

PROCESAMIENTO DE LA SEÑAL DE HABLA

La señal de habla debe preprocesarse antes de ser utilizada, por ejemplo eliminación o atenuación del ruido, compensación de ciertas frecuencias, etc.

1. Filtros

Si queremos extraer ciertos componentes de frecuencia presentes de la señal, por ejemplo eliminar cierto ruido, o procesar la señal de tal manera que ciertas frecuencias significativas brinden información de la señal para su posterior extracción de características se utilizan lo que se conocen como filtros, que no son otra cosa que funciones que tienen cierto comportamiento en el dominio de la frecuencia.

Por ejemplo si deseamos obtener las bajas frecuencias presentes en una señal de habla, lo lógico será transformar la señal de habla al dominio de la frecuencia y aplicar un **filtro paso bajo**, pues este filtro dejaría pasar las bajas frecuencias de la señal y retendría las altas frecuencias de la señal.

Filtro paso bajo ideal

Imaginemos que existe una función $h[n]$ cuya transformada de Fourier es la siguiente:

$$H(e^{j\omega}) = \begin{cases} 1 & |\omega| < \omega_0 \\ 0 & \omega_0 < |\omega| < \pi \end{cases}$$

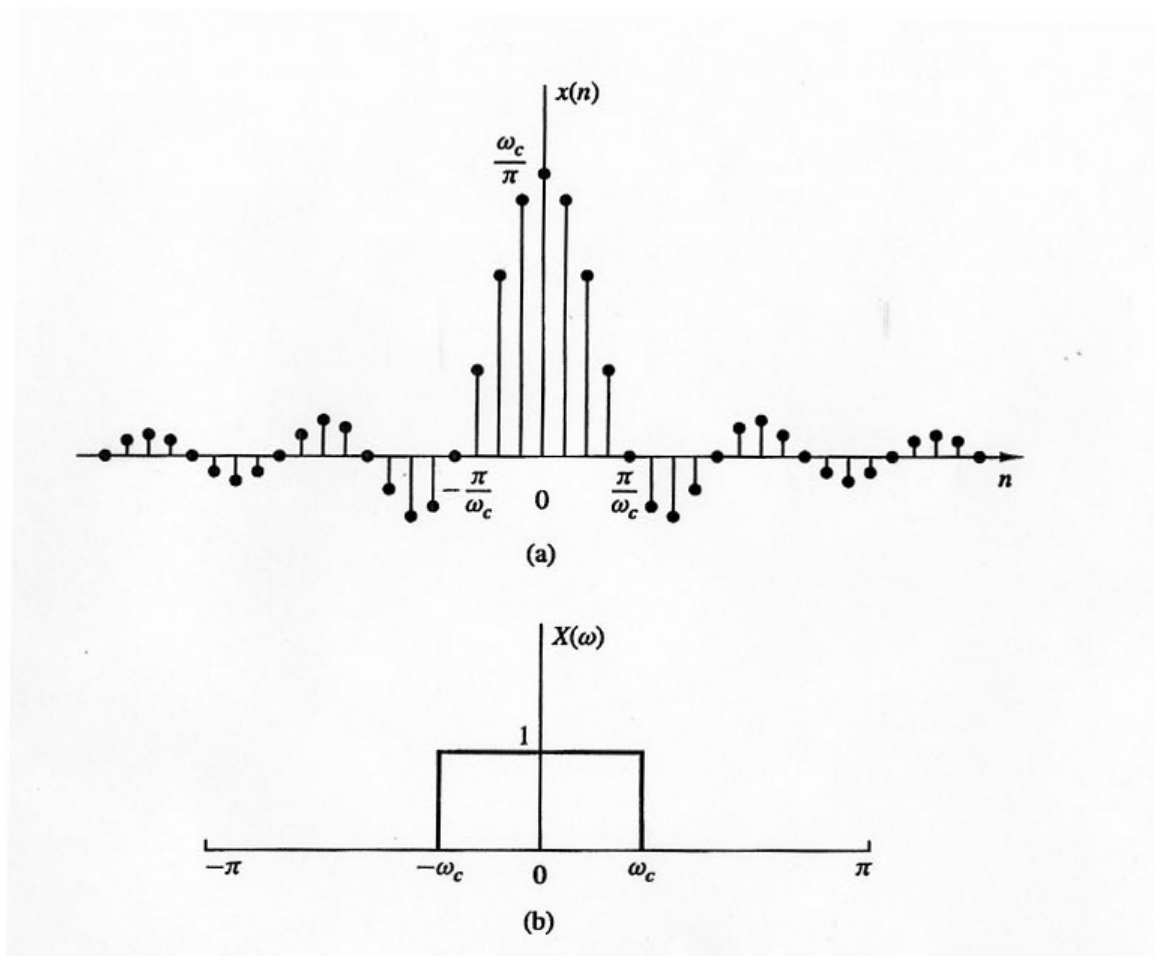
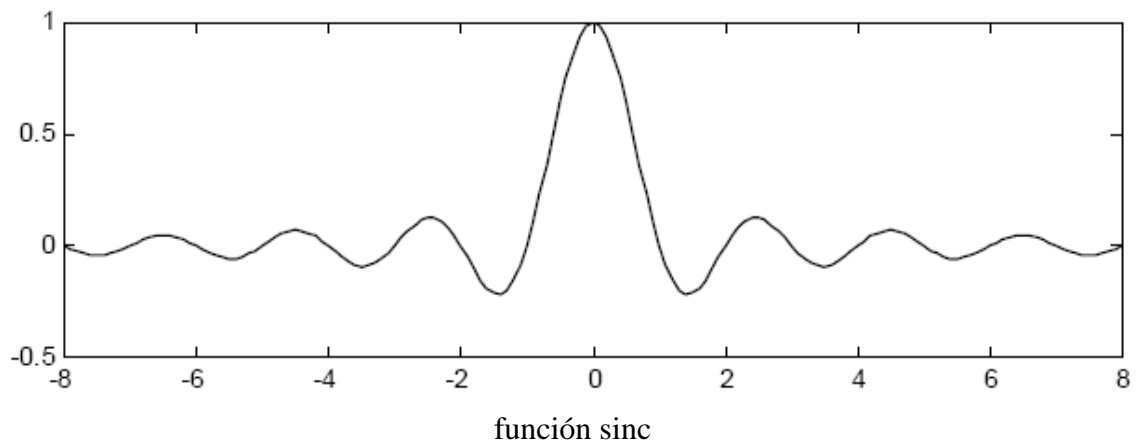
$h[n]$ está trabajando como un filtro paso bajo ideal pues deja pasar todas las frecuencias por debajo de ω_0

Si analizamos su transformada de Fourier obtendremos

$$h[n] = \frac{1}{2\pi} \int_{-\omega_0}^{\omega_0} e^{j\omega n} d\omega = \frac{(e^{j\omega_0 n} - e^{-j\omega_0 n})}{2\pi j n} = \frac{\sin \omega_0 n}{\pi n} = \left(\frac{\omega_0}{\pi} \right) \text{sinc}(\omega_0 n)$$

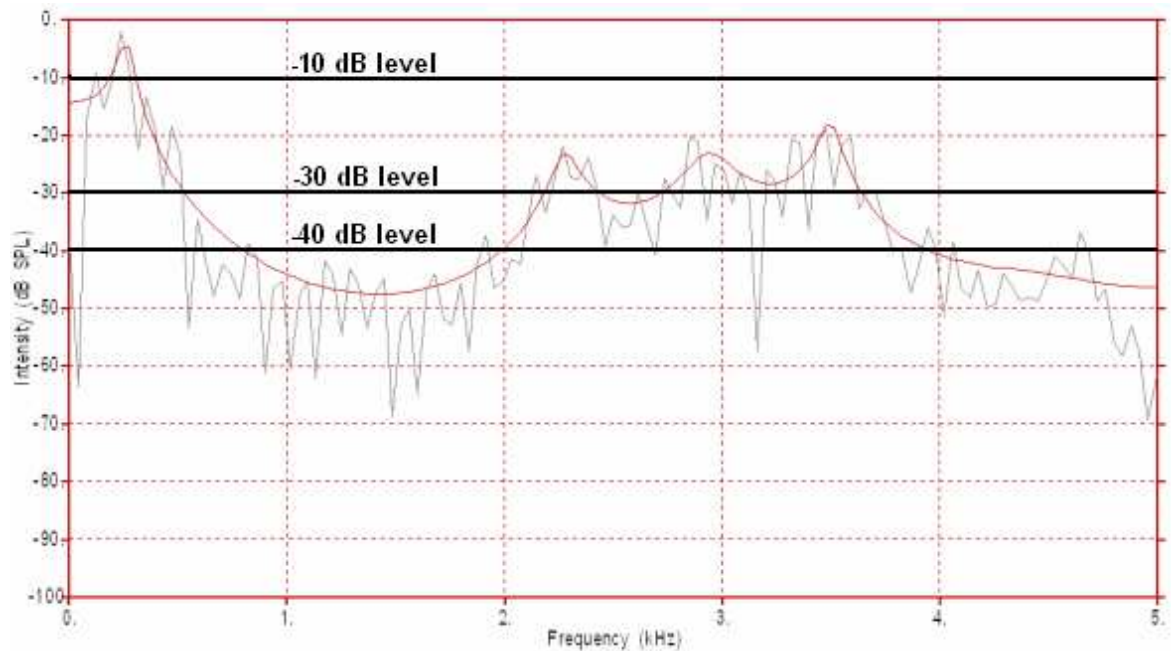
donde

$$\text{sinc}(x) = \frac{\sin \pi x}{\pi x}$$

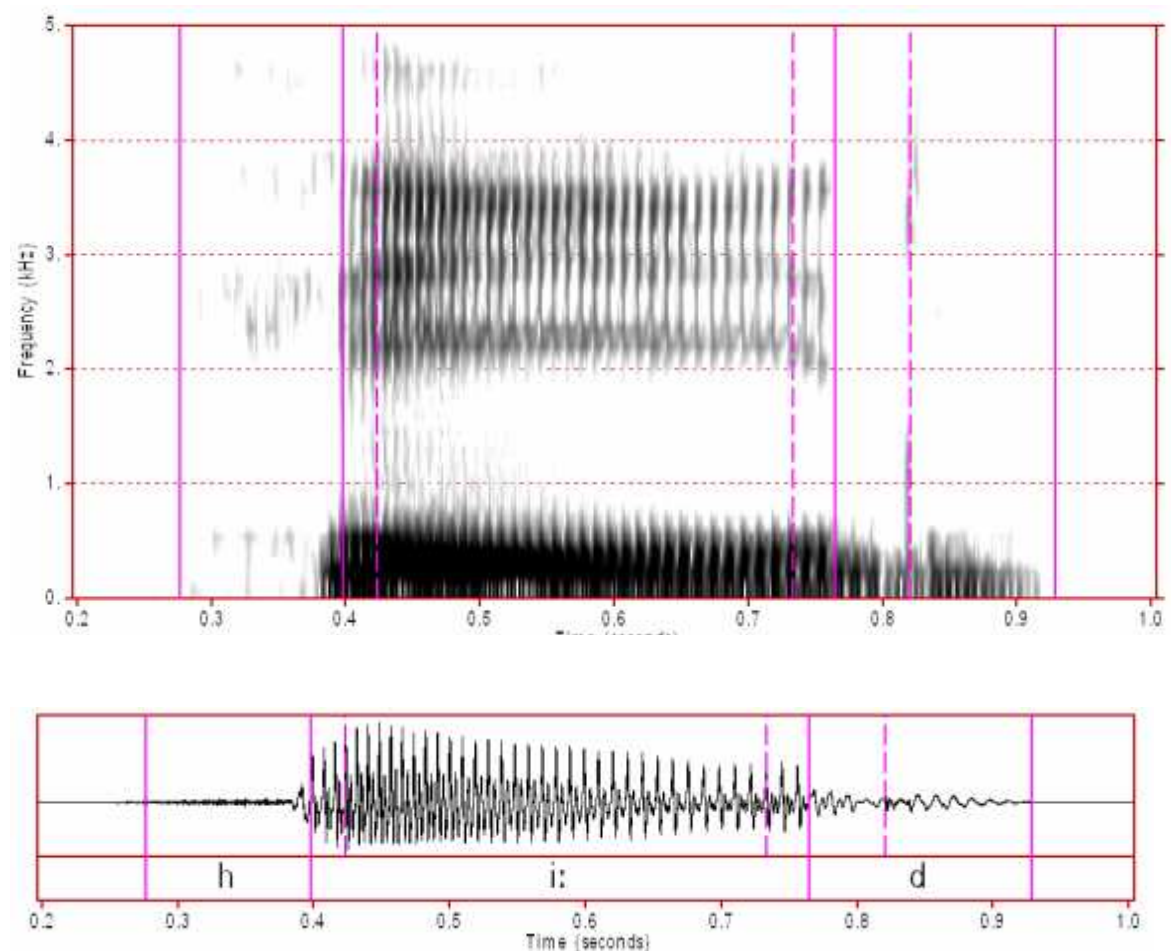


2. Filtro de Pre-enfásis

Antes de empezar a estudiar los algoritmos de extracción de características, analizaremos como se comportan las frecuencias de las voces grabadas.

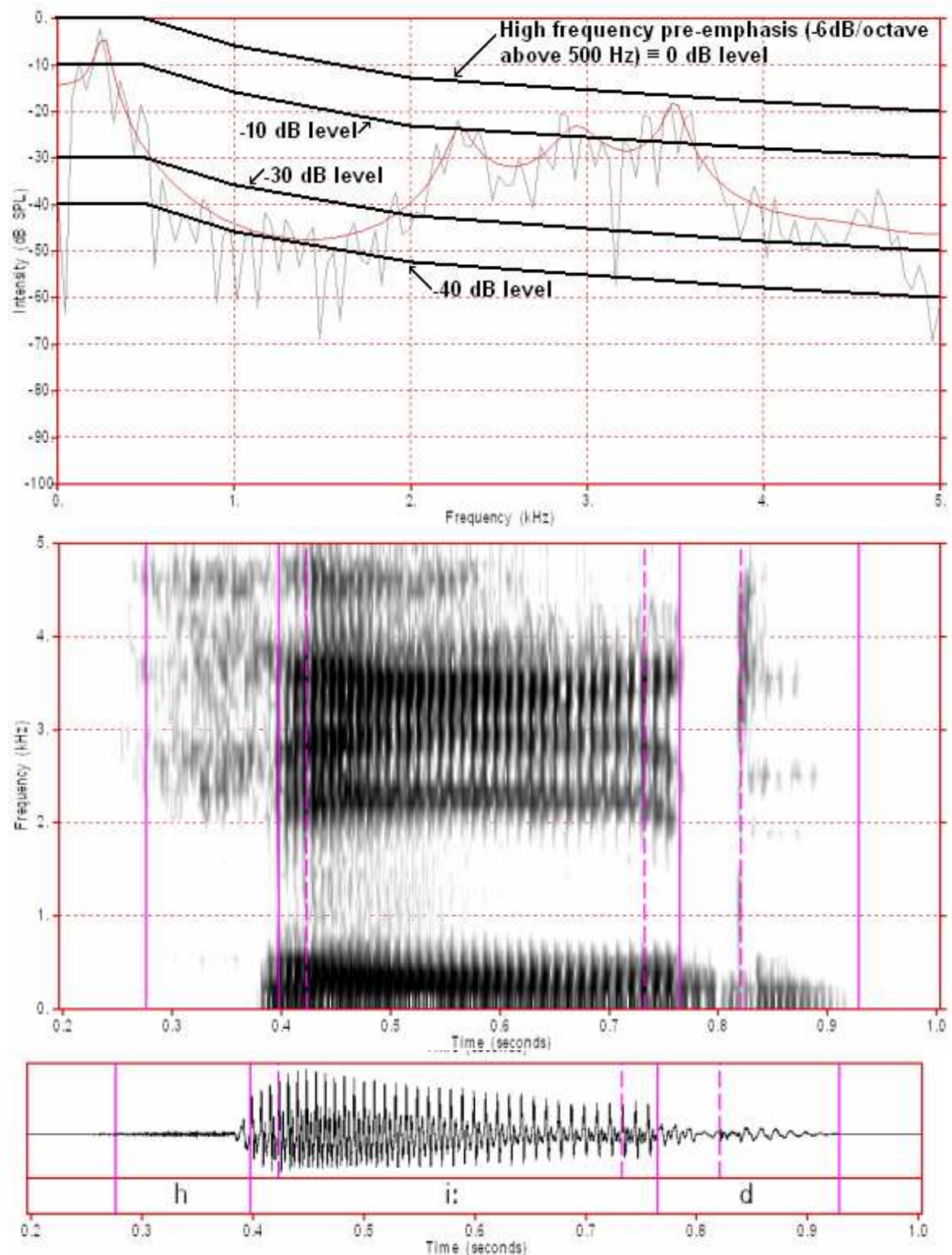


En la figura se muestra la trasformada de Fourier de una palabra grabada,



El problema de esta figura del espectograma es que la primera formante luce mucho mas oscura que las otras formantes, dando la impresión que la primera formate es más,

importante que la segunda formante, lo que no es correcto, para solucionar este problema utilizaremos un filtro llamado filtro pre-énfasis de altas frecuencias



Como se muestra en las imágenes la señal de habla debe ser pasada por un filtro especial llamado filtro pre-énfasis de altas frecuencias y que está motivado por el modelo de

producción de habla pues existe aproximadamente 6dB de decaimiento a la combinación de glotal y los labios, para esto se usa el siguiente filtro

$$y[n] = x[n] + \alpha x[n - 1]$$

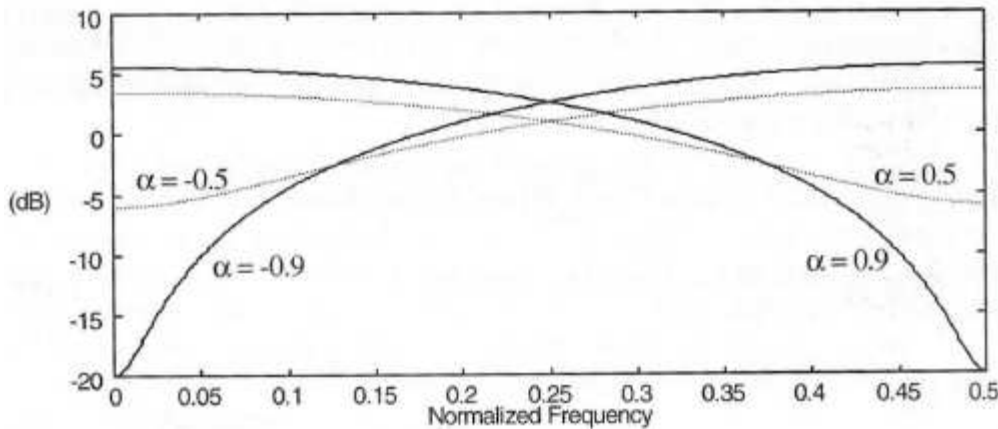


Figure 5.21 Frequency response of the first order FIR filter for various values of α .

cuando $\alpha > 0$ tenemos un filtro paso bajo y cuando $\alpha < 1$ tenemos un filtro paso alto, este filtro es el que permite la compensación en decibeles de las frecuencias de la señal de habla.

3. Ventanamiento

Analizar las frecuencias presentes en la señal de habla muchas veces es mejor hacerlas por tramos en lugar de analizar toda la señal entera, para poder encontrar información de las frecuencias en un dominio del tiempo específico, para esto se procede a realizar un proceso de ventaneamiento de la señal es decir aplicar sobre una parte de la señal una función llamada función ventana, existen diversos tipos de ventana, a continuación estudiaremos 3 de ellas rectangular, hamming y hanning

Ventana Rectangular

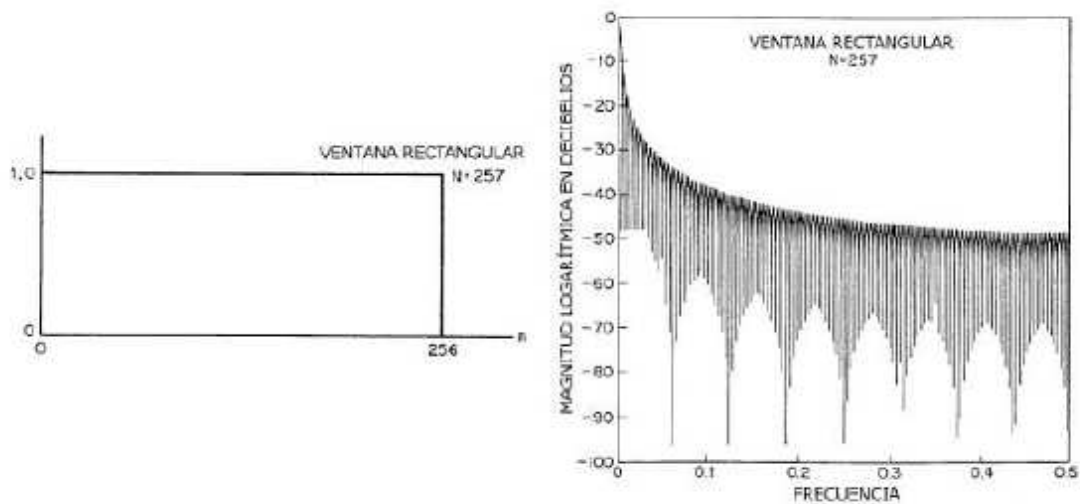
Está definida como:

$$h_n = u[n] - u[n - N]$$

donde u es la función de paso unitario definida como:

$$u[n] = \begin{cases} 1 & \text{si } n \geq 0 \\ 0 & \text{si } n < 0 \end{cases}$$

y N es el tamaño de la ventana



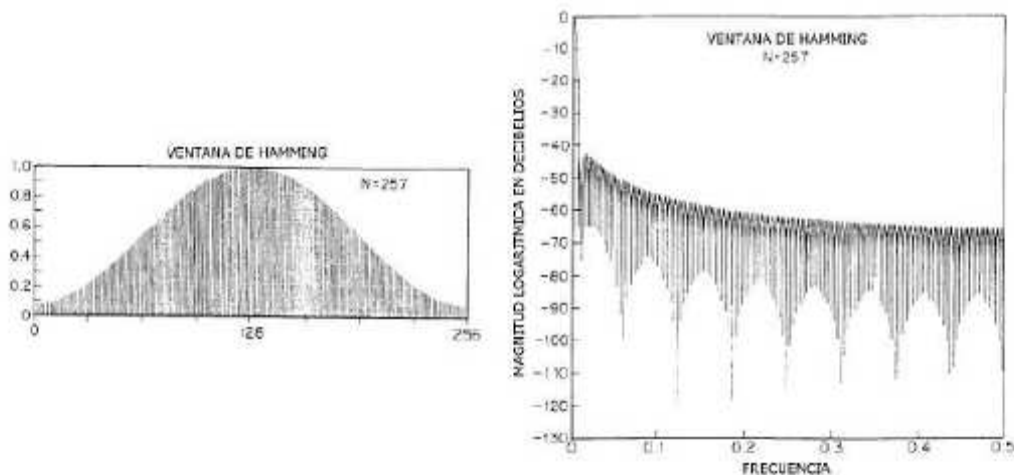
Ventana Generalizada Hamming

Está definida como:

$$h[n] = \begin{cases} (1 - \alpha) - \alpha \cos\left(\frac{2\pi n}{N}\right) & \text{si } 0 \leq n < N \\ 0 & \text{otra manera} \end{cases}$$

Cuando $\alpha = 0.5$ esta función es conocida como ventana Hanning

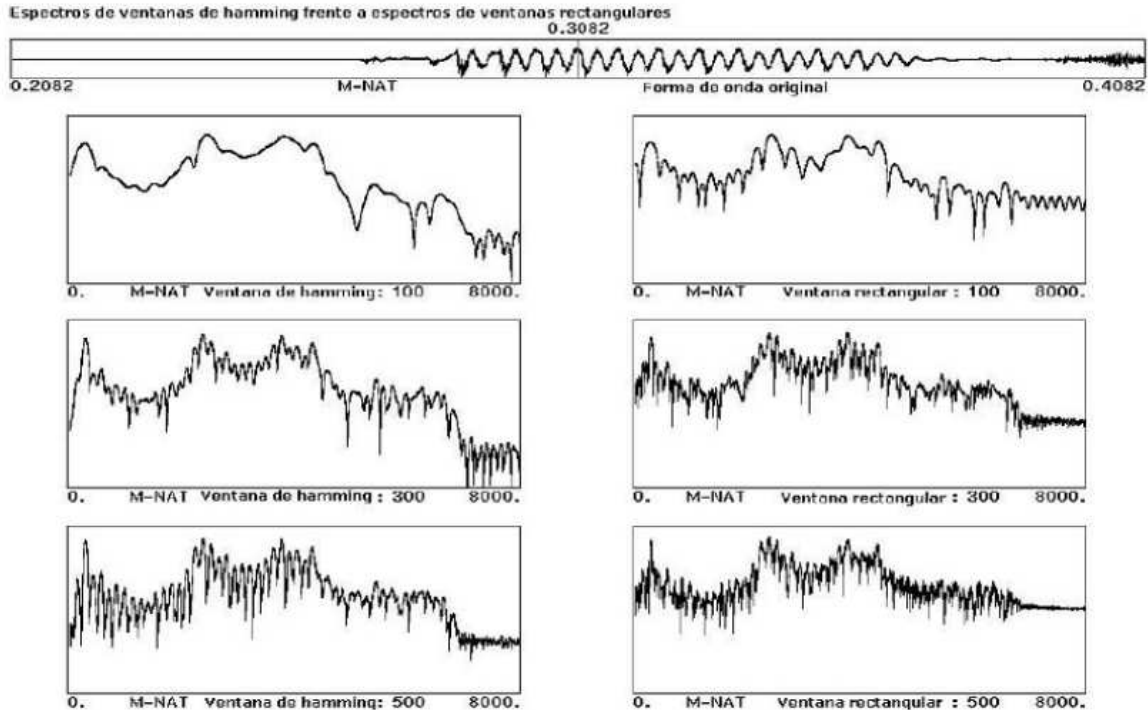
Cuando $\alpha = 0.46$ es conocida como ventana Hamming



Las ventanas rectangulares raramente son usadas para analizar segmentos de habla, pues si bien tienen alta resolución en el tiempo, producen efectos no deseables en las frecuencias obtenidas, mayormente son usadas las ventanas tipo Hamming o Hanning que producen un menor derramamiento espectral que es un efecto asociado a las ventanas rectangulares.

$$h[n] = .5 - .5 \cos 2\pi n/N (\text{Hanning})$$

$$h[n] = .54 - .46 \cos 2\pi n/N (\text{Hamming})$$



Comparación en el dominio de la frecuencia entre la Ventana Rectangular y la Ventana Hamming.

Transformada Corta de Fourier

Descompone la señal en una serie de segmentos y analiza los segmentos independientemente. Dada una señal $x[n]$ se define una señal corta en el tiempo $x_m[n]$ de un segmento m como sigue:

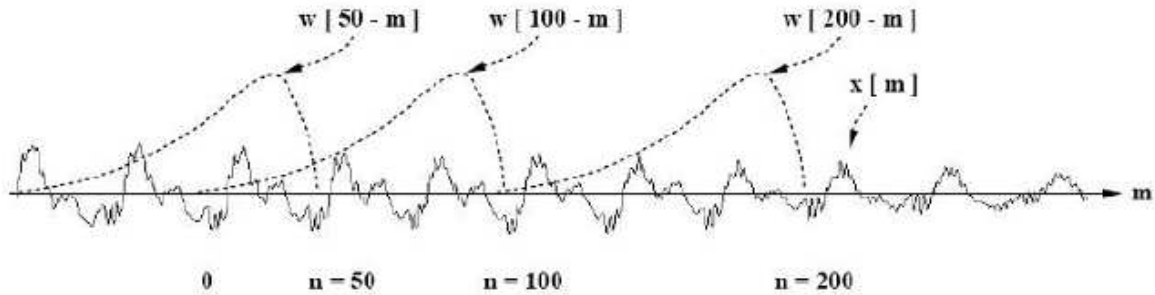
$$x_m[n] = x[n]w_m[n]$$

que es el producto de $x[n]$ por una función ventana $w_m[n]$, luego podremos hacer que la función tenga valores constantes para todos los segmentos:

$$w_m[n] = w[m - n]$$

Finalmente se tiene que la Transformada Corta de Fourier para un segmento m está definida como:

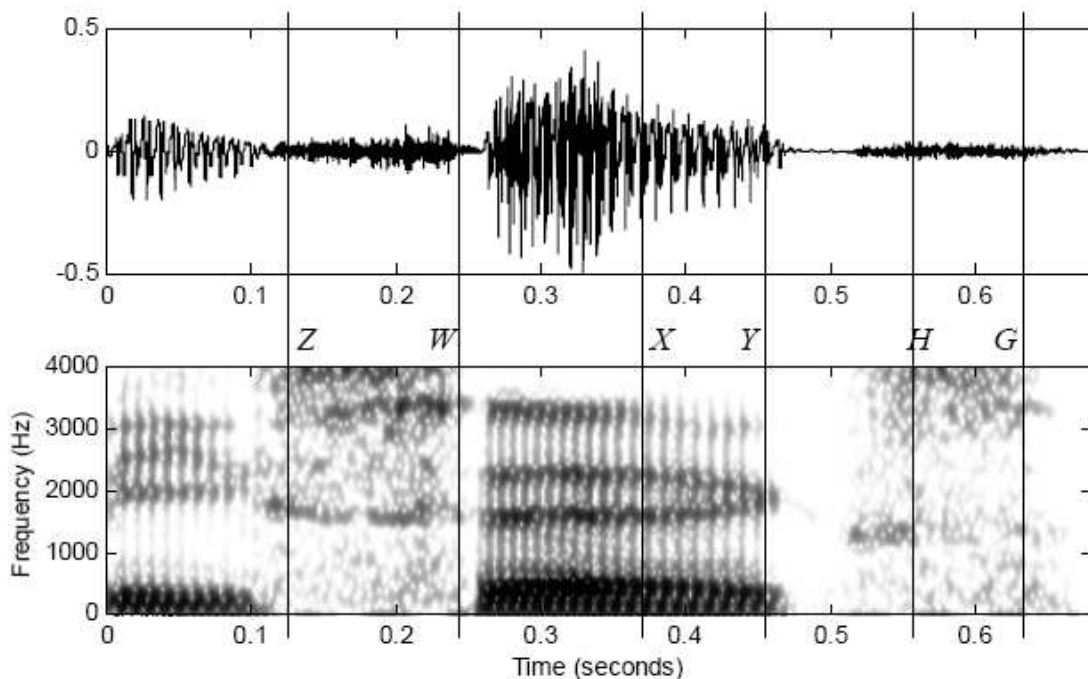
$$X_m(e^{j\omega}) = \sum x_m[n]e^{-j\omega n} = \sum w[m - n]x[n]e^{-j\omega n}$$



Si tenemos una señal de habla se recomienda tener tamaños de ventana entre 20 y 25 ms y espacios entre la aplicación de ventana de 10 ms, para valores mayores de 20 ms de espaciado entre ventana y ventana experimentos muestran que el oído humano aprecia distorsión en la señal

Espectogramas

Son útiles para analizar de manera gráfica las frecuencias de las señales, se grafican en el eje x el tiempo y en el eje y la frecuencia, un punto (x,y), indica la energía en cada punto, usualmente se usan escalas de colores o grises para representarla.



Entonces se puede decir que un espectrograma muestra la energía de los bandos de frecuencia, esta energía se calcula en ventanitas pequeñas (entre 5 y 25 milisegundos) de la siguiente manera

$$\log |X[k]|^2 = \log (X_r^2[k] + X_i^2[k])$$

Referencias bibliograficas

- [1] Huang, X., Acero, A., and Hon, H. 2001 *Spoken Language Processing: a Guide to Theory, Algorithm, and System Development*. 1st. Prentice Hall PTR.