

Extracción de Características en el Procesamiento Digital de una Señal para el Mejoramiento del Reconocimiento Automático de Habla usando Wavelets

Jorge Luis Guevara Díaz

Universidad Nacional de Trujillo, Escuela de Informática,
Trujillo, Perú
jorge.jorjasso@gmail.com

and

Juan Orlando Salazar Campos

Universidad Nacional de Trujillo, Escuela de Informática,
Trujillo, Perú
josco_orlando@hotmail.com

Resumen

La presente investigación hace un estudio en la aplicación de diversas wavelets y wavelets packets en el procesamiento digital de la señal para el reconocimiento automático del habla, se hacen diversos experimentos para medir el desempeño entre diversos tipos de wavelets, también se mide el desempeño de éstas frente a la técnica de MFCC que es una de las técnicas más robustas que existen; finalmente se mide el desempeño de las wavelets frente a la transformada de Fourier.

Se han experimentado con la transformada wavelet de Haar, Daubechies y Coiflets cuyos algoritmos tienen una complejidad computacional de $O(n)$, como también con los wavelets packets cuya complejidad computacional es de $O(n \log n)$, obteniendo mejores resultados para los wavelets packets de Daubechies 6, haciendo un análisis aproximado de las frecuencias tal como lo hace la escala Mel. En una futura investigación se probará con las wavelets continuas.

Palabras clave: Muestreo, Wavelets, cepstrum, Mel, Fourier, Dynamic Time Warping, Procesamiento Digital de Señales.

1. Introducción

La presente investigación esta centrada en el Reconocimiento Automático del Habla (RAH) y mas concretamente en el uso de la transformada Wavelet en la parte del procesamiento digital de la señal y su influencia en la extracción de características.

Las Wavelets han sido utilizadas ampliamente en diversos campos como procesamiento de imágenes y señales, física, matemática incluso economía; los algoritmos rápidos de bajo costo computacional es lo que las hace tan atractivas. En el presente trabajo se diseñó un algoritmo basado en características perceptuales inspirados en los Coeficientes Cepstrales en Escala Mel (MFCC), mediante wavelets packets, también se experimentó con diversos tipos de wavelets discretos.

El resto de éste paper está organizado de la siguiente manera. En la sección 2 se muestra algunos trabajos previos realizados, la sección 3 describe el método MFCC para extracción de características. Se describen las wavelets, además, se trata la Transformada Wavelet frente a la Transformada de Fourier en la sección 4. En la sección 5 se muestra el modelo propuesto por este trabajo de investigación en el uso de la Transformada Wavelet en el Reconocimiento Automático del Habla. En la sección 6 se muestra el modelo propuesto en el uso de las wavelets Packets. En la sección 7 van los experimentos y resultados. En la sección 8 se encuentran las conclusiones del presente trabajo. Finalmente en la sección 9 se analizan futuras investigaciones que pueden surgir del presente trabajo de investigación.

2. Trabajos Previos

Existen diversos trabajos realizados, entre los más importantes detallaremos a continuación los siguientes. En [13], se describen diversos métodos de procesamiento digital de la señal, pero sin involucrar el uso de las wavelets En [1] describe a manera de introducción la posible aplicación de los Wavelets en el RAH, los trabajos de sarikaya, [16], [15], son unos de los mas importantes en lo que se refiere a la aplicación de Wavelets en la parte del procesamiento digital de la señal para el reconocimiento del Hablante, entre otros trabajos no menos importantes, tenemos [17], que hace uso de la transformada Wavelet en el reconocimiento de fonemas y el trabajo de, [10], que hace uso de la aplicación de Wavelets también en el reconocimiento del hablante. En la parte de emparejamiento de patrones uno de los trabajos mas importantes es el trabajo de Sakoe and Chiba, [14] que muestran un algoritmo optimizado para el reconocimiento de palabras haciendo uso de la programación dinámica. Una de las técnicas más usadas hoy en día son los Modelos Ocultos de Markov [3], existen también un sinnúmero de trabajos en Redes Neuronales, Máquinas de Soporte Vectorial aplicadas en la etapa de reconocimiento. Otro trabajo importante es el de [9] quien hace un descripción muy amplia de muchos algoritmos utilizados y una descripción muy detallada de los sistemas informáticos del lenguaje hablado, incluyendo temas como reconocimiento automático del habla, síntesis del habla y entendimiento del lenguaje natural.

Por el lado de las wavelets éstas han sido introducidos recientemente a principios de los años ochenta y han llegado a ser de gran interés en diversas disciplinas, pero sus raíces datan de mucho tiempo atrás. En la actualidad las wavelets han tomado una enorme popularidad. Sin embargo, sus raíces datan de 1873, cuando el trabajo de Karl Weierstrass describió una familia de funciones que son construidas por una superposición de copias escaladas de una función base dada. Trabajos importantes son los de Haar, el trabajo de Dennis Gabor quién describió una base no-ortogonal de lo que ahora se llaman wavelets con soporte no acotado, basado en funciones gaussianas trasladadas, También se deben citar diversos trabajos de Daubechies[4], Morlet y Grossmann [5], Mallat[12], Yves Meyer y muchos otros investigadores quienes han aportado mucho al desarrollo de este campo de estudio.

3. Procesamiento de la señal

La extracción de características de manera tradicional se hace utilizando algunas técnicas, las mas usadas son los coeficientes de predicción lineal Cepstrales LPC-cepstrum [2], coeficientes de predicción lineal perceptuales PLP [8], los coeficientes MFCC.

En el presente trabajo haremos nuestra comparación con la técnica de MFCC, pues esta técnica está basada en un modelo de percepción del habla, y la técnica a mostrar utiliza Wavelet y también esta basada en la percepción humana del habla, como también compararemos los resultados obtenidos mediante el uso de las Wavelets frente al uso de la Transformada de Fourier calculando las energías de diversos bandos de frecuencias.

3.1. Coeficientes MFCC

La señal de voz una vez capturada y digitalizada, es segmentada en frames muy pequeños según algunos experimentos neurofisiológicos en la codificación del habla sugieren que esta segmentación debe ocurrir alrededor de 10 mS, pues alrededor de 20 mS el oído empieza a oír una cierta distorsión,

$$\chi^m[n] = \chi[n - mF]\omega[n]. \quad (1)$$

donde "m" es el frame a procesarse F es el espaciamiento entre frames y $\omega[n]$, es una ventana de longitud N, esta ventana puede ser rectangular o del tipo hamming o hanning, una ventana rectangular produce grandes ondas laterales, y da el máximo ajuste la ventana hamming no tiene tanta precisión frecuencial peso provoca efectos mucho menores

$$\omega[n] = 0,54 - 0,46 \cos \frac{2\Pi n}{N} \text{Hamming} \quad (2)$$

luego una trasformada discreta de fourier es aplicada a cada frame para obtener los componentes de frecuencia de la señal, generalmente se usa un algoritmo que implemente de manera eficiente la Trasformada Rápida de Fourier

$$X_m(e^{jw}) = \sum \chi_m[n]e^{-jw} = \sum \omega[m - n]\chi[n]e^{-jw} \quad (3)$$

una vez llevado al dominio de la frecuencia cada frame es pasado a la escala Mel , la escala Mel es una escala construida en base a la percepción humana del habla, y sus valores han sido dados tras experimentos fisiológicos de muchos investigadores quienes han construido escalas de frecuencias basadas en la respuesta natural del sistema de audición humano, pues el complejo sistema auditivo trata las frecuencias de entrada en una manera no lineal , sino mas bien de una manera casi logarítmica, la escala en frecuencia Mel esta dada por:

$$\beta(f) = 1125 \ln\left(1 + \frac{f}{700}\right) \quad (4)$$

y para traspasar las frecuencias obtenidas por medio del Algoritmo Rápido de la Trasformada Discreta de Fourier se procede hacer un ventanamiento llamado "bins" de la siguiente manera:

$$H = \begin{cases} 0 & \text{si } k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{k-f(m-1)}{f(m+1)-f(m)} & f(m) \leq k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad (5)$$

donde

$$f(m) = \frac{N}{F_s} \beta^{-1}\left(\beta(f_1) + m \frac{\beta(f_h) - \beta(f_1)}{M+1}\right) \quad (6)$$

una vez obtenidas las frecuencias bineadas en escala Mel , se procede a calcular el cepstrum cuya justificación viene del hecho que se tiene el modelo de fuente de voz dice que el timbre y tonalidad de la señal de voz estan en convolución con el filtro del tracto vocal, las partes altas en el tiempo del cepstrum viene a ser el timbre o inflexiones de voz propias de cada persona al hablar , y las partes bajas corresponde a la información proveniente del tracto vocal.

La salida del filtro para cada frame esta dado como sigue:

$$S(m) = \ln\left(\sum |X(k)|H_m(k)\right), 0 < m < M \quad (7)$$

donde X(k) es la salida de la Trasformada Discreta de Fourier.

Finalmente una Trasformada Discreta del Coseno-II es Calculada, por concentrar su energía alrededor de las frecuencias mas bajas.

$$c(m) = \sum S(m) \cos\left(\pi n \left(\frac{m + \frac{1}{2}}{M}\right)\right) \quad (8)$$

Estos coeficientes resultantes son los vectores de características resultantes de la parte del procesamiento digital de la señal , estos coeficientes toman el nombre de Coeficientes Cepstrales en Escala Mel MFCC, nombre obtenido por el procedimiento antes mencionado.

4. Transformada Wavelet

La Transformada Wavelet es una herramienta matemática que corta los datos, funciones o operadores en diferentes componentes de frecuencia [4] y estudia cada componente a una resolución ubicada a esa escala.

4.1. Transformada Wavelet Continua

Restringiendo a una dimensión y estableciendo los parámetros de dilatación y traslación a y b que varían continuamente sobre \mathfrak{R} con la restricción de $a \neq 0$, la transformada wavelet continua de una función f está dada por:

$$\begin{aligned}(T^{wav} f)(a, b) &= \int \delta x f(x) |a|^{-\frac{1}{2}} \psi \\ (T^{wav} f)(a, b) &= \langle f, \psi^{a,b} \rangle \left(\frac{x-b}{a} \right)\end{aligned}\tag{9}$$

la familia de wavelets se puede construir dilatando y trasladando

$$\psi^{a,b}(x) = |a|^{-\frac{1}{2}} \psi\left(\frac{x-b}{a}\right)\tag{10}$$

la función f puede ser recuperada de su transformada wavelet como sigue:

$$f = C_{\psi}^{-1} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\delta a \delta b}{a^2} \psi(T^{wav} f)(a, b) \psi^{a,b}\tag{11}$$

4.2. Comparación de la Transformada de Fourier con la Transformada Wavelet

A continuación se muestran las diferencias y similitudes del análisis Wavelet frente al análisis con Fourier.

$$T^{win}(w, t) = \int \delta s f(s) g(s-t) e^{-i\omega s}\tag{12}$$

Transformada Ventaneada de Fourier.

$$CWT(a, b) = \frac{1}{\sqrt{a}} \int f(x) \psi\left(\frac{x-b}{a}\right) \delta x\tag{13}$$

Transformada Wavelet.

La transformada Wavelet provee una descripción similar Tiempo - Frecuencia. Una similitud entre la Transformada Wavelet y la Transformada Ventaneada de Fourier sería en que ambas toman el producto interno de la función f con una familia de funciones $g(s-t)e^{-i\omega s}$ y con $\psi\left(\frac{x-b}{a}\right)$, donde las funciones $\psi^{a,b}$ son llamadas Wavelets.

La diferencia entre la Transformada Wavelet y la Transformada Ventaneada de Fourier está dada en el hecho de la manera en como analizan las funciones [4], la función g analiza utilizando la misma forma para las frecuencias altas y las frecuencias bajas, la función ψ analiza las altas frecuencias con pequeñas formas y las bajas frecuencias con tamaño mucho mayores.

Para más detalle sobre Wavelets ver [4]

5. Extracción de características utilizando Wavelets

De la misma manera que los Coeficientes Cepstrales en Escala Mel hacen un filtrado de los componentes de frecuencia, y luego una decorrelación del cepstrum mediante una transformada del coseno, la extracción de características en el procesamiento digital de la señal mediante wavelets se procedió a hacer algo similar.

Primero se hizo una descomposición wavelet de cada frame hasta j niveles que corresponderán a un análisis multiresolución, donde la señal es proyectada a cada nivel de resolución obteniendo al final j espacios W correspondientes diversos rangos de frecuencia y un espacio V correspondiente al nivel mas bajo de frecuencia de la señal, para esto se utilizó las wavelets discretas:

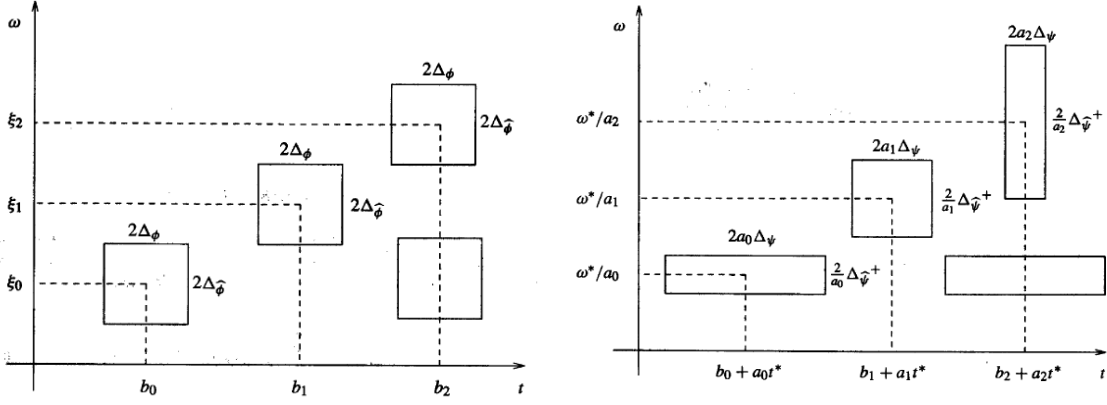


Figura 1: Ventana Tiempo-Frecuencia para la Transformada de Fourier y para la Transformada Wavelet. Fuente: [11]

$$\begin{aligned}\psi^{m,n}(x) &= a_0^{-\frac{m}{2}} \psi(a_0^{-m}(x - nb_0 a_0^m)) \\ \psi^{m,n}(x) &= a_0^{-\frac{m}{2}} \psi(a_0^{-m}x - nb_0)\end{aligned}\quad (14)$$

en particular para este trabajo se eligió $a_0 = 2$ y $b_0 = 1$ entonces:

$$\psi^{m,n}(x) = 2^{-\frac{m}{2}} \psi(2^{-m}x - n) \quad (15)$$

En el espacio V y en cada espacio W , tendremos coeficientes escala ϕ y coeficientes wavelet ψ correspondientes, en este caso solo se utilizarán los coeficientes $C_{j,k}$ y $d_{j,k}$ para determinar información importante en determinado espacio de tiempo en la señal de habla, coeficientes con altos valores nos indicarán la presencia de información importante.

Para proceder a obtener la extracción de características mediante wavelets, se hizo un ventaneamiento de la señal utilizando una ventana Hamming de $16mS$ para ciertos experimentos y también ventanas Hamming de tamaño $32mS$, también se utilizó un tamaño de paso F de $10mS$

Una vez obtenidos los valores del ventaneamiento se procedió a hacer una descomposición wavelet de cada segmento obtenido del ventaneamiento, hasta un nivel $j=7$ para las wavelets de haar $j=6$ para las wavelets de Daubechies 4, $j=5$ para las wavelets de Daubechies 6 y Coiflets 6. El proceso de descomposición utilizó el algoritmo de banco de filtros

5.1. Algoritmo de banco de filtros

Desde que $\phi \in V_0 \subset V_{-1}$ y que $\phi_{-1,n}$ son bases ortonormales en V_{-1} , tenemos:

$$\phi(x) = \sqrt{2} \sum_n h_n \phi(2x - n) \quad \text{con} \quad h_n = \langle \phi, \phi_{-1,n} \rangle \quad \text{y} \quad \sum_{n \in \mathbb{Z}} |h_n|^2 = 1 \quad (16)$$

esto indica que la función escala en cierta nivel m puede ser expresada en términos de funciones escalas trasladadas en la siguiente escala mas pequeña.

Similarmente podemos expresar la función wavelet en cierto nivel en términos de funciones escaladas y trasladadas en la siguiente escala mas pequeña.

$$\psi(x) = \sqrt{2} \sum_n g_n \phi(2x - n) \quad \text{con} \quad g_n = \langle \psi, \phi_{-1,n} \rangle = (-1)^n h_{-n+1} \quad (17)$$

lo cual implica los dos importantes resultados:

Si una función f puede ser representada por funciones escala en el nivel m

$$f(t) = \sum_n C_n \phi_{-1,n} \quad \text{con} \quad C_n = \langle f, \phi_{-1,n} \rangle \quad (18)$$

y en términos de wavelets

$$f(t) = \sum_n d_n \psi_{-1,n} \quad \text{con} \quad d_n = \langle f, \psi_{-1,n} \rangle \quad (19)$$

finalmente se puede establecer lo siguiente:

$$C_m = \sum_l h(l-2n)C_{m-1}(l) \quad , \quad d_m = \sum_l g(l-2n)d_{m-1}(l) \quad (20)$$

Estas dos últimas ecuaciones nos dicen que los coeficientes wavelets y escala en cierto nivel m pueden ser encontrados de manera iterativa, por ejemplo empezando de $\langle f, \phi_{0,n} \rangle$ podremos calcular $\langle f, \psi_{1,n} \rangle$ y $\langle f, \phi_{1,n} \rangle$ y así sucesivamente.

Los valores para h y g actúan como filtros de paso baja y filtros de paso alto respectivamente y llamaremos a h el filtro escala y a g el filtro wavelet.

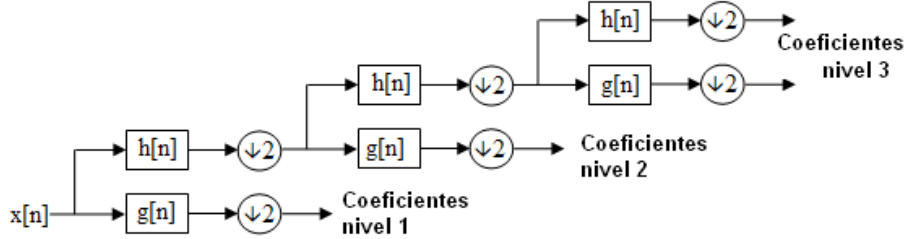


Figura 2: Implementación del banco de filtros iterativo.

5.2. Complejidad Computacional del algoritmo de banco de filtros

La complejidad computacional de este algoritmo es la siguiente:

$$T[n] = \begin{cases} T(\frac{n}{2}) + Cn & \text{si } 2^n \geq 2 \\ 0 & n = 1 \end{cases} \quad (21)$$

resolviendo la ecuación de recurrencia se tiene que este algoritmo para la transformada wavelet con banco de filtros tiene una complejidad de $O(n)$, y para el caso de las wavelets packet este algoritmo tiene una complejidad de $O(n \log n)$, mas detalle en [7]

6. Modelo propuesto para la Extracción de Características basadas en Wavelets Packets Perceptuales

En este modelo que proponemos hacemos uso de los wavelet Packet, pero no nos interesa toda la descomposición del wavelet packet, solo deseamos obtener los coeficientes que están en determinado nivel de resolución, cuyos componentes de frecuencia son aproximadamente iguales a la escala Mel.

Para este caso utilizamos un tamaño de frame igual a $24mS$ y un tamaño de paso igual que el caso anterior de $16mS$.

6.1. Obtención de las Características

Una vez computados los valores de los coeficientes para cada nivel de descomposición se procedió a calcular la energía de cada nivel con la finalidad de comprender el aporte del nivel en el tiempo, la energía de cada nivel se calculó mediante la siguiente expresión

$$E_i = \frac{\sum_{j=1}^N (W_i^p f(j))^2}{N_i} \quad (22)$$

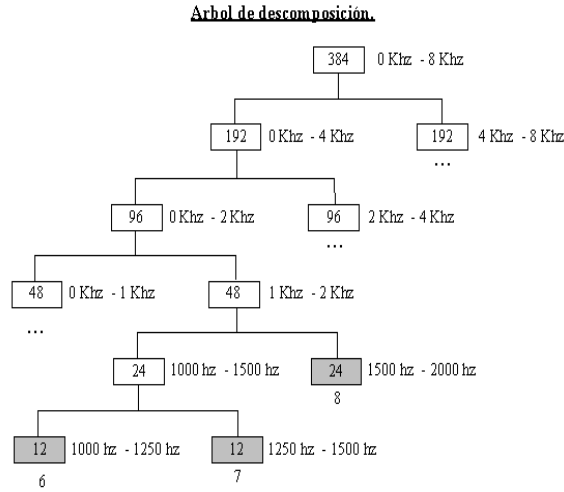


Figura 3: Arbol de descomposición espacios de resolución 6,7 y 8 para las wavelet packet

Y finalmente se aplica una transformada discreta del coseno al logaritmo de las energías para cada bando de frecuencia, estos valores son los que constituirán los vectores de características de la señal

$$F(i) = \sum_{i=1}^N \log E_n \cos\left(\frac{i\left(\frac{n-1}{2}\right)}{N}\right) \quad (23)$$

La complejidad de todo el algoritmo para las wavelets discretas es $O(n)$ y para el caso de las Wavelets Packets es $O(n \log n)$.

7. Experimentos y Resultados

Se realizaron varios experimentos con varios tipos de wavelets que a continuación se detallan. Wavelet de Haar

$$h(n) = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right] \quad (24)$$

Wavelets de Daubechies 4

$$h(n) = \left[\frac{1 + \sqrt{3}}{4\sqrt{2}}, \frac{3 + 3\sqrt{3}}{4\sqrt{2}}, \frac{3 - 3\sqrt{3}}{4\sqrt{2}}, \frac{1 - \sqrt{3}}{4\sqrt{2}} \right] \quad (25)$$

Wavelets de Daubechies 6

$$h(n) = [0,3326, 0,8068, 0,4598, -0,1350, -0,0854, 0,0352] \quad (26)$$

Wavelet Coiflet 6

$$h(n) = \left[\frac{1 - \sqrt{7}}{16\sqrt{2}}, \frac{5 + \sqrt{7}}{16\sqrt{2}}, \frac{14 + 2\sqrt{7}}{16\sqrt{2}}, \frac{14 - 2\sqrt{7}}{16\sqrt{2}}, \frac{1 - \sqrt{7}}{16\sqrt{2}}, \frac{-3 + \sqrt{7}}{16\sqrt{2}} \right] \quad (27)$$

Siendo los filtros g correspondientes: $g_n = (-1)^n h_{-n+1}$

El tipo de pruebas que se realizaron fueron independiente del hablante, tomando como plantillas para el reconecedor 125 palabras de entrenamiento pertenecientes a 5 personas. (25 palabras por persona). Para los casos de test se utilizaron 800 palabras (15 palabras por 60 personas) por método implementado. haciendo un total de 8100 pruebas realizadas; las palabras utilizadas fueron: abajo, abrir, adiós, arriba, caminar, cancelar, cerrar, coger, cuatro, derecha, detener, dos, eliminar, error, guardar, hola, iniciar, izquierda, pez, salir, terminar, test, tres, tres(repetición), uno. Para las pruebas se utilizaron 15 palabras por persona, las

palabras fueron: abrir, cerrar, coger, cuatro, dos, eliminar, error, hola, izquierda, pez, salir, terminar, tres, tres (repetición), uno. Para realizar las pruebas desarrollamos el software LORITO [6] en el que se implementaron todos los algoritmos estudiados y gracias al cual pudimos efectuar nuestras tomas de muestras y el respectivo procesamiento de las mismas.

El método estadístico que nos permitió analizar nuestros resultado fue la Prueba JI Cuadrado de MC Nemar-Datos Correlacionados

<i>Método</i>	<i>Tasa aceptación</i>	<i>Error</i>
Coefficientes Cepstrales en Escala Mel	85.32%	14.68%
Wavelet Haar	34.47%	65.53%
Wavelet Daubechies 4	51.79%	48.21%
Wavelet Daubechies 6	61.32 %	38.68%
Wavelet Coiflets 6	55.46 %	44.54 %
Wavelet Packet Perceptuales Walsh	55.79 %	44.21 %
Wavelet Packet Perceptuales Daubechies 4	69.14 %	30.86%
Wavelet Packet Perceptuales Daubechies 6	74.43 %	25.57%
Wavelet Packet Perceptuales Daubechies 4 (22)	71.21 %	28.79%

Figura 4: Datos obtenidos utilizando la técnica de DTW como reconocedor, con distancia Chebyshev y con Slope Constrain P=1. Se observa la mejor performance en las Wavelet Packet Perceptuales Daubechies 6, y la mas pobre en las Wavelet Haar. Datos obtenidos con software LORITO [6]

	<i>MFCC</i>	<i>W. Haar</i>	<i>W. Db4</i>	<i>W. Db6</i>	<i>W. Coif6</i>	<i>WP Walsh</i>	<i>WP Db4</i>	<i>WP Db 6</i>	<i>WP Perc.</i>
<i>Arriba</i>	86%	17%	41%	48%	34%	31%	66%	72%	62
<i>Cerrar</i>	100%	45%	79%	86%	76%	97%	100%	97%	97%
<i>Coger</i>	90%	17%	17%	31%	17%	28%	52%	69%	62%
<i>Cuatro</i>	97%	48%	28%	31%	24%	62%	76%	83%	83%
<i>Dos</i>	86%	28%	38%	48%	48%	59%	69%	69%	72%
<i>Eliminar</i>	72%	24%	34%	52%	48%	24%	48%	76%	45%
<i>Error</i>	97%	24%	66%	93%	83%	93%	97%	97%	93%
<i>Hola</i>	83%	38%	48%	72%	55%	38%	52%	59%	52%
<i>Izquierda</i>	86%	31%	48%	62%	59%	24%	59%	72%	69%
<i>Pez</i>	62%	24%	52%	41%	66%	62%	59%	59%	55%
<i>Salir</i>	86%	66%	79%	79%	72%	76%	76%	79%	76%
<i>Terminar</i>	62%	24%	24%	31%	28%	41%	55%	55%	55%
<i>Tres</i>	83%	24%	52%	59%	45%	41%	66%	62%	66%
<i>Tres</i>	59%	24%	45%	59%	41%	48%	55%	59%	62%
<i>Uno</i>	76%	31%	79%	83%	86%	76%	72%	66%	59%

Figura 5: Tasa de reconocimiento de las palabras por método. Datos obtenidos con software LORITO [6]

8. Conclusiones y futura investigación

El mejoramiento del espectro se da gracias al análisis tiempo frecuencia de las wavelets, con las wavelets podemos saber aproximadamente el aporte de las frecuencias por nivel de tiempo en las señales de habla, pues analiza con pequeñas wavelets componentes de alta frecuencia en la señal y con wavelets mas grandes componentes de baja frecuencia presentes en la señal, esto se traduce en la tasa de reconocimientos que proporcionan los wavelets. Una extracción de características usando solamente la Transformada de Fourier no muestra buenos resultados, pues ésta hace un análisis tiempo frecuencia de la señal con ventanas del mismo tamaño para todos los niveles de frecuencia sacrificando resolución de tiempo o frecuencia según se empequeñezca o agrande la ventana de análisis empobreciendo de esta manera la resolución espectral

o temporal. La efectividad de reconocimiento de habla del MFCC radica en que es un buen modelo de representación de la producción y percepción de habla, el cual es obtenido gracias a la agrupación de diversos métodos como el cepstrum, la escala Mel, Transformada de Fourier, etc, el mismo hecho de agrupar varios métodos para obtener mayor efectividad en el reconocimiento, eleva el tiempo de ejecución del algoritmo, llegando a obtener una complejidad tiempo de $O(n \log n)$, por frame de tamaño n , y un tiempo de ejecución mucho mayor. Las wavelets pueden ser utilizados alternativamente, para el procesamiento digital de la señal de habla. aprovechando el análisis que permiten y su rápida implementación computacional. Las wavelets discretas implementadas con el algoritmo de banco de filtros, para la extracción de características brindan una tasa de reconocimiento bajo, por lo cual se recurren a las wavelets packets que tienen mayor resolución en frecuencias, a las cuales hemos adaptado de tal manera de que en los espacios de resolución las wavelets tengan una frecuencia aproximada a la de la escala Mel. La complejidad computacional de los algoritmos de extracción de características usando las wavelets y las wavelets packets es de $O(n)$ y de $O(n \log n)$ respectivamente. La complejidad computacional de los MFCC y de las wavelets packet es la misma $O(n \log n)$, pero el menor tiempo de ejecución corresponde a las wavelets packet, debido al menor número de procedimientos utilizados. La ventaja de utilizar wavelets radica, en la variedad de funciones wavelet que se puede escoger, además de sus formas discretas y continuas.

Se pueden utilizar otra gamma de wavelets teniendo en consideración los resultados obtenidos en el presente trabajo, como también el uso de wavelets continuos. Se puede también segmentar la señal de una manera no uniforme, con tamaños variables de frames obtenidos gracias a las variaciones locales en la señal. La tasa de reconocimiento puede variar según el reconocedor que se use, en este caso hemos hecho uso de un reconocedor basado en un algoritmo optimizado de programación dinámica DTW, pero puede utilizarse otros reconocedores basados en Redes Neuronales, uso de probabilidades, Modelos Ocultos de Markov, Redes Bayesianas, Máquinas de Soporte Vectorial, etc. La tasa de reconocimiento que se ha mostrado en los métodos implementados se ha obtenido construyendo un reconocedor del tipo independiente del hablante, esta tasa puede variar muy notablemente en un sistema dependiente del hablante, donde es de esperarse tasas de reconocimientos mucho mayores.

Referencias

- [1] ABOUFADEL, E. A wavelets approach to voice recognition. *Grand Valley State University* (2001).
- [2] ATAL, AND SCHROEDER. Predictive coding of speech signals. *Report of the 6th Int. Congress on Acoustics, Tokio, Japan* (1968).
- [3] BAUN, L., AND EAGON, J. Perceptual linear predictive analysis of speech. *RBulletin of American Mathematical Society, 1967, 73, pp. 360-363* (1968).
- [4] DAUBECHIES, I. *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.
- [5] GROSSMANN, A., AND MORLET, J. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis* 15, 4 (1984), 723–736.
- [6] GUEVARA, J. L. *Lorito, speech recognition software*, v 1.0 ed. Universidad Nacional de Trujillo, Trujillo, Enero 2007.
- [7] GUEVARA, J. L., AND SALAZAR, J. O. *Extracción de Características en el Procesamiento Digital de una Señal para el Mejoramiento del Reconocimiento Automático de Habla usando Wavelets*. Tesis, Trujillo, Enero 2007.
- [8] HERMANSKY, H. A an inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. *J. Acoust. Soc. Am.* (2005).
- [9] HUANG, X., ACERO, A., AND HON, H.-W. *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*. Prentice Hall PTR, April 2001.
- [10] M. SIAFARIKAS, TODOR GANCHEV, N. F. Objective wavelet packet features for speaker verification, 2000.

- [11] MALLAT. Multiresolution approximation and wavelets. *Trans. Amer. Math. Soc.*, 315, pp. 69-68. (1989).
- [12] MALLAT, S. G. A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 11, 7 (1989), 674-693.
- [13] MANTHA, V., R. Y. A. J. Implementation and analysis of speech recognition frontends.
- [14] SAKOE, H., AND CHIBA, S. Dynamic programming algorithm optimization for spoken word recognition. 159-165.
- [15] SARIKAYA, R., AND HANSEN, J. High resolution speech features parametrization for monophone based stressed speech recognition packet transform features with application to speaker identification., 2000.
- [16] SARIKAYA, R., PELLOM, B., AND HANSEN, J. Wavelet packet transform features with application to speaker identification, 1998.
- [17] TAN, B. T., FU, M., SPRAY, A., AND DERMODY, P. The use of wavelet transforms in phoneme recognition. In *Proc. ICSLP '96* (Philadelphia, PA, 1996), vol. 4, pp. 2431-2434.

Apéndice

Imágenes de LORITO [6] Software desarrollado para realizar los experimentos

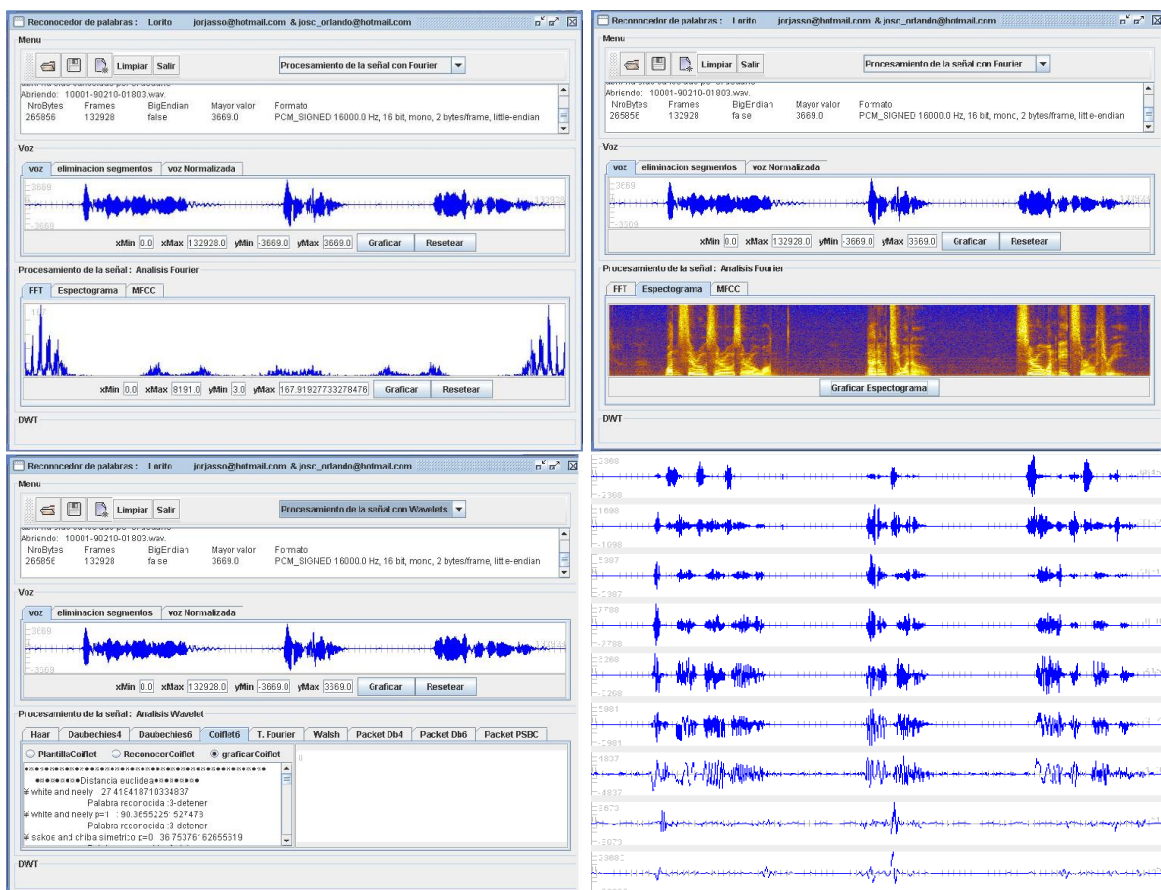


Figura 6: LORITO [6] en accion, Trasformada de Fourier de una palabra: parte superior izquierda, Espectrograma: parte superior derecha, Reconociendo una palabra: parte inferior izquierda, Análisis mediante Wavelets: parte inferior derecha