se debe poner un C_r en $\{t_a, t_b, t_c\}$ como se muestra en la figura 4 para que los T-semiflujos que pasan por $t_i, t_j, \dots t_q$ se sumen creando un nuevo T-semiflujo que contenga transiciones en $\bullet S_{ti} \cap S_{tj} \bullet$ y el sistema se vuelva diagnosticable, es decir forzar la diagnosticabilidad. La parte resaltada en negro es el CR, si se dispara primero t_b luego se dispara t_a y por último t_c . Esto provoca que sólo una entrada esté habilitada a la vez, teniendo los coches que buscar dicha entrada y si están distantes entre sí, es inconveniente.

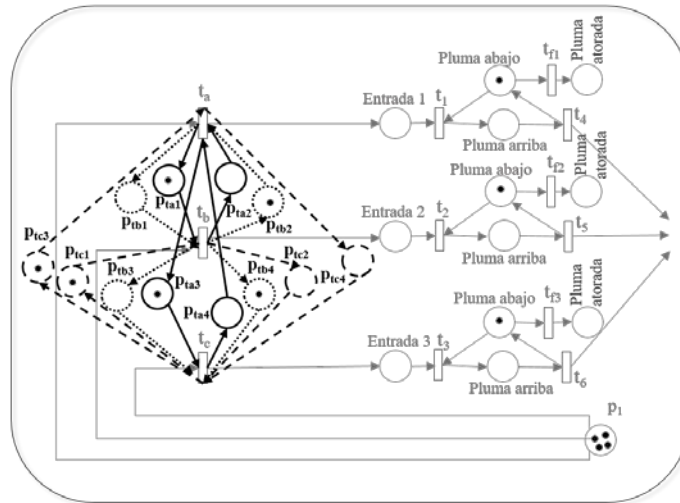


Figura 3 Modelo del estacionamiento con el C_r .

Circuito de Regulación Inteligente

Como se indicó anteriormente, forzar la diagnosticabilidad significa la eliminación de ciertas secuencias infinitas que incluyen las transiciones de pre-riesgo, pero que no incluyen las transiciones de post-riesgo. Durante este proceso pueden ocurrir dos cosas, la primera es que se eliminen secuencias de más y que pueden ser importantes para el sistema, y la segunda, que se generan secuencias que bloquean al sistema.

El CR presentado en trabajos anteriores introduce una solución que garantiza la vivacidad, pero establece una $D_H(t_i, t_j)=1$, eliminando la flexibilidad en el sistema. Además, introduce un orden estricto en el disparo de $t_i, t_j, \dots t_q$, lo cual es limitante

en aplicaciones prácticas como el caso del estacionamiento. Ahora se propone un Circuito de Regulación Inteligente (CRI), que no impone ningún orden en las transiciones del sistema. El circuito sólo actúa cuando se requiere detectar si hay alguna falta. Esto ocurre cuando el circuito detecta el disparo de alguna transición de pre-riesgo y no se ha detectado el disparo de su transición de post-riesgo después de la ocurrencia de un número preestablecido de eventos.

Antes de la definición formal de CRI, se necesita la definición del conjunto de transiciones de regulación de falta, que es el conjunto de transiciones cuyo disparo se puede manipular para asegurar la detección de una falta.

Definición 9. Sea (Q, M_0) una RP y t_{fi} una transición de falta del sistema. Sea $T_{t_{fi}} = \{t_i, t_j, \dots, t_q\} \subseteq T$ un conjunto de transiciones tales que $\bullet t_i = \bullet t_j = \dots = \bullet t_q$. El conjunto es un conjunto de transiciones de regulación de la falta t_{fi} si existe al menos una $t_a \in T_{t_{fi}}$ tal que $D_H(t_a, t_x) = \infty$, donde t_x es la transición de post-riesgo de t_{fi} .

El disparo de las transiciones de este conjunto es el que se puede controlar para reducir la distancia relativa entre transiciones y detectar la falta. La siguiente proposición muestra que si existe una falta no diagnosticable t_{fi} entonces también existe el conjunto $T_{t_{fi}}$.

Proposición 2. Sea (Q^N, M_0^N) una RP viva y acotada y fuertemente conexa. Sea $t_{fi} \in T^F$ y t_x su transición post-riesgo tal que existe t_j con $D_H(t_x, t_j) = \infty$ y $t_j \in T^N$. Entonces existe $T_{t_{fi}}$ que es el conjunto de transiciones de regulación de la falta t_{fi} .

Demostración. Tomar t_x la transición de post-riesgo de la falta t_{fi} para construir una trayectoria de nodos ascendente de la siguiente forma. Tomar los caminos desde t_x recorriendo la RP en sentido inverso a sus arcos hasta encontrar una transición t_a a la que se le puede encontrar un conjunto de transiciones $T = \{t_a, t_b, \dots, t_q\}$ tales que $\bullet t_a = \bullet t_b = \dots = \bullet t_q$. Tal transición existe, de lo contrario cada transición en el camino tiene exactamente un lugar de entrada y estos lugares sólo pueden habilitar las transiciones del camino. Como la red es fuertemente conexa, eventualmente se regresará a t_x formando un ciclo, aunque no necesariamente mínimo. Como la red es viva, se puede proponer un marcado inicial M_0 acotado que hace viva a la red. De este marcado inicial se puede hacer evolucionar a la red. Como los lugares sólo habilitan transiciones del camino y el camino es finito (los conjuntos de

transiciones y lugares en una red son finitos), eventualmente se deberá disparar t_x , es decir $D_H(t_x, t_j) < \infty$, una contradicción. Por lo tanto, existe el conjunto T y éste es el conjunto de transiciones de regulación de la falta t_{fi} . i.e. existe el conjunto $T_{t_{fi}}=T$.

La demostración de la proposición anterior nos sugiere un algoritmo para construir los conjuntos de transiciones de regulación para la falta t_{fi} . Note que si se construye un conjunto $T_{t_{fi}}$ y se agrega un circuito de regulación como en [Hernández et al., 2015] a este conjunto, podría resultar en que todavía existen transiciones con distancia relativa infinita hacia la falta t_{fi} , entonces, por la proposición anterior, debe existir otro conjunto $T_{t_{fi}2}$ con otras transiciones de regulación de falta. Este procedimiento se debe repetir tantas veces como sea necesario, hasta que la falta t_{fi} sea diagnosticable.

Ahora ya se puede definir el Circuito de Regulación Inteligente.

Definición 10. Sea (Q^N, M_0^N) una RP viva, acotada y fuertemente conexa. Sea t_{fi} una transición de falta del sistema. Sea $T_{tk} = \{t_a, t_b, \dots, t_x\} \subseteq T$ un conjunto de transiciones de regulación de la falta t_{fi} . Un CRI para el conjunto T_{tk} está formado para cada $t \in T_{tk}$ por un lugar de auto-lazo p_{ai} para una transición $t \in T_{tk}$, un lugar de salida p_{ci} llamado contador para una transición $t \in T_{tk}$; para cada $t_j \in T^R$ un p_j^R lugar de salida para la transición de pre-riesgo de t_j , para cada $t_z \in T^{PR}$ un p_z^{PR} lugar de post-riesgo de salida a t_z y un algoritmo de toma de decisiones (STD) que calcula el marcado de los lugares agregados. En el marcado inicial todos los lugares de auto-lazo tienen una marca y los lugares contadores y de post-riesgo están desmarcados.

Los lugares de pre-riesgo están inicialmente marcados sólo si los lugares de entrada a la falta están inicialmente marcados. En la figura 5 se muestra el esquema del CRI. Los lugares mostrados son los añadidos. Hay un circuito por cada conjunto de transiciones de regulación de la falta f_i .

Funcionamiento del Circuito de Regulación Inteligente para la Falta X

Sea $T_{tx} = \{t_1, t_2, \dots, t_{nc}\}$, p_{ci} =lugar del contador i -ésimo, p_{+i} = lugar pre-riesgo i -ésimo, p_{-i} lugar post-riesgo i -ésimo, p_{ai} = lugar de auto-lazo i -ésimo, k = el número máximo de veces que se pueden disparar algunas de las transiciones en el

conjunto de transiciones de regulación de la falta sin disparar alguna otra del mismo conjunto.

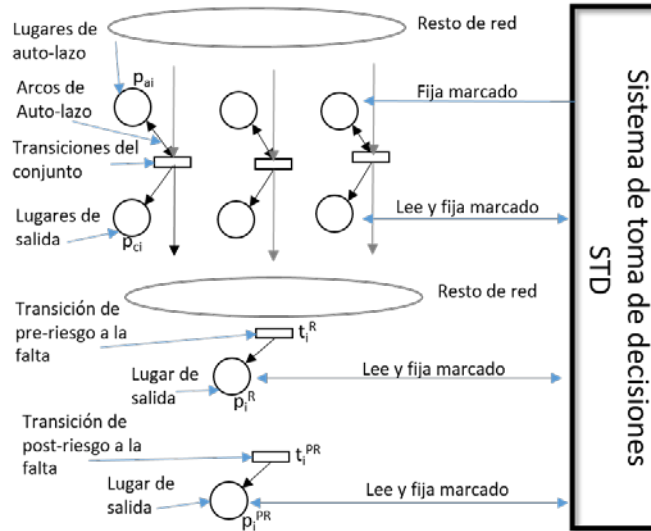


Figura 5 Circuito de regulación inteligente.

Etapa de Diagnostico activo

- Si $\forall p_{ci} M(p_{ci}) > 0$ entonces $\forall p_{ci} M(p_{ci}) = 0, \forall p_{ai} M(p_{ai}) = 1$.
- Si $\sum_{i=1}^{nc} M(p_{ci}) \geq k$ entonces $M(p_{ai}) = 0$ para $M(p_{ci}) > 0$.

Etapa de Diagnostico

- Si $M(p_{+i}) = 1$ entonces activar el disparo de t_i^{PR} (transición de post-riesgo de la falta i).
 - a. Si se intentó disparar t_i^{PR} y $M(p_{-i}) = 0$ entonces error. Dejar $M(p_{ai}) = 0$ falta permanente.
 - b. Si se intentó disparar t_i^{PR} y $M(p_{-i}) = 1$ entonces todas bien. $M(p_{-i}) = 0$.

Proposición 3. Sea (Q^N, M_0^N) una RP viva, acotada y fuertemente conexa. Sea $t_{fi} \in T^F$ no diagnosticable y $T_{t_{fi}} = \{t_a, t_b, \dots, t_q\}$ uno de sus conjuntos de transiciones de regulación. Si se le añade un CRI a $T_{t_{fi}}$, entonces la RP con CRI también es viva.

Demostración. Como la RP original es viva, entonces existen secuencias de disparo de transiciones desde el marcado inicial que marcan el lugar p , que es de

entrada a todas las transiciones de T_i (por definición todas las transiciones de T_i tienen el mismo lugar de entrada).

Aseveramos que como $\bullet t_a = \bullet t_b = \dots \bullet t_q$, entonces el disparo de dos transiciones es independiente, es decir, se puede disparar una de ellas, por ejemplo, t_a , un número infinito de veces, sin necesidad de disparar alguna otra en el conjunto, por ejemplo, t_b . Suponer lo contrario, es decir que después de dispararse t_a un número máximo $k_a < \infty$ de veces se necesita el disparo de alguna otra transición, por ejemplo, t_b . Si t_a se dispara k_a veces y se marca de nuevo p , entonces se puede disparar cualquier transición en T_{ti} . En especial se puede disparar t_a nuevamente. Dos casos ocurren, que k_a no era el número máximo de veces que se dispara t_a o que se bloquee la red después del disparo de t_a . Como la red es viva y k_a ya era el máximo, ambos casos son una contradicción. Entonces el número de veces que se puede disparar cada transición sin disparar otra en T_{ti} es infinito.

Si las secuencias $t_{a\sigma_1}, t_{b\sigma_2}, \dots, t_{q\sigma_q}$, son disparadas en la RP, entonces un subconjunto de ellas se dispara cuando se activa el CRI porque algunos lugares de auto-lazo tienen cero marcas. Como los disparos de t_a, t_b, \dots, t_q son independientes y vienen de secuencias vivas, entonces dichas secuencias permiten que el lugar p se siga marcando frecuentemente. Cada vez que se marca p una transición $t_q \in T_{ti}$ con $M(p_{aq})=1$ se dispara. Después del disparo $M(p_{aq})=0$, por lo que en la siguiente vez que se marque p se disparará una nueva transición y así hasta que todas las transiciones en T_{ti} se hayan disparado al menos una vez y en este momento el CRI para todas las $M(p_{ai})=1$. Cuando todos los lugares de auto-lazo están marcados, entonces se vuelve a tener todo el lenguaje de la RP y la red es viva.

Proposición 4. Sea (Q^N, M_0^N) una RP viva, acotada y fuertemente conexa. Sea $t_{fi} \in T^F$ no diagnosticable con $t_x \in T^{PR}$ y $T_{t_{fi}} = \{t_a, t_b, \dots, t_q\}$ uno de sus conjuntos de transiciones de regulación. Si se le añade un CRI a $T_{t_{fi}}$, entonces t_{fi} se vuelve diagnosticable.

Demostración. Se sabe que $D_H(t_a, t_x) = \infty$ para una alguna transición en $t_a \in T_{t_{fi}}$. Cuando el CRI detecta el disparo de la transición de pre-riesgo de t_{fi} , éste quita las marcas de los lugares auto-lazos de entrada a $t_q \in T_{t_{fi}}$, siempre y cuando $D_H(t_q, t_x) =$

∞ . Es decir, $t_q \in T_{t_{fi}}$ ya no se puede disparar mientras que no se intente disparar t_x . Como la red es viva por la proposición anterior, la red no se bloqueará. Por lo tanto, la red se vuelve diagnosticable.

Proposición 5. Sea (Q^N, M_0^N) una RP viva, acotada y fuertemente conexa. Sea $t_{fi} \in T^F$ no diagnosticable con $t_x \in T^{PR}$ y $T_{t_{fi}} = \{t_a, t_b, \dots, t_q\}$ uno de sus conjuntos de transiciones de regulación. Si se le añade un CRI a $T_{t_{fi}}$, entonces la ocurrencia de t_{fi} se detecta y diagnostica.

Demostración. El circuito de regulación inteligente detecta cuando se dispara la transición de pre-riesgo de t_{fi} . En este estado, el CRI reduce la distancia relativa de t_{fi} a todas las transiciones en uno por modificar los marcados en los lugares de auto-lazo. También detecta si se intenta disparar t_x . Si ésta se dispara entonces no hay falta, si ésta no puede dispararse, entonces no tiene marcas en sus lugares de entrada, esto sólo se debe a que ocurrió la falta t_{fi} . Por lo tanto, la falta se detecta y diagnostica.

3. Resultados

Ahora si se observa el modelo del estacionamiento de la figura 1 se puede notar que $T_{t_{f1}}=T_{t_{f2}}=T_{t_{f3}}= \{t_a, t_b, t_c\}$ por lo que solo se requiere un CRI. La figura 6 muestra el CRI para el estacionamiento. Los lugares remarcados en oscuro son los agregados por el circuito, el resto, lugares claros, ya pertenecían al modelo en RP. En este caso los lugares p_{a1} , p_{a2} y p_{a3} son los lugares de auto-lazo, note que están inicialmente marcados permitiendo que las transiciones del sistema se disparen conforme lo requiera el sistema. Los lugares p_{c1} , p_{c2} y p_{c3} son contadores de ejecución de las transiciones a las que se conectan. Las transiciones $t_a, t_b, t_c \in T_{t_k}$ son las transiciones que conforman los conjuntos de transiciones de regulación. Los lugares $\{p_{+1}, p_{+2}, p_{+3}\}$ y $\{p_{-1}, p_{-2}, p_{-3}\}$ son los lugares de pre-riesgo y post-riesgo a las faltas. Los lugares de pre-riesgo están marcados inicialmente porque las condiciones iniciales del sistema marcan los lugares pluma abajo que son de riesgo, aquí es donde puede ocurrir que la pluma se quede atascada provocando una falta.

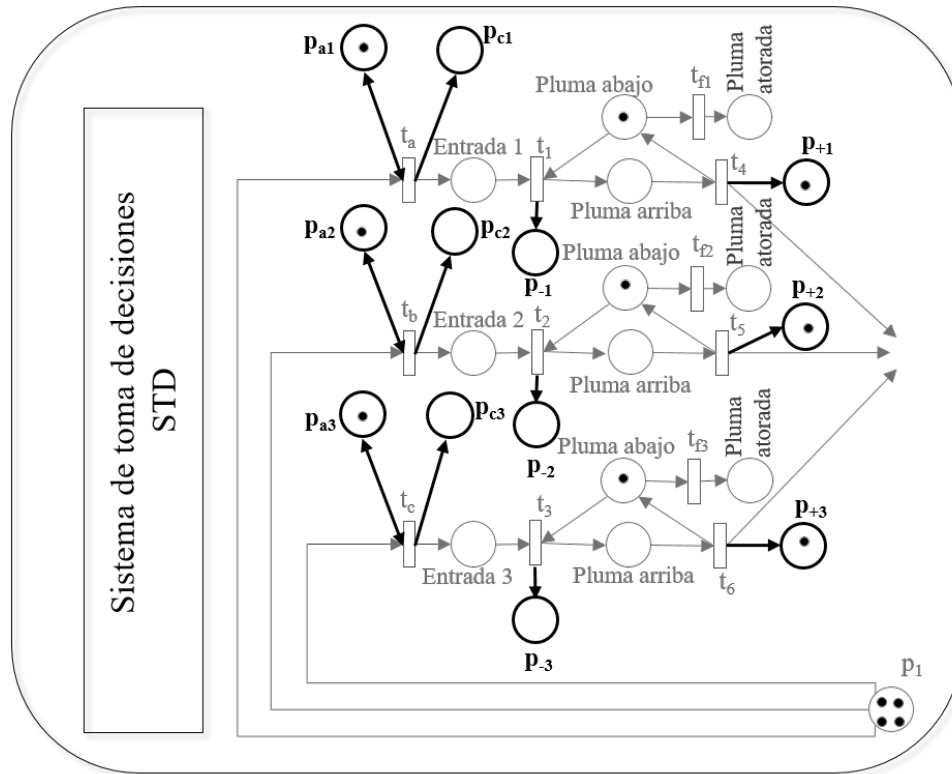


Figura 6 Modelo del estacionamiento don un CRI.

Se agrega un CRI, de acuerdo con la definición del CRI se tiene la siguiente forma de operar el conjunto con el CRI. Suponer que para cada falta se tiene la misma constante $k=3$, es decir a lo más tres coches pueden entrar al estacionamiento sin permitir el no uso de alguna de las entradas. Después de este número el STD debe trabajar para garantizar la diagnosticabilidad de las faltas. Suponer que entran dos carros por la entrada 1 y dos por la entrada 2, entonces el marcado de p_{c1} es igual a dos, lo mismo para p_{c2} . Entonces la suma de estos marcados es cuatro, indicando que cuatro coches entraron al estacionamiento sin usar la entrada 3, i.e. más de lo permitido. Entonces el STD desmarca los lugares p_{a1} y p_{a2} inhabilitando la entrada 1 y 2. Por lo que el siguiente coche que quiera entrar al estacionamiento deberá hacerlo por la entrada 3. Como el lugar p_{+3} tiene marca, el STD está consciente de la posible falta. Como las distancias relativas de todas las transiciones son finitas en la condición de los lugares p_{a1} y p_{a2} desmarcado, entonces eventualmente se intentará el disparo de t_c . Si se marca p_{-3} entonces el disparo de t_c se realizó y la pluma fue levantada indicando que no hay falta. En

este caso el STD pone los marcados a condiciones iniciales en los lugares de auto-lazo y contadores. También remueve la marca del lugar p_{-3} . Permitiendo nuevamente la máxima flexibilidad en el sistema y detectando que no hubo falta alguna. Si por el contrario el lugar p_{-3} no se marca entonces el STD marca que la falta t_{f3} está presente, por lo tanto, hay un error en la pluma de la tercera entrada. En este caso el STD pone el marcado de los lugares añadidos a condiciones iniciales, excepto p_{a3} , que lo pone en cero con el objetivo de no permitir acceso por esta entrada, ya que está dañada la pluma.

4. Discusión

Con base en los resultados se verifica que con el CRI no se requiere un orden estricto en el disparo de las transiciones a diferencia de tener un CR. Con el CR sólo se puede usar una entrada a la vez y dependiendo de la entrada que se seleccione primero se establece un orden para usar las otras entradas, pero con el CRI las tres entradas se pueden usar indistintamente, es decir, están disponibles en todo momento a excepción del instante en que se desee verificar si hay alguna falta en el sistema, especialmente si se detecta que una de las entradas no se utiliza. El CRI sólo actúa cuando se requiere detectar si hay alguna falta. Por otro lado, en la figura 4 se nota que un CR implica colocar más lugares a diferencia del uso del CRI de la figura 6 y otra cosa que se puede notar es que el CR modifica la estructura de la RP, pasando de 12 T-semiflujos a uno solo y con el CRI se mantiene la misma cantidad. Además, si se observara el lenguaje de la RP con el CRI se notaría que realmente no se restringe, sigue siendo el mismo y sólo se limita cuando se requiere verificar si existe o no una falta. Sin embargo, el lenguaje se limita drásticamente cuando se usa el CR porque sólo se puede usar una entrada a la vez. Algo importante que señalar es que el CR se propuso para RP binarias y aunque se ve en el ejemplo que se puede usar para RP no binarias no está analizado el caso, pero el CRI si se analiza para RP no binarias.

El sistema continúa conservando las propiedades de vivacidad y puede ser diagnosticable con la propuesta de la diagnosticabilidad activa. El diagnóstico activo se aplica cuando no hay seguridad de que haya ocurrido una falta, pero es

probable que haya sucedido, esto ocurre cuando los usuarios eligen una sola entrada o dejan de usar una de las entradas y se puede verificar si todo está en orden o si puede haber ocurrido una falta.

El STD sirve para realizar el diagnóstico activo y considera las condiciones de funcionamiento del CRI, es el que controla cuándo revisar si ocurrió una falta después de que se usen las entradas k veces.

Es posible que los mismos usuarios ayuden a verificar el sistema cuando alguien elija otra entrada diferente a la que usan muchos usuarios y eso evitaría parar el sistema por un momento.

5. Conclusiones

Este trabajo reporta un diagnóstico activo para los SEDs y aplica los resultados al problema de diagnóstico de faltas de los sistemas de estacionamiento con tres entradas y una salida. Las principales contribuciones en el área son: 1) el diagnóstico se realiza usando un Circuito de Regulación Inteligente, 2) se introduce una definición de diagnóstico activo para detección de faltas para SED controlables, y 3) los resultados son usados para diagnosticar un estacionamiento. Como trabajo futuro se considera extender el diagnóstico de otros tipos de faltas y a otras clases de RP, y contar con un algoritmo.

6. Bibliografía y Referencias

- [1] Cabasino M.P., Lafortune S. and Seatzu C. Optimal sensor selection for ensuring diagnosability in labeled Petri nets. *Automatica*, vol. 49. Pp.2372-2383, 2013.
- [2] Densel J. and Esparza J. *Free Choice Petri Nets*. University Press. Cambridge, 1995.
- [3] Murata T. Petri nets: properties, analysis and applications. *Proceedings of IEEE*, vol.77. No.4. Pp.541-580, 1989.
- [4] Dotoli M., Fanti M.P., Mangini A.M. and Ukovich W. On-line Fault Detection in Discrete Event Systems by Petri nets and Integer Linear Programming. *Automatica*, vol. 45. no. 11. Pp. 2665-2672, October 2009.

- [5] Hernández-Rueda K., Meda-Campaña M.E. and Arámburo-Lizárraga J. Enforcing Diagnosability in Interpreted Petri Nets. *IFAC-Papers On Line*, vol. 48. No. 7. Pp. 58-63. DOI: 10.1016/j.ifacol.2015.06.473, 2015
- [6] Lafortune S. and Genc S. Distributed diagnosis of place-bordered petri nets. *IEEE Transactions on Automation Science and Engineering*, vol.4. No. 2. April. Pp.206-219, 2007.
- [7] Lefebvre D. and Leclercq E. Stochastic Petri nets identification for the fault detection and isolation of discrete event systems. *IEEE Transactions on Systems, Man, Cybernetics, A., Syst. Humans*, vol. 41. No. 2. Pp. 213-225, 2011.
- [8] Ramírez-Treviño A., Ruiz-Beltrán E., Arámburo J. and López-Mellado E. Structural Diagnosability of DES and Design of Reduced Petri Net Diagnosers. *IEEE Transactions on Systems, Man and Cybernetics*, vol. 42. No.2. Pp. 416-429, 2012.
- [9] Rivera-Rangel I., Ramírez-Treviño A., Aguirre-Salas L.I. and Ruiz-León J. Geometrical characterization of Observability in Interpreted Petri Nets. *Kybernetika*, vol. 41. Pp. 553-574, 2005.
- [10] Ruiz-Beltrán E., Ramírez-Treviño A. and Orozco-Mora J.L. Formal Methods in Manufacturing: Fault Diagnosis in Petri Nets. Edited by Javier Campos, Carla Seatzu and Xiaolan Xie. CRC Press Taylor-Francis Group. Boca Raton, FL. Pages 728, 2014.
- [11] Ruiz-Beltrán E., Ramirez-Treviño A., López-Mellado E. and Arámburo-Lizárraga J. A Structural Characterization of Diagnosable Petri Net Models. *Proceedings of the 3rd Annual IEEE Conference on Automation Science and Engineering*. Scottsdale, AZ, USA. Pp.1137-1142. Sept 22-25, 2007.
- [12] Sampath M., Sengupta R., Lafortune S., Sinnamohideen and K., Teneketzis D.C. Diagnosability of discrete event systems. *IEEE Transactions on Automatic and Control*, vol.4. No.9. Pp.1555-1575, 1995.
- [13] Sampath M., Sengupta R., Lafortune S., Sinnamohideen and K., Teneketzis D.C. Diagnosis of Discrete-Event Systems. *IEEE Transactions on Automatic and Control*, vol. 43. No.7. Pp. 908-929, 1998.

- [14] Sampath M., Sengupta R., Lafortune S., Sinnamohideen K. and Teneketzis D.C. Failure Diagnosis Using Discrete-Event Models. *IEEE Transactions on Control Systems Technology*, vol.4. No. 2. Pp.105-124, 1996.
- [15] Seatzu C. and Giua A. Fault Detection for Discrete Event Systems using Petri nets with unobservable transition. *IEEE CDCD*. Pp.6323-6328, December 2005.
- [16] Wu Y. and Hadjicostis C. N. Algebraic approaches for fault identification in discrete-event systems. *IEEE Trans. Robotics and Automation*, vol. 50. No.12. Pp. 2048–2053, 2005.
- [17] Ziqiang C., Feng L., Caisheng W. G., Wang L. Y. and Min X. Active Diagnosability of Discrete Event Systems and its Application to Battery Fault Diagnosis. *IEEE Transactions on Control Systems Technology*, vol.22. No.5. Pp.1892-1898, 2014.

CONTROL DEL FLUJO DE POTENCIA HACIA LA RED ELÉCTRICA DE UN SISTEMA DE GENERACIÓN EÓLICA EMPLEANDO UN GENERADOR DE INDUCCIÓN DE DOBLE ALIMENTACIÓN

Pedro Hernández Tenorio

IPN, Escuela Superior de Ingeniería Mecánica y Eléctrica
phernandezt1003@alumno.ipn.mx

Jaime José Rodríguez Rivas

IPN, Escuela Superior de Ingeniería Mecánica y Eléctrica
jjrodriguezr@ipn.mx

Oscar Carranza Castillo

IPN, Escuela Superior de Cómputo y Escuela Superior de Ingeniería Mecánica y Eléctrica
ocarranzac@ipn.mx

Rubén Ortega González

IPN, Escuela Superior de Cómputo y Escuela Superior de Ingeniería Mecánica y Eléctrica
rortegag@ipn.mx

Resumen

En este trabajo se presenta el control del flujo de potencia hacia la red eléctrica, a través del Convertidor del Lado de la Red (CLR) de un Sistema de Generación Eólica (SGE), con el objetivo de ser utilizado en un Convertidor Back to Back (CBB) que se requiere conectar a un Generador de Inducción de Doble Alimentación (GIDA). Se diseñan los controladores Proporcionales-Integrales requeridos para controlar la tensión del bus de CD, así como, la Potencia Activa y Potencia Reactiva del SGE. Se simula el CLR en el programa Simulink de Matlab®, donde se valida el funcionamiento del convertidor en los dos modos de operación: como rectificador y como inversor. Aplicando la técnica del control

vectorial se obtiene control total en el intercambio de potencias entre el SGE y la red eléctrica.

Palabras Claves: Control Vectorial, control proporcional-integral, convertidor del lado de la red, flujo de potencia.

Abstract

In this paper is presented the control of the power flow to the electrical network, through the Grid-Side Converter (GSC) of a Wind Generation System (WGS) with the objective of being used in a Back to Back Converter that is required to be connected to a Doubly Fed Induction Generator (DFIG). The Proportional-Integral controls are designed in order to control the voltage on the DC Link, as well as, the Active and Reactive Power of the WGS. The GSC is simulated in the program Simulink of Matlab®, where the operation of the converter is validated in the two modes of operation: as a rectifier and as an inverter. Applying the vector control technique is obtained the total control in the power exchange between the WGS and the electrical network.

Keywords: *Grid side converter, power flow, proportional-integral control, vector control.*

1. Introducción

El continuo cambio climático, así como, el uso desmedido de combustibles fósiles propicia la búsqueda de nuevas formas de producir energía eléctrica para abastecer la constante demanda. La energía fotovoltaica, así como, la energía eólica se ha colocado como fuentes de energía alternativas a las convencionales, ya que no producen gases de efecto invernadero, además de que son provenientes de fuentes renovables, como es la energía del sol y la energía del viento [Hamdan, 2014]. Este trabajo se enfoca en la generación de energía eléctrica por medio de la energía eólica.

En las últimas décadas el uso de la energía eólica ha incrementado considerablemente, a tal grado que se considera que la energía eólica es la fuente renovable con mayor crecimiento [Mathew, 2011]. El crecimiento más significativo

se ha dado en los países de los Estados Unidos, Alemania, España y la India [U.S. Department of Energy, 2006]. Los SGE se han orientado al tipo de velocidad variable, ya que presentan menor desgaste en los componentes, menor fluctuación de la potencia inyectada a la red y mejor desempeño en mayor rango de velocidades de viento, lo cual hace posible maximizar la extracción de energía a diferentes velocidades del viento [Chen, 2007].

Para este trabajo se considera un SGE basado en un generador del tipo GIDA como se muestra en la figura 1. En esta topología de generación se tiene una turbina eólica de velocidad variable, la cual se conecta a un GIDA a través de una caja de engranes o multiplicadora, el devanado del estator del GIDA se conecta a un devanado de un transformador trifásico, mientras que el devanado del rotor, se conecta a un Convertidor Electrónico de Potencia (CEP), mejor conocido como Convertidor Back-to-Back (CBB).

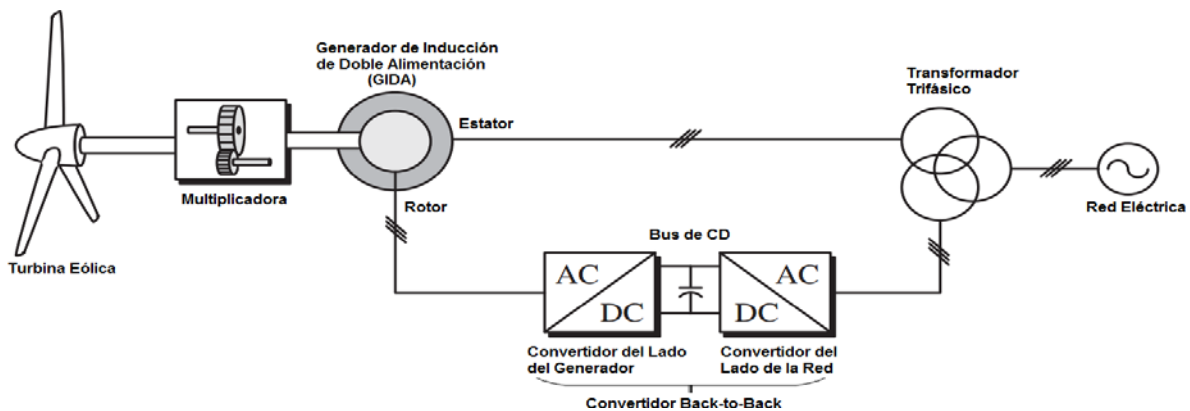


Figura 1 Sistema de Generación Eólica con un generador tipo GIDA [Abu-Rub, 2014].

El CBB está conformado por dos convertidores trifásicos, que se unen mediante un bus de CD, donde el Convertidor del Lado del Generador (CLG) se conecta al devanado del rotor, trabajando como rectificador, convirtiendo voltajes de CA a CD, y proporcionando el voltaje del bus de CD, mientras que el Convertidor del Lado de la Red (CLR) se alimenta mediante el voltaje en el bus de CD, y trabaja como inversor, convirtiendo voltaje de CD a CA. En ambos casos se aplican técnicas de modulación del ancho de los pulsos (PWM del inglés Pulse Width Modulation), con el objetivo de producir formas de onda trifásicas sinusoidales,

con la menor cantidad de distorsión armónica e inyectar a la red la mayor cantidad de potencia [Mohan, 2003].

El CBB se conecta al segundo devanado del transformador, donde finalmente el tercer devanado del transformador se interconecta con la red eléctrica [Abu-Rub, 2014]. El CBB, también es conocido como convertidor reversible, debido a que el flujo de potencia puede ir en sentido contrario, es decir, desde la red eléctrica hacia el devanado del rotor del generador. Por lo cual, ambos convertidores tienen la capacidad de trabajar como rectificador o como inversor, dependiendo del sentido del flujo de potencia. El CBB garantiza la generación de energía a frecuencia nominal de la red y voltaje nominal de la red, independientemente de la velocidad a que gire el rotor [Abad, 2011].

El SGE de la figura 1, se conoce como un Esquema de Generación Dividida (EGD), ya que la mayor parte de potencia generada, es inyectada mediante el devanado del estator del generador, y aproximadamente de un 25-30% de la potencia total, es inyectada a través del CBB a la red trifásica mediante el devanado del rotor del generador [Chen, 2007]. Al distribuirse la potencia generada, a través del estator y otra parte a través del rotor, se reduce el costo y tamaño del convertidor de potencia requerido.

El avance en los dispositivos de electrónica de potencia, así como, el uso de diferentes tipos de generadores ha permitido la implementación de diversas topologías de Sistemas de Generación Eólica, que presentan ventajas y desventajas entre sí. En general el tipo de sistema se caracteriza de acuerdo al generador eléctrico y al modo en que se conecta el Convertidor Electrónico de Potencia. La mayoría de las turbinas eólicas instaladas son de generación a velocidad variable basadas en GIDA, compartiendo el mercado con el Generador Síncrono de Rotor Devanado y esquemas de generación basadas en Generador Síncrono de Imanes Permanentes [Abad, 2011]. Algunas de las características de la topología con un generador GIDA son: el rango de velocidad está limitado a un $\pm 30\%$ de la velocidad síncrona, el convertidor es de un 25 a 30% la potencia nominal del generador, presenta un tamaño reducido del convertidor, lo que se traduce en un costo menor y menos pérdidas de potencia, control completo de

potencia activa y reactiva con la red eléctrica a través del flujo del rotor, utiliza escobillas en el rotor que requieren mantenimiento continuo y utiliza multiplicadora, que también requiere mantenimiento regular [Chen, 2007].

Este trabajo se enfoca en el control del Convertidor del Lado de la Red (CLR), analizando sus dos modos de operación: como fuente de alimentación de voltaje de CD (modo rectificador), donde la potencia fluye desde la red hacia el CBB, y en el modo de operación como inversor, donde el flujo de potencia va desde el CBB hacia la red.

Para llevar a cabo el control del CLR, se utiliza el Control Vectorial y se diseñan los controladores de tensión y corriente, encargados de controlar el voltaje en el bus de CD, y controlar la potencia activa y reactiva generada. Además, para la sincronización con la red eléctrica se utiliza un lazo de enganche de fase (PLL, Phase Locked Loop). El convertidor es validado mediante la simulación.

2. Métodos

Primeramente, se requiere modelar la red eléctrica en un Marco de Referencia Síncrono (MRS), para posteriormente aplicar el Control Vectorial, por lo que se requiere transformar las coordenadas de un Marco Estacionario Trifásico (X_{abc}) a un MRS (X_{dq}), utilizando las transformaciones de Clarke y Park, respectivamente.

En la figura 2, se muestra el diagrama eléctrico de la red trifásica y el CLR. Se tiene el bus de CD, el convertidor CLR, un filtro L trifásico y los voltajes de la red.

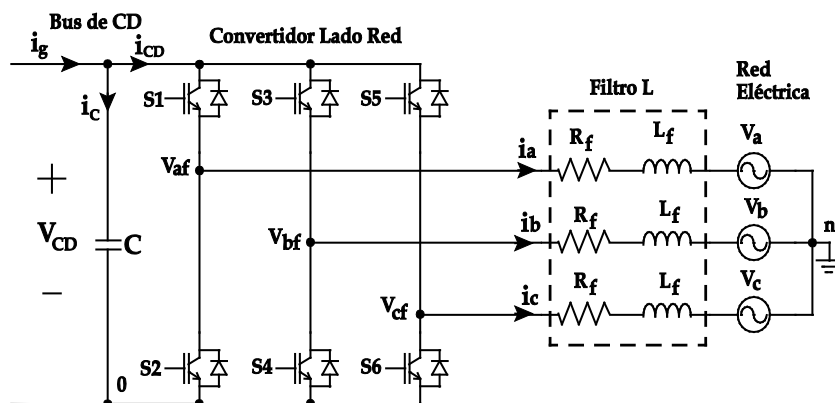


Figura 2 Diagrama eléctrico de la red trifásica y convertidor del lado de la red.

El convertidor es modelado con interruptores bidireccionales, donde el interruptor ideal normalmente es creado por un semiconductor con un diodo en antiparalelo, que permite el flujo de corriente en ambas direcciones, en este trabajo se utiliza el Transistor Bipolar de Compuerta Aislada o IGBT (Insulated Gate Bipolar Transistor) con un diodo en antiparalelo. Para llevar a cabo la conexión del CLR con la red eléctrica se utiliza un filtro L por cada fase de salida del convertidor, donde se incluye la resistencia propia del inductor. En el lado de CA del convertidor, se requiere que el filtro L reduzca los armónicos de la corriente generados por el convertidor, para que no entren en el sistema de la red [Mohan, 2003]. Los voltajes del convertidor en forma vectorial están dados por ecuación 1.

$$\vec{V}_{abc f} = R_f \vec{i}_{abc} + L_f \frac{d(\vec{i}_{abc})}{dt} + \vec{V}_{abc} \quad (1)$$

Donde: $\vec{V}_{abc f}$ son los voltajes trifásicos de salida de convertidor, en (V), \vec{i}_{abc} son las corrientes trifásicas provenientes del convertidor, en (A), L_f es la inductancia del filtro de conexión con la red, en (H), R_f es la resistencia propia del filtro de la red, en (Ω) y \vec{V}_{abc} son los voltajes de la red, en (V). Para llevar a cabo el modelado del convertidor, se consideran que los voltajes son sinusoidales y balanceados. Aplicando la transformación de Clarke al sistema trifásico, se obtiene la ecuación 2, luego al aplicar la transformada de Park se obtiene la ecuación 3.

$$\vec{V}_{\alpha\beta f} = R_f \vec{i}_{\alpha\beta} + L_f \frac{d(\vec{i}_{\alpha\beta})}{dt} + \vec{V}_{\alpha\beta} \quad (2)$$

$$\vec{V}_{dq f} = R_f \vec{i}_{dq} + L_f \frac{d(\vec{i}_{dq})}{dt} + j\omega_{sinc} L_f \vec{i}_{dq} + \vec{V}_{dq} \quad (3)$$

Donde ω_{sinc} es la velocidad angular de la red eléctrica, en (rad/s). La ecuación 3 representa el sistema trifásico abc en el Marco de Referencia Síncrono (MRS) con componentes dq. La expresión 3 se compone por una parte real y una parte imaginaria [Bose, 2002], como se muestra en las ecuaciones 4 y 5.

$$V_{df} = R_f i_d + L_f \frac{d(i_d)}{dt} - \omega_{sinc} L_f i_q + V_d \quad (4)$$

$$V_{qf} = R_f i_q + L_f \frac{d(i_q)}{dt} + \omega_{sync} L_f i_d + V_q \quad (5)$$

Se observa que en las ecuaciones 4 y 5, se tiene un acoplamiento debido a las corrientes i_q en el voltaje V_{df} y la corriente i_d en el voltaje V_{qf} , aplicando el control vectorial se logra obtener el desacople de dichas corrientes. Para hacer esto posible, se realiza la alineación del eje d del marco síncrono con el vector espacial del voltaje de la red \vec{V}_{dq} [Abad, 2011]. Debido a que la componente V_d está alineado con el eje d del marco síncrono, el vector espacial de la red no tiene proyección en el eje q, por lo cual $V_q=0$, y la amplitud del voltaje V_d es igual a la amplitud del vector espacial del voltaje de la red, estas dos consideraciones son vitales para el control vectorial, ecuaciones 6 y 7.

$$V_d = |\vec{V}_{dq}| \quad (6)$$

$$V_q = 0 \quad (7)$$

Debido a las ecuaciones 6 y 7, las expresiones 4 y 5 se modifican a ecuaciones 8 y 9.

$$V_{df} = R_f i_d + L_f \frac{d(i_d)}{dt} \quad (8)$$

$$V_{qf} = R_f i_q + L_f \frac{d(i_q)}{dt} \quad (9)$$

De las ecuaciones 8 y 9, se obtienen los controladores de corriente requeridos para controlar la corriente del CLR, debido a que las dos funciones son idénticas, se procede a realizar el diseño para un controlador y aplicarlo al otro.

Diseño de controlador PI de corriente

Aplicando la transformada de Laplace a la ecuación 8, se obtiene la Función de Transferencia (FT) del controlador de corriente del CLR, para ambos ejes d y q, ecuación 10.

$$\frac{I_d(s)}{V_d(s)} = \frac{1}{sL_f + R_f} = \frac{I_q(s)}{V_q(s)} \quad (10)$$

La FT a lazo abierto (G_{la_i}) y la FT a lazo cerrado (G_{lc_i}) de corriente están dadas por ecuaciones 11 y 12.

$$G_{la_i}(s) = \left(k_{pi} + \frac{k_{ii}}{s} \right) \left(\frac{1}{sL_f + R_f} \right) \quad (11)$$

$$G_{lc_i}(s) = \left(\frac{G_{la_i}}{1 + G_{la_i}} \right) \quad (12)$$

Para el desarrollo de los controladores requeridos, se utiliza el método del lugar de las raíces, diseñado por W. R. Evans para encontrar las raíces de la ecuación característica de la función de lazo cerrado, este método se utiliza ampliamente en la Ingeniería de Control [Ogata, 2002]. Se sigue el criterio de estabilidad de Nyquist, el cual indica que se puede averiguar la estabilidad relativa y absoluta de los sistemas lineales en lazo cerrado, a partir del conocimiento de sus características de frecuencia en lazo abierto [Ogata, 2002]. Para este método se debe cumplir con las siguientes condiciones de estabilidad, ecuaciones 13 y 14.

$$|G_{la_i}(s)|_{s=j\omega_c} = 1 \quad (13)$$

$$\angle G_{la_i}(s)|_{s=j\omega_c} = MF - \pi \quad (14)$$

Donde ω_c es la frecuencia de cruce (o ancho de banda) del controlador, el cual establece la velocidad de respuesta del lazo de control, y MF es el Margen de Fase de la función de transferencia. Aplicando las ecuaciones 13 y 14 a la expresión 11, se obtienen la ganancia proporcional (k_{pi}) e integral (k_{ii}), ecuaciones 15 y 16.

$$k_{pi} = \frac{k_{ii}}{\omega_{c_i}} \left(\tan \left(MF - \frac{\pi}{2} + \tan^{-1} \left(\frac{\omega_{c_i} L_f}{R_f} \right) \right) \right) \quad (15)$$

$$k_{ii} = \omega_{c_i} \sqrt{\frac{(\omega_{c_i} L_f)^2 + R_f^2}{1 + \tan^2 \left(MF - \frac{\pi}{2} + \tan^{-1} \left(\frac{\omega_{c_i} L_f}{R_f} \right) \right)}} \quad (16)$$

Se conocen los valores de $R_f=0.5585 \Omega$ y $L_f=9.0897 \text{ mH}$. Para el controlador de corriente se propone una frecuencia de cruce $\omega_{c_i}=300 \text{ Hz}$ y un $MF=60^\circ$. En

Electrónica de Potencia, se suele elegir un MF superior a 50° y un Margen de Ganancia (MG) superior a 6 dB [Garcerá, Figueres & Abellán, 1998], con los datos que se tienen se calculan las ganancias proporcional e integral del controlador de corriente. Con los valores de las ganancias $k_{pi}=14.5589$ y $k_{ii}=17060.0$, se obtiene el diagrama de Bode la FT de lazo abierto del controlador de corriente (G_{la_i}), que se muestra en la figura 3a, donde se tiene una frecuencia de corte a los 300 Hz y un margen de fase de 60° , obteniendo un lazo de control estable. Se cierra el lazo de control y se calcula la FT a lazo cerrado, de la cual se obtiene la gráfica del lugar de las raíces, esta se muestra en la figura 3b, donde se observa que todos sus polos se encuentran en el lado negativo de plano complejo, lo que indica que la FT del controlador de corriente a lazo cerrado es estable.

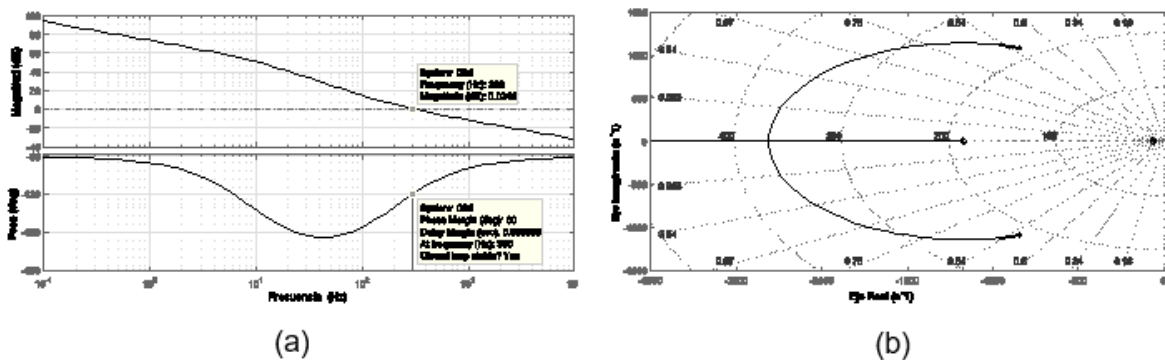


Figura 3 (a) Diagrama de Bode de G_{la_i} , (b) Lugar de las raíces de G_{la_i} .

Diseño de Controlador PI de Tensión

El bus de CD está compuesto por un capacitor que une al convertidor del lado del generador con el convertidor del lado de la red, proporcionando un voltaje de CD constante que requiere el convertidor para operar. El bus de CD está en constante intercambio de energía, su principal propósito es mantener en sus terminales el voltaje establecido, garantizando que la potencia del sistema de CD sea igual a la potencia del sistema de CA, por lo cual las potencias de los dos sistemas deben de ser iguales, como se establece en ecuación 17.

$$V_{CD} i_g - V_{CD} C \frac{d(V_{CD})}{dt} = V_d i_d + V_q i_q \quad (17)$$

El convertidor conmutado del lado de la red es un sistema no lineal, como se observa en la ecuación 17. Para garantizar que una fuente de alimentación proporcionará una tensión de salida regulada, establecida por una señal de consigna o referencia, se debe tener un lazo cerrado. Para obtener un control realimentado lineal a partir de un circuito no lineal, como lo es un convertidor conmutado, se debe linealizar la etapa de potencia. Ante pequeñas perturbaciones el convertidor tiene un modelo lineal de pequeña señal, que permite obtener reguladores lineales para cerrar lazos de control, por lo cual se requiere obtener el circuito equivalente de pequeña señal y baja frecuencia, pues con este circuito se puede obtener la FT del sistema a controlar [Garcerá, 1998].

Aplicando la técnica de pequeña señal a la ecuación 17, se obtiene la etapa de potencia del CLR linealizada, ecuación 18.

$$-V_{CD}C \frac{d(\hat{v}_{CD})}{dt} + V_{CD}\hat{i}_g + I_g\hat{v}_{CD} = V_d\hat{i}_d + \hat{v}_d I_d + V_q\hat{i}_q + \hat{v}_q I_q \quad (18)$$

Considerando las ecuaciones 6 y 7, y aplicando la transformada de Laplace, se obtiene la función de transferencia del regulador de tensión del bus de CD, ecuación 19.

$$\left. \frac{\hat{v}_{CD}(s)}{\hat{i}_d(s)} \right|_{\hat{v}_d=\hat{i}_g=0} = \frac{V_d}{-sCV_{CD} + I_g} = \frac{-V_d}{sCV_{CD} - I_g} \quad (19)$$

Para el diseño del controlador de tensión, se tiene un sistema en cascada, debido a que internamente se tiene la FT de lazo cerrado del controlador de corriente, por lo cual la FT a lazo abierto del bus de CD (G_{la_CD}), considera todos los elementos que se encuentran en su trayectoria directa, ecuación 20.

$$G_{la_CD}(s) = \left(k_{pCD} + \frac{k_{iCD}}{s} \right) \left(\frac{G_{la_i}}{1 + G_{la_i}} \right) \left(\frac{-V_d}{sCV_{CD} - I_g} \right) \quad (20)$$

Donde k_{pCD} es la ganancia del controlador proporcional de tensión y k_{iCD} es la ganancia integral del controlador de tensión, V_{CD} es el voltaje en el bus de CD, C es el valor de la capacitancia del capacitor en el bus de CD, V_d es el valor de tensión de línea a línea (V_{LL}) e I_g es la corriente que proviene del convertidor del lado del generador.

Debido al esquema en cascada, se requiere que el controlador de tensión actúe más lentamente con respecto al controlador de corriente, además se tiene que el bus de CD es un sistema inestable, por lo tanto, su ancho de banda debe ser lo suficientemente bajo para evitar entrar en inestabilidad [Ogata, 2002]. Por lo anterior, se propone un ancho de banda para el lazo de tensión, 10 veces menor al ancho de banda del controlador de corriente. Por lo que la frecuencia de cruce es $\omega_{c_CD} = 30$ Hz y se establece un MF de 60° . Se sabe que $V_{LL} = 220$ V, $V_{CD} = 360$ V, $C = 2200$ μ F, y se tienen los controladores PI de corriente de la sección anterior. Aplicando las ecuaciones 13 y 14 a la expresión (20), se obtienen los valores de las ganancias del controlador de tensión: $k_{pCD} = -0.5804$ y $k_{iCD} = -61.8415$.

Con los valores de las ganancias PI de tensión se obtiene la FT de lazo abierto del controlador de tensión y su diagrama de Bode donde se tiene una frecuencia de corte a los 30 Hz y un MF de 60° , obteniendo un lazo de control estable. Se cierra el lazo de control y se calcula la FT de tensión a lazo cerrado (G_{lc_CD}), de la cual se obtiene el lugar de las raíces, teniendo los polos en el lado negativo de plano complejo, por lo que la FT del controlador de tensión a lazo cerrado es estable.

Diseño de Controlador PI de Lazo de Enganche de Fase (PLL)

El CLR requiere el uso de un PLL (Phase Locked Loop por sus siglas en inglés), o bien un Lazo de Enganche de Fase, que permite la sincronización con la red, ya que, para inyectar potencia a la red, se requiere conocer la secuencia de fases y estar sincronizados a una fase de la red. El PLL permite obtener el valor de la posición angular theta (θ) que se requiere para diversas operaciones del control.

El funcionamiento del PLL se basa en el uso de las coordenadas dq del voltaje de la red, alineando la componente d del MRS con el eje d del voltaje de la red, lo que significa que el ángulo del marco síncrono será modificado, con el propósito de alinear ambos ejes d, hasta que la componente q del voltaje de la red sea cero, en ese momento, se puede decir que el MRS (dq) y el vector espacial del voltaje de la red han sido sincronizados y alineados al eje d.

La FT del PLL a lazo abierto y la FT a lazo cerrado están dadas por ecuaciones 21 y 22.

$$G_{I\alpha_PLL}(s) = \left(k_{pPLL} + \frac{k_{iPLL}}{s} \right) \left(\frac{V_{LL}}{s} \right) \quad (21)$$

$$G_{I\alpha_PLL}(s) = \left(\frac{G_{I\alpha_PLL}}{1 + G_{I\alpha_PLL}} \right) \quad (22)$$

Aplicando las ecuaciones 13 y 14 a la expresión 21, se obtienen las funciones para calcular la ganancia proporcional e integral del controlador del PLL, ecuaciones 23 y 24.

$$k_{pPLL} = \frac{k_{ii}}{\omega_{c_PLL}} \tan(MF) \quad (23)$$

$$k_{iPLL} = \frac{\omega_{c_PLL}^2}{V_{LL} \sqrt{1 + \tan^2(MF)}} \quad (24)$$

Para este trabajo se propone una frecuencia de cruce $\omega_{c_PLL} = 400$ Hz, para que el controlador sea un poco más rápido que el controlador de corriente, el MF se establece a 60° . Con los datos proporcionados se obtiene que $k_{pPLL} = 14356.0$ y $k_{iPLL} = 9.8935$. Con los valores de ganancia proporcional e integral del controlador del PLL se obtienen el diagrama de Bode de la función $G_{I\alpha_PLL}$ (lazo abierto), con una frecuencia de cruce de 400 Hz y con MF de 60° , obteniendo un lazo de control estable. Se calcula la FT a lazo cerrado (G_{Ic_PLL}) y se obtiene el lugar de las raíces, donde los polos se encuentran del lado izquierdo del plano complejo, por lo cual se tiene que la FT del PLL a lazo cerrado es estable.

3. Resultados

A continuación, se presentan los resultados de simulación que se realizaron en este trabajo. En la figura 4a se muestra la simulación del PLL, con el cual se realiza la sincronización a la red. Donde entran los voltajes V_{abc} y se transforman al MRS para obtener las componentes V_{dq} , la componente V_q pasa por el controlador PI diseñado en la sección anterior, se integra la velocidad angular (ω_s) y se obtiene el ángulo Theta (θ). El proceso de sincronización con la red, se muestra en la parte superior de la figura 4b, donde se observa que el PLL se sincroniza con la fase V_a de la red, y se obtiene el ángulo Theta, el cual varía en el

rango de 0 a 2π , reseteando su valor en cada ciclo. En la parte inferior de la figura 4b se observa que el valor de $V_d = 220V$ y $V_q = 0V$, por lo que comprueba que la sincronización con la red, se ha establecido correctamente.

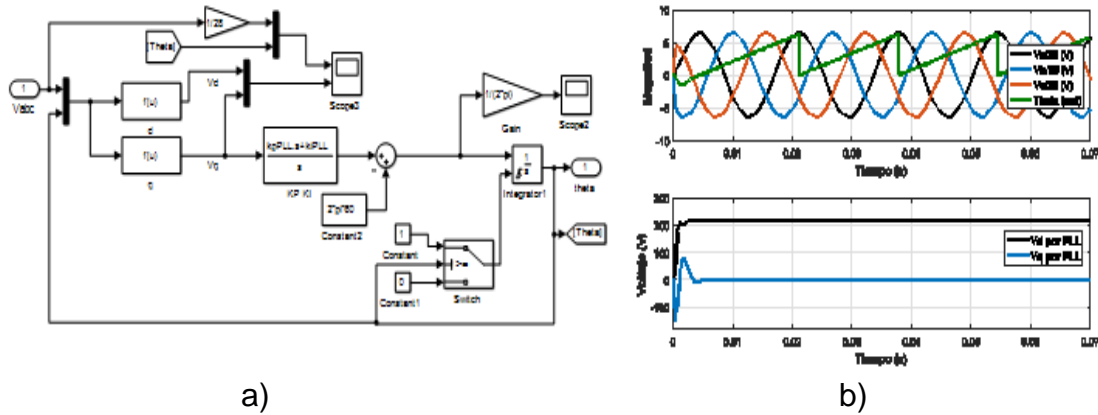


Figura 4 Simulación de PLL.

Convertidor en Modo Rectificador

En la figura 5a se muestra el CLR, en el modo de operación como rectificador, el cual incluye el capacitor del bus de CD, los interruptores de potencia (IGBTs), el filtro L para la conexión con la red y una etapa de medición de voltajes y corrientes. En la figura 5b se muestra el controlador PI de tensión para controlar el voltaje en el bus de CD, los dos controladores de corriente, así mismo, se incluye el desacople de las corrientes en los ejes d y q. A las salidas de los controladores de corriente se obtienen los voltajes de control V_{dqf}^* en el MRS.

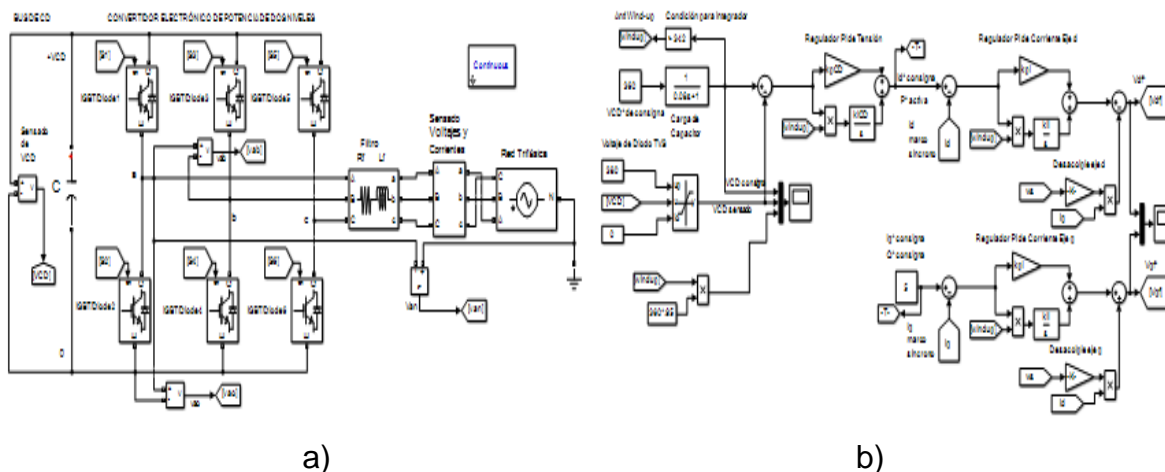


Figura 5 Convertidor trifásico, controladores PI de tensión y de corriente.

Cuando un controlador contiene elementos integradores, al realizar el proceso de integración sobre la señal de error, puede obtenerse un valor muy alto, lo que propicia que el integrador se sature, esto se conoce como efecto “windup” o de saturación [Seok, 2007]. Para este trabajo se aplica una técnica antiwind-up por medio de una integración condicional, esta técnica consiste en deshabilitar la parte integradora hasta que se cumpla una condición establecida. Esto permite que el valor del integrador, este siempre dentro de los valores permitidos establecidos por la condición, evitando el efecto de saturación.

La carga del capacitor del bus de CD, con la condición anti-windup y función Delay, se observa en la figura 6a. En una primera etapa, solo se aplica control proporcional, debido a esto se obtiene un error entre los voltajes de V_{CD} del convertidor y el voltaje V_{CD}^* referencia, el cual está marcado entre dos flechas. Al deshabilitar la condición antiwind-up, en el tiempo 0.24 s, se aplica el control proporcional-integral, propiciando que el error entre ambos voltajes sea prácticamente cero. En la figura 6b se muestran las corrientes de control o de consigna, donde la corriente i_d^* es la salida del regulador de tensión. Se observa que, al inicio de la carga del capacitor, el valor de referencia i_d^* se dispara hasta casi los 90 A, y disminuye su valor lentamente.

Debido a esto, se requiere aplicar la técnica antiwind-up, para evitar que el control integral demande al convertidor una corriente excesiva, la cual pudiera dañar el

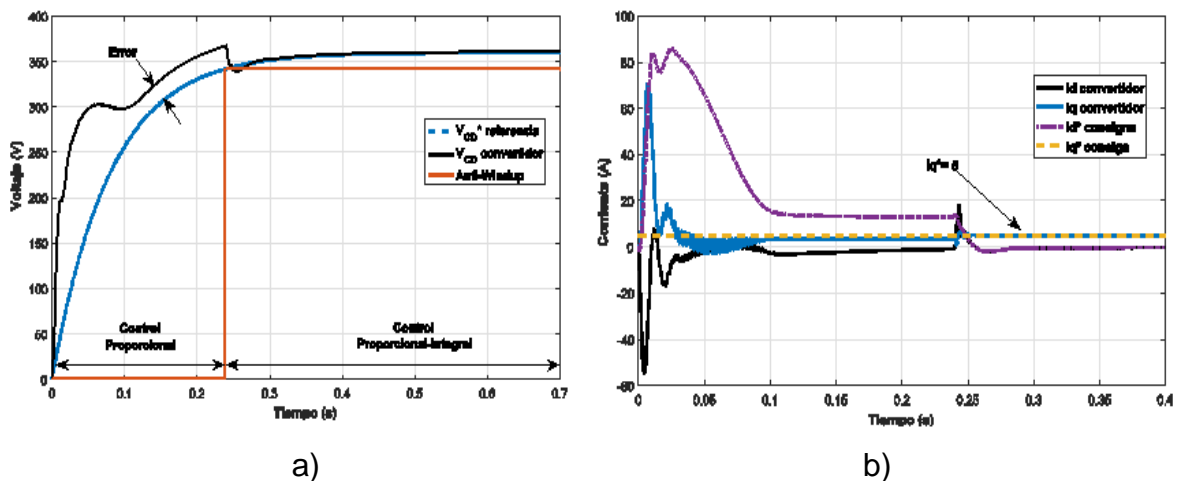


Figura 6 Carga del bus de CD, corrientes de control y del convertidor.

La corriente de consigna hasta el momento en que se habilita el control integral, lo cual pasa en el tiempo 0.24 s, luego en el tiempo 0.3 s se estabilizan las corrientes.

En la figura 7a se muestra la potencia activa y reactiva generada por el convertidor, donde la potencia activa generada en este modo de operación es mínima y cercana a cero ($P = 0W$), debido que la salida del controlador de tensión es cero cuando se aplica el control completo (PI), mientras que la potencia reactiva generada tiene un valor $Q = -1100 VARs$. En la figura 7b se muestran las corrientes generadas por el convertidor y los voltajes de la red, donde las corrientes tienen un desfase de 90° en adelante con respecto a los voltajes.

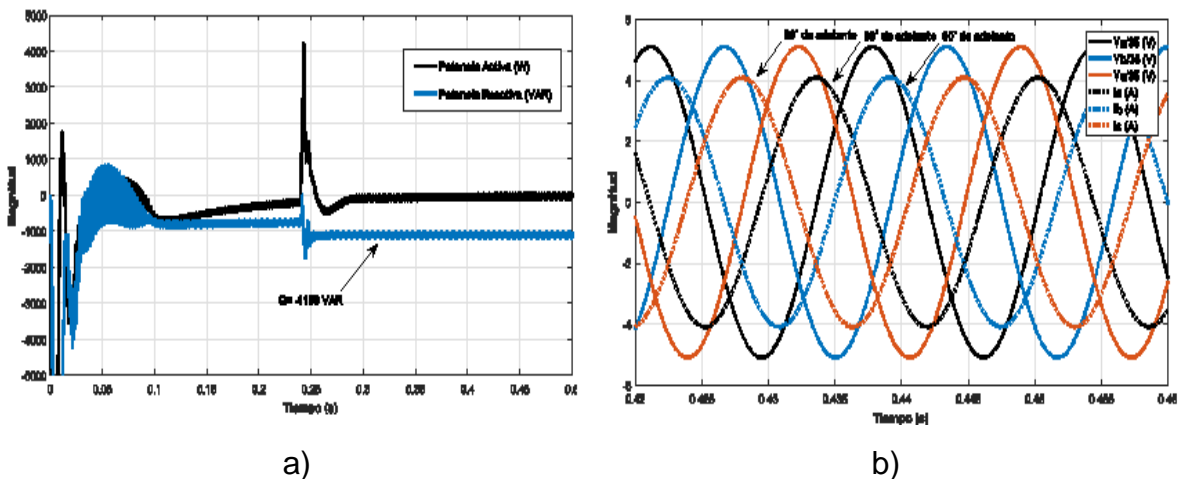


Figura 7 Potencia reactiva generada, Corrientes de convertidor y voltajes de red.

Convertidor en Modo Inversor

El controlador de corriente en el eje d, será responsable de la potencia activa, mientras que el controlador de corriente del eje q, será responsable de la potencia reactiva. En este modo de operación, se tiene control total de ambas potencias: activa y reactiva. En este trabajo, el convertidor se diseña para trabajar con un factor de potencia unitario, con la finalidad de maximizar el flujo de potencia que se inyecta a la red, por lo cual la corriente i_q^* consigna se establece en cero, lo que propicia que la potencia reactiva generada sea cero y que el factor de potencia sea unitario.

En la figura 8a se observa que ambas corrientes de consigna (i_d^* e i_q^*), al inicio tienen un valor de cero, en el tiempo 0.02 s se aplica una corriente de consigna en el eje d con valor de $i_d^* = 5$ A. Se observa que las corrientes d y q se estabilizan aproximadamente en el tiempo 0.03 s.

En la figura 8b se observa que la potencia activa generada por el convertidor es de $P=1100$ W, mientras que la potencia reactiva tiene un valor cercano a cero ($Q=0$ VARs), esto último se debe a que la referencia en el eje q es cero.

En la figura 9a se observa que en el tiempo 0.02 s, el convertidor comienza a generar corrientes, donde las corrientes están en fase con los voltajes de la red. En la figura 9b se observa que el convertidor trabaja con Factor de Potencia Unitario ($FP=1$) a partir del tiempo 0.02 s, cuando la corriente de consigna $i_d^* = 5$ A.

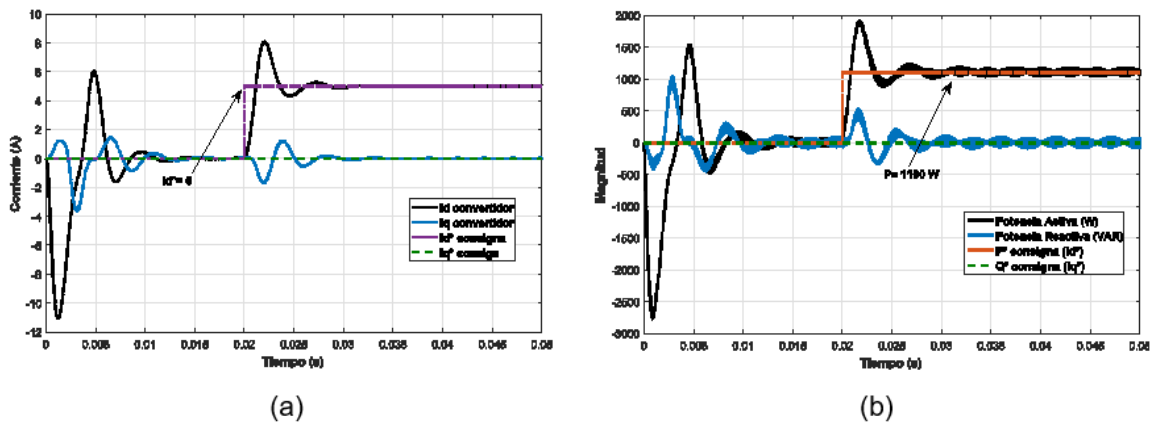


Figura 8 Corrientes de control y del convertidor, potencia activa generada.

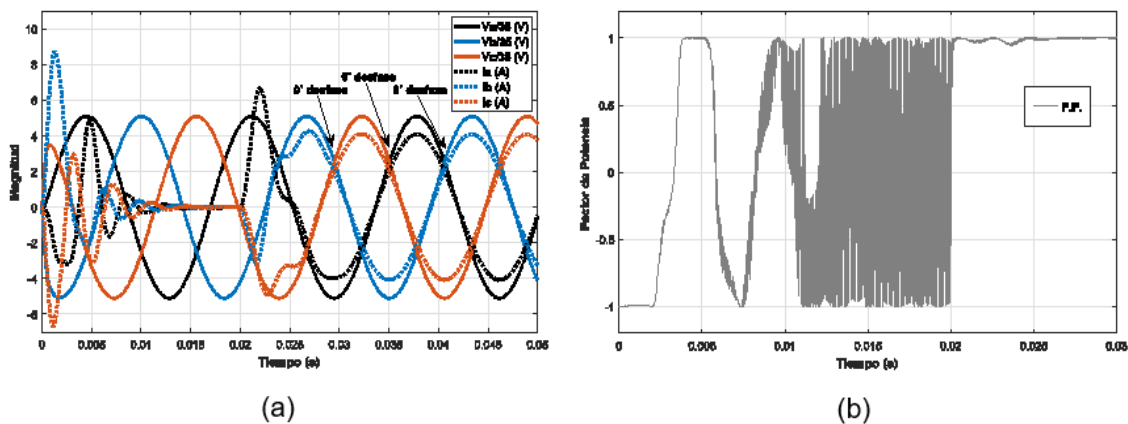


Figura 9 Corrientes del convertidor y voltajes de la red, factor de potencia.

4. Discusión

Para la simulación se consideran los siguientes elementos: los controladores de corriente, tensión y PLL, los parámetros de la inductancia del filtro L (con su respectiva resistencia) y capacitancia del bus de CD.

En el modo de operación como rectificador, la Potencia Activa (P) está controlada de manera indirecta por el valor de salida del controlador de tensión (i_d^* consigna), por lo cual únicamente se puede controlar la Potencia Reactiva (Q) del sistema a través del valor de referencia en i_q^* consigna. Se hace uso de una función de retraso (Delay) para que el voltaje de referencia realice una carga suave en el bus de CD, esto permitirá que el capacitor cargue de manera suave, hasta llegar al voltaje final del bus de CD (360 V).

En el modo de operación como inversor, la potencia proviene del rotor del generador y el CLG opera como rectificador, proporcionando un voltaje constante en el bus de CD, para que el CLR tenga una fuente de alimentación, donde el capacitor se sustituye por una fuente de CD con un valor de 360 V. El objetivo del CLR como inversor, es extraer la potencia en el bus de CD e inyectarla a la red trifásica. Debido a que no se realiza el control del voltaje del bus de CD, el controlador de tensión se omite, quedando únicamente los dos controladores de corriente.

5. Conclusiones

Los controladores de tensión, corriente y PLL diseñados, presentan una respuesta estable, esto se comprueba con el funcionamiento del convertidor, donde se obtienen resultados favorables en los dos modos de operación.

En modo rectificador, se obtiene una carga suave del bus de CD, donde la aplicación de una técnica antiwind-up evita la saturación de los controladores integrales y protege al CLR de generar corriente elevadas, además se tiene control sobre el voltaje del Bus de CD y control de la potencia reactiva.

En modo inversor, se tiene control total sobre las potencias activa y reactiva generadas por el convertidor. La potencia reactiva se lleva a cero y se obtiene un flujo de potencia activa hacia la red trifásica a factor de potencia unitario.

Con la técnica del control vectorial, se logra un correcto desacople de corrientes del convertidor, lo que permite tener control total sobre la potencia activa y reactiva que se intercambia entre el sistema de generación eólica y la red eléctrica.

6. Bibliografía y Referencias

- [1] Abad, G., López, J., Rodríguez, M. A., Marroyo, L., & Iwanski, G., *Doubly Fed Induction Machine, Modeling and Control for Wind Energy Generation*. USA: Wiley, 2011.
- [2] Abu-Rub, H., Malinowski, M., & Al-Haddad, K., *Power electronics for renewable energy systems, transportation and industrial applications*. Ed. Wiley, 2014.
- [3] Bose, B., *Modern Power Electronics and AC Drivers*. Ed. PrenticeHall, 2002.
- [4] Chen, Z., & Li, H., Overview of different wind generator systems and their comparisons. *IET Renewable Power Generation*, 2(2), pp. 123-138, 2007.
- [5] Garcerá, G., Figueres, E., & Abellán, A., *Convertidores conmutados: Circuitos de potencia y control*. Ed. Servicio de Publicaciones, 1998.
- [6] Hamdan, M. O., Hejase, H. A., M. Noura, H., & Fardoun, A. A., *ICREGA '14- Renewable Energy: Generation and Applications*. Ed. Springer, 2014.
- [7] Mathew, S., & Philip, G. S., *Advances in Wind Energy Conversion Technology*. Berlin: Springer, 2011.
- [8] Mohan, N., Undeland, T., & Robbins, W., *Power Electronics: Converters, Applications and Design*. USA: John Wiley & Sons, Inc, 2003.
- [9] Ogata, K., *Modern Control Engineering*. Prentice Hall, Inc, 2002.
- [10] Seok K., Kim, K., J. T. & Lee, C. D., Automatic Mode Switching of P/PI Speed Control for Industry Servo Drives Using Online Spectrum Analysis of Torque Command. *Industrial Electronics, IEEE Transactions*, 54(5), pp. 2642-2647, 2007.
- [11] U.S. Department of Energy, *Annual report on us wind power installation, cost, and performance trends: 2006, USA, 2006*.

OBTENCIÓN DEL MÁXIMO ANCHO DE BANDA PARA LA ADQUISICIÓN Y RECONSTRUCCIÓN DE SEÑALES ANALÓGICAS CON LA TARJETA SPARTAN 3E

Enrique Gerardo Hernández Vega

Tecnológico Nacional de México/Instituto Tecnológico de Chihuahua
ehernand@itchihuahua.edu.mx

Jorge Alberto Ortiz Gallo

Tecnológico Nacional de México/Instituto Tecnológico de Chihuahua
jaortiz@itchihuahua.edu.mx

Daniel Eduardo Morales Fernández

Tecnológico Nacional de México/Instituto Tecnológico de Chihuahua
demorales@itchihuahua.edu.mx

Alejandro Verduzco Hernández

Tecnológico Nacional de México/Instituto Tecnológico de Chihuahua
averduzco@itchihuahua.edu.mx

Resumen

En este trabajo se presenta un sistema en el cual se obtiene el máximo ancho de banda posible para la adquisición y reconstrucción de señales analógicas en la tarjeta de desarrollo Spartan-3E del fabricante Xilinx, utilizando los convertidores que contiene dicha tarjeta, tanto el ADC como el DAC. El sistema está desarrollado en VHDL empleando el concepto de máquina de estados finitos (FSM) y el administrador digital de reloj (DCM) incluido en el FPGA de la tarjeta. En aplicaciones tales como procesamiento digital de señales en tiempo real, comunicaciones digitales y control digital, por mencionar algunas, es muy importante tener un ancho de banda considerable en el sistema. El valor máximo obtenido para el ancho de banda del sistema fue de 161 kHz.

Palabras Claves: ADC, ancho de banda, DAC, FPGA, Spartan-3E.

Abstract

This paper presents a system in which the maximum bandwidth possible for the acquisition and reconstruction of analog signals is obtained in the Spartan-3E development board of the Xilinx manufacturer, using the converters contained in this board, both the ADC and The DAC. The system is developed in VHDL using the concept of finite state machine (FSM) and the digital clock manager (DCM) included in the FPGA. In applications such as digital processing of real-time signals, digital communications and digital control, to mention a few, it is very important to have a considerable bandwidth in the system. The maximum value obtained for the system bandwidth was 161 kHz.

Keywords: ADC, bandwidth, DAC, FPGA, Spartan-3E.

1. Introducción

En los planes y programas de estudio de Ingeniería Electrónica, no sólo en los Institutos Tecnológicos del país sino en todos los otros subsistemas educativos, se considera el uso de VHDL y FPGAs para el desarrollo de sistemas digitales y sus aplicaciones. Es común encontrar en las diversas instituciones educativas tarjetas de desarrollo como la Spartan-3E de Xilinx o sus equivalentes.

En el caso de la tarjeta Spartan-3E, la comunicación entre el FPGA y los convertidores de datos ADC y DAC, es a través de una sola interfaz serial SPI lo cual limita el ancho de banda del sistema. Además, la salida digital de 14 bits del ADC es en complemento a 2, mientras que el DAC es de 12 bits, tal cual se señala en la guía de usuario de la tarjeta [Xilinx, 2011]. Por estas razones, son pocos los trabajos de procesamiento de señales con la tarjeta Spartan-3E que utilicen ambos convertidores, la mayoría ha utilizado convertidores externos, ya sea uno o ambos [Domínguez, 2011].

La guía de usuario de la tarjeta contiene información muy escueta y algunos errores conceptuales como en el caso de la función de transferencia del ADC, lo cual puede generar errores en el cálculo de los valores de salida digital y su linealización dado que está en complemento a 2, como en [Mascharak, 2012].

El cuello de botella en el sistema es la interfaz SPI compartida por ambos convertidores cuyo límite máximo es de 50 MHz según el fabricante de los convertidores, por lo que a mayor frecuencia de la señal de reloj de esta interfaz se tendrá una mayor razón de muestreo y en consecuencia un mayor ancho de banda. En el resto de los trabajos similares consultados, la frecuencia de la señal de reloj de la interfaz SPI es menor a 10 MHz, ejemplo [Khedr, 2013].

El mayor ancho de banda obtenido usando los convertidores de datos, ADC y DAC, en la tarjeta Spartan-3E para la adquisición y reconstrucción de señales analógicas, es de 141 kHz con una razón de muestreo de 282 kmuestras/segundo. La frecuencia de la señal de reloj para la interfaz SPI es de 41.67 MHz para el manejo del ADC y DAC [Silage, 2008]. Es posible aumentar el ancho de banda del sistema de adquisición y reconstrucción de señales mediante el uso de un sintetizador digital de frecuencia a través de los administradores digitales de reloj (DCM), para aumentar la frecuencia de trabajo del SPI. Además, es posible utilizar un protocolo SPI de 24 bits en lugar de 32 bits para la comunicación con el DAC como se establece en la hoja de datos [Linear Technology Corporation, 2004].

2. Métodos

El hardware que se utilizó para llevar a cabo esta implementación fue la tarjeta de desarrollo Spartan-3E con un FPGA XC3S500E. La tarjeta contiene, además, entre otros muchos recursos, un circuito de captura análoga y un circuito de conversión digital a análoga. El diagrama a bloques del sistema se muestra en la figura 1. El software utilizado fue el ambiente de desarrollo ISE 14.7 de Xilinx.

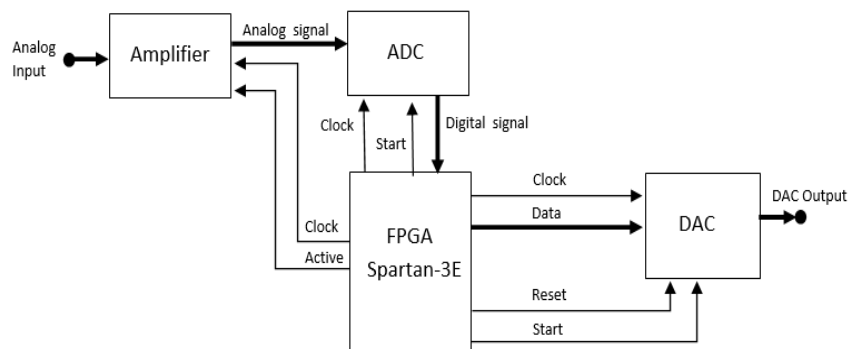


Figura 1 Diagrama a bloques del sistema.

El circuito de captura análoga consiste en un amplificador de ganancia programable (PGA LTC 6912) y de un convertidor análogo a digital (ADC LTC 1407A). Ambos son controlados por el FPGA a través de una interfaz SPI. El diagrama a bloques del circuito de captura análoga se muestra en la figura 2.

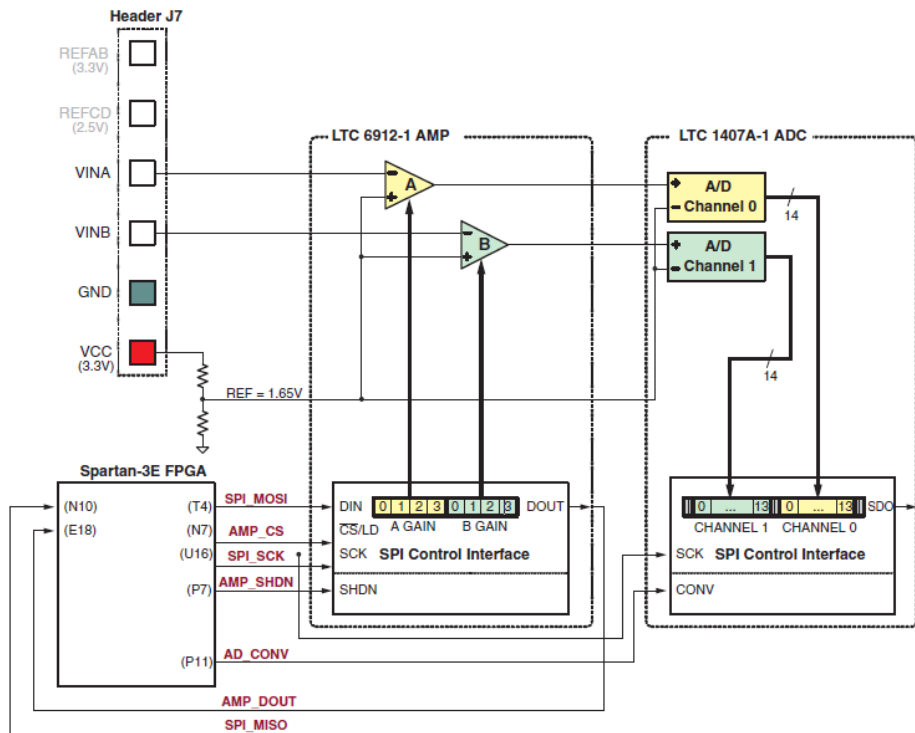


Figura 2 Vista detallada del circuito de captura análoga.

El circuito de captura análoga toma el voltaje de las entradas VINA o VINB y lo convierte a una representación digital de 14 bits, D[13:0], como se expresa en la ecuación 1.

$$D[13:0] = GANANCIA \times \frac{(V_{IN} - 1.65 V)}{1.25 V} \times 8192 \quad (1)$$

El voltaje de referencia del amplificador PGA y del ADC es de 1.65 V. El máximo rango del ADC es de ± 1.25 V, por lo que los límites de voltaje inferior y superior de la señal análoga serían 0.4 y 2.9 V respectivamente, para una ganancia de -1.

El hecho de que se comporte de esta manera hace que la salida digital sea un número con signo, por lo tanto, el bit más significativo será el signo y los 13 bits

restantes serán la magnitud, ya sea magnitud normal si el signo es positivo o magnitud en complemento a 2 si es negativo. Es por esto que la ecuación 1 sólo aplica para los números negativos y para los números positivos se utilizará la ecuación 2.

$$D[13:0] = GANANCIA \times \frac{(V_{IN} - 1.65 V)}{1.25 V} \times 8191 \quad (2)$$

En cambio, la salida del DAC (LTC 2624) es el análogo equivalente de un valor de 12 bits sin signo, escrito por el FPGA al DAC a través de una interfaz SPI, D[11:0], como se observa en la figura 3.

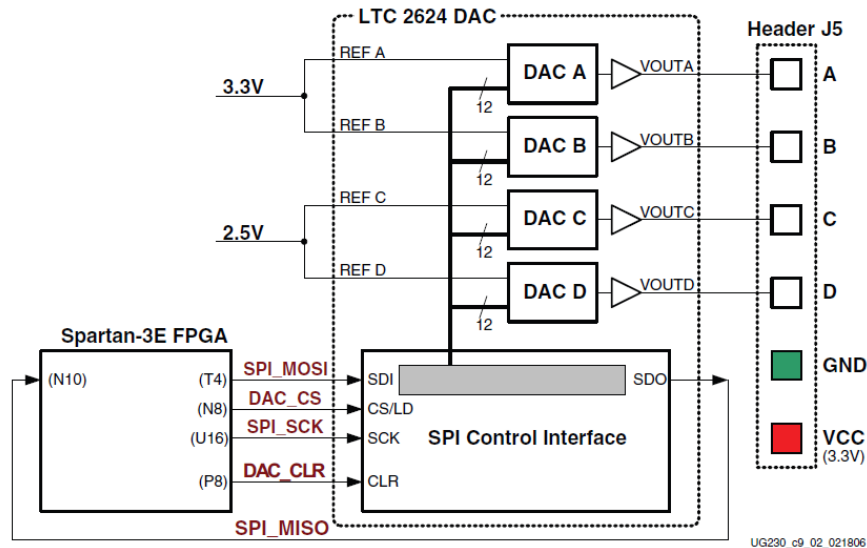


Figura 3 Vista detallada del circuito de conversión digital a análogo.

El voltaje de salida del DAC se describe en la ecuación 3. El voltaje de referencia es diferente en las 4 salidas del DAC. Los canales A y B usan un voltaje de referencia de 3.3 V, mientras que los canales C y D usan un voltaje de 2.5 V de referencia. En este caso se utilizó la salida A, por lo que el voltaje de referencia es de 3.3 V.

$$V_{OUT} = \frac{D[11:0]}{4096} \times V_{REFERENCIA} \quad (3)$$

El principal reto de esta aplicación es el uso de la interfaz SPI, ya que sólo se puede utilizar en un dispositivo a la vez: PGA, ADC o el DAC, por lo que fue

necesario crear una máquina de estados finitos (FSM) para el manejo total del sistema, con la cual fuera utilizada esta interfaz de la mejor manera posible. El diagrama de estados de la FSM se muestra en la figura 4.

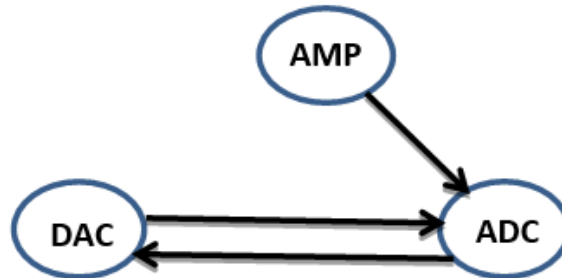


Figura 4 Diagrama de estados del sistema.

En el sistema se utilizó como entrada la señal de reloj de 50 MHz incluida en la tarjeta de desarrollo y a través del uso del administrador digital de reloj (DCM), se generó una señal de reloj de 80 MHz, obteniendo así una velocidad de 40 MHz en la operación de la interfaz SPI. Además, se generó otra señal de reloj de 2 MHz para el manejo del PGA, el cual sólo se configura una sola vez, antes que empiecen a trabajar el ADC y el DAC. El diagrama a bloques de la entidad del sistema se muestra en la figura 5.

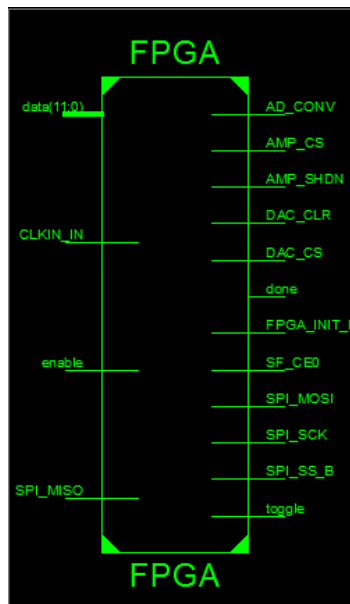


Figura 5 Diagrama de la Entidad.

Cada elemento del sistema (PGA, ADC, DAC) requirió de una FSM para su manejo particular. Para el caso del amplificador fueron necesarios 6 estados como se muestra en la figura 6. Estos estados se basan en el protocolo de la interfaz de comunicación serial SPI tal cual se muestra en la figura 7.

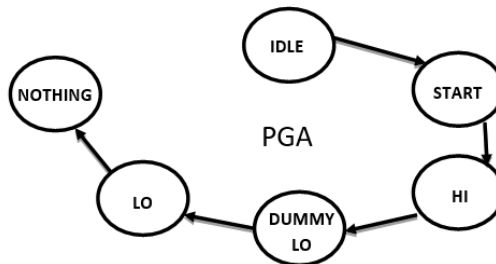


Figura 6 Diagrama de estados para el amplificador.

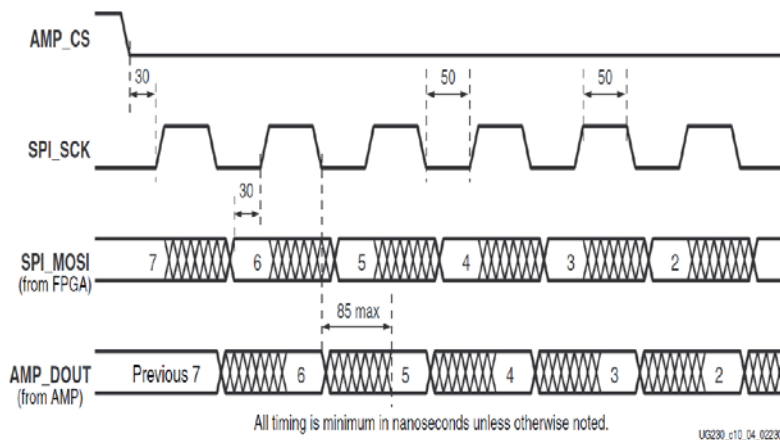


Figura 7 Comunicación SPI con el Amplificador.

En el diagrama de estados del PGA, figura 6, el primer estado, IDLE, se mantiene hasta que se active el bit de habilitación. Una vez que esto sucede pasa al estado START, donde se pone en cero la señal amp_cs, se inicia una cuenta en cero y se hace un retardo para posteriormente pasar al estado HI.

En este estado, se manda el bit más significativo del valor de la ganancia del amplificador, se pone en 1 la señal del reloj SPI, aumenta en uno la cuenta para después pasar al estado de DUMMY_LO.

Aquí, se evalúa que la cuenta no haya llegado a 8, número de bits a enviar, si no ha llegado, pone en 1 el reloj SPI y se pasa al estado LO. En este estado se

selecciona el bit siguiente a enviar, se pone en cero el reloj SPI y se regresa al estado HI donde esta vez se enviará el bit seleccionado.

Cuando la cuenta llega a 8, se han enviado los ocho bits del amplificador, por lo que la señal amp_cs se vuelve a poner en 1, se pasa al estado NOTHING y de ahí se pasa a la máquina de estados del ADC.

En la figura 8 se muestran los 6 estados requeridos para el manejo del ADC. En el estado IDLE_ADC se verifica que el sistema esté habilitado, lo cual se hace a través de un interruptor deslizable de la tarjeta.

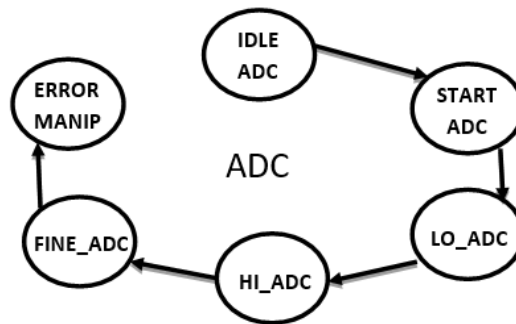


Figura 8 Diagrama de estados para el ADC.

Si no está habilitado el sistema, se regresa al estado inicial para configuración del PGA. En caso contrario se manda al estado de inicio START_ADC además de poner en alto AD_CONV que comienza simultáneamente la lectura de los dos canales del ADC.

En el estado START_ADC se pone en bajo AD_CONV, cambiando así al estado LO_ADC donde se verifica el valor de count para disminuirlo en uno, aumentar counter en 1, y cambiar al estado HI_ADC en el cual se monitorea por medio de rangos que el counter dure 34 ciclos de reloj para que se deje la señal SPI_MISO en alta impedancia, evitando el bloqueo de la comunicación SPI para los demás periféricos, y cambiando al estado FINE_ADC cuando el ADC ha terminado dando paso al estado siguiente el cual es ERROR_MANIPULATE donde se hace la linealización de la lectura, terminando así la máquina de estado del ADC, para luego continuar con la del DAC.

Para realizar la linealización de la lectura del ADC, se tomó en cuenta que la conversión análoga-digital se representa en un número de 14 bits y la salida del

DAC es de 12 bits. Se optó por descartar los valores de los 2 bits menos significativos del valor del ADC. Para linealizar la lectura se realizó el complemento a uno de los 11 bits menos significativos y manteniendo el valor del bit más significativo. En la figura 9 se muestra la salida normal del ADC y en la figura 10 se muestra ya linealizada.

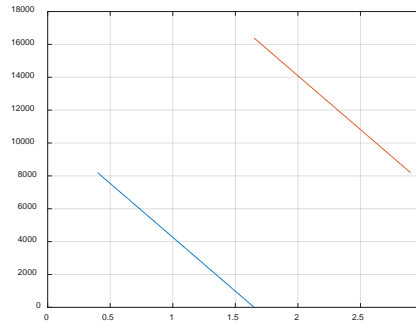


Figura 9 Lectura del ADC sin linealizar.

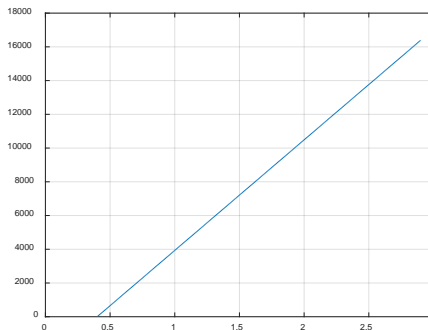


Figura 10 Lectura del ADC ya linealizada.

De igual manera, para el manejo del DAC, se utilizaron 6 estados como se muestra en la figura 11.

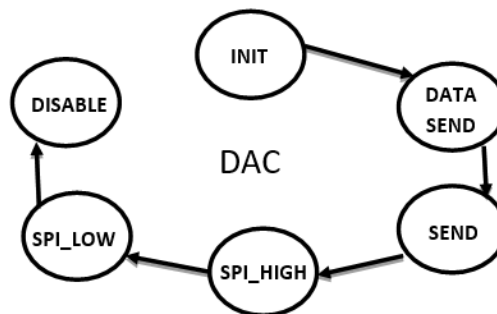


Figura 11 Diagrama de estados para el DAC.

La máquina de estados finitos para el DAC arranca con el estado INIT en el cual se inicia la conversión de datos al poner en alto DAC_CS y se asigna la longitud del dato a enviar, pasando así al estado DATA_SEND que recibe el vector arrojado por el último estado del ADC y este conjunto de bits se concatena con los bits de comando y selección del DAC, continuando al estado SEND, donde por medio de la interfaz SPI se inicia el envío del dato anteriormente mencionado, realizando el envío por medio de los estados SPI_HIGH y SPI_LOW que controlan la señal de reloj de la interfaz, siendo el último estado mencionado, el de DISABLE donde al finalizar el envío de datos se regresa al inicio de la máquina de estados del ADC, al estado IDLE_ADC, mientras que en la máquina del DAC se regresa al estado INIT.

3. Resultados

Una vez terminada la reconfiguración del FPGA, se realizaron las pruebas correspondientes para comprobar el correcto funcionamiento del sistema. Por lo tanto, se utilizó un generador de funciones, un osciloscopio y la tarjeta Spartan-3E que contiene el FPGA ya reconfigurado.

Como patrón de entrada se utilizó el generador de funciones para aplicar la señal analógica a procesar, una señal senoidal en este caso, se ajustó a una frecuencia baja de 100 Hz, de tal forma que se pudiera observar que la señal de salida corresponda a la entrada, dado que el procesamiento fue multiplicar la señal de entrada por 1, como se puede observar en la figura 12.

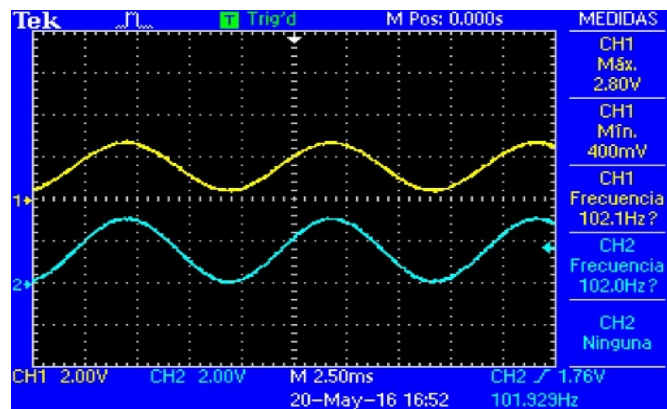


Figura 12 Respuesta del sistema a frecuencias bajas.

En figuras 13 y 14 se puede observar, además de la señal de entrada y la de salida, la señal que representa la activación del ADC que indica la frecuencia de muestreo.

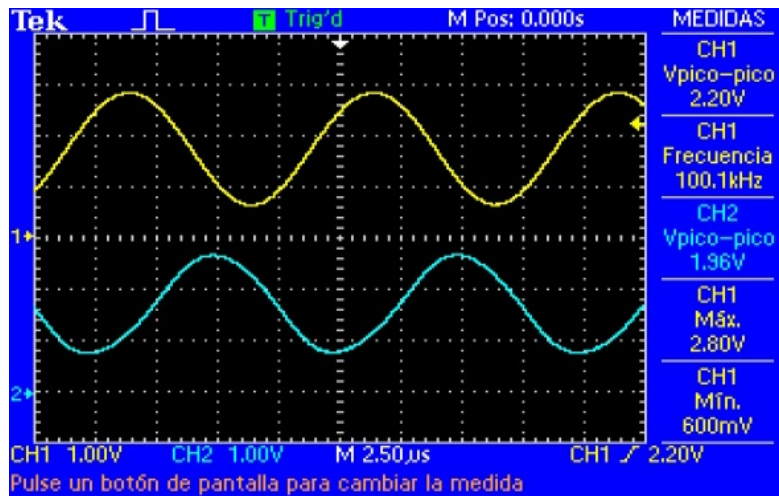


Figura 13 Sistema funcionando a 100 kHz.

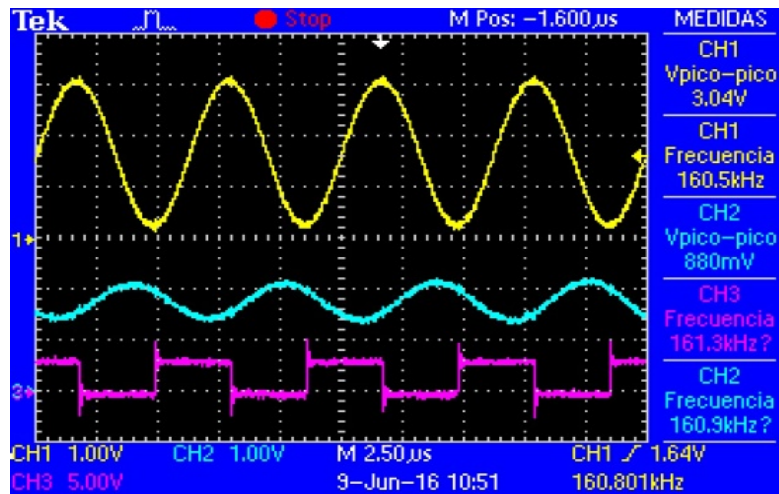


Figura 14 Sistema funcionando a 161 kHz.

4. Discusión

Se observó el rango de valores a los cuales el sistema mantuvo su funcionamiento, tras variar la frecuencia de la señal de entrada obteniendo el ancho de banda máximo soportado por el sistema, el cual fue 161 kHz.

Asimismo, tras aumentar la frecuencia de entrada más allá de la soportada, se generó el fenómeno de “aliasing” como se puede ver en la figura 15, donde se aplica la señal de entrada con una frecuencia de 322.79 kHz y se obtiene a la salida una señal senoidal pero de una frecuencia muy baja, 37.99 Hz, como consecuencia de no cumplir con el teorema de muestreo.

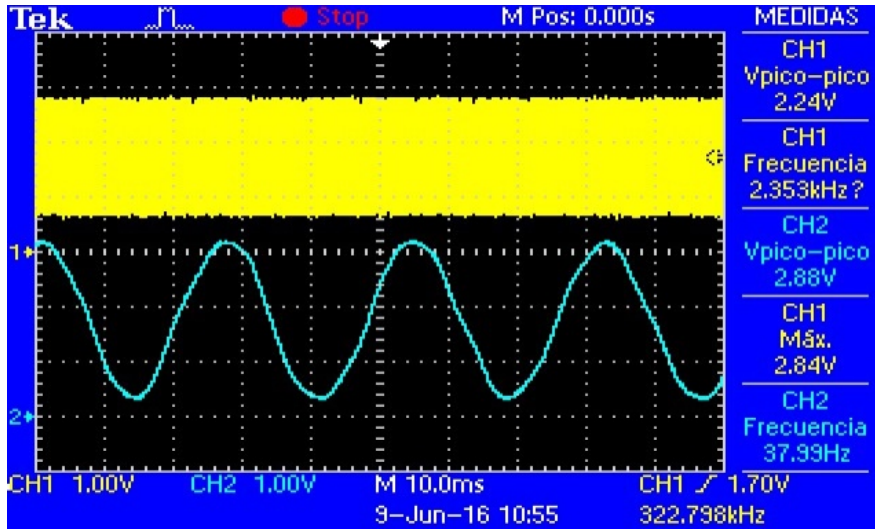


Figura 15 Sistema funcionando a 322 kHz con salida “alias” de 37.99 Hz.

En la figura 16 se aplica una frecuencia de entrada de 641 kHz y se obtiene un alias de 1.48 kHz.

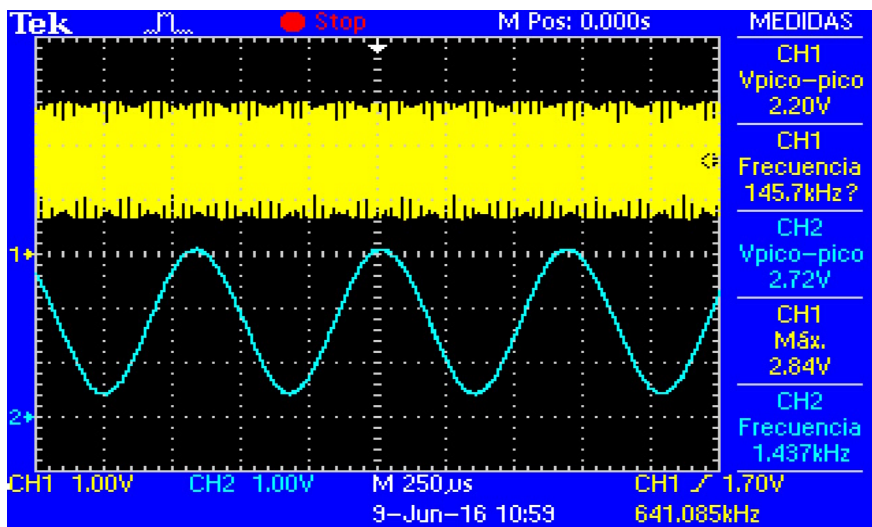


Figura 16 Sistema funcionando a 641 kHz con salida “alias” de 1.43 kHz.

Al inicio, el amplificador se utilizaba dentro del ciclo de captura y reproducción de la señal análoga, pero esto hacía que el ancho de banda se limitara a 10 kHz. Por lo que posteriormente se modificó para que solo se usara una sola vez, también se logró realizar un aumento al hacer uso del protocolo de 24 bits en lugar del de 32 bits que maneja el convertidor digital análogo (DAC). El sintetizador digital de frecuencia se programó a 80 MHz para una frecuencia de operación de la interfaz SPI de 40 MHz.

Gracias a la acción de iniciar el amplificador únicamente al arrancar el sistema y de utilizar el protocolo de 24 bits para el DAC, se obtiene un mayor ancho de banda.

5. Conclusiones

Se logró obtener el máximo ancho de banda permitido por el sistema para el procesamiento de señales hasta una frecuencia de 161 kHz, así como una razón de muestreo de 322 kmuestras/segundo a una velocidad de operación de 40 MHz de la interfaz SPI.

Esto es un aumento de 14% en el ancho de banda con respecto a [Silage, 2008] y a una velocidad 4% menor. La forma de manejo del amplificador y el uso del protocolo de 24 bits del DAC fueron la clave para lograr estos resultados. El manejo del amplificador de ganancia programable no es determinante para el aumento de la frecuencia de operación del sistema ya que sólo se accede a él en una sola ocasión al arranque del sistema.

Un trabajo futuro contemplaría el uso de una señal de reloj externa, en lugar de la de 50 MHz disponible en la tarjeta, con una frecuencia tal que se lleve al límite la velocidad de operación de la interfaz SPI que es de 50 MHz.

6. Bibliografía y Referencias

- [1] Domínguez, I. & Rodríguez, J. Procesador digital sincrónico en tiempo real soportado sobre un circuito FPGA Spartan-3E. *Revista de ciencia y tecnología*, enero/junio, 2011.
- [2] Linear Technology Corporation. LTC2624 DAC Datasheet, 2004.

- [3] Khedr, H.I., & Mostafa, A.G., & Radi, A., & Zidan, W.I. Controlling of Analog Capture Circuit and Digital Analog Converter for Spartan-3E FPGA Starter Kit. *Nature and Science*, 177-182, 2013.
- [4] Mascharak, S. Implementation of the onboard ADC and DAC on the Spartan 3E FPGA platform. Tesis. 2012.
- [5] Silage, Dennis. DSP on the Xilinx Spartan-3E Starter Board, 2008: <http://astro.temple.edu/~silage/pl-edpga.pdf>, Consultado Mayo de 2016.
- [6] Silage, Dennis. Embedded design using programmable gate arrays. Bookstand Publishing, 2008.
- [7] Xilinx Inc. Spartan-3E FPGA Starter Kit Board User Guide, 2011.

MÉTODO DE INSTRUMENTACIÓN INDIRECTA BASADO EN ONDAS ACÚSTICAS DERIVADAS DE VIBRACIONES MECÁNICAS PARA LA ESTIMACIÓN DE VELOCIDAD ANGULAR EN MAQUINARIA ROTATIVA

Enrique Gerardo Hernández Vega

Tecnológico Nacional de México/Instituto Tecnológico de Chihuahua

ehernand@itchihuahua.edu.mx

Sergio Iván Chavaría Estrada

Tecnológico Nacional de México/Instituto Tecnológico de Chihuahua

sergio.i.chavarria@gmail.com

Resumen

El análisis de vibraciones es fundamental en el diagnóstico de maquinaria. Al hacer un análisis de vibraciones, buscamos formas para impedirlos o minimizar sus efectos. Las vibraciones tienen información valiosa sobre el comportamiento de un sistema y hay más de una forma de aprovechar esta información.

Este trabajo explora una manera alternativa de utilizar las vibraciones producidas por un motor eléctrico, para determinar su velocidad angular. Asumiendo una relación directa en la frecuencia fundamental de las vibraciones con la velocidad del motor y utilizando herramientas matemáticas para el análisis de señales, se desarrolla un método para la estimación de la velocidad de un motor.

En los casos explorados, 3 motores eléctricos, los valores de estimación obtenidos mantuvieron una correlación casi perfectamente lineal con una desviación estándar del error no mayor al 1%.

Palabras Claves: Acondicionamiento y procesamiento de señales, detección de tono, frecuencia fundamental, instrumentación acústica, vibraciones mecánicas.

Abstract

Vibration analysis is fundamental in machinery diagnosis. When doing a vibration analysis, we look for forms to minimize their effects. Vibrations have

valuable information regarding the behavior of a system, and there is more than one way to harness this information.

This work explores an alternative way of utilizing the vibrations produced by an electrical motor, to determine its angular speed. Assuming there is a direct relation between the fundamental frequency of the vibrations and the motor's speed, and using mathematical tools for signal processing, develops a method for estimating the speed of a motor.

In the explored cases, 3 electrical motors, the obtained values maintained an almost perfectly linear correlation with a standard deviation of the error no greater than 1%.

Keywords: *Acoustic instrumentation, fundamental frequency, mechanical vibrations, pitch detection, signal conditioning and processing.*

1. Introducción

De la manera más simple se define a la vibración como un movimiento oscilatorio de pequeña amplitud. Los términos vibración y oscilación suelen usarse indiscriminadamente, aunque algunos autores distinguen a la oscilación como un movimiento periódico de baja frecuencia con una amplitud perceptible y regular. Por otro lado el comportamiento de las vibraciones suele ser irregular, incluso aleatorio y difícil de percibir a simple vista [de Silva, 2000]. Se define entonces a las vibraciones mecánicas como el movimiento periódico o armónico de masas, o como la respuesta oscilatoria, repetitiva o periódica de un sistema mecánico.

La respuesta dinámica de un sistema a una acción o estímulo puede generar vibraciones de forma natural, por ejemplo, la respuesta de percusión de un tambor cuya frecuencia natural dependerá solamente de su estructura y no del estímulo aplicado.

Un sistema también puede sufrir vibraciones forzadas, las cuales pueden ser iguales o completamente diferentes a la frecuencia natural del sistema, generadas por el mismo sistema o de manera externa, por ejemplo el eje de un motor al rotar produce vibraciones en todo el cuerpo del motor que coinciden con la frecuencia de rotación. Al igual que la respuesta al escalón o al impulso, la respuesta de un

sistema a un estímulo oscilatorio consta de su componente transitoria y su componente estable [Stoker, 1961] como se muestra en la figura 1.

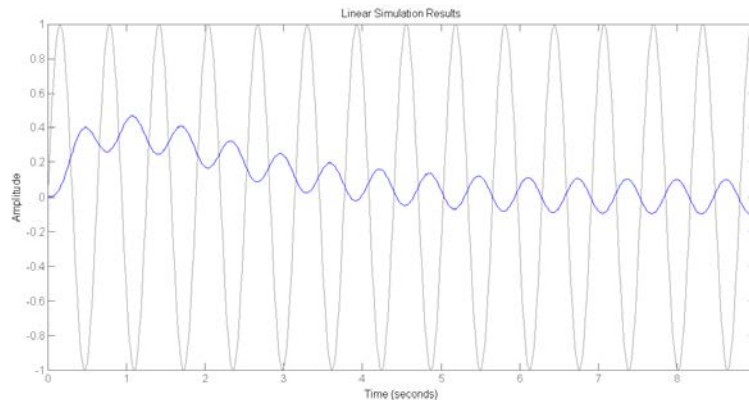


Figura 1 Ejemplo de respuesta transitoria y estable a una señal sinusoidal.

Ya sean naturales o forzadas, las vibraciones pueden ser deseables o indeseables según la naturaleza y objetivo del sistema. Ciertamente, los instrumentos musicales son un caso en el cual las vibraciones son deseables, pero la existencia de vibraciones en sistemas mecánicos puede llegar a ser un fenómeno perjudicial, tanto para la estabilidad del sistema como para su integridad estructural. Ignorar la ubicuidad de las vibraciones en todo sistema mecánico sería entonces un error que puede llevar a fallas imprevistas o incluso consecuencias fatales [Park, 2003]. Actualmente la manera más común y práctica de medir vibraciones mecánicas es transformarlas primero en señales eléctricas. Un sensor de vibraciones es necesario para esta tarea. Un sensor nos proporciona información, normalmente en forma de una señal eléctrica que regularmente debe ser acondicionada para ser procesada de manera digital, lo cual nos permite hacer manipulaciones matemáticas inmensamente complejas en fracciones de segundos [Park, 2003]. De poco sirve capturar y procesar una señal si no se le da una interpretación objetiva, pues dicha señal contiene información que puede ser observada e interpretada por un humano, o por una computadora.

Típicamente se utilizan sensores de movimiento para la detección y medición de vibraciones. Los sensores de movimiento más comunes son el piezoeléctrico, inductivo y capacitivo. Cabe mencionar que el sensor inductivo solía ser el más

popular para la medición de vibraciones, hasta que acelerómetro piezoeléctrico tomó su lugar debido a su respuesta dinámica mejorada y su construcción económica. Sin embargo, el sensor capacitivo no se queda atrás en popularidad, pues es muy utilizado en aplicaciones de acústica.

La naturaleza de una señal de vibraciones es analógica, es decir, es una señal con una infinidad de valores espaciados a intervalos infinitesimalmente pequeños. Para poder procesar la información de una señal de forma digital, es necesario discretizar dicha señal en intervalos finitos para obtener un arreglo de muestras numéricas [Shannon, 1948], [Smith, 1999], ver figura 2.

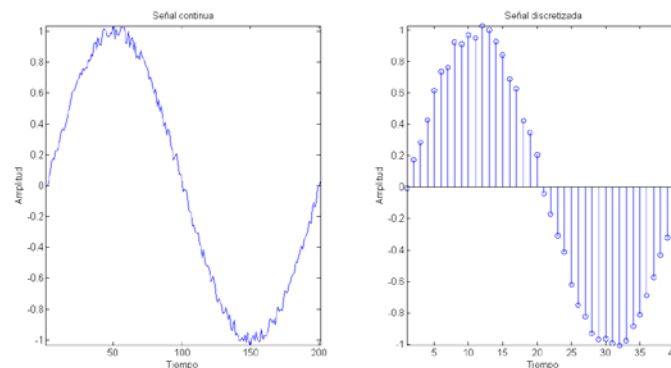


Figura 2 Discretización de una señal con componentes de ruido aleatorio.

Una señal vibratoria normalmente consiste en una superposición de ruido aleatorio, componentes cíclicas y sus armónicas. Mucha de la información contenida en una señal vibratoria es difícil de discernir a simple vista. Es por eso que se recurre a herramientas matemáticas para transformar estas señales que por naturaleza se encuentran en el dominio del tiempo a otros dominios como el de la Frecuencia. La Transformada de Fourier nos permite observar el espectro de frecuencia de una señal continua o discreta, de manera que se abren las puertas a un análisis que sería muy complicado en el dominio del tiempo [Smith, 1999].

2. Métodos

La premisa de este trabajo es la idea de que se pueden aprovechar las inevitables vibraciones mecánicas que produce un motor para extraer información acerca de su velocidad. ¿Qué implicaciones tiene esta declaración? En primer

lugar, se puede construir un tacómetro basado solamente en un sensor de vibración.

La manera más común de medir la velocidad de un motor es con uno o varios puntos de referencia en el rotor o flecha. Este punto, o puntos de referencia son detectados por algún sensor en cada revolución que da el motor. Con la ayuda de un sistema electrónico es posible contar la cantidad de ocasiones que se detecta este punto de referencia con respecto a un intervalo de tiempo. Normalmente en forma de pulsos eléctricos, solo es necesario contar estos pulsos en un intervalo de tiempo para determinar la velocidad, aceleración o posición del motor.

¿Qué ventajas ofrece un tacómetro basado en vibraciones mecánicas? Esta propuesta ofrece un método alternativo, no intrusivo, para la medición de velocidad de un motor de manera que no sea necesario acceder al mecanismo del motor, ni agregar cargas mecánicas. La propuesta a continuación es entonces un método para detección de velocidad para un motor basado en un sensor acústico. Dicho en otras palabras, utilizando algo tan sencillo como un micrófono es posible determinar la velocidad de un motor.

Es difícil ignorar el ruido acústico que produce un sistema vibratorio. ¿Sería ingenuo asumir la existencia de una relación perfecta entre ruido acústico y vibración mecánica? Bajo una perspectiva diferente, podríamos definir a las vibraciones mecánicas como ruido acústico que se propaga en un sólido.

La relación de velocidad de un motor y el ruido acústico que produce no es aparente a simple vista. Podría pensarse que solamente es cuestión de detectar el componente de frecuencia con mayor amplitud para conocer la velocidad de motor. Mientras que puede ser cierto para un sensor puramente vibratorio, acústicamente existen muchos factores diferentes que producen un sonido irregular, donde en la mayoría de los casos el componente de frecuencia de mayor amplitud no es la velocidad del motor.

La figura 3 muestra un ejemplo de la señal de ruido generada por un pequeño motor de corriente directa, así como el espectro de frecuencia correspondiente.

Se puede apreciar un patrón en el espectro de frecuencia, donde los picos son equidistantes uno del otro, lo que refleja una relación armónica. Este punto es de

suma importancia ya que confirma el fundamento teórico sobre la frecuencia fundamental de una señal, donde dicha frecuencia fundamental equivale a la velocidad angular del motor. Sin embargo, aún es necesario confirmar la consistencia de este patrón con un instrumento de medición de velocidad.

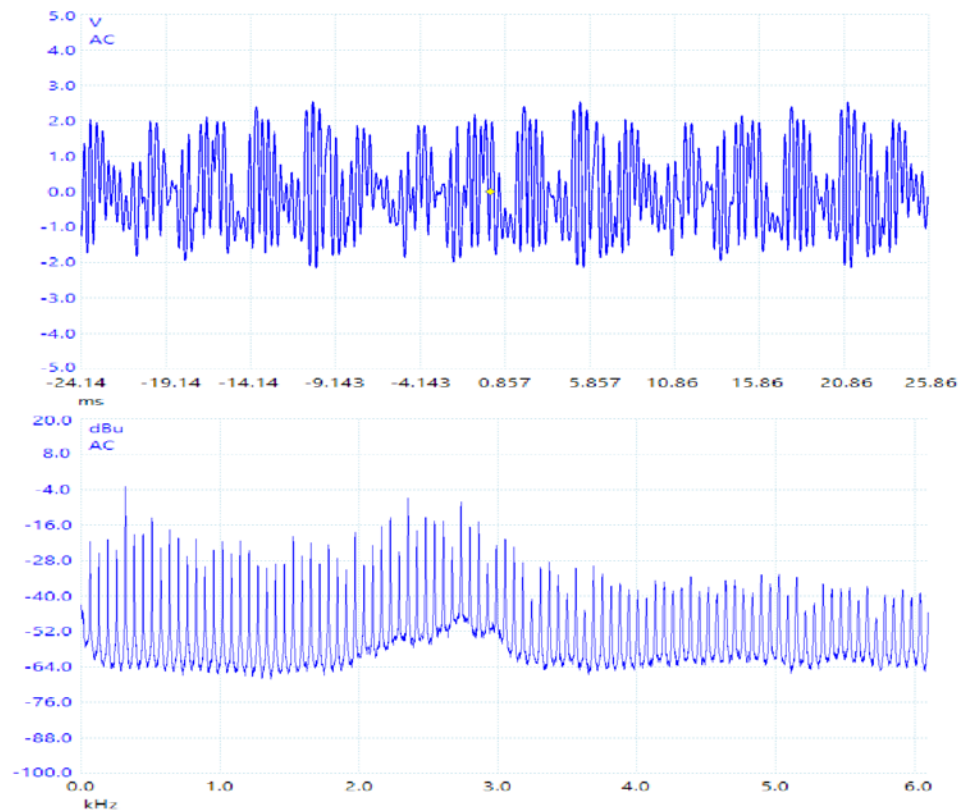


Figura 3 Sonido generado por un motor de corriente directa y su espectro de frecuencia.

Entonces, si se caracteriza el sonido producido por un motor de manera que consista de una frecuencia fundamental equivalente a su velocidad angular, más sus componentes armónicos, es posible determinar la velocidad angular de un motor midiendo las vibraciones o el sonido producido por éstas. Una buena regla general para la detección de la frecuencia fundamental en una señal es mantener una cantidad significativa de armónicos, pues el filtrar excesivamente es prominente a errores, aunque se permita la frecuencia fundamental.

El objetivo básico es extraer la frecuencia fundamental F_0 de una señal de sonido, la cual normalmente es la componente, o parcial, de menor frecuencia, la cual

suele estar relacionada con las parciales mayores. En una señal periódica, la mayoría de las parciales están relacionadas armónicamente, es decir, son múltiplos enteros de dicha frecuencia fundamental.

Existen diversos métodos [Middleton, 2003] para la estimación de la frecuencia fundamental F_0 , siendo cada método útil en diferentes contextos. Muchos de estos métodos toman un enfoque similar, aprovechando la naturaleza periódica de una señal es posible entonces determinar con cierto grado de certeza la frecuencia fundamental. Se debe asumir la periodicidad de la señal, ya que, de lo contrario, cualquier método será propenso a errores. La mayor parte de estos métodos funcionan adecuadamente al ser presentados con una señal periódica limpia, pero cuando la señal es ruidosa, o está compuesta de múltiples tonos, muchos métodos actuales pueden fallar inesperadamente. En la literatura se suele llamar Detección de Tono a un método de Estimación de la Frecuencia Fundamental.

La Correlación es una medida de similitud entre dos señales. La Autocorrelación de una señal es la correlación de dicha señal consigo misma. Similar a la Convolución, es la sumatoria sucesiva de la multiplicación de dos señales, donde para cada valor de la Autocorrelación se desplazan en el tiempo las muestras de una de las señales. En el dominio del tiempo, la Autocorrelación se define por ecuación 1.

$$R_{xx}(n) = \sum_{k=0}^{N-1} x(k)x(k+n) \quad (1)$$

La representación simplificada de la Autocorrelación por ecuación 2.

$$R_{xx}(n) = x(n) * x(n) \quad (2)$$

Donde $*$ es el operador de correlación cruzada.

La Autocorrelación puede ser representada como la Convolución de una señal con su complejo conjugado invertido en el tiempo, ecuación 3.

$$R_{xx}(n) = x(n) * x^*(-n) \quad (3)$$

Donde $*$ es el operador de Convolución.

La Autocorrelación es una operación matemática que permite determinar la periodicidad de una señal, si esta existe. En el proceso del producto por el

conjugado se pierde la información de fase de la señal, dejando solamente información sobre el periodo. Esto resulta muy útil en la estimación de la frecuencia fundamental de una señal. Una ventaja es que el ruido aleatorio es eliminado, asumiendo que existe una correlación nula con la señal.

La figura 4 muestra el resultado de la Autocorrelación de la señal de sonido producida por el motor de la figura 3. El resultado de la Autocorrelación de una señal periódica es en realidad también una señal periódica. Mientras que sí se obtiene el periodo de la frecuencia fundamental, es difícil discernir entre los picos más prominentes. El resultado se encuentra en el dominio del tiempo, aunque no exactamente, pues puede verse como la representación del retraso en tiempo de una señal periódica en lugar del valor absoluto de un valor en un momento determinado.

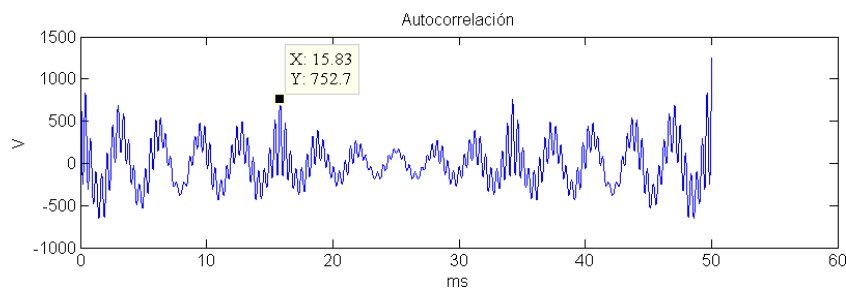


Figura 4 Ejemplo de Autocorrelación.

En [Tolonen, 2000] se describe un algoritmo eficiente para la estimación de múltiples tonos fundamentales en una señal, denominado como ESCAF por Enhanced Summary Autocorrelation Function, derivado del trabajo de [Meddis, 1997], basado en modelos de percepción auditiva humana.

Una implementación parcial [Mazzoni, 2000] del algoritmo ESCAF se encuentra en el software de composición y análisis de audio Audacity [Audacity, 2000], donde se omite la sumatoria de la disección de señales y se aplica el método de rectificación de media onda y sustracción a la autocorrelación de la señal obteniendo resultados muy prometedores.

El diagrama de flujo en la figura 5 muestra a grandes rasgos los pasos a seguir en el algoritmo de Autocorrelación Mejorada. El algoritmo comienza con la FFT de la

señal capturada, para luego determinar la Autocorrelación Generalizada, la cual se define por ecuación 4.

$$r_{xx}(\tau) = F^{-1} [|F[x(n)]|^k] \quad (4)$$

Donde los operadores F y F^{-1} son la *Transformada Discreta de Fourier* y la *Transformada Inversa de Fourier*, respectivamente. En lugar de multiplicar por el conjugado, se eleva el espectro a la potencia k , el cual es un parámetro que determina la compresión en el dominio de la frecuencia. La Autocorrelación Estándar utiliza un valor de $k = 2$ lo que es equivalente al producto por el conjugado.

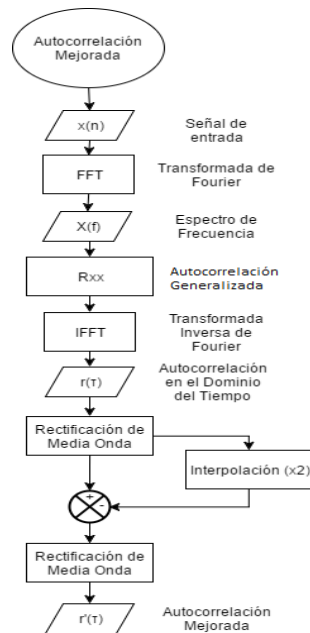


Figura 5 Diagrama de flujo del algoritmo de Autocorrelación Mejorada.

Se recomienda una $k=2/3$ según [Tolonen, 2000], lo que es un buen compromiso entre la sensibilidad al ruido para valores pequeños de k y baja resolución en los picos detectados para valores mayores. El nombre de Autocorrelación Mejorada viene del siguiente paso, el cual tiene el propósito de eliminar los armónicos redundantes en el espectro de la Autocorrelación, siendo estos múltiplos enteros de la o las frecuencias fundamentales. Primero se cortan todos los valores

negativos, rectificación de media onda, igualándolos a cero, se escala en el tiempo en un factor de dos, el resultado se sustrae a la función original con los valores negativos cortados, y por último se remueven de nuevo todos los valores negativos igualándolos a cero. Esto remueve picos repetidos con el doble de tiempo de retraso donde la amplitud del pico básico es mayor a la amplitud del duplicado. También se remueven los retrasos cercanos a cero, los cuales son una consecuencia colateral del algoritmo de Autocorrelación. Esta operación puede repetirse un número de veces con un escalamiento de tiempo de tres, cuatro, cinco, etc., hasta donde se desee, de manera que se eliminen múltiplos mayores de cada pico.

La estimación del tono fundamental es entonces determinada por el pico de mayor amplitud, siendo en el caso de la figura 6 equivalente a un retraso en el tiempo de 15.83 ms, y el recíproco igual a 63.167 Hz.

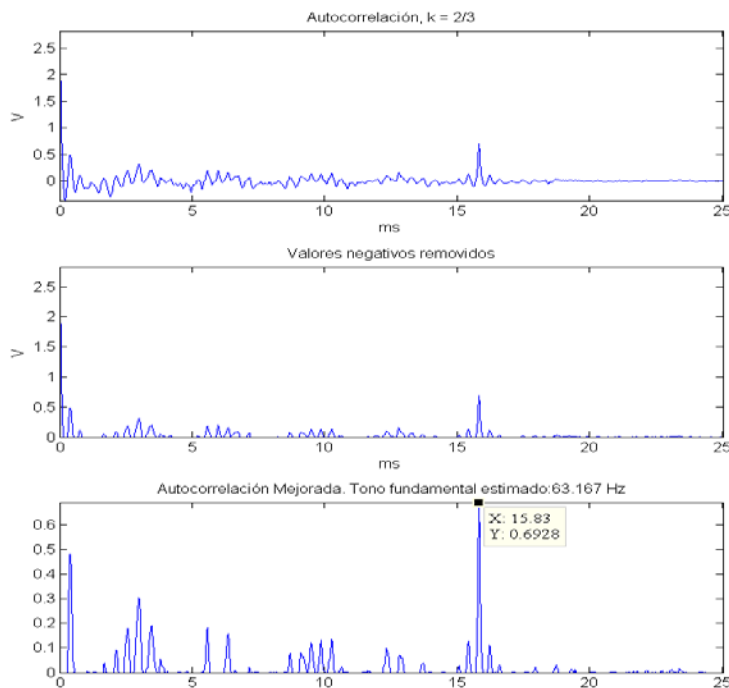


Figura 6 Autocorrelación Mejorada.

La información obtenida del resultado de aplicar la Autocorrelación Mejorada nos da información sobre los armónicos fundamentales más prominentes en una señal. Se puede asumir entonces, con un cierto grado de precisión e incertidumbre

que los componentes de mayor amplitud corresponden directamente a los modos principales de vibración y, aplicando este tipo de análisis a las vibraciones producidas por un motor, se puede determinar si existe una relación directa entre dichos componentes y la velocidad angular del motor.

La estimación de velocidad angular consiste entonces en determinar cuál es la componente de mayor amplitud en el resultado de la Autocorrelación Mejorada. La información que proporciona la Autocorrelación está en el dominio del tiempo y puede representarse como retrasos temporales en la señal. Aplicando el recíproco al valor de tiempo correspondiente a la componente de mayor amplitud nos revela la frecuencia equivalente. Esta frecuencia representará la frecuencia fundamental del motor, asumiendo que el motor es la única o la mayor fuente de ruido acústico capturado.

El micrófono utilizado es del tipo electreto, el cual debe conectarse a un circuito de polarización, el cual se muestra en la figura 7.

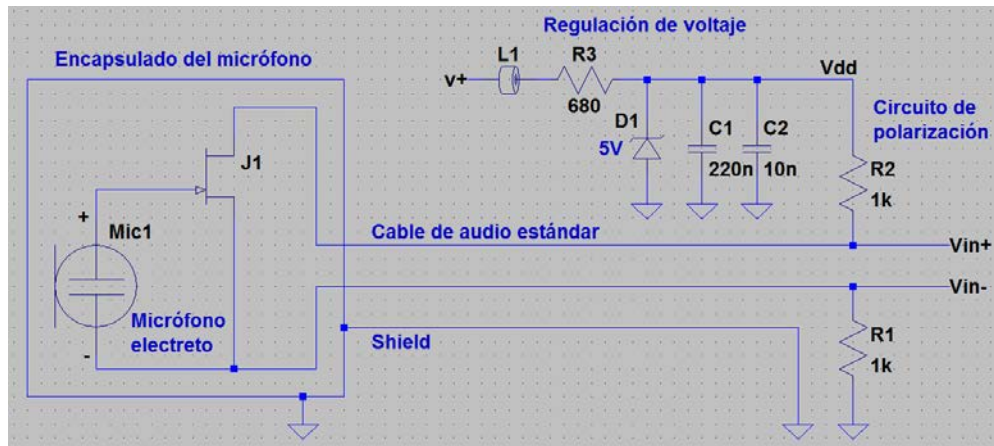


Figura 7 Circuito de polarización del micrófono electreto.

El circuito de acondicionamiento está basado en amplificadores operacionales, constando de un amplificador de instrumentación, un filtro pasa bajas de segundo orden y un filtro pasa altas de cuarto orden, figura 8.

El proceso experimental consistió de los siguientes pasos:

- Realizar las conexiones eléctricas al sistema de adquisición de datos tanto del micrófono de captura, como del optoacoplador, para la comparación de ambos resultados.

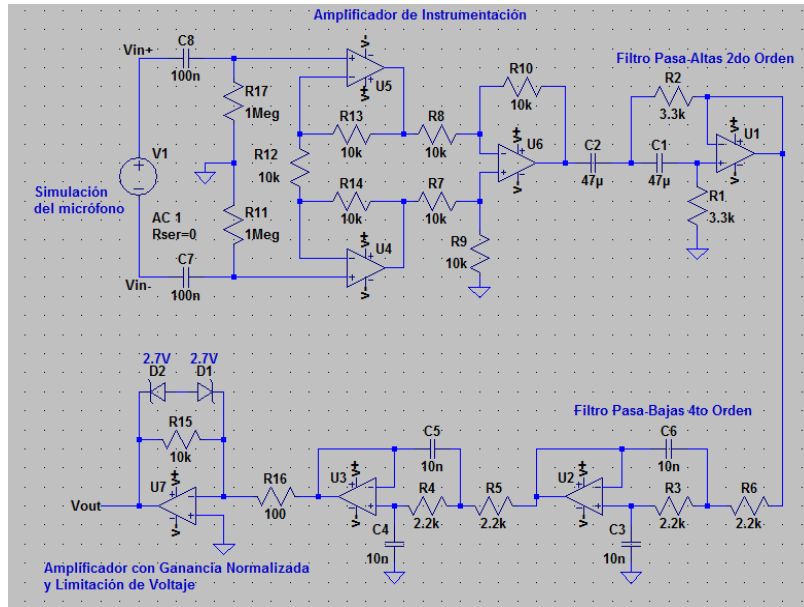


Figura 8 Circuito de Acondicionamiento de la señal del micrófono.

- Se automatizó el proceso de captura usando el Osciloscopio USB picoScope 2204A [Pico Technology, 2017] y aprovechando el API de Matlab que provee Pico Technology. De esta manera se capturan de forma paralela las señales entregadas por el micrófono y el optoacoplador.
- Las señales son procesadas por un script de Matlab. Primero se determina la frecuencia de la señal del optoacoplador y se ajusta a un factor equivalente al número de ranuras que tiene el detector de cuadratura. Por otro lado, se utiliza el algoritmo de estimación de frecuencia fundamental en la señal del micrófono.
- Acumulando 1000 resultados, estos son exportados a una hoja electrónica de datos para caracterizar la relación que existe entre los valores entregados por el optoacoplador y el algoritmo de detección de tono.

En la figura 9 se muestra el diagrama a bloques del sistema.

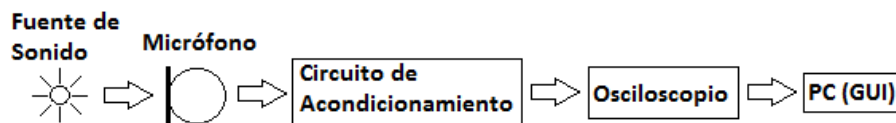


Figura 9 Diagrama del sistema.

3. Resultados

Se utilizaron tres motores de prueba y se observaron resultados muy similares, y una clara relación directa entre la frecuencia fundamental de las vibraciones con la velocidad angular de cada motor. En cada caso, el error de medición se mantuvo dentro de un rango casi constante, con muy poca desviación de la media. Debido a cuestiones de espacio y al hecho de que en 2 motores el experimento fue casi totalmente automatizado, sólo se muestran los resultados para un solo motor. Además, los resultados de este motor fueron los más interesantes por así decirlo. El motor eléctrico es parte del kit didáctico 1405B de la corporación Nida en la tarjeta experimental 130A-255, figura 10, [Nida, 1996]. El tema de la tarjeta es la experimentación con sensores fotoeléctricos para medición de movimiento angular en un motor. La tarjeta cuenta con un optoacoplador con interruptor de haz, y un sensor reflector.

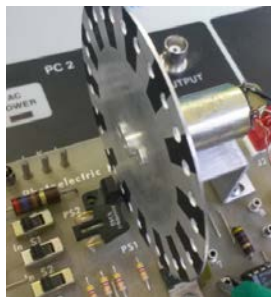


Figura 10 Motor de DC con detector de cuadratura por reflexión e interrupción de luz.

Para el experimento de comparación con la frecuencia fundamental de vibraciones, se utilizó la salida del optoacoplador y se calculó la frecuencia de la señal de salida dividida entre 16, ya que el disco cuenta con 32 orificios. Los resultados de 1000 comparaciones se muestran en la gráfica de la figura 11a, los cuales son filtrados por el método de Theil-Sen [Wikipedia, 2017], el cual consiste en determinar todas las pendientes generadas por todos los pares de datos posibles, para luego seleccionar aquella que se encuentre en la media, de esta manera pueden eliminarse todos los valores atípicos, en este caso manteniendo 828 datos dentro de la tolerancia de 30% en la tendencia lineal principal, ver figura 11b.

También se muestra la distribución del error en la figura 11d. El resumen de resultados se muestra en la tabla 1.

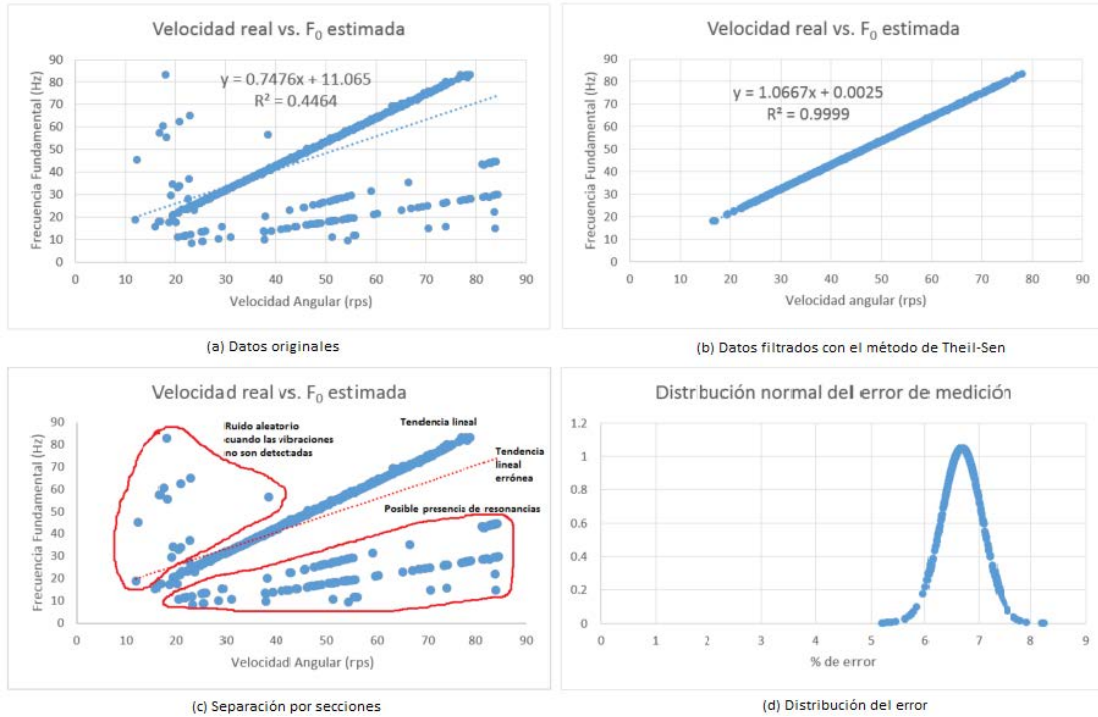


Figura 11 Resultados del experimento.

Tabla 1 Parámetros de captura

Parámetros de captura	
Frecuencia de muestreo F_s	19.6078 kHz
Número de muestras N	$2^{16} = 65536$
Ventana Utilizada	Hamming
Parámetros de comparación	
Motor de prueba	Nida 130A-255
Instrumento de medición de velocidad	Optoacoplador
Número de pruebas	828 filtradas de 1000
Umbral de detección	16.57 rps \Leftrightarrow 78.04 rps
Error mínimo e_{min}	5.20%
Error máximo e_{max}	8.27%
Error promedio \bar{e}	6.68%
Desviación estándar del error σ_e	0.3806 %
Regresión lineal	$y = 1.0667x + 0.0025$
Coeficiente de determinación R	0.9999

Un aspecto interesante en las mediciones de los motores es que entre más grande es el motor, el umbral de detección es mejor para bajas velocidades. Esto tiene sentido si se tiene en cuenta que la energía producida por las vibraciones del motor son una función directa de la masa total o inercia rotacional, así como de la velocidad de rotación. Entonces, la energía de vibraciones producida por dos motores de diferente tamaño, cuya velocidad es igual, no será la misma, pues las vibraciones en el motor más grande serán más prominentes. Por otro lado, un motor bien balanceado produce menos vibraciones por lo que se dificulta la detección de velocidad. Se observan discrepancias de detección cuando se utiliza una carga de inercia en la flecha. Intuitivamente, cuanto mayor es la carga, mayores son las vibraciones, por ende, mayor la sensibilidad en la detección de velocidad. Otra forma de mejorar la sensibilidad es utilizando un número grande de muestras en el análisis de frecuencia mejorando la resolución y aumentando el umbral de detección, permitiendo observar bajas frecuencias.

La colocación del sensor parece afectar enormemente la medición. Hay que tomar en cuenta la transmisibilidad de vibraciones del sistema y colocar el sensor donde sea posible maximizar la detección. Ya que el transductor es un micrófono, en ocasiones no es necesario mantener contacto mecánico directo, pues basta con sostener el sensor en el aire, y en algunas ocasiones esto puede generar mejores resultados. Sin embargo, puede ser difícil determinar el uso correcto del instrumento en cada aplicación.

A pesar de que se depende de la existencia de vibraciones mecánicas, una presencia excesiva puede ser perjudicial en las mediciones. Si la plataforma o el chasis donde esté instalado el motor sufren de vibraciones excesivas, estas vibraciones se superponen en la detección del sensor. Además, el sistema puede entrar en resonancia provocando que se detecte mayormente la frecuencia natural del sistema en lugar de la velocidad angular del motor.

Es claro que la calidad de los resultados mejora entre mayor es el número de pruebas, sin embargo, dado que los últimos dos experimentos se realizaron de forma automatizada, estos se volvieron muy propensos a la presencia de valores atípicos los cuales pueden ser removidos con relativa facilidad, sin embargo, estos

puntos no deseados pueden darnos información sobre comportamientos inesperados en el sistema bajo prueba. Este comportamiento se puede observar en las pruebas sobre el motor mediano y pequeño, pero es mucho más prominente en el motor pequeño. La figura 11c indica una separación por secciones de comportamiento, indicando la presencia de diferentes fenómenos en el experimento.

4. Discusión

Posibles implementaciones y aplicaciones quedan fuera del alcance de este trabajo ya que el objetivo principal fue el discutir solamente la metodología de una forma indirecta de detectar la velocidad de un motor en base a las vibraciones producidas por este. Solamente se discuten algunas sugerencias de trabajos futuros, como desarrollo de productos y/o aplicaciones.

El primer producto obvio que se puede crear basado en este método es un tacómetro de bolsillo, el cual serviría como una herramienta alternativa de diagnóstico para cualquier tipo de motor o sistema rotativo, creando la posibilidad de medir velocidad en motores sin sensores o en lugares difíciles de acceder directamente.

A todo esto, puede aunarse un sistema de diagnóstico de vibraciones, aplicando todo este desarrollo como un sistema interactivo, que dé información detallada al usuario, similar a muchas herramientas de diagnóstico especializadas.

La selección del número de muestras y la frecuencia de muestreo juegan un papel crítico en el proceso de estimación, afectando factores como la resolución y el umbral de detección. Si no se optimizan estos parámetros para el funcionamiento en tiempo real, se puede dificultar enormemente la aplicación de un lazo de control, ya que se requiere de una respuesta dinámica lo suficientemente rápida como para no afectar la estabilidad del sistema. Se propone la posibilidad de no mantener los parámetros de captura fijos, pues dependiendo de la respuesta del sistema, quizá sea conveniente ajustar estos parámetros dependiendo, por ejemplo, de la diferencia necesaria para detectar velocidades bajas o altas. Si se

resuelven estos problemas, teóricamente será posible aplicar este producto a un lazo de control de velocidad.

5. Conclusiones

En base a los resultados de los experimentos, se observa que existe una relación lineal entre la velocidad del motor y la frecuencia fundamental de las vibraciones producidas por éste. A pesar de la presencia de errores de medición, esta relación lineal se mantiene en los tres motores, por lo que se concluye que estos errores pueden deberse principalmente a las incertidumbres de los instrumentos, así como falta de calibración.

El algoritmo de Autocorrelación Mejorada entregó buenos resultados en general, sin embargo, no quedan descartadas posibles modificaciones, así como el uso de otros métodos para la estimación de la frecuencia fundamental.

Este método de medición de velocidad angular ofrece la ventaja ser una medición indirecta no intrusiva. El sensor de vibraciones puede colocarse entonces donde sea conveniente y no estorbe al mecanismo del motor. Además, no requiere de aditamentos extra como indicadores ópticos o acoplamientos mecánicos.

Una desventaja al usar este método es la presencia de un umbral de detección impredecible. Dada la naturaleza de las vibraciones mecánicas, se esperan respuestas dinámicas completamente diferentes para todo tipo de motores. Debido a esto es difícil determinar en primera instancia si los valores devueltos por el instrumento son veraces o errores aleatorios. Sin embargo, es posible mantener la opción de una ganancia variable cuando el ruido producido por un motor sea, o muy tenue, o muy intenso, dando al usuario del instrumento control sobre el umbral de detección.

6. Bibliografía y Referencias

- [1] Audacity, Audacity download | SourceForge.net, 2000: <http://sourceforge.net/projects/audacity/>, 2016.
- [2] De Silva, C. *Vibration Fundamentals and Practice*. CRC Press LLC, 2000.
- [3] Mazzoni, D. *Spectrum.cpp*. GNU General Public License Open Software,

- 2000.
- [4] Meddis, R. & O'mard, L. A unitary model of pitch perception, 1997.
 - [5] Middleton, G. Pitch Detection Algorithms, *OpenStax CNX*, 2003: <http://cnx.org/contents/i5AAkZCP@2/Pitch-Detection-Algorithms>, 2017.
 - [6] Nida, Basic Transducers Experiment Card Set Model 1405B. Nida Corporation, 1996.
 - [7] Park, J. & Mackay, S. Practical Data Acquisition for Instrumentation and Control Systems, pp. 435, 2003.
 - [8] Pico Technology, PicoScope 2000 Specifications: <https://www.picotech.com/oscilloscope/2000/picoscope-2000-specifications>, 2017.
 - [9] Shannon, C. A Mathematical Theory of Communication, *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
 - [10] Smith, S. The Scientist and Engineer's Guide to Digital signal processing, 1999.
 - [11] Stoker, J. Nonlinear Vibrations in Mechanical and Electrical Systems. Interscience Publishers, Inc., 1961.
 - [12] Tolonen, T. & Karjalainen, A. A computationally efficient multipitch analysis model, *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 708–716, 2000.
 - [13] Wikipedia, Theil–Sen estimator, *Wikipedia*, 2017: https://en.wikipedia.org/wiki/Theil–Sen_estimator.

DESIGN, CONSTRUCTION AND SIMULATION OF A UNIFORM MAGNETIC FIELD GENERATOR WITH STEEL NUCLEUS TO DEFLECT COSMIC RAYS

Karla Natalia Herrera Guzmán

University of Guanajuato, Department of Physics
herrera2012@licifug.ugto.mx

Raúl Alejandro Gutiérrez Sánchez

University of Guanajuato, Department of Physics
gutierrezr2012@licifug.ugto.mx

Jorge Luis Arceo Miquel

University of Guanajuato, Department of Physics
miquel@fisica.ugto.mx

Julián Félix

University of Guanajuato, Department of Physics
felix@fisica.ugto.mx

Resumen

La trayectoria de una partícula puede ser determinada midiendo algunos puntos por los que ha pasado. Esto se aplica a cualquier tipo de partículas, incluyendo rayos cósmicos. En este se presenta la construcción de un generador de campo magnético uniforme, dentro del cual se colocará un arreglo de detectores para medir las trayectorias de los rayos cósmicos. Presentamos detalles del diseño, construcción, calibración y algunos resultados preliminares.

Palabras Claves: Bobinas, partícula relativista, rayos cósmicos, simulación.

Abstract

A particle's trajectory can be determined measuring some points where it has passed. This is applied to all kind of particles, including cosmic rays. In this paper

we present the construction of a uniform magnetic field generator. Inside it, an arrangement of cosmic ray detectors will be placed in order to measure cosmic ray trajectories. We present details of the design, construction, calibration and some preliminary physical results.

Keywords: *Cosmic rays, helmholtz coils, relativistic particle, simulation.*

1. Introduction

The cosmic rays are particles coming from outer space and they can be charged or neutral particles (charged cosmic rays at sea level are mostly muons). They were discovered in 1912 by Victor Hess. Since then, a lot of cosmic rays detectors have been built to study the universe. There are two types of cosmic rays: primary and secondary. The primary cosmic rays are generated by astrophysical sources such as supernovae, stars, pulsars, etc. The secondary cosmic rays are generated by collision of primary cosmic rays with interstellar gas, this means that earth atmosphere is a source of secondary cosmic rays [PDG, 2015].

The Cerenkov radiation is produced when a charged particle travels faster than light in a medium (radiator) [Mark, 2017], this radiation can be detected by photomultipliers. There are several types of cosmic ray detectors using Cerenkov radiation [Butslov, 1963], [Aseev, 1992], but most of them are based on transparent materials. New radiator materials could allow us to explore different energetic regions of detection and particles. Cosmic ray detectors usually are made of gases and liquids, among other transparent materials. The aim of this work is to know if Cherenkov radiation is produced in a material like aluminum, if so, it must be detectable. A uniform magnetic field deflects the charged particles trajectories, thus it is possible to use it with a particle detector. The objective is to detect the change in the direction of the cosmic rays and determine their energy, momentum, identity and trajectories. To achieve this, it is necessary to build a base to place an array of detectors inside the magnetic field generator. It is necessary that neither the detector's material nor the base's material distort the magnetic field. In this case we built the system for an arrangement of detectors of 8x8x8 in which consists of 32 Aluminum bars (1x2x8 in).

2. Methods

Analytic Description

Cosmic rays are very energetic, therefore it is considered a relativistic calculation about how they are deflected in a magnetic field. Starting with the Lorentz's principle [Serway, 2005] with no electric field

$$\vec{F}_{electromagnetic} = q\vec{v} \times \vec{B},$$

Where q is the electric charge of the particle, \vec{v} its velocity and \vec{B} the magnetic field. Using the second Newton's principle with the relativistic correction, equation 1.

$$\frac{d\vec{p}}{dt} = q\vec{v} \times \vec{B}, \quad (1)$$

Where $\vec{p} = \gamma m \vec{v}$ is the relativistic momentum ($\gamma = \sqrt{1 - \vec{v} \cdot \vec{v}}$).

Due to that a charged particle moving inside a uniform magnetic field follows a uniform circular motion (which implies that v is constant), γ is a constant. Because of this, and that $\vec{B} = (0, B, 0)$ and $\vec{v} = (v_x, 0, v_z)$, (1) is reduced to the following equations 2.

$$\begin{aligned} \gamma m \frac{dv_x}{dt} &= -qBv_z, \\ \gamma m \frac{dv_z}{dt} &= qBv_x. \end{aligned} \quad (2)$$

Defining $\alpha = \frac{\gamma m}{qB}$ and solving equation 2 we found equations 3 y 4.

$$x = \alpha \left[A \sin\left(\frac{t}{\alpha}\right) - B \cos\left(\frac{t}{\alpha}\right) \right] + K \quad (3a)$$

$$z = -\alpha \left[A \cos\left(\frac{t}{\alpha}\right) + B \sin\left(\frac{t}{\alpha}\right) \right] + C \quad (3b)$$

$$v_x = A \cos\left(\frac{t}{\alpha}\right) + B \sin\left(\frac{t}{\alpha}\right) \quad (4a)$$

$$v_z = A \sin\left(\frac{t}{\alpha}\right) - B \cos\left(\frac{t}{\alpha}\right) \quad (4b)$$

Applying the initial conditions $x = 0$, $z = z_0$ and $v_x = 0$, $v_z = v_{z0}$ in $t = 0$ to equations 3a, 3b, 4a and 4b, the solution is equations 5a and 5b.

$$x = \alpha v_{z0} \cos\left(\frac{t}{\alpha}\right) - \alpha v_{z0} \quad (5a)$$

$$z = \alpha v_{z0} \operatorname{sen}\left(\frac{t}{\alpha}\right) + z_0. \quad (5b)$$

The equations 5a and 5b have to be solved to find the maximum and minimum incidence speed in which the incident particle is going to be deflected in the desired way to find the lower and upper bounds for the energy. In order to solve this equations it is necessary to give the B , m , q y z_0 values, those are the magnetic field generated by the coils, the muon mass, the muon charge and the point in which the particle enters in the magnetic field.

Design

The prototype design consists in a pair of Helmholtz coils with 210 turns each and a ferric nucleus. Both coils are joined by steel bars to close the magnetic field. An aluminum base, which will be used to place detectors inside the magnetic field, was designed too. Figure 1 shows the design of the coils and the Aluminum base (drawn in SketchUp).

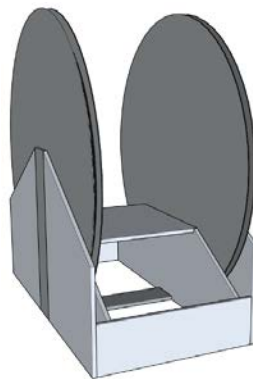


Figure 1 System perspective view.

In addition, the ferric pieces that makes the coils' reels, as well as the pieces that close the magnetic field, are made of ferric sheets to reduce possible eddy currents. The dimensions of the pieces in the design are shown in table 1. The shape and dimensions of some pieces are shown in figure 2.

Table 1 Pieces of the coil's reels and Aluminum base.

PIECES	MATERIAL	SIZE (cm)
4	Aluminum	Figure 4.1
2	Aluminum	Figure 4.2
2	Steel	27.22 x 5.00 x 0.40
1	Steel	23.72 x 5.00 x 0.40
1	Aluminum	20.32 x 20.32 x 0.40
2	Aluminum	23.72 x 10.00 x 0.40
PIECES	MATERIAL	DIAMETER (cm)
2	Steel	47.04
10	Steel	43.04

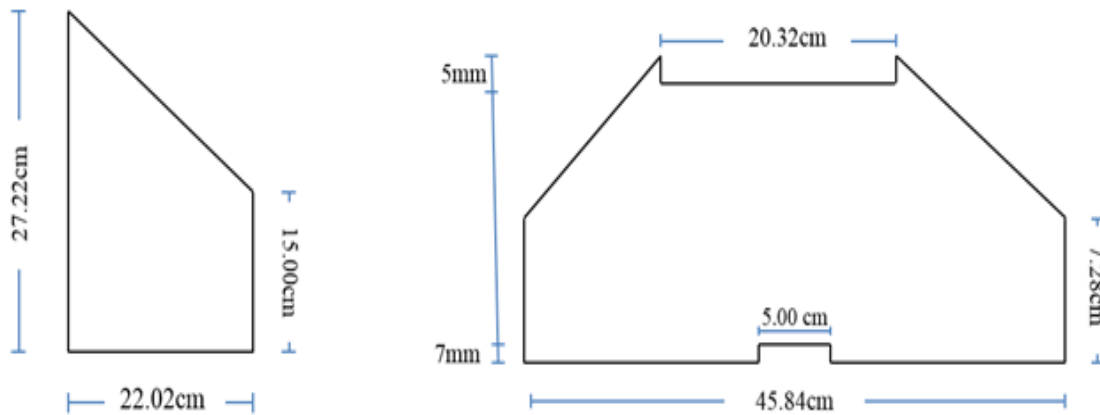


Figure 2 Dimensions and shape of two pieces of the Aluminum base.

All pieces in the design are joined together with nuts, bolts and Aluminum angles. Steel bolts are used to join steel pieces together and Aluminum bolts to join all Aluminum pieces.

This prototype is supposed to hold everything, including the electronic cards. Thus, a few Aluminum plates must be added. Those extra Aluminum plates are shown in figure 3. This figure also shows the way the 32 Aluminum bars will be placed inside the coils. Number 1 is an Aluminum plate needed to hold the connection strips (white rectangular prisms) needed to supply the voltage to the discriminator cards (dark green plates). Number 2 is an Aluminum angle that will be used to keep the photodiode cards (green plates) in its place (the latter ones will be screwed to the first one). Number 3 is an Aluminum plate and an Aluminum “C” that will support the discriminator cards outside the magnetic field.

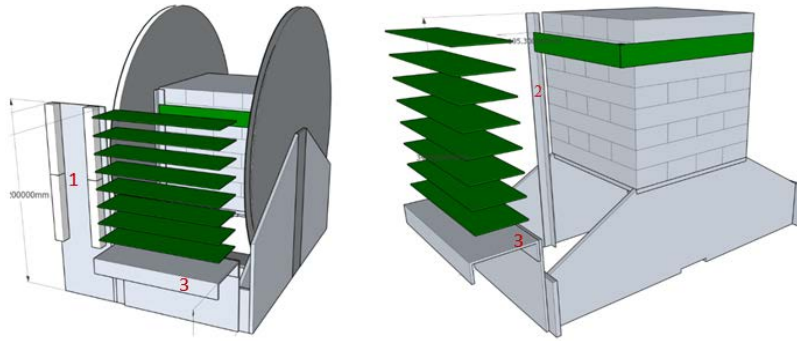


Figure 3 Design with aluminum pieces added to hold all electronic components.

Construction

All pieces in table 1 were cut with water jet. The steel pieces (Carbon steel A36) were cut from a steel sheet 18 AWG (0.91 mm) and the Aluminum pieces from a 5 mm thick aluminum plate.

Base Construction

The Aluminum angles and pieces were drilled. As the Aluminum bolts were too long to press the aluminum pieces together, we had to make washers. These washers were made with 5 mm thick black rubber. The washers used to fix the Aluminum square on top of the base, due to the limited space, are pre-made washers 1.2 mm thick.

Coils' Reels Construction

The steel laminas were glued together with JB Weld to form bars. The result was two bars of 27.5 x 5.0 x 0.6 cm and one of 23.4 x 5.0 x 0.6 cm (for future references this pieces are called L and K pieces respectively). The discs were perforated with 5 holes of 1/2 inch diameter. With the drilled discs were made 2 reels, each one containing two 23.4 cm radio discs and eleven 21.4 cm radio discs. The discs were glued together by the perimeter.

Assembly and Winding

The steel reels were wined with 210 turns each with 19 AWG wire. In the turn 120 it was necessary to put insulating tape to level them. After finishing the

winding, Qualtex Silicone was used in the top of the reels to protect the coils. L and K pieces are screwed together, so the necessary threaded holes were made. Finally, the reels were painted red color and the L and K pieces matte black color.

Coils' Electric Connection

To switch the magnetic field direction produced by the coils, it was used a two pole two throw switch, female banana connectors and connector strips. The components were connected with 5 mm red and black wire. The magnet wire of the coils was protected using thermofit.

Magnetic Field Mapping System

In order to map the magnetic field, a MakeBlock XY Plotter [MakeBlock, 2017] was adapted. A graduated (in millimeters) acrylic tube was used instead of a pencil. Inside the acrylic tube, a Vernier MG BTA gaussmeter [Vernier, 2017] was placed. With this modification was possible to map the magnetic field in a volume of $20.00 \pm 0.05 \times 19.00 \pm 0.05 \times 20.0 \pm 0.1$ cm in x, y, z directions respectively. Figure 4 shows the measurement system on the prototype. Figure 5 shows the details of the measurements. In figure 5a the measurement system maps flat surfaces from left to right and to the electric connection "c", starting from the point in the red circumference. Figure 5b shows that the surfaces are mapped from the bottom to the top surface with a distance between them of $d = 1$ cm. Each surface is mapped in 20 lines from left to right; each line is made of 19 points. The mapped volume consists of 20 flat surfaces 1 cm apart.

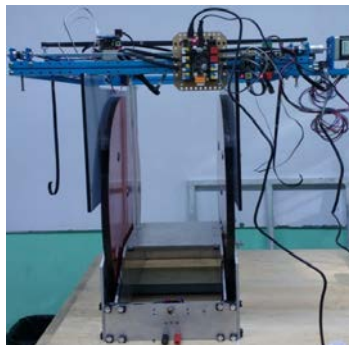


Figure 4 Measurement system on the prototype.

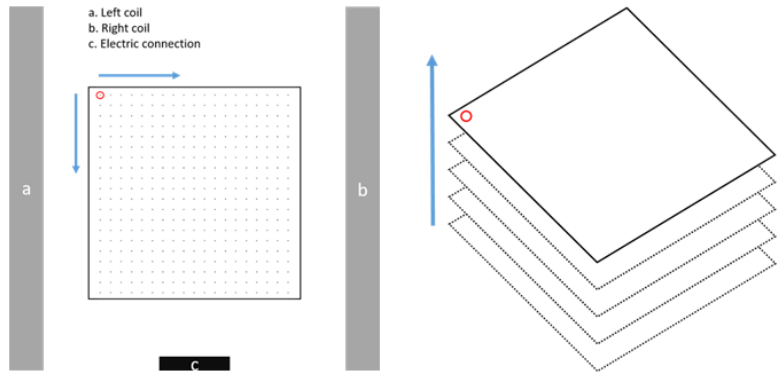


Figure 5 Scheme of the magnetic field mapping.

When the sensor in z direction is moved, it is possible that it got bend. Due to that this is a one direction magnetic field sensor, it is necessary to move it until it's parallel to the reels.

3. Results

Simulation

In order to have a prediction of the magnetic field that is generated by the coils, a model with real dimensions was sketched in Poisson Superfish [Poisson, 2017]. The resulting magnetic field is shown in table 2 when the current passing through the wire is between 1 and 20 amperes (the magnetic field magnitude at the center of the coils' axis). Figure 6 shows the generated graph by Poisson Superfish.

Table 2 Magnetic field magnitude obtained with Poisson Superfish, The relative μ value for the material of the reels is $\mu=250$.

Current (A)	2	4	8	10	12	14	16	18	20
Magnetic Field (mT)	2.9	5.9	11.8	14.7	17.7	20.6	23.6	26.5	29.4

Magnetic Field Measurements and Prediction of the Equations (no core).

The predicted values were obtained by solving the Helmholtz coils equations and considering the middle point between coils on their common axis as origin. The resulting equation 6.

$$B = \frac{8 \cdot \mu_0 \cdot I \cdot N}{5 \cdot \sqrt{5} \cdot a}, \quad (6)$$

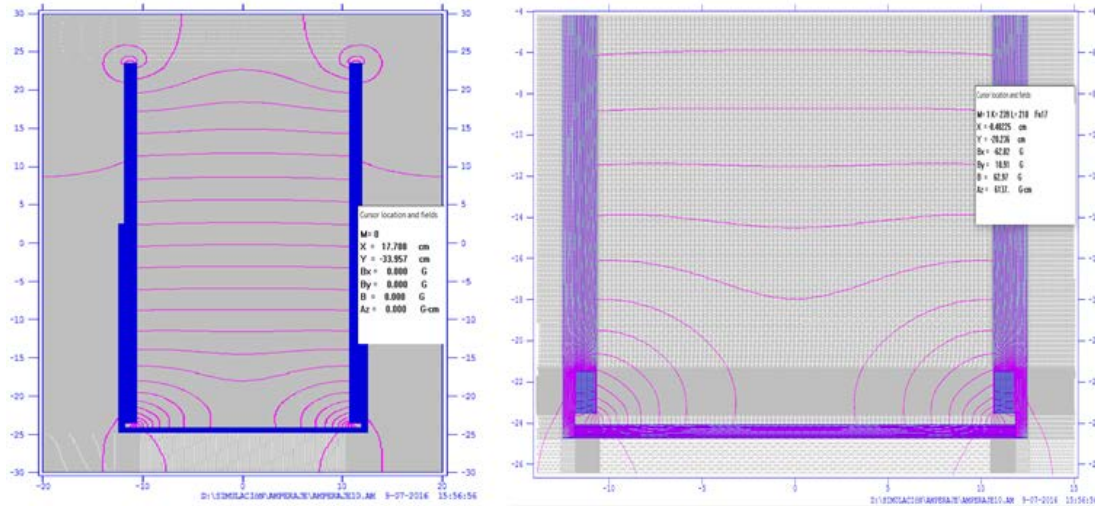


Figure 6 Magnetic field model.

where μ_0 is the vacuum permeability, I the electric current passing through the wire, N the number of turns, and a the radius of coils [Milford, 1995].

In order to obtain the multiplicative factor A that adjusts the simulation to the measurement (the factor that multiplies the measurements so they get as close to the simulation as possible), we start with the method of least squares [Steven, 2017].

$$\chi^2 = \sum_{i=1}^n \frac{(B_{simi} - AB_{mi})^2}{(\Delta B_{mi})^2}, \quad (7)$$

Where B_{simi} is the obtained magnetic field value by the simulation, B_{mi} is the measured magnetic field value. Equating the first derivative of equation 7 to zero and considering an error of 5% ($\Delta B_{mi} = 0.05 \cdot B_{mi}$), A and ΔA are obtained equations 8 y 9.

$$A = \sum_{i=1}^n \frac{B_{simi}}{n \cdot B_{mi}}, \quad (8)$$

$$\Delta A = \sqrt{\sum_{j=1}^n \frac{B_{simi}^2 \cdot 0.0025}{n^2 \cdot B_{mi}^2}}. \quad (9)$$

The multiplying factor that makes the simulation values closer to the measurements (C) is related with A as equation 10.

$$C = \frac{1}{A} \cdot \Delta C = \frac{1}{A^2} \cdot \Delta A. \tag{10}$$

To pass from the predictions factor to measurements factor, the following changes have to be made $B_{simi} = B_{mi}$ and $B_{mi} = B_{pi}$, with the subscript pi denoting the prediction values. Table 3 shows ten measurement points (n=10) with their respective electric current values. In these 10 points, A and C were measured for simulation (subscript s) and prediction (subscript p) cases, these results are given in table 4. Figure 7 shows measurement, simulation and prediction results.

Table 3 Electric current values.

n	1	2	3	4	5	6	7	8	9	10
I (A)	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0

Table 4 Multiplying factors.

	A_s	C_s	A_p	C_p
Value	1.28	0.75	1.38	0.72
Error	0.05	0.03	0.05	0.03

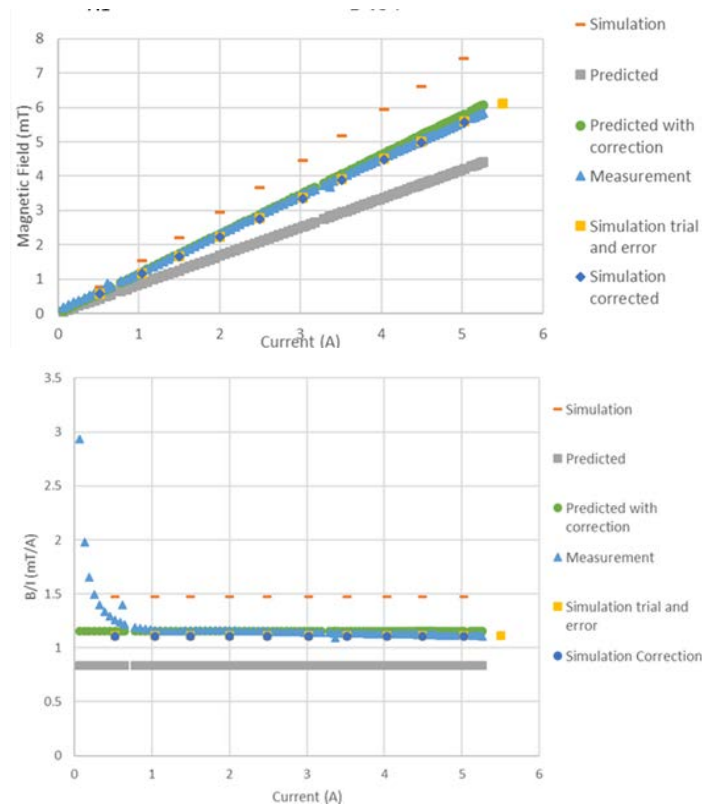


Figure 7 Measurement, simulation and prediction results.

The measurements in figure 7 were made with a VGA AlphaLab sensor, this measurements are consistent with the Vernier sensor. Yellow points are obtained by modifying the μ value in the simulation by trial and error until we found a value of B approximated to the measurements, the μ value obtained is $\mu = 120.05$. The values indicated by "Simulation Correction" (dark blue) are the values obtained by multiplying C_s with the original simulation values to make them fit with the measurement values. The green values are obtained in the same way that the dark blue ones, but this time multiplying the predicted values (gray) with A_p to fit the predicted values to the measurements.

Relativistic Particle Kinematics.

The values needed to solve equations 5a and 5b are:

$B = 29.4429\text{mT}$ (obtained from Poisson Superfish for a 20 Amperes current).

$$m = 1.08838 \times 10^{-28} \text{ kg}$$

$$q = \pm 1.0602 \times 10^{-19} \text{ C}$$

$$z_0 = 23.52 \text{ cm}$$

Defining $\beta = \frac{m}{qB}$, the equations 11.

$$\begin{aligned} x &= \frac{\pm \beta x c}{\sqrt{c^2 - v_{z0}^2}} \cdot v_{z0} \left[\cos \left(\frac{\sqrt{c^2 - v_{z0}^2}}{\beta x c} \cdot t \right) \mp 1 \right], \\ z &= \frac{\pm \beta x c}{\sqrt{c^2 - v_{z0}^2}} \cdot v_{z0} \left[\text{sen} \left(\frac{\sqrt{c^2 - v_{z0}^2}}{\beta x c} \cdot t \right) \mp 0.2352 \right]. \end{aligned} \tag{11}$$

To find the incidence speed of the particle we must give conditions to the energy bounds, see figure 8 which are:

Upper bound; case in which the particle leaves the array of detectors without a detectable deflection (green, the width of one detector is 5.08 cm):

$$x = \pm 0.0508\text{m}, \quad z = -0.1016\text{m}.$$

Lower bound; case in which the particle leaves the array of detectors before crossing its half (red):

$$x = \pm 0.1016\text{m}, z = 0.0254\text{m}$$

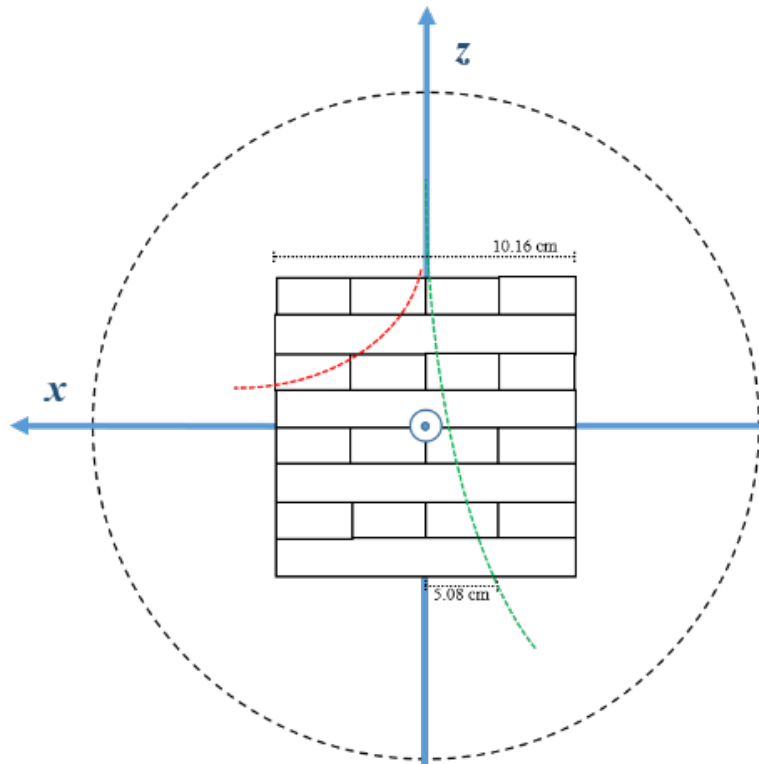


Figure 8 Energy bounds. In red, the trajectory of a particle passing through the lower bound; in green, the one which goes through the upper bound.

As equations 12 are nonlinear equations, it was necessary to use an internet platform called “WolframAlpha” to solve them. The results obtained are:

$$\text{Lower Bound: } v_{z0} = -6.7 \times 10^6 \frac{\text{m}}{\text{s}}, t = 3.6 \times 10^{-8} \text{s.}$$

$$\text{Upper bound: } v_{z0} = -2.8 \times 10^7 \frac{\text{m}}{\text{s}}, t = 1.2 \times 10^{-9} \text{s.}$$

The energy bounds according to these velocities are:

$$E_{max} = 106.3 \text{MeV}, E_{min} = 105.9 \text{MeV.}$$

Automated Magnetic Field Map

When the coils were powered with 100 V, the power supply was giving about 6.6 A, but they started to warm up. This made the 6.6 A to decrease rapidly. The current output stabilized about 1.5 hours later, having around 4.7 A and an

approximated temperature of 78 °C. It is at this point where the magnetic field measurements takes place. It is necessary to point out that the rising of the temperature deforms the acrylic tube. This makes that the measurements in the vicinity of the reels are made with a deformed acrylic tube. The acrylic tube is separated from the aluminum square 1.52 ± 0.01 mm (before deformation). Table 5 shows the initial and final current and temperatures in which the measurements took place for every plane.

Table 1 Values of current and temperature between each plane was mapped.

Plane number	I_i (A)	I_f (A)	T_i (°C)	T_f (°C)
0	4.767	4.695	72.5	77.0
1	4.675	4.669	78.8	78.8
2	4.668	4.665	79.5	79.4
3	4.665	4.660	79.4	79.7
4	4.661	4.660	79.8	80.0
5	4.760	4.690	71.6	77.3
6	4.690	4.680	77.3	77.5
7	4.680	4.666	77.7	78.8
8	4.667	4.662	79.1	79.7
9	4.662	4.657	79.7	80.2
10	4.657	4.657	80.3	80.2
11	4.658	4.659	80.4	80.2
12	4.807	4.713	68.4	75.5
13	4.712	4.686	75.7	76.8
14	4.686	4.674	76.8	77.5
15	4.674	4.666	78.2	78.6
16	4.666	4.665	78.7	78.9
17	4.665	4.663	78.9	79.1
18	4.663	4.665	78.7	78.8
19	4.665	4.673	78.4	78.4
20	4.704	4.690	77.0	78.5

Figure 9 shows the graphs of the initial, final and central mapped plane, the rest of the planes were omitted because this planes are enough to realize the way the magnetic field changes with the z position.

In table 6 it is shown the maximum and minimum of each plane and its proportion. The last row shows the maximum variation of the entire map (21 planes) taking the global maximum and minimum. Figure 10 shows the graphs of the values in table 6.

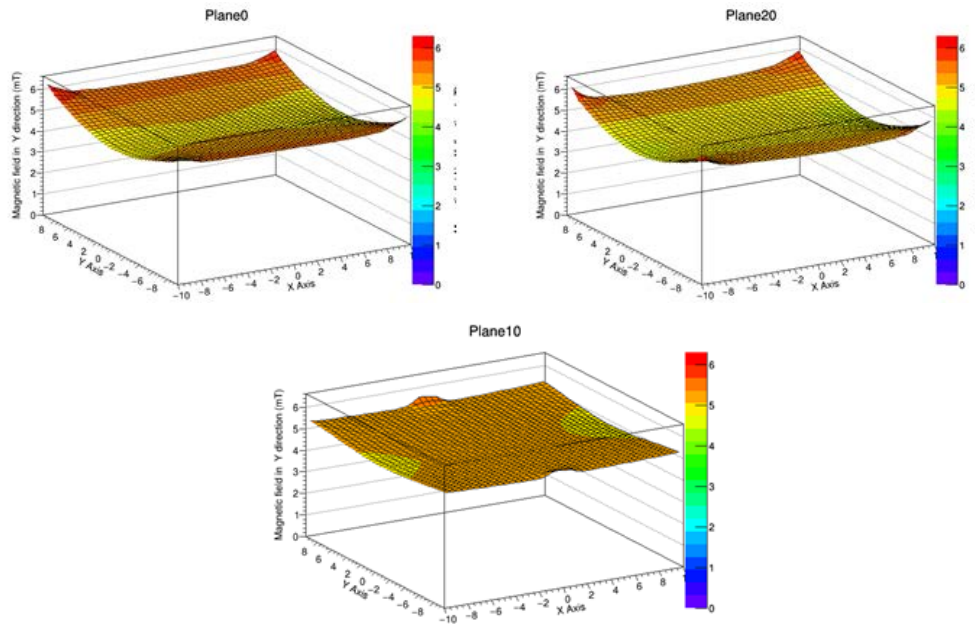


Figure 2 Graphs of each one of some mapped planes.

Table 2 Maximum and minimum values of each plane and its percent variation.

Plane	Max (mT)	Min (mT)	Variation
0	6.3	4.7	1.4
1	5.7	4.7	1.2
2	5.6	4.7	1.2
3	5.5	4.8	1.2
4	5.5	4.8	1.1
5	5.6	4.9	1.1
6	5.5	4.9	1.1
7	5.5	4.9	1.1
8	5.4	4.9	1.1
9	5.8	4.9	1.2
10	5.7	4.9	1.2
11	5.4	4.9	1.1
12	5.5	4.9	1.1
13	5.4	4.9	1.1
14	5.4	4.8	1.1
15	5.3	4.7	1.1
16	5.3	4.6	1.1
17	5.3	4.6	1.2
18	5.7	4.6	1.2
19	6.3	4.5	1.4
20	6.3	4.4	1.4
Maximum variation	6.3	4.4	1.4

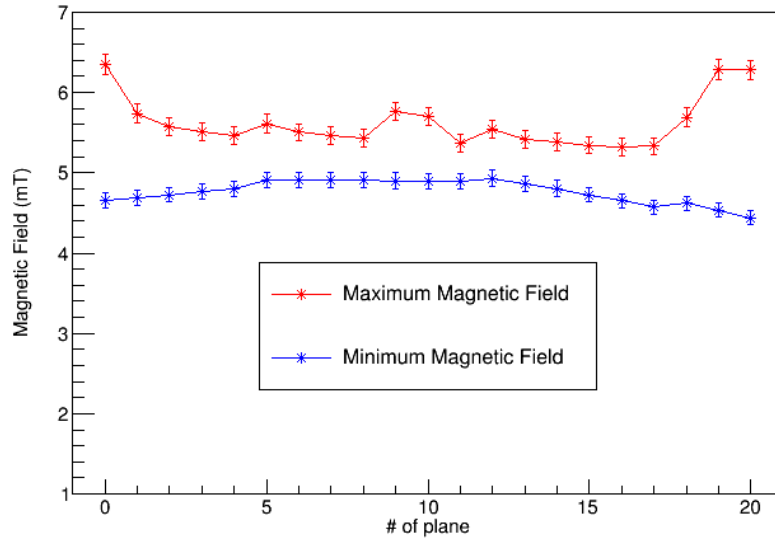


Figure 10 Maximum and minimum magnetic field values.

4. Discussion

This cosmic ray detector will have the unique characteristic that uses a metal as a radiator material for Cherenkov radiation. Some very preliminary tests have been done and we know we measure something in the aluminum bars, but we still have to do more tests to demonstrate that it is Cherenkov radiation. No other country is developing this kind of detectors. One of the advantages about using metals is that you do not have to isolate them of external radiation, they are easily mounted and the maintenance is much easier than for liquids and gases.

5. Conclusions

We have designed, constructed and characterized a device to create a uniform magnetic field. We have simulated the magnetic field as a function of electric current; we have predicted the magnetic field proportional to the applied electric current; we have measured the magnetic field as a function of the applied electric current.

The prediction and simulation results were adjusted to the results of the measurements by least square method and to the measurements to obtain the μ_r of the utilized material of the coils' core resulting $A_p = \mu_r = 1.38 \pm 0.05 * \mu_0$ (figure 7). This is a technique to measure the magnetic permeability of the coils' core.

Incident cosmic particles with energy between $E_{max}=106.311$ MeV and $E_{min}=105.858$ MeV must be detected by this spectrometer. Loss of energy from particles passing through the detectors is not considered.

Magnetic field percent variation decreases as the sensor approaches to the axis of the coils and increases as they move away from it. Also it is observed that this variation increases faster in the upper planes than in the lower planes.

6. Bibliography and References

- [1] Mark Chen, Queen's University. Scintillation and Light Sensitive Detectors. http://neutron.physics.ucsb.edu/docs/scintillation_presentation_info.pdf, March 11, 2017.
- [2] Aseev, E.G. Devitsin, A.A. Komar, V.A. Kozlov, Yu.I. Hovsepyan, S.Yu. Potashov, K.A. Sokolovsky, T.V. Uvarova, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 317 pp. 143-147, 1992.
- [3] Butslov M. M., Medvedev M. N., I V Chuvilo and M V Sheshuno, Nuclear Instruments And Methods 20, pp. 263-266, 1963.
- [4] Makeblock, website, <http://learn.makeblock.com/xy-plotter-robot-kit/>. August 11, 2017.
- [5] PDG (Particle Data Group), <http://pdg.lbl.gov/2011/reviews/rpp2011-rev-cosmic-rays.pdf>, January 2017.
- [6] Poisson Superfis: http://laacg.lanl.gov/laacg/services/download_sf.shtml, August 11, 2017.
- [7] Reitz J. R. y F. J. Milford, Foundations of Electromagnetic Theory (Massachusetts/Adison-Wesley) Capítulo 8, pp. 165-166, 1992.
- [8] Serway R. A. y R J Beichner, Física para ciencias e ingenierías. Tomo II (Mexico/McGraw-Hill) Chapter 29, pp. 922, 2001.
- [9] Steven J. Miller. The method of Least Squares 0. https://web.williams.edu/Mathematics/sjmillier/public_html/BrownClasses/54/handouts/MethodLeastSquares.pdf, August 11, 2017.
- [10] Vernier, <https://www.vernier.com/products/sensors/mg-bta/>, January 2017.

DISEÑO E IMPLEMENTACIÓN DE SENSORES DE GAS QCM DE ALTA SENSIBILIDAD PARA UNA NARIZ ELECTRÓNICA

Juan Jesús Jiménez Arellano

Benemérita Universidad Autónoma de Puebla
juanjesusjimenez@yahoo.com.mx

Severino Muñoz Aguirre

Benemérita Universidad Autónoma de Puebla
smunoz@fcfm.buap.mx

Juan Castillo Mixcoatl

Benemérita Universidad Autónoma de Puebla
smunoz@fcfm.buap.mx

Georgina Beltrán Pérez

Benemérita Universidad Autónoma de Puebla
smunoz@fcfm.buap.mx

José Lorenzo Muñoz Mata

Universidad Tecnológica de Puebla
jose.munoz@utpuebla.edu.mx

Resumen

Este trabajo presenta el incremento de sensibilidad de sensores de gas del tipo microbalanza de cristal de cuarzo (QCM) para una nariz electrónica. Se muestra el diseño e implementación del sensor a través de una variante del circuito oscilador de Colpitts para QCM, primero en su modalidad fundamental y luego en la de tercer sobretono, acondicionando la señal de salida en forma digital. El sensor es colocado dentro de una cámara de medición, donde la temperatura y la humedad internas son continuamente monitoreadas. Después de la estabilización de su

línea-base, se hace una serie de aplicación de muestras del componente orgánico volátil a detectar y las variaciones de frecuencia son medidas por un frecuencímetro de alta resolución, cuyos datos son registrados y almacenados a través de una aplicación en LabVIEW. Finalmente se muestran y analizan los resultados obtenidos realizando un comparativo con sensores del mismo tipo, pero de baja frecuencia.

Palabras Claves: Nariz electrónica, QCM en sobretono, sensor de gas.

Abstract

This work presents the increment of sensitivity to gas sensors of the type Quartz Crystal Microbalance (QCM) for an electronic nose. The design and implementation of the sensor is shown through a variant of the Colpitts oscillator circuit for QCM, first for fundamental mode and then for third overtone, conditioning the output signal in digital form. The sensor is placed inside of a measuring chamber where both internal temperature and humidity are continuously monitored. After the stabilization of its baseline, a series of application of samples of the volatile organic component to be detected is performed and the frequency variations are measured by a high-resolution frequency meter, whose data are recorded and stored through an application in LabVIEW. Finally, the results obtained are shown and analyzed performing a comparison with sensors of the same type, but of low frequency.

Keywords: *Electronic nose, gas sensor, QCM in overtone.*

1. Introducción

Una nariz electrónica es un dispositivo diseñado de manera selectiva para la detección y análisis de componentes orgánicos volátiles cuyo desempeño depende en gran parte del tipo de sensor a utilizar. Entre los sensores más empleados destacan los del tipo QCM debido a su afinidad como sensor microgravimétrico, disponibilidad y bajo costo [Gardner, 1999]. Para este tipo de sensores, el depósito de una película sensible a cierto componente orgánico volátil sobre la superficie de sus electrodos le permite al dispositivo retener una cierta

cantidad de moléculas odorantes, hecho que se manifiesta en cambios en la frecuencia resonante del sensor [Muñoz, 2005]. La ecuación de Sauerbrey (ecuación 1) muestra que la cantidad de masa que pueden retener estos sensores es proporcional al cuadrado de su frecuencia de oscilación [Arnau, 2004].

$$\Delta f = k \frac{\Delta m f_0^2}{A} \quad (1)$$

Donde Δf (Hz) es el cambio en la frecuencia, k ($\text{cm}^2/[\text{g Hz}]$) es una constante de proporcionalidad, Δm (g) es la masa retenida o agregada, f_0 (Hz) es la frecuencia resonante del QCM y A (cm^2) es el área efectiva cubierta por la película sensible sobre el electrodo.

La ecuación 1 expresa que una mayor cantidad de moléculas odorantes retenidas por la película sensible del sensor se ve reflejado en un mayor incremento en los cambios de frecuencia del QCM y por ende un aumento en la sensibilidad del dispositivo. Lo que de inmediato supone depositar películas sensibles con mayor capacidad de retención, hecho que muchas veces no es posible debido a un exceso en la masa sobre el electrodo del QCM imposibilitando así su capacidad de oscilar [Nakamoto, 1996]. Por otro lado, manteniendo la misma cantidad de muestra del componente orgánico volátil en contacto con el sensor y un aumento en la frecuencia oscilante del QCM, implica una mayor respuesta en las variaciones de frecuencia y en consecuencia un aumento significativo en la sensibilidad de la nariz electrónica [Muñoz, 2014], [Nakamoto, 2002], [Stehrer, 2010].

Ante el contexto descrito en los párrafos anteriores, se justifica el propósito del presente trabajo en diseñar e implementar un sensor de gas del tipo QCM para altas frecuencias con la finalidad de ver un incremento significativo en la respuesta del dispositivo ante la presencia de pequeñas muestras de componentes orgánicos volátiles como etanol.

El trabajo de investigación consistió en una revisión acerca de los resonadores de cristal de cuarzo disponible en el mercado por arriba de los 20 MHz, encontrándose para frecuencias desde 30 hasta 200 MHz en modalidad de 3er,

5to y 7mo sobretono. Haciendo referencia a los resultados obtenidos en [Jiménez, 2015], se muestran los criterios de diseño del circuito oscilador para su operación primeramente en 30 MHz tercer sobretono y mediante una simulación en PSpice bajo condiciones ideales, se muestra la funcionalidad de este. Debido a las altas frecuencias que se manejan, fue necesario el diseño de un circuito impreso para optimizar las conexiones entre los componentes y reducir los efectos del ruido. Se implementó físicamente el circuito oscilador realizando los ajustes necesarios para sintonizar los filtros a la frecuencia deseada del cristal bajo operación.

Se describe el procedimiento para la implementación de los QCM, es decir, el proceso para el depósito de la película sensible, la variación en el valor de la frecuencia oscilante antes y después del depósito y el cálculo del espesor de dicha película. Se realizaron las mediciones de respuesta a etanol de los sensores con películas sensibles de diferentes espesores utilizando un frecuencímetro de alta resolución de 1 Hz [Muñoz, 2012]. Los datos obtenidos fueron registrados y almacenados en una computadora para el posterior análisis. Finalmente se realizó una comparación de las respuestas obtenidas con las de sensores del mismo tipo, pero de una frecuencia más baja.

2. Métodos

Para el desarrollo del presente trabajo, se muestran en la figura 1 los elementos esenciales que conforman el sistema de medición de respuesta para sensores de gas del tipo QCM.

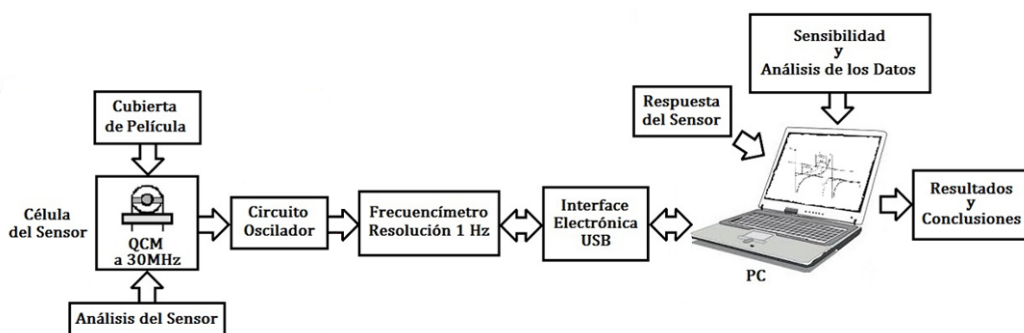


Figura 1 Sistema implementado para medición de respuesta de los QCM.

Diseño del oscilador. Dado que los fabricantes de los cristales de cuarzo ofrecen estos dispositivos en dos modalidades: fundamental para frecuencias por debajo de los 30 MHz y sobretono para frecuencias por arriba de los 30 MHz hasta los 200 MHz. Los cristales fabricados en modalidad de sobretono sí es posible hacerlos oscilar en modo fundamental, pero a la inversa no es posible. El circuito oscilador propuesto para este trabajo es una variante del circuito oscilador de Colpitts para QCM en modo fundamental como se muestra en la figura 2, cuyo análisis fue considerado en [Jiménez, 2015]. Después de proponer el punto de operación Q de $I_{CQ} = 5mA$ y $V_{CQ} = 2.5V$ con una fuente de alimentación para el circuito de $V_{CC} = 5V$, se obtuvieron los valores de las resistencias de polarización para tal punto. Para poder sintonizar este oscilador, el análisis del trabajo anterior dio como resultado que $C_1 = 3.18pF$ y $C_2 = 1.59pF$. Estos valores son teóricos y para evitar ambigüedades se utilizaron dos capacitores variables de 100 pF de tal manera que se fueron ajustando hasta obtener la frecuencia fundamental de 10 MHz y una forma de onda lo más cercana a una senoidal.

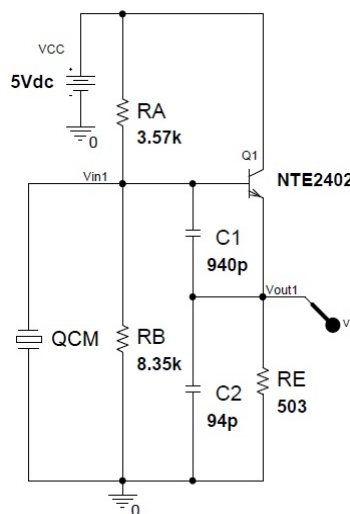


Figura 2 Variación del circuito oscilador de Colpitts para QCM en modo fundamental.

Una vez realizado el análisis para el oscilador en la modalidad de fundamental, se procedió a extender los resultados teóricos a la modalidad de sobretono. Para tal propósito se reconfiguró el circuito de la figura 2 como se muestra en la figura 3. El

punto de operación Q para este circuito se propuso de $I_{CQ} = 10mA$ y $V_{CQ} = 2.5V$ con una fuente de alimentación para el circuito de $V_{CC} = 5V$. Nuevamente, empleando los resultados del análisis hecho en [Jiménez, 2015], se obtuvieron los valores de las resistencias de polarización para tal punto. El circuito de la figura 3 presenta un circuito LC adyacente a la salida del mismo, cuya función es la de un filtro pasa altas para bloquear la frecuencia de oscilación fundamental del cristal y dejar pasar el sobretono. El análisis matemático del filtro muestra como estimar los valores de sus componentes mediante la ecuación 2.

$$f_{bloq} = \frac{1}{2\pi\sqrt{L_{trp}C_{trp}}} \quad (2)$$

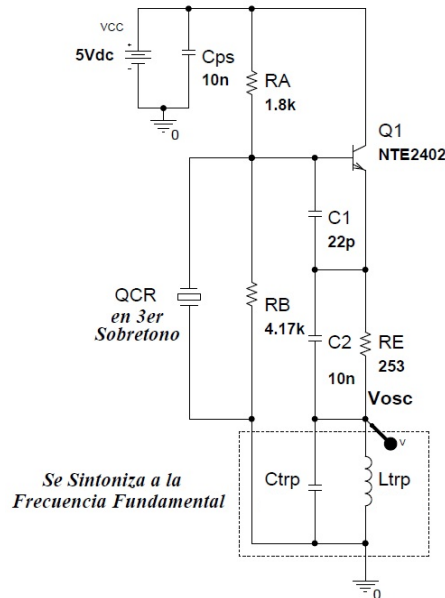


Figura 3 Variante del circuito oscilador de Colpitts para QCM en modo de sobretono.

Si $f_{bloq} = 10MHz$ y $L_{trp} = 5.56\mu H$ es una bobina de valor real, empleando la ecuación 2 resulta que $C_{trp} = 45.56pF$ por lo que en el arreglo experimental se empleó un capacitor variable de 100 pF.

Simulación del oscilador en PSpice. Finalizada la etapa de diseño se realizó una simulación en PSpice del oscilador en modalidad de sobretono debido al interés de manejar las frecuencias altas del cristal. La figura 4 y figura 5 muestran el

diagrama eléctrico del circuito a simular y el resultado de dicha simulación respectivamente. El cristal fue sustituido por su equivalente eléctrico [Arnau, 2004], es decir, por un arreglo en paralelo de circuitos RLC cuyos valores son estimados por algunos fabricantes. De hecho, dichos valores en la simulación se eligieron muy cercanos a los reportados para otras frecuencias en tercer sobretono cercanas a 30 MHz.

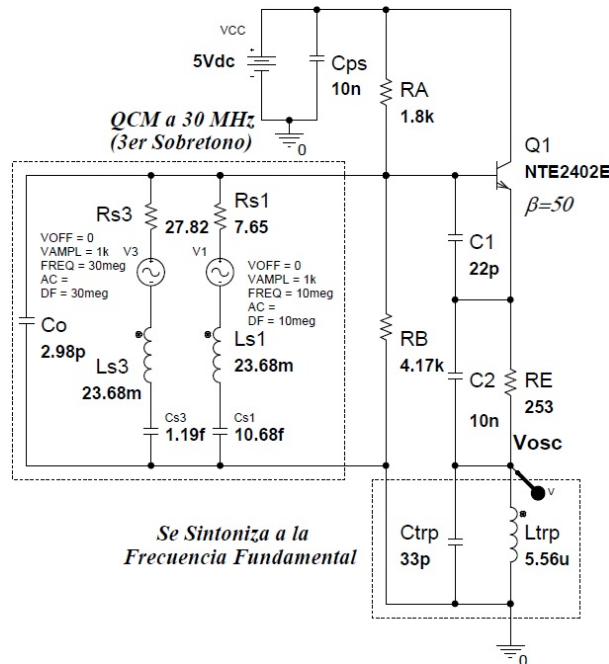


Figura 4 Circuito de simulación en PSpice del oscilador en modalidad de sobretono.

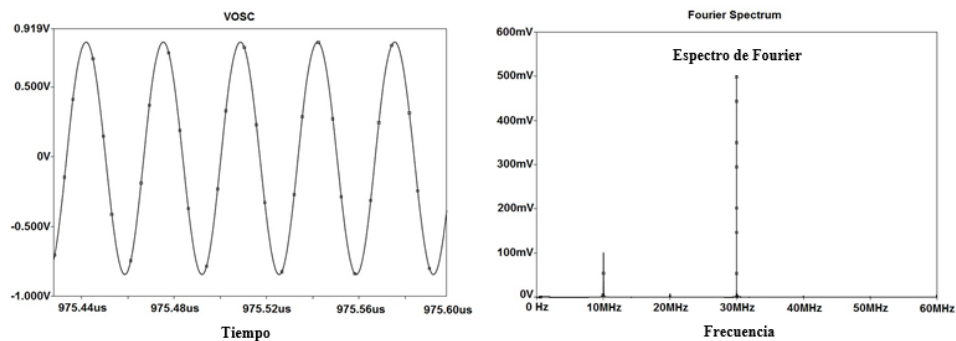


Figura 5 Respuesta de la simulación en PSpice del oscilador y su espectro de Fourier.

En la figura 5 se puede observar el comportamiento del oscilador en la simulación donde la gráfica de la izquierda muestra la señal en el dominio del tiempo cuyo

periodo es aproximadamente de 33 ns y, en la gráfica de la derecha se muestra el espectro de Fourier en el dominio de la frecuencia donde la contribución del tercer sobretono es mucho mayor que la del fundamental.

Implementación del sensor QCM. Para este proceso, se utilizó el método de casting para el depósito de película sensible sobre el electrodo del cristal para ambas caras como lo muestra la figura 6. Para este trabajo se empleó una solución de etil-celulosa disuelta en cloroformo en una proporción de 2 mg/ml.

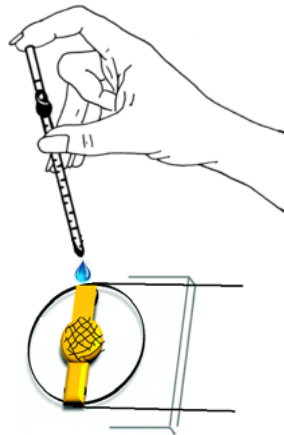


Figura 6 Método de casting para el depósito de película sobre el electrodo del cristal.

Posterior al proceso del depósito de la película sensible, siguió la etapa de la estimación del espesor de dicha película. El proceso es muy simple, primero se quitó el blindaje protector que trae de fabricación el cristal por medio de un corte en la base, después se conectó el dispositivo al oscilador y se midió su frecuencia a través de un frecuencímetro comercial. Luego se hizo el depósito de la película sensible como se describió en el párrafo anterior y se esperó un tiempo razonable para que se adhiriera de una forma más adecuada la película a los electrodos del cristal. Después se midió nuevamente la frecuencia de oscilación del sensor y el resultado que se espera es una frecuencia menor a cuando no tenía película. La diferencia entre estos valores de frecuencia Δf es proporcional al espesor de dicha película como lo indica la ecuación 3.

$$l_{ps} = \frac{\Delta f}{k' \rho_{ec} f_o^2} \quad (3)$$

Donde Δf es el cambio en la frecuencia, $k' = 2/\sqrt{\rho_q \mu_q}$ es una constante que depende de las propiedades intrínsecas del cuarzo, ρ_{ec} es la densidad de masa volumétrica de la etil celulosa y f_o es la frecuencia de oscilación del QCM.

3. Resultados

En la gráfica de la izquierda en la figura 7 se muestra la señal de salida del circuito oscilador cuyos datos fueron leídos desde un osciloscopio Tektronix TDS 3034B donde se puede observar la frecuencia resonante del tercer sobretono aproximadamente a 29.94 MHz. La gráfica de la derecha corresponde al espectro de Fourier calculado en un programa de computadora a partir de los datos obtenidos en el dominio del tiempo donde se pueden observar las contribuciones de las componentes fundamental y de sobretono en múltiplos enteros del fundamental.

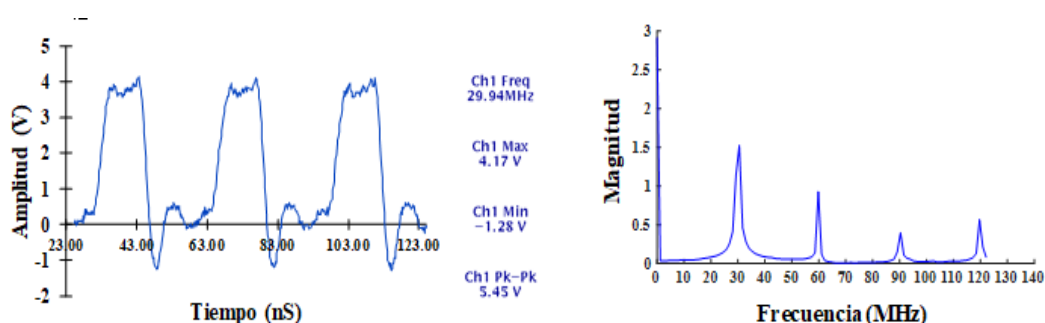


Figura 7 Señal de salida tercer sobretono obtenida y correspondiente espectro de Fourier.

Una vez comprobado el buen funcionamiento del oscilador, se procedió al depósito de la película sensible sobre los electrodos del cristal para 3 sensores mediante la técnica expuesta en la sección anterior. A través de la ecuación 3 se estimó el espesor de dichas películas como se muestra en la tabla 1.

Tabla 1 Estimación del espesor de la película sensible para 3 QCM.

No.	Frecuencia Sin Película Sensible f_o (MHz)	Frecuencia Con Película Sensible (MHz)	Δf (KHz)	l_{sf} (μm)
1	29.99460	29.81518	179.42	77.30
2	29.99503	29.83746	157.57	67.88
3	29.99432	29.62606	368.26	158.66

Finalmente se utilizó el arreglo experimental de la figura 8 para medir la respuesta del QCM expuesto a etanol. El QCM se colocó dentro de una cámara de medición inmersa en un baño térmico a una temperatura de aproximadamente 25 °C. Las temperaturas y humedades externas e internas a la cámara fueron continuamente monitoreadas, con el propósito de observar que se mantuvieran relativamente constantes durante el proceso de medición. La señal de salida del oscilador se conecta a un canal de un frecuencímetro de 8 canales de alta resolución cuyas lecturas son registradas y almacenadas en una interfaz de usuario gráfica.

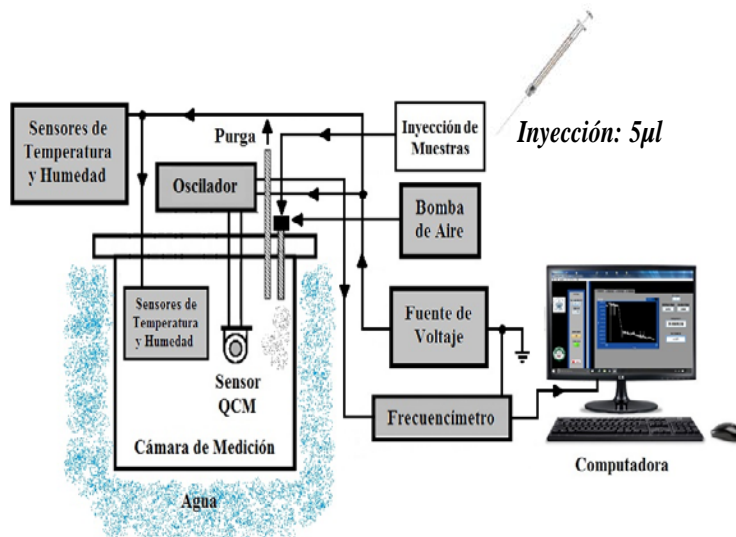


Figura 8 Arreglo experimental para la medición de la respuesta del sensor.

El procedimiento de medición consistió en que una vez inicializada la prueba, esperar el tiempo necesario hasta que lectura de la frecuencia del QCM dentro de la cámara fuera estable, a este comportamiento se le conoce como línea base. Posteriormente se inyectó el componente orgánico volátil, que para este experimento se aplicaron tres muestras de etanol con una concentración de aproximadamente 1625 ppm cada una. Estas pruebas se realizaron con los tres QCM de la tabla 1 que poseen espesores diferentes.

La figura 9 muestra la respuesta Δf del primer QCM cuyo espesor de película sensible fue aproximadamente de 77.30 μm . Las flechas indican el momento donde se aplicaron las muestras de etanol una vez que la respuesta del QCM

alcanzó una nueva línea base. Se observan Δf de aproximadamente 700 Hz por muestra de etanol, lográndose un máximo de respuesta aproximado de 2100 Hz.

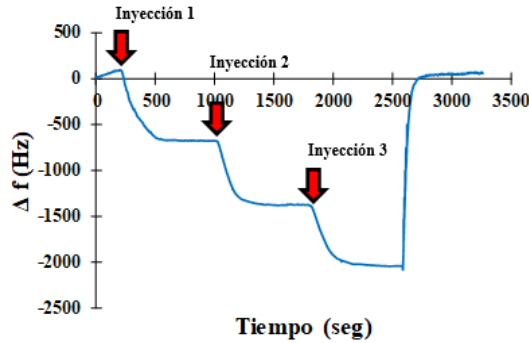


Figura 9 Respuesta obtenida del sensor QCM con un espesor de 77.30 μm .

La figura 10 muestra la respuesta Δf del segundo QCM cuyo espesor de película sensible fue aproximadamente de 67.88 μm . Las flechas indican el momento donde se aplicaron las muestras de etanol una vez que la respuesta del QCM alcanzó una nueva línea base. Se observan Δf de aproximadamente 500 Hz por muestra de etanol, lográndose un máximo de respuesta aproximado de 1500 Hz.

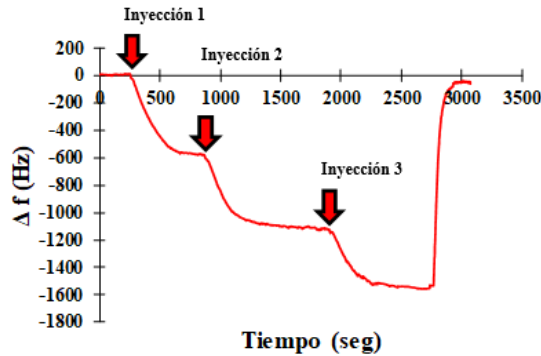


Figura 10 Respuesta obtenida del sensor QCM con un espesor de 67.88 μm .

La figura 11 muestra la respuesta Δf del tercer QCM cuyo espesor de película fue aproximadamente de 158.66 μm . Las flechas indican la aplicación de las muestras de etanol en cada línea base. Se observan Δf de aproximadamente 2100 Hz por muestra de etanol, lográndose un máximo de respuesta aproximado de 6200 Hz.

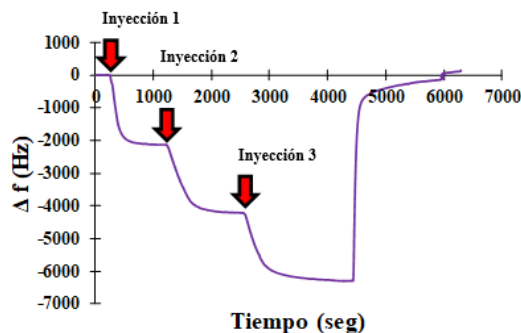


Figura 11 Respuesta obtenida del sensor QCM con un espesor de 158.66 µm.

4. Discusión

Después de realizadas las mediciones correspondientes se procedió a analizar la respuesta en función de la concentración. Se observa un comportamiento lineal para cada una de las respuestas obtenidas como se muestra en la figura 12.

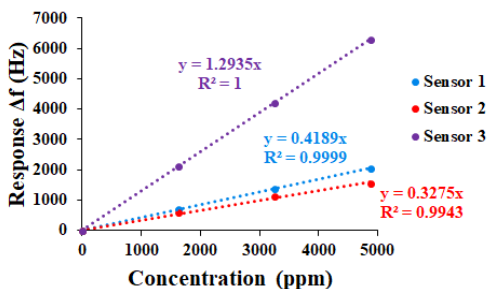


Figura 12 Respuesta en función de la concentración para los sensores analizados.

Como se puede observar, para el sensor 1 se obtuvo un coeficiente de correlación $R^2=0.9999$, el sensor 2 un coeficiente $R^2=0.9943$ y para el sensor 3 un coeficiente $R^2=1$ lo que indica un comportamiento típico para este tipo de sensores. Se puede decir que los resultados obtenidos fueron satisfactorios. Además, se puede asumir que entre mayor sea la pendiente de la recta, mayor es el espesor de la película en el sensor y mayor es su sensibilidad como se define en la ecuación 4.

$$S_{QCM} \equiv \frac{\Delta f}{\Delta c} \quad (4)$$

Donde Δc (ppm) es el cambio en la concentración del componente a detectar, en este caso etanol. Para finalizar se realizó un comparativo con una respuesta de un

sensor del mismo tipo, pero fabricado con un cristal a una frecuencia de 12 MHz en modo fundamental al cual se le sometió a la misma concentración de la muestra de etanol, la figura 13 ilustra este comparativo. Nótese que debido a la baja sensibilidad del QCM a 12 MHz, el decremento en la respuesta del sensor fue menor por lo que se pueden observar fluctuaciones en la misma. Sin embargo, para el QCM de 30 MHz no se observaron las fluctuaciones debido a su alta sensibilidad. Lo que permitirá registrar concentraciones más pequeñas con este tipo de sensor. Lo anterior se puede justificar viendo a través del registro de los datos de ambos sensores en su primera línea base; lo que muestra que existe una fluctuación $\Delta f_{12M} = 2$ (Hz) y $\Delta f_{30M} = 5$ (Hz). A través de la ecuación 4 se calcula el límite de detección (LOD) Δc (ppm) para ambos casos obteniéndose: $\Delta c_{12M} \approx 295$ (ppm) y $\Delta c_{30M} \approx 15$ (ppm), que indica claramente que el QCM de 30 MHz es mucho más sensible que el de 12 MHz pudiendo detectar concentraciones de etanol del orden de 15 ppm.

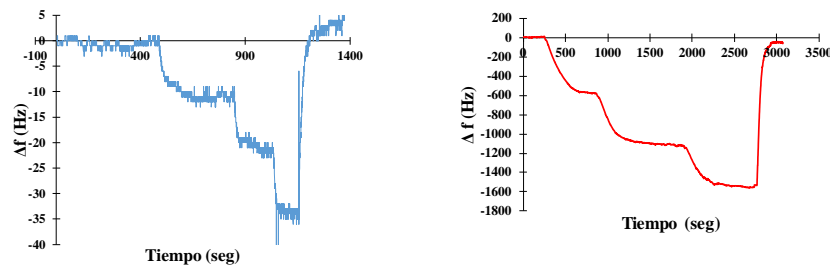


Figura 13 Comparativo entre QCM de baja y alta frecuencia.

5. Conclusiones

- Se ha diseñado e implementado un circuito oscilador de Colpitts para QCM de altas frecuencias en modo de sobretono.
- Se construyeron sensores QCM aplicando películas sensibles de etil celulosa con diferentes espesores.
- Se realizaron las mediciones correspondientes obteniéndose una alta sensibilidad del sensor.
- Se observó en el comparativo que el aumento de la sensibilidad permite medir concentraciones más pequeñas.

- Como trabajo a futuro se plantea construir osciladores y sensores a 40 y 50 MHz para observar su comportamiento a menores concentraciones aplicando películas sensibles más delgadas.

6. Bibliografía y Referencias

- [1] Arnau, A. Piezoelectric Transducers and Applications. Springer-Verlag. Berlin. 1-100, 2004.
- [2] Gardner, J. W., Bartlett, P. N. Electronic Noses: Principles and Applications. Oxford University Press 1-5. Oxford USA, 1999.
- [3] Jiménez-Arellano, J. J., Muñoz-Aguirre, S., Beltrán-Pérez, G., Castillo-Mixcóatl, J., Muñoz-Mata, J. L. Análisis para El Diseño de Circuitos Osciladores de Colpitts con Sensores de Gas QCM. *Pistas Educativas* No.112, 1120-1133. Senie 2015 XI, Noviembre 2015.
- [4] Muñoz-Aguirre, S., López-Casique, A., Alcántara-Iniesta, S., Castillo-Mixcóatl, J., Beltrán-Pérez, G. and Muñoz-Aguirre, N. High-Resolution Gas/Odor Sensors Using High-Frequency Quartz Crystal Microbalance. *Sensors and Materials*, Vol. 26, No. 3, pp. 131–136, 2014.
- [5] Muñoz-Aguirre, S., Nakamoto, T., Moriizumi, T. Study of deposition of gas sensing films on quartz crystal microbalance using an ultrasonic atomizer. *Sensors and Actuators B*. 105, pp. 144-149, 2005.
- [6] Muñoz-Mata, J. L., Muñoz-Aguirre, S., etal. Development and Implementation of a System to Measure The Response of Quartz Crystal Resonator Based Gas Sensor Using a Field Programmable Gate Array. *Measurement Science and Technology* Vol. 35 No. 5 United Kingdom, 2012.
- [7] Stehrer, B. P., Schwödiauer, B. S., Graz, I.M., Pollheimer, P.D., Gruber, H.J. High Frequency QCM Based Sensor System for Sensitive Detection of Dissolved Analytes. *Procedia Engineering* 5, pp. 835-837, 2010.
- [8] Nakamoto, T., Nakamura, K., etal. Study of Oscillator-Circuit Behavior for QCM Gas Sensor. *IEEE Ultrasonics Symposium*, pp. 351-354, 1996.
- [9] Nakamoto, T., Suzuki, Y., Moriizumi, T. Study of VHF-band QCM Gas Sensor. *Sensors and Actuators B* 84, pp.98-105, 2002.

SIMULACIÓN DE ESTRATEGIAS DE BÚSQUEDA EN ANIMALES CON POSIBLES APLICACIONES EN COMPUTACIÓN Y ROBÓTICA

Joel Ricardo Jiménez Cruz

Universidad Autónoma Metropolitana-Iztapalapa

jcjr@xanum.uam.mx

Resumen

En este trabajo se describen y se simulan por medio del lenguaje de programación Netlogo, estrategias de búsqueda que emplean los animales para el forrajeo (recolección de recursos) en su hábitat natural. Se realizaron dos simulaciones, en la primera se simulan las estrategias de movimiento kinésicas (ortokinesis, klinokinesis y klinokinesis adaptativa) y en la segunda, las estrategias Wiggle y Saltatory. El movimiento de los organismos puede determinar una cierta distribución espacial y temporal de las poblaciones en función de la adaptación a su medio ambiente. Estas estrategias de navegación y búsqueda pueden servir de inspiración para implementar algoritmos de forrajeo en agentes computacionales y robots. Por ejemplo, unos nanorobots podrían reparar y mejorar algunas partes del cuerpo humano. Aunque la aplicación para cada una de las estrategias depende del contexto, se observó en los experimentos realizados que klinokinesis adaptativa y Saltatory son estrategias de búsqueda óptimas.

Palabras Claves: Búsqueda, forrajeo, kinesis, navegación, Netlogo.

Abstract

In this work, the search strategies used by animals in foraging (resource collection) in their natural habitat are described and simulated using the Netlogo programming language. Two simulations were performed, in the first one, the kinetic movement strategies (orthokinesis, klinokinesis and adaptive klinokinesis) were simulated and in the second, the Wiggle and Saltatory strategies. The movement of organisms can determine a certain spatial and temporal distribution

of populations depending on the adaptation to their environment. These navigation and search strategies can be used as an inspiration to implement foraging algorithms in computational agents and robots. For example, nanorobots could repair and improve some parts of the human body. Although the application for each of the strategies depends on the context, it was observed in the experiments carried out that adaptive klinokinesis and Saltatory are optimal search strategies.

Keywords: *Foraging, kinesis, navigation, netlogo, search.*

1. Introducción

El movimiento es un proceso biológico importante, presente en todos los organismos y con consecuencias para los individuos, las poblaciones, las especies y las comunidades biológicas. A través del movimiento, los organismos pueden localizar alimento, compañeros, refugio, un lugar donde vivir, etc., o pueden evitar depredadores o condiciones peligrosas. Tales factores afectan la vida de todo tipo de organismos que van desde bacterias, virus y otros organismos "simples" a organismos multicelulares más complejos y a una amplia variedad de grupos de animales [Pyke, 2015].

Generalmente los nutrientes se encuentran esparcidos temporal y espacialmente en el medio ambiente y los organismos deben implementar estrategias de búsqueda que les permitan obtener, por ejemplo, los recursos alimenticios en forma eficiente. A su vez, el movimiento de los organismos determinar la distribución espacio-temporal las poblaciones y especies, y por lo tanto, en última instancia, los patrones espaciales exhibidos por las comunidades biológicas.

El estudio y simulación de las bioestrategias de navegación y búsqueda son útiles para conocer y entender el comportamiento de los animales y con la posibilidad de aplicarlas en el desarrollo de algoritmos inteligentes en artefactos computacionales o robóticos.

Para la navegación y búsqueda de agentes o robots hay algoritmos clásicos de la Inteligencia Artificial como el algoritmo A* [Fernández, 2005] y otros basados en la Cibernética que están inspirados en la biología. En este trabajo se reportan bioestrategias de navegación y búsqueda simples basadas en la kinesis. En

particular se describen y simulan las estrategias de orthokinesis (orientación basada en la velocidad), klinokinesis (orientación basada en el ángulo de giro) y klinokinesis adaptativa (giro en función de la concentración), Saltatory (movimiento intermitente) y Wiggle (movimiento oscilatorio). La simulación de estos mecanismos se lleva a cabo en Netlogo. Netlogo es un lenguaje de programación multiagente que permite la exploración de fenómenos emergentes de una gran variedad de campos del conocimiento [Poza, 2009].

Uno de los problemas fundamentales de supervivencia de los organismos es el hallazgo, captura, consumo y utilización de fuentes de energía. Estas fuentes tienen un determinado valor nutritivo y se distribuyen en forma espacial y temporal y son limitadas en cantidad, representando un costo energético. Debido a que un posible consumidor tiene un tiempo y energía limitados, éste tiene que tomar decisiones de diversos tipos que pueden afectar su supervivencia. Existen algunas estrategias generales que han sido adoptadas por diversas especies, que permiten resolver eficientemente los problemas de forrajeo, siempre y cuando exista un balance entre el gasto de energía y la ganancia de ella [Gutiérrez, 1998].

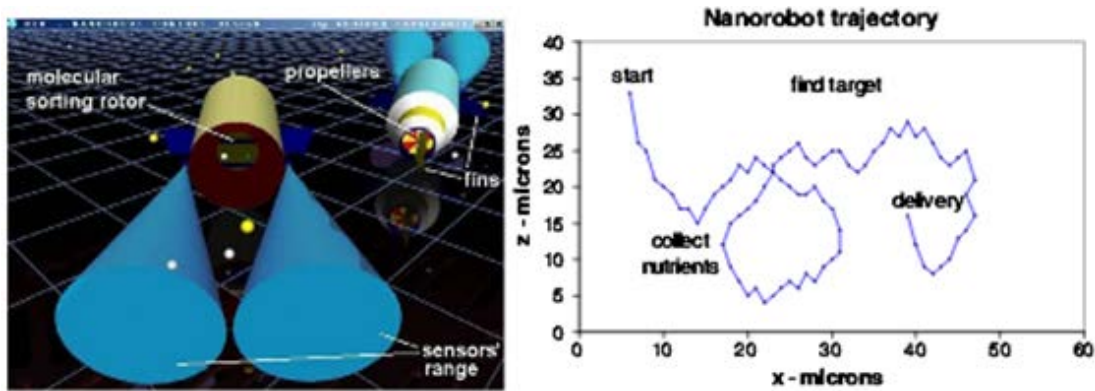
La comprensión de estas estrategias movimiento, búsqueda y desplazamiento que siguen los animales en su hábitat natural nos pueden permitir diseñar aplicaciones artificiales como en el movimiento y la búsqueda inteligente en agentes y robots.

Sus aplicaciones en agentes pueden encontrarse en la búsqueda de artefactos en Internet y en robótica pueden utilizarse para encontrar y recolectar objetos, por ejemplo, los nanorobots podrían reparar y mejorar el cuerpo humano [Jones, 2008]. En la figura 1 se muestran nanorobots virtuales equipados con rotores moleculares, aletas, propelas y sensores que siguen una trayectoria para la búsqueda de un objetivo [Calvalcanti et al, 2008].

En la figura 2 se muestra la imagen de un nanorobot futurista, diseñado por [Svidinenko 2015], el cual realiza el monitoreo, localización y posible cirugía celular en el interior de una arteria con flujo sanguíneo.

En la siguiente sección se describen las características de la navegación y estrategias kinésicas, Saltatory y Wiggle que realizan los animales en ambientes dinámicos.

En la tercera sección se estudia una simulación de estas estrategias. En las últimas secciones se realizan la discusión y las conclusiones del trabajo.



[Tomada de Calvacanti et al, 2008]

Figura 1 Nanorobots virtuales y trayectoria de un nanorobot que busca un objetivo.



Figura 2 Nanorobots en el flujo sanguíneo para cirugía celular [Svidinenko, 2015].

2. Métodos

En esta sección se explica lo que se programó y simuló en Netlogo, basándose en los fundamentos de la navegación y búsqueda de los animales en medios ambientes dinámicos.

La navegación de los animales en entornos dinámicos tiene varios propósitos, entre ellos el de forrajeo (recolección de recursos) y la búsqueda de su nido, de una pareja o de un sitio específico. Se han identificado 3 estrategias o mecanismos de búsqueda individual y 3 patrones poblacionales generados por la distribución de la comida [Mueller, 2008].

Las estrategias individuales de movimiento se pueden clasificar en:

- Mecanismos sin orientación.

- Mecanismos con orientación.
- Mecanismos de memoria.

Los mecanismos sin orientación implican movimientos como la difusión y la kinesis que parecen movimientos aleatorios. Con estos mecanismos, los estímulos externos provocan una alteración en el ángulo de giro, la velocidad o la frecuencia del movimiento. Matemáticamente se pueden representar como paseos aleatorios correlacionados.

Los mecanismos con orientación se basan en la percepción de señales alejadas de la posición del animal y provocan un movimiento predecible que les permite a los animales acercarse a los recursos o a la ubicación de su destino. No se conocen bien los rangos de percepción y si estos operan a escalas temporales o espaciales en función de los cambios de disponibilidad de los recursos.

En los mecanismos basados en una memoria existe información acerca de la ubicación del objetivo. Esta información proviene de la experiencia, comunicación con sus congéneres o la genética. Estos mecanismos se basan en la integración de una ruta por medio de una brújula o por medio de mapas cognitivos que se obtienen a base de puntos de referencia conocidos. Las aves utilizan señales celestes, olfativas, coordenadas geomagnéticas y lugares de referencia que facilitan su navegación y regreso a su nido. Para los herbívoros, la investigación se ha centrado en el aprendizaje espacial de la localización de los recursos.

Para entender cómo se mueven los organismos, se requiere de un enfoque que compare estos tres mecanismos. Algunos modelos combinan la memoria con la información espacial y social y la evasión de depredadores.

La distribución dinámica de los recursos provoca que las estrategias de movimiento inmersas en la navegación individual de un grupo de individuos generen 3 distribuciones o patrones poblacionales:

- Sedentarismo. Se da con recursos que tienen poca variabilidad. El sedentarismo comprende estrategias en que los residentes se encuentran en un territorio establecido y cuando los recursos se encuentran disponibles durante mucho tiempo.

- Migraciones. Se produce con recursos de variabilidad estacional. La migración se define como un patrón de movimiento de distancias grandes y se observa en variaciones estacionales regulares. Son de naturaleza periódica y temporal dependiendo de su época de reproducción.
- Patrones nómadas. Los recursos aparecen con una distribución impredecible. El nomadismo se presenta como un patrón de movimiento que no se repite durante un periodo largo. Se produce cuando los recursos fluctúan de manera irregular o impredecible en grandes áreas geográficas.

La estructura del territorio es un factor importante para determinar la eficiencia de los mecanismos de movimiento y los patrones de distribución de la población. La disponibilidad de recursos puede deberse a su configuración espacial, variabilidad temporal y previsibilidad.

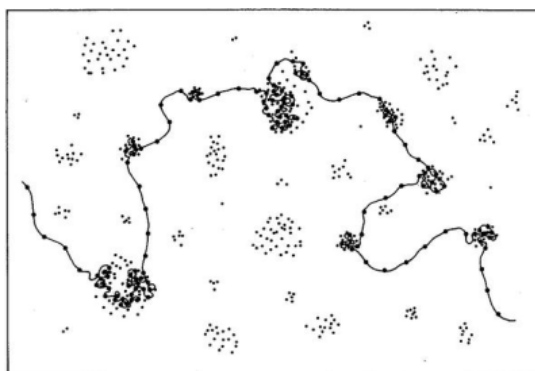
La distribución de los recursos afecta la eficacia de los mecanismos de movimiento y búsqueda. En los territorios con poca variabilidad, la memoria es importante. Una experiencia previa es una fuente importante de información para movimientos futuros.

De estos mecanismos de navegación y búsqueda, en este trabajo, nos enfocamos en simular inicialmente los mecanismos sin orientación que involucran movimientos simples como la difusión (concentración) y la kinesis (gradientes) y que dan como resultado una decisión de movimiento. En particular estudiamos y simulamos las siguientes estrategias simples de movimiento: orthokinesis, klinokinesis y klinokinesis adaptativa. También se simularon las estrategias saltarory (intermitente) y Wiggle (oscilatoria).

La ortokinesis, klinokinesis y klinokinesis adaptativa son modelos probabilísticos que corresponden a reacciones motoras elementales respecto a la posición de los estímulos y que regulan el cambio de movimiento del animal en base de su velocidad en el caso de la ortokinesis y de la dirección en el caso de la klinokinesis. Si no existe una adaptación (mecanismo absoluto) la regulación es una función del valor real de la intensidad del estímulo y cuando el animal se adapta (mecanismo diferencial), la regulación es una función de las variaciones en

la intensidad del estímulo que son percibidas durante el movimiento, como en el caso de la klinokinesis adaptativa [Benhamou, 1989].

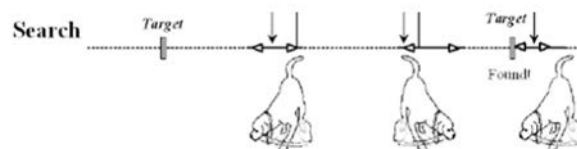
En la figura 3 se muestran los movimientos de forrajeo de un animal en un entorno de parcelas. Los puntos pequeños representan la posición de las presas y los puntos más grandes la posición del animal. Se observa que los mecanismos ortokinésicos y klinokinésicos producen una conducta de búsqueda en un área determinada que les permite a los animales permanecer por más tiempo en las regiones de un alto contenido de presas. Entre las simulaciones que se muestran en la sección de resultados, se reproducen los mecanismos que se observan en esta figura.



[Tomada de Benhamou, 1989]

Figura 3 Movimientos de forrajeo ortokinésicos y klinokinésicos de un animal en un entorno de parcelas.

Por otro lado, en este trabajo también se programaron en Netlogo las estrategias de búsqueda Saltatory y Wiggle que son mecanismos de movimiento que han sido observados en diferentes organismos para encontrar recursos. En la estrategia de movimiento Wiggle, el animal realiza rutinas de giros oscilatorios y redundantes sobre pequeñas áreas de búsqueda y durante un periodo suficiente de tiempo que le permite escanear el medio de manera sensitiva. En la estrategia de movimiento Saltatory se extiende el movimiento Wiggle integrando un movimiento de desplazamiento después de una rutina de oscilación (Wiggle) y repitiendo la serie, figura 4. En la estrategia Saltatory, los animales buscan de una manera intermitente: avanzan, hacen una breve pausa y avanzan otra vez [Anderson et al, 1997].



[Tomada de Anderson et al, 1997]

Figura 4 Movimiento Saltatory.

3. Resultados

En esta sección se presentan los resultados de las dos simulaciones que se realizaron. Por una parte se simularon las estrategias de movimiento kinésicas y por otro, las estrategias Wiggle y Saltatory. En ambas se pretende encontrar cuál es la mejor estrategia de movimiento de acuerdo a la distribución de los recursos en el medio ambiente.

Simulaciones de las Estrategias Kinésicas

Se realizaron 4 simulaciones que comparan las estrategias de deambular azarosamente (wander), klinokinesis, ortokinesis y klinokinesis adaptativa [Nelson, 2017]. En el programa implementado en Netlogo se van escogiendo 4 escenarios de distribución de recursos: aleatorio (random), en un agrupamiento (cluster), en franja como en un fluido (stream) o en varios agrupamientos (clusters). Este último nos permitirá simular la figura 3 del artículo de Benhamou y Bovet [1989].

En cada una de las simulaciones el programa se corre 5 veces con un tiempo de 5000 ticks para cada uno de los escenarios (aleatorio, cluster, fluido y clusters) y se anota en cada uno de ellos el número de porciones de comidas (pellets) que el agente consumió.

Simulación 1. El agente Deambula (wander) en Diferentes Escenarios

En la figura 5 se muestra la corrida en los 4 escenarios cuando el agente deambula en su medio ambiente. En el escenario aleatorio (random) los recursos se distribuyen en forma estocástica mientras que en el agrupado se da una concentración aleatoria circular en la parte central del territorio. En el escenario como una franja los recursos están esparcidos en forma de un río. Cuando hay

varias conglomeraciones o parcelas de recursos, los agentes ajustan sus estrategias de búsqueda.

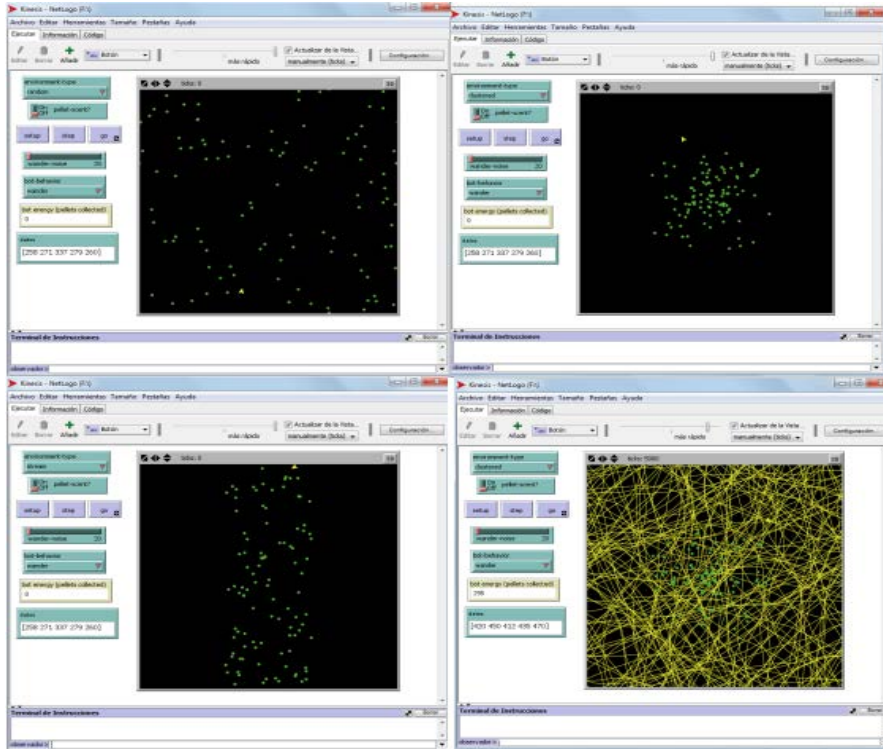


Figura 5 Comportamiento del agente que deambula en los cuatro escenarios.

En la tabla 1 se muestran cantidades de pellets consumidos utilizando la estrategia de movimiento azaroso en cada uno de los escenarios y su promedio y desviación estándar.

Tabla 1 Pellets consumidos en cada uno de los medios ambientes, cuando el agente utiliza el movimiento de deambular.

Escenario	Número de pellets	Promedio, \pm des.est
Aleatorio (random)	[356 348 320 326 356]	341 \pm 8
Agrupado (cluster)	[296 291 223 276 307]	279 \pm 15
Fluido (stream)	[258 271 337 279 260]	281 \pm 15
Varios grupos (clusters)	[111 104 84 109 113]	104 \pm 5

Utilizando un movimiento errático, se puede observar que en el escenario aleatorio (random) se obtiene una mayor recolección de alimento. Es decir, en un ambiente

donde los recursos se hayan distribuidos aleatoriamente, la mejor estrategia a utilizar es el vagar o deambular por el ambiente.

Simulación 2. Klinokinesis

En la figura 6 se muestra la corrida en los 4 escenarios cuando el agente utiliza la estrategia de klinokinesis en su medio ambiente.

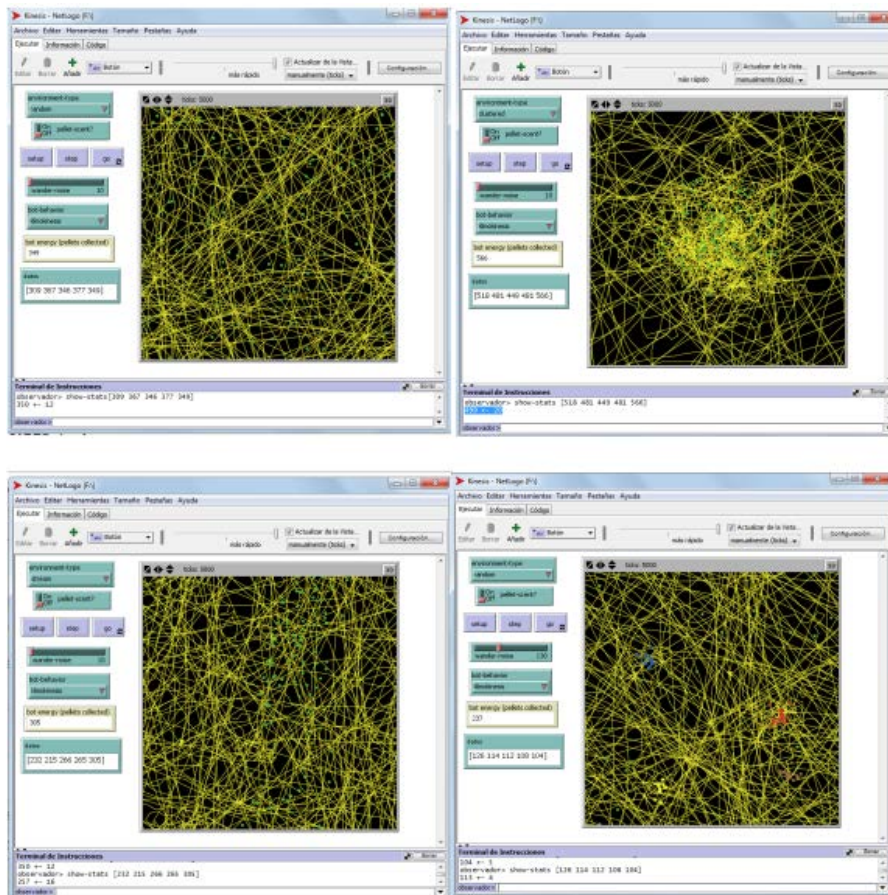


Figura 6 Aplicación de la estrategia de klinokinesis en cuatro diferentes ambientes.

En la tabla 2 se muestran cantidades de pellets consumidos utilizando la estrategia de movimiento de klinokinesis en cada uno de los escenarios y su promedio y desviación estándar. En esta simulación se observa que la estrategia de klinokinesis es útil cuando los recursos se encuentran distribuidos en una parcela o en varias de ellas.

Tabla 2 Pellets consumidos por el animal utilizando la estrategia de klinokinesis

Escenario	Número de pellets	promedio, ±
Aleatorio (random)	[309 367 346 377 349]	350 ± 12
Agrupado (cluster)	[126 114 112 108 104]	113 ± 4
Fluido (stream)	[232 215 266 265 305]	257 ± 16
arios grupos (clusters)	[518 481 449 481 566]	499 ± 20

Simulación 3. Ortokinesis

En la figura 7 se muestra la corrida en los 4 escenarios cuando el agente emplea el mecanismo de ortokinesis en su medio ambiente.

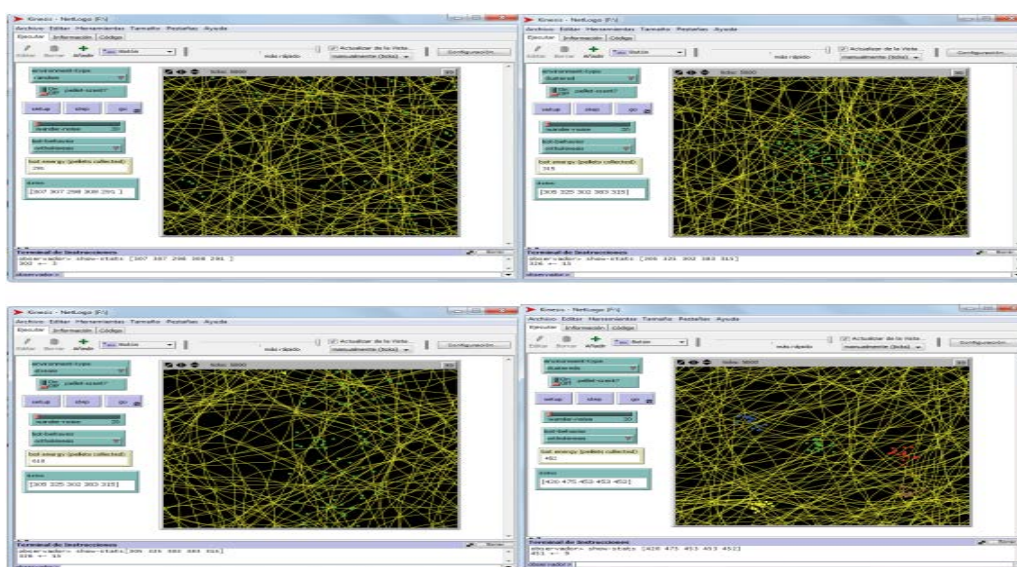


Figura 7 Aplicación de la estrategia de ortokinesis en 4 medios ambientes.

En la tabla 3 se muestran cantidades de pellets consumidos utilizando la estrategia de movimiento de ortokinesis en cada uno de los escenarios y su promedio y desviación estándar.

Tabla 3 Cantidad de recursos obtenidos en los 4 escenarios utilizando ortokinesis.

Escenario	Número de pellets	promedio, ± des.est
Aleatorio (random)	[307 307 298 308 291]	302 ± 3
Agrupado (cluster)	[420 475 453 453 452]	451 ± 9
Fluido (stream)	[305 325 302 383 315]	326 ± 15
Varios grupos (clusters)	[305 325 302 383 315]	326 ± 15

Con los valores obtenidos en la tabla 3 se observa que con la estrategia de ortokinesis los agentes obtienen una mayor recolección recursos en un ambiente agrupado.

Simulación 4. klinokinesis Adaptativa ('run-tumble': voltereta)

En la figura 8 se muestra la corrida en los 4 escenarios cuando el agente realiza la estrategia de klinokinesis adaptativa en su medio ambiente. Esta estrategia se puede observar en las bacterias.

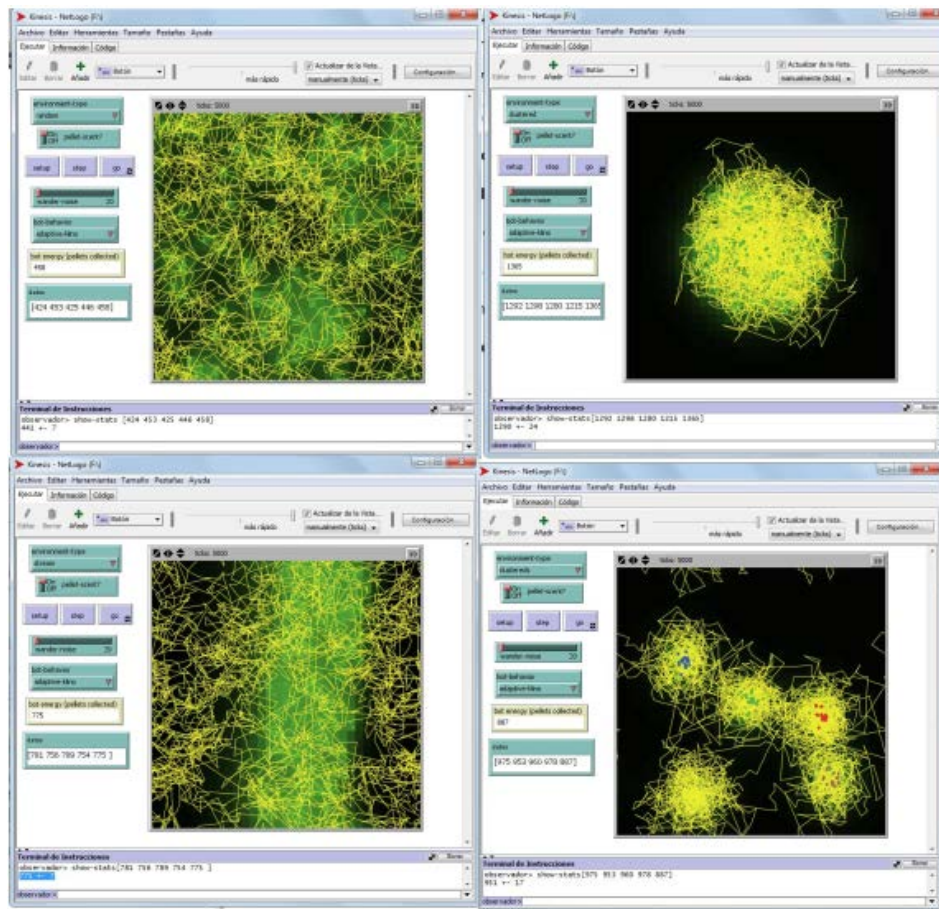


Figura 8 Aplicación del mecanismo de Klinokinesis adaptativa en 4 diferentes entornos.

En la tabla 4 se muestran cantidades de pellets consumidos utilizando la estrategia de movimiento de klinokinesis adaptativa en cada uno de los escenarios y su promedio y desviación estándar.

Tabla 4 Cantidad de pellets recogidos utilizando la estrategia de klinokinesis adaptativa.

Escenario	Número de pellets	promedio, \pm des. est
Aleatorio (random)	[424 453 425 446 458]	441 \pm 7
Agrupado (cluster)	[975 953 960 978 887]	951 \pm 17
Fluido (stream)	[781 756 789 754 775]	771 \pm 7
Varios grupos (clusters)	[1292 1298 1280 1215 1365]	1290 \pm 24

En la tabla 4 se puede determinar que la estrategia de klinokinesis adaptativa es particularmente útil cuando los recursos se encuentran en uno o varios amontonamientos. También se puede apreciar que esta estrategia es útil cuando los recursos se hayan esparcidos en una franja o fluido.

En la tabla 5 se muestra un resumen de la cantidad de pellets recogidos en las simulaciones realizadas de acuerdo a las estrategias y a los escenarios utilizados. Se observa que en todos los escenarios, la mejor estrategia sería la klinokinesis adaptativa.

Tabla 5 Pellets recogidos utilizando los diversos escenarios y técnicas de recolección.

entorno	aleatorio	flujo	cluster	clusters
vagar	341 \pm 8	281 \pm 15	104 \pm 5	279 \pm 15
klinokinesis	350 \pm 12	257 \pm 16	113 \pm 4	499 \pm 20
orthokinesis	302 \pm 3	326 \pm 15	451 \pm 9	326 \pm 15
klino adaptiva	441 \pm 7	771 \pm 7	951 \pm 17	1290 \pm 24

Simulaciones de las Estrategias Wiggle y Saltatory

En esta sección se presentan las simulaciones que comparan las estrategias Wiggle y Saltatory en una situación médica donde un grupo de nanorobots tratan de encontrar productores cancerígenos dentro de los vasos sanguíneos. Se simula en el medio ambiente del vaso sanguíneo los glóbulos rojos, los nanorobots y elementos productores de una sustancia química (e-cadherin) asociada con la producción de cáncer. La función de los nanorobots es encontrar a estos productores con el fin de eliminarlos [Cuevas, 2008].

En la simulación intervienen los siguientes elementos:

- Ambiente: Simulación del torrente sanguíneo en el cual interactúan los productores, glóbulos y buscadores.

- **Productores:** Representan al productor de e-cadherin que se encuentra en el vaso.
- **Glóbulos:** Glóbulos rojos que forman parte del torrente del vaso sanguíneo.
- **Buscadores:** Nanorobots de búsqueda que son insertados dentro del vaso sanguíneo.
- **Spinners:** Reloj que lleva un control del tiempo (en tick's) que transcurrirá durante las diferentes pruebas, en las cuales los buscadores localizan al productos.

En Netlogo estos elementos quedan representados de la siguiente manera:

- breed [productores productor]
- breed [globulos globulo]
- breed [buscadores buscador]
- breed [spinners spinner]

El movimiento de la estrategia wiggle en Netlogo quedó definido con el procedimiento: Wiggle {right random 15 forward .0005 left random 15 forward .0005}. En la figura 9 se muestra una corrida de la simulación con el movimiento Wiggle.

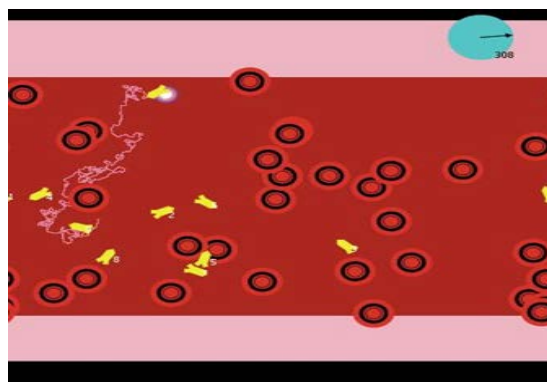


Figura 9 Simulación de la estrategia de movimiento Wiggle implementada en nanorobots dentro de un vaso sanguíneo.

Para implementar el movimiento Saltatory se modifica ligeramente el comportamiento oscilatorio (Wiggle), agregando un pequeño desplazamiento

considerando que los sensores del nanorobot solo perciben el área o patch en la que se encuentran.

La simulación del movimiento de la estrategia Saltatory quedó definida con el procedimiento: Saltatory { forward .05 Wiggle}. En la figura 10 se muestra una corrida de su simulación.

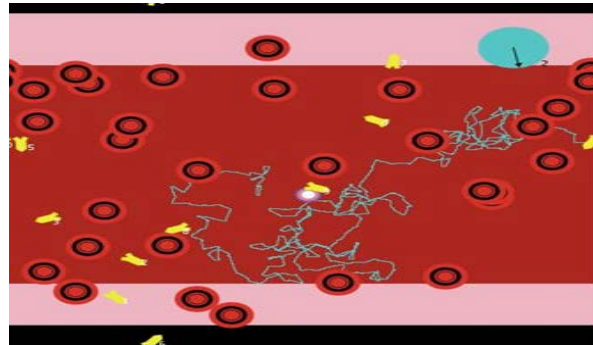


Figura 10 Simulación de nanorobots dentro de un vaso sanguíneo, estrategia de movimiento Saltatory.

En Piña, Rechy y García [2008] se presenta un refinamiento de esta simulación agregando parámetros para la densidad y viscosidad del fluido, comparando las estrategias Wiggle, Saltatory y deambular.

4. Discusión

La detección de recursos distribuidos aleatoriamente en el espacio es una tarea común tanto en agentes y robots como en sistemas biológicos.

Si bien los algoritmos que se han presentado muestran un grado adecuado de resolución de problemas, se puede discutir que existen otros algoritmos sin inspiración biológica que dan iguales o mejores resultados. Lo que sucede es que los algoritmos biológicos buscan y tienen generalmente un mayor grado de adaptabilidad cuando suceden cambios inesperados.

En las corridas que se hicieron de las simulaciones se observó que los mecanismos aleatorios realizan bien su trabajo y son relativamente más simples de implementar y menos costosos computacionalmente hablando. Es por ello que se pregunta uno si vale la pena indagar y construir otros algoritmos más

sofisticados. Los mecanismos aleatorios les han servido a los animales para sobrevivir en diferentes entornos y bajo muy diversas circunstancias. Aun así, han especializado sus mecanismos y estrategias cuando se requiere más rapidez y eficacia en la búsqueda.

Uno de los problemas principales en la implementación de algoritmos de búsqueda con inspiración biológica es la construcción de sensores que imiten a lo biológico. Entre mayor sea la inteligencia que se quiera implementar en un artefacto artificial, mayor debe ser la similitud biológica de los sensores.

5. Conclusiones

En este trabajo se han mostrado las simulaciones en Netlogo de las bioestrategias de movimiento y búsqueda que utilizan los organismos y que se podrían implementar en el caso de agentes computacionales o robots.

Se pudo comprobar que en un ambiente donde los recursos se hayan distribuidos aleatoriamente, la mejor estrategia a utilizar es el vagar o deambular por el ambiente. Se observó que los mecanismos ortokinésicos y klinokinésicos permiten a los animales permanecer por más tiempo en las regiones de un alto contenido de recursos. Klinokinesis es útil cuando los recursos se encuentran distribuidos en una parcela o en varias de ellas, mientras que con ortokinesis se obtiene una mayor recolección recursos en un ambiente agrupado. La estrategia Wiggle permite obtener una mejor percepción local mientras que Saltatory es más rápida y permiten ir recorriendo y abarcando una mayor área de búsqueda.

Aunque la aplicación para cada una de las estrategias depende del contexto, se observó en términos generales en los experimentos realizados que klinokinesis adaptativa y Saltatory son estrategias de búsqueda óptimas.

Hay una serie de algoritmos en los que se puede continuar esta investigación y que consisten en utilizar estrategias intermitentes de búsqueda como Saltatory que combinan fases de movimiento lento, con fases de movimiento rápido, minimizando el tiempo de búsqueda [Bénichou, 2011].

También se puede explorar estrategias metaheurísticas que buscan un equilibrio entre la exploración y la explotación y que en ciertos casos son mejores que la

estrategia de búsqueda intermitente [Xin-She, 2013]. En otros casos, se pueden utilizar estrategias de búsqueda aleatorias combinadas que generan predicciones sobre cómo deben comportarse el agente en cada modo de búsqueda y cuándo debe cambiar el método de búsqueda. Varios organismos utilizan señales sensoriales no direccionales sin precisión para mejorar la búsqueda aleatoria y la teoría tradicional del forrajeo de parcelas [Noltinga, 2015].

Algunas aplicaciones de estos algoritmos se encuentran en robots autónomos que basan sus estrategias de búsqueda en las caminatas aleatorias de Lévy [Krivonosov, 2016]. Otra aplicación sería en redes complejas como en Internet, el tráfico urbano y el cerebro, donde se estudian estrategias de búsqueda en donde el agente (por ejemplo, un impulso eléctrico, una excitación, un animal o un individuo humano, tal como un surfista de la web), situado en un nodo de la red puede saltar a un nodo vecino, siempre que exista un enlace como se especifica en la matriz de adyacencia asociada con el gráfico. El caminante se desliza por la red a través de una secuencia de pasos, que permiten una exploración y búsqueda [Di Patti, 2015].

6. Bibliografía y Referencias

- [1] Anderson, J. Stephens, D., and Dunbar S., Saltatory search: a theoretical analysis, *Behavioral Ecology*, Vol 8 No. 3: pp. 307-317, 1997. https://www.researchgate.net/publication/249282814_Saltatory_search_A_theoretical_analysis.
- [2] Bénichou, O., Loverdo, C., Moreau, M., Voituriez, R., Intermittent search strategies. *Review of modern physics*, volume 83, pp. 81-129, 2011, <http://www.normalesup.org/~loverdo/art/RevModPhys.pdf>.
- [3] Cuevas, M., Comparación de dos Estrategias bioinspiradas de búsqueda de objetivos aplicadas a la navegación nanorobótica. Caso de Estudio: Prevención del Cáncer. Tesis para obtener el título de Licenciado en Sistemas Computacionales Administrativos, Facultad de Contaduría y Administración, Universidad Veracruzana, Agosto 2008, <http://cdigital.uv.mx/bitstream/123456789/28492/1/Cuevas%20Perez.pdf>.

- [4] Benhamou, S., Bovet, P., How animals use their environment: a new look at kinesis. *Anim. Behav.*, 38: pp. 375-383, 1989. http://www.cefe.cnrs.fr/images/stories/DPTEEvolution/Ecomportementale/chercheurs/simon_benhamou/Benhamou%20&%20Bovet%201989AB.pdf.
- [5] Cavalcanti, A., Shirinzadeh B., Freitas, R. A. Hogg, T., Nanorobot architecture for medical target identification. *Nano-technology* 19: 015103, 2008, <http://www.nanorobotdesign.com/papers/nanorobot-architecture.pdf>.
- [6] Di Patti, F., Fanelli, D., Piazza, F. (2015). Optimal search strategies on complex multi-linked networks. *Nature Scientific Reports* 5, number: 9869, pp. 1-6. https://www.nature.com/articles/srep09869?WT.ec_id=SREP-639-20150512&spMailingID=48636705&spUserID=MzcxNDE0MDA3NzMS1&spJobID=681447381&spReportId=NjgxNDQ3MzgxS0.
- [7] Fernández, M., Algoritmos de búsqueda heurística en tiempo real. Aplicación a la navegación en los juegos de video." 34 Jornadas Argentinas de Informática e Investigación Operativa, 2005, <http://www.exa.unicen.edu.ar/catedras/aydalgo2/docs/TFca06aCompleto.pdf>.
- [8] Krivonosov, M., Denisov, S., & Zaburdaev, V. (2016). Lévy robotics. arXiv preprint arXiv:1612.03997, pp. 1–6, 2016, <https://arxiv.org/pdf/1612.03997.pdf>.
- [9] Pyke, G. (2015). Understanding movements of organisms: it's time to abandon the Lévy foraging hypothesis. *Methods in Ecology and Evolution*. Volume 6, Issue 1, Pages 1–16. Pages 1–16. Consultado el 31-05-17. <http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12298/full>
- [10] Gutiérrez, G. (1998). Estrategias de forrajeo. En R. Ardila, W. López, A.M. Pérez, R. Quiñones, & F. Reyes (Eds.). *Manual de Análisis Experimental del Comportamiento*, pp. 359-381., 1998, Madrid: Librería Nueva, <http://www.docentes.unal.edu.co/gahermannr/docs/1998%20Estrategias%20de%20Forrajeo.pdf>.
- [11] Mueller, T., Fagan, W., Search and navigation in dynamic environments from individual behaviors to population distributions. *Oikos* 117, pp. 654-664: <http://www.clfs.umd.edu/biology/faganlab/pdf/MuellerFagan2008.pdf>.

- [12] Jones, R., Rupturing The Nanotech Rapture. Biological nanobots could repair and improve the human body, but they'll be more bio than bot. IEEE Spectrum's Special Report: The Singularity, 2008, <http://spectrum.ieee.org/semiconductors/nanotechnology/rupturing-the-nanotech-rapture/0>.
- [13] Nelson, M. (2017). Homework: week04_hw_template (kinesis strategies). Brain, Behavior & Info Processing Course, 2017, <http://output.jsbin.com/vecixe>.
- [14] Noltinga, B., Hinkelmanb, T., Brassilc, C., Tenhumberg, B., Composite random search strategies based on non-directional sensory cues. *Ecological Complexity*, Volume 22, pp. 126-138, 2015.
- [15] Piña, C., Rechy, E., García V., Comparing Three Simulated Strategies for Cancer Monitoring with Nanorobots. En A. Gelbukh and E.F. Morales (Eds.): MICAI 2008, LNAI 5317, Springer-Verlag Berlin Heidelberg, pp. 1020–1030, 2008, http://link.springer.com/chapter/10.1007%2F978-3-540-88636-5_96.
- [16] Poza, D., Manual de Netlogo en español, 2009, <https://sites.google.com/site/manualnetlogo/>.
- [17] Svidinenko, Y., Simulation of a simple mobile cell-repair nanorobot and some of it's subsystems. *Nanotech*, 2015, http://www.nanotech-now.com/Art_Gallery/svidinenko-yuriy.htm.
- [18] Xin-She Yang and Xing shi He, Firefly Algorithm: Recent Advances and Applications, *Int. J. Swarm Intelligence*, Vol. 1, No. 1, pp. 36–50, 2013, https://www.researchgate.net/publication/255971821_Firefly_Algorithm_Rece nt_Advances_and_Applications.

EMULACIÓN EN FPGA DE TECNICA PARA CORRECCION DEL DESEQUILIBRIO I/Q APLICADO EN UN MODULADOR DIGITAL 256-QAM

Sergio Alberto Juárez Cazares

Instituto Politécnico Nacional, IPN-CITEDI

sjuarez@citedi.mx

Aldo Bonilla Rodríguez

Instituto Politécnico Nacional, IPN-UPIITA

aldo.bonilla.r@gmail.com

José Cruz Núñez Pérez

Instituto Politécnico Nacional, IPN-CITEDI

nunez@citedi.mx

Resumen

En este trabajo se presenta la metodología de diseño e implementación de un sistema para corrección del desequilibrio I/Q basado en una tarjeta DSP-FPGA. Este sistema utiliza series de Volterra para modelar el comportamiento no lineal del desequilibrio I/Q. El desempeño del sistema se demuestra utilizando una señal compleja 256-QAM con desequilibrio en fase y amplitud. El sistema desarrollado tiene como característica un bajo costo de implementación y alta flexibilidad del diseño, lo que permite modificaciones o expansiones futuras. Se utiliza la tarjeta de desarrollo Stratix III de Altera para implementación práctica y verificación de los resultados experimentales del sistema propuesto. El sistema desarrollado es capaz de corregir el desequilibrio I/Q satisfactoriamente, tanto en fase como en amplitud. Este trabajo puede ser considerado como una alternativa de bajo costo para corregir el desequilibrio I/Q ya que no requiere de algoritmos complejos o equipo de medición adicional.

Palabras Claves: 256-QAM, corrección, desequilibrio, FPGA, modulador.

Abstract

In this paper the design methodology for a I/Q imbalance correction system is presented based on a DSP-FPGA board. This system employs Volterra series to model the non-linear behavior of the I/Q imbalance. The system performance is verified using a 256-QAM complex signal with phase and amplitude imbalance. The implemented system has the advantage of having low implementation cost and a high design flexibility, which allows for future revisions or enhancement. The Stratix III FPGA board from Altera is employed for the practical implementation and result verification of the system. The developed system can compensate the I/Q imbalance, in amplitude and phase. This work can be considered as a low-cost alternative for I/Q imbalance correction given that it doesn't require additional measurement equipment nor uses complex algorithms.

Keywords: 256-QAM, correction, imbalance, FPGA, modulator.

1. Introducción

Actualmente, la industria de las telecomunicaciones demanda sistemas de modulación digital con una alta confiabilidad. Los avances en manufactura han permitido el desarrollo de circuitos integrados de alta densidad, haciendo posible la integración de sistemas digitales para telecomunicaciones de alto desempeño en forma de circuitos de aplicación específica (del inglés ASIC), System on Chip (SoC), entre otros [Dick, 2004]. En contexto, los dispositivos FPGA modernos poseen ventajas significativas para implementación en hardware, ofreciendo así una alternativa que permite una extensa verificación del diseño y así evitar implementaciones de diseños poco confiables.

Los sistemas de modulación digital disponibles en el mercado deben contar con una baja producción de espurias, de lo contrario se provocarán errores en el espectro de la señal tanto en el transmisor como en el receptor ocasionando alta tasa de error de bit (del inglés BER). Estas imperfecciones son más significativas en sistemas de comunicaciones de banda ancha [Erdogan, 2008] y en sistemas con modulación compleja causan el desequilibrio I/Q. El desequilibrio I/Q se produce cuando existe un desbalance en fase y amplitud de los canales I y Q. En

el estado del arte se reportan diversos trabajos de medición y compensación del desequilibrio de I/Q [Erdogan, 2008].

El método descrito en [Asami, 2007] proporciona una metodología para corregir la cuadratura utilizando un detector de envolvente en el convertidor de bajada. Sin embargo, necesita una estimación previa de las imperfecciones y posee un alto costo computacional ya que utiliza un algoritmo iterativo. En [Cavers, 1997] se presenta un algoritmo sencillo que permite cancelar las espurias producidas en el oscilador local (del inglés LO). No obstante, requiere de un analizador de espectro lo cual se traduce en gastos adicionales en hardware. En [Nash, 2009] se propone una metodología de calibración en moduladores QAM, se explora la presencia de un bloque extra en la retroalimentación para corregir el desbalance I/Q.

Actualmente, el mercado no cuenta con soluciones directas para corregir el desequilibrio I/Q. Sin embargo, existen diversas soluciones para este problema donde se utilizan métodos de calibración para la etapa de transición y recepción, como se detalla en [Nash, 2009].

En síntesis, la tendencia actual es compensar el desequilibrio I/Q haciendo uso de sistemas digitales [Snelgrove, 1999], [Anttila, 2007], [Anttila, 2009], [Mattera, 2008]. Es por esto que utilizar dispositivos FPGA permite verificar el diseño y detectar errores a bajo costo comparado con los sistemas de diseño asistido por computadora (del inglés CAD), para el diseño de circuitos [Cong, 2011].

Por lo anterior, en este trabajo se presenta la metodología de diseño de un sistema para la corrección del desequilibrio I/Q basado en series de Volterra, y su implementación en FPGA. Este trabajo está organizado de la siguiente manera: en la sección 2 se explica la metodología y los fundamentos básicos del desequilibrio I/Q. La sección 3 describen los resultados de la implementación en FPGA. En la sección 4 se realiza la discusión de los resultados obtenidos. Finalmente, en la sección 5 se desglosan las conclusiones.

2. Métodos

Un sistema de comunicaciones moderno [Sen, 2012] se utiliza el dominio digital y análogo para el procesamiento de la señal. Este sistema se conoce como

transceptor y está conformado por un transmisor, un canal de transmisión y un receptor, el diagrama de bloques del sistema se muestra en la figura 1.

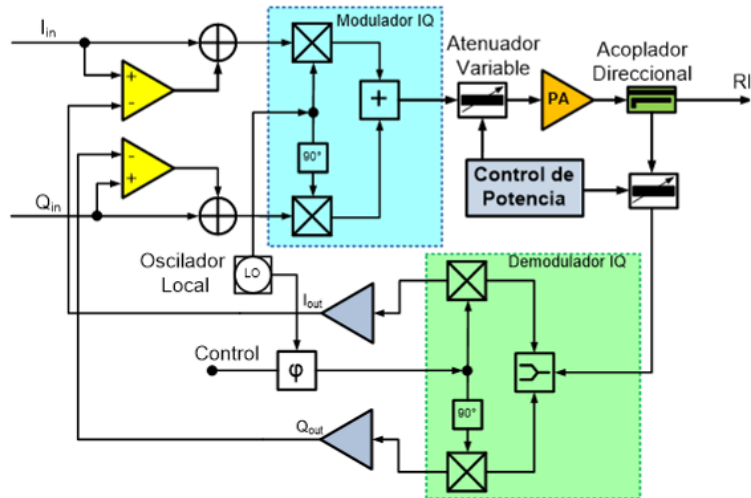


Figura 1 Diagrama de bloques de un sistema de telecomunicaciones.

Los circuitos modernos de modulación comprenden modulaciones complejas basadas en componentes de variable compleja I/Q como 256-QAM para codificar información [Klymyshyn, 2002]. La modulación QAM requiere cambiar la fase y amplitud de una señal senoidal. Se utilizan dos señales portadoras desfasadas 90 grados entre sí. Estas señales representan la componente real (I) e imaginaria (Q) de la señal, las cuales se representan mediante las ecuaciones 1 y 2, en donde A representa la amplitud y ω la fase.

$$I(t) = A \cos(\omega t) \quad (1)$$

$$Q(t) = A \sin(\omega t) \quad (2)$$

La metodología utilizada para la generación de una señal M-QAM rectangular, donde M representa 8, 16, 64 ó 256-QAM, por esta razón se utilizará la configuración 16-QAM para ilustrar su generación. En la ecuación 3 se define el tamaño del símbolo necesario para acceder a los M elementos de la constelación.

$$k = \log_2(M) \quad (3)$$

Para el caso de 16-QAM son necesarios 4 bits, la representación I-Q de la constelación se muestra en la figura 2.

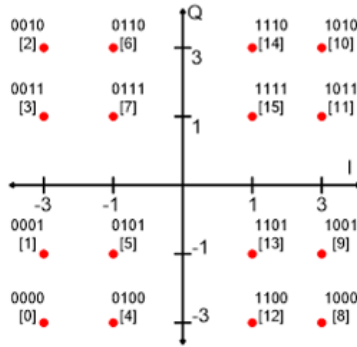


Figura 2 Representación de una constelación 16-QAM.

Donde los puntos rojos representan la señal QAM, en cualquier constelación M-QAM es necesario asignar de manera no consecutiva los símbolos de cada punto de la constelación para restringir las decisiones erróneas de símbolo que provocaría un error de bit menos significativo (del inglés LSB). Por esto, se convierten los símbolos de entrada en símbolos codificados mediante código Gray usando mapas de Karnaugh, para asignar un símbolo a cada punto de la constelación, como se muestra en la tabla 1.

Tabla 1 Mapa de Karnaugh para 16-QAM.

AB\CD	-3	-1	+1	+3
-3	0000 [0]	0001 [1]	0011 [3]	0010 [2]
-1	0100 [4]	0101 [5]	0111 [7]	0110 [6]
+1	1100 [12]	1101 [13]	1111 [15]	1110 [14]
+3	1000 [8]	1001 [9]	1011 [11]	1010 [10]

El desequilibrio I/Q, se produce por una diferencia de ganancias en amplitud o fase entre las señales I y Q del esquema de modulación. Es decir, cualquier desequilibrio en fase será reflejado en la amplitud de la constelación y afectará su cuadratura, un ejemplo de este comportamiento se muestra en la figura 3.

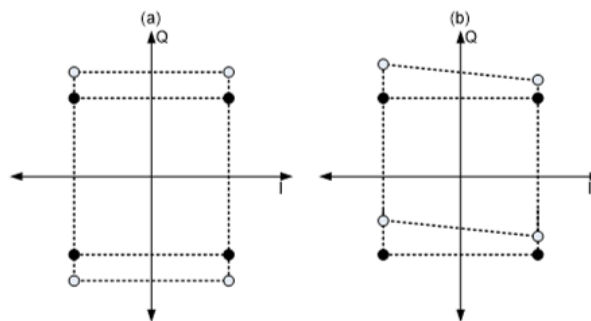


Figura 3 Desequilibrio I/Q en (a) fase y (b) amplitud.

En una transmisión I/Q, la ganancia y el desequilibrio pueden ser medidas fácilmente con un detector de potencia, el cual se encuentra en cualquier sistema de comunicaciones a la salida del amplificador de potencia (PA) [Niubó, 2015], véase la figura 1. Tomando $f(\theta, \phi)$ como la señal a la entrada al amplificador, tenemos ecuaciones 4.

$$\begin{aligned} f(\theta, \phi) &= I_t \cos(\omega t) - Q_t \sin(\omega t + \phi) \\ I_t &= \cos(\omega_b t), \quad Q_t = \sin(\omega_b t + \phi) \end{aligned} \quad (4)$$

Donde ϕ representa el desequilibrio en fase, ω_b la frecuencia en rad/s de la señal moduladora, ω frecuencia de la señal portadora, θ desequilibrio entre las señales en banda base. Se obtiene ecuación 5.

$$\begin{aligned} f(\theta, \phi) &= \cos(\omega_b t) \cos(\omega t) - \sin(\omega_b t + \phi) \sin(\omega t + \phi) \\ &= \cos(\omega t) [\cos(\omega_b t) - \sin(\omega_b t + \phi) \sin(\phi)] - \sin(\omega t) [\sin(\omega_b t + \phi)] \cos(\phi) \end{aligned} \quad (5)$$

En la ecuación 5 se observa que la envolvente contiene la información del desequilibrio I/Q y la transformación de la fase a amplitud se tiene inherentemente en la arquitectura I/Q. Se debe encontrar θ donde la sensibilidad de salida es maximizada para cualquier cambio en ϕ . Considerando la ganancia G_{PA} del PA la función del $E(\theta, \phi)$ se reescribe como se muestra en la ecuación 6.

$$\begin{aligned} E(\theta, \phi) &= envelope[G_{PA} * f(\theta, \phi)] \\ &= G_{PA} \sqrt{[\cos(\omega_b t)]^2 - 2 \sin(\omega_b t + \phi) \sin(\phi) \cos(\omega_b t) + [\sin(\omega_b t + \phi)]^2} \end{aligned} \quad (6)$$

Al diferenciar $E^2(\theta, \phi)$ con respecto a ϕ se obtiene la sensibilidad $S(\theta, \phi)$, ecuación 7.

$$S(\theta, \phi) = -2G_{PA}^2 * \sin(\omega_b t + \phi) \cos(\omega_b t) \cos(\phi) \quad (7)$$

En la ecuación 7 se demuestra que la sensibilidad de potencia alcanza su máximo cuando $I_t = Q_t$ o $Q_t = -Q_t$. En este punto la sensibilidad es máxima para variaciones de la potencia de salida y cualquier cambio en ϕ . La amplitud para ϕ positivas se tiene cuando $I_t = \cos(\omega_b t) = -Q_t$.

$$Tx_{out(\phi)} = G_{PA} * \cos(\omega_b t) \left[2 \sin \left(\omega_b t + 45^\circ \frac{\phi}{2} \right) \sin \left(45^\circ + \frac{\phi}{2} \right) \right] \quad (8)$$

Para obtener los picos de la señal se calcula el módulo como se muestra en la ecuación 9.

$$\begin{aligned} |Tx_{out(\emptyset)}| &= 2G_{PA} * \sin\left(45^\circ + \frac{\emptyset}{2}\right) \\ |Tx_{out(\emptyset)}| &= M * \sin\left(45^\circ + \frac{\emptyset}{2}\right) \end{aligned} \quad (9)$$

Para corregir el desequilibrio I/Q en ganancia, se tiene:

$$I_t = A \cos(\omega_b t) \text{ y } Q_t = 0 \text{ con } P = P_I \quad I_t = 0 \text{ y } Q_t = A \cos(\omega_b t) \text{ con } P = P_Q$$

Donde P_I y P_Q representan la potencia de salida del detector cuando solo está activo I, Q de manera respectiva. La ganancia de desacople puede calcularse mediante ecuación 10.

$$G_{desacople} = \sqrt{\frac{P_I}{P_Q}} \text{ y } K = \sqrt{\frac{2P_I}{G_{PA}^2}} \quad (10)$$

Para corregir el desequilibrio en fase se tiene que al aplicar $I_t = A \cos(\omega_b t)$ y $Q_t = -G_{desacople} * A \cos(\omega_b t)$. se garantiza que las señales serán sumadas con la misma amplitud y la salida se escribe en ecuación 11.

$$|Tx_{out(\emptyset)}| = M * \sin\left(45^\circ + \frac{\emptyset}{2}\right) \quad (11)$$

Conociendo M es fácil encontrar \emptyset .

$$\emptyset = \arcsin\left(\frac{|Tx_{out}|}{|M|}\right) - 45^\circ \quad (12)$$

La ecuación 12 demuestra que se puede calcular el corrimiento fase conociendo el factor M y del módulo de la señal a transmitir.

Sin embargo, es posible modelar el desequilibrio I/Q utilizando series de Volterra, las cuales se utilizan para describir la relación entre la entrada y salida de un sistema no lineal [Nuñez, 2013]. Siendo una alternativa ideal para modelar el comportamiento del desequilibrio I/Q. El modelo del desequilibrio I/Q en banda base complejo se muestra en la figura 4.

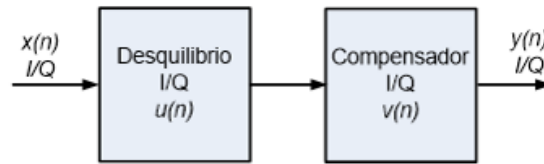


Figura 4 Modelo desequilibrio I/Q en banda base.

Donde $x(n)$ es la señal transmitida, $y(n)$ es la señal distorsionada y, $u(n)$ describe el comportamiento del desequilibrio I/Q y $v(n)$ son los coeficientes del sistema de compensación, descrito en la ecuación 13.

$$x(n) = \sum_{k=0}^{L-1} u(k)s(n-k) + \sum_{k=0}^{L-1} v(k)s^*(n-k) \quad (13)$$

Al rescribir la ecuación 1 en su forma matricial se obtiene la ecuación 14.

$$y_{IQ} = S_{IQ}C_{IQ} \quad (14)$$

Donde C_{IQ} es un vector columna que contiene los coeficientes del modelo, y_{IQ} es el vector de salida del sistema, debido a que se desea que la salida sea lineal respecto a la entrada se dice que $y_{IQ} = x(n)$. S_{IQ} es una matriz que describe la señal con desequilibrio I/Q descrita en las ecuaciones 15 y 16.

$$S_{IQ} = [S_{IQ}(n), S_{IQ}(n-1), \dots, S_{IQ}(n-L)]^T \quad (15)$$

$$S_{IQ}(i) = [s(i), s(i+1), \dots, s(N), s^*(i), s^*(i+1), \dots, s^*(N)]^T \quad (16)$$

En la ecuación 4 N representa el número de muestras de $s(n)$, L es la longitud del vector de coeficientes C_{IQ} . Para resolver el sistema se utiliza el método de mínimos cuadrados. La ecuación 17 muestra el cálculo de los coeficientes.

$$C_{IQ} = (S_{IQ}^H S_{IQ})^{-1} * S_{IQ}^H * y_{IQ} \quad (17)$$

3. Resultados

La implementación se realizó utilizando el entorno de Matlab/Simulink con DSP Builder. Para demostrar el funcionamiento del sistema se utilizará una señal modulada en 256-QAM cuya generación se explica detalladamente en [Juárez, 2017]. En la figura 5 se observa la constelación.

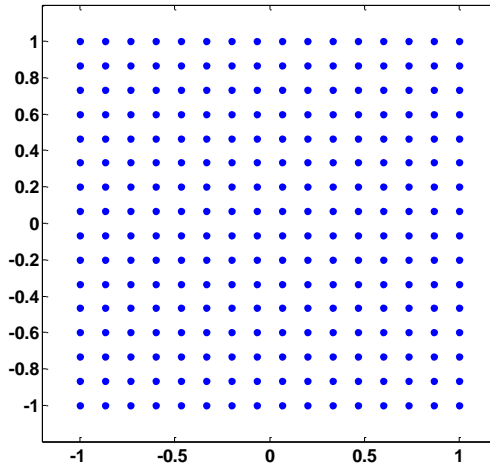


Figura 5 Constelación I/Q 256-QAM.

Se tomarán en cuenta 3 casos de desequilibrio I/Q, descritos en la figura 6:

- Desequilibrio IQ en amplitud -3dB.
- Desequilibrio IQ en fase de 60 grados.
- Desequilibrio IQ en amplitud +1dB.

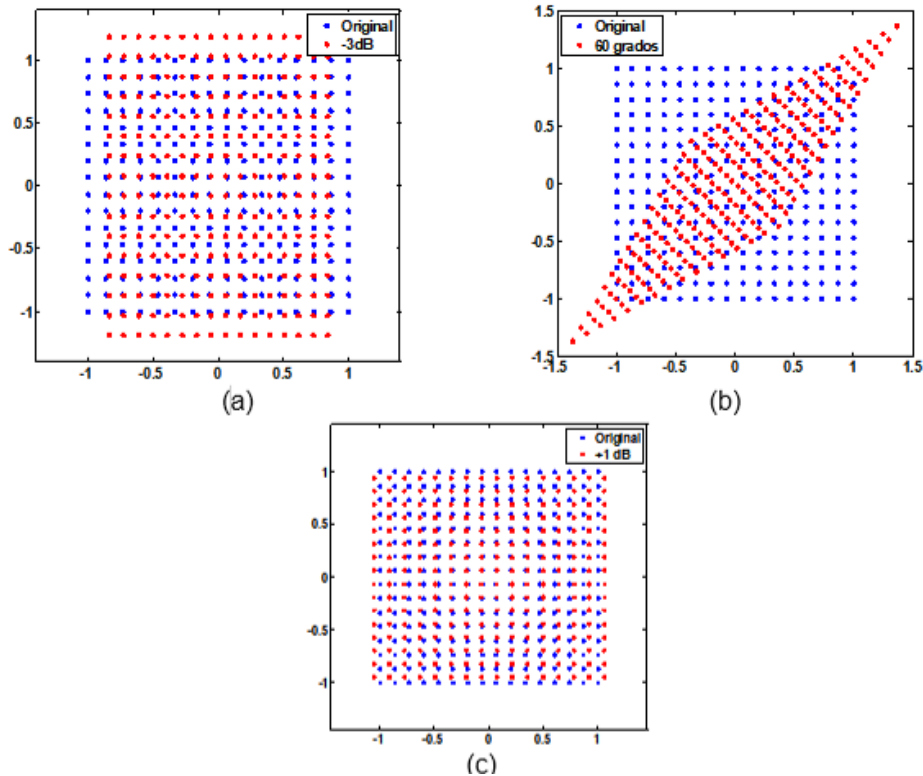


Figura 6 Casos del desequilibrio I/Q.

Como se puede notar en la figura 6 se observa una comparación del desequilibrio I/Q presente en la constelación 256-QAM contra la constelación 256-QAM ideal, en el caso (a) se muestra la compresión de -3dB, en (b) se aprecia un desfase de 60 grados y finalmente en (c) se observa el caso de saturación +1dB.

Definidos los casos del desequilibrio, se realizó la implementación del modelo para obtener la salida del compensador descrito en la ecuación 14 utilizando DSP Builder, como se muestra en la figura 7.

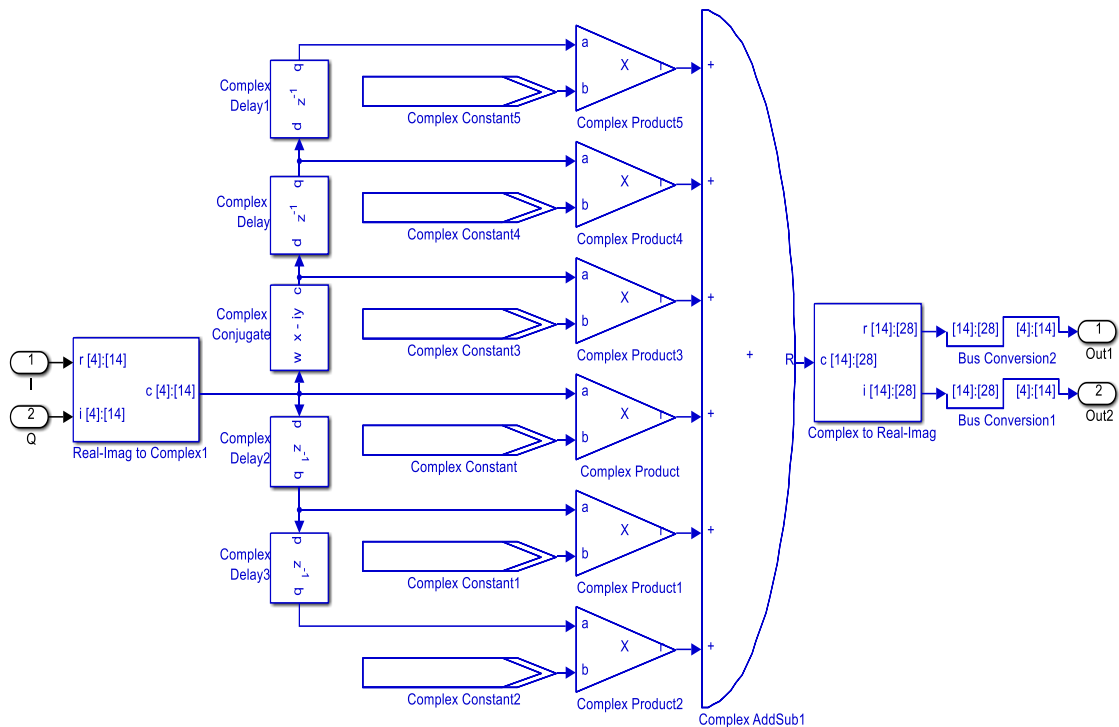


Figura 7 Módulo de corrección del desequilibrio I/Q en DSP Builder.

Una vez realizada la etapa de corrección se integra el proyecto completo en DSP Builder, donde se programan cada uno de los casos y se introduce la etapa de corrección del desequilibrio I/Q como se muestra en la figura 8.

En la entrada del modulador se introduce una señal interna de escalera la cual pasa por todos los valores desde 0 hasta 255, idealmente se debe utilizar la señal que se desea modular, en este caso se optó por la señal escalera para mostrar el correcto funcionamiento del sistema. Esta implementación trabaja en una frecuencia de reloj de 50 MHz, la tarjeta de adquisición de datos Terasic Modelo

MNL-01016-1.0 la cual cuenta con 2 convertidores A/D y 2 D/A con una resolución de 14 bits. En la figura 9 se observa el banco de pruebas y el sistema funcionando.

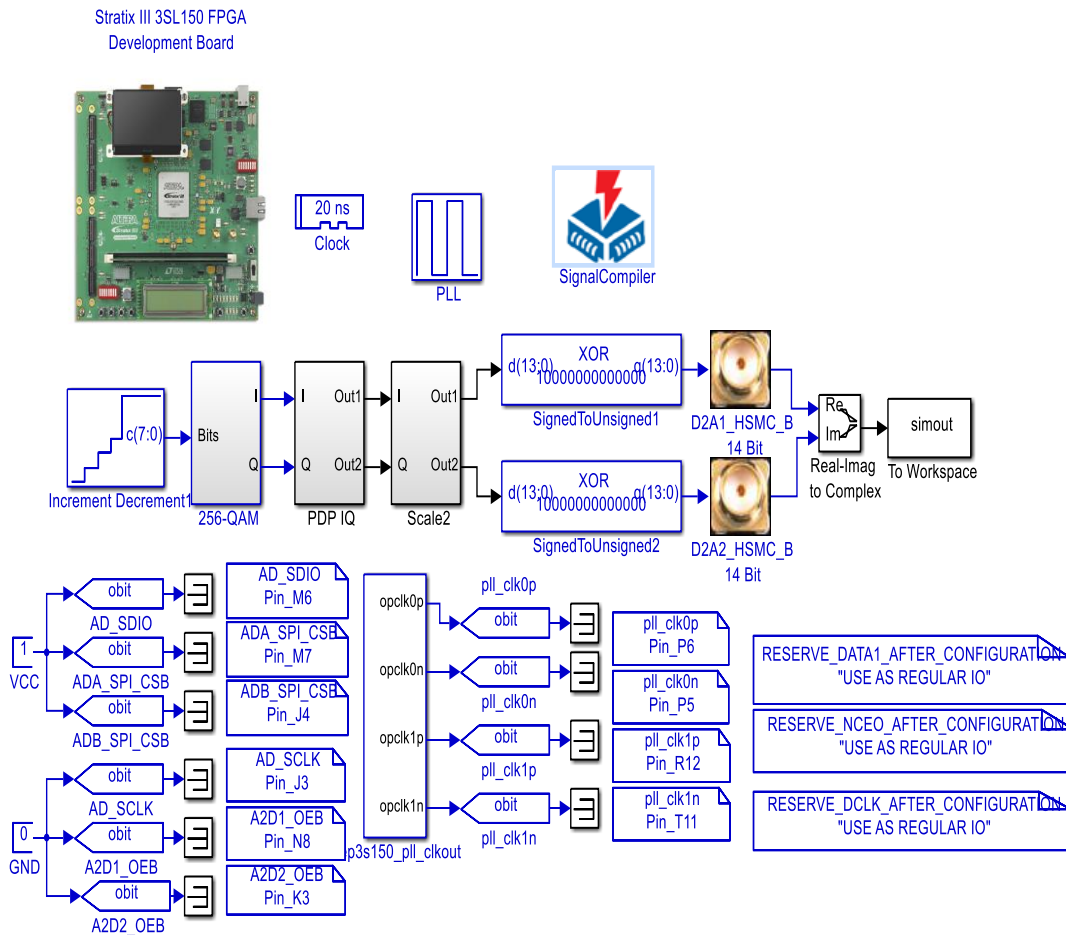


Figura 8 Implementación en DSP Builder.



Figura 9 Implementación del Sistema en FPGA.

4. Discusión

El procedimiento para la corrección del desequilibrio I/Q en amplitud y fase presente en una señal 256-QAM se realizó siguiendo la metodología propuesta en la sección 2. Para calcular los coeficientes del modelo no-lineal basado en series de Volterra se utilizó una longitud de entrenamiento de $L=2$, el cual genera 6 coeficientes. En la figura 10a se observa la corrección del desequilibrio I/Q en -3 dB de amplitud, el cual es corregido satisfactoriamente como se muestra en la figura 9b. Asimismo, la figura 11a posee desequilibrio I/Q en fase de 60 grados, mismo que es corregido mediante el sistema como se presenta en la figura 11b. Finalmente, en la figura 12a se aprecia desequilibrio I/Q en amplitud de +1dB, mientras que en la figura 12b se obtiene el resultado sin desequilibrio I/Q.

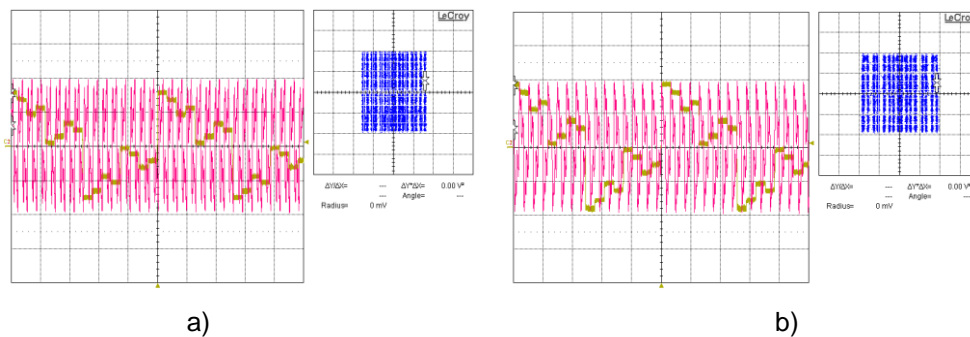
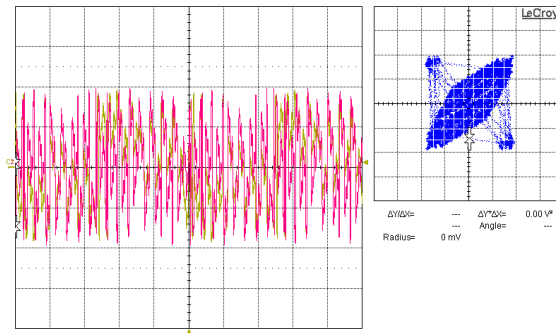


Figura 10 Señales I (rosa) y Q (amarillo) respecto al tiempo y Constelación (I vs Q) vistas en osciloscopio.

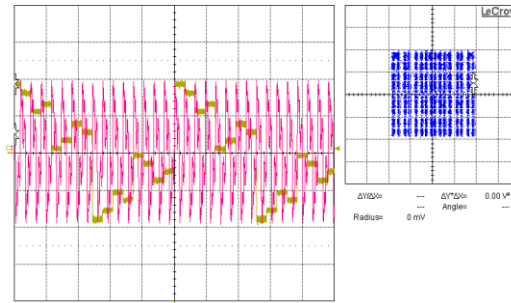
Los recursos utilizados por esta implementación en la tarjeta de desarrollo Stratix III de Altera se muestran en la tabla 1.

Como se observa en la Tabla anterior la cantidad de recursos utilizados están en rangos bajos de 14%, lo cual se traduce en un bajo costo de implementación y una alta flexibilidad, en una posible fabricación en circuito integrado.

El sistema desarrollado demuestra la capacidad de corrección del algoritmo tanto en amplitud como en fase. Cabe señalar que en el futuro se evaluara una verificación del error vector magnitud (del inglés EVM) y una verificación de la tasa de error de bit (del inglés BER). No obstante, es posible corroborar los resultados obtenidos debido a que la forma del trazo en modo X-Y recupera su cuadratura original.

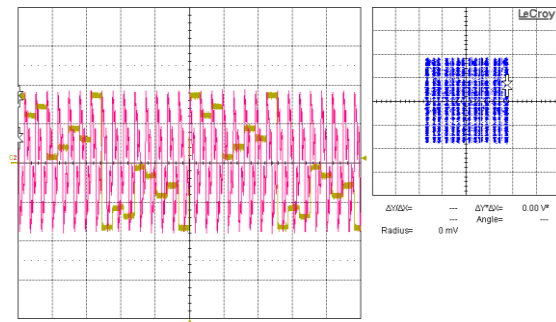


(a)

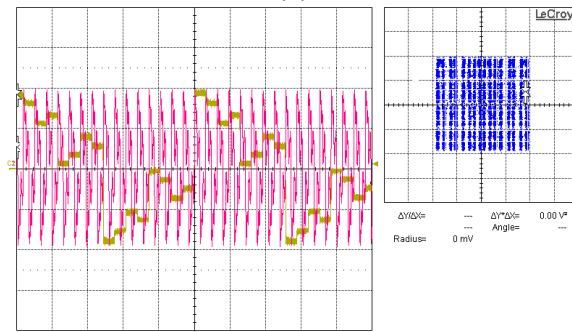


(b)

Figura 11 Señales I (rosa) y Q (amarillo) respecto al tiempo y Constelación (I vs Q) corregida 60° . Señal I/Q y Constelación, vistas en osciloscopio.



(a)



(b)

Figura 12 Señales I(rosa) y Q(amarillo) respecto al tiempo y Constelación (I vs Q) Corrección de 1dB, vistas en osciloscopio.

Tabla 1 Resumen de recursos utilizados del FPGA.

Bloques Lógicos	Utilizados	Disponibles	Total
Funciones Combinacionales	14,351	119,088	13%
ALUTs de Memoria	161	56,800	<1%
Numero de registros lógicos	15,679	113,600	14%
Numero de pines	51	744	7%
Bits de memoria	0	5,630,976	0%
Bloques DSP de 18 bits	38	384	10%
PLLs	1	8	13%
DLLs	0	4	0%

5. Conclusiones

Este artículo presenta la metodología completa de diseño de un sistema para corrección del desequilibrio I/Q digital implementado en una tarjeta DSP-FPGA. El sistema fue implementado en la tarjeta FPGA Stratix III de Altera y se puede considerar como alternativa de bajo costo para la corrección del desequilibrio I/Q y la industria de fabricación de demoduladores. Se desarrolla una metodología de diseño integral capaz de funcionar con cualquier tipo de señal I/Q, además, es capaz de corregir el desequilibrio tanto en amplitud como en fase, dotándolo con una alta flexibilidad de implementación, lo cual provee al sistema una ventaja en cuanto a expansiones futuras, o integración con otros sistemas se refiere. Los resultados de la implementación demuestran su bajo costo y efectividad ya que se requiere una cantidad mínima de recursos lógicos para su puesta en marcha, obteniendo resultados precisos.

6. Bibliografía y Referencias

- [1] Anttila, L., Valkama, M. & Renfors, M., Circularity-Based I/Q Imbalance Compensation in Wideband Direct-Conversion Receivers, *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2099–2113, 2008.
- [2] Cavers, J. K., New methods for adaptation of quadrature modulators and demodulators in amplifier linearization circuits, *IEEE Trans. Veh. Technol.*, Vol. 46, num. 3, pp. 707–716, 1997.
- [3] Anttila, L., Valkama, M. & Renfors, M., Blind Compensation of Frequency-Selective I/Q Imbalances in Quadrature Radio Receivers: Circularity -Based

- Approach, IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07, vol. 3, pp. 245–248, 2007.
- [4] Asami, K., An algorithm to evaluate wide-band quadrature mixers”, 2007 IEEE International Test Conference, pp. 1–7, 2007.
- [5] Cong, J., Liu, B., et al, High-Level Synthesis for FPGAs: From Prototyping to Deployment, IEEE Trans. Computer-aided Design of Integrated Circuits and Systems., vol. 30, no. 4, pp. 473-791, 2011.
- [6] Dick, C., Harris, F., & Rice, M., FPGA Implementation of Carrier Synchronization for QAM Receivers, Journal of VLSI Signal Processing 36, pp. 57-71, 2004.
- [7] Erdogan, E. S., & Ozev, S., Single-Measurement Diagnostic Test Method for Parametric Faults of I/Q Modulating RF Transceivers, 26th IEEE VLSI Test Symposium (vts 2008), pp. 209–214, 2008.
- [8] Juarez-Cazares, S. A., Nuñez-Perez, J.C., et al, Sistema de Modulación Digital Compleja 256-QAM Basado en FPGA, Revista Aristas: Investigación Básica y Aplicada., Vol. 6, Núm. 11, pp. 99-105, 2017.
- [9] Klymyshyn, D. M., FPGA implementation of multiplierless M-QAM modulator, in Electronics Letters, vol. 38, no. 10, pp. 461-462, 2002.
- [10] Lee, C. P., Behzad, A., et al, A Highly Linear Direct-Conversion Transmit Mixer Transconductance Stage with Local Oscillation Feedthrough and I/Q Imbalance Cancellation Scheme, IEEE International Solid State Circuits Conference - Digest of Technical Papers, pp. 1450–1459, 2006.
- [11] Mattera, D., Paura, L., & Sterle, F., MMSE WL Equalizer in Presence of Receiver IQ Imbalance, IEEE Trans. Signal Process., vol. 56, no. 4, pp. 1735–1740, 2008.
- [12] Sen, S., Devarakond, S. K., & Chatterjee, A., Phase Distortion to Amplitude Conversion-Based Low-Cost Measurement of AM-AM and AM-PM Effects in RF Power Amplifiers, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 20, no. 9, pp. 1602-1614, 2012.
- [13] Nash, E., Correcting imperfections in IQ modulators to improve RF signal fidelity, 2009.

- [14] Niubó-Aleman, T., Nuñez-Perez, J.C, et al, Diseño e implementación en un FPGA de un detector de fase para corregir el desequilibrio en señales I/Q, *ELECTRO*, Vol. 37, pp. 98-103, 2015.
- [15] Nuñez-Perez, J.C., Cardenas-Valdez, J.R., et al, Flexible test bed for the behavioural modelling of power amplifiers, *COMPEL*, vol. 33, no. 12, pp. 355-375, 2013.
- [16] Snelgrove, W. M., A novel adaptive mismatch cancellation system for quadrature IF radio receivers, *IEEE Trans. Circuits Syst. II Analog Digit. Signal Process.*, vol. 46, no. 6, pp. 789–801, 1999.