

# CONTROL DE ILUMINACIÓN CON RECONOCIMIENTO REMOTO DE VOZ

LIGHTING CONTROL WITH REMOTE VOICE RECOGNITION

Recibido: 03 de abril del 2011

Aprobado: 15 de mayo del 2011

DOLLY P. CAICEDO-ZAMORA\*  
JUAN C. MARTÍNEZ-QUINTERO\*\*  
JAIME E. ANDRADE-RAMÍREZ\*\*\*

## Resumen

En el artículo los autores enseñan cómo se desarrolló un sistema capaz de reconocer cinco palabras aisladas en un ambiente de bajo nivel de ruido. El sistema creado permite el control electrónico de iluminación en el que, usando la capacidad de procesamiento de un PC, se implementan algoritmos referentes al tratamiento y reconocimiento de la señal de voz. Para este proceso se usa el algoritmo denominado “transformada rápida de Fourier” y redes neuronales artificiales. La implementación del control electrónico incluye el uso del protocolo *zigbee* para comunicar el dispositivo de adquisición de voz del usuario y la computadora. El resultado es un sistema de reconocimiento de voz para el control de luces portátil, limitado por el alcance de los módulos de transmisión-recepción empleados para este fin. El presente trabajo es desarrollado por el grupo de automatización industrial de la Universidad Cooperativa de Colombia, en el marco del proyecto de investigación “Diseño y construcción de un módulo genérico de control activado por voz para personas con discapacidad”, aprobado en el 2009.

**Palabras clave:** procesamiento de señales de voz, protocolo *zigbee*, reconocimiento de voz, red neuronal artificial (RNA), transformada rápida de Fourier (FFT).

## Abstract

In this paper, the authors show the development of a system able to recognize five separate words in a low noise environment. The created system allows electronic control of lighting, using the computational power of a personal computer to run the algorithms involved in the processing and recognition of the voice signal. For this process, an algorithm called Fast Fourier Transform and artificial neural networks are used. Implementing the electronic control includes using the *zigbee* protocol to communicate between the user's voice acquisition device and the computer. The result is a portable voice recognition system for lighting control, which is only limited by the coverage of the transmission-reception modules used. This work is being developed by the industrial automation research group of the Universidad Cooperativa de Colombia as part of the project “Design and construction of a generic voice activated control module for people with disabilities”, approved in 2009.

**Keywords:** voice signal processing, *zigbee* protocol, voice recognition, Artificial Neural Networks (RNA), Fast Fourier Transform (FFT).

• Cómo citar este artículo: Dolly P. Caicedo-Zamora, Juan C. Martínez-Quintero y Jaime E. Andrade-Ramírez. “Control de iluminación con reconocimiento remoto de voz”. *Revista Ingeniería Solidaria*, vol. 7, núms. 12-13, 2011, pp 35-45.

\* Ingeniera de Sistemas de la Universidad de Cundinamarca. Docente de tiempo completo de la Facultad de Ingeniería de la Universidad Cooperativa de Colombia, sede Bogotá.

Correo electrónico: dolly.caicedo@campusucc.edu.co

\*\* Ingeniero electrónico de la Universidad Manuela Beltrán. Especialista en Seguridad Física e Informática de la Escuela de Comunicaciones Militares. Docente de tiempo completo de la Facultad de Ingeniería de la Universidad Cooperativa de Colombia, sede Bogotá. Correo electrónico: juan.martinez@campusucc.edu.co

\*\*\* Ingeniero electrónico de la Universidad Manuela Beltrán. Especialista en Gerencia Integral de las Comunicaciones de la Escuela de Comunicaciones Militares. Docente de tiempo completo de la Facultad de Ingeniería de la Universidad Cooperativa de Colombia, sede Bogotá. Correo electrónico: jaime.andrade@campusucc.edu.co

## Introducción

Uno de los principales objetivos del reconocimiento del habla es permitir el acceso a las máquinas mediante comandos de voz [1]. Esta idea es la tendencia moderna de los dispositivos electrónicos en las diversas áreas de aplicación. Algunas de estas aplicaciones están orientadas al desarrollo de sistemas que permiten a personas con algún tipo de discapacidad tener una mejor calidad de vida.

En este documento se describe un sistema útil para el control de iluminación en forma remota haciendo uso de tecnologías inalámbricas, tratamiento digital de la señal de voz, redes neuronales artificiales y microcontroladores.

Este desarrollo le permite al usuario portarlo de un lugar a otro al interior de un recinto sin la necesidad de conexiones tangibles, puesto que usa un modulo de comunicación inalámbrica basado en el protocolo *zigbee*. La comunicación inalámbrica brinda la posibilidad de utilizar todos los recursos del PC sin tener que llevarlo consigo. Esta capacidad de procesamiento es aprovechada en la identificación de la señal de habla mediante algoritmos implementados en las herramientas de programación *Labview* y *Matlab*. Los algoritmos desarrollados se encargarán de procesar la señal de voz reci-

bida y extraer algunos de los parámetros más relevantes de esta; después de lo anterior, se identificará mediante el algoritmo de una red neuronal artificial (RNA) si la interpretación que la red genera a partir del patrón de entrada es acorde con una de las cinco palabras aisladas que el sistema puede reconocer. Esta interpretación se convierte en una orden que nuevamente es enviada en forma inalámbrica hacia el sistema de adquisición de usuario, en el cual se emitirán órdenes que actuarán sobre el proceso de control de iluminación. Además de las tareas ya mencionadas, la tarjeta de adquisición de usuario permite la detección, digitalización y transmisión al PC de la señal de voz.

Una de las grandes ventajas del sistema es que permite a personas con algún tipo de discapacidad motora controlar en forma inalámbrica una variable de su entorno y es el comienzo de un sistema que permitiría una mayor interacción del discapacitado con el medio que lo rodea.

## Metodología

Para el desarrollo propuesto, a continuación se describirá cada una de las etapas que intervienen en el proceso (ver figuras 1 y 2).

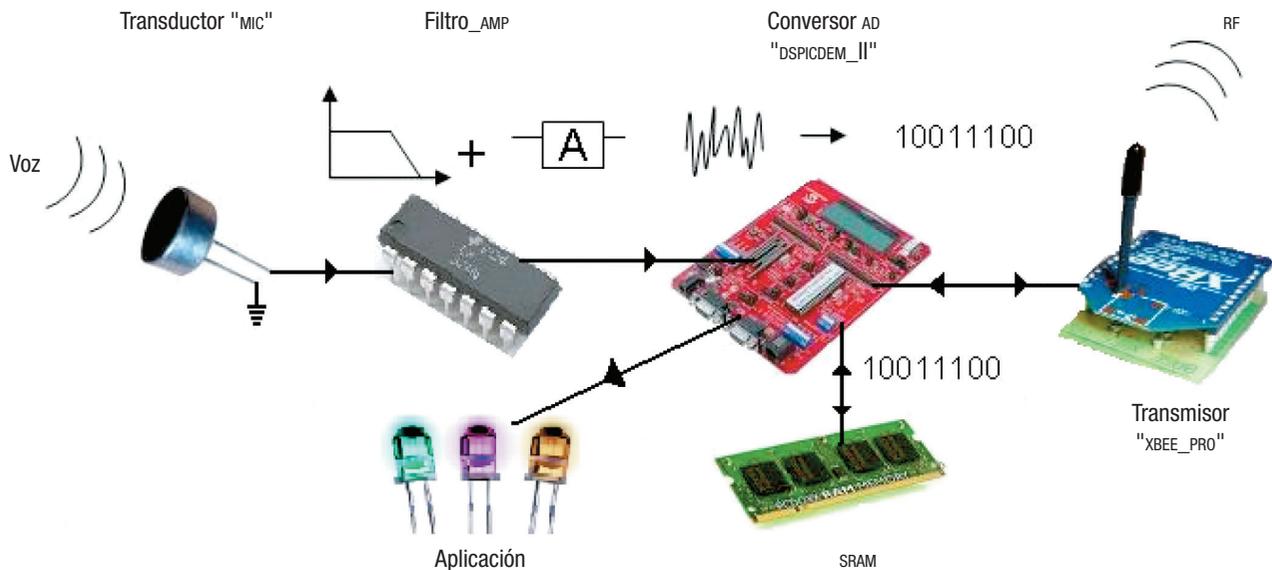


Figura 1. Diagrama general del módulo transmisor (módulo de adquisición de usuario)

Fuente: los autores

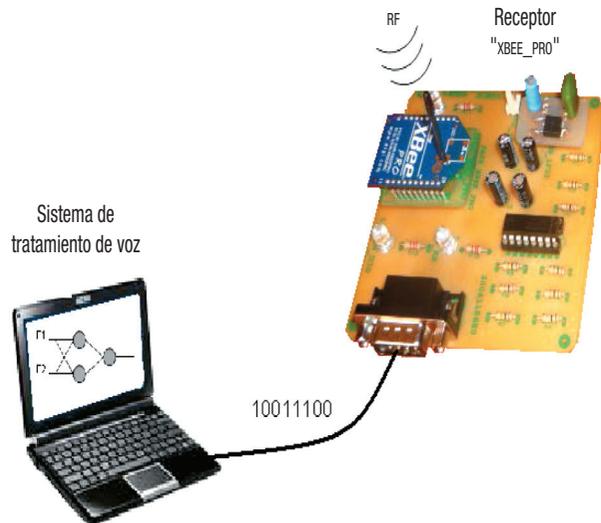


Figura 2. Diagrama general del módulo receptor (sistema de identificación y reconocimiento)

Fuente: los autores

### Captura de la voz en forma analógica

Se configura e implementa una etapa analógica de polarización y pre-amplificación para el transductor de entrada (micrófono electrec), y se establece el filtro adecuado con ganancia DC (corriente directa) variable para sujetar la señal de voz que ingresará al sistema de adquisición digital.

### Conversión de la señal de voz

Se realiza la conversión de la señal de voz analógica en señal digital, mediante la programación del microcontrolador Dspic 30f4013, incorporado en el módulo dspicdem II de Microchip.

### Almacenamiento de las muestras digitales

Utilizando la técnica de nivel de energía se detecta la presencia de la señal de voz en la entrada del conversor, y es allí donde se habilita el sistema para tomar las muestras que son almacenadas temporalmente en la memoria SRAM (memoria estática de acceso directo).

### Manejo de los módulos *xbee*

Con las muestras almacenadas en la SRAM, luego de la conversión y mediante la programación del dspic, se extraen las muestras de la memoria y se transmiten en forma serial desde el microcontrolador hacia el módulo de transmisión inalámbrica *xbee pro*, el cual está programado mediante el software x-ctu para trabajar full dúplex,

permitiendo la conexión con el otro *xbee pro* en el PC en el que se realizará el procesamiento.

### Descripción del software de extracción de patrones y reconocimiento de voz

Se describen los algoritmos que se van a utilizar para la FFT y la RNA, luego se desarrollan y se implementan dichos algoritmos en los programas *Matlab* y *Labview* para proceder a tomar en forma serial las muestras de voz provenientes del módulo de adquisición de usuario; luego estas son identificadas y el software permite retornar un código de identificación al módulo de adquisición.

### Validación

Los códigos retornados al módulo de adquisición de usuario permitirán el control de iluminación planteado como aplicación de este trabajo y la toma de datos como soporte del funcionamiento del sistema en general. Se obtiene el porcentaje de acierto de las palabras reconocidas y se procede a determinar la fiabilidad del sistema.

A continuación, se mostrará en detalle el conjunto de métodos, técnicas y herramientas que se emplearon en el proceso de implementación del sistema propuesto.

### Captura de la voz en forma analógica

Las señales de entrada al sistema corresponden a las palabras aisladas “uno”, “dos”, “tres”, “cuatro” y “cinco” pronunciadas por un único hablante, en un ambiente de bajo nivel de ruido. En el análisis de estas señales se debe tener en cuenta un procesamiento en el dominio del tiempo y de la frecuencia. Las figuras 3, 4, 5, 6 y 7 muestran el comportamiento de las palabras mencionadas en el dominio de la frecuencia.

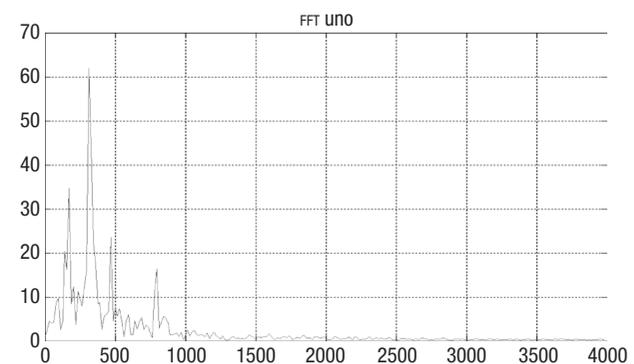


Figura 3. Palabra “uno” en el dominio de la frecuencia

Fuente: los autores

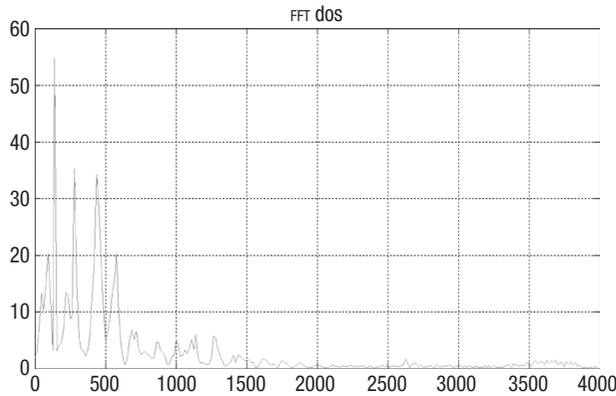


Figura 4. Palabra “dos” en el dominio de la frecuencia

Fuente: los autores

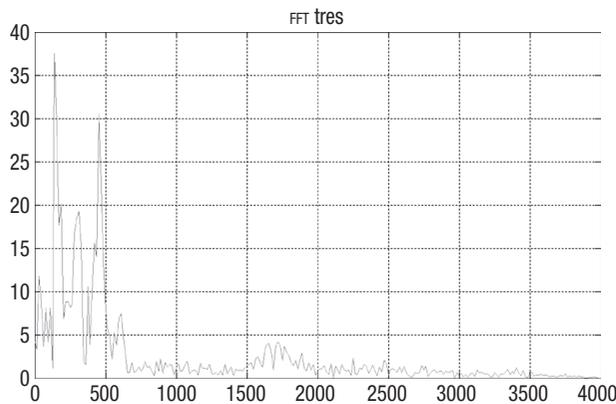


Figura 5. Palabra “tres” en el dominio de la frecuencia

Fuente: los autores

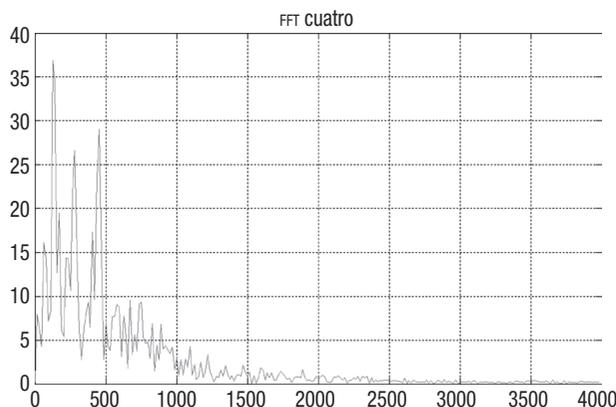


Figura 6. Palabra “cuatro” en el dominio de la frecuencia

Fuente: los autores

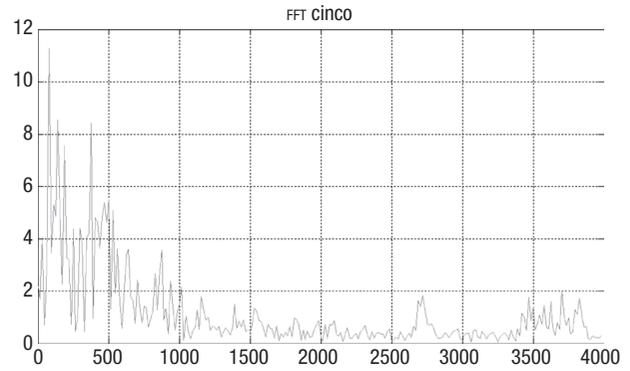


Figura 7. Palabra “cinco” en el dominio de la frecuencia

Fuente: los autores

La adquisición de la señal en el tiempo inicia en el transductor de entrada (micrófono electret), el cual convierte la energía acústica de la señal de voz en energía eléctrica. Físicamente, este micrófono requiere ser polarizado mediante una fuente de alimentación de corriente directa.

Un ambiente de conversación se puede caracterizar por un medio de comunicación específico y un nivel o condición de ruido [2], por lo cual la señal eléctrica que entrega el micrófono es una señal variable que puede llegar a contener frecuencias por encima del valor máximo permitido para un canal de voz, además de ser una señal pobre en amplitud debido a la naturaleza del transductor. Es por esto que se hace necesario implementar un selector de frecuencias (filtro) que permita obtener los principales componentes de dicho canal hasta 4kHz [3] y rechazar las que estén por encima, con unos niveles de voltaje adecuados, que permitirán ajustar la señal a los valores permitidos por el módulo de conversión análogo a digital. Dicho filtro presenta la siguiente función de transferencia:

$$H_{PB}(s) = \frac{1}{\frac{s^2}{\omega_0^2} + \frac{s}{Q\omega_0^2} + 1}$$

Ecuación 1

Donde  $\omega_0$  es la frecuencia natural de la función y  $Q$  es el factor de calidad.

La respuesta del filtro seleccionado en el dominio de la frecuencia se presenta en la figura 8.

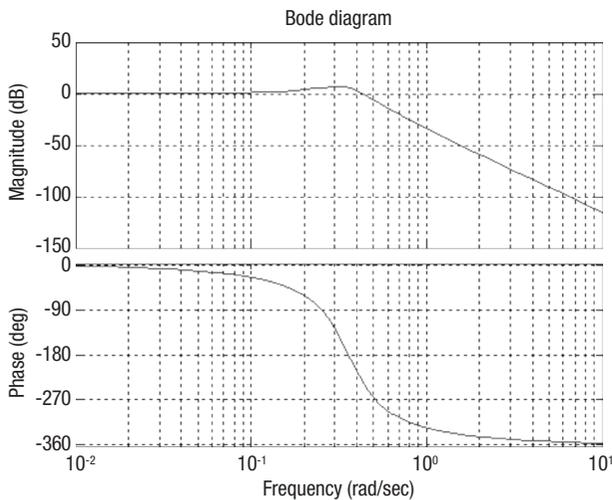


Figura 8. Respuesta en frecuencia del fpb implementado

Fuente: los autores

## Conversión de la señal de voz

Se requiere procesar la señal en un sistema que permita realizar operaciones con velocidad elevada, tamaño reducido y alta capacidad de memoria; estas prestaciones las proporcionan los microprocesadores. Para realizar el procesamiento de la señal sobre un sistema digital de este tipo se requiere hacer una discretización con un convertidor de analógico a digital. En el mercado son diversas las opciones que permiten realizar esta labor, sin embargo, se busca un dispositivo que cumpla con las velocidades necesarias para el muestreo de la voz. El teorema de muestreo de Nyquist se muestra en la *ecuación 2* [4].

$$F_m \geq 2F$$

Ecuación 2

Donde  $F$  es la frecuencia máxima de la señal a muestrear y  $F_m$  es la frecuencia de muestreo. Teniendo en cuenta la frecuencia superior de corte del filtro, se tendrá una frecuencia de muestreo de 8 khz, lo cual implica que se deberán tomar 8.000 muestras por segundo con un periodo de muestreo de 125 microsegundos.

Existen dispositivos que permiten integrar unidad de procesamiento, memoria de conversión analógico-digital y otra serie de recursos, adecuándose en gran medida a varios de los requerimientos de este trabajo. En este caso, se hace referencia a los microcontroladores. Estos dispositivos digitales se fabrican para diferentes aplicaciones con diferentes niveles de complejidad. La línea DSC (digital signal controller) está diseñada, entre otras, para

el diseño de aplicaciones relacionadas con adquisición y procesamiento del audio.

Debido a esto y a muchas otras características, se seleccionó el dspic 30f4013, incluido en el modulo dspicdem II, del cual se presentan algunas características.

- Arquitectura de 16 bits
- Velocidad CPU: 30 MIPS
- Memoria de PGM: 48 kb
- RAM: 2048 b
- Pines I/O: 30
- Conv. A/D: 12b@200 ksp/s
- Múltiples secuencias de conversión, entre otras

A partir de estos y otros aspectos, el dspic cumple con la mayoría de criterios de diseño; sin embargo, el tamaño de la RAM es reducido para esta aplicación, ya que se requiere almacenar como mínimo 8.000 muestras que correspondan a un segundo de grabación de voz, por tanto, debe usarse una RAM externa que permita dicho proceso de almacenamiento.

Una vez determinado el microcontrolador, el proceso que sigue es el de convertir la señal de analógico a digital, lo cual implica hacer muestreo, retención, cuantificación y codificación de la señal en un rango de conversión de 0 a 5 voltios. Este resultado permite convertir la señal analógica de entrada en números binarios de 12 bits, de los cuales solo se utilizan los ocho bits de más peso, obteniendo una resolución menor que no afecta en gran medida el reconocimiento de voz, pero que facilita el almacenamiento y la transmisión de las muestras.

Este convertidor provee un periodo de muestreo mínimo de cinco microsegundos y tiene hasta 16 entradas analógicas multiplexadas en un amplificador de muestreo y retención. El módulo se basa en el manejo de registros de propósito específico que controlan la operación del convertidor. En particular, los registros ADCHS Y ADPCFG seleccionan los canales de entrada y configuran los pines del puerto como entradas analógicas o entradas/salidas digitales, respectivamente. Por medio de la configuración de estos registros se pueden escoger diferentes formas de manejar el convertidor analógico digital [5].

## Almacenamiento

Considerando la cantidad de muestras que se utilizaran y la velocidad a la que deberán ser procesadas para evitar retardos significativos en el desempeño del sistema, se elige la memoria SRAM ut62256, cuyo diagrama de bloques es mostrado en la figura 9.

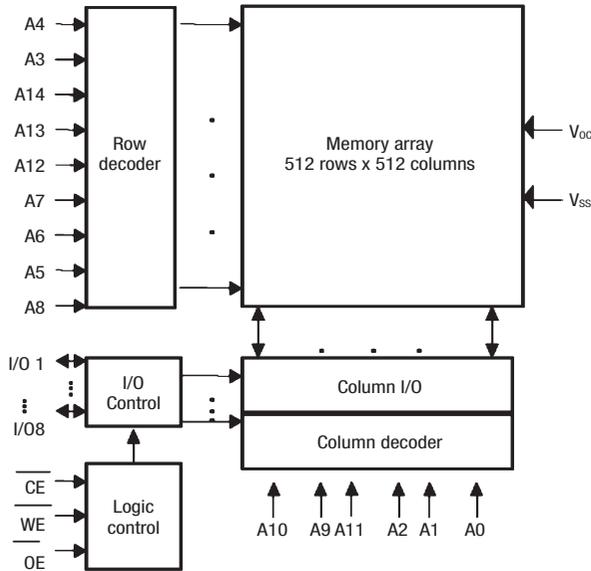


Figura 9. Esquema general Memoria sram ut62256

Fuente: Utron [6]

Esta elección se debe a que cumple con todos los requerimientos para el almacenamiento de datos (32 K x 8) y los tiempos de lectura y escritura son del orden de 35 a 70 nanosegundos. A continuación se muestra un resumen de las características:

- Tiempos de acceso de 35 a 70 nseg máx
- Bajo consumo de energía
- Todas las I/O con tecnología TTL
- Tamaño de 32.768 palabras de 8 bits
- Fuente de alimentación de 5 v

El algoritmo desarrollado para el acceso a la memoria debe cumplir con los tiempos estipulados por el fabricante. En la figura 10 se muestra el diagrama de tiempos para el proceso de lectura controlando los pines CE (chip enable) y OE (output enable):

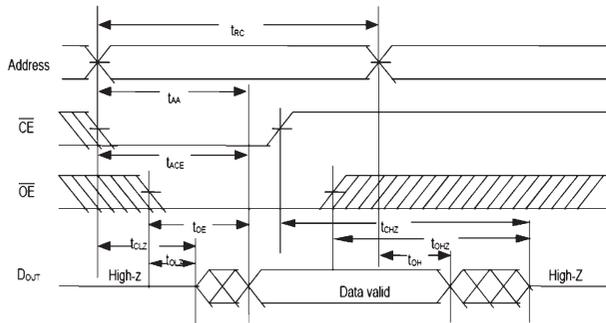


Figura 10. Esquema general Lectura Memoria SRAM ut62256

Fuente: Utron [5]

En la figura 11 se muestra el diagrama de tiempos para el proceso de escritura controlando el pin: WE (write enable):

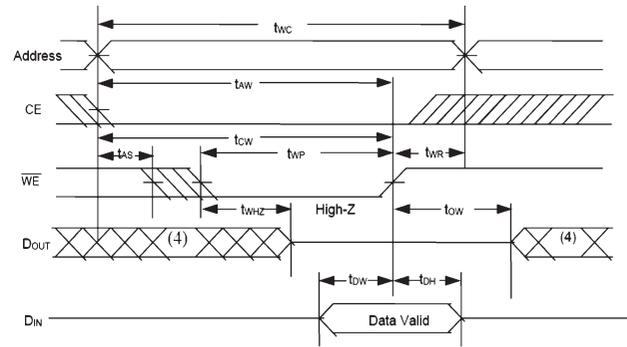


Figura 11. Esquema general Escritura Memoria sram ut62256

Fuente: Utron [5]

### Configuración del módulo de transmisión-recepción inalámbrica xbee-pro

Uno de los requerimientos clave para el desarrollo de este trabajo es la comunicación entre el módulo de adquisición y almacenamiento temporal de datos con el sistema de procesamiento y reconocimiento (computador), debido a que se encuentran distantes entre sí e incluso pueden cambiar de posición en un espacio determinado. Una de las soluciones que ofrece el mercado se basa en el protocolo de comunicación inalámbrica zigbee; este protocolo se basa en el estándar IEEE\_802.15.4. para redes inalámbricas [7].

Entre las ventajas del uso de dispositivos basados en comunicación zigbee, tal como el xbee-pro, se encuentran la operación en la banda libre de frecuencias de 2.4 GHz, el bajo consumo de energía, las distancias de los enlaces (pueden llegar incluso hasta una milla), pueden trabajar múltiples dispositivos en red, instalación económica y simple, y bajo costo, entre otras.

Algunas de las características técnicas del xbee-pro utilizado son:

- Alcance espacio cerrado 100 metros
- Alcance espacio abierto 1600 metros
- Potencia de salida +18dbm
- Tasa de transferencia 250kbps
- Voltaje de alimentación 3 a 3.4 Vdc

Como puede observarse, la tasa de transferencia de datos hace el módulo ideal para la transmisión de voz. El requerimiento de este diseño es de 64 kbps, ya que se

maneja una frecuencia de muestreo de 8.000 muestras por segundo y cada muestra tiene un tamaño de 8 bits. Además, el alcance en espacio cerrado es apropiado para aplicaciones domóticas y, en este caso, orientadas a personas con discapacidad. La velocidad utilizada fue 115.200 bps, ya que debe estar acorde con el estándar rs232 usado por el puerto del PC.

En la figura 12 se muestra el módulo de comunicación elaborado para este desarrollo, haciendo uso del *xbee-pro*.

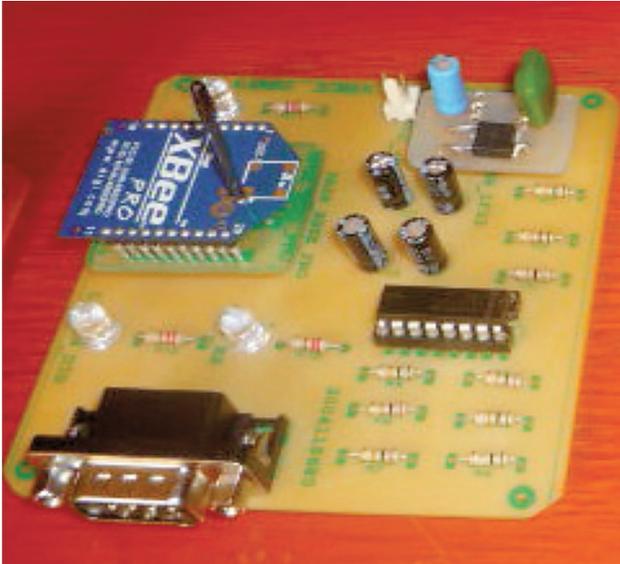


Figura 12. Módulo transceptor elaborado

Fuente: los autores

## Descripción del software de extracción de patrones

En el procesamiento de la señal de voz, se implementa una parte en el módulo de hardware de adquisición de

usuario, y otra se implementa en las herramientas de software utilizadas *labview* y *matlab*.

En el módulo hardware se implementan los algoritmos de detección de inicio y fin de la palabra. Para esta tarea, se emplea el microcontrolador dspic 30f4013, que también se encarga de realizar la digitalización de la señal analógica de entrada.

Para detectar el inicio de la palabra pronunciada, se emplea la técnica de nivel de energía, que consiste en determinar si el nivel de energía de la señal ha superado un umbral determinado. La energía de una señal discreta es definida como [8]:

$$E = \sum_{m=-\infty}^{\infty} x^2[n]$$

Ecuación 3

El nivel de energía umbral depende del nivel de ruido del recinto. Una vez se ha detectado el inicio de la palabra, comienza el almacenamiento de las muestras de voz en la memoria; en total, se almacenan 8.000 muestras después de detectado el inicio, es decir, un segundo de grabación, tiempo durante el cual se estima que se ha pronunciado la totalidad de la palabra. Al tener las muestras almacenadas en la memoria, se procede a ejecutar un algoritmo también basado en la técnica del nivel de energía, estableciendo el final de la palabra y limitando la cantidad de muestras transmitidas en forma inalámbrica al PC. En el algoritmo de detección de final de palabra se divide la señal en ventanas de 20 milisegundos, que corresponde a tomar 160 muestras para determinar su energía y nuevamente verificar si están por debajo de un umbral establecido; si se encuentra una cantidad de ventanas consecutivas por debajo del umbral, se concluirá que la palabra ha terminado (ver figuras 13 y 14).

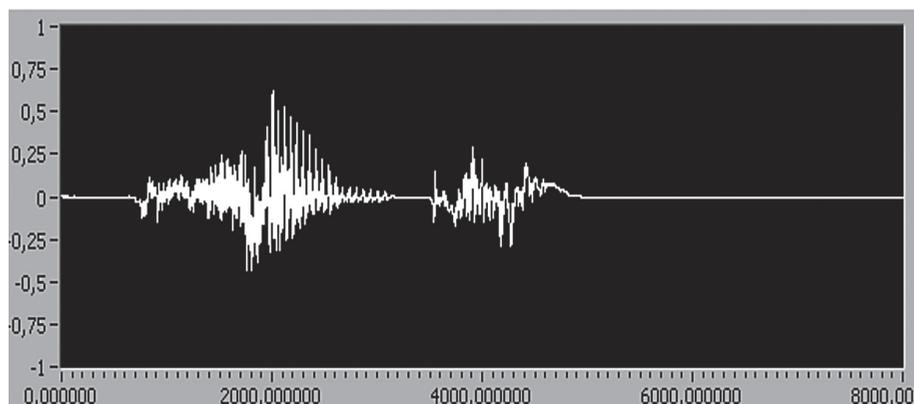


Figura 13. Palabra "cinco" sin recortar en el tiempo

Fuente: los autores

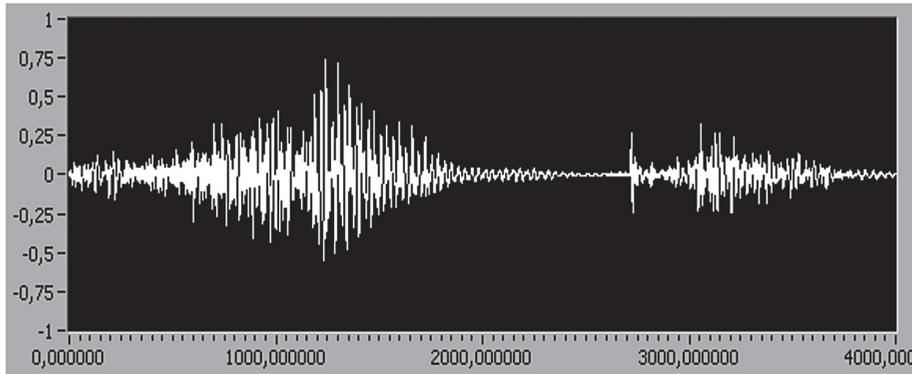


Figura 14. Palabra “cinco” recortada en el tiempo

Fuente: los autores

Posteriormente a la finalización del algoritmo para recorte de la palabra, se leen los datos de la memoria correspondientes a la información útil de la palabra y se envían en forma serial desde uno de los módulos uart del microcontrolador al módulo *xbee-pro*, quien se encarga de transmitirlos al PC en forma inalámbrica.

Los datos son recibidos por un segundo *xbee-pro* que se encuentra conectado al PC mediante el puerto serial, de ahí en adelante el procesamiento y la identificación de la palabra se realizará únicamente sobre el software desarrollado para tal fin.

Labview permite recoger los datos por medio de instrumentos virtuales que vienen definidos para tal fin. Una vez se recibe la totalidad de las muestras, se busca suavizar los extremos de la palabra mediante la ventana hamming descrita por la ecuación 4 [9].

$$h[n] = 0,54 - 0,46 \times \cos\left(2\pi \times \frac{n}{N}\right), \text{ si } 0 \leq n \leq N$$

$$h[n] = 0 \text{ en otros casos}$$

Ecuación 4

El siguiente paso es la extracción del patrón que posteriormente será identificado. Existen diferentes técnicas para la extracción de estos patrones, en este caso se utilizó la transformada de Fourier en tiempo discreto para obtener el espectro de frecuencias y determinar parámetros tales como frecuencias fundamentales y formantes. La transformada discreta de Fourier se puede obtener mediante la ecuación 5 [10].

$$X(k) = \sum_{n=0}^{N-1} x(n) \times e^{-j(2\pi \times n/N)}$$

Ecuación 5

La transformada rápida de Fourier (FFT) es un algoritmo que permite calcular la transformada discreta de Fourier para encontrar el espectro de una señal a un coste computacional bajo, esta implementación se realiza sobre  $2^n$  muestras. Este trabajo se desarrolló con una transformada rápida de Fourier de 512 puntos. El espectro obtenido se convierte a escala logarítmica para realzar las componentes de baja amplitud, se normaliza y se divide en 20 ventanas, cada una con el mismo ancho de frecuencia; para cada ventana se obtiene la energía total de ese segmento y se normaliza nuevamente. Al final de este proceso se obtienen 20 valores normalizados de energía, correspondientes a los 20 intervalos tomados del espectro.

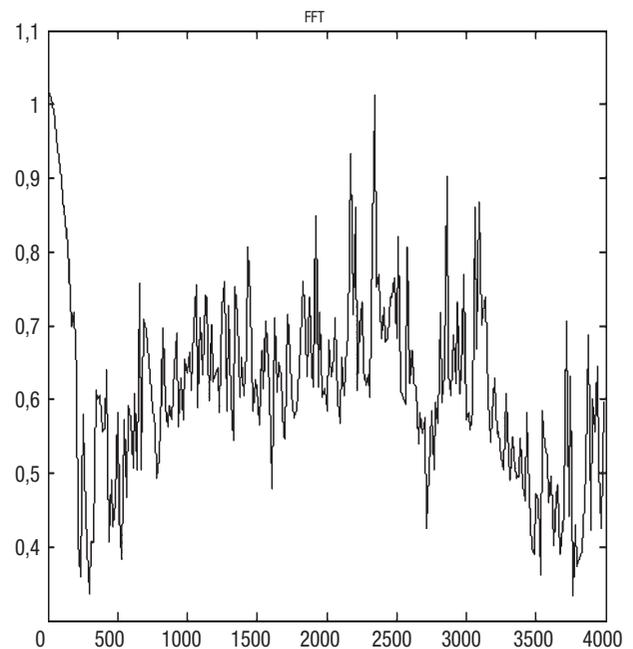


Figura 15. Espectro en escala log y normalizado palabra “cinco”

Fuente: los autores

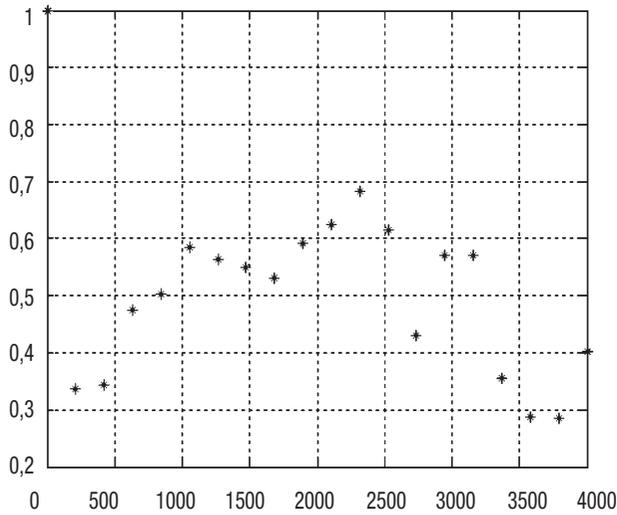


Figura 16. Patrón de energía normalizado palabra "cinco"

Fuente: los autores

Estos 20 valores representan el patrón que será identificado. En la figura 15 se observa el espectro de una palabra pronunciada, en contraste con la figura 16 que muestra el patrón de energía obtenido después del proceso antes mencionado. Con algún nivel de análisis, podrían determinarse los formantes de la palabra a partir de la figura 16. Cada una de las palabras pronunciadas (uno, dos, tres, cuatro, cinco) presenta un patrón en el cual se hace evidente la variación de la energía en las diferentes ventanas. En las figuras 17 y 18 se presenta uno de los espectros y los niveles de energía obtenidos.

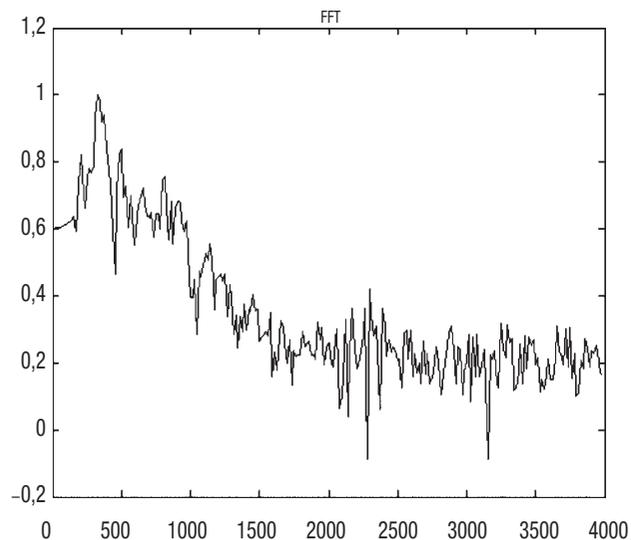


Figura 17. Espectro en escala log y normalizado palabra "uno"

Fuente: los autores

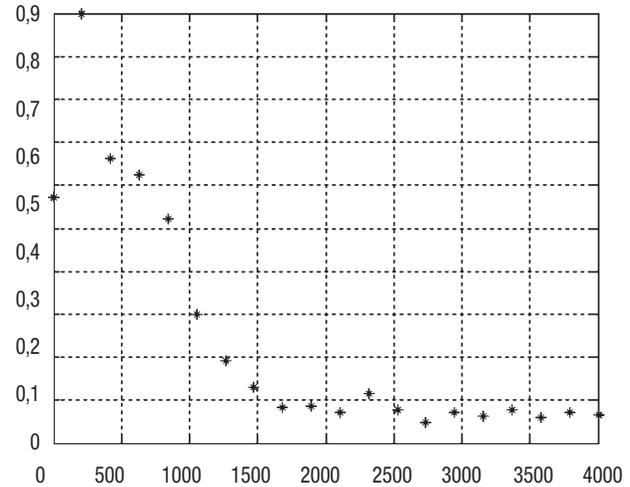


Figura 18. Patrón de energía normalizado palabra "uno"

Fuente: los autores

## Reconocimiento del patrón de voz

Finalmente, el patrón de energía obtenido se ingresa a una red neuronal artificial (RNA). Las redes neuronales artificiales son algoritmos que tienen que ver con procesamiento inteligente de datos o patrones. Este tipo de algoritmos pretende, en parte, imitar el funcionamiento del cerebro humano, por este motivo las redes neuronales artificiales son capaces de aprender de la experiencia, pueden extraer características importantes de diferentes patrones de entrada, mediante una etapa de entrenamiento inicial pueden llegar a aprender a realizar una tarea, son tolerantes a fallos y a pérdidas de información, permitiendo reconocer patrones incompletos con un buen nivel de acierto [11].

Existen diferentes tipos de redes neuronales artificiales, para este trabajo se utilizó una red neuronal *backpropagation* (o de retropropagación del error). Esta red neuronal es de tipo *off line*, es decir, requiere una etapa de entrenamiento previa a su funcionamiento normal. Durante este proceso de aprendizaje se modifican los pesos de la red. Estos pesos se asemejan a las conexiones entre las neuronas biológicas y es allí donde finalmente se almacena el conocimiento de la red. Para una red neuronal artificial, los pesos son números que se encuentran entre las conexiones de las neuronas. En el proceso de entrenamiento se presenta una cantidad de patrones con sus respectivas salidas deseadas. Durante un proceso iterativo la red neuronal actualiza sus pesos hasta obtener un error bajo para cada uno de los patrones de entrada con relación a la salida deseada.

Para esta red neuronal se utilizaron tres capas: la capa de entrada es en la que se recibe el patrón de 20 parámetros, por tanto, la capa debe tener 20 neuronas. La capa intermedia u oculta puede variar en cantidad de neuronas dependiendo de la respuesta que se tenga después de un entrenamiento; esto indica que el proceso para determinar la cantidad de neuronas se hace a prueba y error. El mejor resultado para este trabajo fueron 20 neuronas en la capa oculta. La cantidad de neuronas de la capa de salida depende en este caso del número palabras a identificar, porque a pesar de tener una salida sigmoideal para todas las neuronas, en la capa de salida se aproxima la salida de cada neurona a “cero” o “uno”, obteniendo

una salida binaria. Como el sistema debe identificar cinco palabras, entonces se tendrán tres neuronas de salida que permiten  $2^3$  combinaciones posibles, como se muestra a continuación:

- 000 salida deseada de la palabra “uno”
- 001 salida deseada de la palabra “dos”
- 010 salida deseada de la palabra “tres”
- 011 salida deseada de la palabra “cuatro”
- 100 salida deseada de la palabra “cinco”

El diagrama de la red neuronal implementada se muestra en la figura 19.

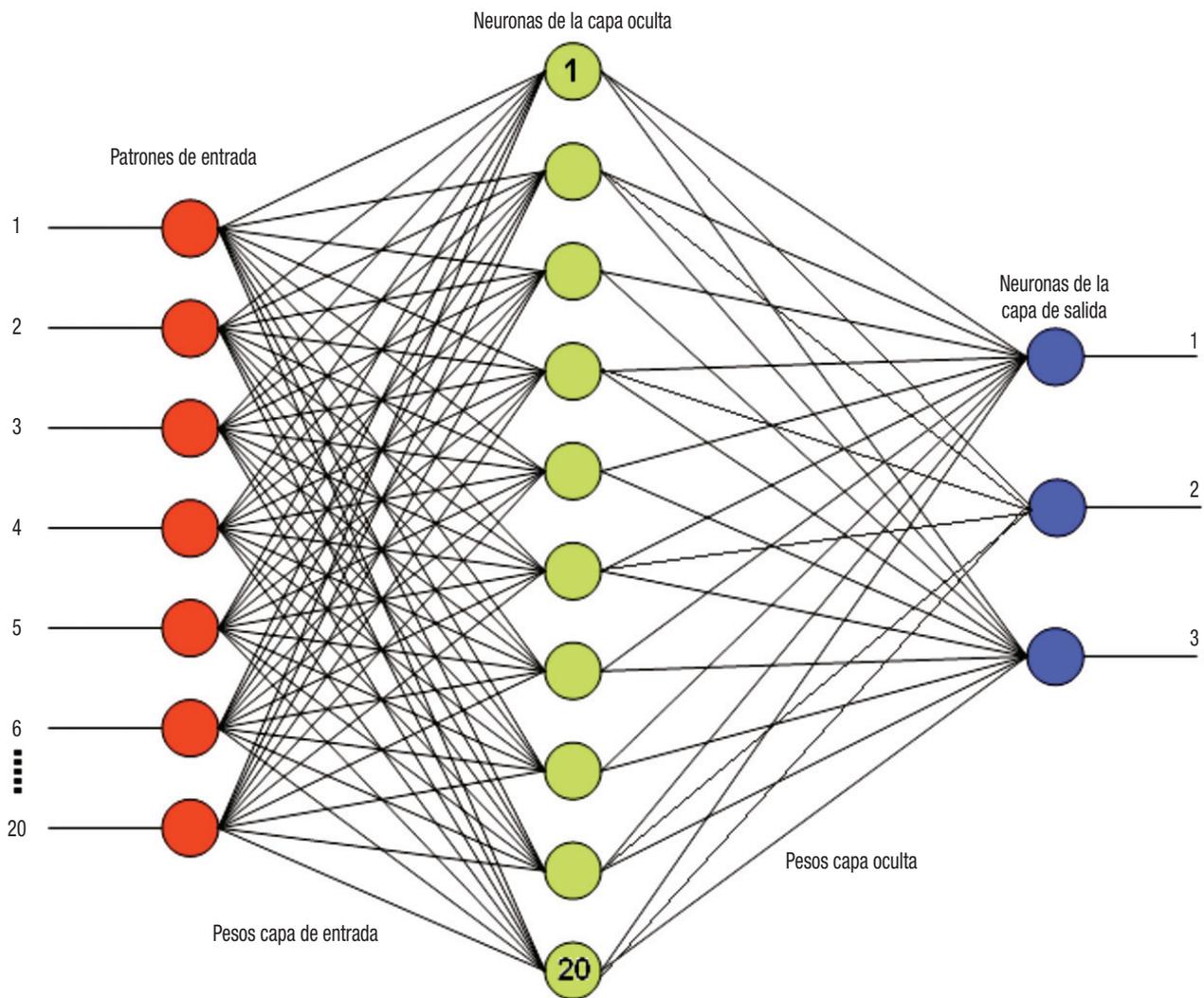


Figura 19. Diagrama red neuronal *backpropagation* implementada  
Fuente: los autores

Una vez identificada la palabra, se envía al módulo de adquisición de usuario el código binario obtenido y este toma la decisión de encender o apagar la luz a la cual se hace referencia.

## Resultados

Se tomó un total de 100 muestras por cada palabra pronunciada por un único hablante hombre en un ambiente poco ruidoso, obteniendo como resultado un 92% de acierto en general para la identificación de las cinco palabras. En la tabla 1 se muestran los resultados específicos de cada palabra.

Tabla 1. Resultados específicos por palabra

	Uno	Dos	Tres	Cuatro	Cinco
Uno	80%		10%	5%	5%
Dos	10%	90%			
Tres			100%		
Cuatro		5%	5%	90%	
Cinco					100%

Fuente: los autores

Se encuentra, en general, un error del 8%, que puede deberse a ruido externo, problemas de vocalización, cansancio del locutor, entre otras.

## Conclusiones

El sistema funciona en forma adecuada, con un bajo nivel de error generado por el módulo de identificación de voz, que no afecta en gran medida el normal funcionamiento de la aplicación. De esta manera, se concluye que el sistema puede ser utilizado por personas con discapacidad motora, elevando en alguna medida su calidad de vida.

Este sistema puede ser empleado para controlar remotamente otro tipo de dispositivos, logrando controlar diferentes variables del entorno y electrodomésticos, entre otros. Para esto, se hace necesario ampliar el vocabulario de palabras reconocidas y, por tanto, se debe profundizar en el estudio de otras técnicas de procesamiento y reconocimiento de la señal de voz que permitan hacerlo más eficiente y robusto.

La selección de la cantidad de neuronas en la red es clave para incrementar hasta cierto punto el porcentaje de aciertos, pero a su vez se convierte en un proceso demorado debido a que cada cambio en la estructura de la

red requiere que se haga un nuevo entrenamiento y un nuevo proceso de prueba de esta.

La transformada rápida de Fourier (FFT), a pesar de ser un método básico en el tratamiento de señales, se convierte en una herramienta poderosa para el proceso de identificación de voz.

La evolución de tecnologías inalámbricas permite el desarrollo de aplicaciones en diferentes áreas y, en este caso, permite la distribución de tareas facilitando y agilizando los procesos.

## Referencias

- [1] L. Rabiner. "Applications of voice processing to telecommunications", *Proc. IEEE*, vol. 82, núm. 2, 1994, pp. 199±230.
- [2] Maren, A.J., Harston C.T., R.M. Pap. *Handbook of neuronal computing applications*. Academic Press. 1990, pp. 20-40.
- [3] L. Cruz, M. Acevedo. *Reconocimiento de voz usando redes neuronales artificiales back propagation y coeficientes LPC*. Ciudad de México: CICos. 2008, pp. 89-99. ISBN:978-607-00-0165-9.
- [4] M.E. Esparza, B. Avalos. *Reconocimiento de voz, conciencia tecnológica*. Vigésimo segunda edición. México: Instituto Tecnológico de Aguas Calientes. 2003.
- [5] Z. García, I. Begoña, I. Angulo, J.M. Martínez, U. Angulo. *Microcontroladores avanzados DSPIC. Controladores digitales de señal. Arquitectura, programación y aplicaciones*. Madrid: Paraninfo. 2006, pp. 112-200.
- [6] "Utron technology inc". Utron. Consultado: 15 de enero del 2011, disponible en: <http://www.aufzu.de/semi/utron.html>
- [7] J.C. Valverde. *El estándar inalámbrico Zigbee*, Perú: Universidad Nacional de Trujillo. 2007. Disponible en: <http://seccperu.org/files/ZigBee.pdf>.
- [8] G. Santos. *Inteligencia artificial y matemática aplicada: reconocimiento automático del habla*. Segunda edición. Valladolid: Universidad de Valladolid. 2001, pp. 40-80.
- [9] B. Planer. *An introduction to speech recognition*. Munich Germany. 2005. Consultado: 10 de enero del 2011. Disponible: [www.speech-recognition.de/pdf/introsr.pdf](http://www.speech-recognition.de/pdf/introsr.pdf)
- [10] C. Torres, L. Mattos, G. Perpiñan, J. Castro, J. Pardo. "Reconocimiento automático del habla utilizando la transformada rápida de Fourier y redes neuronales", *Revista Colombiana de Tecnologías de Avanzada*, vol. 2, núm. 6, 2005, pp. 1595-1598.
- [11] J.R. Hiler, V.J. Martínez. *Redes neuronales artificiales, fundamentos, modelos y aplicaciones*. Madrid: Alfa Omega. 2000, pp. 11 – 65.