

Reconocimiento automático del habla III

**Extracción de
características
(dominio del tiempo)**

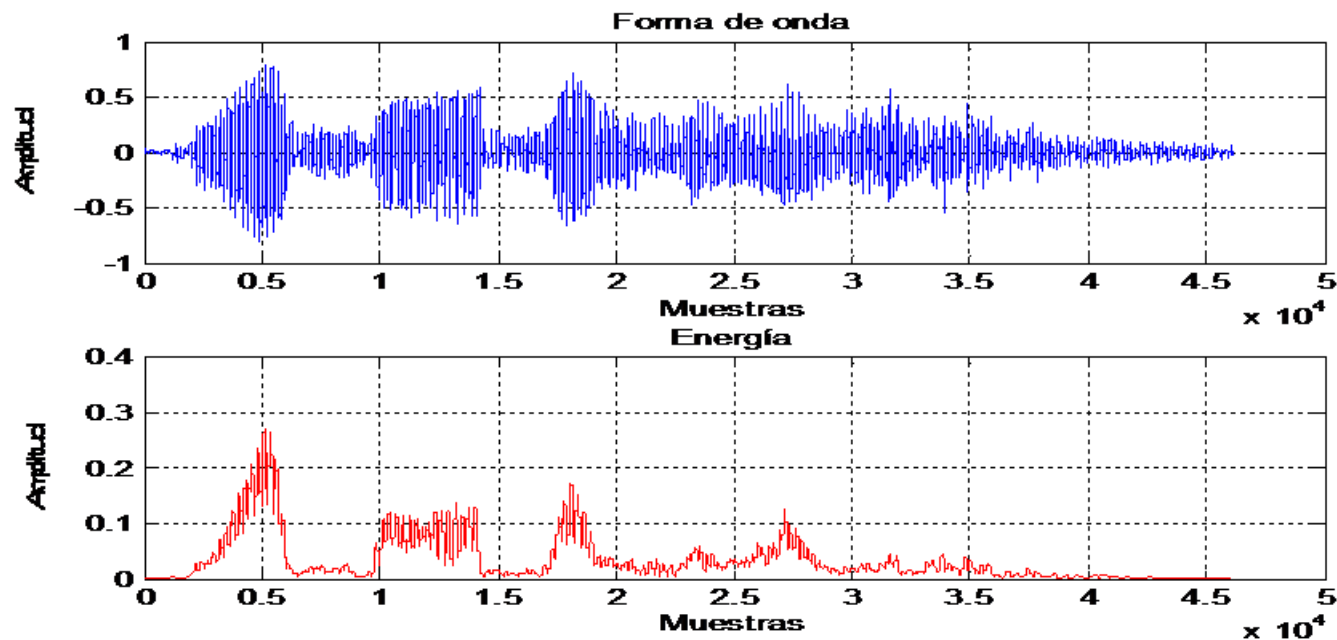
Extracción de características (Energía localizada)

$$E[m] = \sum_{n=-\infty}^{+\infty} [x[n]w[n-m]]^2 = \sum_{n=0}^{N-1} x[n]^2 w[n-m]^2$$

- Nos permite distinguir con fiabilidad:
 - ◆ La existencia de voz del silencio.
 - ◆ Los sonidos sordos de los sonoros.

Extracción de características (Energía localizada)

$$E[m] = \sum_{n=-\infty}^{+\infty} [x[n]w[n-m]]^2 = \sum_{n=0}^{N-1} x[n]^2 w[n-m]^2$$



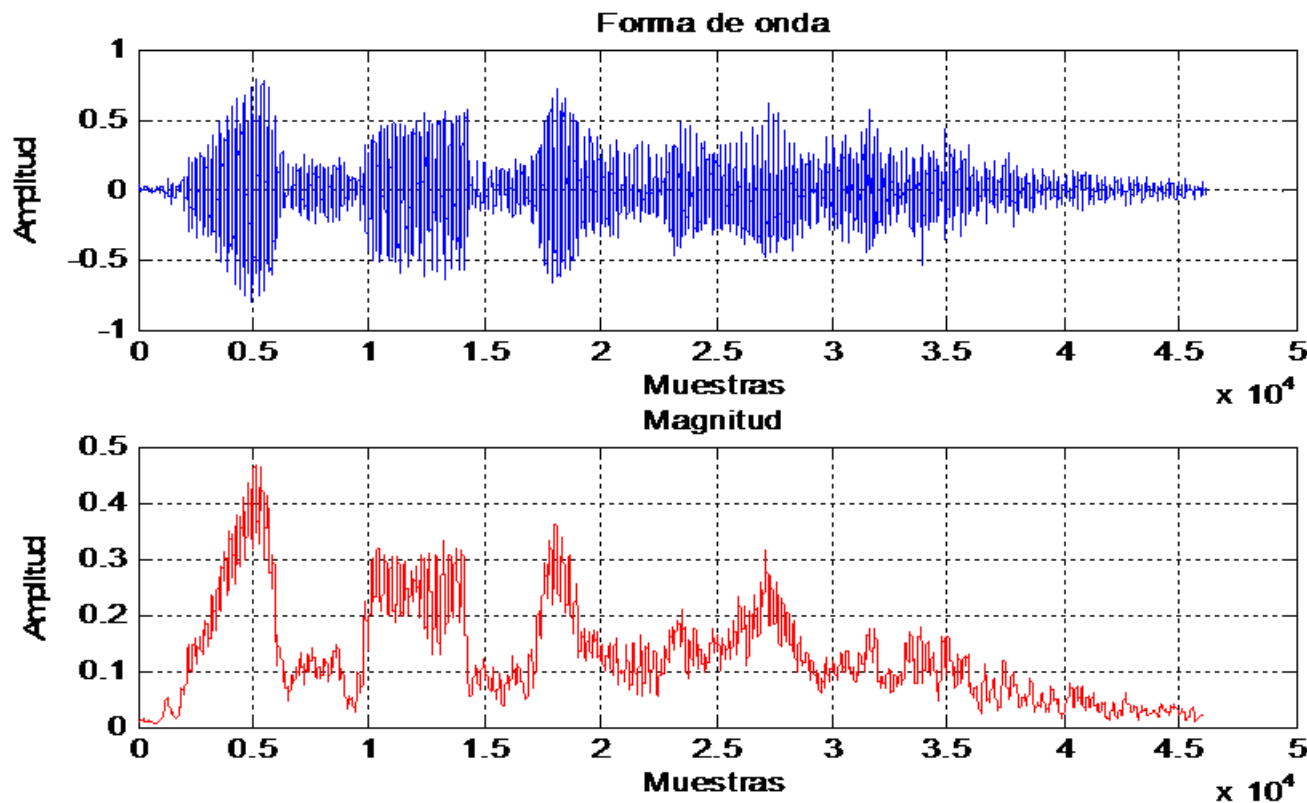
Extracción de características (Magnitud)

$$M[m] = \sum_{n=0}^{N-1} |x[n]| w[n-m]$$

- Parámetro alternativo a la Energía.
- Ventajas con respecto a la energía:
 - ◆ Menor complejidad.
 - ◆ Menor margen dinámico (muestras con valores de amplitud elevado pueden desvirtuar el valor de la energía al estar elevado al cuadrado)

Extracción de características (Magnitud)

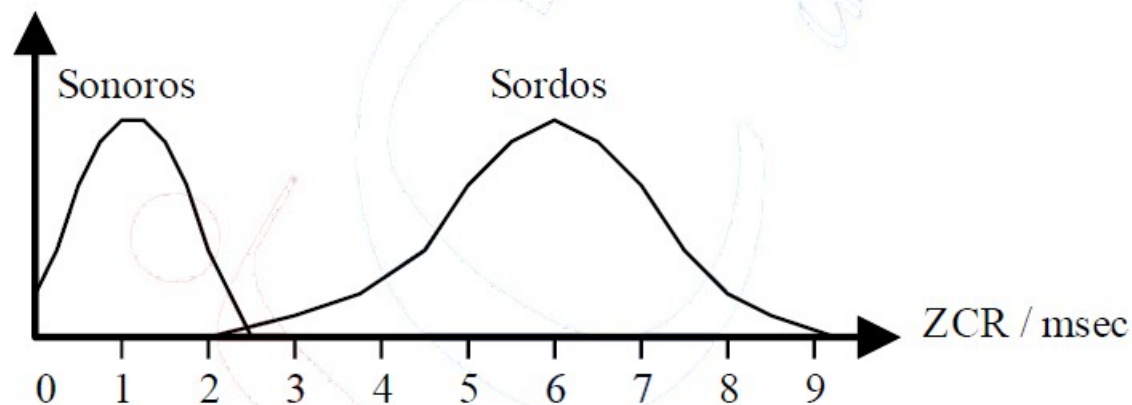
$$M[m] = \sum_{n=0}^{N-1} |x[n]| w[n-m]$$



Extracción de características (Tasa de cruces por cero)

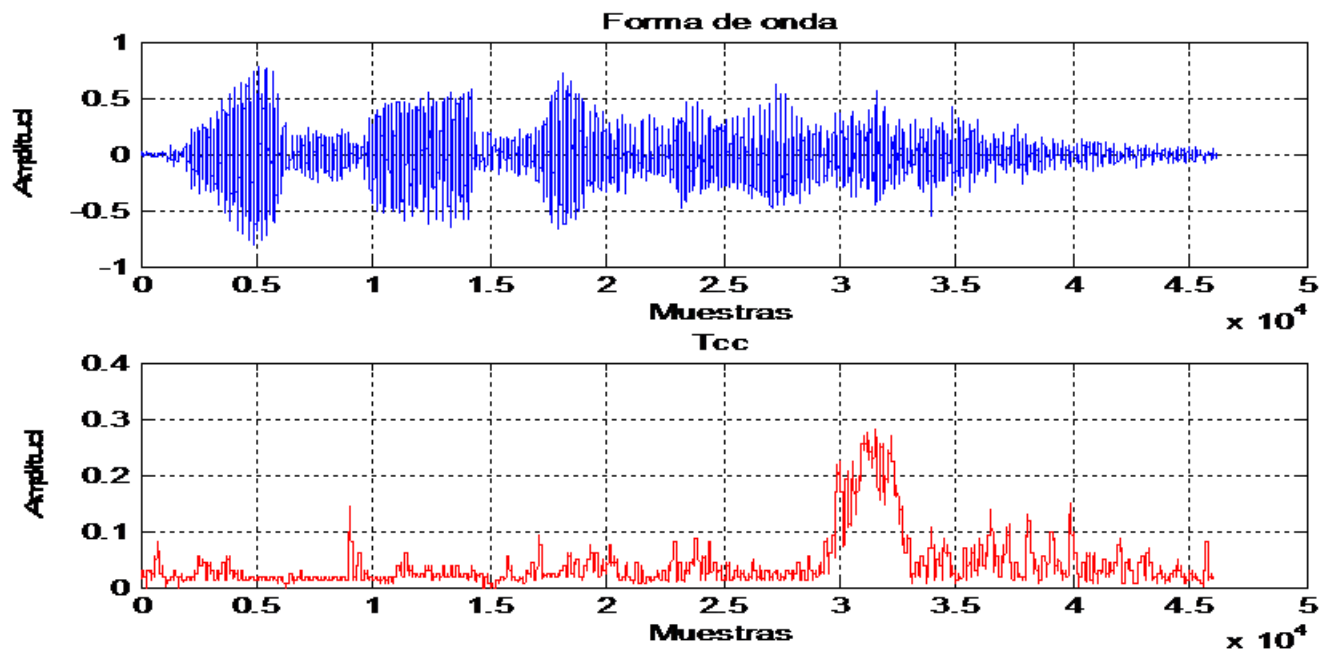
$$Tcc[m] = \frac{1}{N} \sum_n \frac{1}{2} |\text{sgn}(x[n]) - \text{sgn}(x[n-1])| w(m-n)$$

- Da una idea del contenido en frecuencia de una señal.
- Las señales sonoras dan una tasa menor que las señales sordas.



Extracción de características (Tasa de cruces por cero)

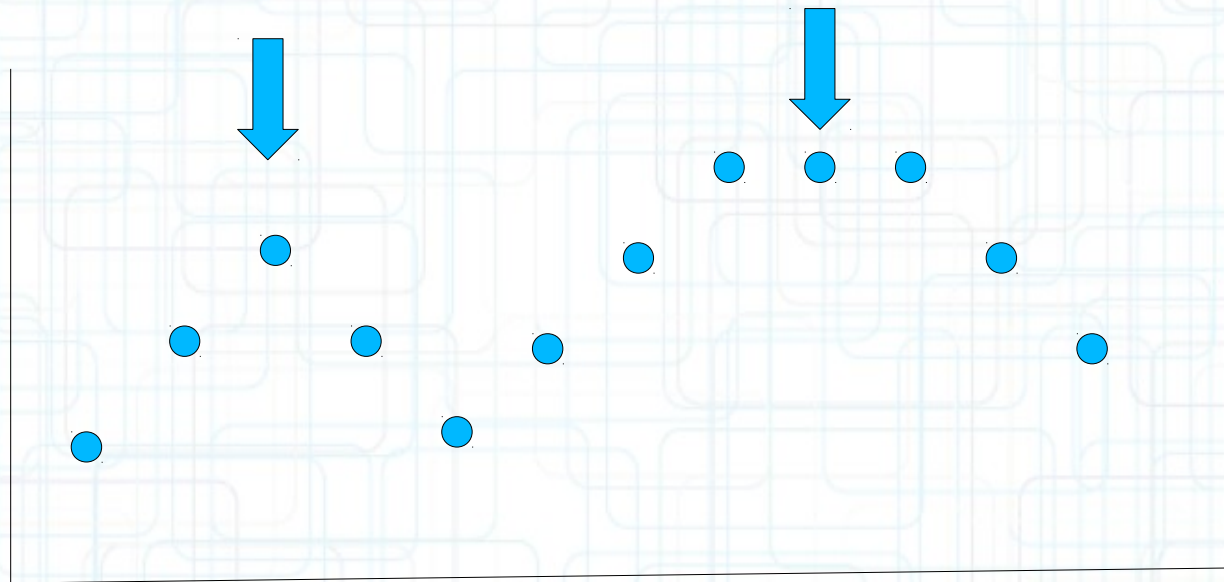
$$Tcc[m] = \frac{1}{N} \sum_n \frac{1}{2} |\text{sgn}(x[n]) - \text{sgn}(x[n-1])| w(m-n)$$



Extracción de características (Cálculo de máximos/mínimos locales)

- Busca el número de máximos/mínimos locales en una ventana.
- Existe un máximo local cuando una muestra tiene un valor superior a la anterior o posterior.
- Si un conjunto de muestras tuvieran el mismo valor y fueran superiores a las anteriores y posteriores, la muestra que se encuentra en medio corresponde al máximo local.

Extracción de características (Cálculo de máximos/mínimos locales)



Extracción de características (Autocorrelación)

$$R_m[k] = \sum_{n=0}^{N-1} \{w[m-n]x[n]\}\{w[m-(n+k)]x[n+k]\}$$

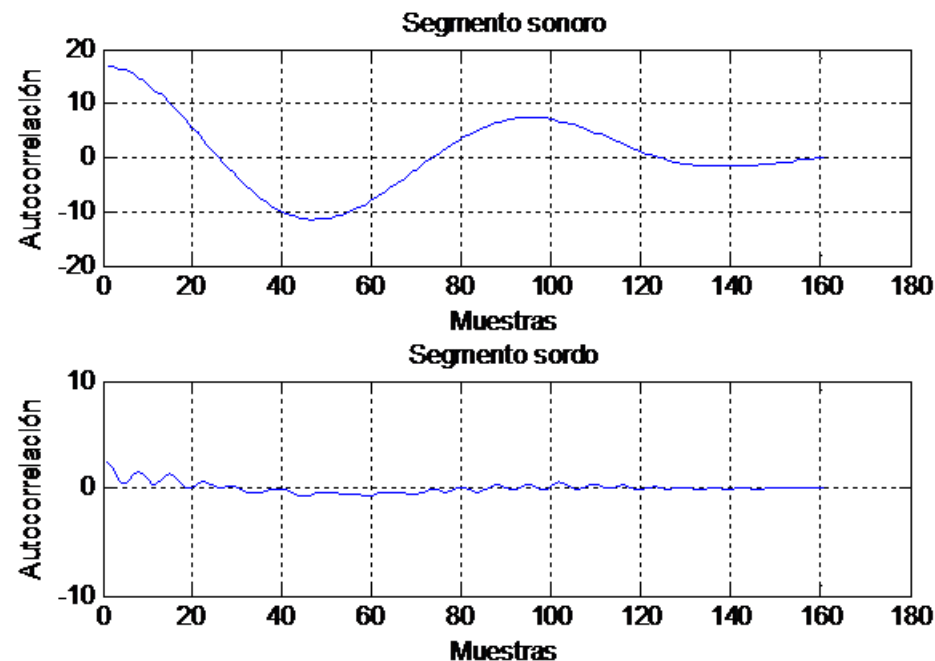
$k = 0, 1, 2, \dots, p.$

- Es una función par.
- Tiene un máximo en $k = 0$, que corresponde con el valor de la energía.
- La autocorrelación de una señal periódica es periódica.
- Para desplazamientos de k igual al período de la señal, la autocorrelación tiene máximos locales.

Extracción de características (Autocorrelación)

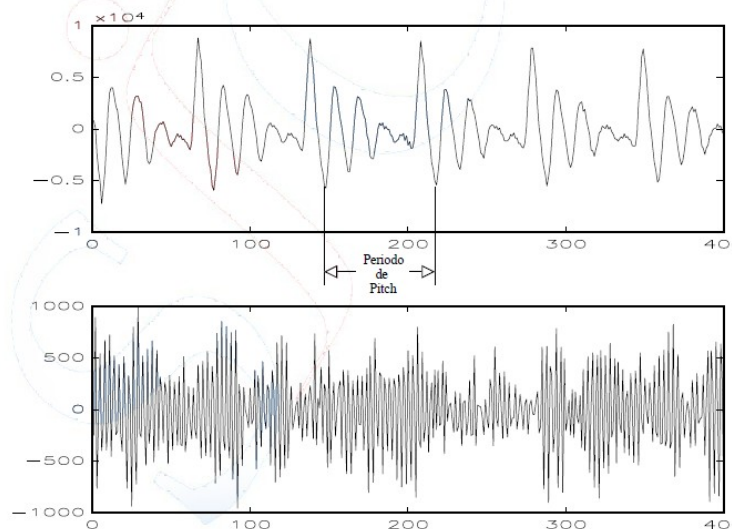
$$R_m[k] = \sum_{n=0}^{N-1} \{w[m-n]x[n]\} \{w[m-(n+k)]x[n+k]\}$$

$$k = 0, 1, 2, \dots p.$$



Extracción de características (Pitch)

- La frecuencia fundamental de una señal, o *pitch*, se define como la frecuencia aparente que tiene la señal.
- Las señales sonoras tienen un pitch muy claro, sin embargo, las señales sordas carecen de periodicidad.

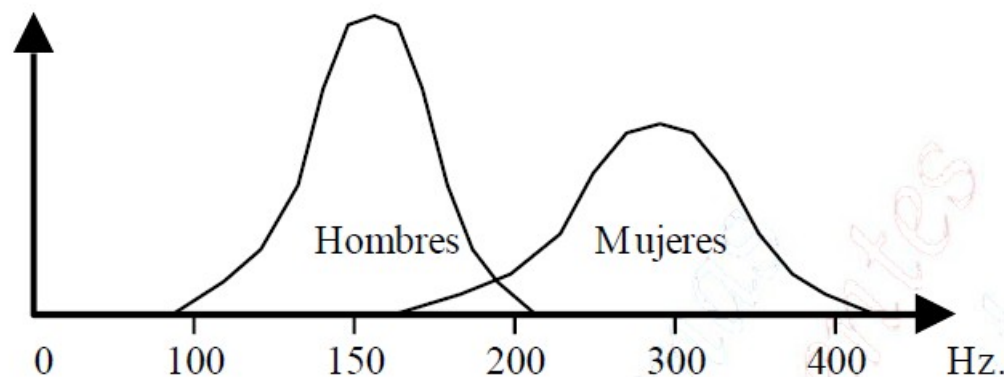


Señal sonora

Señal sorda

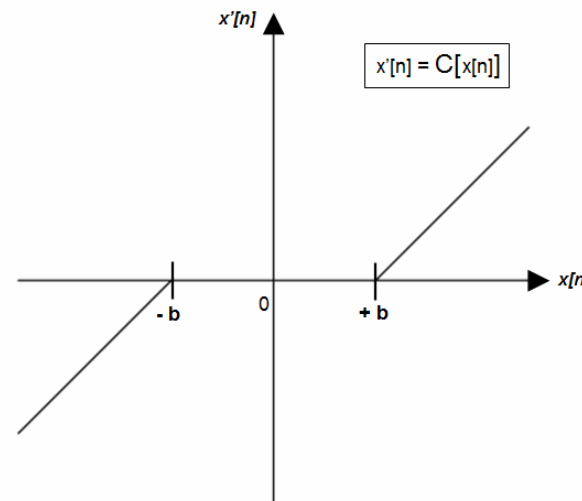
Extracción de características (Pitch)

- El pitch varía mucho entre diferentes locutores (incluso para el mismo locutor).
- Junto con la energía son factores determinantes en la entonación de las frases.
- Es una magnitud muy fiable para la discriminación hombre-mujer:

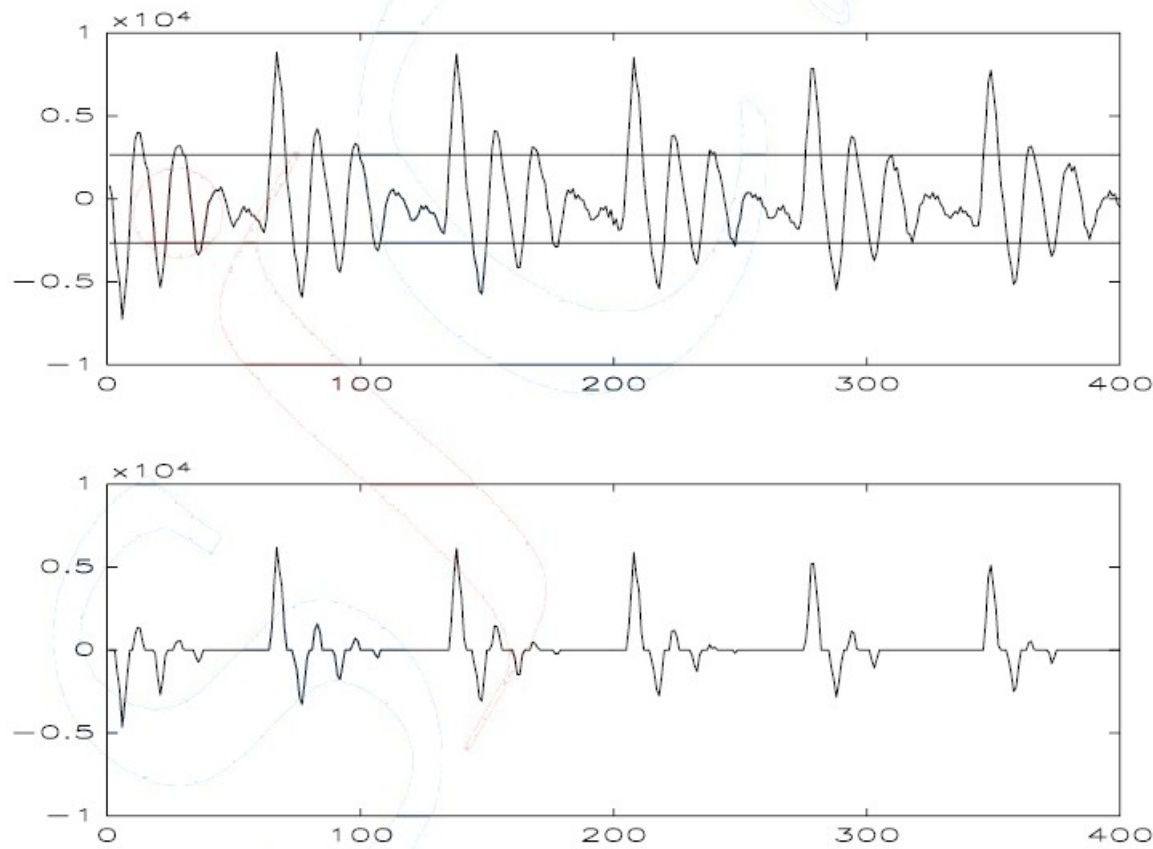


Extracción de características (Pitch)

- Estimación del período del *pitch* mediante autocorrelación:
 1. Aplicar función de transformación Center-Clipping para eliminar toda la señal de entrada que no sobrepase un determinado umbral (suele ser un 30% del máximo de la señal).



Extracción de características (Pitch)

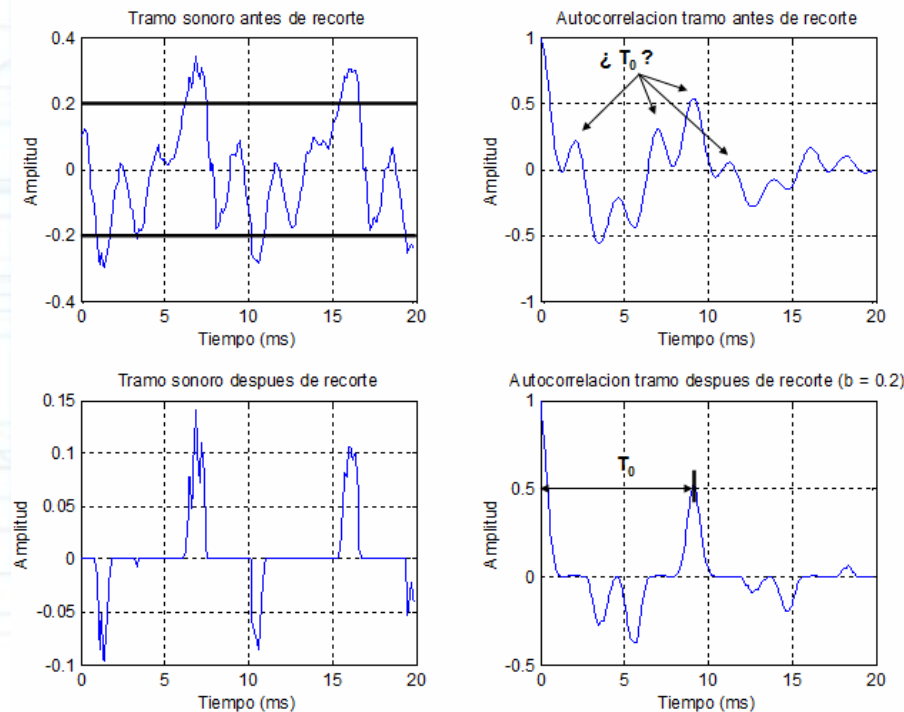


Center
Clipping

Extracción de características (Pitch)

- Estimación del período del *pitch* mediante autocorrelación:

2. Autocorrelación de la señal recortada.



Extracción de características (Pitch)

- Estimación del período del *pitch* mediante AMDF:
- AMDF, *Average Magnitude Difference Function*
 - Estima del pitch empleando la Magnitud en vez de la correlación
 - Menor complejidad y coste computacional
 - En este caso en vez de buscar máximos se deben buscar mínimos

$$AMDF[m, \tau] = \sum_{n=m+\tau}^{m+N-1} |s[n]w[n-m] - s[n-\tau]w[n-m-\tau]|$$

Extracción de características (Pitch)

- Estimación del período del *pitch* mediante AMDF:

