Programación Matemática y Software (2009)
Vol. 1. No 1. ISSN: 2007-3283

Recibido: 1 de Mayo de 2008/Aceptado: 20 de Mayo de 2008
Publicado en línea: 26 de Junio de 2009

Aplicación del Reconocimiento de Voz de un Hablante Mediante una Red Neuronal Artificial Backpropagation y Coeficientes LPC sobre un Canal Telefónico

Luis. A. Cruz-Beltrán[A] and Marco. A. Acevedo-Mosqueda
[1] SEPI-Telecomunicaciones ESIME IPN Unidad Profesional "Adolfo López Mateos".
Col. Lindavista, 07738, México. D. F
lcruzb06@ipn.mx, macevedo@ipn.mx

Abstract. En esta investigación se presenta un método para el reconocimiento de personas en un canal telefónico. El cual se basa en el comportamiento de las Redes Neuronales Artificiales (RNA), en particular, sobre la arquitectura del Perceptrón Multicapa mediante el algoritmo de aprendizaje Backpropagation aunado con la aplicación de la teoría de las Wavelets (Haar, Daubechies y Coiflet) y el empleo de los Coeficientes de Código de Predicción Lineal (LPC) como parámetro para extraer las principales características de cada hablante. En este trabajo se utilizan archivos de voz con formato estándar *.Wav. El método propuesto es implementado mediante el software de programación de Matlab.

Keywords: RNA, Backpropagation, Wavelets Haar, Coiflet y Daubechies, LPC, Wav.

1 Introducción

El reconocimiento automático de hablantes es un campo aún abierto a la investigación a nivel mundial debido a sus grandes complicaciones aleatorias que presenta. Actualmente existen diversos sistemas comerciales de reconocimiento de palabras continuas, y discretas implementados con diferentes métodos como lo son las Cadenas Ocultas de Markov [1-2], Análisis de Principales Componentes [3], Distorsión Dinámica Temporal ó Dynamic Time Warping (DTW) [4], Redes Neuronales [5-9] etc., pero con problemas de entrenamiento y falta de robustez en situaciones reales, por lo que sigue la búsqueda de técnicas adecuadas en la selección de parámetros, que permita hacer más óptima su compresión y clasificación.

[A] Autor de correspondencia

En los sistemas de reconocim iento de voz no se int enta, como mucha gente piensa, reconocer lo que el usuario dice, sino identificar una serie de sonidos y sus características principales para saber si el hablante es quien dice s er. El tamaño de la "frase" en el reconocimiento de v oz afecta su com plejidad, requerimientos de proceso y precisión del sistema. De esta manera surgen dos tipos de sistemas para el reconocimiento de voz, los cuales son: discreto ó sistema dependiente del locutor (frases cortas) y continúo ó siste ma independiente del locutor (frases largas).

Sin embargo también es i mportante tomar en cue nta los siguientes cinco factores que determinan la complejidad de un siste ma de reconocimiento de voz.

Locutor: Es uno de los factores que introduce mayor variabilidad en la forma de la señal de voz, y por lo tanto se requiere que el siste ma sea altamente robusto. Debido a que u na persona no pronuncia siempre de la misma forma las palabras, esto se debe a distintas situaciones físicas y psicológicas est e efecto se conoce como (variabilidad intr a-locutor). Existe además una gran variedad entre distintos locutores (hombr es, mujeres, niños, etc.), diferencias según la edad ó la región de origen (variabilidad interlocutor).

La forma de hablar: Es el segundo factor que determina la complejidad de un reconocimiento de voz. El hablante pronuncia las palabras de una forma continúa, y de acuerdo a la inercia de los órgano s articulatorios, que n o pueden moverse instantáneamente. Ello, aunado a las variaciones introducidas por la prosodia, hace que una palabra al principio de una frase s ea diferente que cuando se dice en medio, ó que sea diferente d ependiendo de que sea lo que le precede o le sigue de acuerdo al contexto.

El Vocabulario: Es el número de palabras diferentes que debe reconocer el sistema, mientras mayor es el número de ellas más difícil es l a implementación, por dos motivos principales. El primero porque al au mentar el número de las palabras es más fácil que aparezcan palabras parecidas entre si, y el segundo porque el tiempo de procesamiento incrementa al aumentar el número de palabras con las que comparar.

La Gramática: Es el conjunto de reglas que lim ita el número de combinaciones permitidas de las pal abras del vocabulario. En general la existencia de una gramática en un reconocedor de voz ayuda a mejorar la tasa de reconocimiento, al eliminar am bigüedades y puede ayudar a di sminuir la necesidad de cálculo, al limitar el número de palabras en una determinada fase del reconocimiento

El entorno físico: Es una parte tan i mportante como las anteriores, esto s e debe al hecho de que no es lo m ismo un sistema que funciona en un ambiente poco ruidoso, como puede ser el despac ho de un medico, ó en contraparte al que tiene que funcionar en un coche, en una fabrica, la línea telefónica, etc.

El objetivo de esta investigación es la im plementación de un siste ma de reconocimiento voz de tipo discreto para la verificación o identificación de personas en un canal telefónico, em pleando su patrón de voz, par a lo cual se propone una metodología que emplea las características del patr ón de voz, Wavelets, LPC y una RNA mediante al algoritm o de aprendizaje Backpropagation. Para poder solucionar ef icientemente un caso jurídico en el cual se encuentra inmiscuida una grabación telefónica del a cusado como prueba del caso y se necesita de un sistema que autentifique y corrobore si efectivamente la voz que se encuentra en la grabación pertenece al inculpado que se encuentra en el proceso penal.

Este documento se divide en las siguientes sec ciones. En la Sección 2 se presenta un panorama general de las RNA y los coeficientes LPC. La Sección 3 muestra la arquitectura de la RNA Backpropagation. En la Sección 4 se da una explicación detallada del sistema propuesto. En la Sección 5 se presentan los resultados obtenidos. Final mente, la Se cción 6 es dedicada a las conclusiones de este artículo.

2 Redes Neuronales Artificiales

Las RNA son sistemas de procesamiento de información cuy a estructura y funcionamiento están inspirados en las redes neuronales biológicas [6]. En todo modelo de RNA se tienen cuatro elementos básicos.

- Un conjunto de conexio nes, pesos o sinapsis que determ inan el comportamiento de la ne urona, las cuales pueden ser excitado ras, presentan un signo positivo (conexiones positivas) y las inhibidoras presentan un signo negativo (conexiones negativas).
- Una función que se encarga de sumar todas las entradas multiplicadas por sus pesos correspondientes.
- Una función de activación que puede ser lineal o no lineal empleada para limitar la amplitud de la salida de la neurona.
- Una ganancia exterior qu e determina el umbral de activación de la neurona.

Desde que el psicólogo Frank Rosenblatt en 1957 [6] introdujo el modelo del perceptrón de una sola capa, el cual so lo resolvía problemas de carácter lineal, marcó la pau ta para que se implementaran diferentes arquitecturas y diseños de las RNA una de ellas la más popular es la arquitectura del perceptrón Multicapa con el alg oritmo de aprendizaje Backpropagation [7-9], convirtiéndose así en una herramienta poderosa para solucionar diversos tipos de problemas relacionados con la clasificación, estimación funcional y optimización del reconocimiento de patrones.

El modelo propuesto se o bserva en la ecuación (1), donde $x_{p1},\ldots, x_{pj}$ son las unidades de entrada, $w_{ji},\ldots,w_{ji}$ son los pesos de la RNA, b_i es la ganancia ó umbral de activación, N_{pj} es el producto de los pesos con respecto a la entrada, f es la función de activación de la RNA y finalmente y_{pj} es la salida de la RNA, estas variables se relacionan en la siguiente expresión:

$$y_{pj} = f(N_{pj} = \sum_{i=1}^{m} x_{pi} w_{ji} + b_i), \quad para \quad m \in IR,\ m < \infty. \qquad (1)$$

2.1 Código de Predicción Lineal

Una gran parte de las aplicaciones rel acionadas con el trata miento del habla, están basadas en el anál isis de LPC, dado que es capa z de extraer la

información lingüística y eliminar la correspondiente a la persona particular. La predicción lineal modela la zona vocal humana como una respuesta al impulso infinita, que produzca la señal de voz.

El término predicción lineal se refiere al método para predecir ó aproximar una muestra de una señal en el dominio del tiempo *s[n]* basada en varias muestras anteriores *s[n - 1], s[n - 2], s[n -M]*.

$$s[n] \approx \hat{s}[n] = -\sum_{i=1}^{M} a_i s[n-i]. \qquad (2)$$

donde *s[n]* es llamada señal muestreada, y a_i, $i = 1,2,\ldots,M$ son los predictores ó coeficientes LPC. Un pequeño número de coeficientes LPC $a_1, a_2,\ldots,a_M$ pueden ser usados para representar eficientemente una señal *s[n]* [8-9]. Los valores $a_1, a_2,\ldots,a_M$ son la base para la realización de este trabajo debido a que nos ayudan a modelar los parámetros de la voz de cada uno de los hablantes que se emplean en este sistema propuesto.

3 La Red Backpropagation

En 1986, Rumelhart, Hinton y Williams, basados en otros trabajos formalizaron un método para que una red neuronal aprendiera la asociación que existe entre los patrones de entrada y las clases correspondientes, utilizando más niveles de neuronas que los que utilizó Rosenblatt para desarrollar el Perceptrón.

Este método es conocido como Backpropagation (retropropagación del error) que es un tipo de red con aprendizaje supervisado, el cual emplea un ciclo propagación-adaptación de dos fases.

Una vez aplicado un patrón de entrenamiento a la entrada de la red, este se propaga desde la primera capa a través de las capas subsecuentes de la red, hasta generar una salida, la cual es comparada con la salida deseada y se calcula una señal de error para cada una de las salidas, a su vez esta es propagada hacia atrás, empezando de la capa de salida, hacia todas las capas de la red hasta llegar a la capa de entrada, con la finalidad de actualizar los pesos de conexión de cada neurona, para hacer que la red converja a un estado

que le permita clasificar correctamente todos los patrones de entrenamiento. La estructura general se muestra en la Figura 1.

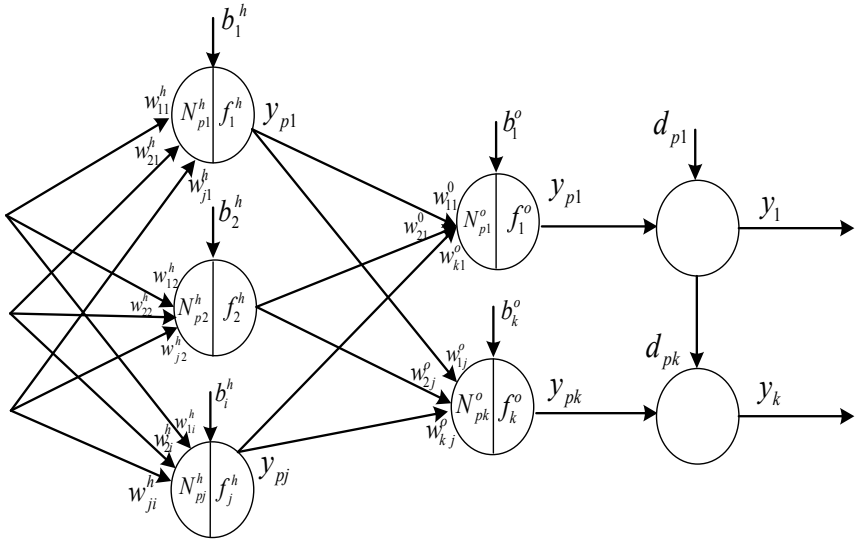

Fig. 1. Modelo de la RNA Backpropagation.

3.1 Algoritmo de entrenamiento de la Red.

A continuación se presenta el algoritmo empleado para el entrenamiento de la RNA Backpropagation. [6-7], [10-11].

1. Inicializar los pesos de la red (w) con valores aleatorios pequeños.
2. Mientras la condición de paro sea falsa realizar los pasos (3-6).
3. Se presenta un patrón de entrada, ($x_{p1}, x_{p2},\ldots,x_{pi}$) y se específica la salida deseada que debe generar la red ($d_{p1}, d_{p2,}\ldots,d_{pk}$).
4. Se calcula la salida actual de la red, para ello se presentan las entradas a la red y se va calculando la salida que presenta cada capa hasta llegar a la capa de salida ($y_1,y_2,\ldots,y_k$). Los pasos son los siguientes:
   a) Se determinan las entradas netas para las neuronas ocultas procedentes de las neuronas de entrada.

$$N_{pj}^{h} = \sum_{i=1}^{m} w_{ji}^{h} x_{pi} + b_{i}^{h}. \tag{3}$$

b) Se aplica la función de activación a cada una de las entradas de la neurona oculta para obtener su respectiva salida.

$$y_{pj} = f_{j}^{h}(N_{pj}^{h} = \sum_{i=1}^{m} w_{ji}^{h} x_{pi} + b_{i}^{h}). \tag{4}$$

c) Se realizan los mismos cálculos para obtener las respectivas salidas de las neuronas de la capa de salida.

$$\begin{aligned} N_{pk}^{o} &= \sum_{j=1}^{m} w_{kj}^{o} y_{pj} + b_{k}^{o}; \\ y_{pk} &= f_{k}^{o}(N_{pk}^{o} = \sum_{j=1}^{m} w_{kj}^{o} y_{pj} + b_{k}^{o}). \end{aligned} \tag{5}$$

5. Determinación de los términos de error para todas las neuronas:

a) Cálculo del error (salida deseada–salida obtenida).

$$e = (d_{pk} - y_{pk}). \tag{6}$$

b) Obtención de la delta (pro ducto del error con la derivada de la función de activación con respecto a los pesos de la red).

$$\delta_{pk}^{o} = e * f_{k}^{o'}(N_{pk}^{o}). \tag{7}$$

6. Actualización de los pesos. Se emplea el algoritmo recursivo del gradiente descendente, comenzando por las neuronas de salida y trabajando hacia atrás hasta llegar a la capa de entrada.

a) Para los pesos de las neuronas de la capa de salida:

$$\begin{aligned} w_{kj}^{o}(t+1) &= w_{kj}^{o}(t) + \Delta w_{kj}^{o}(t+1); \\ \Delta w_{kj}^{o}(t+1) &= miu\delta_{pk}^{o} y_{pj}. \end{aligned} \tag{8}$$

b) Para los pesos de las neuronas de la capa oculta:

$$w_{ji}^{h}(t+1) = w_{ji}^{h}(t) + \Delta w_{ji}^{h}(t+1);$$
$$\Delta w_{ji}^{h}(t+1) = miu\delta_{pj}^{h} x_{pi}\,. \tag{9}$$

7. Se cumple la condición de pa ro (error mínimo ó número de iteraciones alcanzado logrado).

4 Algoritmo Empleado

La Figura 2, muestra el sistema propue sto, el cual consiste de tres etapas: la etapa de la captura de l a señal de v oz del canal telefónico, la etapa de preprocesamiento de la señal y finalmente la etapa de verif icación del hablante usando las características extraídas en las dos primeras etapas.

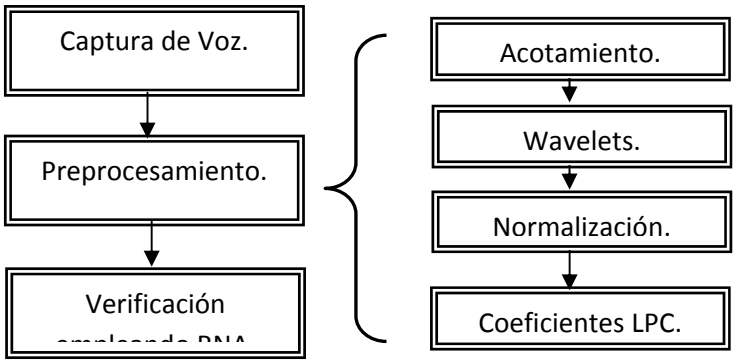

Fig. 2. Sistema Propuesto.

4.1 Captura de Voz

Con la finali dad de cumplir con l os parámetros mencionados arriba para obtener un sistema de reconocim iento fiable, pr oponemos que para la realización de la c aptura de la voz se registren 5 veces la fr ase "Zoológico" con cinco personas difere ntes, cada una de ellas gr abó la m isma frase con diversos estados de ánimos (alegre, tr iste, eufórico, melancólico etc.). Lo s

hablantes fueron Luis, Orlando, Alejandro, Diana y Leydi de 23, 29, 30, 5 y 22 años de edad, respectivamente.

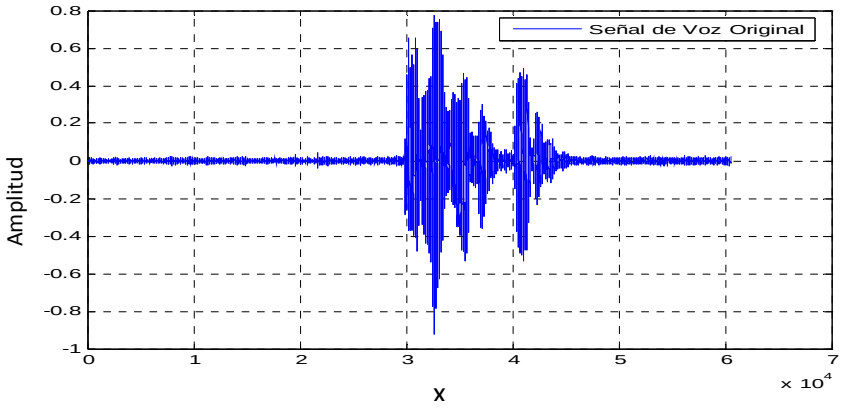

Fig. 3. Señal de voz grabada**.**

Se escogió la palabra "Zoológico" debido a que esta contiene la mayor parte de los formantes de la voz, gran cantidad de características espectrales. El procedimiento se llevó a cabo de la siguiente manera: Se instaló el software Mercury en una LapTop Dell con Windows XP a 1.73 GHz y 730 MB de RAM, conectada a la línea telefónica vía MODEM. Después, cada uno de los 5 hablantes realizó una llamada telefónica al número de la casa donde se encontraba conectada la PC a la línea telefónica. Con el software Mercury se grabó en la PC la conversación con la palabra Zoológico pronunciada desde la caseta telefónica. La conversación contiene el ruido ambiental y el ruido del canal telefónico. Este proceso se realizó 5 veces para cada uno de los 5 hablantes, obteniendo así 25 archivos monofónicos (de un solo canal) de audio que fueron convertidos en formato *.Wav por su versatilidad de manejo con el software Matlab, cada uno de los 25 archivos tienen las características mostradas en la Tabla 1.

Cabe resaltar que se empleó una velocidad de muestreo de sonido de 11 KHz, con la finalidad de cumplir con el criterio de Nyquist que es mayor ó igual a 2 veces la frecuencia de muestreo, que para nuestro caso pertenece a la del canal telefónico que aproximadamente tiene un rango de frecuencias que se encuentra comprendida entre los 300 Hz y 3,400 Hz.

Tabla 1. Características de cada archivo de voz.

Velocidad de Transmisión	128 Kbps.
Tamaño de muestra de sonido	16 bits.
Tipo de canal	Monofónico
Velocidad de muestreo de sonido	11 KHz.
Formato de audio	*.Wav

4.2 Preprocesamiento

El objetivo de esta etapa es acondicionar la señal de entrada preservando la mayor cantidad posible de las características más relevantes de la misma, para que esta pueda ser procesada adecuadamente por la RNA y no se presenten problemas de inestabilidad en la etapa de entrenamiento, aprendizaje y reconocimiento de la RNA. Para poder llevar a cabo estas tareas primero acotamos la señal de voz eliminando la parte inicial y final de la misma ya que estas partes no contienen ninguna información relevante del hablante, de esta manera obtenemos la señal de voz a la cual le aplicaremos las Wavelets, con la finalidad de compactar y eliminar parte del ruido presente en la señal, empleando solo la subseñal de bajas frecuencias que nos da como resultado de aplicar las Wavelet, para finalmente normalizar esta señal y extraerle los coeficientes LPC que nos arrojaran los coeficientes con las principales características de la voz los cuales serán introducidos como los parámetros de las neuronas de entrada para la primera capa de la RNA a implementar.

La etapa del preprocesamiento de la señal de voz consiste de los pasos mostrados en la Figura 2, a continuación detallamos brevemente cada uno de estos pasos los cuales son:

Acotamiento de la señal: En esta etapa se eliminan las muestras de tiempo que solo contienen "silencios" diferentes a las características acústicas de cada hablante, como lo son principalmente el ruido ambiental, ruido eléctrico,

ruido térmico etc., por lo general estas se encuentran al principio y al final de los archivos de audio como se observa en la Figura 4 donde apreciamos dos silencios el A y el B.

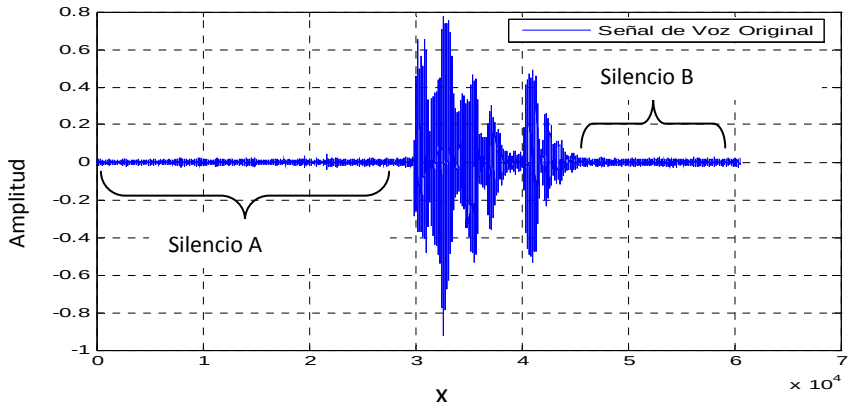

Fig. 4. Silencios en el archivo de audio.

El procedimiento para cancelar estos silencios consiste en calcular la energía del archivo de audio, ya que de esta manera claramente podemos observar en que partes se encuentran los segmentos sordos y los segmentos sonoros de nuestra grabación de voz, debido a que los valores de ambas características varían en amplitud de acuerdo a las peculiaridades de cada hablante como lo son su entonación, volumen de voz, ritmo, las pausas y su estado de ánimo. Una vez que ya hemos calculado la energía de nuestra señal calculamos el valor de la amplitud máxima de la energía de la señal y proponemos un umbral que en nuestro caso corresponde al 10% del valor de la amplitud máxima encontrada y en base a este valor hacemos un recorrido de derecha a izquierda de la señal de audio original hasta encontrar un valor que sea mayor al umbral propuesto, una vez encontrado ese valor lo tomamos como referencia de inicio de nuestra nueva señal eliminando todos los componentes que se encuentren por debajo de ese umbral con lo cual cancelamos los silencios A de nuestra señal de audio. De igual manera realizamos un recorrido de derecha a izquierda con la nueva señal y procedemos a eliminar todas las muestras de derecha a izquierda a partir del punto mayor a nuestro umbral propuesto eliminando así el silencio B de la señal de voz. El resultado de este proceso se observa claramente en la Figura 5.

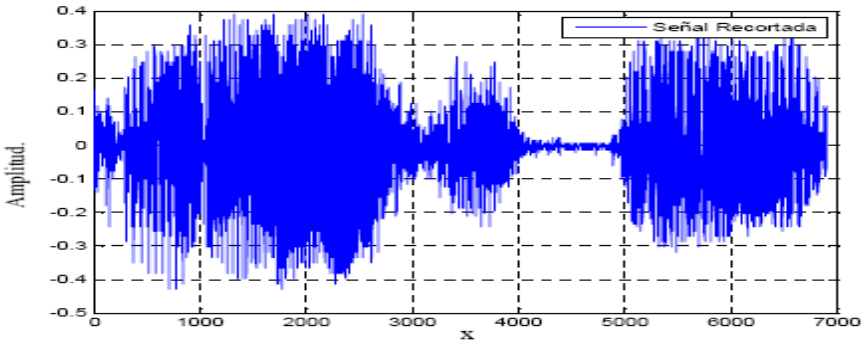

Fig. 5. Recorte de la señal de audio.

Wavelets: La Transformada Wavelet (W T), es una herramienta de las matemáticas aplicadas, esta técni ca fue des arrollada a partir de la transformada de Fourier (FT) hacia fi nales de los 80's. Se utiliza para l a descomposición de señales sobre un conjunto de funciones base obtenidas por dilataciones y traslaciones de una fun ción principal llamada wavelet madre [12].

Toda señal d e voz en la Naturaleza se encuentra af ectada por ru ido, y la señal de voz del canal telefónico no es la excepción. Por tal motivo se emplean las Wavelets para reducir este efecto. De la Figura 6 podemos observar que $x[n]$ corresponde a la señal acotada. Aplicando la Wavelet a la señal $x[n]$ obtenemos dos subseñale s las cual es son: la subseñal $a[n]$ que correspondiente a las bajas frecuencias de la señal de voz do nde se localiza la mayor cantidad de energí a e infor mación de la misma, despreciándose la subseñal $b[n]$ para nuestros cálculos, ya que corresponde a las altas frecuencias y es en esta parte donde se en cuentra la mayor cantidad de ruido de la señal (ruido am biental, el ruido del canal telefónico, etc.). Obteniendo así una señal de voz compacta y filtrada con respecto a la original.

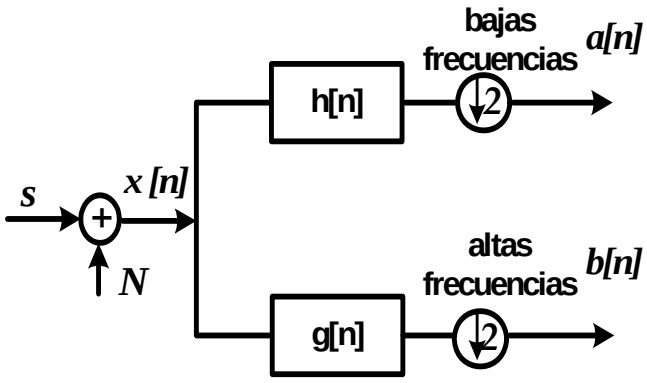

Fig. 6. Estructura de la Wavelet.

En este trabajo se propone emplear tres tipos de Wavelets las cuales son Haar, Coiflet y Daubechies 45. Observando de acuerdo a la Tabla 2 que la mejor de ellas es la Wavelet Daubechies 45 debido a que, presenta el mayor porcentaje de compactación de energía en la subseñal $a[n]$ para cada uno de los veinticinco archivos de nuestra base de datos, de acuerdo a los resultados presentes en la tabla podemos observar claramente que la Wavelet Haar con respecto a la Coiflet prácticamente tienen el mismo porcentaje de compactación de energía lo que nos permite descartarlas y poder emplear la Wavelet Daubechies 45 por su mayor compactación de energía en cada uno de los 25 archivos de la base de datos, en comparación con las dos anteriores Wavelets.

Tabla 2. Características de cada archivo de voz.

COMPRESION DE ENERGÍA CON WAVELETS			
Nombre de Archivo.	Haar	Coiflet	Daubechies 45
Lb1	80.27%	89.45%	202.35%
Lb2	74.94%	80.72%	200.45%
Lb3	83.13%	93.21%	212.00%
Lb4	83.47%	87.48%	201.15%

Lb5	80.37%	87.45%	198.56%
Aa1	98.01%	101.13%	114.64%
Aa2	98.76%	105.47%	122.73%
Aa3	97.72%	102.40%	120.44%
Aa4	98.07%	101.88%	116.63%
Aa5	96.33%	99.70%	122.55%
Le1	91.79%	96.67%	141.86%
Le2	91.99%	96.89%	139.70%
Le3	86.53%	94.78%	173.36%
Le4	92.33%	97.01%	134.93%
Le5	93.97%	97.87%	126.85%
Oc1	95.84%	98.71%	117.98%
Oc2	96.02%	98.71%	118.01%
Oc3	96.31%	98.78%	116.50%
Oc4	96.24%	98.73%	117.51%
Oc5	96.31%	98.78%	116.50%
Dd1	54.98%	64.53%	194.51%
Dd2	61.79%	76.92%	272.13%
Dd3	62.99%	75.49%	243.80%
Dd4	84.90%	94.76%	210.59%
Dd5	63.18%	76.50%	315.09%

Normalización: La normalización consiste en ajustar todos los parámetros a una sola escala para que al momento de ser utilizados por la RNA no causen

problemas de estabilidad, en este caso la escala empleada se encuentra dada por los parámetros de la función de activación de la RNA que es una tangente bipolar sigmoidal y trabaja con valores de [-1,1], por lo tanto cada uno de los 25 archivos que previamente fueron compactados y filtrados por medio de las Wavelets son normalizados a esta escala, como se observa en la ecuación (10) donde se propone que los datos normalizados tengan la siguiente características, la media sea igual a cero y la desviación estándar la unidad, donde los datos que se quieren normalizar se encuentran dentro del vector $x[i]$, con i=$1,\ldots,n$. El procedimiento a seguir es el siguiente:

a) Se calcula la media (u) y la desviación estándar (σ) del vector $x[n]$.

b) Se normalizan los datos según la relación:

$$\dot{x}[n] = \frac{x[n]-\mu}{\sigma} \tag{10}$$

c) Se calculan el máximo y el mínimo del vector $\dot{x}[n]$, se divide por el de mayor valor absoluto y los datos normalizados caen dentro del intervalo $[-1,1]$.

Es importante normalizar previamente los datos de entrada para evitar saturaciones en las neuronas de la capa oculta, otra ventaja de esta normalización es que permite que la búsqueda del gradiente durante la etapa de aprendizaje se realice de manera más efectiva ya que las pendientes de la función del gradiente son mayores en comparación de las pendientes si los datos no estuvieran normalizados. Los resultados se ilustran en la Figura 7.

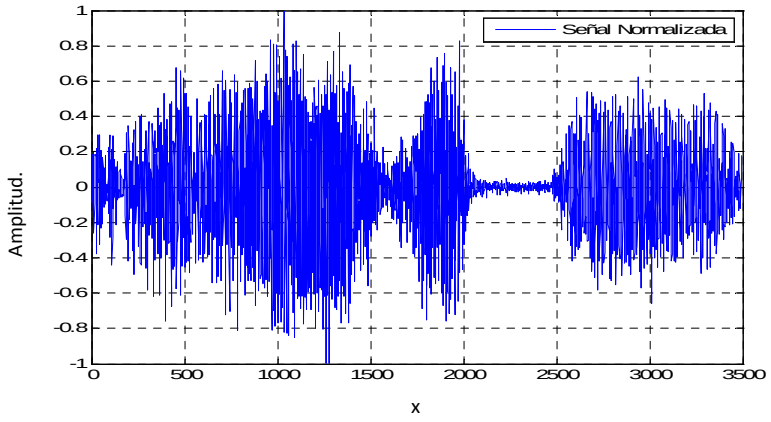

Fig. 7. Señal de audio normalizada.

Extracción de los coeficientes LPC: Debido a las propiedades mencionadas en la Sección 2 de los coeficientes LPC, en particular a que son capaces de modelar con gran aproximación la información lingüística y la zona vocal humana, en este trabajo se propone emplear diferentes números de coeficientes descritos en la ecuación (2) formando así una matriz de *25 x n* elementos que corresponden a la extracción de *n* coeficientes para cada uno de los 25 archivos formados por los hablantes, creando de esta manera el patrón de entrenamiento de la RNA, por consiguiente se cambian el número de capas de entrada y oculta de la RNA con los siguientes resultados, como se observa en la Tabla 3.

Tabla 3. Coeficientes LPC.

COEFICIENTES LPC			
LPC	**Capa Entrada**	**Capa Oculta**	**Efectividad**
2	3	8	60%
3	4	9	80%
4	5	10	100%
5	6	10	100%
6	7	10	100%
10	11	8	100%
15	16	8	100%

4.3 Diseño de la Red Neuronal Artificial

Esta etapa consiste en dos partes, la primera de ellas es el entrenamiento de la RNA, la cual se lleva a cabo con la finalidad de modificar los pesos de la red en cada una de las capas, de manera que coincida la salida deseada por el usuario con la salida obtenida por la red ante la presentación de un determinado patrón de entrada.

La segunda consiste en una fase de validación de la red frente a cualquier patrón de entrada que le sea presentado. Se empleó una arquitectura Backpropagation con tres capas, la capa de entrada, oculta y la de salida.

Fase de entrenamiento: Para el correcto desempeño de est a fase se emplearon, los valores establecidos para la capa de entrada y oculta mostrados en la Tabla 3 con los siguientes parámetros:

1) Neuronas de la capa de entrada=25.
2) Neuronas de la capa oculta=21.
3) Neuronas de la capa de salida=5.
4) Número de entrenamientos=25.
5) Número de épocas=700.
6) Pesos de la c apa de entrada y de la capa oculta. (valores dentro de un rango de ([2.4 -2.4]) /Neuronas de entrada).[10]
7) Patrón de entrenamiento.
8) Salida deseada.
9) Error cuadrático medio requerido=0.005.
10) Tasa de aprendizaje =.009, .05, 0.02.

Bajo estos parámetros y basándonos en la sección 3 en donde se explica detalladamente el funcionamiento de la RNA se entrenó a la misma, una vez entrenada se evalúa la RNA con el número de entrenamientos propuestos para generar y guardar los pesos de la capa oc ulta y de salida ya entrenados para emplearse en la próxima etapa.

Fase de evaluación: Se abren los pesos guardados obtenidos para la capa oculta y de salida del proceso de entrenamiento, se definen los puntos (1-4,6 y 7) de la fase de entrenamiento, se evalúa la red con un solo patrón de entrenamiento el cual es el objetivo a iden tificar dentro de nuestra RNA, si el patrón de entrenam iento se encuentra la RNA lo id entifica con uno de los posibles hablantes e mpleados en el entrena miento de acuerdo a la s características de los valores de sus p esos, sino se encuentra dentro de los hablantes empleados en el entrenamiento de la red se e mite un mensaje de error indicando que la persona no ha podido ser identificada.

5 Resultados obtenidos

Tabla 4. Resultados obtenidos.

PRUEBAS CON DIFERENTES ESTADOS DE ANIMO					
Hablantes	Alejandro	Leydi	Orlando	Diana	Luis
	A1=Ide	Le1=Ide	O1=Ide	D1=Ide	Lb1=Ide
	A2=Ide	Le2=Ide	O2=Ide	D2=Ide	Lb2=Ide
	A3=Ide	Le3=Ide	O3=Ide	D3=Ide	Lb3=Ide
	A4=Ide	Le4=Ide	O4=Ide	D4=Ide	Lb4=Ide
	A5=Ide	Le5=Ide	O5=Ide	D5=Ide	Lb5=Ide
Reconocimiento	100%	100%	100%	100%	100%
Efectividad=					100%

La Tabla 4 muestra los resultados obtenidos en esta investigación tomando en cuenta los valores propuestos en la fila 3 de la Tabla 3, con lo cual observamos que nuestros resultados son bastante ideales debido a que obtenemos una efectividad del 100%.

Continuando con nuestras pruebas, al momento de evaluar la RNA con los archivos de voz sin que estos hayan pasado por la etapa del preprocesamiento los resultados obtenidos en efectividad disminuyen del 100% al 96%.

Graficando las variaciones de la tasa de aprendizaje obtenemos diferentes valores de error para el proceso de entrenamiento de la RNA, que se muestran en la Figura 8. De la gráfica observamos que los mejores valores para la construcción de la RNA, son los de la línea de color azul (alfa=0.009,0.05, 0.02) ya que con ellos obtenemos los mínimos errores en la RNA.

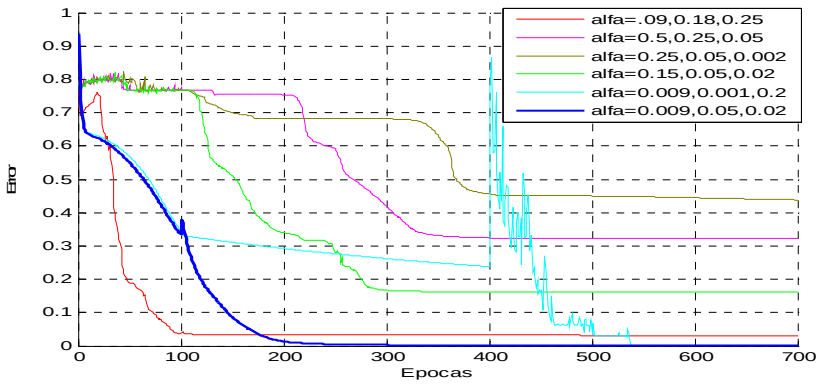

Fig.8. Gráfica del error.

6 Conclusiones

Con el sistema propuesto en este trabajo se tiene un buen funcionamiento para esta aplicación ya que obtenemos un reconocim iento de l os hablantes del 100% como se da a notar en la tabla de resultados.

Cabe destacar que el procedim iento para la obtención de los valores de los parámetros empleados en el diseño de la RNA Backpropagatio n no existen como tales bien definidos, sin embargo, los valores propuestos en este trabajo fueron obtenidos a prueba y error dándon os cuenta de cómo la tasa d e aprendizaje y la elección c orrecta de los pesos iniciales influy e mucho en el resultado obtenido.

Podemos observar que con sólo 4 c oeficientes es ideal para aproxim ar correctamente una señal de voz. El siste ma propuesto presenta una estructur a fácil de desarrollar y su complejidad matemática es mínima, por lo que puede tener diversas aplicaciones en el campo de la identificación y verificación del hablante.

Cabe mencionar que em pleando este método los resultados obtenidos son viables debido a que e n trabajos anteriores en este t ema con otras metodologías propuestas como lo son las cadenas ocultas de Markov, la cuantización vectorial etc, los resultad os obtenidos no superan el 95 % de efectividad.

Referencias

1. L. R. Rab iner. *A tutorial on hidden markov models and selected applications in speech recognition*. Proc. of IEEE, Vol. 77, No. 2, pp. 257–286, Febrero 1989.

2. A. Amano, N. Hataoka, H. Kokubo. *Development of robust speech recognition middleware on microprocessor*. Proc. of IEEE, Vol. 26, pp.837–840, 1998.

3. David Mustafa Kay nak and Q. Zhi. *Analysis of lip geometric features for audio-visual speech recognition*. IEEE Transactions on Systems, Vol. 34, No. 4, pp.564–570, July 2004.

4. B. H. Juang. *The past, present, and future of speech processing*. Proc. of IEEE S ignal Processing, pages 24–48, May 1998.

5. R. Gallart, V . Moreno J. Bestard y M. Laucirica. *Algunas experiencias sobre reconocimiento de fonemas utilizando redes neuronales artificiales*. Bioingeniería: Ingeniería Electrónica, Automática y Comunicaciones, Vol. 21, No. 2, pp.79– 84, Mayo 2000.

6. José R Hilera Martínez, "R edes Neuronales Artificiales, Fundamentos, Modelos y Aplicaciones", Alfa Omega, México, (2000)

7. Simon Haykin, "Neural Networks", Prentice - Hall, New Jersey, (1999)

8. Sadaoki Furui, " *Digital Speech Processing Synthesis, and Recognition*, Cambridge University Press, (2001).

9. José Luis Oropeza Rodr íguez y Sergio Suárez Guerra. *Algorithms and methods for the automatic speech recognition in spanish language using syllables*. Computación y Sistemas, Vol. 9, No. 3, pp. 270–286, 2006.

10. Bonifacio Martín del Rio, Alfredo Sanz Molina. " *Redes Neuronales y Sistemas Borrosos*", Ra-Ma, Madrid, (2001)

11. Laurene Fausett. " *Fundamentals Neuronal Network, architectures, algorithms, and applications",* Prentice – Hall, New Jersey, (1995).

12. Stephane Mallat. *A Wavelet Tour of SignalProcessing,* Second Edition, New York. Academic Press (1999).