UN MODELO PARA EL RECONOCIMIENTO DEL HABLA UTILIZANDO WAVELETS

Ronald W. León Navarro
Docente del Departamento de Matemáticas
rowlen@yahoo.com

Jorge L. Guevara Díaz
Escuela de Informática
jorjasso@hotmail.com

Juan O. Salazar Campos
Escuela de Informática
josc_orlando@hotmail.com

Facultad de Ciencias Físicas y Matemáticas de la Universidad Nacional de Trujillo
Av. Juan Pablo II S/N; Ciudad Universitaria, Trujillo – Perú, América del Sur

RESUMEN

El reconocimiento del habla es un área de investigación en Inteligencia Artificial. En los últimos años muchos trabajos de investigación están contribuyendo en la solución del problema, aún con los avances alcanzados todavía no se logra la alta fidelidad deseada. Una de las herramientas matemáticas cuyo aporte en la solución de muchos problemas que se presentan en la industria y en las ciencias son las wavelets, particularmente son útiles para la representación y segmentación de señales no estacionarias, análisis tiempo - frecuencia, y son de fácil implementación, de rápidos algoritmos computacionales. En el presente trabajo utilizamos la transformada rápida wavelet de Haar para construir un patrón de características de la señal que produce cuando una persona habla. Esta señal previamente es tratada para eliminar partes inservibles en el proceso de reconocimiento. Se usa segmentación uniforme y redes neuronales de propagación hacia atrás (backpropagation) en la construcción del sistema de reconocimiento. El primer prototipo construido es solo para ser utilizado por personas entre 18 y 65 años, y para reconocer solo vocales y números. Se proyecta un segundo prototipo para reconocer palabras aisladas.

1. INTRODUCCIÓN

A todo lo largo de historia, el género humano ha utilizado su oído como un cuerpo receptor y analizador de sonidos.

El reconocimiento del habla ha sido una importante parte de la civilización por mucho tiempo. Nosotros dependemos de reconocer sonidos inteligentes para comunicarnos. Si cada silaba no es enunciada justo en la forma correcta nosotros seremos incapaces de vocalizar convenientemente nuestras ideas, pensamientos y emociones

Nuestros oídos reúnen información audible y luego lo pasan al cerebro para su procesamiento. Un elemento de información que el oído recoge es la amplitud del sonido, que es el nivel de excitación de los folículos que en la cóclea del oído experimentan. El segundo elemento de información que el oído recoge es el tono (tonalidad, inflexión de voz, modulación, timbre), esto es la medida de frecuencia, y es determinado en el oído como el número de cambios de presión que ocurre en un período de tiempo dado. El tercero, y quizás el elemento de información menos intuitivo que el oído recolecta, es el tiempo en que cada amplitud y frecuencia hacen su aparición, desde entonces nosotros tenemos una percepción mas general del paso del tiempo.

La importancia de la construcción de reconocedores de voz radica en la utilidad práctica que pueda dárselos, por ejemplo en la ejecución de una tarea especifica por parte de un computador o robot después de haberle "hablado" o dado una orden especifica mediante la voz; también establecer las bases para la construcción de reconocedores de voz sería útil en un sistema biométrico; un sistema reconocedor de voz, podría utilizarse para personas con discapacidad en hablar y ayudarles a expresar sus ideas, como también en los comunes procesadores de texto; es decir se podrían encontrar un sin número de aplicaciones útiles en diferentes campos de investigación

En consecuencia el objetivo de este proyecto es recalcar la importancia e incentivar la investigación en sistemas reconocedores de voz, presentando un método que permita que el computador reconozca voz humana, lo que se lograría en nuestro caso con la ayuda de las wavelets.

Ciertas señales cuya amplitud varía en forma rápida y abrupta en el tiempo o señales cuyo contenido de frecuencia es variable de un instante de tiempo a otro, las cuales son más conocidas como señales no estacionarias, no son analizadas completamente mediante la transformada de Fourier, debido a ciertas limitaciones de este análisis en el campo tiempo - frecuencia. Es en estos términos de análisis donde la utilidad de una nueva herramienta matemática llamada wavelet, cobra importancia.

La transformada wavelet es el resultado de un gran número de investigaciones y constituye una técnica de análisis reciente. Inicialmente un geofísico francés llamado Jean Morlet1 investigaba un método para modelar la propagación del sonido a través de la corteza terrestre. Como alternativa a la transformada de Fourier, Morlet utilizó un sistema basado en una función prototipo, que cumpliendo ciertos requerimientos matemáticos y mediante dos procesos denominados dilatación o escalamiento y traslación, formaba bases que permitían representar las señales de propagación con la misma robustez y versatilidad que la transformada de Fourier, pero sin sus limitaciones. La simplicidad y elegancia de esta nueva herramienta matemática fue reconocida por un matemático francés llamado Yves Meyer quien descubrió que las wavelets formaban bases ortonormales de espacios ocupados por funciones cuyo cuadrado es integrable, lo que traducido al lenguaje del procesamiento de señales, corresponde a funciones o señales cuyo contenido energético es finito. En este momento ocurrió una pequeña explosión de la actividad en esta área, ingenieros e investigadores comenzaron a utilizar la transformada wavelet para aplicaciones en diferentes campos tales como astronomía, acústica, ingeniería nuclear, detección de terremotos, compresión de imágenes, reconocimiento de voz, visión humana, neurofisiología, óptica, resonancia magnética, radar, etc. El término wavelet se define como una "pequeña onda" o función localizable en el tiempo, que visto desde una perspectiva del análisis o procesamiento de señal puede ser considerada como una herramienta matemática para la representación y segmentación de señales, análisis tiempo - frecuencia, y fácil implementación de rápidos algoritmos computacionales.

Las características propias de la transformada wavelet nos otorgan la posibilidad de representar señales en diferentes niveles de resolución, representar en forma eficiente señales con variaciones abruptas, analizar señales no estacionarias permitiéndonos saber el contenido en frecuencia de una señal y cuando estas componentes de frecuencia se encuentran presentes en la señal.

En nuestro ámbito local y en la parte norte del país, no existe un trabajo de tal tipo hecho por alguna universidad o centro de investigación y nos atreveríamos a decir que a nivel nacional no se ha hecho aportes significativos para la construcción de reconocedores de voz, el software existente es comprado del extranjero como es el caso de procesadores de textos que trabajan con patrones de voz; en el campo de la electrónica y la robótica, en el Perú no existen evidencias de trabajo alguno de tal tipo, y los trabajos de investigación en este tema en particular son escasos y por no decirlos nulos; esperamos contribuir con la presente investigación a la fomentación de creación de software que procese señales de voz humana, estableciendo tanto las bases teóricas y prácticas, que contribuirían con la ciencia y la industria pues tendría muchas aplicaciones prácticas, algunas de las cuales se mencionaron líneas arriba, y se espera aportar al inmenso campo de la Inteligencia Artificial, en su eterna e incansable búsqueda de simular el comportamiento humano a que algún día se establezca comunicación "verbal", por así decirlo, entre un computador y los humanos, y que en combinación con áreas como visión computacional, electrónica, biología, psicología y muchas otras, permitan construir "máquinas inteligentes" .

2. WAVELETS DE HAAR

Si bien las wavelets de Haar, que a continuación brevemente desarrollamos, no son la clase de wavelets que producen los mejores resultados, son las mas fáciles de estudiar e implementar, lo que nos facilitará en la construcción de nuestro primer prototipo, el cual nos servirá para estudiar y comprender algunos aspectos no considerados en el reconocimiento del habla propuesto. Luego podremos reemplazar este tipo de wavelets por otros, como las wavelets B-splines, en el prototipo construido para dar lugar a uno nuevo mejorado.

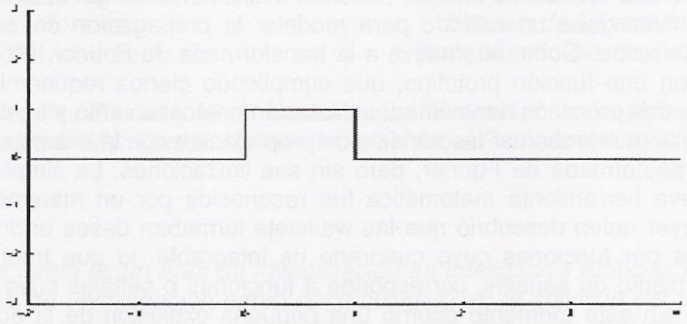
En esta sección desarrollaremos la teoría para las wavelets de Haar, en el caso continuo, para esto descompondremos una función $f \in L^2(R)$.

2.1 FUNCIÓN ESCALA DE HAAR

Sea la función $\varphi(t)$ perteneciente a $L^2(R)$, definida de la siguiente forma:

$$\varphi(t)=\begin{cases}1, & si\ 0\le t<1\\ 0, & otro\ valor\end{cases}$$

Esta función la denominaremos *función escala*, cuya gráfica se presenta en la siguiente figura.



Definimos entonces un conjunto de funciones escalas en términos de traslaciones enteras de la función básica de escalamiento $\varphi(t)$:

$$\varphi_k(t)=\varphi(t-k)=\begin{cases}1, & si\ t_1=k\le t<k+1=t_2\\ 0, & caso\ contrario\end{cases}$$

donde $k\in Z$ y $\varphi\in L^2(R)$.

El subespacio de $L^2(R)$ generado por estas funciones es definido como:

$$V_0=\overline{span_{k\in Z}}\{\varphi_k(t)\},$$

donde el superrayado denota la clausura, esto significa que:

$$f(t)=\sum_{k\in Z}c_k\varphi_k(t)$$

para cualquier $f\in V_0\subset L^2(R)$, en otras palabras cualquier función $f(t)$ que esté en V_0 puede ser representada por una combinación lineal del conjunto de funciones $\varphi_k(t)$.

Para un rápido cálculo de los coeficientes c_k es necesario que $\{\varphi_k(t)\}$ sea ortonormal. En efecto, si definimos la familia de funciones:

$$\varphi_m(t)=\varphi(t-m)=\begin{cases}1, & si\ t_3=m\le t\le m+1=t_4\\ 0, & en\ caso\ contrario\end{cases}$$

Es fácil verificar que:

$$<\varphi_k(t),\varphi_m(t)>=\int_R\varphi_k(t)\varphi_m(t)dt=0,$$

para todo $k\neq m$, lo que demuestra que la familia de funciones es ortogonal.

2.2 FUNCIONES WAVELETS

Se obtiene una mejor aproximación de la señal utilizando las funciones escala que ocupan el espacio V_1 que aquellas que pertenecen al espacio V_0. Sin embargo, las características de una señal pueden ser mejor descritas, no incrementado el tamaño del espacio de las funciones escala, sino definiendo espacios de funciones W_i levemente diferentes a los espacios escala, que representen la diferencia que existe entre un espacio V_{i+1} y un espacio V_i, o bien:

$$V_{i+1} = V_i \oplus W_i$$

por lo que podemos de decir que el espacio W_0 corresponde al complemento del espacio V_0 en el espacio V_1.

La función que genera el espacio W_0 se conoce como función wavelet, y se define de la forma:

$$\psi(t) = \begin{cases} 1, & si\ 0 \le t < \frac{1}{2} \\ -1, & si\ \frac{1}{2} \le t < 1, \\ 0, & otro\ valor \end{cases}$$

que al igual que la función escala que genera el espacio V_0, puede ser representada sobre el intervalo [0,1] como una combinación lineal de las funciones escalas que generan el espacio V_1.

Como se sabe los espacios V_0 y V_1 son ortogonales y por lo tanto cualquier espacio V_j, con i = 0; ±1; ±2; ... también lo es, entonces el espacio W_0 al ser el complemento de V_0 en V_1 es ortogonal. Por lo tanto, al igual que con la función escala, es posible obtener una representación de la diferencia que existe entre aproximar una señal con un nivel de resolución j y aproximar la misma señal con un nivel de resolución j + 1, mediante el producto interno de esta señal con un conjunto de funciones que generen el espacio W_j donde j será elegido de acuerdo al grado de aproximación que se desee.

Si definimos la función wavelet $\psi_k(t)$ como:

$$\psi_k(t) = \begin{cases} 1, & si\ k \le t < k + \frac{1}{2} \\ -1, & si\ k + \frac{1}{2} \le t < k + 1, \\ 0, & otro\ valor \end{cases}$$

donde $k \in Z$ y $\psi_k(t) \in L^2(R)$ corresponde a la misma función pero desplazada en el tiempo por una constante k. Definimos de la misma forma otra función $\psi_m(t)$ con $m > k$, entonces realizamos el producto entre ellas obteniendo:

$$< \psi_k(t), \psi_m(t) >= \int_k^{k+1} \psi_k(t).0dt + \int_n^{n+1} 0.\psi_m(t)dt = 0$$

con lo que demostramos la propiedad de ortogonalidad.

3. RECONOCIMIENTO AUTOMÁTICO DEL HABLA USANDO WAVELETS DE HAAR

En el presente trabajo se han tomado muestras de personas comprendidas entre los 18 y 65 años de edad, debido a que los niños por ejemplo presentan un timbre de voz muy fino y por ahora aun no hemos estudiado esos casos, dichas muestras comprenden las vocales y en algunos casos los números del 1 al 10.

A continuación explicaremos las fases que comprende el desarrollo del trabajo.

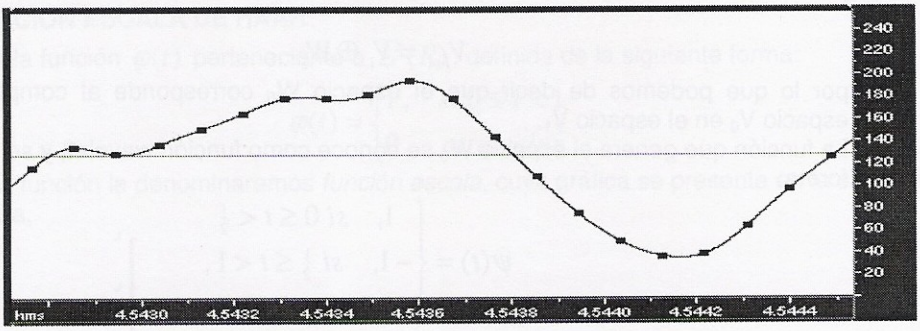
1. PROCESAMIENTO DE LA SEÑAL:

1.1. Captura de la señal:

La señal de voz básicamente está constituida por ondas de presión producidas por el aparato fonador humano. La manera obvia de capturar este tipo de señal se realiza mediante un micrófono, el cual se encargará de convertir la onda de presión sonora en una señal eléctrica.

A partir de la señal analógica obtenida se hace necesario convertir la señal a formato digital para poder procesarla en la computadora.

Por ejemplo tenemos una señal la cual queremos digitalizarla. Como podemos observar (ver figura), tomamos un valor en voltaje equivalente a la onda real el que será cuantificado, es decir transformado a digito, así su correspondiente valor binario es el que será almacenado en la memoria.



Todo este proceso se resume en:

- **Muestreo:**

Para realizar esto debemos saber que la señal vocal tiene componentes frecuenciales que pueden llegar a los 10 khz., sin embargo la mayor parte de los sonidos vocales tienen energía espectral significativa hasta los 5 khz.

El muestreo de una señal consiste en el paso de la señal de la forma analógica al ámbito discreto, es decir viene a ser el proceso de captura de puntos (muestras) que sean necesarios para poder representar la señal en una unidad de tiempo (segundo), para esto debemos de tener muy en cuenta el siguiente teorema de muestreo:

TEOREMA: Si $F_{\max} = B$ es la frecuencia más alta de la señal analógica $x(t)$ y esta se muestra a una velocidad $F_s > 2F_{\max} = 2B$, entonces $x(t)$ se puede recuperar totalmente a partir de sus muestras mediante la siguiente función de interpolación.

$$g(t) = \frac{sen(2\pi Bt)}{2\pi Bt}$$

- **Cuantificación:**

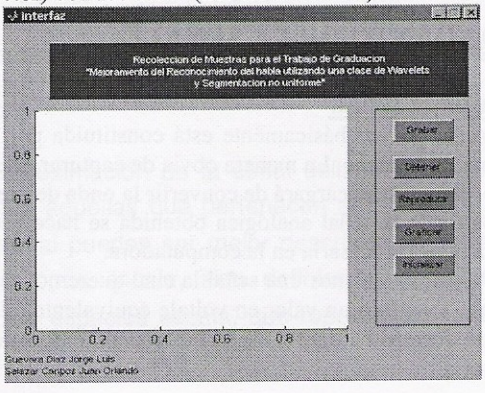
Otra consideración que se debe tener en cuenta es la cuantificación de la señal, la cual involucra la conversión de la amplitud de los valores muestreados a forma digital usando un número determinado de bits. El número de bits usados afectará la calidad de la voz muestreada y determinará la cantidad de información a almacenar.

La señal de voz exhibe un rango dinámico de unos 50 a 60 dB, por lo que resultaría suficiente una cuantificación de 8 a 9 bits para una buena calidad de voz.

El proceso de captura de la señal lo hemos efectuado con:

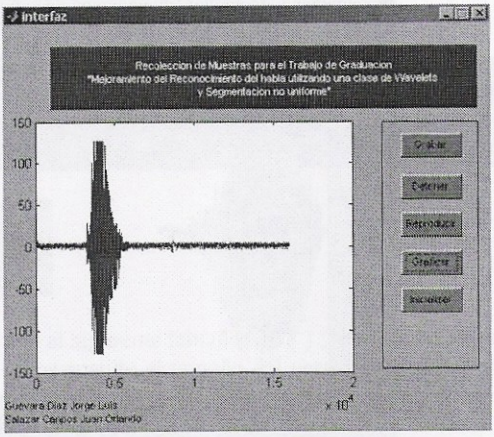
- ✓ Un micrófono M 750 H (V) Dynamic Stereo HeadPhone Microphone Combo, con control de volumen, un rango de frecuencias de 20 Hz a 20 KHz, una impedancia de 32 Ohms, una sensibilidad de –58 db.
- ✓ Una tarjeta de sonido SoundMax integrada a una placa Intel 850EMV2.
- ✓ La frecuencia fue de 8000 Hz.
- ✓ El tamaño para cada amplitud fue de 8 bits, con lo que podemos obtener 256 amplitudes diferentes.

Luego, diseñamos e implementamos la siguiente interfaz, la cual nos devolverá un archivo (conteniendo un vector) con los datos (números enteros) de la señal digitalizada:



Por ejemplo algunas muestras de vocales graficadas son:

Vocal "a:



1.2. Selección y Eliminación de Segmentos Inservibles:

Una vez que hemos obtenido un vector con los datos de la señal digitalizada tenemos que eliminar las secciones que no contienen información válida para nuestros fines tales como los valores del inicio y final captados ya sea por la demora en pronunciar una vocal o por la demora en detener la grabación.

Para esto obtenemos un umbral al cual le sumamos cierto valor, con el cual vamos a comparar las muestras; si existen valores que estén por debajo de éste se eliminarán.

El umbral lo obtenemos de la siguiente manera:

$$\text{valor} = \frac{1}{radio} \sum_{i}^{radio+i} |x_i|$$

donde:

radio = número de muestras que se están evaluando.

umbral = (*valorI* + *valorF*)/2 + *n*

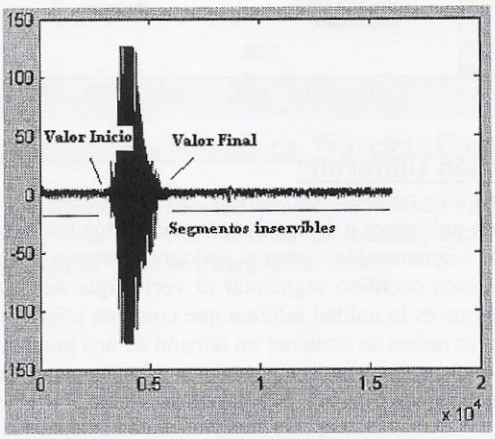
valorI = valor de las x primeras muestras.

valorF = valor de las x últimas muestras.

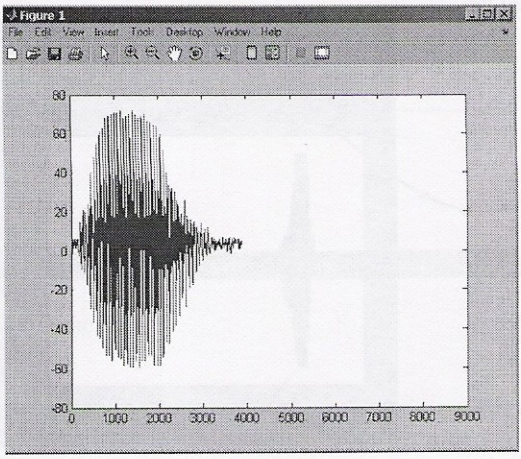
n = un valor fijo.

En el trabajo obtuvimos el umbral de aproximadamente 14, del promedio de los valores absolutos de las 10 primeras muestras y de las 10 últimas, pues con seguridad éstos valores representan momentos de silencio los cuales no son significativos, el valor que le sumamos (*n*) fue de 5, de esta forma recorremos el vector con un radio de 5 elementos, se empieza desde el inicio de la señal, si su promedio es menor que el umbral se descartan caso contrario se fija el ***inicio*** de la señal que deseamos. Para encontrar el final, de la misma manera empezamos a buscar los promedios de los valores absolutos de las muestras pero empezando por el final hasta encontrar el valor al que llamaremos ***final.***

Una vez que tenemos los valores de inicio y final recortamos el vector de muestras.



Por ejemplo la vocal i quedaría de la siguiente manera:



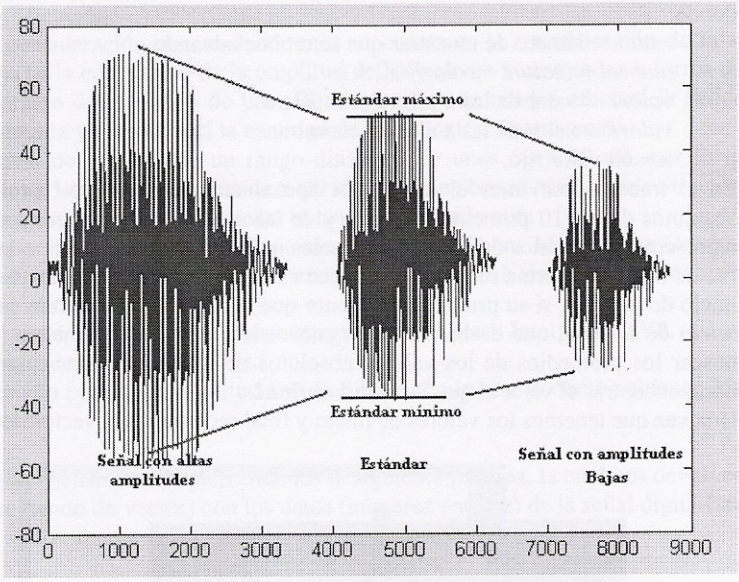
1.3. Normalización:

Luego de tener solamente la información necesaria, tenemos que normalizar para poder trabajar uniformemente, esta fase trata de llevar a un determinado rango y sus equivalentes todos los valores de las muestras. Así, realizamos la siguiente transformación:

[a,b] [c,d] [m, n]

$T(y)=Ay+B$, donde $A = (n-m)(d-c)$ y $B=m-(n-m)/(d-c)$

Luego de haber normalizado tenemos los valores de las señales entre un determinado rango para todos.



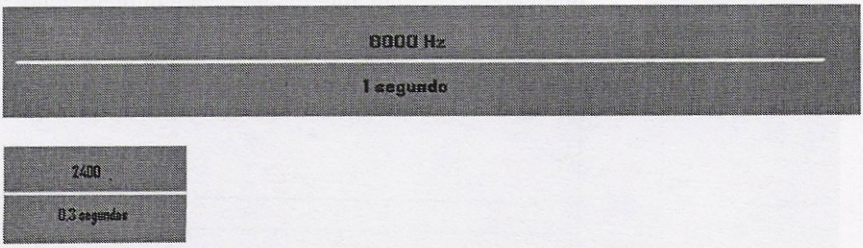
1.4. Segmentación Uniforme:

La segmentación consiste en dividir un vector en varias partes, pueden ser iguales (segmentación uniforme) o desiguales (segmentación no uniforme).

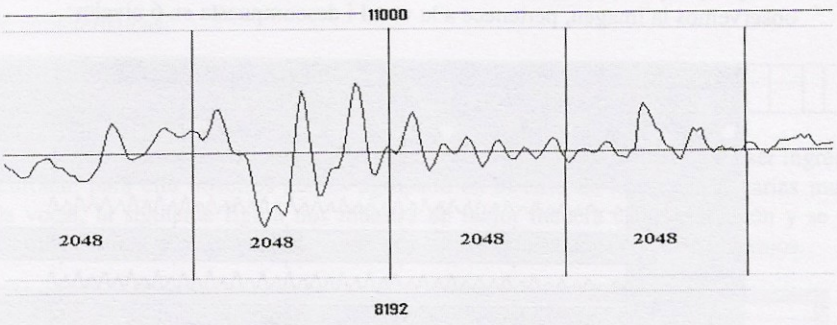
Mediante una segmentación nosotros podremos obtener una mejor caracterización de la señal, por lo que hemos decidido segmentar el vector que nos queda en partes equivalentes a 0.3 segundos, ya que es la unidad mínima que contiene información válida, teniendo cuidado que estos segmentos deben de contener un tamaño de una potencia de 2.

Si tenemos que en 1 segundo capturamos 8000 muestras en 0.3 segundos tendremos 2400 muestras, pero la potencia de 2 más próxima es $2^{11} = 2048$. En el caso de las vocales nos puede

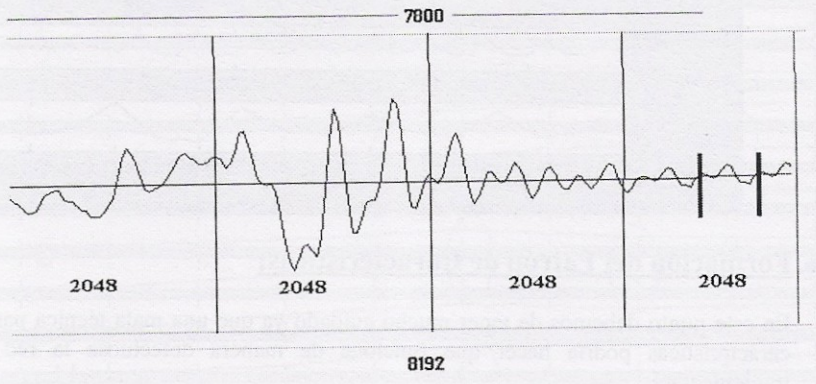
salir un solo segmento pero en palabras grandes tendremos más, pudiendo así caracterizarlas de mejor manera la señal



Obteniendo un vector con el siguiente tamaño: (2^{11}).(Número de Segmentos).
Acá se presentan dos casos cuando el tamaño del vector es superior al que deseamos obtener, calculamos el tamaño que deseamos y descartamos los últimos valores.

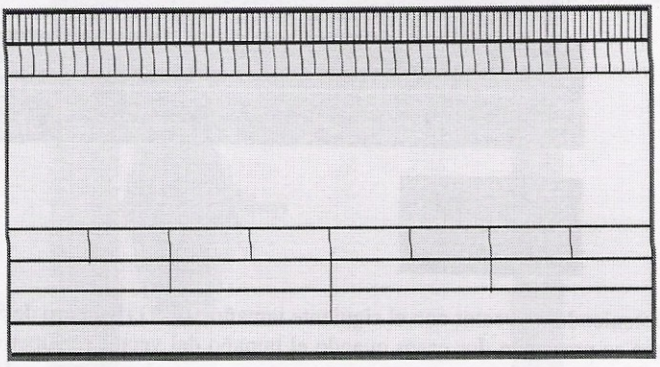


El otro caso es cuando la cantidad de valores del vector no superar el tamaño deseado, en este caso cogemos los últimos 5 valores y los duplicamos hasta completar el tamaño deseado.

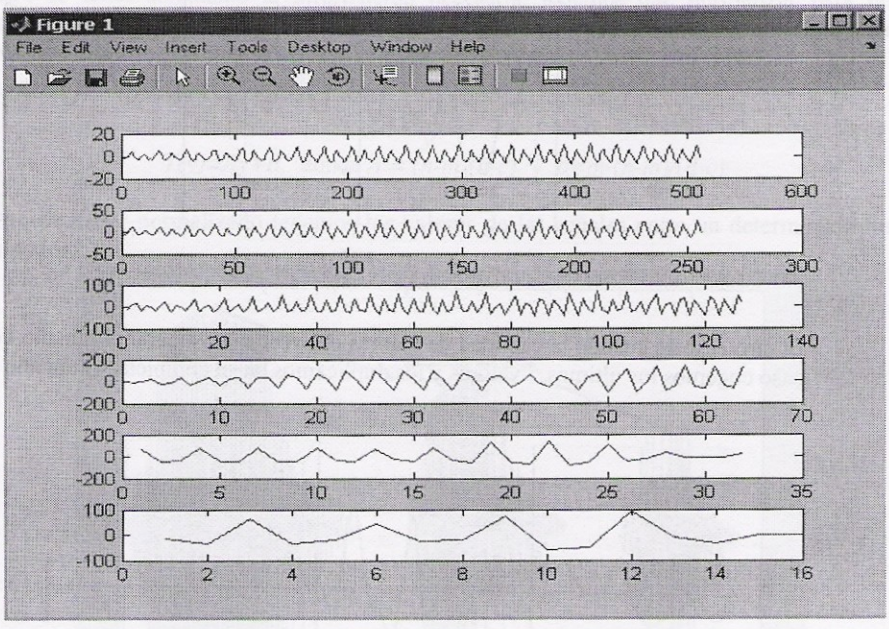


1.5. Aplicación de la Transformada Rápida de Wavelets Haar:

Una vez que tenemos los segmentos con los tamaños adecuado aplicamos la Transformada Rápida Wavelet de Haar a cada uno con un nivel de descomposición 6, ya que entre estas frecuencias se encuentra la mayor parte de información válida.



Por ejemplo para las vocales solamente tendremos un segmento, lo que no pasa con los números, observemos la imagen, pertenece a la vocal i descompuesta en 6 niveles:



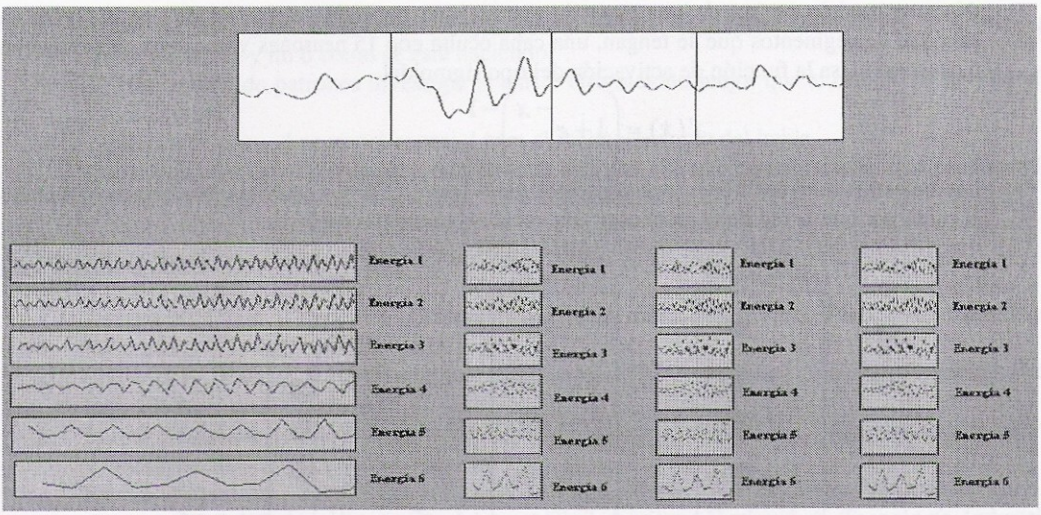
1.6. Formación del Patrón de Características:

En este punto debemos de tener mucho cuidado ya que una mala técnica para la extracción de características podría hacer que funcione de manera defectuosa la red neuronal que se implementará.

Hemos decidido que el patrón de características este formado por la energía de cada nivel, por lo que tendremos 6 energías por cada segmento. Luego para formar el patrón general de la señal concatenaremos estas energías (si hubiesen varias), de izquierda a derecha.

$$\text{Energía} = \frac{1}{k}\sum_{i=1}^{k}(x_i^{\,2})$$

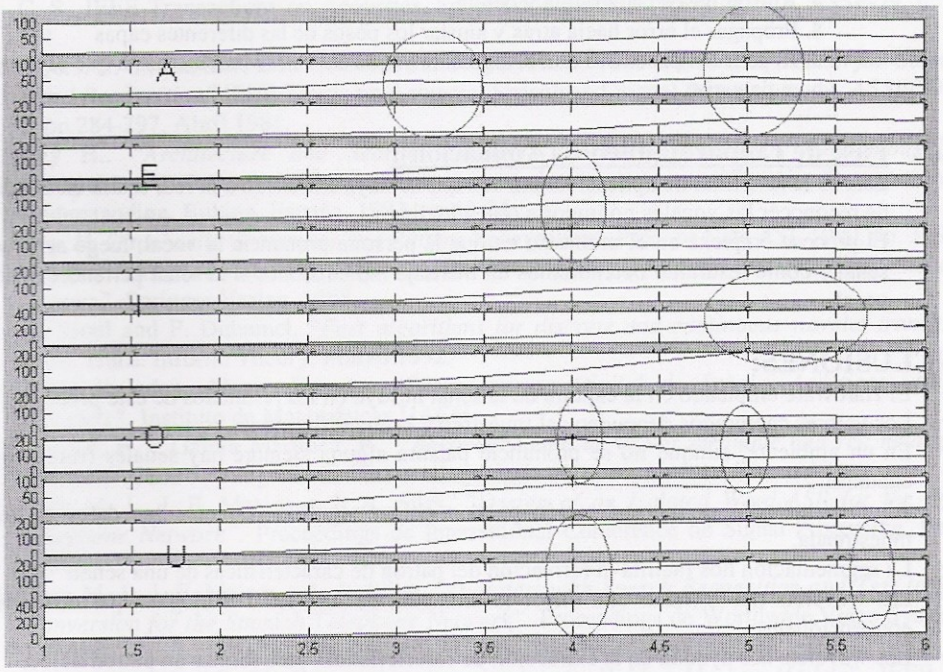
k = Cantidad de elementos del nivel de descomposición



Patrón de características:

Energía 1	Energía 2	Energía 3	...	...	Energía 6	...	...	...	...	Energía 1	Energía 2	Energía 3	...	...	Energía 6	...	...	...					

Una vez que se formado el patrón de características (vector), este queda listo para ser ingresado a la red neuronal, para ello nosotros hemos agrupado en el caso de las vocales, varias muestras para cada vocal, la siguiente figura nos muestra de mejor manera esta agrupación y se puede notar las características propias de cada vocal que han sido resaltadas con círculos rojos.



2. FASE DE RECONOCMIENTO:

Para esta fase se tomaron 5 muestras por persona y por vocal, es decir 5 muestras de la vocal a por una persona, 5 de la e y así hasta completar las vocales, un proceso similar se hizo para algunos números.

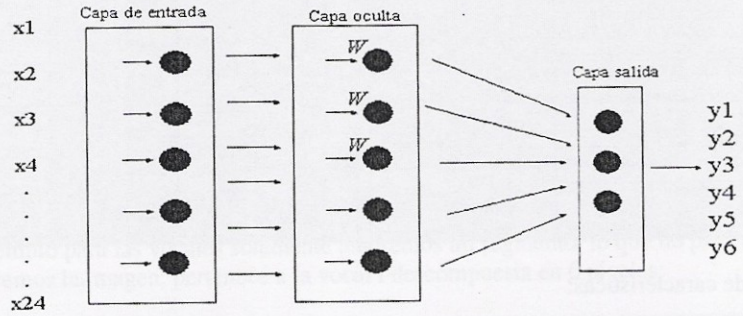
2.1. Fase de entrenamiento.

Para el presente trabajo hemos optado por una Red Neuronal Backpropagation por sus buenos resultados en las diferentes aplicaciones que se le ha dado.

Nuestro modelo consta de una capa de entrada que depende del patrón de entrada y de la cantidad de segmentos que se tengan, una capa oculta con 15 neuronas y una capa de salida con 6 neuronas. Usa la función de activación del tipo sigmoidal.

$$f(x) = \left(1 + e^{-x}\right)^{-1}$$

Las muestras tomadas tienen que pasar por todas las fases antes descritas. No hemos eliminado el ruido por que la red debe reconocer una vocal en ambientes ruidosos.



Algoritmo de backpropagation:

1. Estructura de la red
2. Escoger funciones de activación
3. Dar vectores de entrenamiento
4. Inicializar los pesos
5. Calcular la salida de la red "O"
6. Calcular los deltas para la capa salida y oculta
7. Ajustar pesos capa salida y oculta
8. Propagar el error hacia atrás y ajustar los pesos de las diferentes capas
9. Repetir los pasos 3-8 con el siguiente vector de entrenamiento hasta que el error sea bien pequeño

2.2. Fase de Emparejamiento (Aplicación):

Esta es la fase final de aplicación una vez que se haya entrenado a la red, estará en condiciones de reconocer las vocales no importa quien lo diga.

El proceso empieza en el momento en que la persona pronuncia la vocal, luego se procesa la señal y como resultado obtendremos un mensaje indicándonos si la señal pertenece o no a una vocal .

CONCLUSIONES:

- El Hardware empleado en la captura de la señal influye en los resultados de éste proceso.
- Los valores capturados dependen del muestreo y la cuantificación.
- En un ambiente, aunque no se pronuncie palabra alguna siempre hay señales (ruido) que son captadas.
- Algunos métodos de normalización pueden alterar significativamente los datos de la señal muestreada.
- La segmentación nos facilita la formación del patrón de características de una señal.
- Las wavelets nos brindan una mejor información de una señal analógica que otros métodos convencionales.
- Los wavelets nos ayudan en la extracción de características para formar un patrón único.
- Una red neuronal back-propagation puede trabajar adecuadamente pero el costo de aprendizaje es alto.

SUGERENCIAS:

De acuerdo a las experiencias adquiridas en la ejecución de la primera etapa de nuestro proyecto nos permitimos sugerir lo siguiente:

- Mejorar los métodos de Eliminación de Segmentos Inservibles y de Transformación de la señal a un Rango Determinado.
- Implementar una segmentación no uniforme que es uno de los fines del trabajo.

- Probar si se puede obtener un mejor patrón de características usando los 9 niveles de descomposición y no 6 como se está haciendo.
- En la selección de patrones investigar si existe otro método mejor que el cálculo de las energías por cada nivel.
- Implementar una red neuronal especial para el reconocimiento del habla.
- Probar con otro tipo de wavelets, que permita abaratar el costo computacional que producen las wavelets de Haar. Por ejemplo, utilizar las wavelets B-splines.

REFERENCIAS

1. Andrew K. Chan y Jaideva C Goswanmi, *"Fundamental of Wavelets. Theory, Algorithms and Aplications"*, Texas A&M Univertisy.
2. Bellman, R., Kalaba, R. *"Dynamic Programming and Modern Control Theory"*. Academic Press Inc., 1965.
3. Bernal Bermúdez Jesús, Bobadilla Sancho Jesús, Gómez Vilda Pedro., *"Reconocimiento de Voz y fonética acústica"*. Printed in México.
4. Fabián Acquaticci., Sergio Gwirc, Diego Brengi, *"Aplicación de Redes Neuronales para el Control de Calidad de Productos Lácteos Uht"*. Instituto Nacional de Tecnología Industrial, Centro de Investigación en Tecnología Electrónica e Informática, Bueno Aires.
5. Forney G. D. *"The Viterbi Algorithm"*. Proceedings of the IEEE, Vol. 61, Mar. 1973, pp 268.
6. L. G. Weiss, *"Wavelets and wideband correlation processing"*. IEEE Signal Process. Magazine, Enero 1994.
7. Linde Y., A. Buzo y R. M. Gray, *"An Algorithm for Vector Quantizer Design"*, IEEE Transactions on Communications, Vol. COM-28, pp 84.
8. M. J. Shensa, *"The discrete wavelet transform: Wedding the a trous and Mallat algorithms"*, IEEE Trans. Signal Process, Octubre 1992.
9. Myers y L. R. Rabiner, *"Connected Digit Recognition Using a Level-Building DTW Algorithm"*, C. S., IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-29, n° 3, pp 351.
10. Myers, C. S., Rabiner, L.R. *"A Level Building Dynamic Time Warping Algorithm for Connected Word Recognition"* IEEE Trans. on Acoustic, Speech and Signal Processing, vol. ASSP-29, num. 2, pp.284-297, Abril 1981.
11. Ney H., *"Architecture and Search Strategies for Large-Vocabulary Continuous-Speech Recognition"*, Proc. of Nato Advanced Study Institute on Speech Recognition and Understanding, Bubion, España, 1993, pp.59-84.
12. Ney, H. *"Stochastic Grammars and Pattern Recognition"*. Proc. of Nato ASI, 1990, pp. 319-344.
13. Niemann H., M. Lang & G. Sagerer, *"Recent advances In speech understanding and dialog systems"*. Springer Verlag, 1988.
14. O. Rioul and P. Duhamel, *"Fast algorithms for discrete and continuous wavelet transforms"*. IEEE Trans. Inform. Theory, Marzo 1992.
15. Pablo Faundez, Alvaro Fuentes, *"Procesamiento Digital de Señales Acústicas utilizando Wavelets"*. Instituto de Matemáticas UACH.
16. Priegue, R. y García Martínez, R., *"Reconocimiento de la voz mediante una red neuronal de kohonen"*.Centro de Ingeniería del Sw e Ingeniería del conocimiento, Bueno Aires.
17. Poza M. J., J. F. Mateos y J. A. Siles, *"Design of an Isolated Word ASR for the Spanish Telephone Network"*. Proceedings de International Conference on Signal Processing, Beiging, 1990.
18. Poza M. J., J. F. Mateos y J. A. Siles, *"Audiotext with Speech Recognition and Text to Speech Conversion for the Spanish Telephone Network"*. Proceedings de Worldwide Voice Systems'90, London.
19. Rabiner L. R. y B. H. Juang, *"An introduction to Hidden Markov Models"*. IEEE ASSP MAGAZINE, Enero 1986.
20. Rabiner L. R., *"A Tutorial on, LIMM and Selected Applications M Speech Recognition"*. Proceedings of the IEEE, Vol. 77, n° 2, pp 257.
21. Richard P. Loppmann, *"Neural Nets for Computing"*. ICASSP 1989. 7. Trends In speech recognition, W. A. LEA, Prentice Hall, 1980.
22. Sakoe, H., *"Two-Level DP-Matching- A Dynamic Programming - Based Pattern Matching Algorithm for Connected Word Recognition"*, IEEE Trans. on Acoustic, Speech and Signal Processing, vol. ASSP-27, num.6, pp.588-595, Diciembre 1979.