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

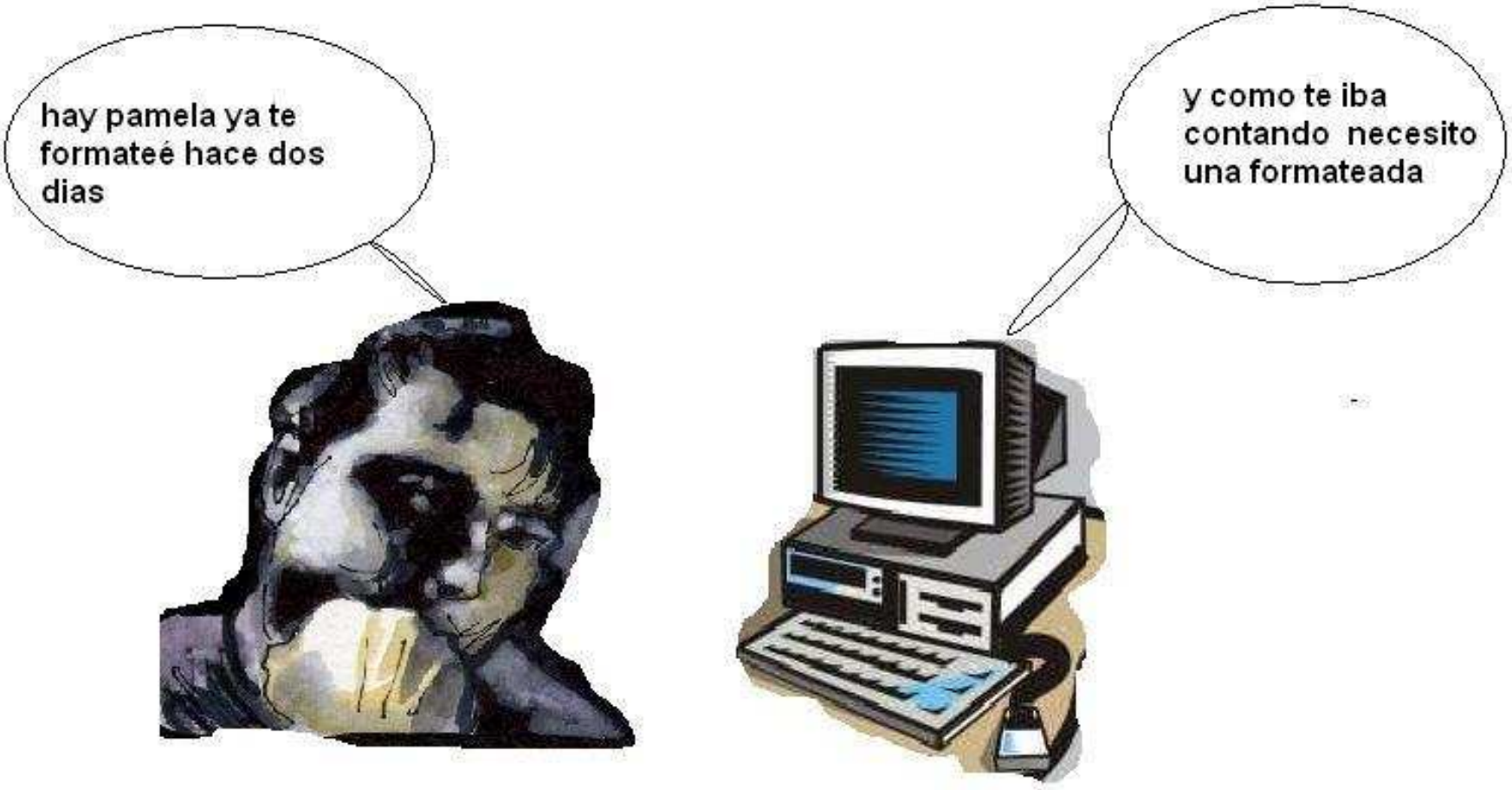
Juan Orlando Salazar Campos
Universidad Nacional de Trujillo, Escuela de Informática,
Trujillo, Perú
josc_orlando@hotmail.com

Que veremos?

1. Introducción
2. Trabajos Previos
3. Procesamiento de la Señal
4. Coeficientes MFCC
5. Trasformada Wavelet
6. Extracción de características usando wavelets
7. Experimentos y Resultados
8. Conclusiones

1. Introducción

Podriamos conversar con las maquinas como lo hacemos con los humanos?



1. Introducción

- **Speech Recognition**

¿Cómo hacer que las computadoras puedan convertir a texto la palabra hablada?

Problemas:

algoritmos de bajo costo computacional

extracción de “buenas” características

correcta clasificación

1. Introducción

- **Extracción de características**

Complejidad computacional

Cual es la mejor representación de características?

Reducción de la dimensionalidad

conjunto mas pequeño que contenga la información mas esencial presente en los atributos originales

$$\mathbf{x} = (x_1, \ldots, x_n) \longmapsto \boldsymbol{\phi}(\mathbf{x}) = (\phi_1(\mathbf{x}), \ldots, \phi_d(\mathbf{x})), \quad d < n,$$

2. Trabajos Previos

2. Trabajos Previos

- **Diversas Técnicas**

Bandos de Energias de la Trasformada de Fourier

LPC Coeficientes de Prediccion Lineal [Atal, and Schroeder]

LPC-Cepstrum [Atal, and Schroeder] [Bogert and Tukey]

PLP Coeficientes de Prediccíon Lineal Perceptuales

MFCC Coeficientes Cepstrales en Frecuencia Mel [Davis and Mermelstein]

Propuestas Basada en Wavelets

3. Procesamiento Digital de la Señal de Habla

3. Procesamiento Digital de la Señal de Habla

- Diversos algoritmos para procesar la señal digital de habla en una computadora

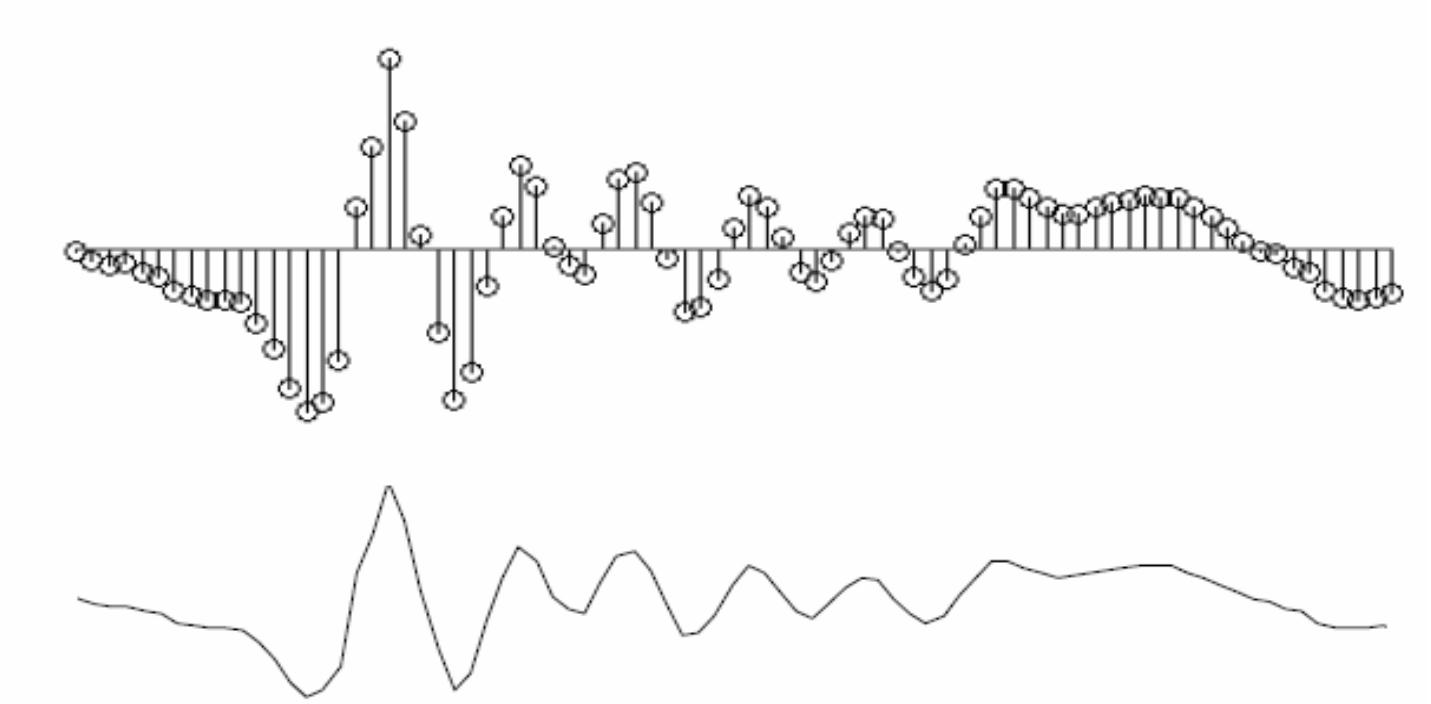
- Ejemplo:

  Eliminacion de ruido, analisis de frecuencias, etc

3. Procesamiento Digital de la Señal de Habla

Capturar la señal analógica y digitalizarla para poder usarla en la computadora

$$x[n] = x_0(nT).$$



3. Procesamiento Digital de la Señal de Habla

3.1 Transformada de Fourier

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t}$$

Donde :

$$e^{j\phi} = \cos\phi + j\sin\phi$$

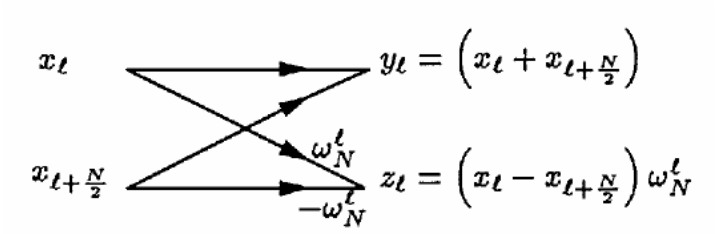
Transformada Discreta de Fourier.

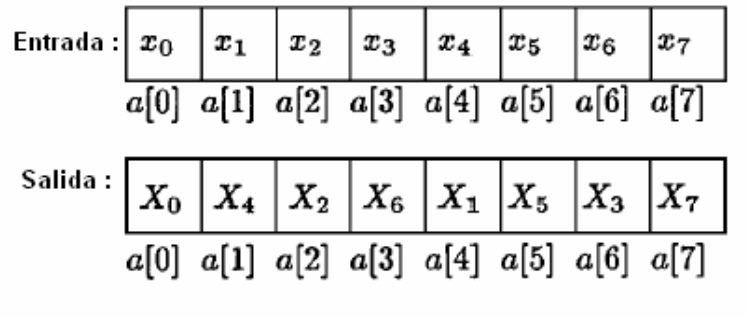
$$X[k] = \sum_{n=0}^{N-1} x[n] e^{\frac{-j2\pi nk}{N}}$$

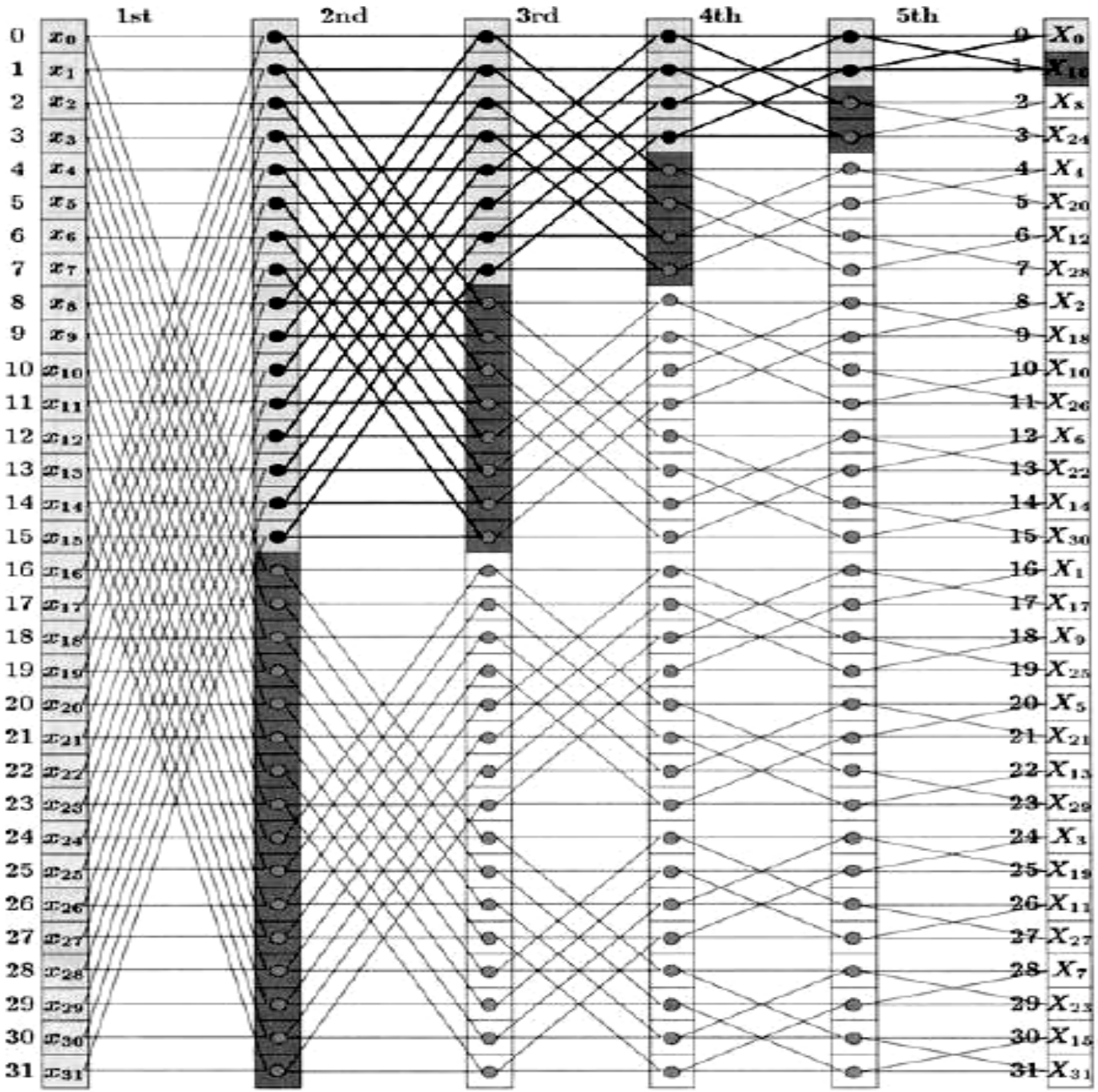
Complejidad computacional : O(n²)

Transformada Rápida de Fourier.

Algoritmo radix-2 con diezmado en frecuencia y reordenamiento de la salida de bits mezclados, cuya complejidad es O(n log n).






1st
2nd
3rd
4th
5th
0 x_0
1 x_1
2 x_2
3 x_3
4 x_4
5 x_5
6 x_6
7 x_7
8 x_8
9 x_9
10 x_{10}
11 x_{11}
12 x_{12}
13 x_{13}
14 x_{14}
15 x_{15}
16 x_{16}
17 x_{17}
18 x_{18}
19 x_{19}
20 x_{20}
21 x_{21}
22 x_{22}
23 x_{23}
24 x_{24}
25 x_{25}
26 x_{26}
27 x_{27}
28 x_{28}
29 x_{29}
30 x_{30}
31 x_{31}
0 X_0
1 X_{16}
2 X_8
3 X_{24}
4 X_4
5 X_{20}
6 X_{12}
7 X_{28}
8 X_2
9 X_{18}
10 X_{10}
11 X_{26}
12 X_6
13 X_{22}
14 X_{14}
15 X_{30}
16 X_1
17 X_{17}
18 X_9
19 X_{25}
20 X_5
21 X_{21}
22 X_{13}
23 X_{29}
24 X_3
25 X_{19}
26 X_{11}
27 X_{27}
28 X_7
29 X_{23}
30 X_{15}
31 X_{31}

Complejidad computacional.

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

Resolviendo la ecuación de recurrencia se tiene:

$$O(n \log n)$$

3. Procesamiento Digital de la Señal de Habla

3.2 Ventaneamiento

Se puede cortar la señal por partes para un análisis más cómodo

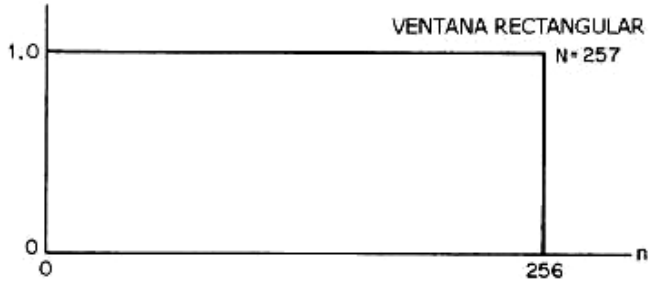
Idea: utilizar ventanitas

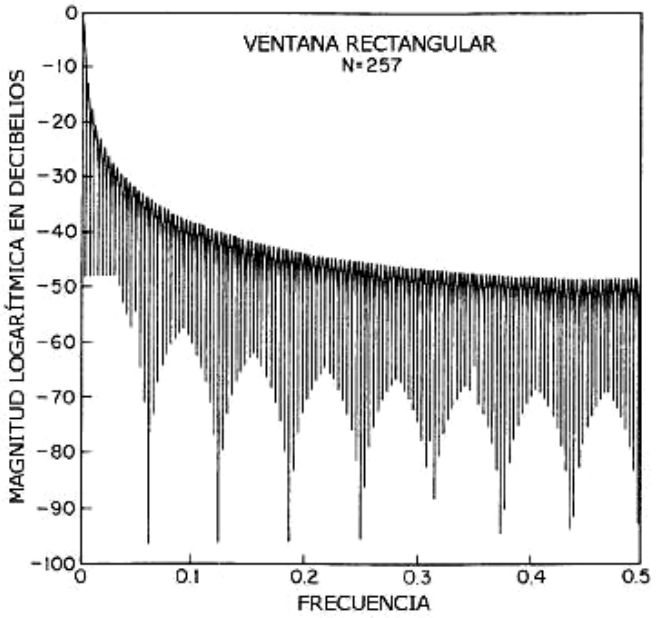
Problema : ¿Qué tipo de ventana usar?

3. Procesamiento Digital de la Señal de Habla

- Caso ventana rectangular

$$w[n] = 1, \qquad 0 \le n \le N-1$$

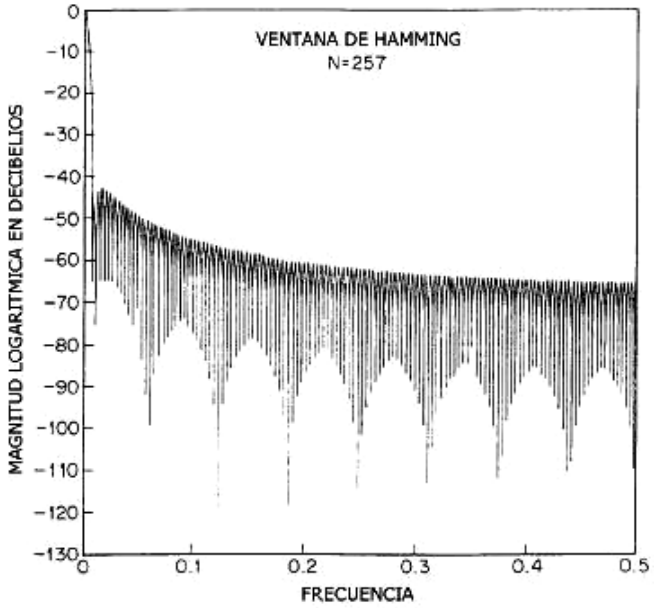




3. Procesamiento Digital de la Señal de Habla

- Caso ventana Hamming

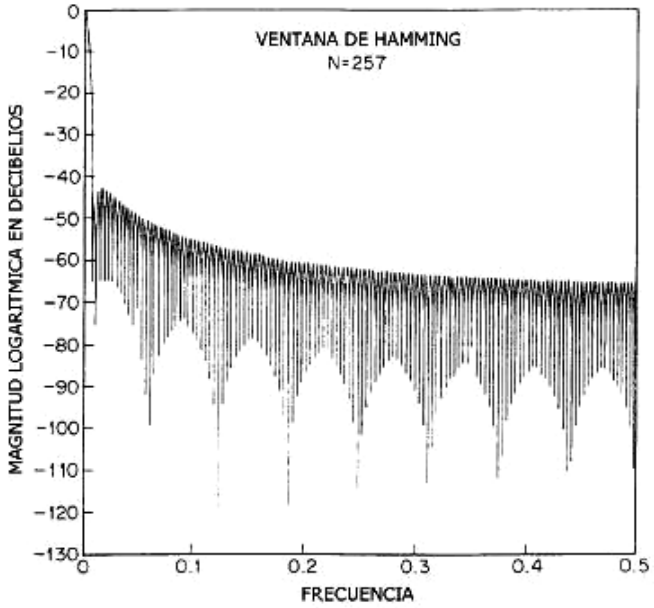
$$w[n] = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), \qquad 0 \le n \le N-1$$



3. Procesamiento Digital de la Señal de Habla

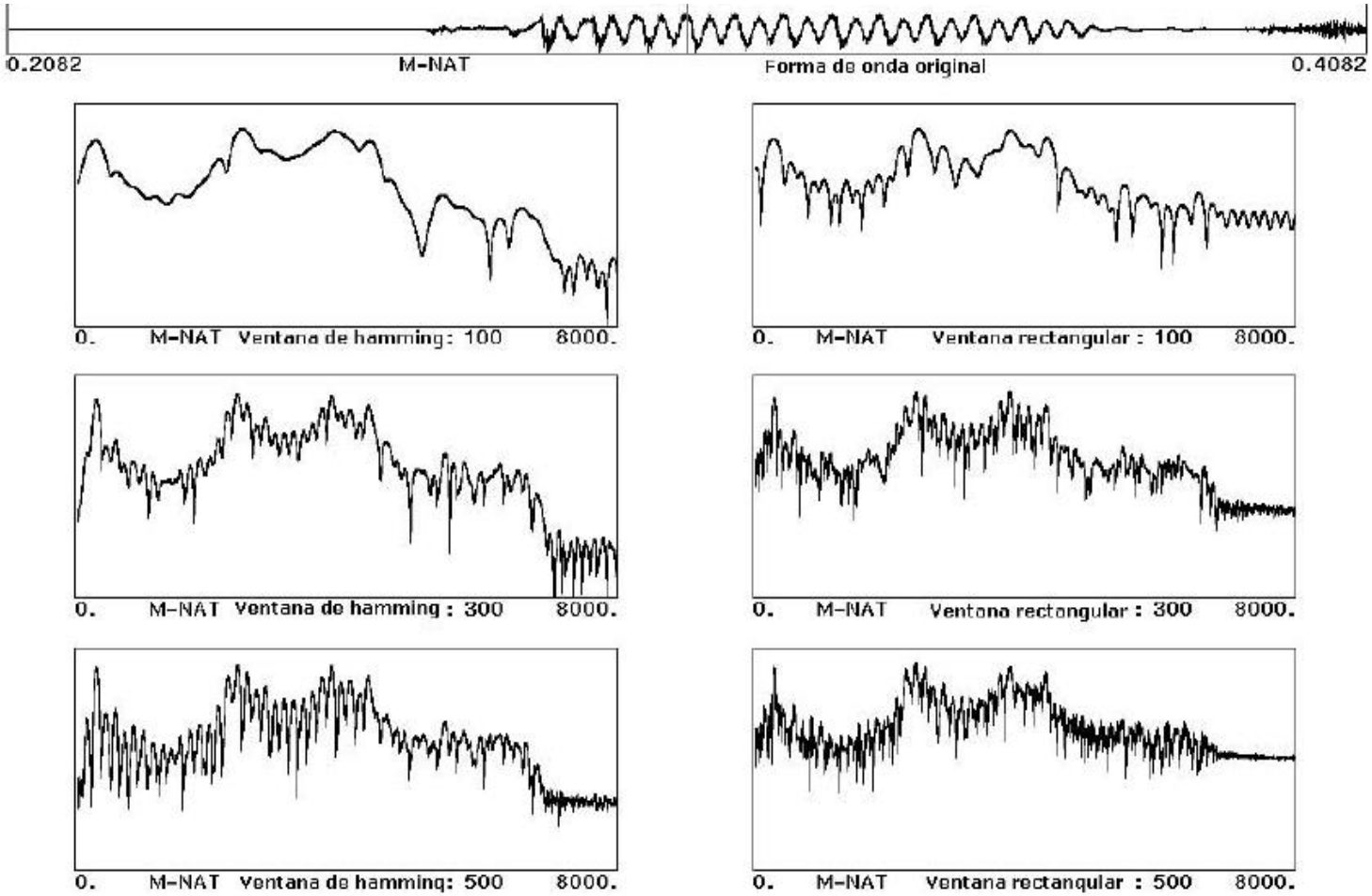
- Por que hamming? Caso ventana rectangular

$$w[n] = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), \qquad 0 \le n \le N-1$$



3. Procesamiento Digital de la Señal de Habla

Comparación de Ventanas.



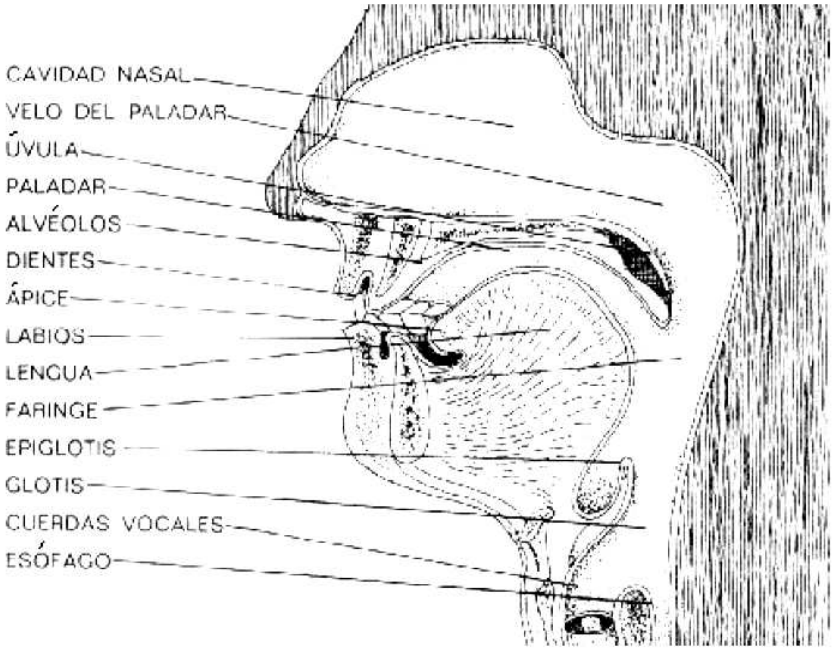
4. Coeficientes MFCC

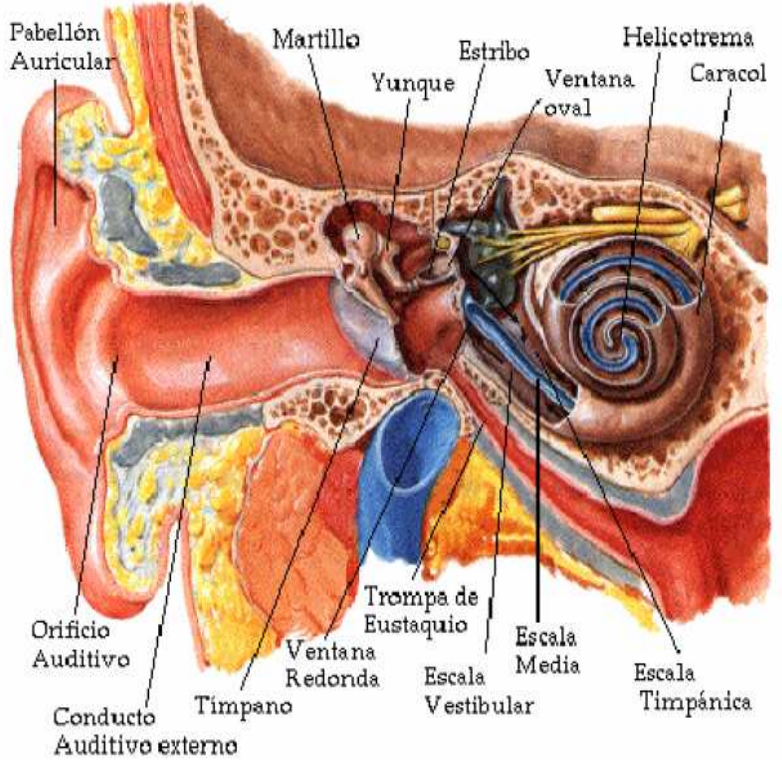
4. Coeficientes MFCC

- Es un método (el más famoso) para extracción de características
- La idea esta inspirada en un modelo biológico
- Usa Trasformada de Fourier
- Complejidad Computacional O(n Log n)

4. Coeficientes MFCC

Producción y Percepción del Habla

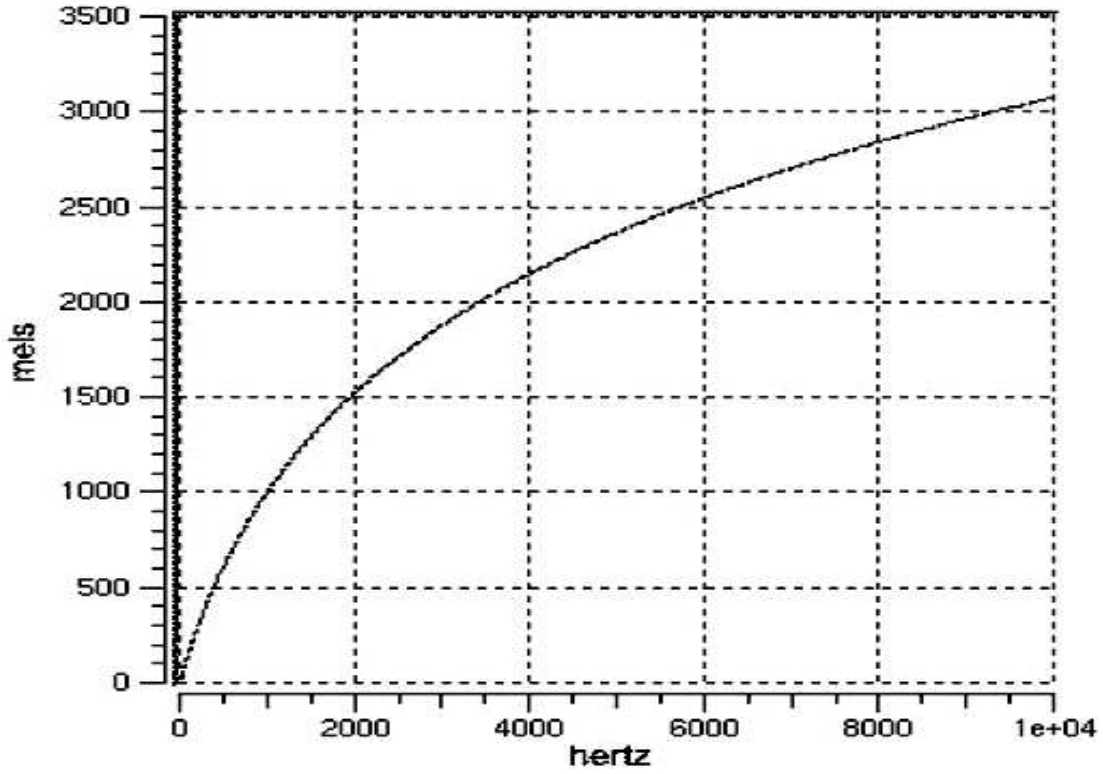




4. Coeficientes MFCC

Frecuencia Mel.

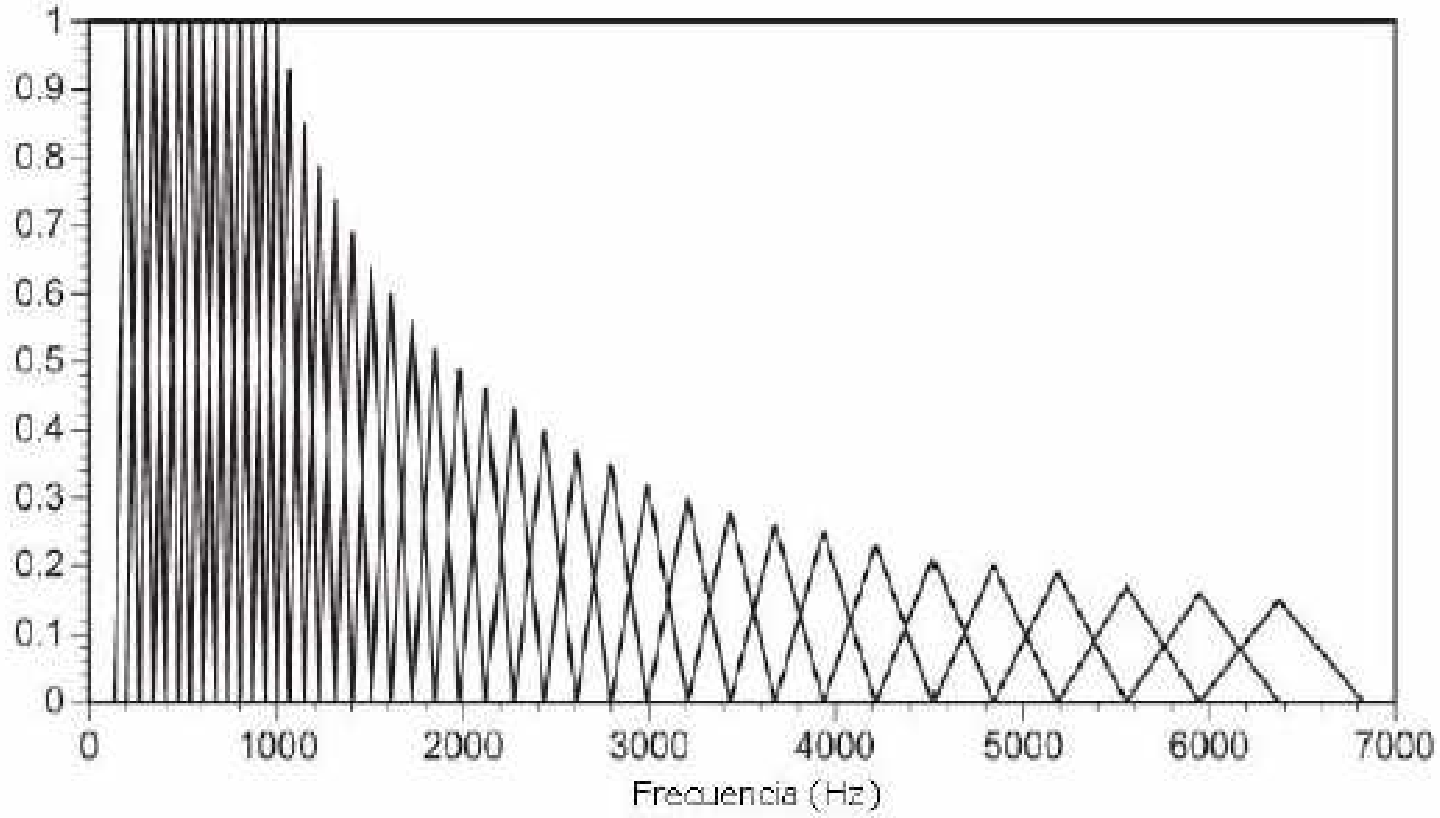
Es una escala basada en como oímos, y se ha construido , a través de experimentos fisiológicos.



$$\beta(f) = 1125 \ln(1 + \frac{f}{700})$$

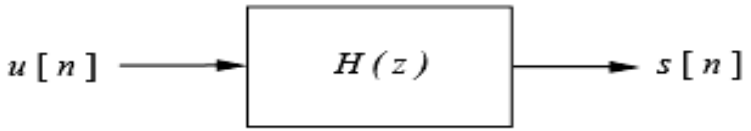
4. Coeficientes MFCC

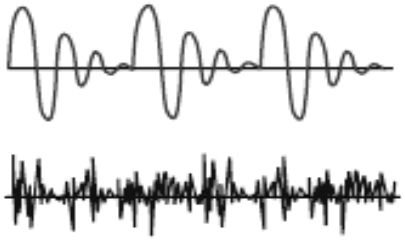
Frecuencia Mel.



Cepstrum.

Si imaginamos la señal de voz como producto de la convolución del aire que fluye de nuestros pulmones y varios filtros correspondientes al tracto vocal.



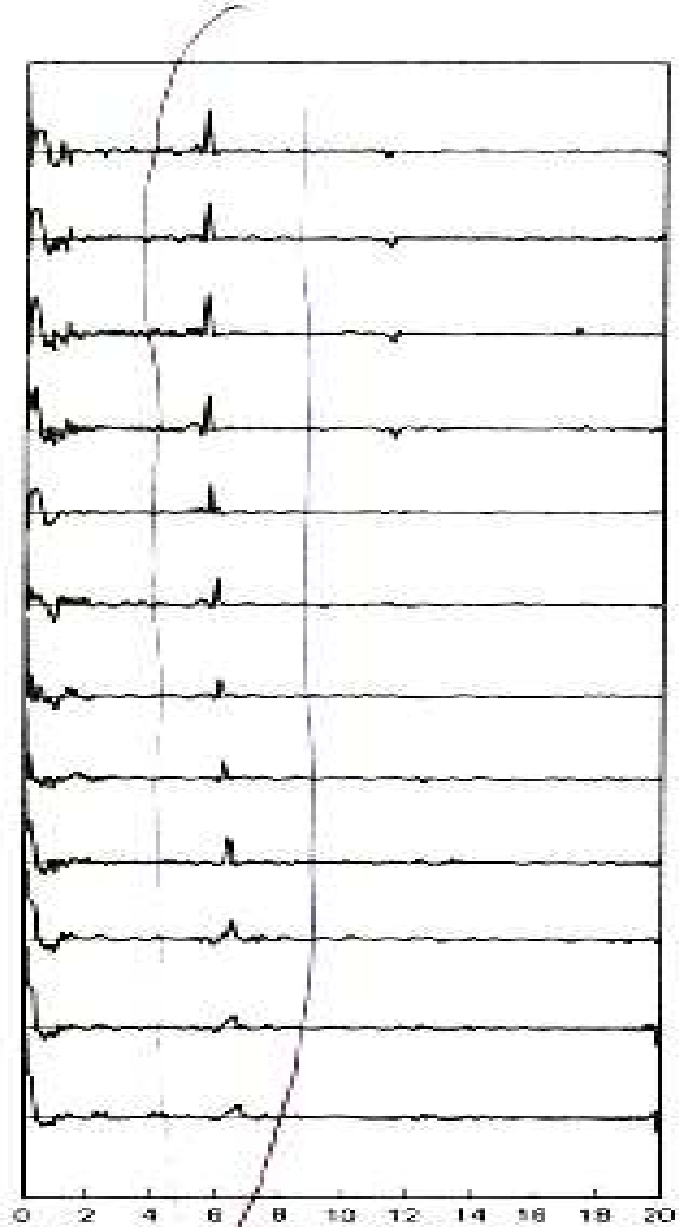


$$x[n] = e[n] * h[n]$$

$$\hat{x}[n] = \hat{e}[n] + \hat{h}[n]$$

Objetivo: Desconvolucionar la señal de voz

$$\hat{x}[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln|X(e^{jw})|e^{jwn}$$


0
2
4
6
8
10
12
14
16
18
20

4. Coeficientes MFCC

- **Algoritmo**
  - Se hace un análisis por frames de la señal

$$\chi^m[n] = \chi[n - mF]\omega[n].$$

Con una ventana Hamming

$$\omega[n] = 0{,}54 - 0{,}46\cos\frac{2\Pi n}{N}.$$

4. Coeficientes MFCC

- **Algoritmo**
  - Se aplica una Trasformada de Fourier a cada Frame (Trasformada Corta de Fourier) con un algoritmo rápido O(n log n)

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

En nuestro caso un algoritmo Radix-2 con decimación en frecuencia y reordenamiento de bits mezclados

4. Coeficientes MFCC

- **Algoritmo**
  - Se traspasa de la escala de frecuencias a la escala Mel, mediante un ventaneamiento con ventanas triangulares (bins)

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

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

4. Coeficientes MFCC

- **Algoritmo**
  - Se obiente el Cepstrum de las frecuencias en escala Mel

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

  - Finalmente se una trasformada de Coseno II es calculada

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

5. Trasformada Wavelet

5. Trasformada Wavelet

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

Ingrid Daubechies

Ten Lectures of Wavelets

5. Trasformada Wavelet

transformada wavelet continua de una función f está dada por:

$$(T^{wav}f)(a,b) = |a|^{-\frac{1}{2}} \int \delta t f(t)\psi(\frac{t-b}{a})$$

la familia de wavelets se puede construir dilatando y trasladando

$$\psi^{a,b}(x) = |a|^{-\frac{1}{2}}\psi(\frac{t-b}{a})$$

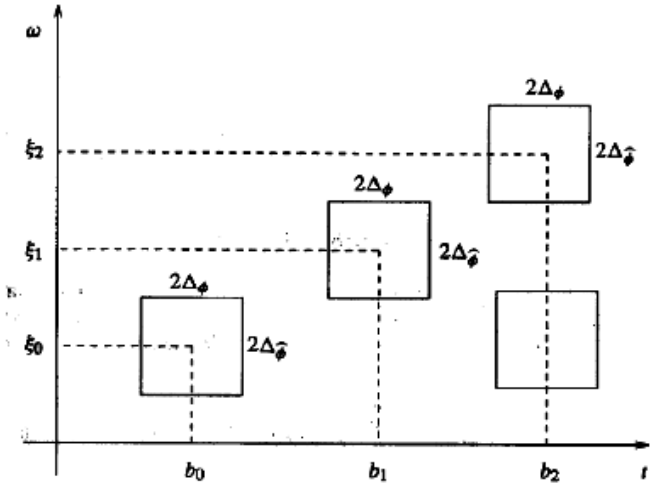
$$f = C_{\psi}^{-1} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\delta a \delta b}{a^2} \psi(T^{wav}f)(a,b)\psi^{a,b}$$

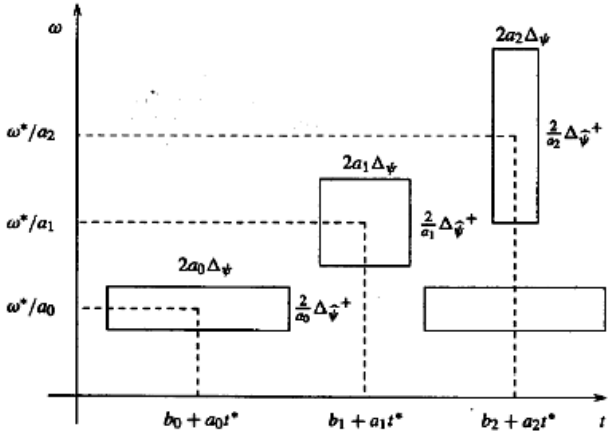
5. Trasformada Wavelet

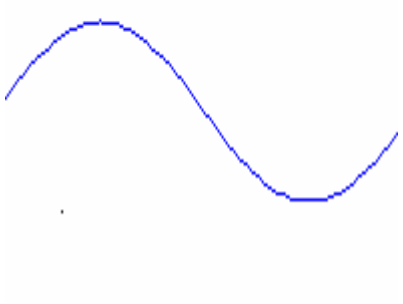
T. Fourier vs T. Wavelet

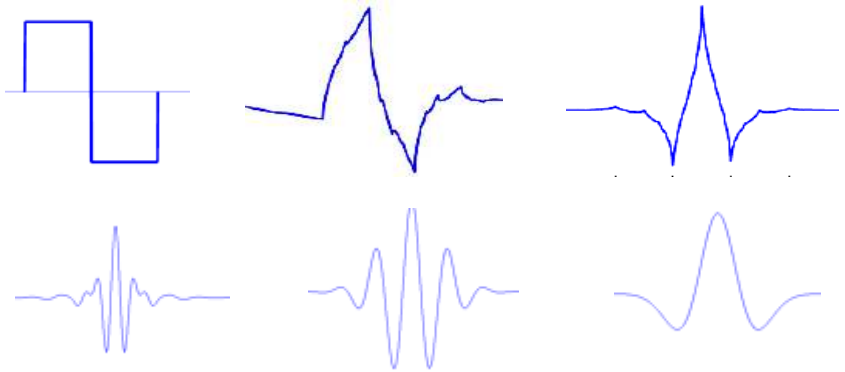
$$T^{win}(w,t) = \int \delta s f(s) g(s-t) e^{-i\omega s}$$

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int x(t) \psi(\frac{t-b}{a}) \delta t$$



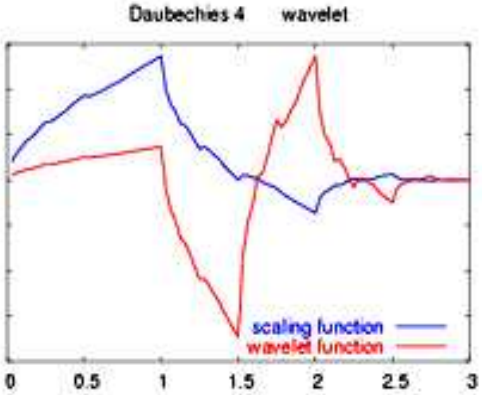


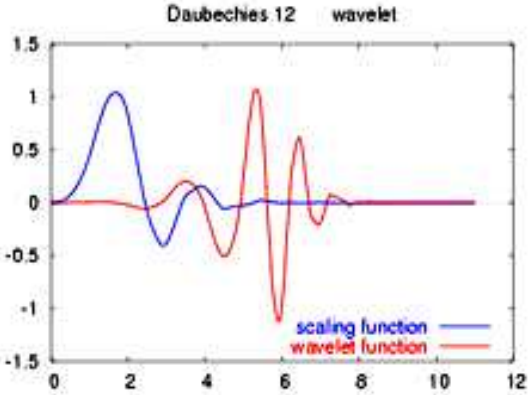


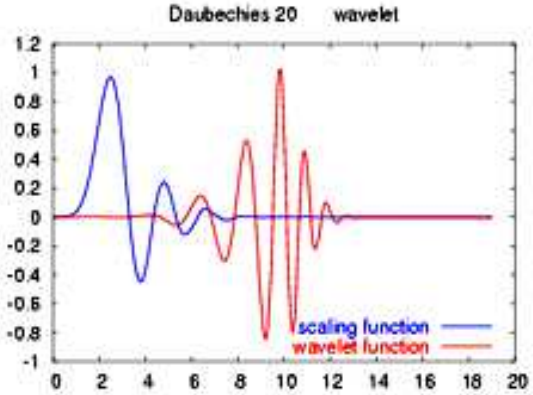


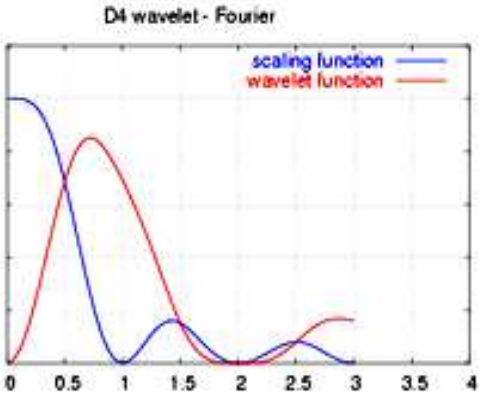
5. Trasformada Wavelet

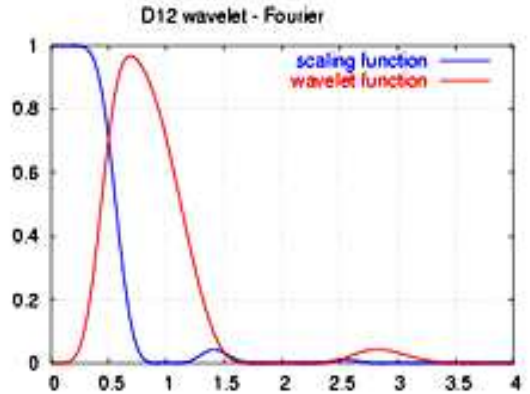
Wavelets en el Dominio de la Frecuencia

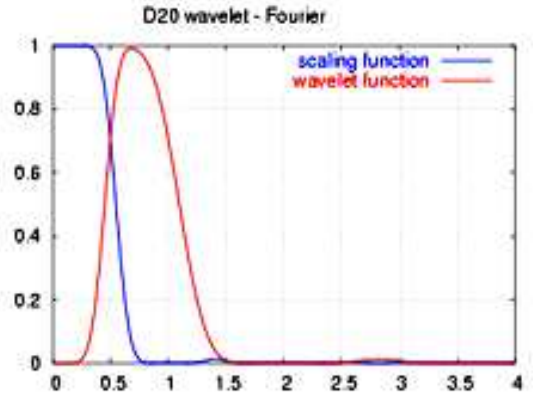












5. Trasformada Wavelet

Wavelets Discretas.

$$\psi^{m,n}(x) = a_0^{-\frac{m}{2}} \psi(a_0^{-m}(x - nb_0a_0^m))$$

$$\psi^{m,n}(x) = a_0^{-\frac{m}{2}} \psi(a_0^{-m}x - nb_0)$$

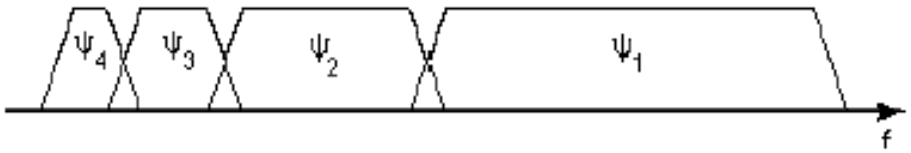
en particular si escogemos $a_0 = 2$ y $b_0 = 1$ entonces:

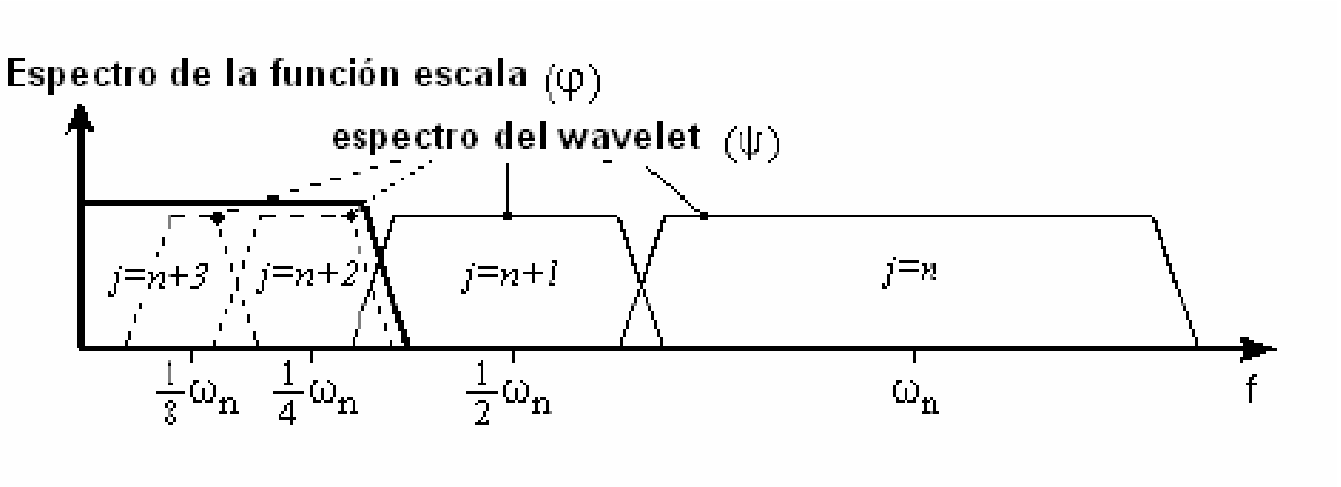
$$\psi^{m,n}(x) = 2^{-\frac{m}{2}} \psi(2^{-m}x - n)$$

5. Trasformada Wavelet

Filtro Pasa Banda

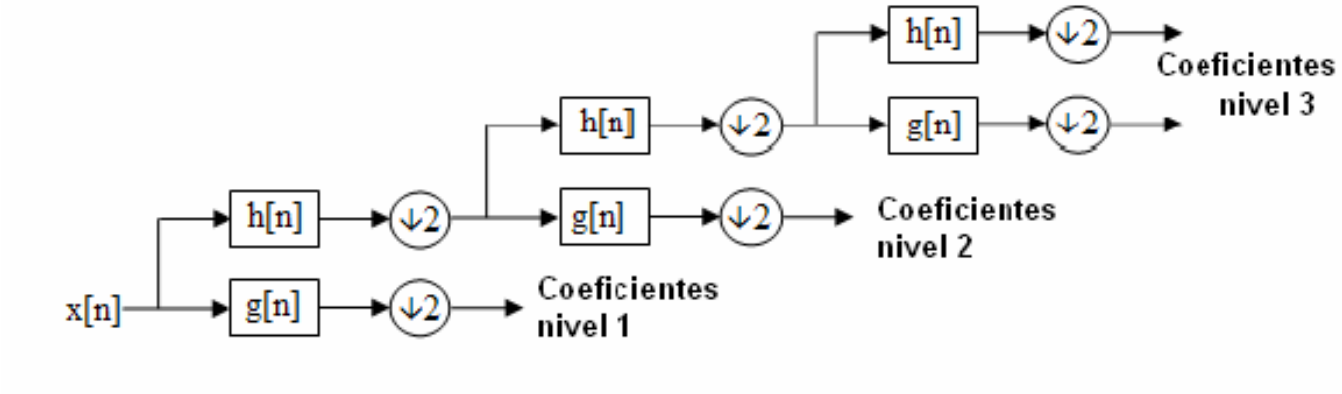
$$H(f(at)) = \frac{1}{|a|} H(\frac{\omega}{a})$$





5. Trasformada Wavelet

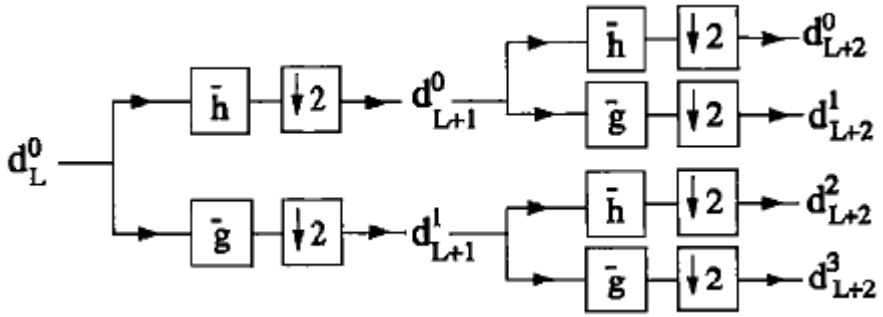
Algoritmo de Banco de Filtros Iterativo

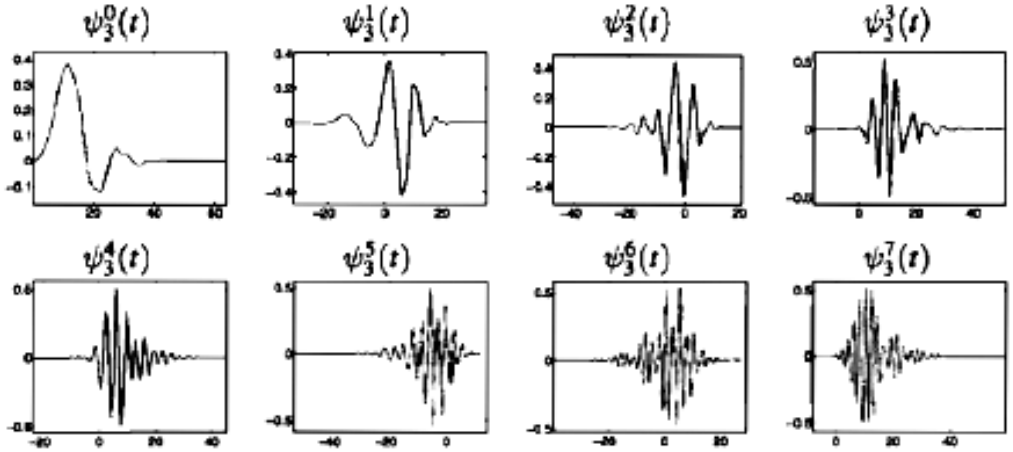


$$T(n) = 2cn - 2c$$

La Transformada wavelet con banco de filtros tiene una complejidad de $O(n)$.

Wavelets Packets

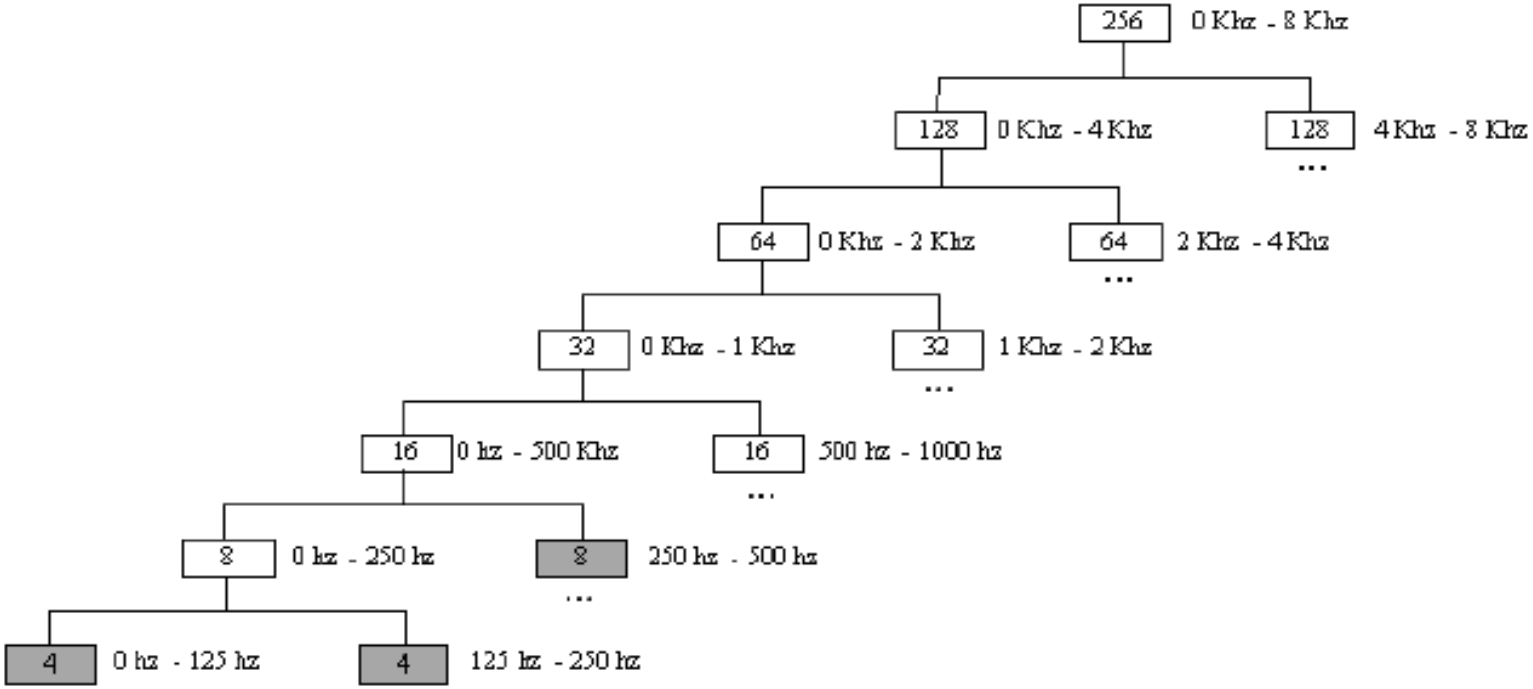




6. Extracción de Características usando Wavelets

Extracción de características con Wavelets

Arbol de descomposición.



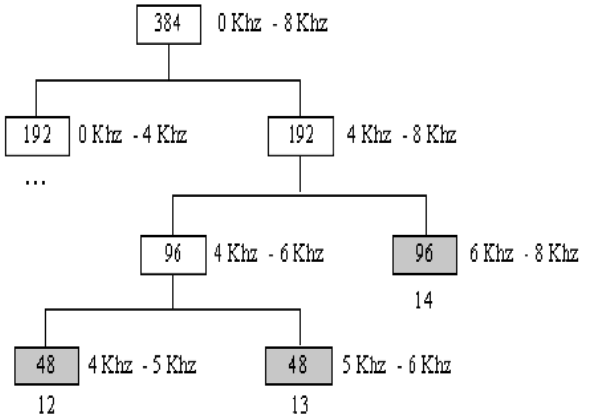
Extracción de Características con Wavelets Packet Perceptuales

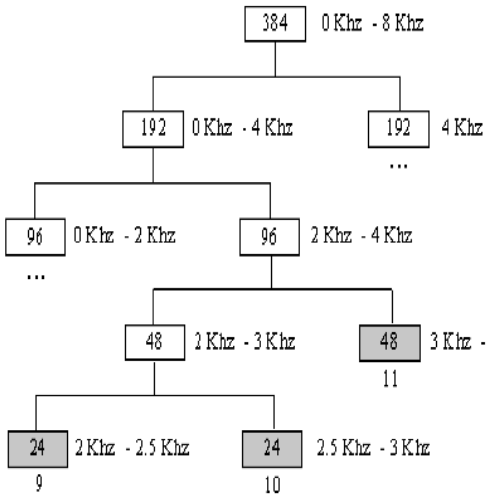
Análisis con Wavelet Packet paso (202.45) | M

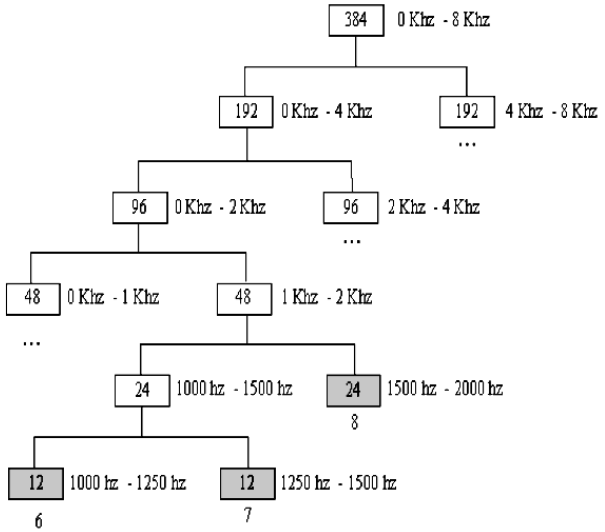
Equivalencias escala MEL a Frecuencias

	MEL	Frecuencias
1	0	0
2	202.10	135
3	404.21	300
4	606.31	500
5	808.42	735
6	1010.53	1030
7	1212.63	1375
8	1414.73	1750
9	1616.85	2235
10	1818.95	2830
11	2021.06	3500
12	2223.16	4333
13	2425.27	5310
14	2627.38	6500
	2829.48	8000

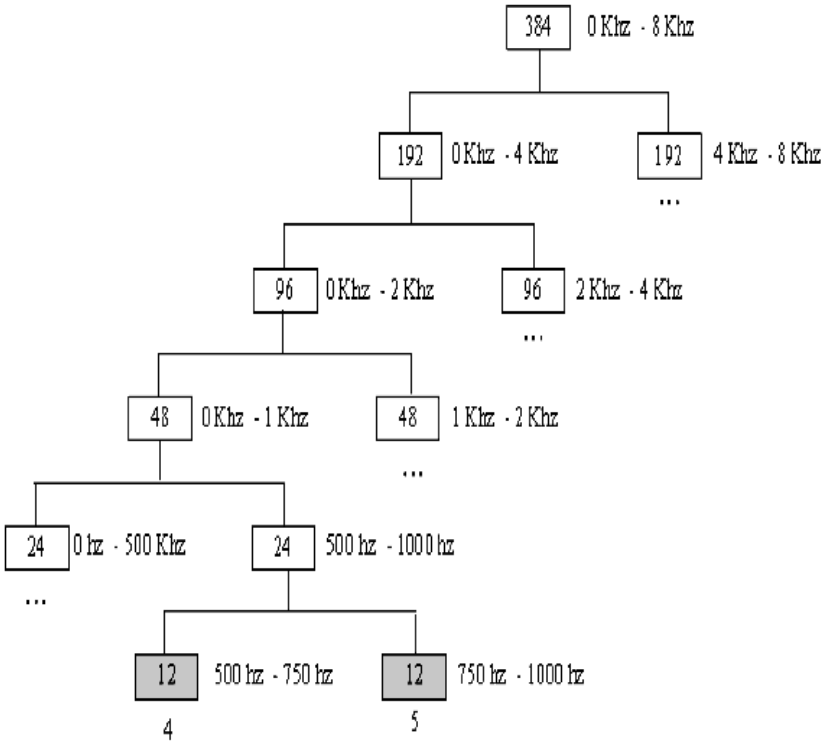
Arbol de descomposición.

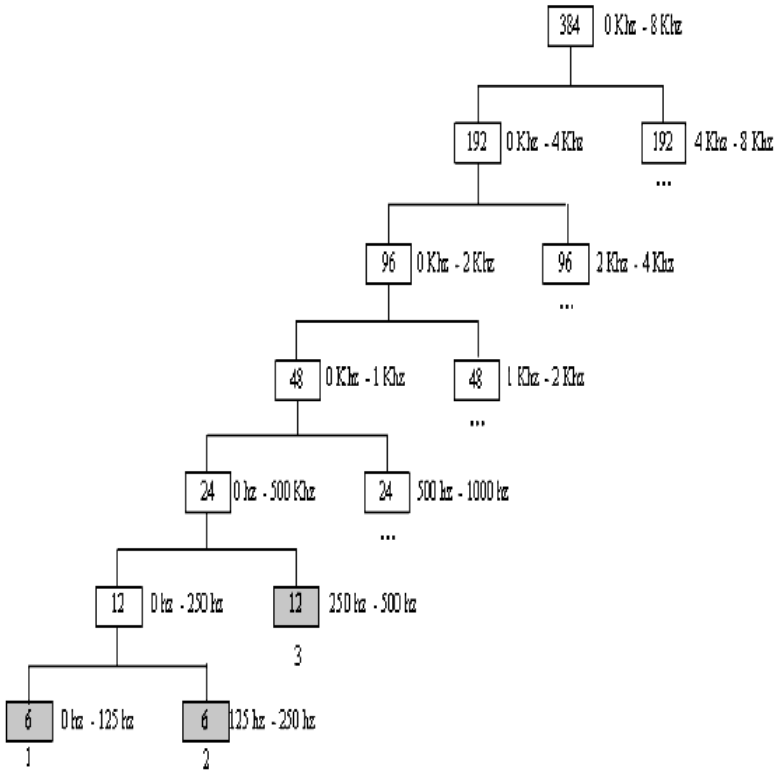






Extracción de Características con Wavelets Packet Perceptuales





Filtros Usados

Haar

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

Wavelets de Daubechies 4

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

Wavelets de Daubechies 6

$$h(n) = [0{,}3326, 0{,}8068, 0{,}4598, -0{,}1350, -0{,}0854, 0{,}0352]$$

Wavelet Coiflet 6

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

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

Extracción de características

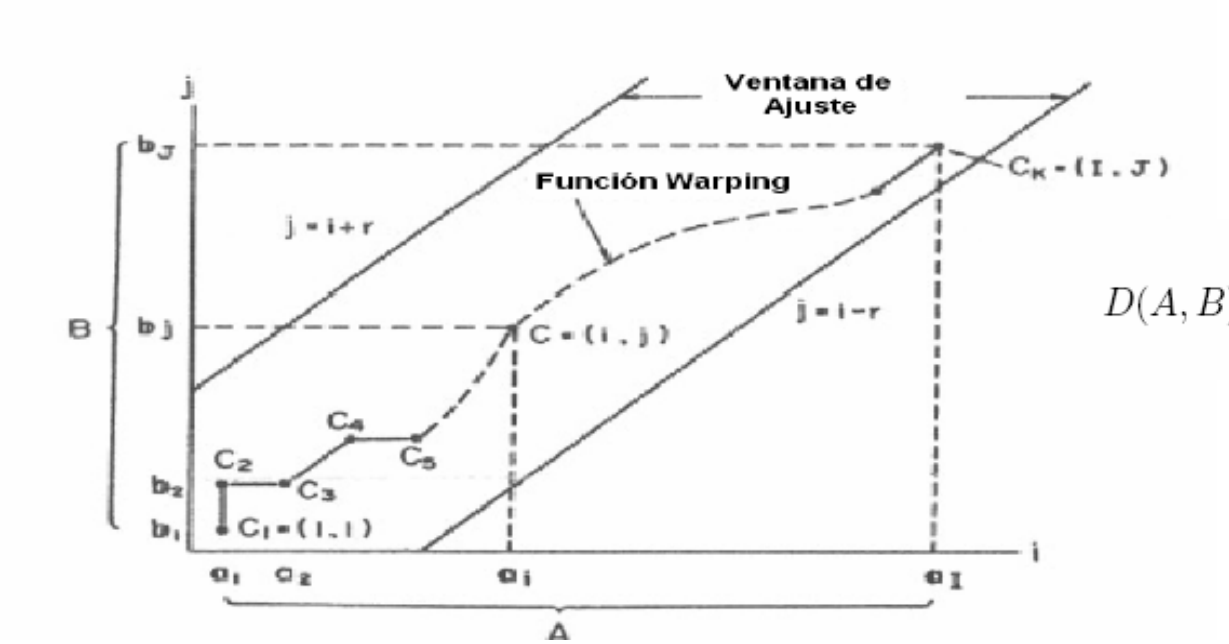
- Cálculo de energias por nivel del resolución aproximadamente igual a la escala Mel

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

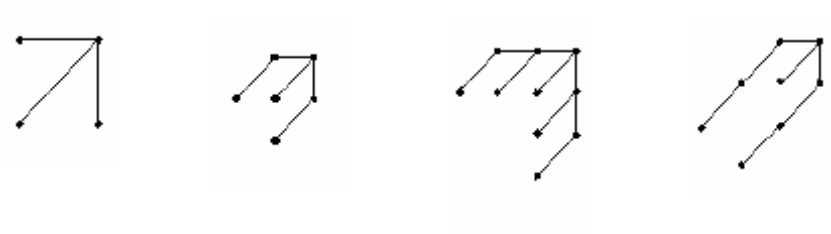
- Aplicación de un “Cepstrum”

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

Dynamic Time Warping DTW



$$D(A,B) = \min\left[\frac{\sum_{k=1}^{K} d(c(k)).w(k)}{\sum_{k=1}^{K} w(k)}\right]$$



Experimentos y Resultados

Resultados Wavelets O(n)

DISTRIBUCION DE 900 PALABRAS PRONUNCIADAS POR 60 PERSONAS SEGUN SU IDENTIFICACION Y NO IDENTIFICACION POR LA COMPUTADORA CON EL MÉTODO DE FOURIER Y EL MÉTODO PROPUESTO BASADO EN LAS WAVELET HAAR.

TRUJILLO 2006.

MÉTODO WAVELET HAAR		MÉTODO DE FOURIER: Palabras identificadas	MÉTODO DE FOURIER: Palabras no identificadas	TOTAL
-	Palabras identificadas	262	13	275
-	Palabras no identificadas	172	453	625
	TOTAL	434	466	900

χ^2_{Mc} : Prueba ji-cuadrado de Mc Nemar para datos correlacionados

χ^2_{Mc} = 136.65	p < 0.01

Eficiencia relativa de aciertos = 275 / 434 x 100

EF :63.36%

DISTRIBUCION DE 900 PALABRAS PRONUNCIADAS POR 60 PERSONAS SEGUN SU IDENTIFICACION Y NO IDENTIFICACION POR LA COMPUTADORA CON EL METODO DE FOURIER Y EL MÉTODO PROPUESTO BASADO EN LAS WAVELETS DE DAUBECHIES 6 .

TRUJILLO 2006.

MÉTODO WAVELET DAUB 6		MÉTODO DE FOURIER: Palabras identificadas	MÉTODO DE FOURIER: Palabras no identificadas	TOTAL
-	Palabras identificadas	430	93	523
-	Palabras no identificadas	4	373	377
	TOTAL	434	466	900

χ^2_{Mc} : Prueba ji-cuadrado de Mc Nemar para datos correlacionados

χ^2_{Mc} = 81.66	p < 0.01

Eficiencia relativa de aciertos = 523 / 434 x 100

ER-aciertos = 120.51

Resultados Wavelets Packets O(n log n)

DISTRIBUCION DE 900 PALABRAS PRONUNCIADAS POR 60 PERSONAS SEGUN SU IDENTIFICACION Y NO IDENTIFICACION POR LA COMPUTADORA CON EL METODO DE FOURIER Y EL MÉTODO PROPUESTO BASADO EN LAS WAVELETS PACKET DUABECHIES 6.

TRUJILLO 2006.

MÉTODO WAVELET PACKET DAUB6	MÉTODO DE FOURIER		TOTAL
	Palabras identificadas	Palabras no identificadas	
- Palabras identificadas	425	220	645
- Palabras no identificadas	9	246	255
TOTAL	434	466	900

χ^2_{Mc} : Prueba ji-cuadrado de Mc Nemar para datos correlacionados

χ^2_{Mc} = 194.41	p < 0.01

Eficiencia relativa de aciertos = 645 / 434 x 100

ER-aciertos = 148.62

DISTRIBUCION DE 900 PALABRAS PRONUNCIADAS POR 60 PERSONAS SEGUN SU IDENTIFICACION Y NO IDENTIFICACION POR LA COMPUTADORA CON EL METODO DE FOURIER Y EL METODO PROPUESTO BASADO EN LAS WAVELETS PACKET PERCEPTUAL CON DAUBECHIES 4.

TRUJILLO 2006.

MÉTODO WAVELET PACKET PERCEPTUAL	MÉTODO DE FOURIER		TOTAL
	Palabras identificadas	Palabras no identificadas	
- Palabras identificadas	420	183	603
- Palabras no identificadas	14	283	297
TOTAL	434	466	900

χ^2_{Mc} : Prueba ji-cuadrado de Mc Nemar para datos correlacionados

χ^2_{Mc} = 144.98	p < 0.01

Eficiencia relativa de aciertos = 603 / 434 x 100

ER-aciertos = 138.94

Comparaciones con MFCC

Método	*Tasa aceptación*	*Error*
Coeficientes 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%

Tabla 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

Tasa de Reconocimiento por Palabra

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

8. Conclusiones

Conclusiones.

-El mejoramiento del espectro se da gracias al análisis tiempo frecuencia de las wavelets.

-Una extracción de características usando solamente la Trasformada de Fourier no brinda buenos resultados.

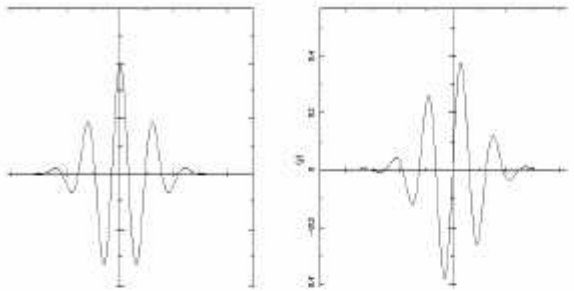
-Los wavelets pueden ser utilizados alternativamente, para el procesamiento digital de la señal de habla.

-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 ventaja de utilizar wavelets radica, en la variedad de funciones wavelet que se puede escoger.

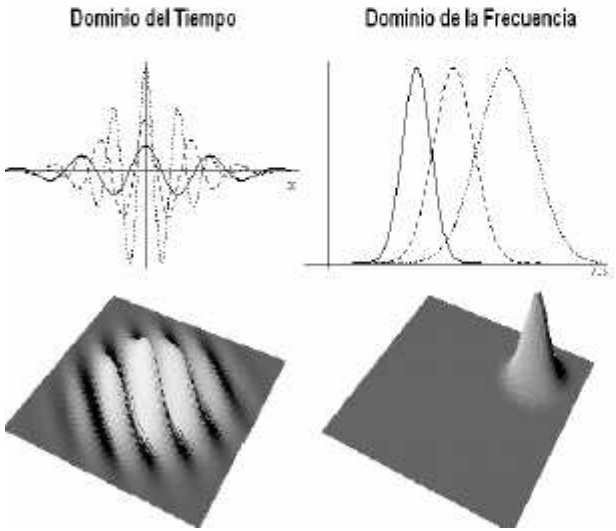
- Las wavelets que mejor funcionan, son aquellos que tienen su espectro parecido a un filtro paso de banda ideal.

Proyecciones



Parte real e imaginaria de las Wavelets de Morlet

$$\psi(x) = \pi^{-\frac{1}{4}}(e^{-i\omega x} - e^{-\frac{\pi^2\alpha^2}{4}})e^{-\frac{x^2}{\alpha^2}}$$



Método	*Tasa aceptación*	*Error*
Coeficientes Cepstrales en Escala Mel	85.32%	14.68%
Wavelet Continuo de Morlet	87.32%	12.68%

$$O(n \log n).$$

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. Objetive wavelet packet features for speaker verication, 2000.

Software de pruebas

LORITO version 3.14