

Reconocimiento automático del habla II

**Captura y
preprocesamiento de la
señal**

Captura de la señal

- En esta etapa tiene lugar la adquisición, la cuantificación y el muestreo de la señal de audio.
- `audiorecorder (Fs, num_bits, Ch):`
 - F_s: Frecuencia de muestreo
 - num_bits: nº de bits para almacenar cada muestra.
 - Ch: tipo de canal: mono (1) o estéreo (2).

Preprocesamiento (Preénfasis)

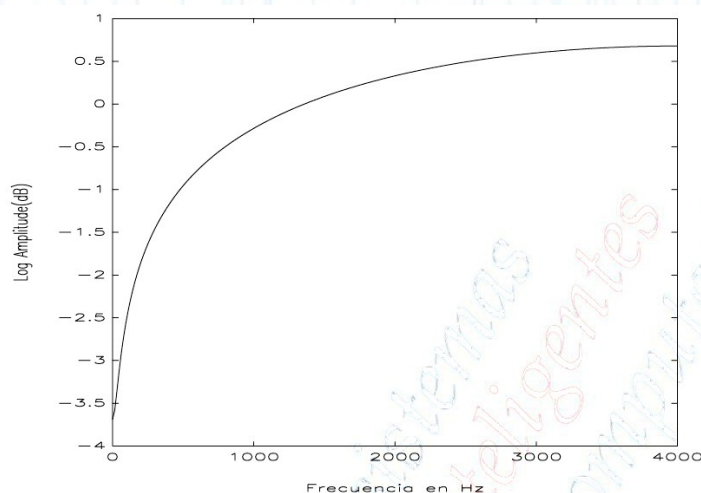
- La señal de voz se atenúa a medida que aumenta la frecuencia.
- Es necesario reducir el rango dinámico de las señales alisando el espectro incrementando la relevancia de las frecuencias altas con un filtro denominado *Filtro de Preénfasis*.

$$s'(n) = s(n) - a \cdot s(n-1)$$

donde a refleja el grado de preénfasis y suele estar en el rango de 0.9 a 1.0

Preprocesamiento (Preénfasis)

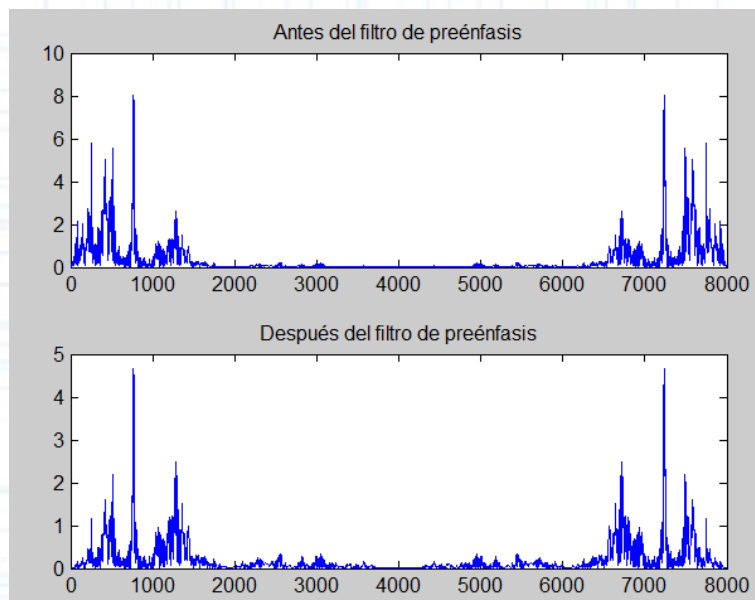
- Respuesta en frecuencia típica de un filtro de preénfasis:



$$a = 0.975$$

- Sólo se debería aplicar a las señales sonoras, pero se aplica a todas por:
 - ◆ Por la pequeña distorsión que se introduce en las señales aperiódicas.
 - ◆ Por simplificar el RAH.

Preprocesamiento (Preénfasis)

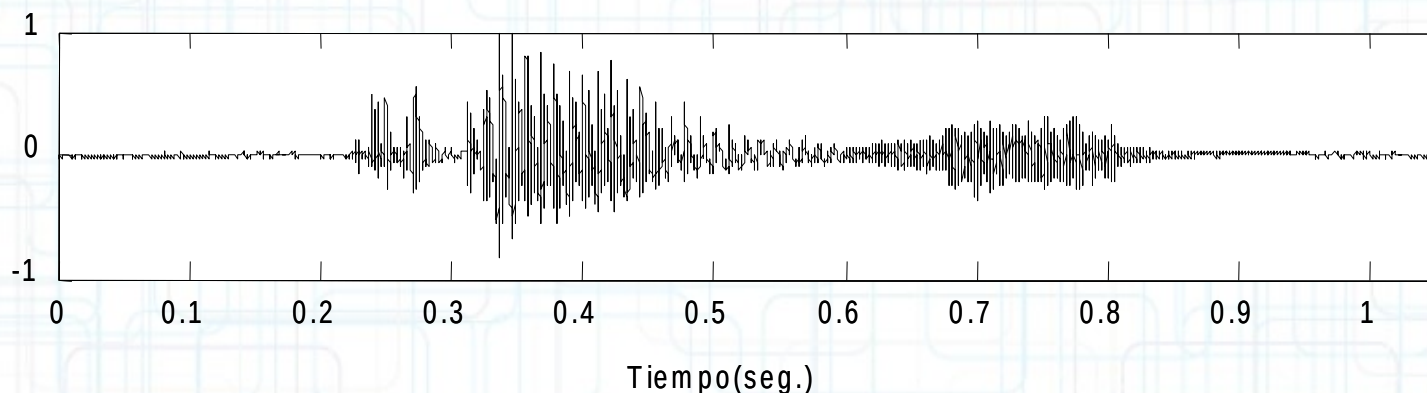


```
% Preénfasis  
a = 0.95;  
y = filter([1 -a], 1, y);
```

Preprocesamiento (Segmentación)

- Las propiedades de la señal de voz cambian con el tiempo.

Sonido: "Tres"



- Para poder realizar el procesamiento de la voz, se realiza la suposición de que las propiedades de la señal cambian relativamente despacio en el tiempo.

Preprocesamiento (Segmentación)

- A largo plazo (segundos) como a medio plazo (cientos de milisegundos) la señal de voz no es estacionaria.
- A corto plazo (decenas de milisegundos), la señal es quasi-estacionaria.

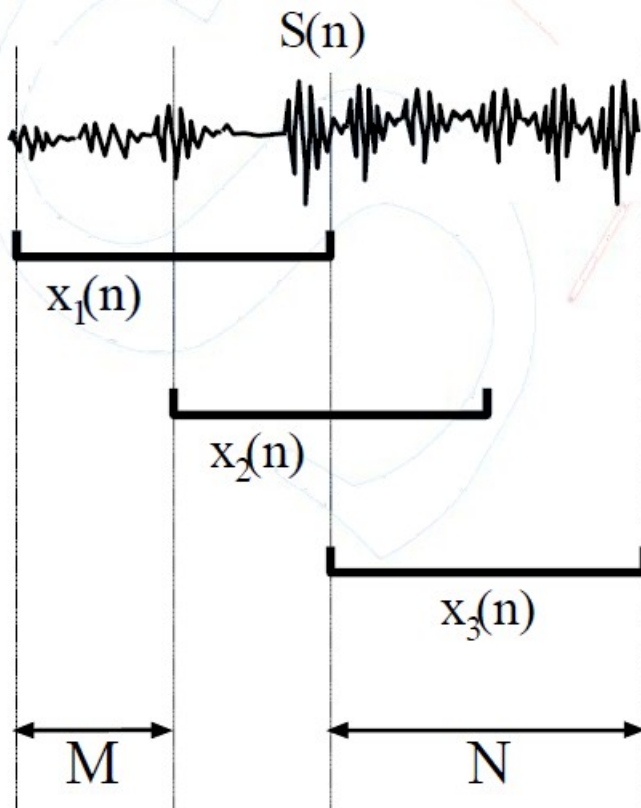
Una señal estacionaria es constante en sus parámetros estadísticos en el tiempo.

Preprocesamiento (Segmentación)

- Es necesario procesar la señal de voz en segmentos de corta duración: análisis localizado o en tiempo corto (short-time).

Preprocesamiento (Segmentación)

- Breves segmentos de la señal (tramas) se aíslan y procesan de forma independiente.



$$x_1(n) = \text{[segment 1 waveform]}$$

$$x_2(n) = \text{[segment 2 waveform]}$$

$$x_3(n) = \text{[segment 3 waveform]}$$

Preprocesamiento (Segmentación)

- Casos dependiendo de la longitud de la trama (N) y del desplazamiento (M):
 - ◆ $N < M$
 - ◆ No hay solapamiento entre tramas sucesivas.
 - ◆ Se pierde parte de la señal.
 - ◆ No se suele utilizar.

Preprocesamiento (Segmentación)

- Casos dependiendo de la longitud de la trama (N) y del desplazamiento (M):
 - ◆ $N = M$
 - ◆ No hay pérdida de señal.
 - ◆ No hay correlación en los valores espectrales de tramas consecutivas.
 - ◆ Desaconsejable.

Preprocesamiento (Segmentación)

- Casos dependiendo de la longitud de la trama (N) y del desplazamiento (M):
 - ◆ $N > M$
 - ◆ Las tramas adyacentes se solapan.
 - ◆ Los valores espectrales tendrán cierta correlación entre tramas consecutivas (si $N \gg M$, las variaciones entre valores espectrales son muy suaves).
 - ◆ Es el caso habitual.

Preprocesamiento (Segmentación)

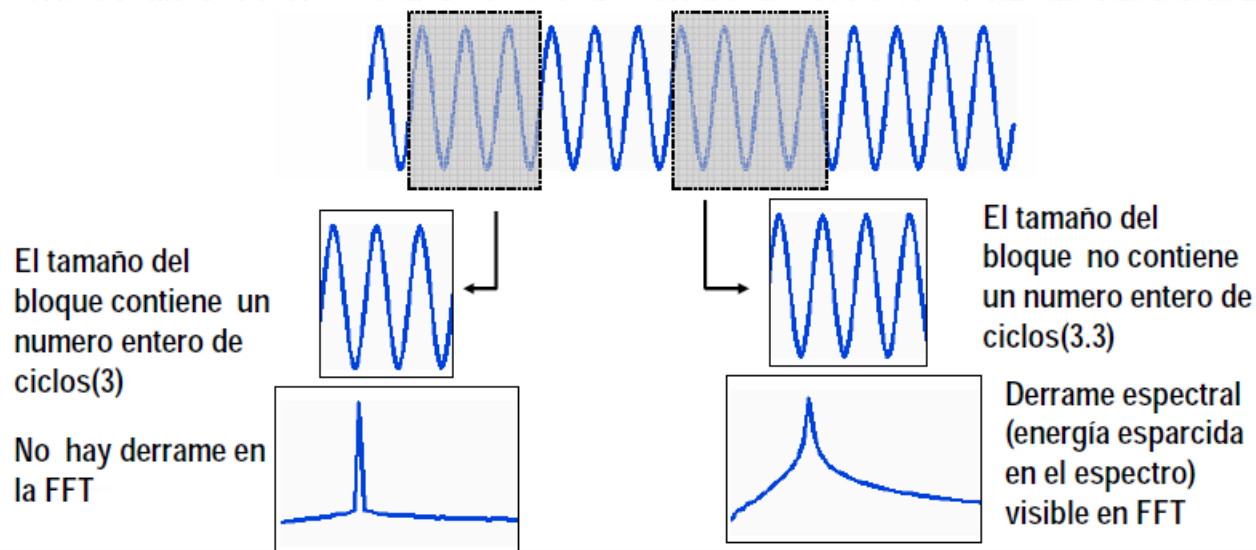
- Valor de N:
 - ◆ El valor óptimo sería el período de la señal (un valor menor produciría un análisis distorsionado, ya que la señal analizada estaría truncada).
 - ◆ Se suelen utilizar valores de N que incluyan varios períodos de señal (los menos posibles).

Preprocesamiento (Segmentación)

- Valor de N:
 - ◆ Problema: El período es variable entre diferentes locutores e incluso variable en el tiempo para un locutor.
 - ◆ Los valores del período de una señal varían desde los 40 Hz (hombre) a los 100 Hz (mujer) \Rightarrow la longitud de la trama estaría entre 10 y 25 msec.

Preprocesamiento (Segmentación)

- Problemas de la segmentación:
 - ◆ Produce el derrame espectral (*spectral leakage*).

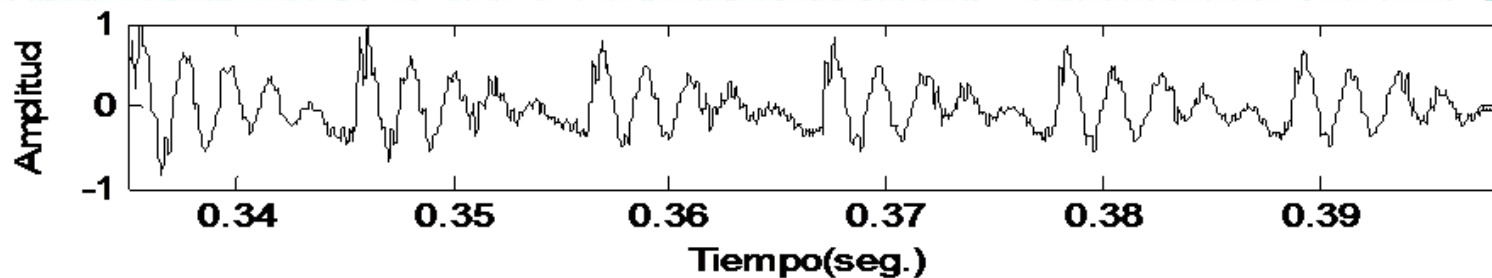


Preprocesamiento (Segmentación)

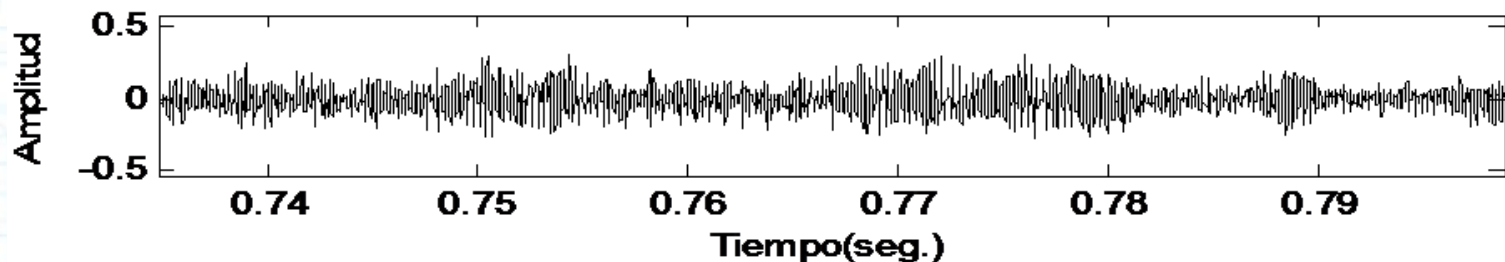
- Valor de M:
 - ◆ $M = K * N$, donde $K = 1/2, 1/3, 1/4, \dots$
 - ◆ Menor desplazamiento \Rightarrow más suaves las fluctuaciones de los parámetros obtenidos en la fase de *Extracción de características*.

Preprocesamiento (Segmentación)

- Tipos de tramos:
 - ◆ Pseudo-periódicos.



- ◆ Ruidosos.



Preprocesamiento (Segmentación)

`BUFFER (señal, tam_seg, tam_solap, opciones)`

Particiona la señal de entrada en segmentos de datos:

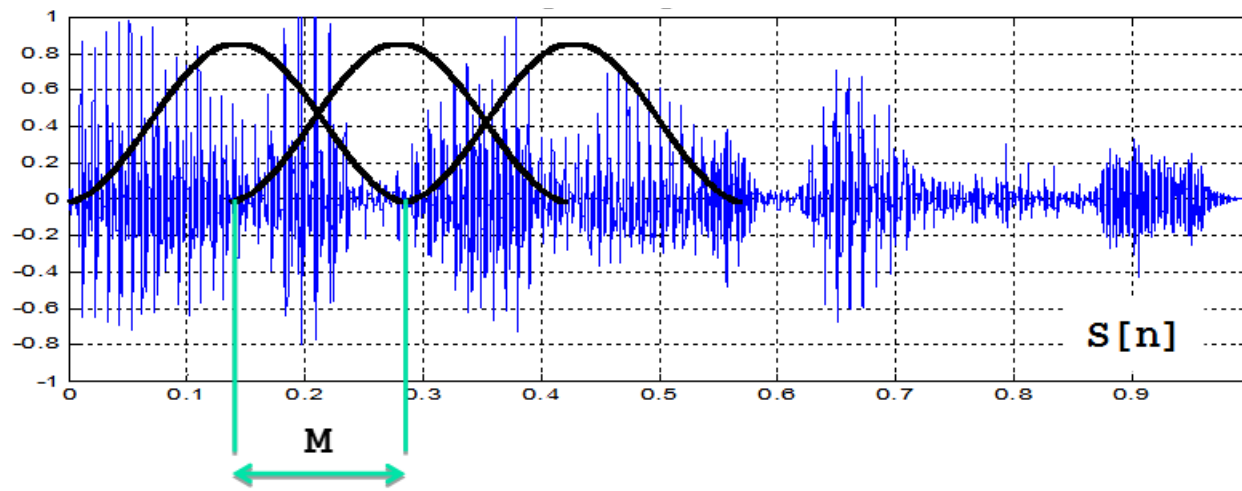
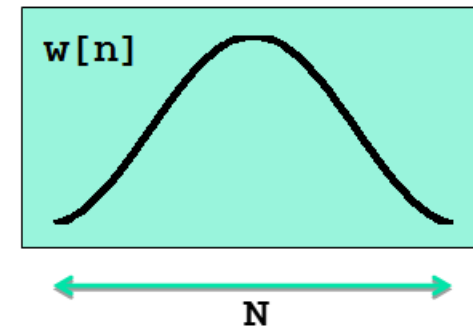
- `señal`: señal a segmentar.
- `tam_seg`: longitud de los segmentos.
- `tam_solap`: nº de muestras del final del segmento anterior que se repite en el siguiente.
- `opciones`: Probar la diferencia entre omitir esta opción o darle el valor 'nodelay'.

Preprocesamiento (Enventanado)

- Su objetivo es minimizar las discontinuidades que la señal pueda tener al comienzo y al final de cada trama.
- La utilización de una ventana rectangular (es decir, segmentación sin enventanado) hace que en los extremos de dicha ventana la función decaiga rápidamente y se produce el efecto Gibbs.

Preprocesamiento (Enventanado)

- $S[n]$: Señal de voz
- $w[n]$: Ventana de análisis
 - N : Tamaño de la ventana
 - M : Desplazamiento

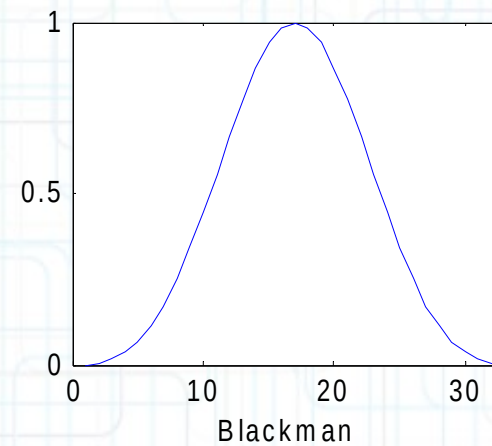
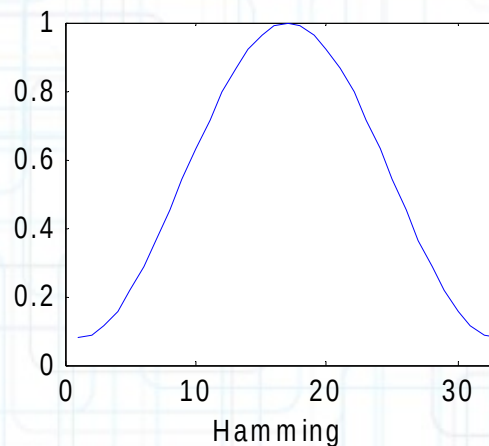
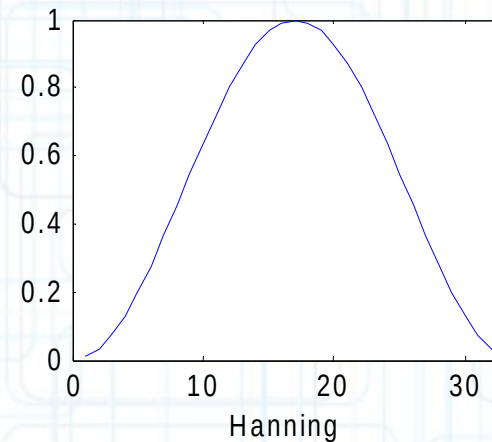
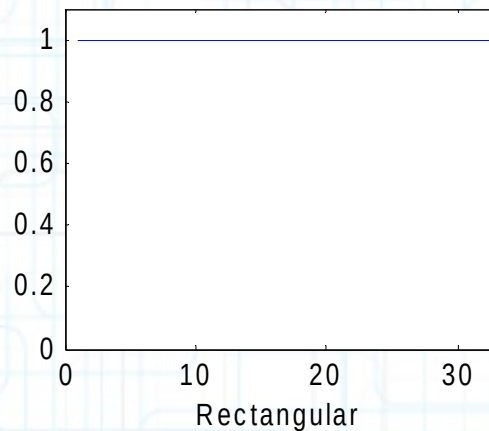


Preprocesamiento (Enventanado)

- $x[n] = x[n] w[n]$ $n = 1, 2, \dots, N$
- La ventana $w[n]$ suele tener un valor próximo a cero en los extremos y es simétrica respecto al centro de la misma.
- Efectos del enventanado:
 - ◆ Atenuar de forma gradual la señal en los extremos de la trama.
 - ◆ Una convolución de los espectros de la ventana y la señal.

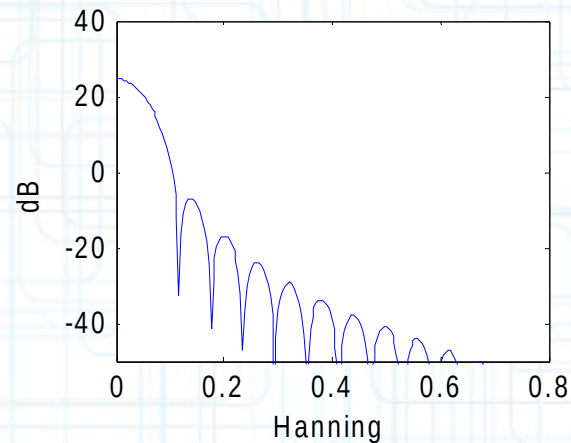
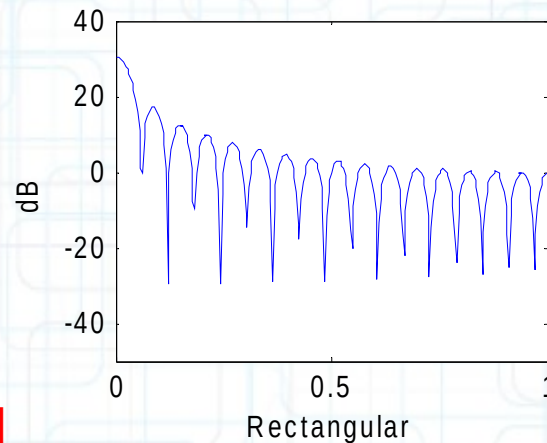
Preprocesamiento (Enventanado)

- Algunos ejemplos de ventanas:

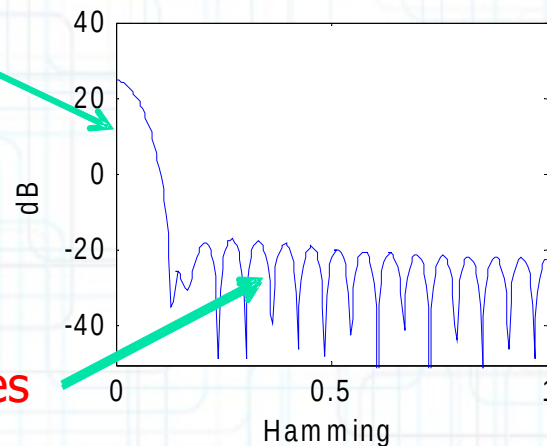
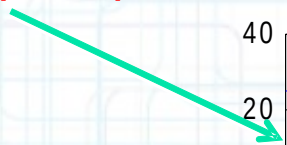


Preprocesamiento (Enventanado)

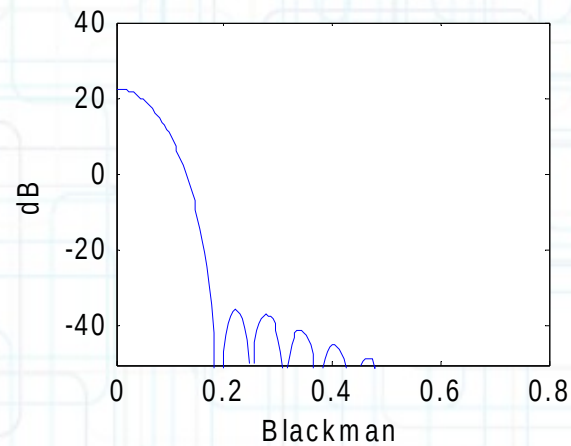
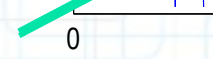
- Espectros de las ventanas:



Lóbulo principal



Lóbulos laterales



Preprocesamiento (Enventanado)

- Espectros de las ventanas (código MATLAB):
¿cómo afecta el n^o de puntos usados en la fft?

```
num_muestras = 51;

ventana = hamming(num_muestras);

fftlongitud_maxima = 1024;

fftlongitudes = round(linspace(num_muestras,fftlongitud_maxima,...
    round(fftlongitud_maxima/num_muestras)));

for fftlongitud = fftlongitudes,
    amplitudes = abs(fft(ventana, fftlongitud));
    subplot(2,1,1);
    plot(linspace(0,0.5,ceil(fftlongitud/2)), amplitudes(1:ceil(fftlongitud/2)));
    ylabel('Amplitud');
    xlabel('Frecuencia normalizada');
    legend ([num2str(fftlongitud), ' puntos']);
    subplot(2,1,2);
    plot(linspace(0,0.5,ceil(fftlongitud/2)), 20*log10(amplitudes(1:ceil(fftlongitud/
2)))));
    ylabel('Amplitud (dB)');
    xlabel('Frecuencia normalizada');
    legend ([num2str(fftlongitud), ' puntos']);
    pause(1);
end
```


Preprocesamiento (Enventanado)

- Problemas del enventanado:
 - ◆ El lóbulo principal dificulta la identificación de frecuencias cercanas entre sí.
 - ◆ Los lóbulos laterales introducen señal en frecuencias donde no debería haber nada.

Preprocesamiento (Enventanado)

- La ventana debe satisfacer dos características para reducir la distorsión espectral producida por el enventanado:
 - ◆ Lóbulo principal estrecho y agudo.
 - ◆ Gran atenuación de los lóbulos secundarios.
- Son características contrapuestas \Rightarrow hay que buscar un compromiso entre ambas \Rightarrow se suelen minimizar los lóbulos laterales.

Preprocesamiento (Enventanado)

- Ventanas (MATLAB):

- ◆ rectwin(N)
- ◆ hamming(N)
- ◆ hanning(N)
- ◆ bartlett(N)
- ◆ blackman(N)
- ◆ boxcar(N)
- ◆ triang(N)
- ◆ gausswin(N)
- ◆ blackmanharris(N)
- ◆ kaiser(N, BETA)
- ◆ chebwin(N, R)