

# Control de un prototipo Mecatrónico usando Redes Neuronales

Yolanda Pérez Pimentel<sup>1</sup>, Ismael Osuna Galán<sup>1</sup>, Rodolfo E. Ibarra Orozco<sup>1</sup>  
Juan Villegas Cortez<sup>2</sup>

<sup>1</sup>Universidad Politécnica de Chiapas, Eduardo J. Selvas S/N, Tuxtla Gutiérrez, Chiapas, (961)6128404,  
[ypimentel@upchiapas.edu.mx](mailto:ypimentel@upchiapas.edu.mx), [ribarra@upchiapas.edu.mx](mailto:ribarra@upchiapas.edu.mx), [iosuna@upchiapas.edu.mx](mailto:iosuna@upchiapas.edu.mx)

<sup>2</sup>Universidad Autónoma Metropolitana, Azcapotzalco. Departamento de Electrónica. San Pablo Xalpa No. 180,  
Col. Reynosa Tamaulipas, CP 2200, México D.F., [juanvc@azc.uam.mx](mailto:juanvc@azc.uam.mx)

## Resumen

*El reconocimiento de voz presenta algunos desafíos aún no resueltos en su implementación en controladores en tiempo real con sistemas reconfigurables de entradas y salidas (Reconfigurable Inputs/Outputs o RIO por sus siglas en inglés). Recientemente los métodos de reconocimiento de voz han logrado grandes avances resolviendo problemas usando técnicas cada vez más complejas. En el aspecto teórico, los diferentes criterios de aprendizaje discriminativo usando redes neuronales han sido introducidos para tener una visión que unifique los criterios en el reconocimiento de patrones asociados a la voz. En el campo práctico se han popularizado nuevas y cada vez más poderosas computadoras al alcance de la población en general, con lo que se puede realizar aplicaciones efectivas aplicadas al control de dispositivos. En este artículo se muestra un proyecto que consiste en la aplicación del reconocimiento de voz en el control de sistemas mecatrónicos, en éste caso una silla de ruedas, a la que se le proveerá de una interfaz electrónica y un controlador de automatización programable que permitan su desplazamiento utilizando comandos de voz básicos.*

**Palabras clave:** Análisis de Voz, Redes Neuronales, LabVIEW.

## 1. Introducción

Un sistema de tiempo real debe procesar la información y producir una respuesta en un plazo determinado, es decir, el funcionamiento de un sistema no sólo depende del resultado que devuelve la computadora, también depende del tiempo en que se produce ese resultado. Por otra parte, los Sistemas de Reconocimiento de Voz (SRV) se han vuelto muy populares en los sistemas operativos de computadoras personales, smartphones y de dispositivos de entretenimiento por mencionar algunos ejemplos. Una de las técnicas para el análisis y reconocimiento

de comandos de voz es el Análisis Multirresolución de Wavelets (MRA) [3]. En años recientes se han incorporado técnicas cada vez más complejas como lógica difusa, redes neuronales (RN) y Modelos Ocultos de Markov (HMM). La metodología básica para aplicar éstas técnicas consiste en extraer sus características principales, usando para ello métodos multirresolución para representar una señal dinámica en el tiempo en términos de sus componentes principales con lo que se tiene un análisis en tiempo real, por último, mediante una red neuronal se realiza la clasificación.

La implementación de un sistema de reconocimiento de voz usando MRA y redes neuronales en un FPGA es lo mostrado en el presente artículo. En la sección 2 se presenta el planteamiento del desarrollo de nuestra solución, en cinco diferentes etapas de la implementación con detalle, para en la sección 3 mostrar los resultados obtenidos, en la sección 4 se brinda una discusión de éstos y, finalmente la sección 5 muestra las conclusiones.

## 2. Metodología

La propuesta de este trabajo consta de tres partes fundamentales:

- Reconocimiento de voz: Consistente en analizar una señal de voz, primero extrayendo características de ella y después aprenderla y reconocerla mediante una Red Neuronal con el algoritmo de aprendizaje Backpropagation. Se utiliza para la implementación la plataforma LabVIEW debido a la posibilidad que ofrece de ser embebido dentro de un PLC o FPGA de la empresa National Instruments o de terceros, en este caso una CompactRIO.
- Diseño del prototipo: Una silla de ruedas estándar modificada con motores eléctricos y operada con una tarjeta CompactRIO NI 9074 en la cual se descargará el software.
- Integración Mecatrónica: La parte final del proyecto consiste en el diseño y construcción del prototipo. Una silla de ruedas con motores acoplados,

baterías y tarjeta CompactRIO con el programa embebido para realizar el control mediante el reconocimiento de Voz.

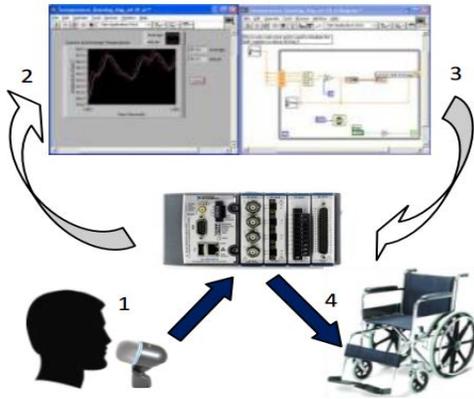


Fig. 1. Esquema General del proyecto.

La figura 1 ilustra el esquema de manera general en qué consiste el proyecto, en cuatro etapas, tal como se describe a continuación: (1) La voz es adquirida por medio de un micrófono a través de una entrada analógica del módulo NI 9234. (2) Esa señal adquirida es procesada por medio del programa embebido desarrollado en LabVIEW. (3) Después del procesamiento, se envía una señal de activación mediante un módulo NI 9481 de salidas a relés. (4) Esta activación es la que controla los dos motores que se han incorporado a la silla de ruedas, y hacen que avance, se detenga, o gire hacia la izquierda o la derecha dependiendo del resultado del procesamiento.

### 3. Reconocimiento de Voz

En la primera parte de la propuesta se diseñó un Programa de Reconocimiento de Voz mediante análisis multirresolución y Redes Neuronales. La metodología en dos etapas, aprendizaje y prueba, se muestra en la Figura 2.

Metodología	
Aprendizaje	Prueba
Adquisición de Datos	Adquisición de Datos
Limpieza	Limpieza
Extracción de Características	Extracción de Características
Entrenamiento de la Red Neuronal	Clasificación

Fig. 2. Esquema de entrenamiento de la RN.

El primer paso es la adquisición de datos. Para formar una base de datos que será usada tanto en la parte del entrenamiento como en las pruebas.

Las muestras de audio se adquirieron mediante un micrófono incorporado en una diadema. Se tomaron 4000 muestras a una frecuencia de 10 KHz.

La base de datos está conformada por 25 archivos de las palabras “Avanza”, “Para”, “Izquierda” y “Derecha” en idioma español, de un sólo hablante. De esos 25 archivos, 10 se utilizaron en la etapa de entrenamiento y 15 se reservaron para la etapa de prueba.

#### 3.1 Pretratamiento de la señal.

Después de la adquisición de datos se observa que cada muestra contiene datos que resultan inútiles, por lo cual el siguiente paso es la eliminación de tales segmentos, lo cual se lo logra tomando como referencia los niveles inicial y final de la muestra para obtener un promedio:

$$v = \frac{1}{\# \text{datos}} \sum_{i=1}^{\# \text{datos}+1} |x|$$

donde:

v es el valor que se desea eliminar

#datos es el número de muestras que se están evaluando.

$$\text{umbral} = \frac{\text{ValI} + \text{ValF}}{2} + n$$

ValI es el Valor de los primeros datos

ValF es el Valor de los últimos datos

n = 2000 datos inútiles.

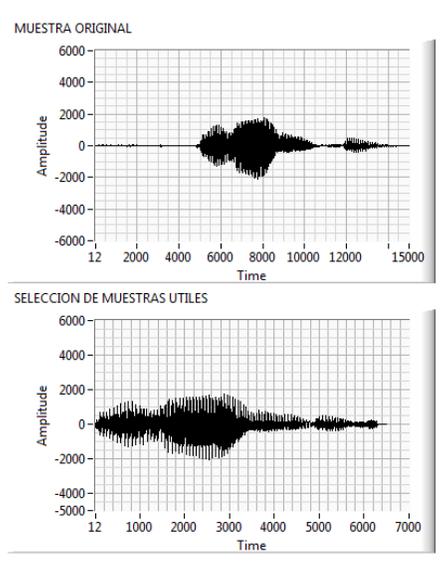


Fig. 3. Datos obtenidos.

En la Figura 3 se aprecia una muestra completa y una muestra en la que se ha realizado la selección de datos, ambas correspondientes a la palabra “Avanza”.

Se busca realizar el reconocimiento de palabras sin segmentar las muestras. Dado que se trata de un sistema que requiere identificar pocas palabras, se da inicio utilizando toda la extensión de la palabra, que para la palabra “Avanza” es de 6500 muestras, mientras que Para tiene únicamente 4500 muestras. Sin embargo, se ha considerado que las muestras deben ser uniformes, y en escala de  $2n$ , por lo que se ha tomado el menor valor que cubra ambas palabras,  $212 = 4096$ . De ésta forma, las muestras quedan uniformes en tamaño.

### 3.2 Extracción de características

Entre las técnicas usadas más para extraer características están la Transformada Rápida de Fourier y la Transformada con Wavelets. En este trabajo se explora el uso de la Transformada con Wavelets debido a su efectividad para realizar el análisis de la señal tanto en el dominio de la frecuencia como del tiempo [7]. La implementación de la Transformada Wavelet Discreta (TWD) se realiza mediante la técnica de Análisis Multiresolución (MRA) ilustrada en la Figura 5, con seis niveles de descomposición

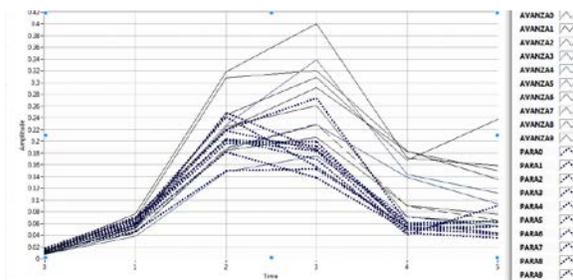


Fig. 4. Niveles de descomposición usando MRA.

Como se puede observar los niveles 4, 5 y 6 de energía, proveen una mayor información para la clasificación. Para corroborar lo anterior, se procede a graficar dichos niveles por parejas. La Figura 7 muestra la gráfica realizada para los niveles 4 y 5 de las palabras “Avanza” y “Para” obtenidos por MRA.

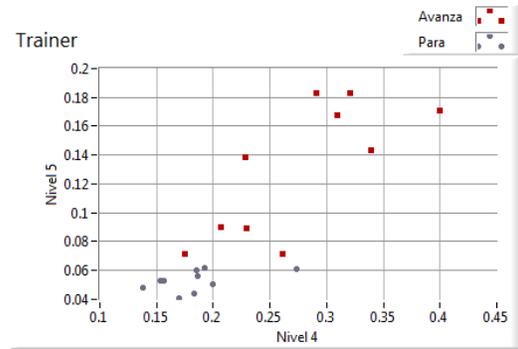


Fig. 5. Energía correspondientes a los niveles 4 y 5.

### 3.3 Red Neuronal

Se procede al entrenamiento de la Red Neuronal utilizando un vector de entradas con tres características. Derivado del entrenamiento del Perceptrón, se obtiene un vector de pesos ajustados que se utilizan para la etapa de prueba.

Se hizo un Perceptrón Multicapa que permita realizar una mejor clasificación de clases a cuatro por el momento, que serían: “Avanza”, “Para”, “Derecha” e “Izquierda” [8].

Teniendo ya como meta la implementación de un Perceptrón Multicapa que clasifique las cuatro palabras propuestas, lo primero que se hace es graficar las instancias para ver su disposición espacial.

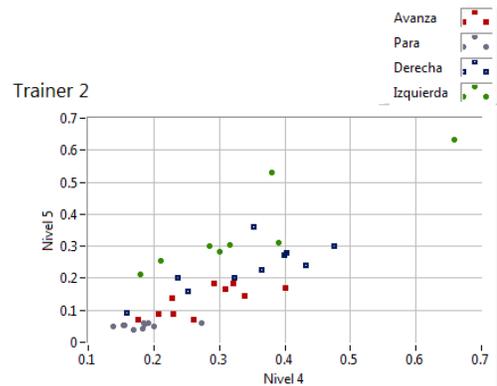


Fig. 6. Energía correspondientes a los niveles 4 y 5.

La figura 7 muestra 10 instancias de las palabras “Avanza”, “Para”, “Derecha” e “Izquierda”, que serán utilizadas en la parte del entrenamiento de la Red Neuronal con el algoritmo de aprendizaje Backpropagation.

Se probaron alternativamente, Estado Máximo de Energía de cada nivel (EME), Amplitud, RMS,

Estado Mínimo de Energía (EmE), y se procedió a entrenar la Red Neuronal y a probarlos con el vector de pesos resultantes.

El Estado Máximo de Energía fue el que ofreció mejores resultados, mejorando notablemente los resultados de la etapa de prueba.



**Fig. 7. Prototipo mecatrónico.**

Es importante hacer notar que LabVIEW no cuenta con un módulo de Redes Neuronales. Por lo que para tener código embebido en la CompactRIO se realizaron un total de 12 programas. Los más importantes son:

- *SeleccionMuestra.vi*
- *MatrizCaracteristicas.vi*
- *PerceptronTrainer.vi*
- *RNPrincipal.vi*

El resto de ellos, aunque se encuentran anexos al proyecto son de menor jerarquía. A continuación se presentan los detalles operacionales de cada programa:

*SeleccionMuestra.vi*: Permite seleccionar las muestras útiles de cada archivo, guardándolos en un nuevo archivo con extensión \*.tdms.

*MatrizCaracteristicas.vi*: Realiza el MRA, calcula el nivel de energía por nivel de cada muestra, selecciona el Estado Máximo de Energía de cada nivel, y guarda N muestras, que pueden ser de N muestras y N clases, en un solo archivo \*.tdms. Adicionalmente, se guardó un archivo en \*.lvm, el cual puede ser abierto con un block de notas y editado como archivo \*.arff.

*PerceptronTrainer.vi*: Consiste en un Perceptrón simple que se entrega para clasificar dos clases diferentes. El programa se detiene después de piteraciones, las cuales corresponden al número de muestras a clasificar.

*RNPrincipal.vi*: Es el programa general que se ejecutará en la CompactRIO, y dónde se realizará todo el procesamiento en Tiempo Real y el control de la silla de ruedas.

#### 4. Prototipo

Una vez que el programa y la silla fueron probados de manera independiente se procedió a integrar todos los componentes del sistema. Hubo inconvenientes menores al acoplar los motores debido a que la silla es estándar no eléctrica; se hizo un análisis estructural para que este acople se realizara con las modificaciones mínimas. Y por otro lado, la selección del micrófono adecuado para trabajar en ambientes sin control de ruido.

Para reducir costos se decidió usar silla de ruedas estándar no eléctrica. Además se ha considerado que pueda mover al menos 120 kg de peso, en lo cual están incluidos, el mismo peso de la silla, la CompactRIO, motores, baterías y la persona que va a controlar la silla. De acuerdo a un análisis previo, se usarán dos motores de 12VCD que en conjunto proveen una potencia de 80 W, con el sistema de transmisión de movimiento serán capaces de mover el peso calculado a una velocidad razonable. El control es de lazo abierto y es realizada mediante una CompactRIO NI 9074, lo que permite que trabaje de forma autónoma.

El sistema CompactRIO contiene dos módulos:

● NI 9234 es un módulo de cuatro canales para adquisición de señal dinámica para realizar medidas de frecuencia de audio de alta precisión, desde sensores piezoeléctricos electrónicos integrados, proporciona rango dinámico de 102 dB e incorpora acondicionamiento de señales seleccionable por software para micrófonos. Los cuatro canales de entrada digitalizan señales simultáneamente a rangos de hasta 51.2 kHz por canal con filtros integrados anti-aliasing que ajustan automáticamente su rango de muestreo.

● NI 9481 es un módulo de cuatro salidas digitales. Cada canal proporciona acceso a un relé electromecánico para conectar señales hasta 60 VDC o 250 VAC a 2A y ofrece aislamiento entre canales de 250 Vrms y permite conexión directa a dispositivos como lo son motores DC. e utilicen subsecciones, la separación entre el título de la subsección y los párrafos superiores e inferiores será de un renglón. La numeración de las subsecciones se efectuará con números arábigos. El título de cada subsección se escribirá con letras mayúsculas y

minúsculas, de acuerdo al caso usando el tipo new roman normal de 10 puntos.

## 5. Análisis de Resultados

Se logró un sistema básico implementado en LabVIEW y embebido en una CompactRIO para el reconocimiento de comandos de voz usando redes neuronales mediante el análisis multiresolución con Wavelets. A pesar de que no se ha logrado el 100% de certidumbre al momento de clasificación, los resultados son bastante alentadores.

Otra técnica que se desea explorar son los Modelos Ocultos de Markov, [4]. Además, se desea que el sistema funcione con múltiples hablantes y pueda ser implementado para el control de una silla de ruedas mediante el uso de una CompactRIO.



**Fig. 8. Prototipo mecatrónico.**

El sistema presentado es de lazo abierto, con algunas mejoras se asegura el movimiento completo de la silla de ruedas operada por comandos de voz. Para ello se proponen ciertas mejoras: Aumentar el vocabulario que pueda reconocer el sistema añadiendo las palabras “Lento” y “Rápido”, instrumentar sensores de proximidad, encoders y de inclinación, con el fin de que la silla responda no sólo ante los comandos de voz sino a otras variables del entorno. Por ejemplo, en una pendiente ascendente será necesario mantener las condiciones de velocidad y torque para responder a las mismas órdenes “Avanza” y “Para”. De esta forma se proporciona al usuario un mejor control sobre el movimiento de la silla de ruedas.

## 6. Conclusiones

La implementación de una red neuronal para el control de un dispositivo por medio de comandos de voz es viable, a mediano plazo se pretende conseguir un sistema híbrido, que use no solamente señales de voz sino también señales electromiográficas o subvocales,

## 7. Referencias

- [1] B. J.B. “Small Vocabulary Recognition Using Surface Electromyography in an Acoustically Harsh Environment”. NASA/TM-2005-213471. Nov 2005.
- [2] E. López Larraz, O. Martínez Moroz. “Diseño de un sistema de reconocimiento del habla mediante electromiografía”. Actas del XXVII Congreso Anual de la Sociedad Española de Ingeniería Biomédica. Año 2009. 601-604.
- [3] E. Pinto Moreno, M. J. Gómez García. Análisis y Evaluación de los niveles de energía obtenidos mediante WPT para defectología de Ejes. Universidad Carlos III de Madrid. Octubre 2012.
- [4] J. L. Oropeza Rodríguez, S. Suárez Guerra. “Algoritmos y Métodos para el Reconocimiento de Voz en Español mediante Sílabas”. Computación y Sistemas. Volumen 9. Año 2006. Págs. 270-286.
- [5] J. Travis, J. Kring. LabVIEW for Everyone. 3rd. Ed. 2007. Pearson Education, Inc. Crawfordsville, Indiana.
- [6] LabVIEW Real-Time Application Development Course Manual. National Instruments Corporation. México. 2007.
- [7] M. Weeks. Digital Signal Processing Using Matlab y Wavelets. Año 2007. Infinity Science Press. Hingham Massachusetts.
- [8] V. Kecman. Learning and Soft Computing. Support Vector machines, Neural Networks and Fuzzy Logic Models. The MIT Press. Cambridge, London, England. Págs. 5.

## 8. Autores

**M. I. D. S. Yolanda Pérez Pimentel** obtuvo su título de Maestría en Ingeniería y Desarrollo de Puebla por el Colegio de Posgrado en Desarrollo de Software de Puebla. Obtuvo la Certificación como Desarrolladora Asociada en LabVIEW por National Instruments. Actualmente se encuentra realizando el Doctorado en Sistemas Computacionales en la Universidad del Sur en Chiapas. Su línea de investigación es en Redes

Neuronales Aplicadas al Reconocimiento de Voz. Es Profesora-Investigadora en la Universidad Politécnica de Chiapas adscrita al área de Mecatrónica.

**Dr. Ismael Osuna Galán** es Doctor en Matemáticas por la Universidad Autónoma Metropolitana. Labora en la Universidad Politécnica de Chiapas adscrito al área de Mecatrónica. Tiene la Certificación como Desarrollador Asociado en LabVIEW. Su trabajo de investigación está enfocado en el Control Inteligente con Lógica Difusa y Redes Neuronales e instrumentación virtual.

**Dr. Rodolfo Esteban Ibarra Orozco** es doctor en Ciencias Computacionales por el Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Estado de México. Actualmente trabaja como Profesor-Investigador en la Universidad Politécnica de Chiapas adscrito al área de Desarrollo de Software. Sus áreas de investigación son Memorias Asociativas, Redes Neuronales y Sistemas Inteligentes.

**Dr. Juan Villegas Cortez** obtuvo su doctorado en Ciencias de la Computación en el Instituto Politécnico Nacional en 2009. Actualmente labora como Profesor-Investigador en el Departamento de Electrónica de la Universidad Autónoma Metropolitana, Unidad Azcapotzalco, Ciudad de México, México. Su trabajo de investigación se enfoca en Reconocimiento de Patrones, Análisis de Imágenes, Redes Neuronales Artificiales, Memorias Asociativas y Aplicaciones de Reconocimiento de Fuentes.