Procesamiento de la señal de voz

Leandro Vignolo
Diego Milone

Procesamiento Digital de Señales
Ingeniería Informática FICH-UNL

9 de mayo de 2013

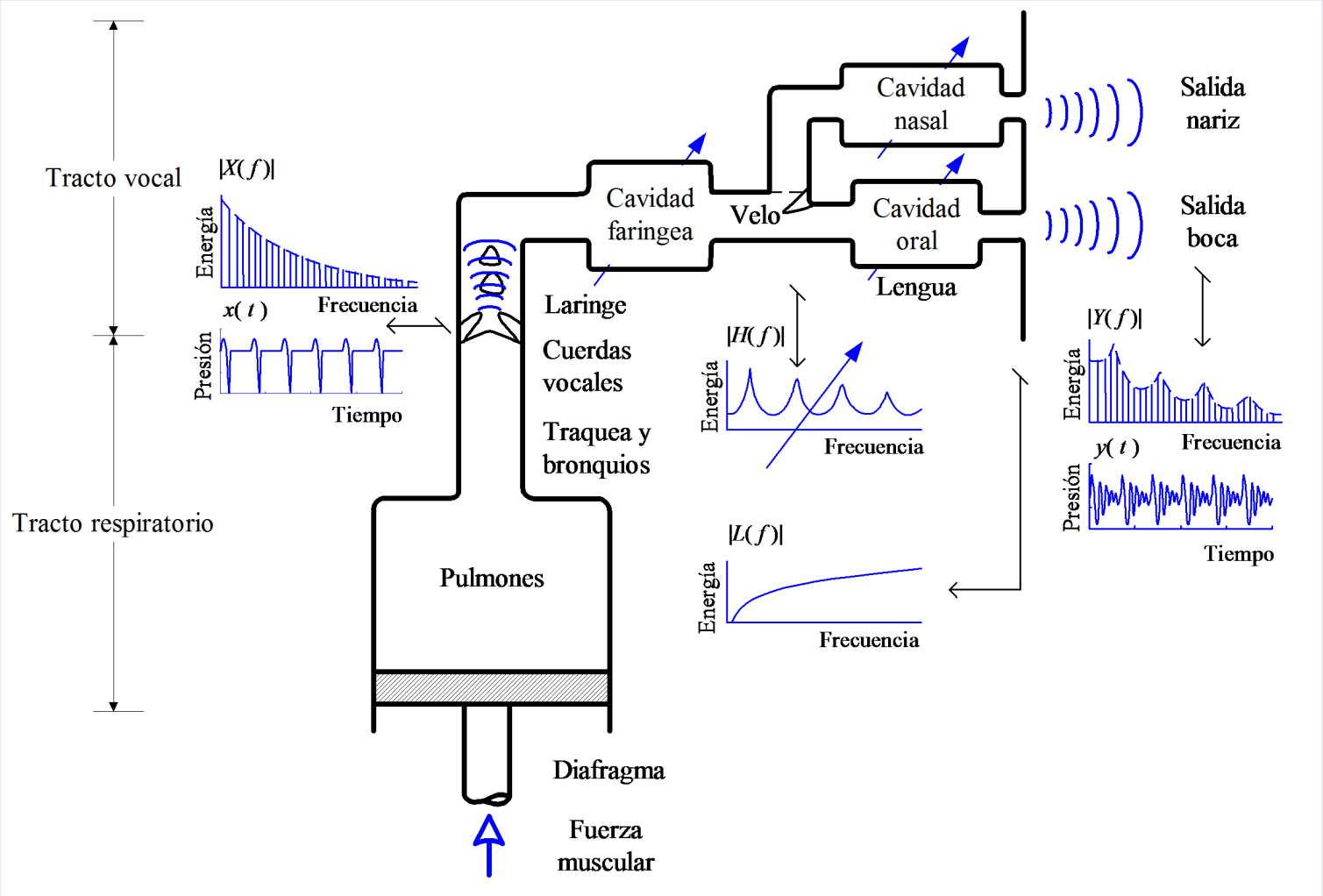
Organización de la clase

1. Producción y percepción de la voz
   - Generalidades del aparato fonador
   - Fuentes y modificadores del sonido de la voz
   - Generalidades del oído
   - Percepción del sonido
2. Organización estructural del habla
   - Niveles de la estructura
   - Análisis por tramos
3. Procesamiento homomórfico
   - Definición de los coeficientes cepstrales
   - Procesamiento homomórfico de la voz
   - Estimación de F0

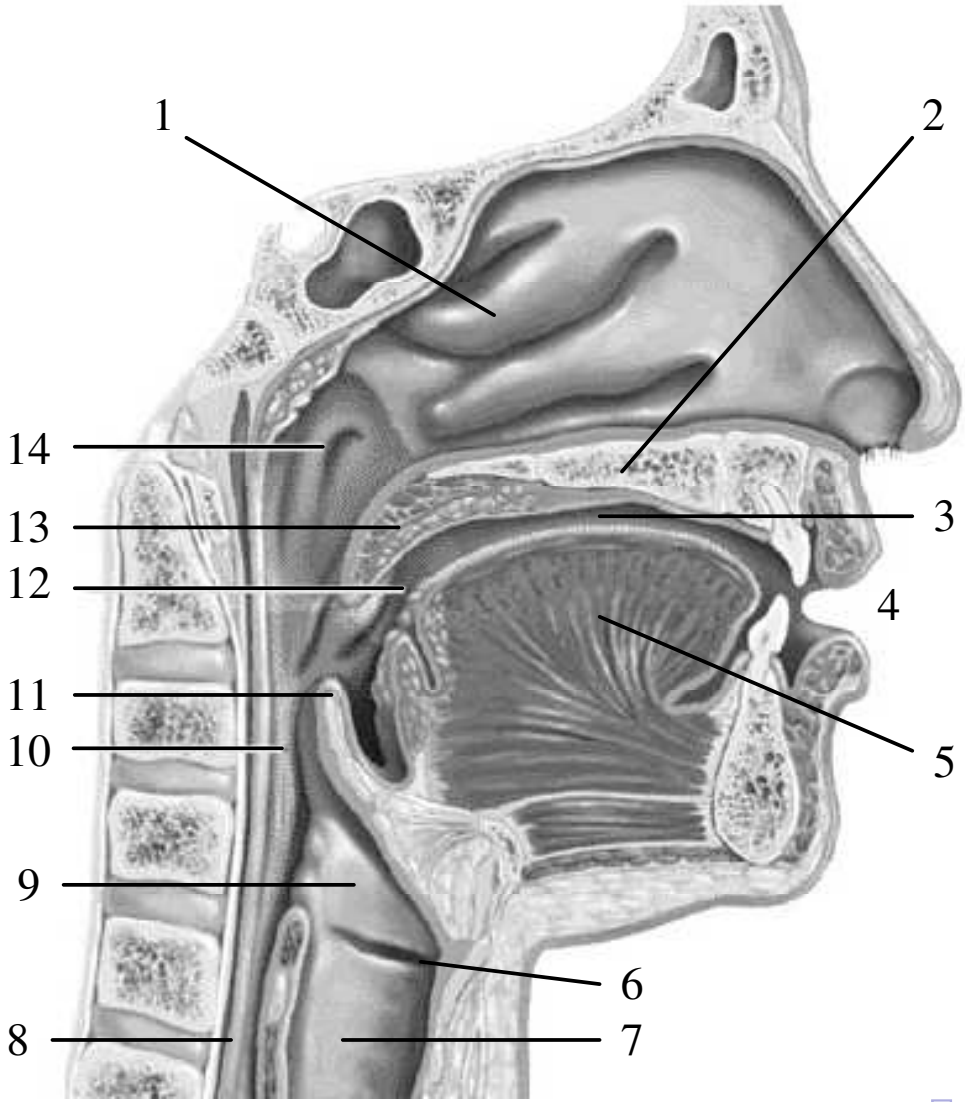
Organización de la clase

Aparato fonador

Diagrama esquemático del aparato fonador



Estructura anatómica del tracto vocal



Fuentes principales del sonido

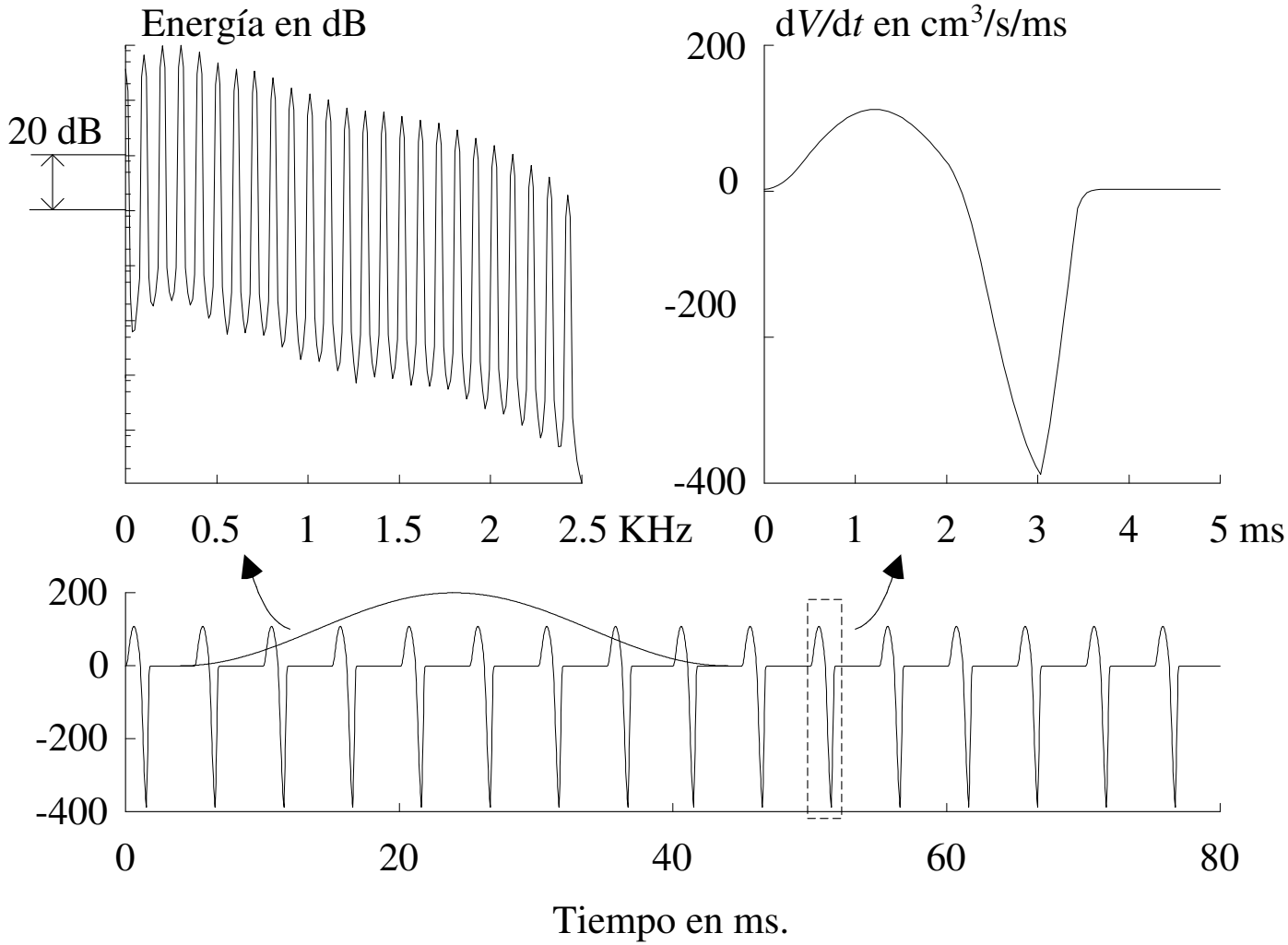
Tipos de entrada

- Tren de pulsos cuasiperiódicos (sonidos sonoros)
- Ruido de banda ancha

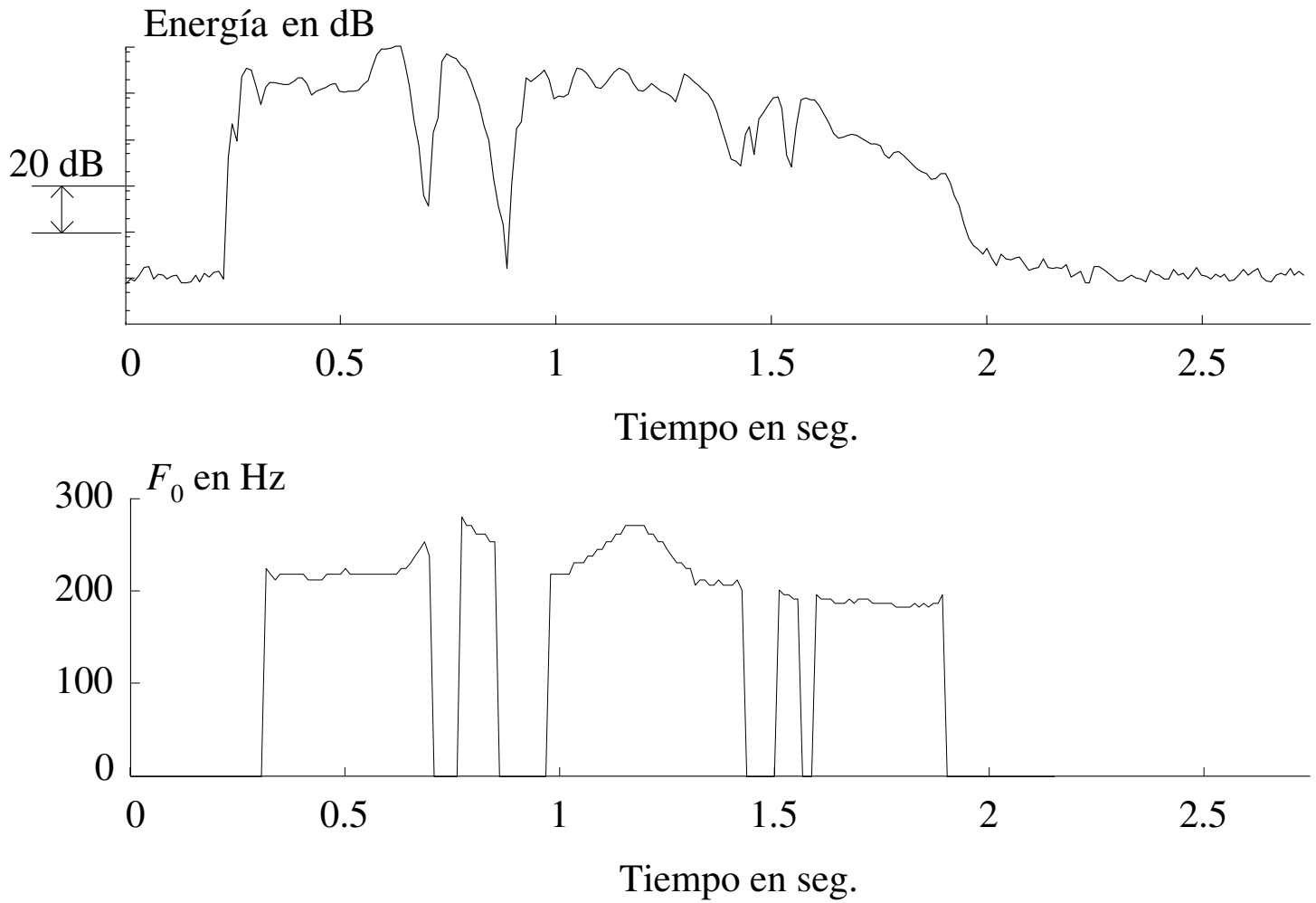
Modificadores del sonido

- Restricciones en el flujo de aire
- Labios, lengua, dientes, etc.

Pulsos glóticos



Energía y entonación



Modificadores del sonido

- Morfología del tracto vocal
- Circuito nasal
- Radiación en los labios
- Posición de la lengua

Análisis de la señal de voz

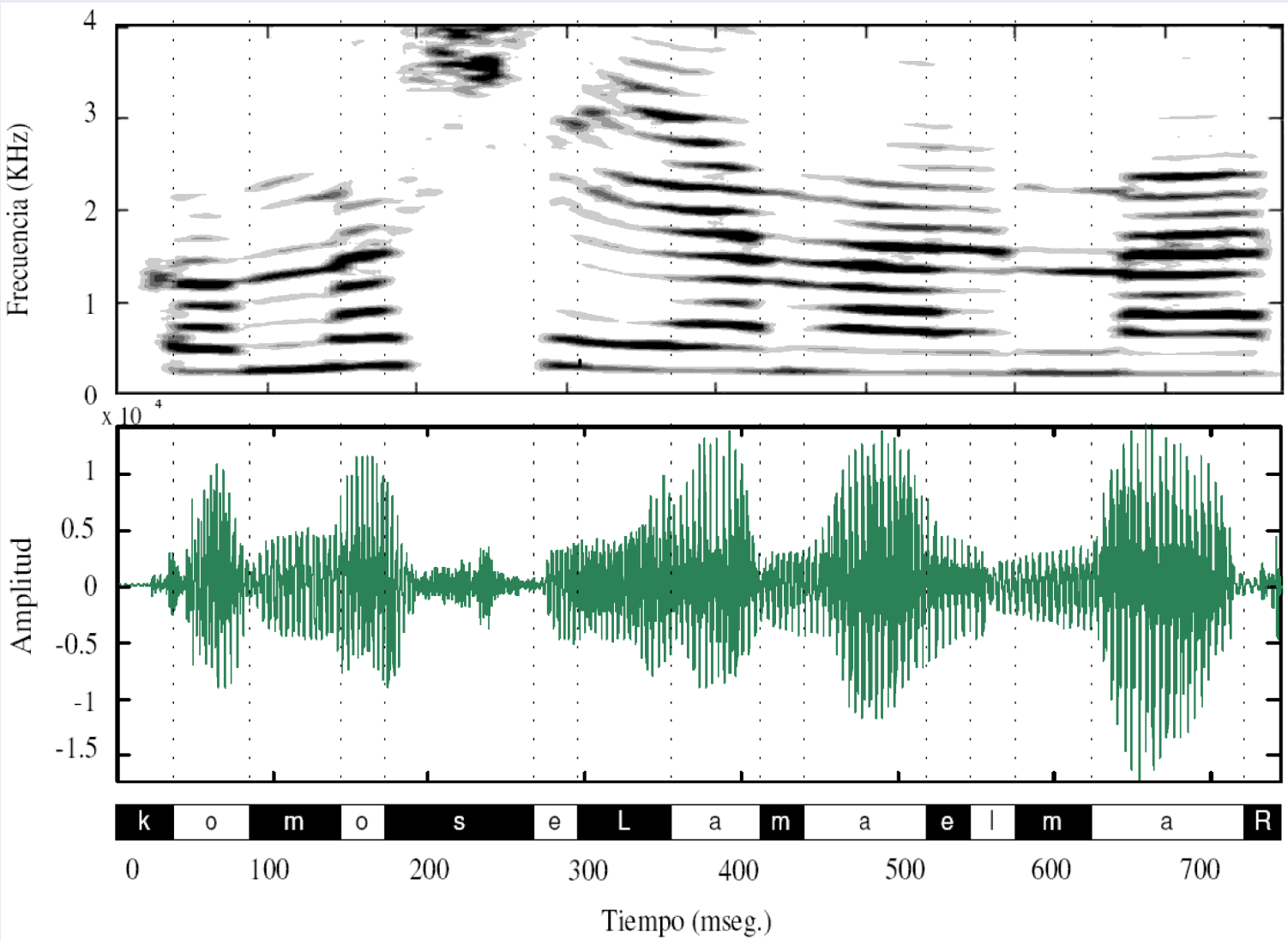
Vocal sostenida - Período y Frecuencia fundamental (F_0) - Formantes

La frecuencia fundamental F_0 corresponde a la frecuencia glótica, presente en los fonemas sonoros, y es una componente importante de la entonación en el habla.

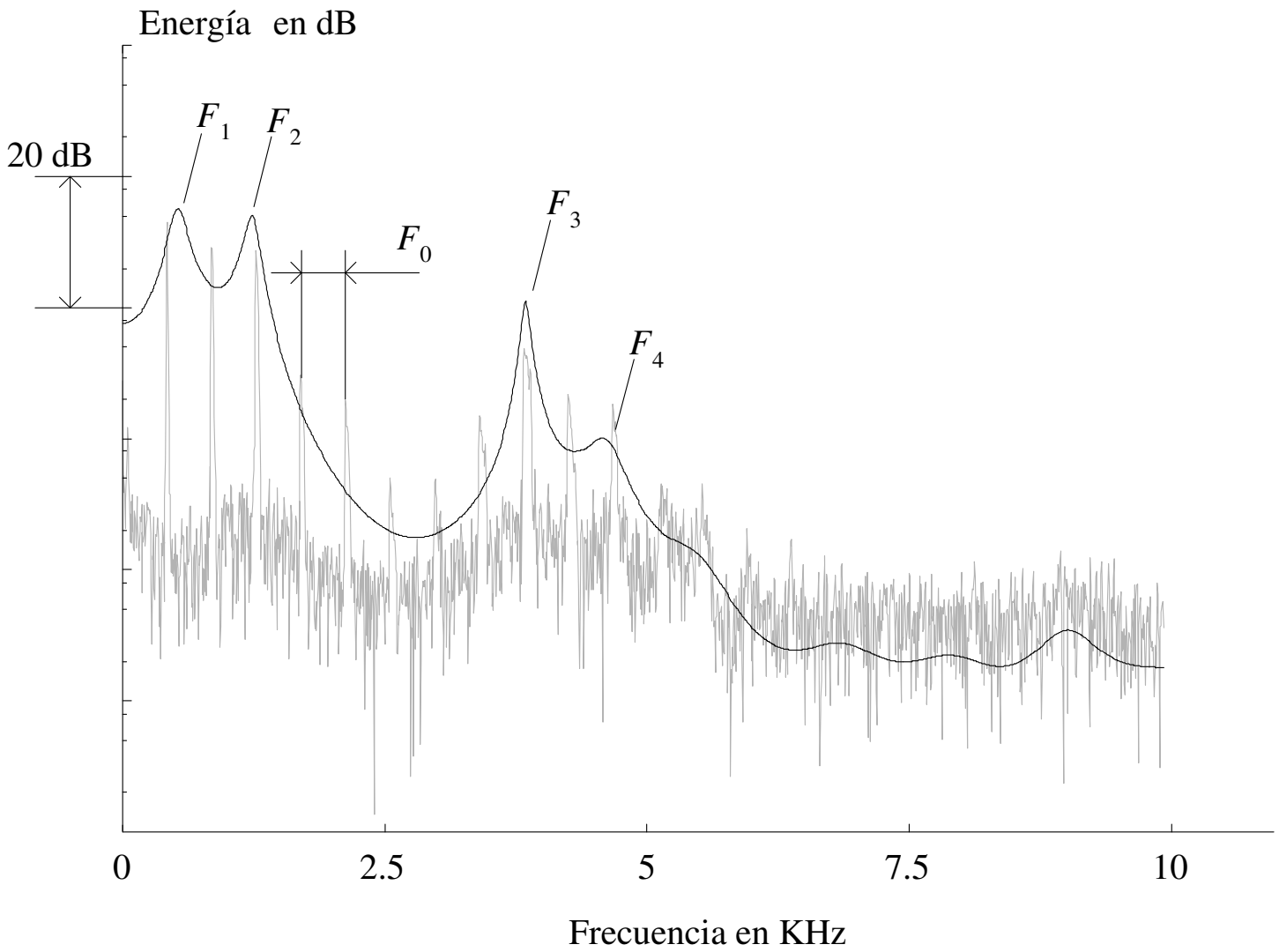
Las frecuencias formantes $(F_1, F_2, F_3, \cdots)$ permiten discriminar entre las vocales. Su variación temporal posibilita también diferenciar entre los diferentes fonemas sonoros.

Análisis de la señal de voz

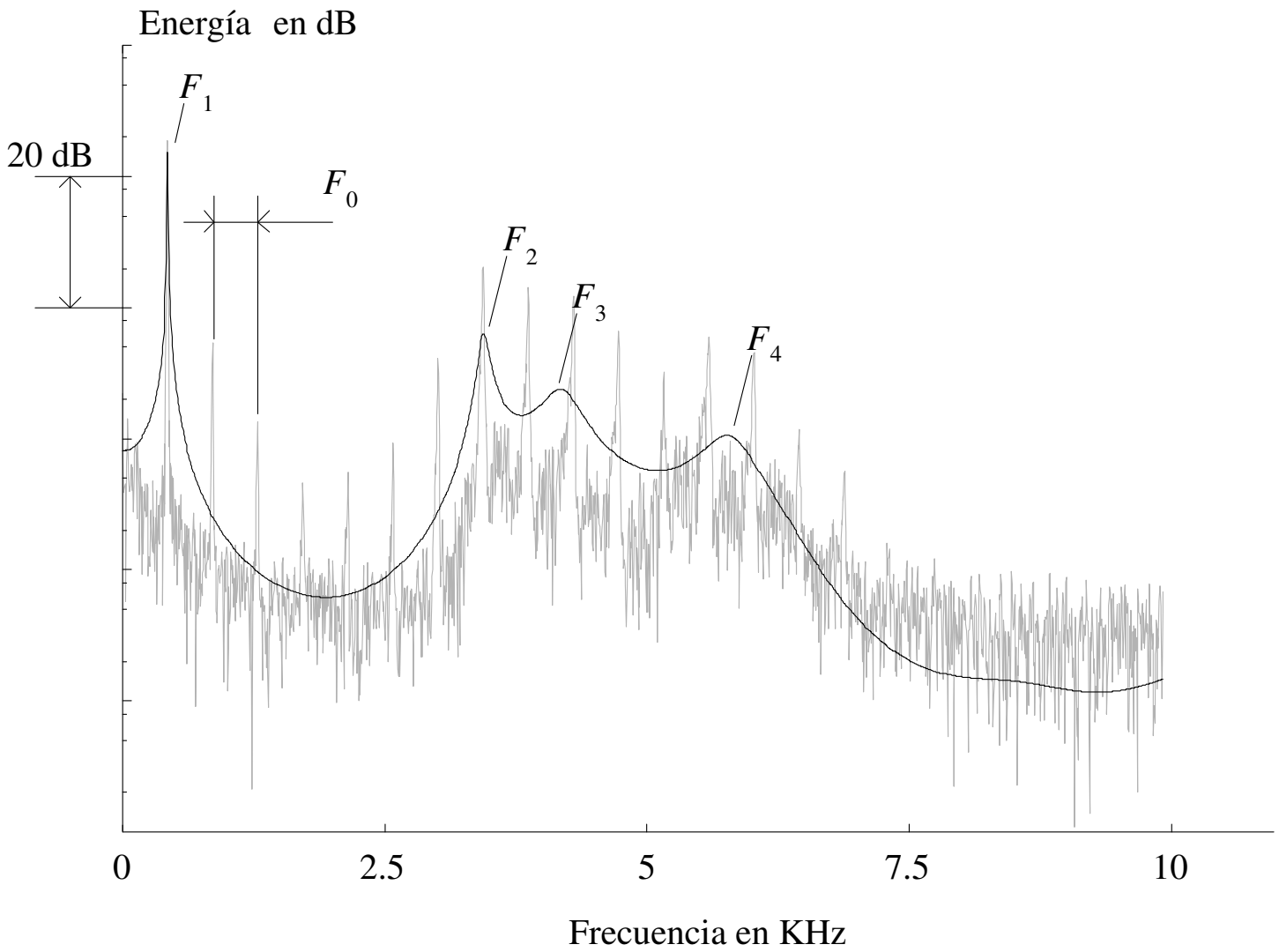
Sonograma y espectrograma



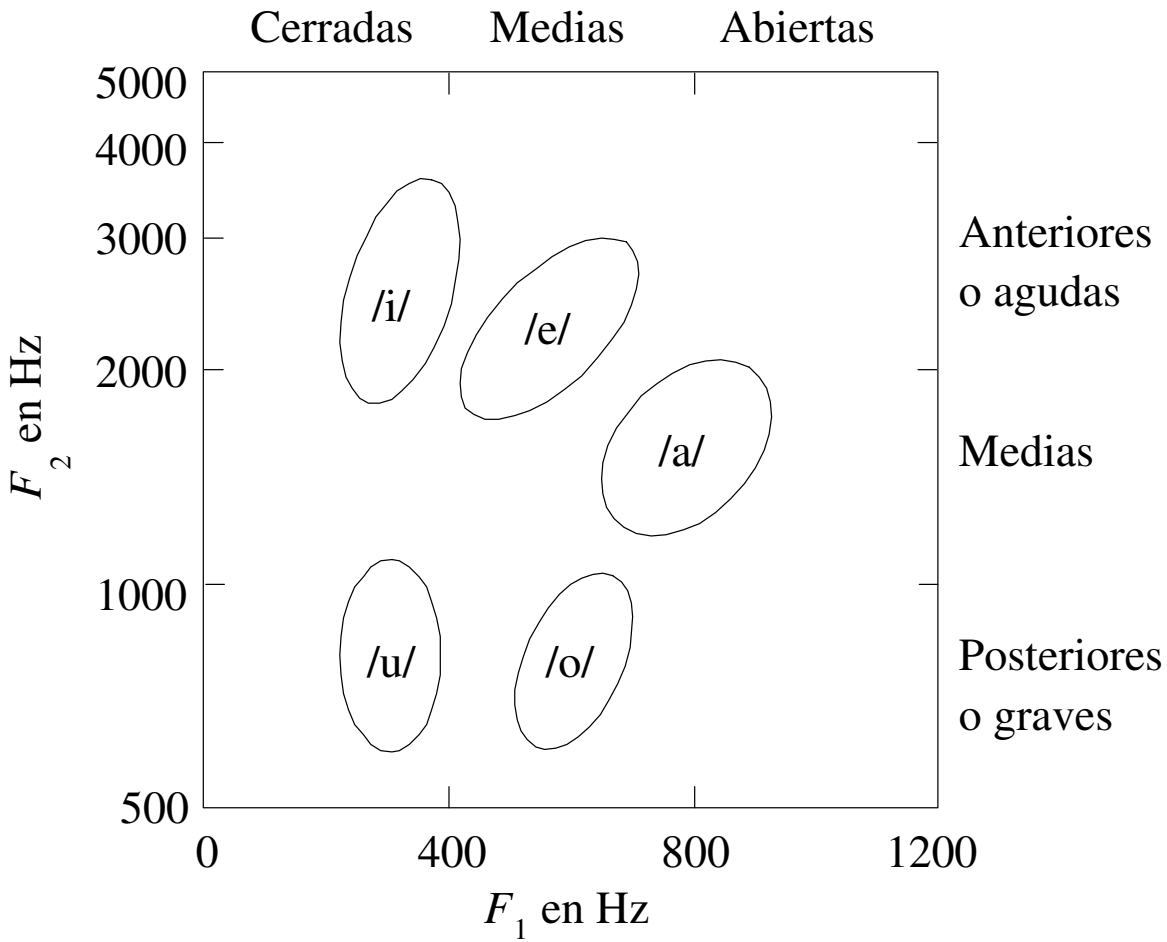
Espectro de una vocal



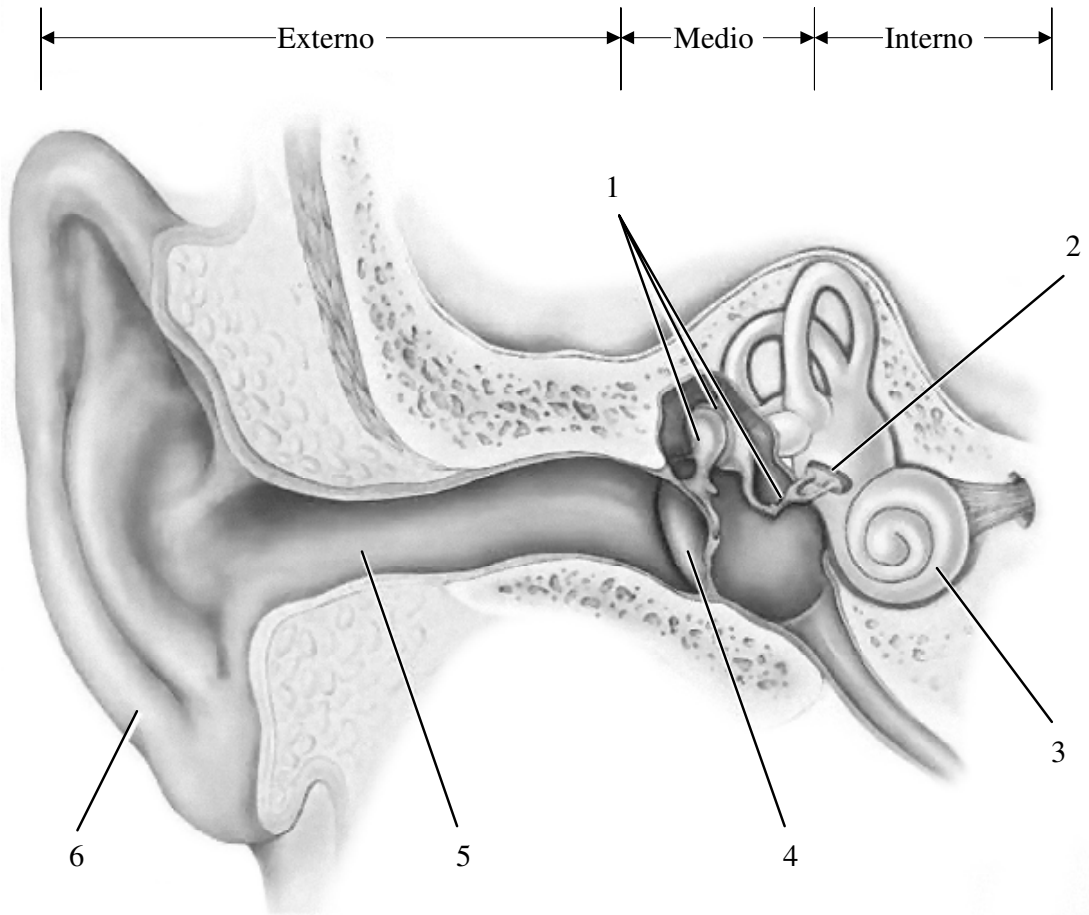
Espectro de una vocal



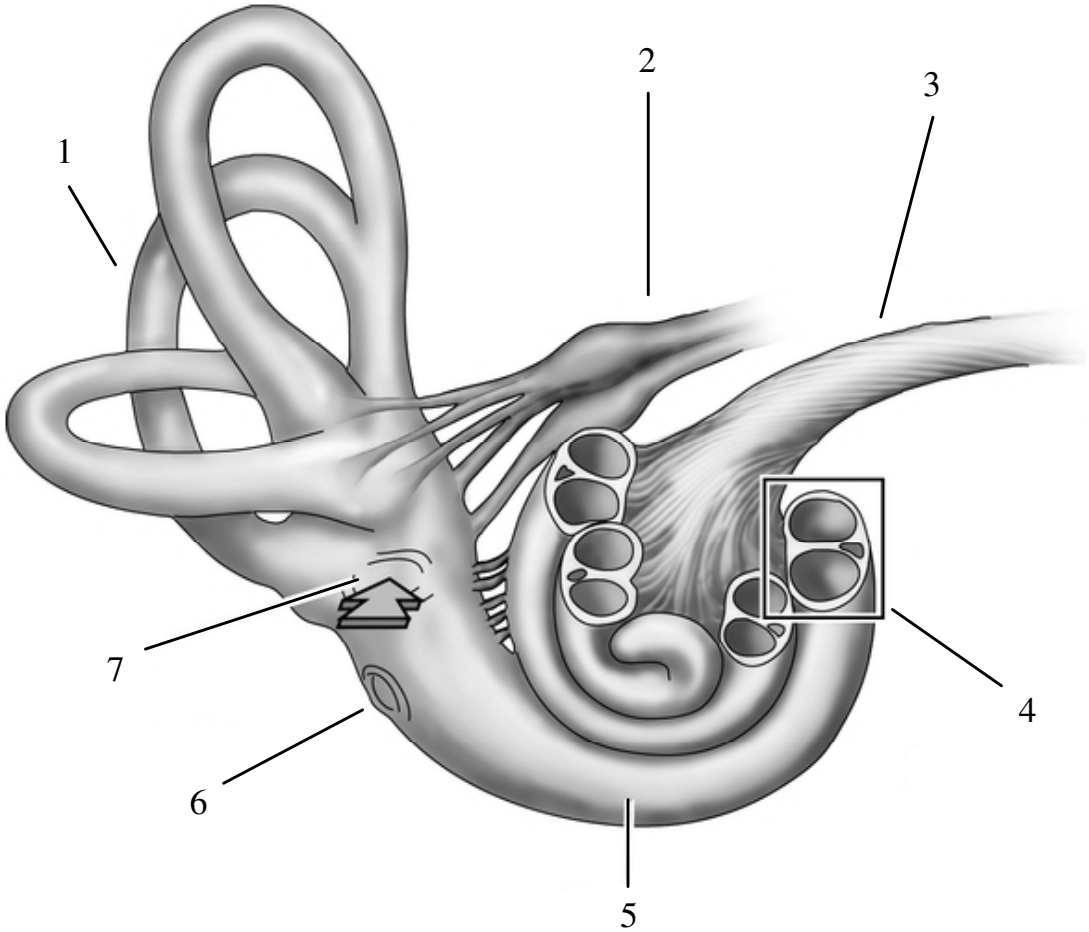
Triángulo de las vocales



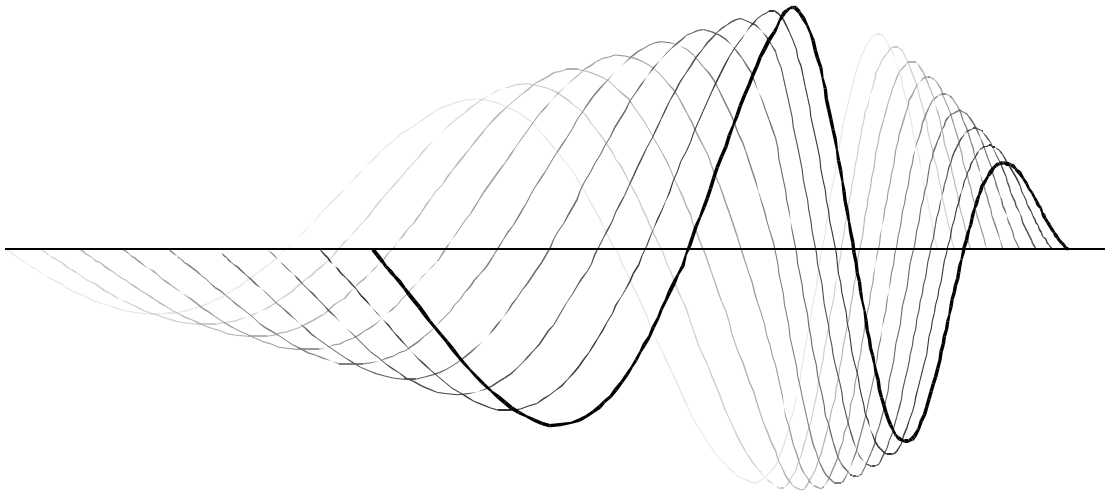
Partes del oído



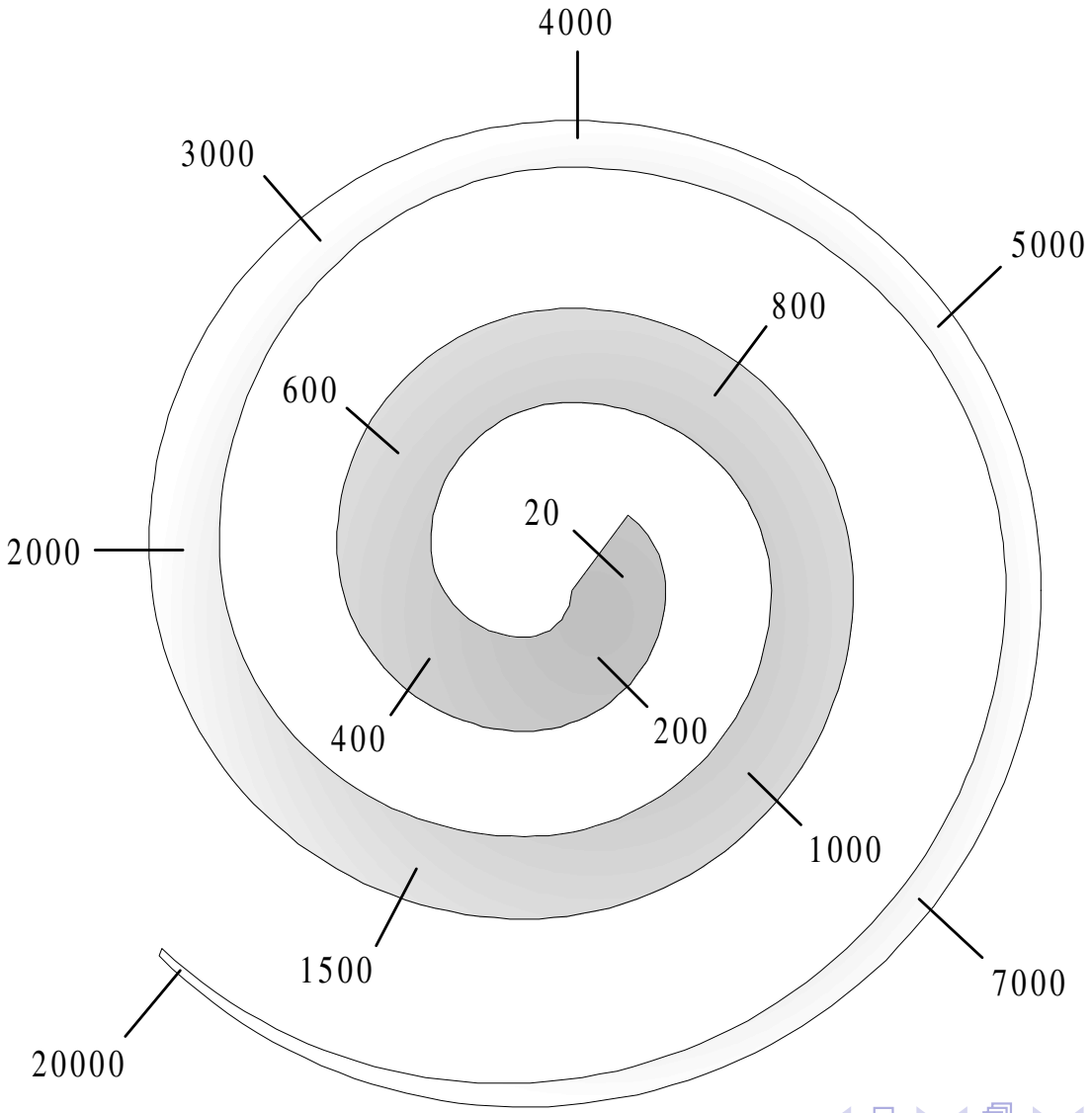
Cóclea



Onda viajera



Principio de tonotopía



Frecuencia y Pitch

- A menudo confundidos en la literatura, el pitch no es igual a la frecuencia fundamental.
- La frecuencia, intensidad y las propiedades espectrales de un sonido interactúan en formas muy complejas para dar una percepción de pitch que puede ser un reflejo muy pobre de la F_0. El pitch percibido cambia con la intensidad.
- El pitch se refiere a un atributo perceptual del sonido, mientras que a frecuencia es un atributo físico de las señales.

Escala de mel

Mel

La unidad del pitch percibido de un tono puro es el mel. No se corresponde linealmente con la frecuencia física del tono. Stevens y Volkman (1940) establecieron arbitrariamente: 1000 Hz $= 1000$ mel.

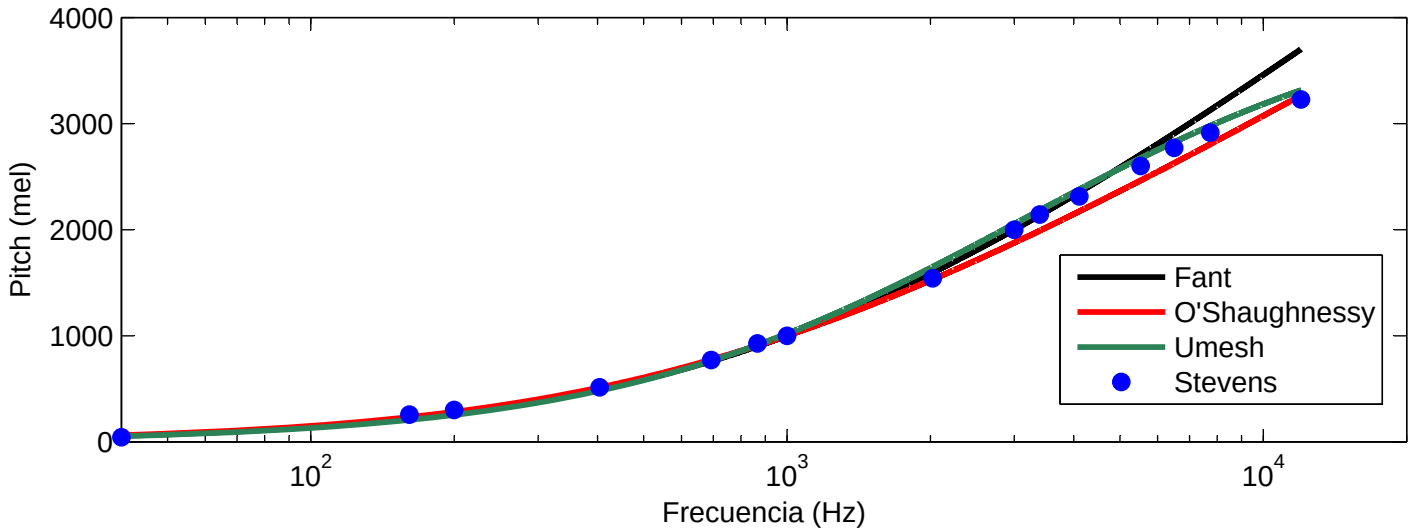
Escala de mel

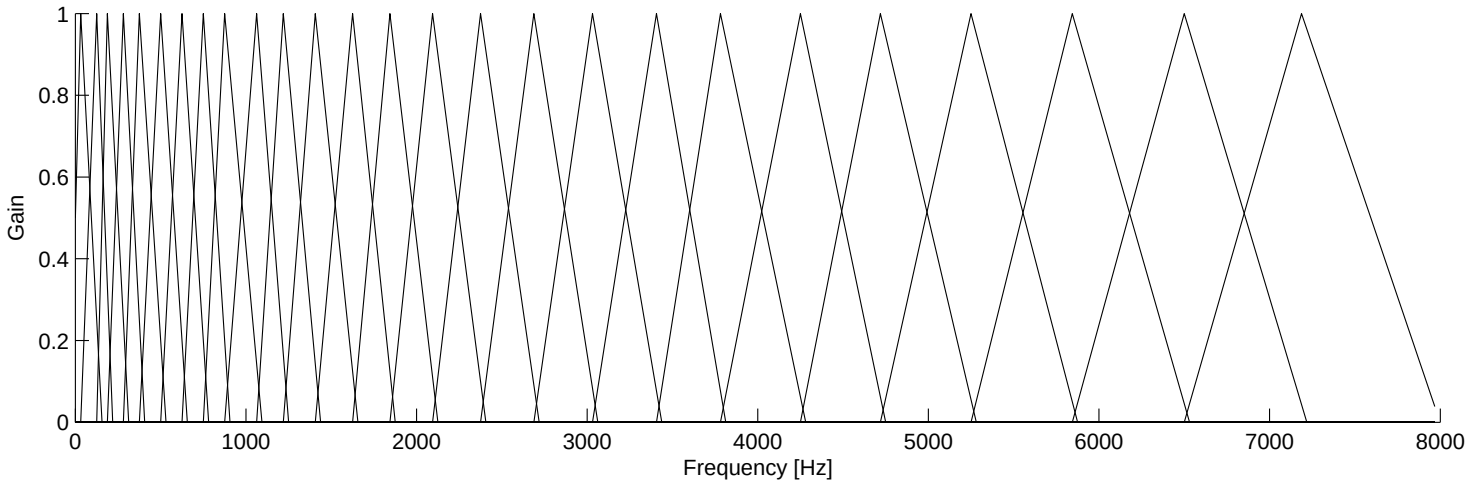
$$F_{mel} = \frac{1000}{\log(2)} \log\left(1 + \frac{F_{Hz}}{1000}\right) \qquad \text{(Fant, 1973)}$$

Otras variantes

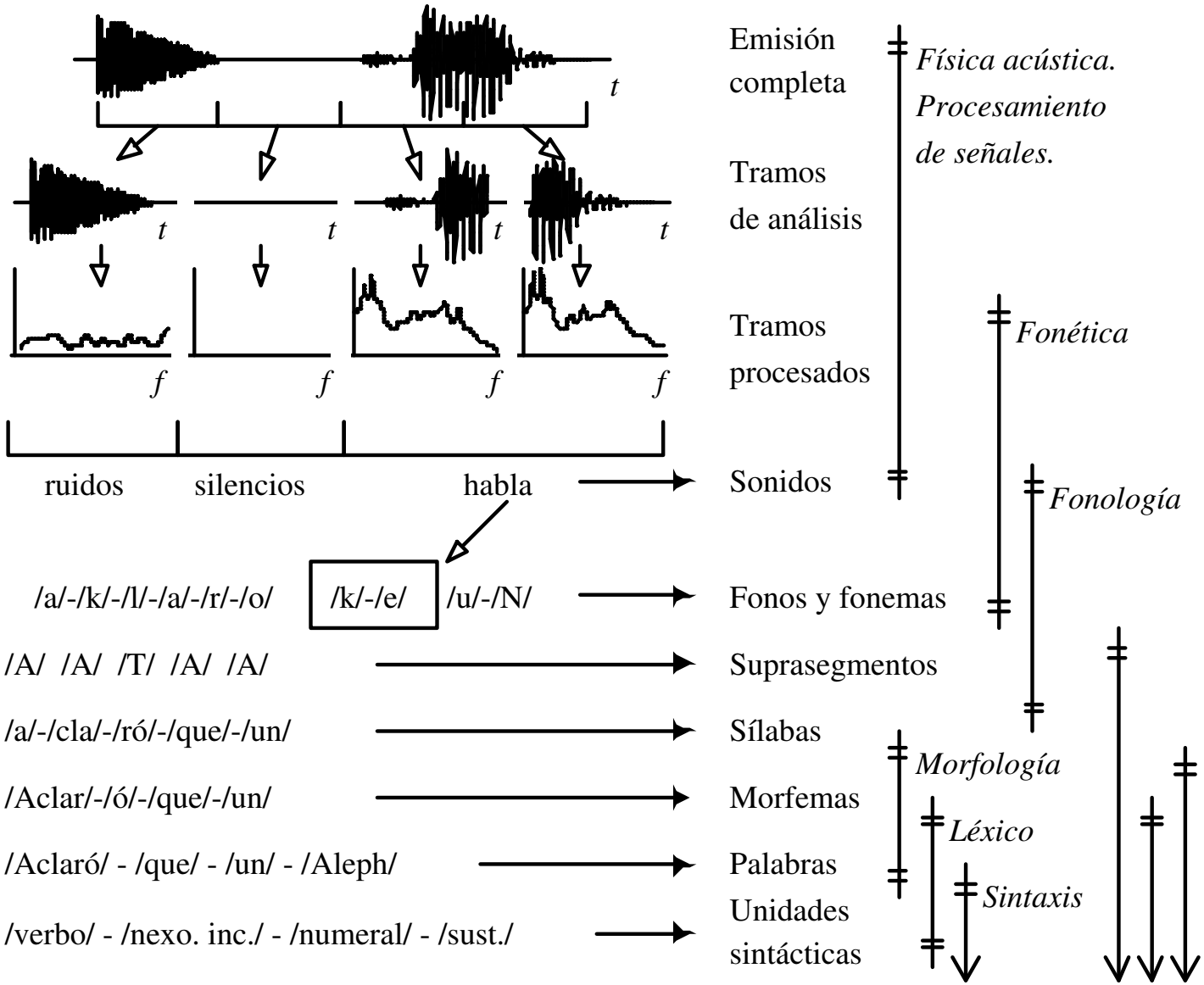
- O'Shaugnessy (1987)
- Umesh (1999)

Banco de filtros en escala de mel





Organización de la clase


t
t
t
t
t
f
f
f
f
ruidos
silencios
habla
/a/-/k/-/l/-/a/-/r/-/o/
/k/-/e/
/u/-/N/
/A/ /A/ /T/ /A/ /A/
/a/-/cla/-/ró/-/que/-/un/
/Aclar/-/ó/-/que/-/un/
/Aclaró/ - /que/ - /un/ - /Aleph/
/verbo/ - /nexo. inc./ - /numeral/ - /sust./
Emisión completa
Tramos de análisis
Tramos procesados
Sonidos
Fonos y fonemas
Suprasegmentos
Sílabas
Morfemas
Palabras
Unidades sintácticas
Física acústica. Procesamiento de señales.
Fonética
Fonología
Morfología
Léxico
Sintaxis

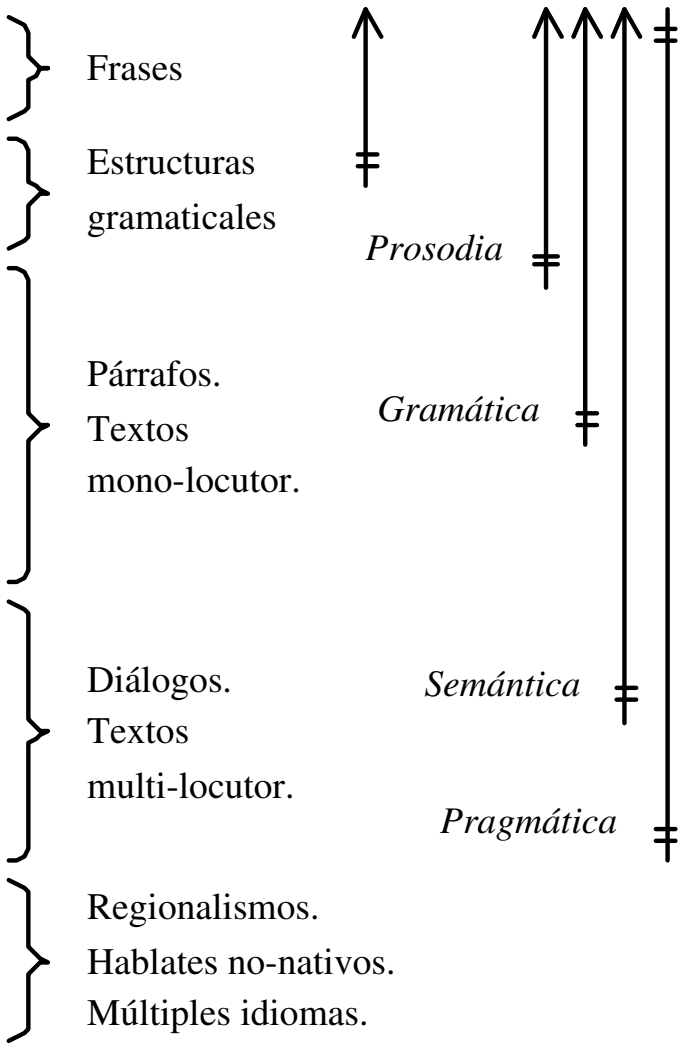
Aclaró que un Aleph es uno de los puntos del espacio que contiene a todos los puntos.

¿/pred./-/sujeto/-/pred./? ¿/pred./? /pred./-/subordinadas/-/sujeto/.

¿Existe ese Aleph en lo íntimo de una piedra? ¿Lo he visto cuando vi todas las cosas y lo he olvidado? Nuestra mente es porosa para el olvido; yo mismo estoy falseando y perdiendo, baja la trágica erosión de los años, los rasgos de Beatriz.

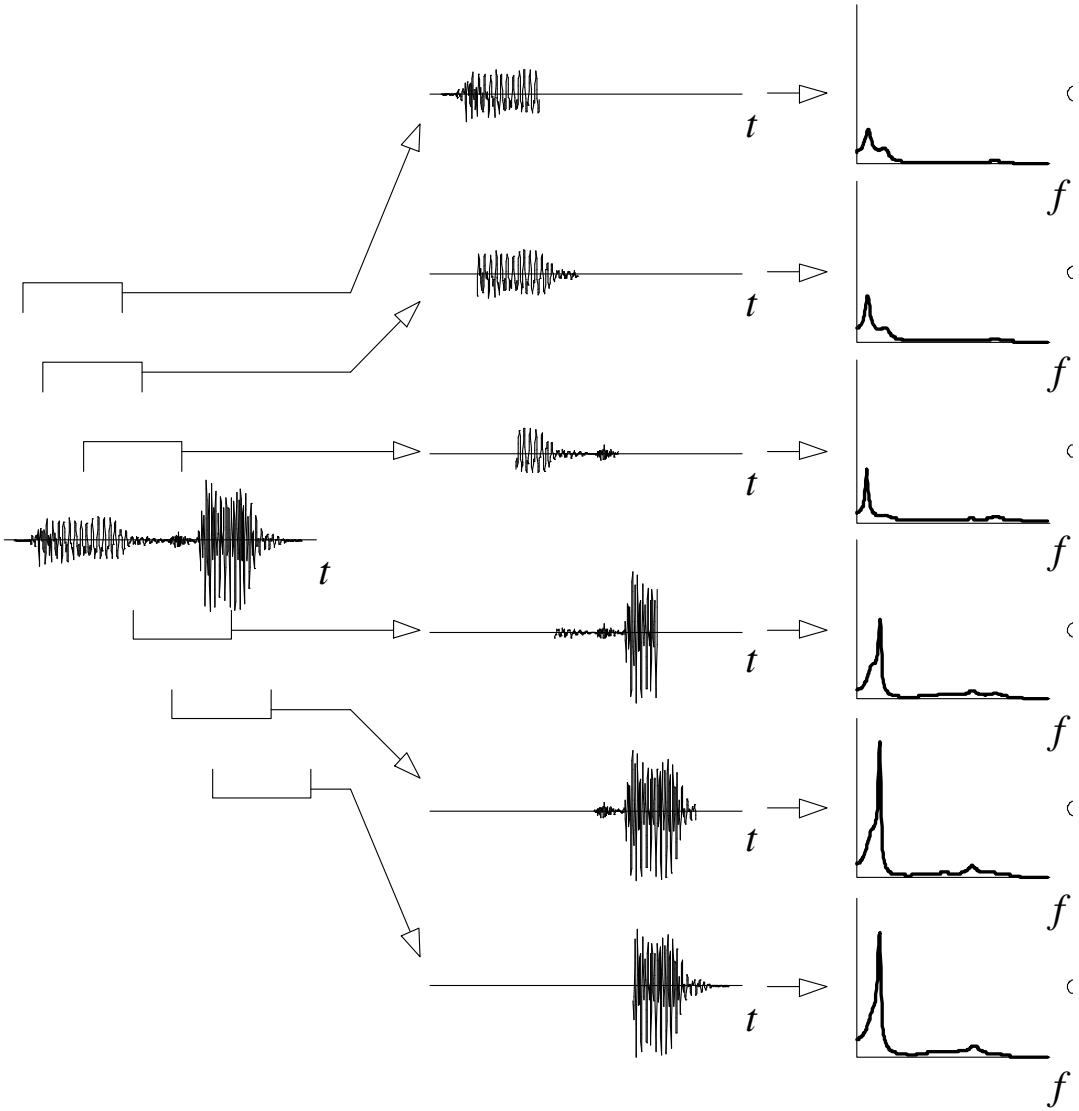
- Pero, ¿no es muy oscuro el sótano?
- La verdad no penetra en un entendimiento rebelde. Si todos los lugares de la tierra están en el Aleph, ahí estarán todas la luminarias...

¡Qué observatorio formidable, che Borges! O God!, I could be bounded in a nutshell, and count myself a King of infinite space...



Análisis por tramos

- Necesidad
- Ventanas cuadradas
- Técnicas de ventaneo
- Solapado en el tiempo
- Análisis de las ventanas independientes


t
f
t
f
t
f
t
f
t
f
t
f
t
f

Ventaneo

$$
\begin{aligned}
v(t;n) &= \omega(n;N_\omega)x(tN_d+n), \quad 0<n\leq N_\omega \\
&\omega_H(m;N_\omega) = \frac{27}{50} - \frac{23}{50}\cos(2\pi m/N_\omega)
\end{aligned}
$$

$$
X(t;k) = \mathcal{T}(k)\left\{v(t;n)\right\}, \quad 0<k\leq N_x
$$

Transformaciones de dominio

I) CE:

$$\mathbf{x}_t = [u(t;k)] = \mathcal{T}_F(k)\{v(t;n)\},$$

II) CPL:

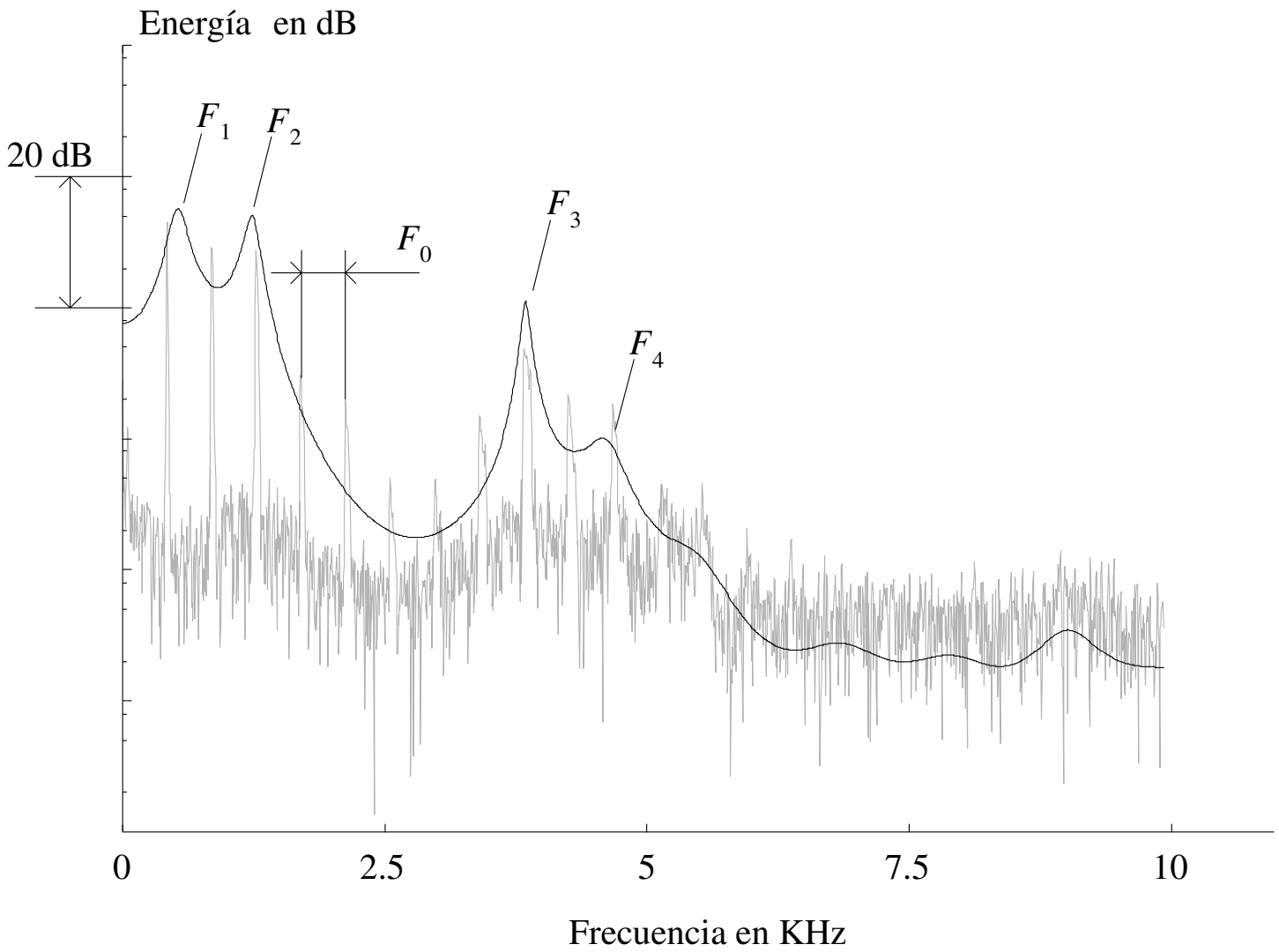
$$\mathbf{x}_t = [a(t;k)] = \mathcal{T}_L(k)\{v(t;n)\},$$

III) CC:

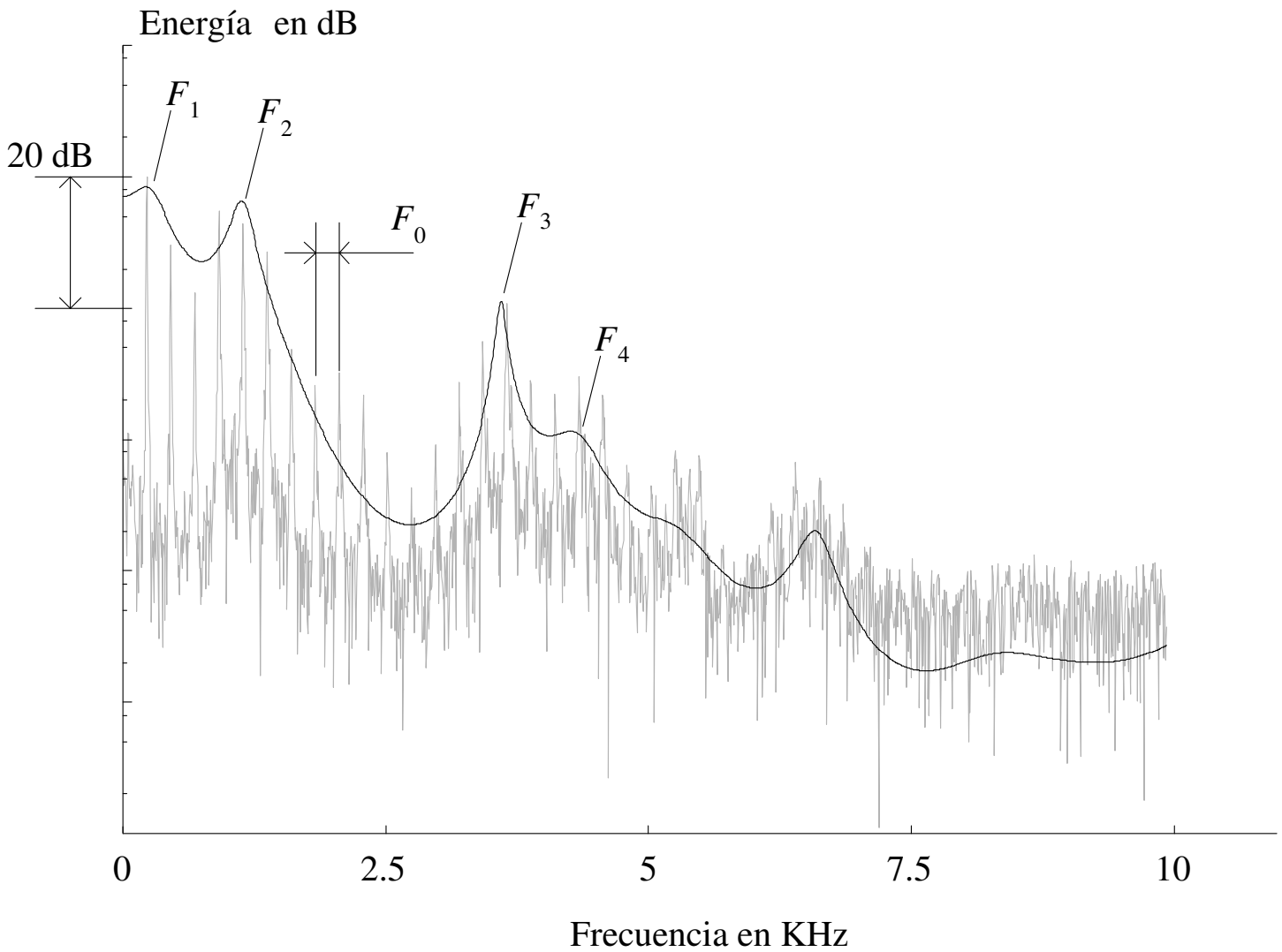
$$\mathbf{x}_t = [c(t;k)] = \mathcal{T}_C(k)\{v(t;n)\}$$

Organización de la clase

Espectro de una vocal



Otra elocución de la misma vocal



Coeficientes cepstrales

$$c(m) = \mathcal{T}_F^{-1}\left\{\log\left|\mathcal{T}_F\left\{v(m)\right\}\right|\right\}$$

- Espectral $\rightarrow$ Cepstral
- Espectro $\rightarrow$ Cepstro
- Frcuencias $\rightarrow$ Cuefrencias
- Filtro, filtrado $\rightarrow$ Liftro, liftrado
- Armónicas $\rightarrow$ Ramónicas

Coeficientes cepstrales

$$c(m) = \mathcal{T}_F^{-1}\left\{\log\left|\mathcal{T}_F\left\{v(m)\right\}\right|\right\}$$

Espectral $\rightarrow$ Cepstral

Espectro $\rightarrow$ Cepstro

Frcuencias $\rightarrow$ Cuefrencias

Filtro, filtrado $\rightarrow$ Liftro, liftrado

Armónicas $\rightarrow$ Ramónicas

Separación de fuentes y modificadores del sonido

$$\hat{v}(n) = g(n) * h(n)$$

$$\hat{V}(k) = G(k) \times H(k)$$

$$\hat{\log}\,|V(k)| = \log|G(k) \times H(k)|$$

$$\hat{\log}\,|V(k)| = \log|G(k)| + \log|H(k)|$$

$$\hat{v}(m) = \mathcal{T}_F^{-1}\{\log|G(k)|\} + \mathcal{T}_F^{-1}\{\log|H(k)|\}$$

Separación de fuentes y modificadores del sonido

$$\hat{v}(n) = g(n) * h(n)$$

$$\hat{V}(k) = G(k) \times H(k)$$

$$\hat{\log}\,|V(k)| = \log|G(k) \times H(k)|$$

$$\hat{\log}\,|V(k)| = \log|G(k)| + \log|H(k)|$$

$$\hat{v}(m) = \mathcal{T}_F^{-1}\{\log|G(k)|\} + \mathcal{T}_F^{-1}\{\log|H(k)|\}$$

Separación de fuentes y modificadores del sonido

$$\hat{v}(n) = g(n) * h(n)$$

$$\hat{V}(k) = G(k) \times H(k)$$

$$\hat{\log}\,|V(k)| = \log|G(k) \times H(k)|$$

$$\hat{\log}\,|V(k)| = \log|G(k)| + \log|H(k)|$$

$$\hat{v}(m) = \mathcal{T}_F^{-1}\{\log|G(k)|\} + \mathcal{T}_F^{-1}\{\log|H(k)|\}$$

Separación de fuentes y modificadores del sonido

$$\hat{v}(n) = g(n) * h(n)$$

$$\hat{V}(k) = G(k) \times H(k)$$

$$\hat{\log}\,|V(k)| = \log|G(k) \times H(k)|$$

$$\hat{\log}\,|V(k)| = \log|G(k)| + \log|H(k)|$$

$$\hat{v}(m) = \mathcal{T}_F^{-1}\{\log|G(k)|\} + \mathcal{T}_F^{-1}\{\log|H(k)|\}$$

Separación de fuentes y modificadores del sonido

$$\hat{v}(n) = g(n) * h(n)$$

$$\hat{V}(k) = G(k) \times H(k)$$

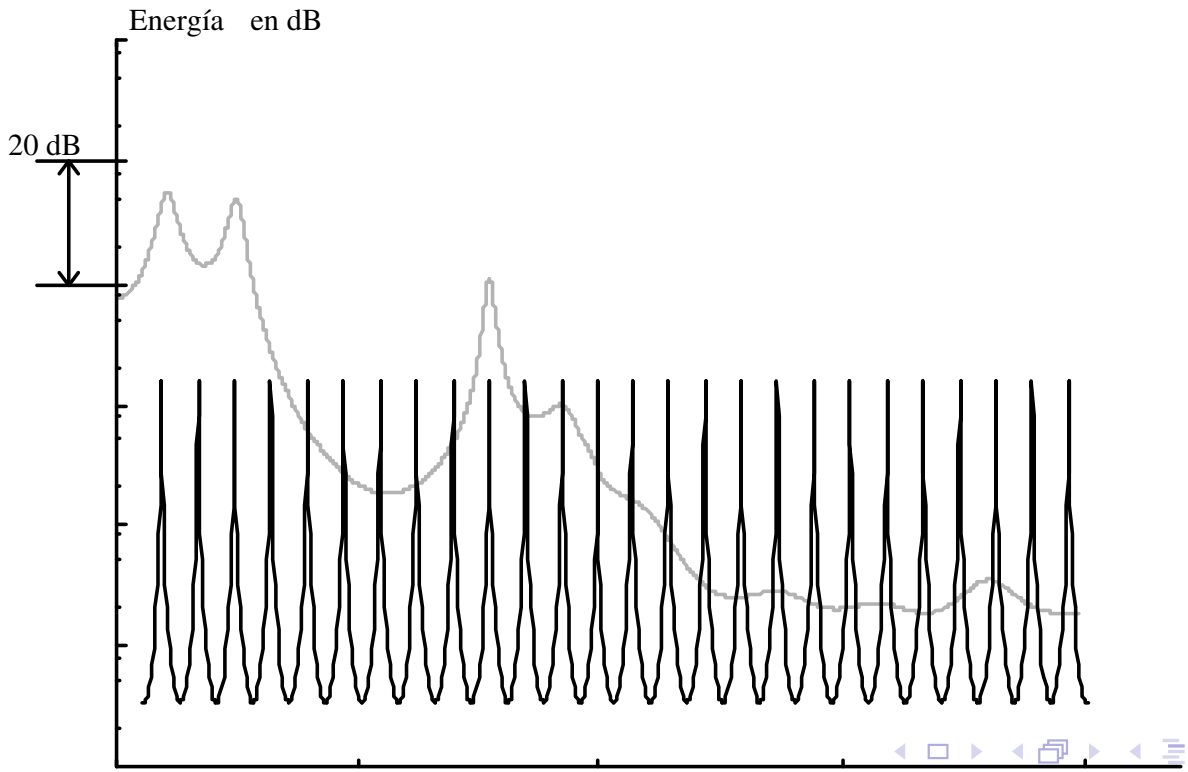
$$\hat{\log}\,|V(k)| = \log|G(k) \times H(k)|$$

$$\hat{\log}\,|V(k)| = \log|G(k)| + \log|H(k)|$$

$$\hat{v}(m) = \mathcal{T}_F^{-1}\{\log|G(k)|\} + \mathcal{T}_F^{-1}\{\log|H(k)|\}$$

Separación de fuentes y modificadores del sonido

$$\hat{v}(m) = \mathcal{T}_F^{-1}\{\log|G(k)|\} + \mathcal{T}_F^{-1}\{\log|H(k)|\}$$

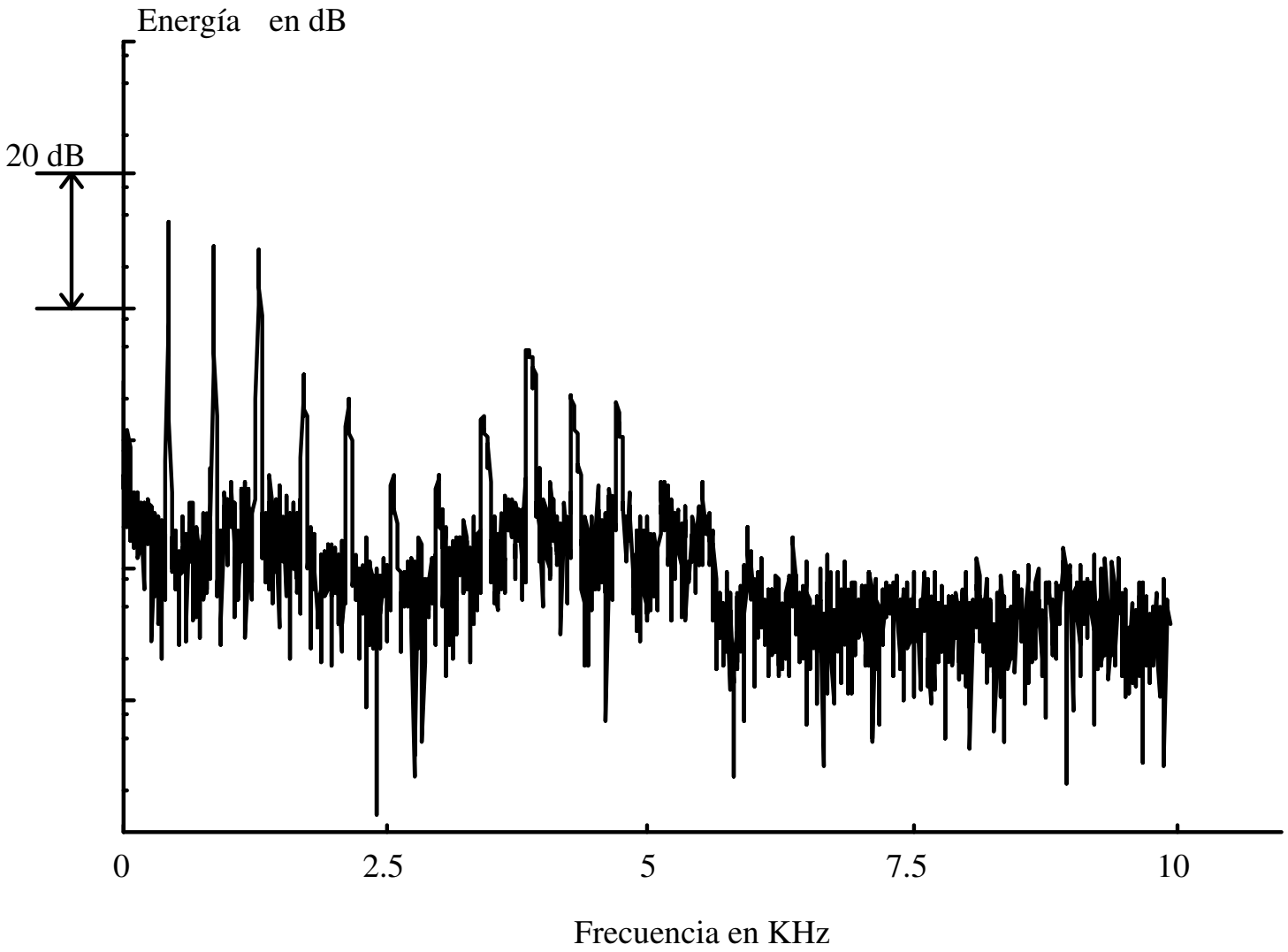


Separación de fuentes y modificadores del sonido

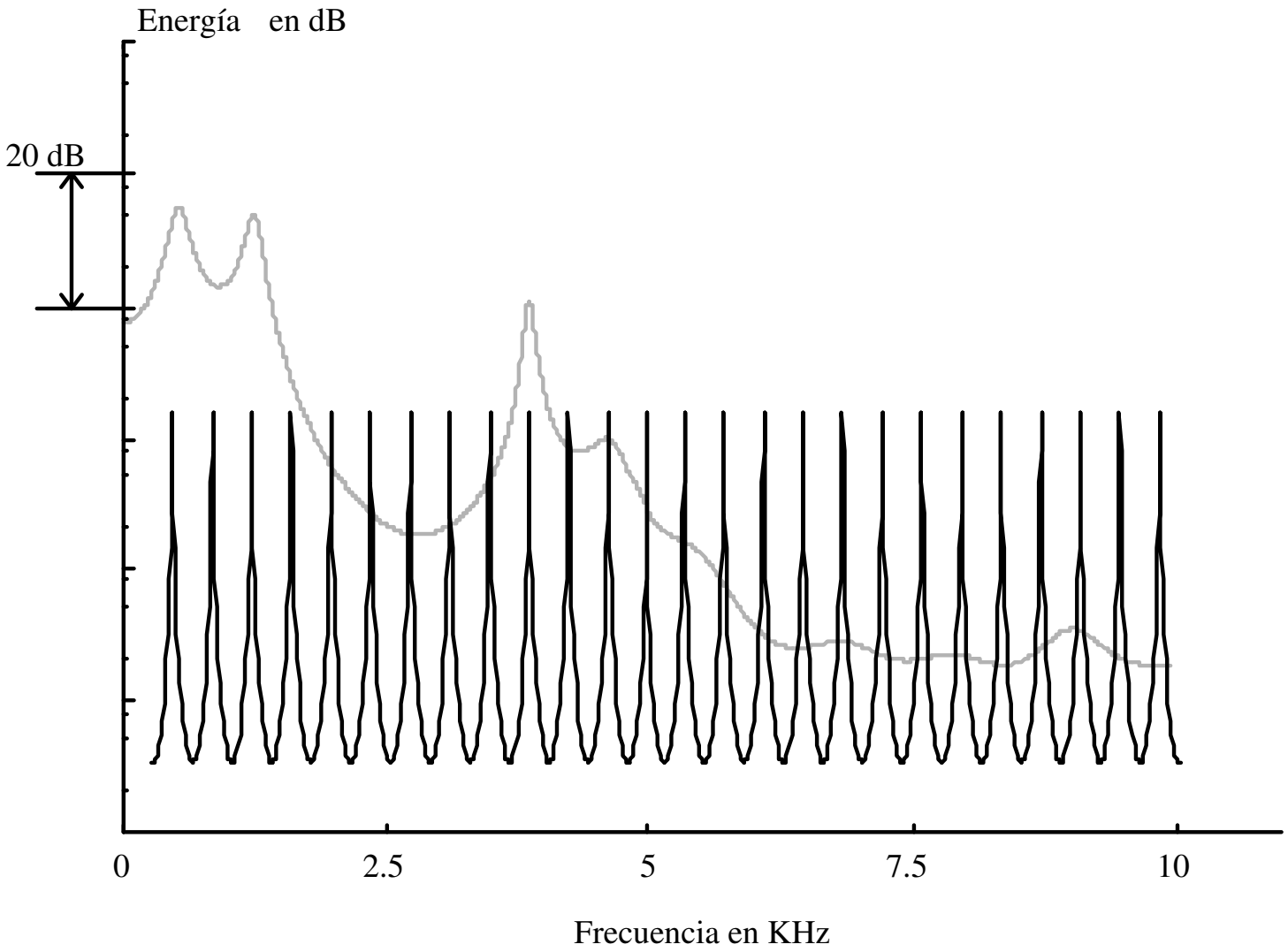
$$\hat{v}(m) = \mathcal{T}_F^{-1}\left\{\log|G(k)|\right\} + \mathcal{T}_F^{-1}\left\{\log|H(k)|\right\}$$

G y H ocupan partes diferentes del eje de cuefrencias. Podemos separar la parte que varía rápidamente (correspondiente a la excitación del tracto vocal) de la que varía lentamente (la respuesta en frecuencia del tracto).

Fuentes y modificadores de sonido en el espectro

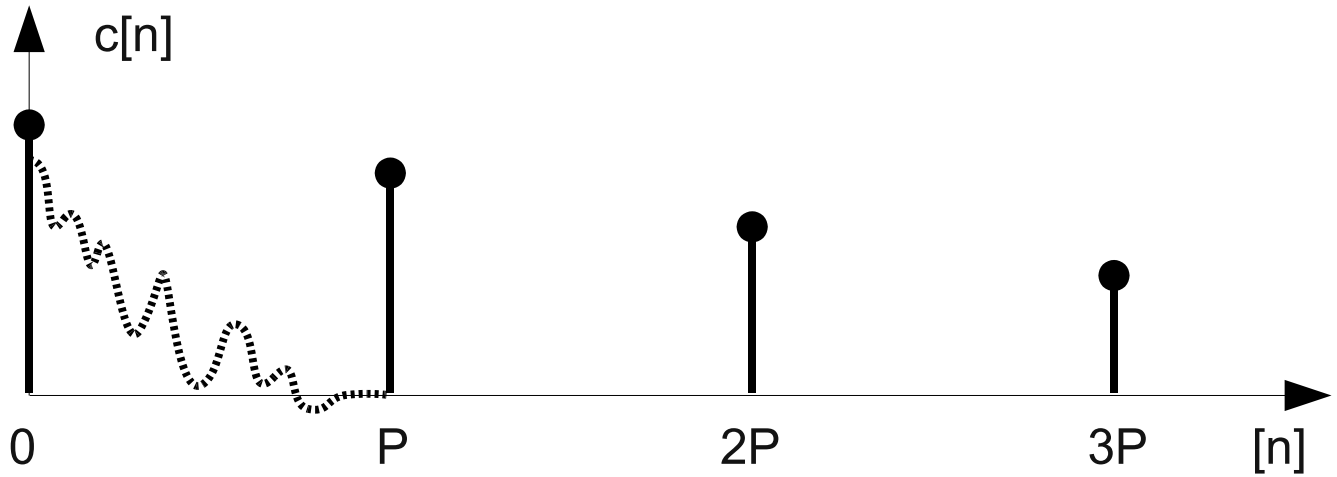


Fuentes y modificadores de sonido en el espectro

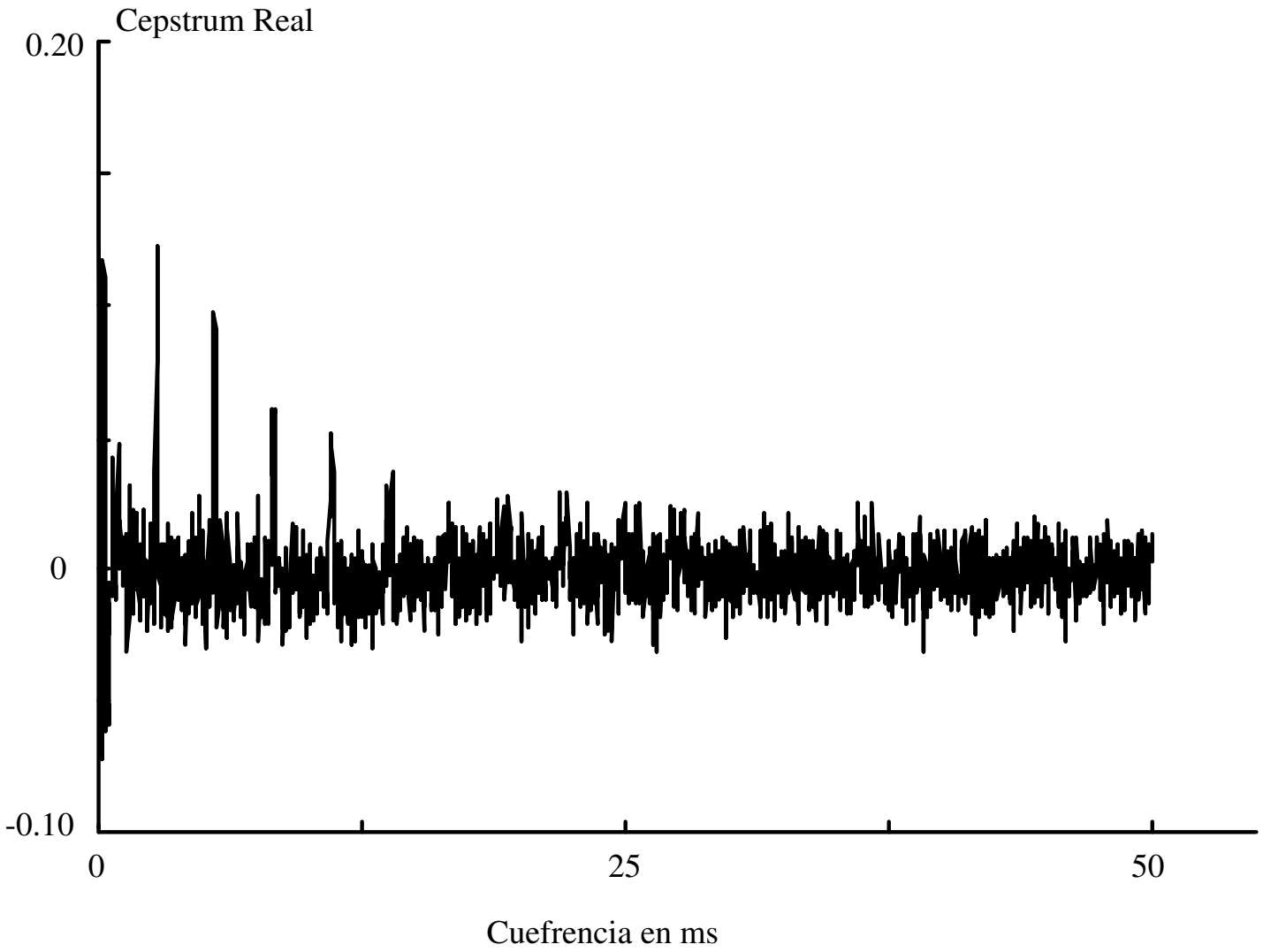


Cepstrum de una vocal

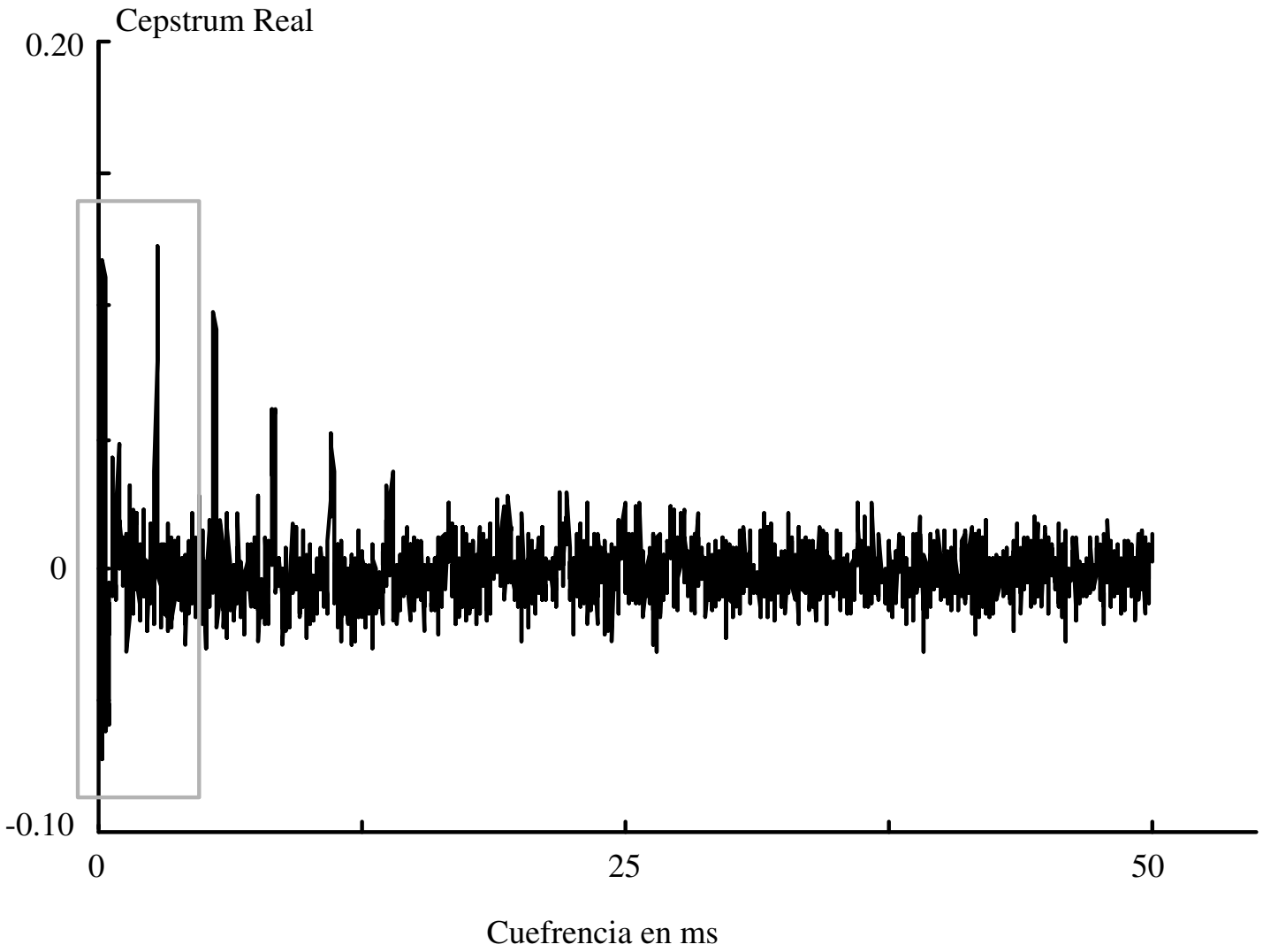
(esquema representativo)



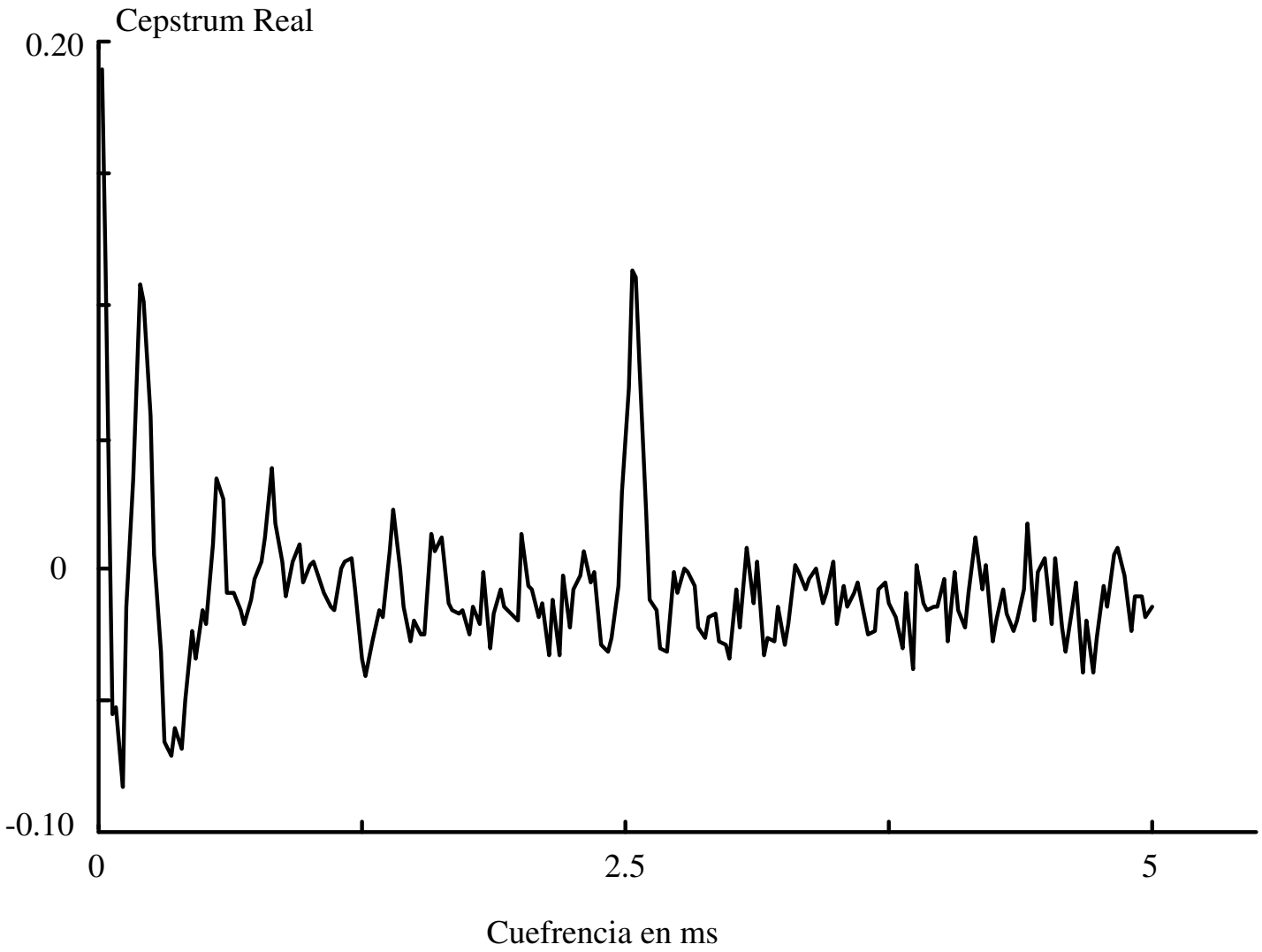
Cepstrum de una vocal



Cepstrum de una vocal



Cepstrum de una vocal



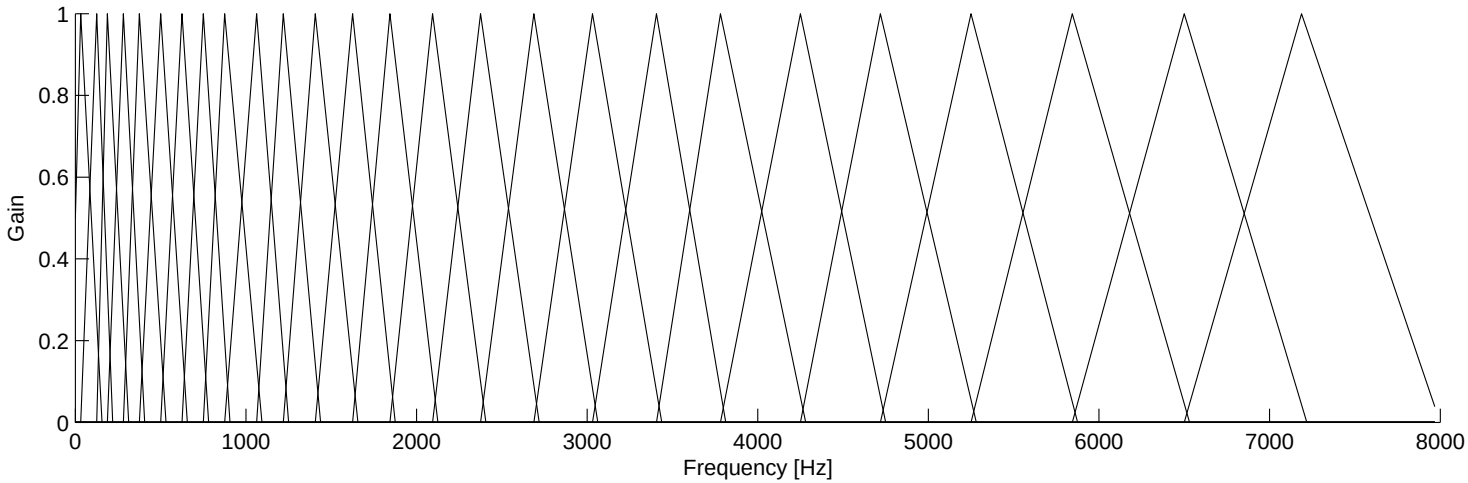
Coeficientes cepstrales en escala de mel

- Banco de filtros en escala de mel
- Integración por bandas del espectro
- Coeficientes de energía por cada banda
- Transformación inversa

Coeficientes cepstrales en escala de mel

Escala de mel

$$F_{mel} = 1000 \log_2 \left(1 + \frac{F_{Hz}}{1000}\right)$$



Coeficientes cepstrales en escala de mel

El espectro de magnitud

$$X[k] = \log_e |TDF\{x[n]\}|,$$

es integrado en bandas

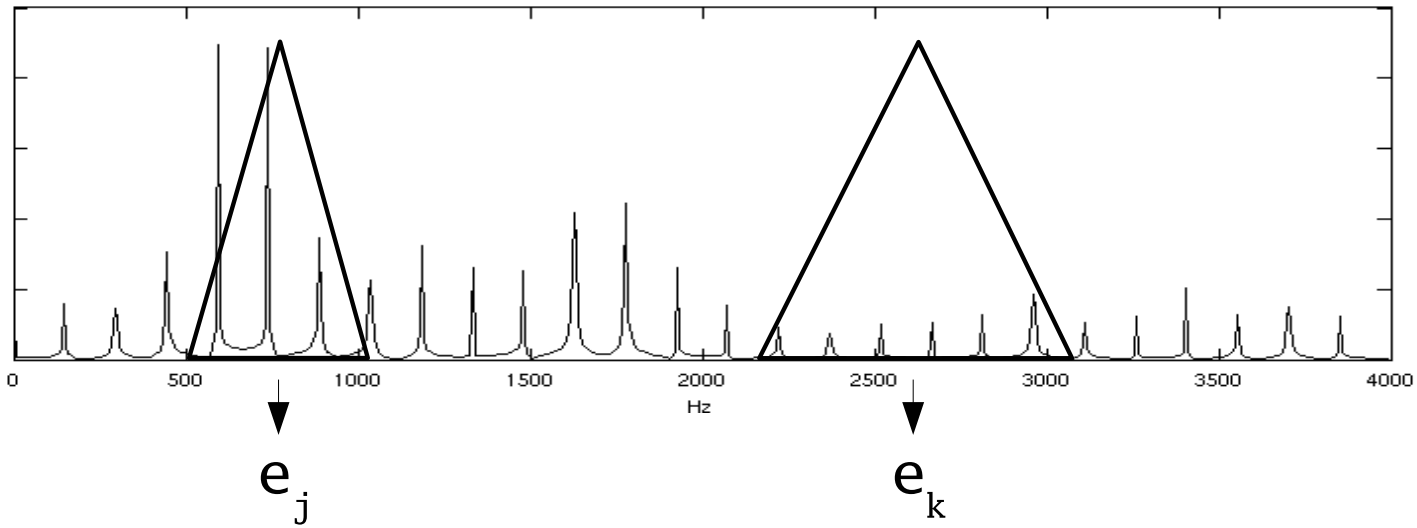
$$U[i] = \sum_k W_i[k]X[k],$$

y luego se calcula la transformada inversa

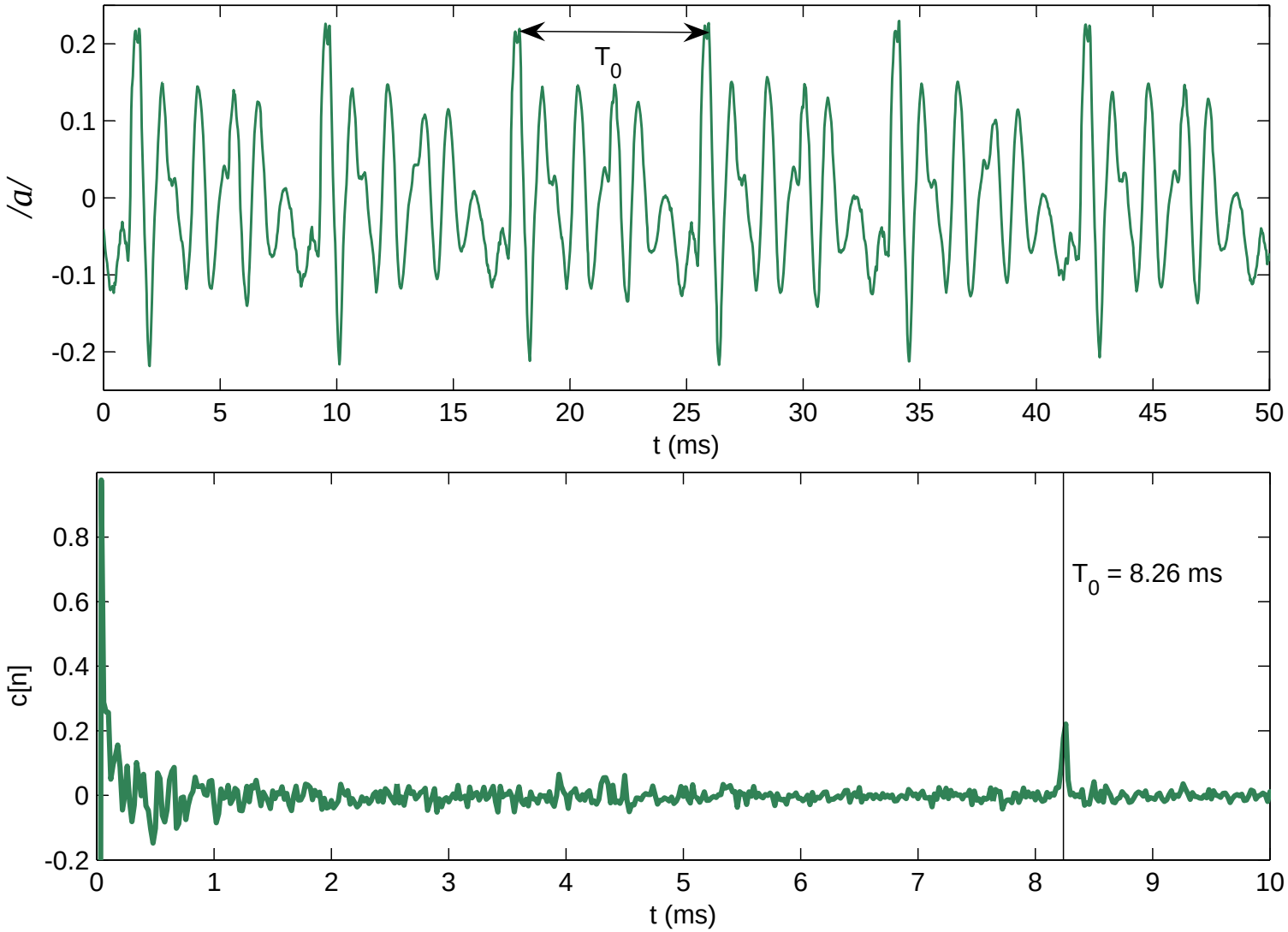
$$C = TDFI\{U\}.$$

Coeficientes cepstrales en escala de mel

Integración por bandas



Estimación de F0 por cepstrum



Estimación de F0 por autocorrelación

Bibliografía básica

- L. R. Rabiner y B. Gold, Theory and Application of Digital Signal Processing, Prentice Hall, 1975.
  Secciones: 12.1, 12.2, 12.3 y 12.13.
- J. R. Deller, J. G. Proakis, J. H. Hansen, Discrete-Time Processing of Speech Signals, Prentice Hall, 1993.
  Secciones: 4.1, 4.2.1, 4.2.2, 6.1 y 6.2.
  → Error en la figura 6.3 (c), pp 361.
- H.L. Rufiner, "Análisis y modelado digital de la voz: Técnicas recientes y aplicaciones",
  Editorial UNL, 2009. (Capítulo 3).
- J. Makhoul, "Linear Prediction: A Tuturial Review," Proc. IEEE, vol 63, no. 4,
  páginas 561-580, 1975.

Bibliografía básica

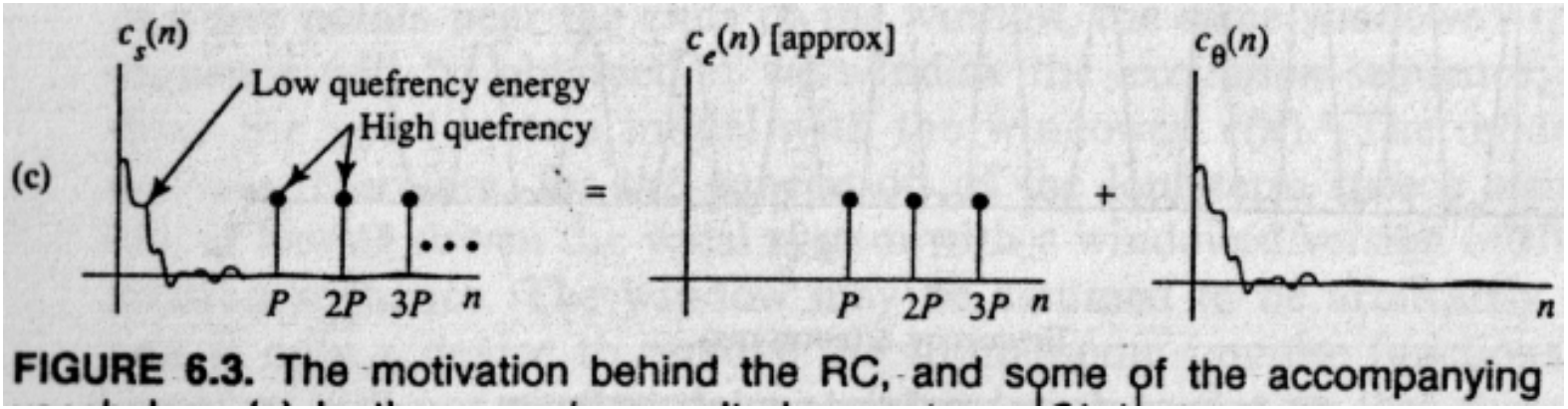
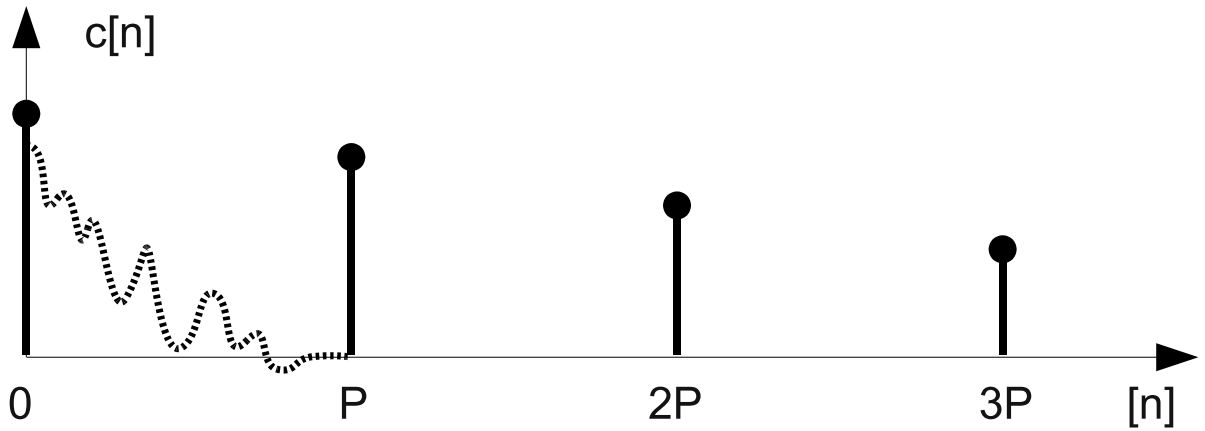


FIGURE 6.3. The motivation behind the RC, and some of the accompanying



Bibliografía básica

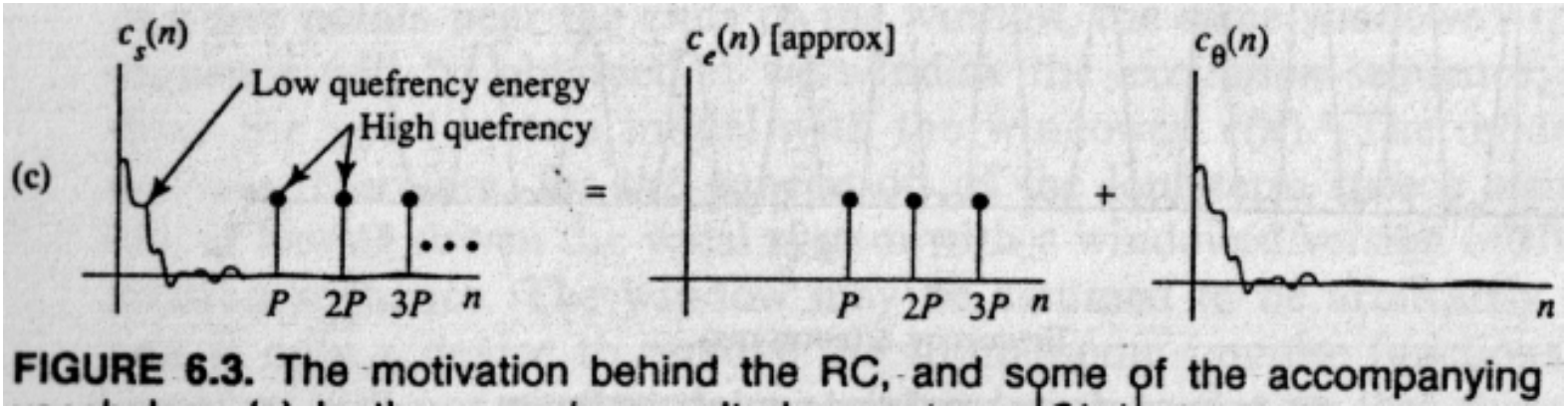


FIGURE 6.3. The motivation behind the RC, and some of the accompanying