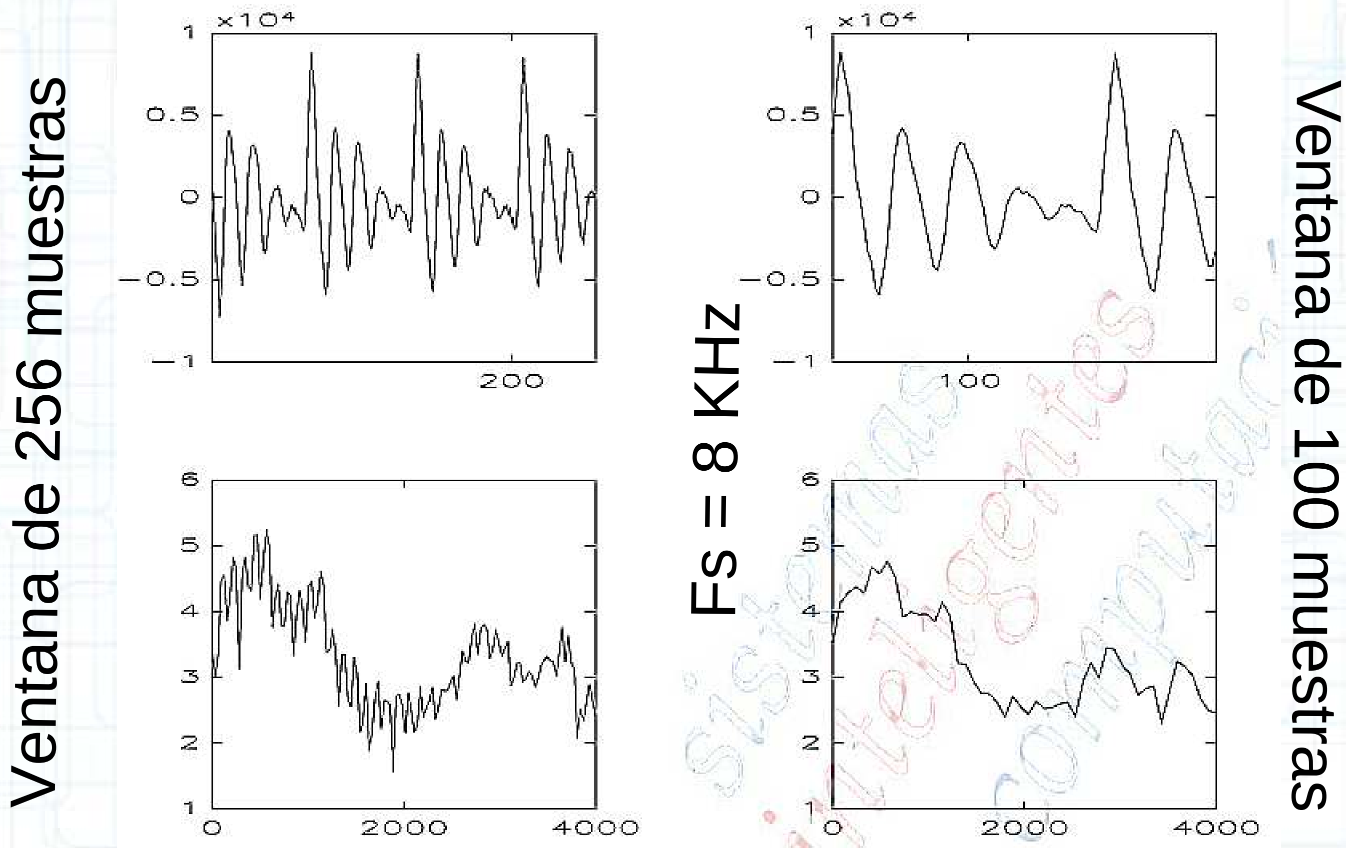
Reconocimiento automático del habla III

Extracción de características (dominio de la frecuencia)

Análisis localizado en frecuencia

- Calcular la transformada de Fourier de una trama de señal enventanada.



Análisis localizado en frecuencia

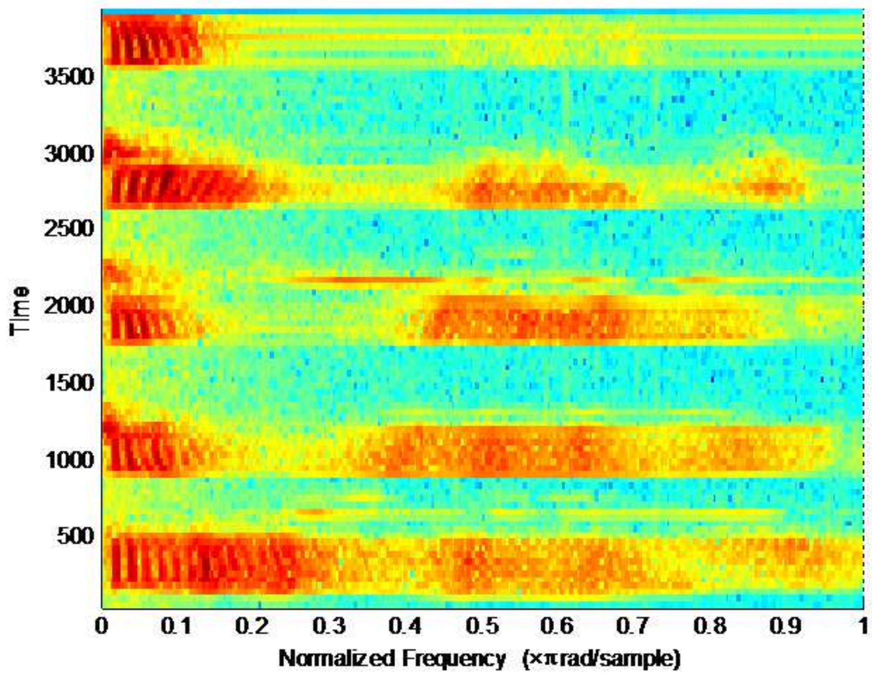
- Calcular la transformada de Fourier de una trama de señal enventanada:
  - Si el número de muestras es grande respecto al pitch, en el espectro se puede analizar muy bien los armónicos correspondientes al pitch.
  - Si el número de muestras es pequeño respecto al pitch, la señal tiene poca resolución en frecuencia, pero la envolvente espectral es muy limpia.

Espectrogramas (sonogramas)

- Representan la evolución del espectro en el tiempo.
- Ambas variables son inversas (ganar resolución en una de ellas implica perderla en la otra).

Espectrogramas (sonogramas)

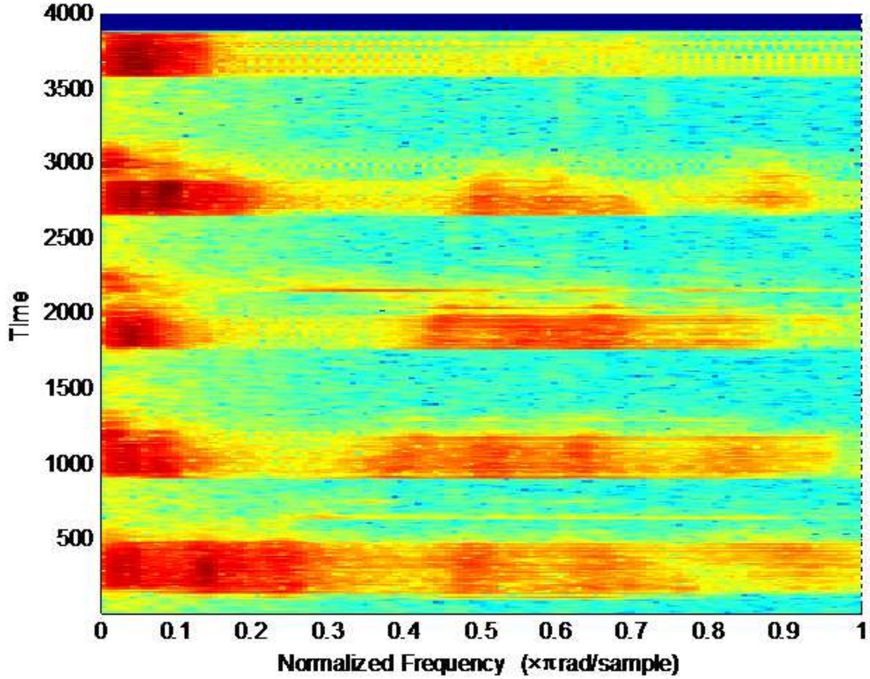
- Tipos de espectrogramas:
  - Banda estrecha:
    - poca resolución temporal (ventanas temporales grandes)
    - alta resolución en frecuencia



Permite apreciar los detalles de los componentes (por ejemplo, los armónicos en una señal periódica).

Espectrogramas (sonogramas)

- Tipos de espectrogramas:
  - Banda ancha:
    - buena resolución temporal (ventanas temporales pequeñas)
    - poca resolución en frecuencia



Permite apreciar mejor las relativamente amplias zonas de frecuencias que tienen mayor energía.

Análisis de predicción lineal

- Queremos realizar una estimación del valor de una muestra en el instante n como una combinación lineal de k muestras anteriores:

$$\tilde{x}(n) = \sum_{i=1}^{k} a_i * x(n-i)$$

Análisis de predicción lineal

- El conjunto de coeficientes óptimo será aquel que haga que el error cuadrático sea mínimo:

$$L = \sum_n e^2(n) = \sum_n [x(n) - \tilde{x}(n)]^2 = \sum_n \left[x(n) - \sum_{i=1}^{k} a_i * x(n-i) \right]^2$$

Análisis de predicción lineal

- Para obtener el valor mínimo de L, calculamos las derivadas parciales respecto a cada una de las variables y se igual a cero :

$$\frac{dL}{da_j} = \frac{d\sum_n \left[x(n) - \sum_{i=1}^{k} a_i * x(n-i)\right]^2}{da_j} = 0$$

$$\frac{dL}{da_j} = 2\sum_n \left(x(n) - \sum_{i=1}^{k} a_i * x(n-i)\right) * (0 - x(n-j)) = 0$$

$$\frac{dL}{da_j} = \sum_n \left(x(n) - \sum_{i=1}^{k} a_i * x(n-i)\right) * (x(n-j)) = 0$$

para $1 \leq j \leq k$

Análisis de predicción lineal

- Para obtener el valor mínimo de L, calculamos las derivadas parciales respecto a cada una de las variables y se igual a cero:

$$\frac{dL}{da_j} = \frac{d\sum_n \left[x(n) - \sum_{i=1}^{k} a_i * x(n-i)\right]^2}{da_j} = 0$$

$$\frac{dL}{da_j} = 2\sum_n \left(x(n) - \sum_{i=1}^{k} a_i * x(n-i)\right) * \left(0 - x(n-j)\right) = 0$$

$$\frac{dL}{da_j} = \sum_n \left(x(n) - \sum_{i=1}^{k} a_i * x(n-i)\right) * \left(x(n-j)\right) = 0$$

$$\text{para } 1 \leq j \leq k$$

Análisis de predicción lineal

- Desarrollamos la ecuación:

$$\sum_{n} x(n-j) * x(n) - \sum_{i=1}^{k} a_i * \sum_{n} x(n-j) * x(n-i)$$

$$C_{ij} = \sum_{n} x(n-j) * x(n-i)$$

$$C_{j0} - \sum_{i=1}^{k} a_i * C_{ji}$$

Análisis de predicción lineal

- Método de autocorrelación:

$$C_{ij} = C_{ji} = r_{|i-j|} = \text{coeficientes de correlación}$$

$$\sum_n x(n-j) * x(n-i) = \sum_n x(n) * x(n+|i-j|) = r_{|i-j|}$$

$$C_{j0} - \sum_{i=1}^{k} a_i * C_{ji} \Rightarrow \sum_{i=1}^{k} r(|j-i|) * a_i = r(j), \quad 1 \leq j \leq k$$

Análisis de predicción lineal

- Método de autocorrelación:

$$\sum_{i=1}^{k} r(|j-i|) * a_i = r(j), \quad 1 \leq j \leq k$$

Forma matricial

$$\begin{bmatrix} r_n(0) & r_n(1) & r_n(2) & \cdots & r_n(k-1) \\ r_n(1) & r_n(0) & r_n(1) & \cdots & r_n(k-2) \\ r_n(2) & r_n(1) & r_n(0) & \cdots & r_n(k-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_n(k-1) & r_n(k-2) & r_n(k-3) & \cdots & r_n(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} r_n(1) \\ r_n(2) \\ r_n(3) \\ \vdots \\ r_n(k) \end{bmatrix}$$

Análisis de predicción lineal (MATLAB)

- `coeficientes = lpc (señal, p)`
  - Calcula los coeficientes de predicción lineal minimizando el error cuadrático medio.
  - El parámetro *p* es el número de coeficientes LPC.
  - La ecuación de diferencias del sistema es:

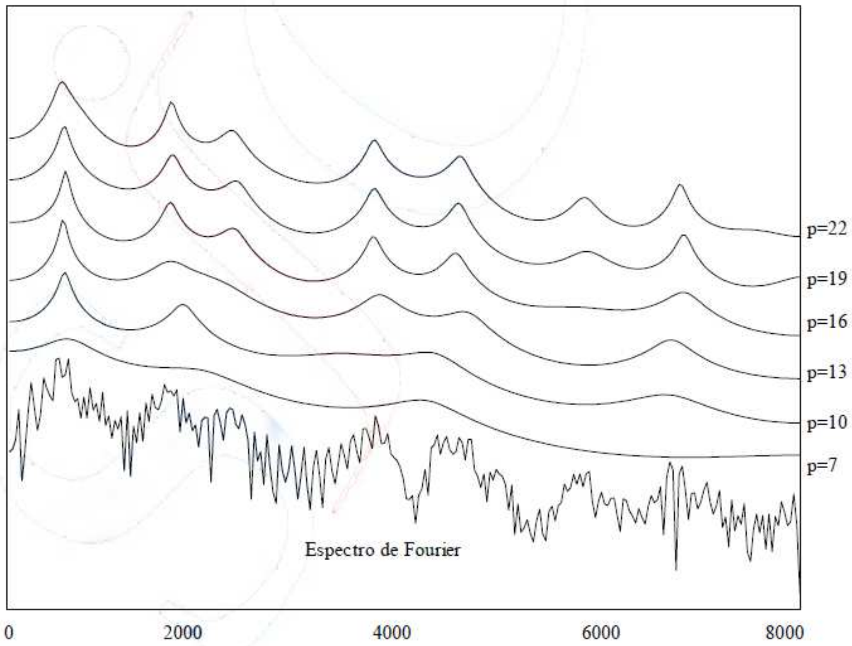
$$\hat{x}(n) = -a(2)x(n-1) - a(3)x(n-2) - \cdots - a(p+1)x(n-p)$$

Análisis de predicción lineal (MATLAB)

```
Fs = 100;
n= 0:500;

A1 = 5;
F1 = 2;
Fase1 = 0;
A2 = 10;
F2 = 4;
Fase2 = 0;

orden = 10;

y = A1 * sin(2*pi*F1/Fs*n + Fase1) + A2 *
sin(2*pi*F2/Fs*n + Fase2);

subplot (3,1,1);
plot(n/Fs,y);
xlabel('Función');

% cálculo de los coeficientes
lpcoefs = lpc(y,orden);
% Estimación de la señal
estsenal = filter([0 -lpcoefs(2:end)],1,y);
subplot(3,1,2);
plot(n/Fs,estsenal);
xlabel('Estimación de la senal');
% Error
error = y - estsenal;
subplot(3,1,3);
plot(n/Fs,error);
xlabel('Error de estimación');
```

Análisis de predicción lineal

- El número de coeficientes determina la resolución con la que el análisis de predicción lineal va a representar la envolvente espectral de la señal.



Un número reducido de coeficientes implica poca resolución.
Un valor excesivo implica cierta distorsión debido a que se considera la estructura fina del espectro, no sólo su envolvente.

En general, se consideran entre 10 y 14 coeficientes.

Análisis de predicción lineal

- Para calcular la envolvente espectral de la señal, obtenemos la función de transferencia del filtro LPC:

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} a_k \cdot z^{-k}}$$

- donde:
  - G es la ganancia del filtro y se calcula como: `sqrt(sum(error .^ 2))`
  - El denominador es el vector devuelto por la función *lpc*.

Análisis de predicción lineal

- Y calculamos su respuesta en frecuencia con la función freqz de MATLAB:

```
H = freqz(G,lpcoefs,puntos)
```

Análisis de predicción lineal (MATLAB)

Muestra las envolventes de las tramas de una vocal

```
% Capturar el sonido de la vocal durante un segundo
Fs = 8000;
tiempo = 1;
canales = 1;
tipo_dato = 'double';
senal = wavrecord(tiempo*Fs,Fs,canales,tipo_dato);

% Detectar el inicio y fin de la palabra
long_ventana_tiempo = 0.03; % segundos
long_ventana_muestras = long_ventana_tiempo * Fs;
desplazamiento = round(0.5 * long_ventana_muestras);
num_ventanas_ruido = 10;

[senal] = deteccion_inicio_fin_palabra (senal, long_ventana_muestras,
                     desplazamiento, num_ventanas_ruido);

% Preénfasis
a = 0.98;
senal = preenfasis(senal,a);

% Segmentación y enventanado
ventana = 'hamming';
tramas = segmentacion(senal,long_ventana_muestras, desplazamiento);
tramas_enventanadas = enventanado(tramas,ventana);

% Cálculo de los coeficientes de predicción lineal
num_puntos = 1024;
transFourier = abs(fft(tramas_enventanadas,num_puntos));

%orden = [4 10 14 16 28];
orden = [10 16];
```

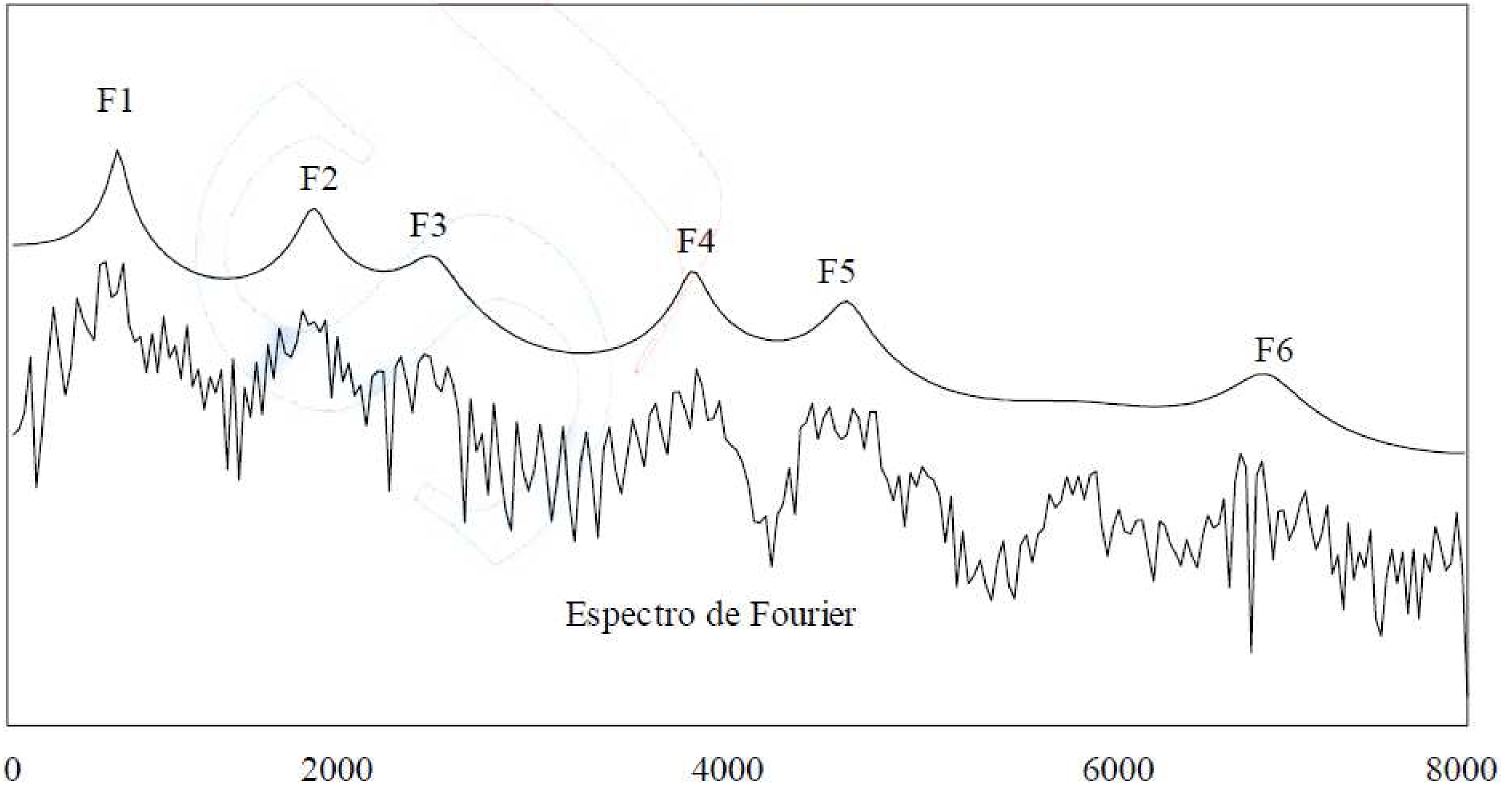
Análisis de predicción lineal (MATLAB)

```
for i=1:size(tramas_enventanadas,2),
       subplot(length(orden)+1,1,1);
       plot(tramas_enventanadas(:,i));
   for j=1:length(orden),
       subplot(length(orden)+1,1,j+1);
       plot(linspace(0,Fs/2,round(num_puntos/2)+1),
             20*log10(transFourier(1:round(num_puntos/2)+1,i)));
       % Cálculo de los coeficientes
       lpcoefs = lpc(tramas_enventanadas(:,i),orden(j));
       % Estimación de la señal
       estsenal = filter([0 -lpcoefs(2:end)],1,
                        [tramas_enventanadas(:,i);zeros(orden(j),1)]);
       % Error
       error = [tramas_enventanadas(:,i);zeros(orden(j),1)] - estsenal;

       % Ganancia del filtro LPC
       G = sqrt(sum(error .^ 2));

       % Respuesta en frecuencia del sistema
       H = freqz(G,lpcoefs,round(num_puntos/2)+1);
       hold on;
       plot(linspace(0,Fs/2,round(num_puntos/2)+1),20*log10(abs(H)),'r');
       legend([num2str(orden(j)),' coefs']);
   end
   pause;
   close;
end
```

Los formantes

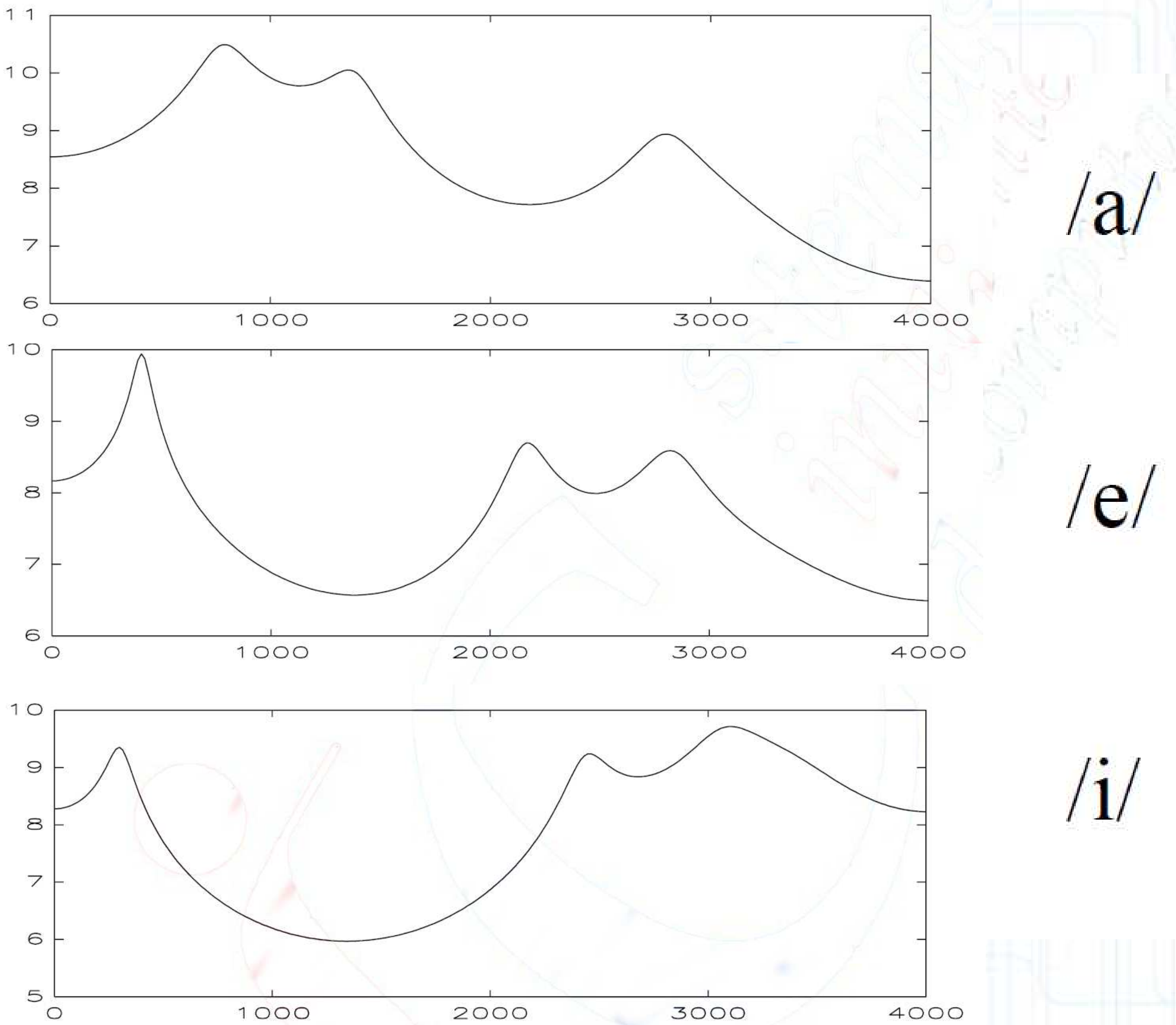
- Corresponden (aproximadamente) con los máximos de la envolvente espectral.
- Son zonas de resonancia en las que se pone de relieve un conjunto determinado de armónicos.

Los formantes

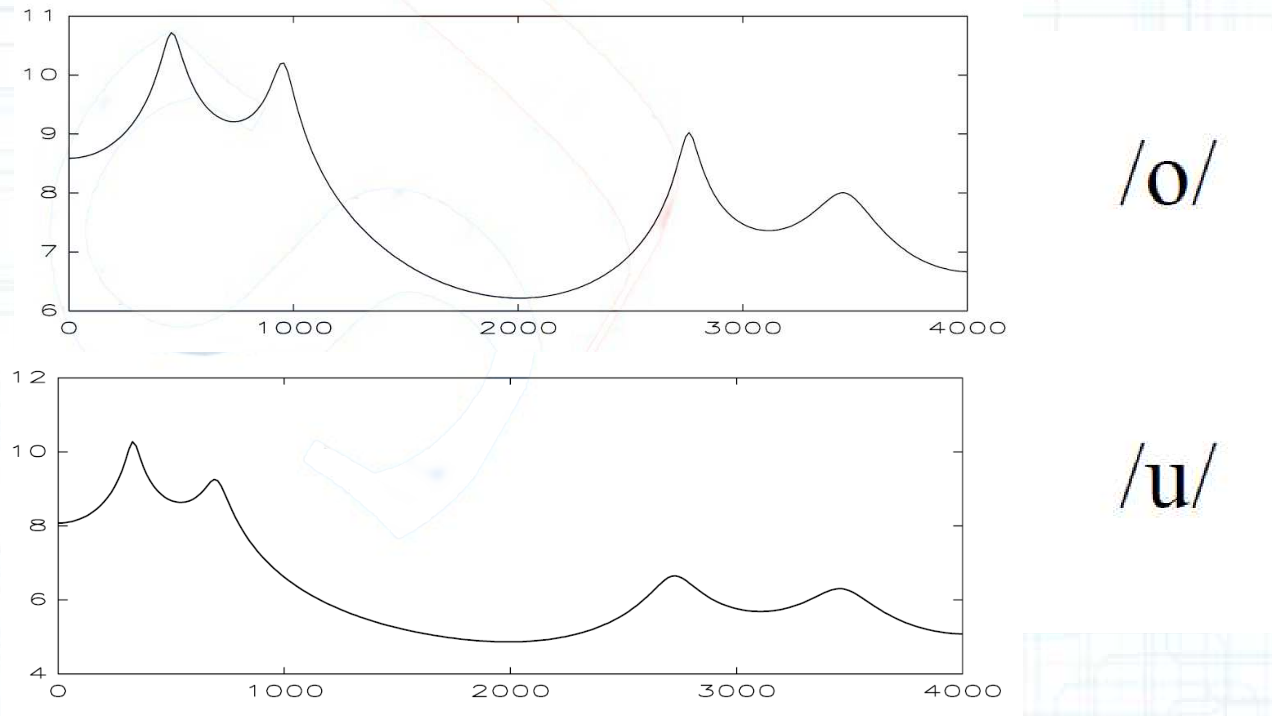


- F1: Relacionado con la abertura del canal bucal (más grande la apertura, más alta la frecuencia de F1)
- F2: Se modifica por la posición de la lengua (más elevada se halle, mayor será) y la posición de los labios (más redondeados y abocinados, más bajo será).
- F3: Se eleva su frecuencia cuando se produce un descenso del velo del paladar.

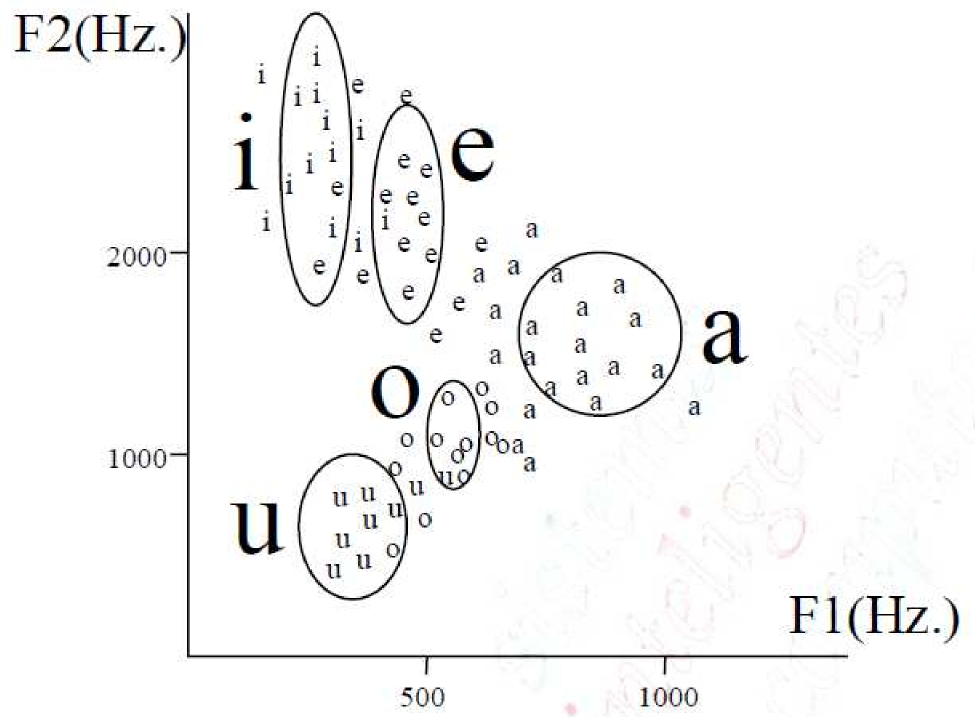
Los formantes (discriminación de las vocales)



Los formantes (discriminación de las vocales)



Los formantes (discriminación de las vocales)



Con sólo F1 y F2 podemos diferenciar las vocales del alfabeto.

Análisis cepstral

- El *cepstrum* o coeficiente cepstral se define como la transformada inversa de Fourier del logaritmo del módulo espectral.
- El término proviene de la inversión de la primera parte de la palabra *spectrum* (espectro).
- La variable independiente en el dominio cepstral se denomina *quefrency*.

Análisis cepstral

- La característica principal del análisis cepstral es que permite separar la estructura fina del espectro y la envolvente espectral.
- Permite separar la señal de excitación de la respuesta del filtro del tracto vocal.

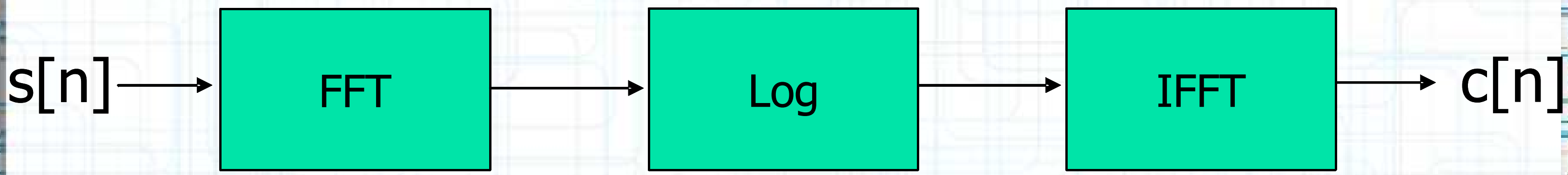
Análisis cepstral

- Un segmento sonoro es la convolución entre:
  - La señal de excitación glotal e[n]
  - El filtro del tracto vocal h[n]
  - $s[n] = e[n] * h[n]$
- La convolución en el dominio del tiempo corresponde a un producto en el dominio de la frecuencia:
  - $S[k] = E[k] H[k]$

Análisis cepstral

- Si aplicamos logaritmos sobre el módulo:
  - $\log(|S[k]|) = \log(|E[k]|) + \log(|H[k]|)$
- Calculamos ahora la transformada de fourier inversa y obtenemos el cepstrum (en el dominio cepstral):
  - $IDFT(\log(|S[k]|)) = IDFT(\log(|E[k]|)) + IDFT(\log(|H[k]|))$

Análisis cepstral

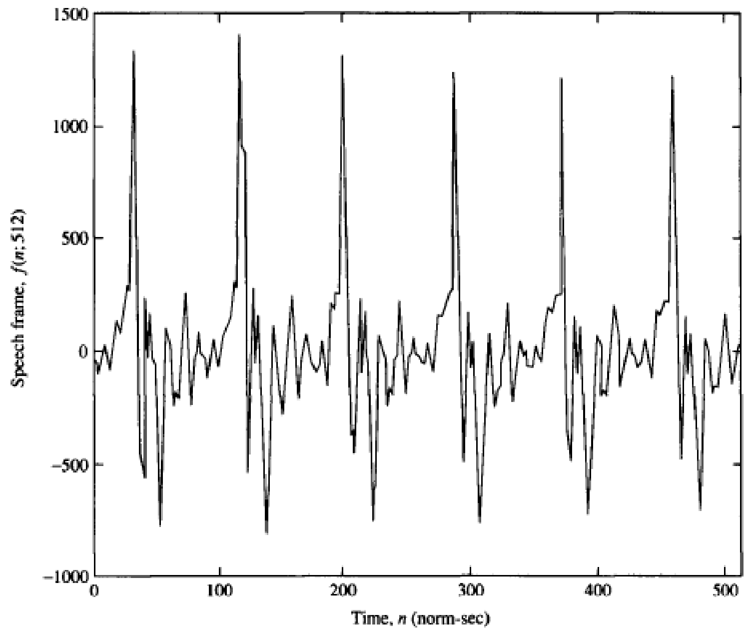


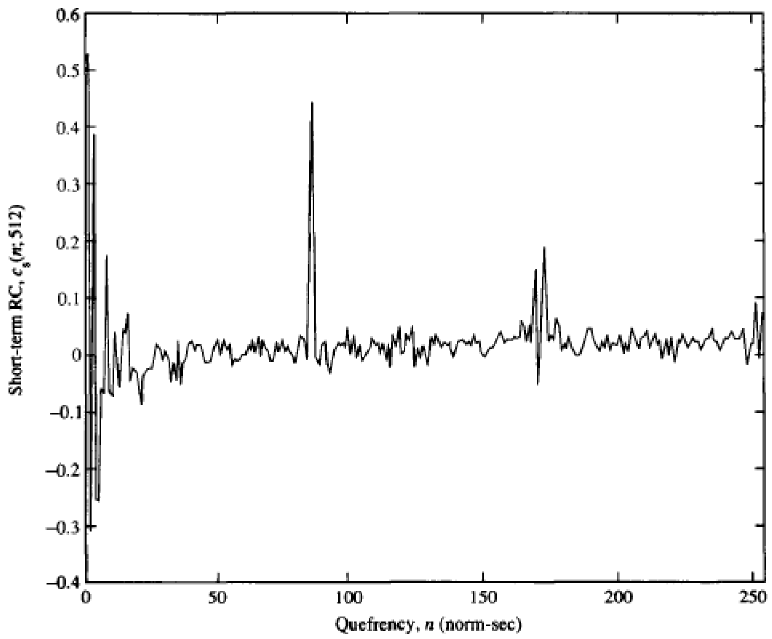
Análisis cepstral

- En el dominio cepstral:
  - Las componentes de la estructura fina y la envolvente espectral aparecen como sumandos.
  - Las componentes debidas a la estructura armónica aparecen como picos equiespaciados de altas *quefrencies* (separados por el periodo de un *pitch*).
  - La respuesta del tracto vocal aparece en baja *quefrencies* como señal impulsiva que abarca los primeros coeficientes cepstrales.

Análisis cepstral

- Trama sonora y su cepstrum:

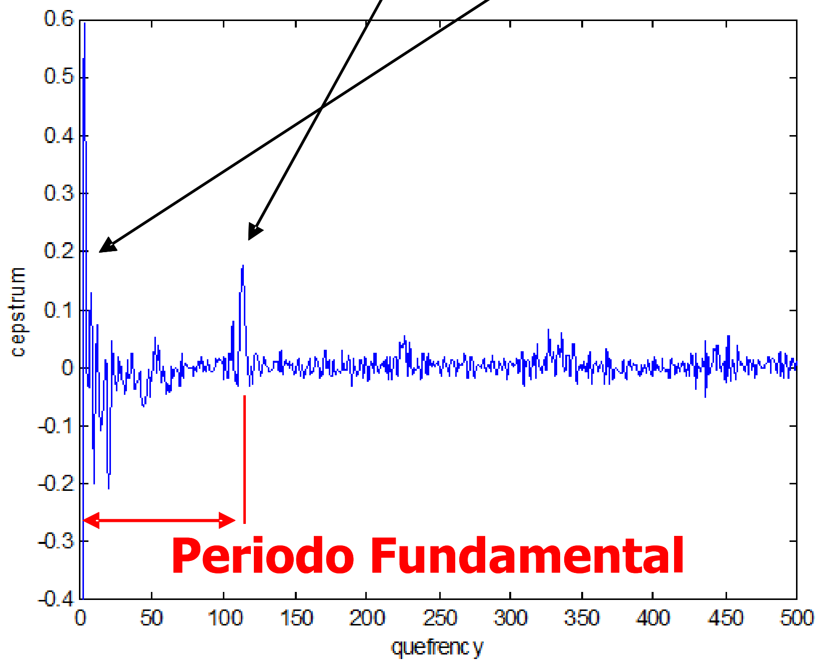




Análisis cepstral

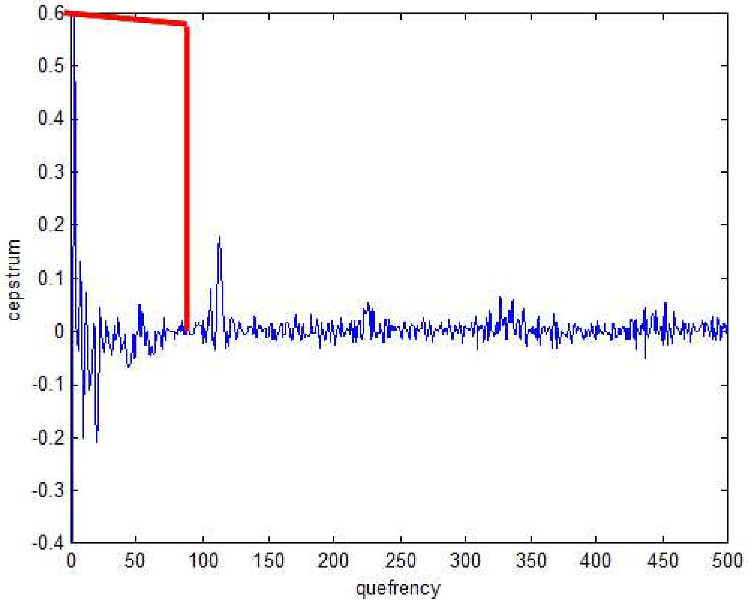
$$s[n] = e[n] * h[n] \longrightarrow c[n] = c_e[n] + c_h[n]$$

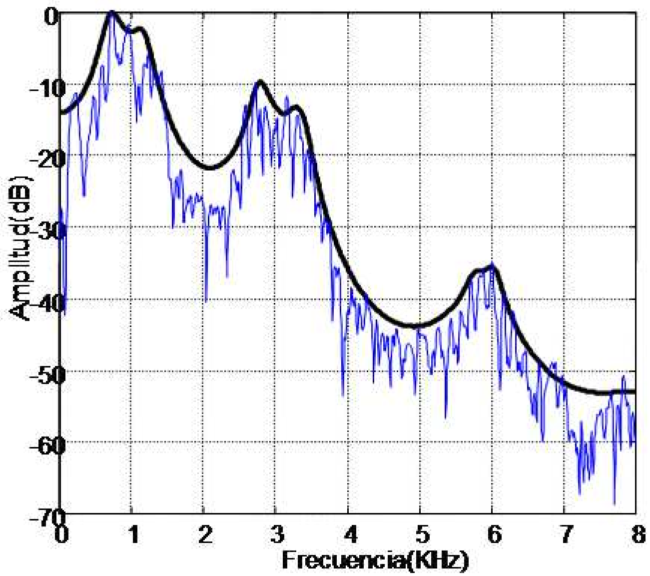
c_e y c_h son separables



Análisis cepstral

- Con un filtrado paso bajo seleccionamos los primeros coeficientes cepstrales, calculamos la transformada de Fourier y tendremos la estructura de la envolvente espectral.





- Si nos quedamos con los coeficientes altos, tendremos una estimación precisa del *pitch*.

Análisis cepstral

- El cepstrum puede ser:
  - Complejo:
    - Se aplican logaritmos del espectro completo.
    - Permite la reconstrucción de una señal, ya que contiene información sobre la magnitud y la fase.
  - Real:
    - Se aplica logaritmo al módulo (la magnitud) del espectro.
    - No se puede reconstruir la señal al no contener información de fase.
- En reconocimiento de voz se suele utilizar el cepstrum real.

Análisis cepstral (MATLAB)

- cceps: Cepstrum complejo.
- icceps: Inversa del cepstrum complejo.
- **rceps: Cepstrum real.**

```
rcepstrum = real(ifft(log(abs(fft(senal)))));
```

Coeficientes cepstrales con frecuencia en escala de Mel (MFCC)

- La escala de frecuencias MEL es una escala de frecuencias de distribución no lineal que responde al mecanismo de percepción auditiva.
- Se trata de obtener el mejor compromiso entre la resolución frecuencia-tiempo:
  - Usa un ancho de banda pequeño en baja frecuencia (permite resolver armónicos).
  - Usa un ancho de banda más grande en alta frecuencia (permite buena resolución en ráfagas temporales)

Coeficientes cepstrales con frecuencia en escala de Mel (MFCC)

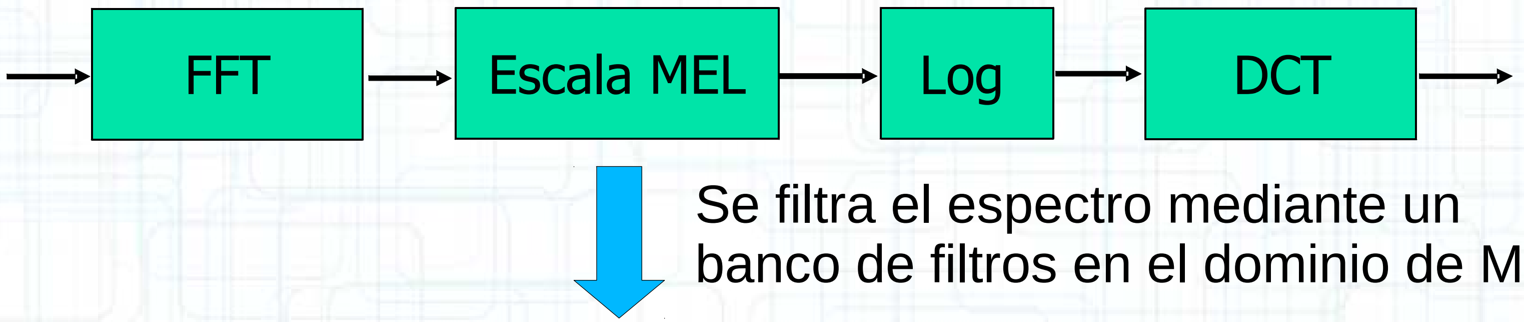
- En esta escala, la frecuencia se mide en MELs.
- La conversión de Hz a MELs viene dada por la fórmula:

$$m = 2595 \,.\, \log(1 + f/700) \qquad ó$$

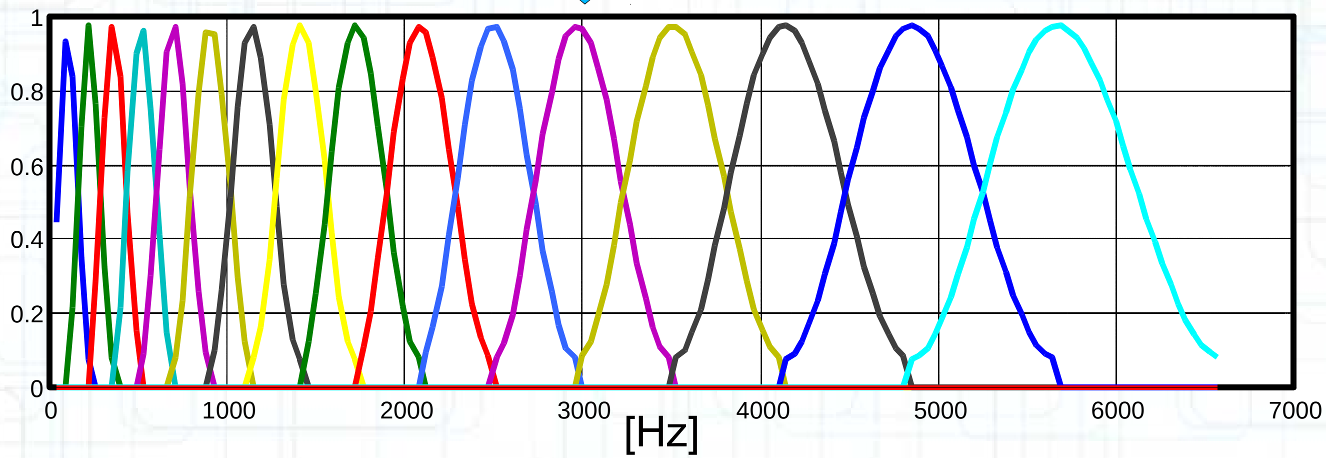
$$m = \frac{1000}{\ln(1 + 1000/700)} \,.\, \ln(1 + f/700)$$

Coeficientes cepstrales con frecuencia en escala de Mel (MFCC)

- Pasos obtener los coeficientes MEL-Cepstrum:

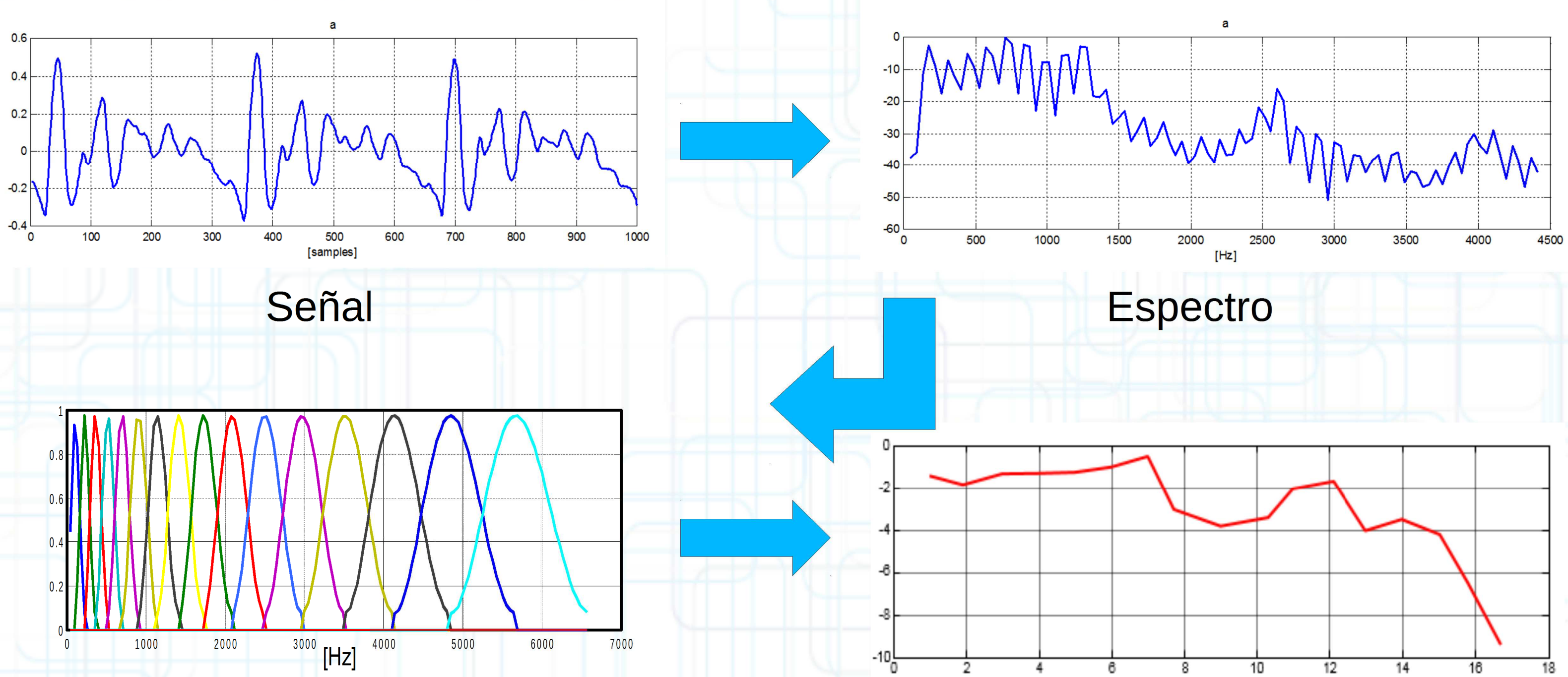


Se filtra el espectro mediante un banco de filtros en el dominio de Mel



Coeficientes cepstrales con frecuencia en escala de Mel (MFCC)

- Pasos obtener los coeficientes MEL-Cepstrum:



Señal

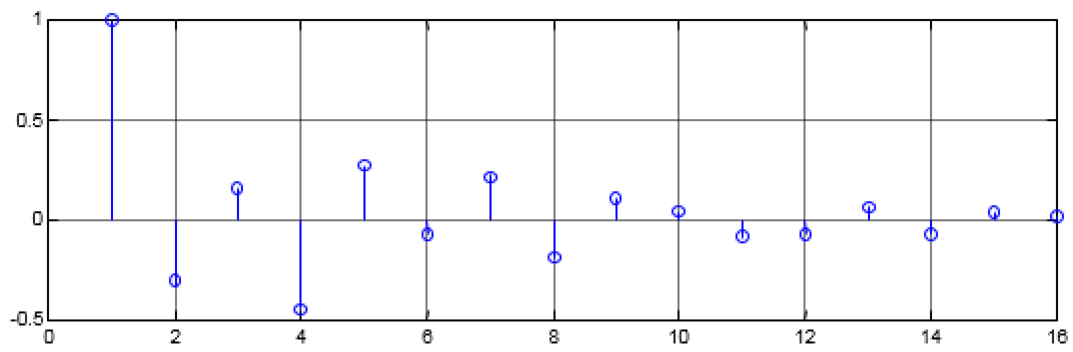
Espectro

Banco de filtros

Espectro suavizado

Coeficientes cepstrales con frecuencia en escala de Mel (MFCC)

- Pasos obtener los coeficientes MEL-Cepstrum:
  - El nº de coeficientes es muy inferior.
  - El cepstrum obtenido es una aproximación.



Cepstrum LPC (LPCC)

- Podemos obtener los coeficientes cepstrales a partir de los coeficientes LPC:

$$c(1) = -a_1$$

$$c(n) = -a_n - \sum_{m=1}^{n-1}\left(1-\frac{m}{n}\right)a_m c(n-m) \qquad n = 2..P$$

$$c(n) = -\sum_{m=1}^{P}\left(1-\frac{m}{n}\right)a_m c(n-m) \qquad n > P$$