

Procesamiento digital de señales de audio

Procesamiento en el dominio del tiempo

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

1 Introducción

① Introducción

② Energía y magnitud de tiempo corto

- ① Introducción
- ② Energía y magnitud de tiempo corto
- ③ Tasa de cruces por cero

- ① Introducción
- ② Energía y magnitud de tiempo corto
- ③ Tasa de cruces por cero
- ④ Estimación de frecuencia fundamental
 - Autocorrelación
 - Función diferencia

① Introducción

② Energía y magnitud de tiempo corto

③ Tasa de cruces por cero

④ Estimación de frecuencia fundamental

Autocorrelación

Función diferencia

Procesamiento en el dominio del tiempo

Introducción

- procesamiento para estimar parámetros de análisis y realizar diversas tareas (voz/silencio, sonoro/sordo, formantes, fonemas, f_0 , etc.)

Procesamiento en el dominio del tiempo

Introducción

- procesamiento para estimar parámetros de análisis y realizar diversas tareas (voz/silencio, sonoro/sordo, formantes, fonemas, f_0 , etc.)
- **procesamiento en el tiempo**: operaciones sobre la forma de onda

Procesamiento en el dominio del tiempo

Introducción

- procesamiento para estimar parámetros de análisis y realizar diversas tareas (voz/silencio, sonoro/sordo, formantes, fonemas, f_0 , etc.)
- **procesamiento en el tiempo**: operaciones sobre la forma de onda
- **procesamiento en frecuencia**: sobre alguna representación espectral

Procesamiento en el dominio del tiempo

Introducción

- procesamiento para estimar parámetros de análisis y realizar diversas tareas (voz/silencio, sonoro/sordo, formantes, fonemas, f_0 , etc.)
- **procesamiento en el tiempo**: operaciones sobre la forma de onda
- **procesamiento en frecuencia**: sobre alguna representación espectral

Procesamiento en el tiempo

- las propiedades de la señal varían con el tiempo (e.g. 20 sonidos/seg)

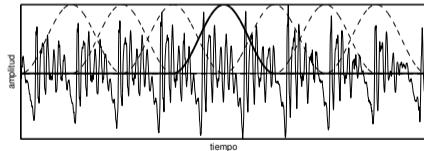
Procesamiento en el dominio del tiempo

Introducción

- procesamiento para estimar parámetros de análisis y realizar diversas tareas (voz/silencio, sonoro/sordo, formantes, fonemas, f_0 , etc.)
- **procesamiento en el tiempo**: operaciones sobre la forma de onda
- **procesamiento en frecuencia**: sobre alguna representación espectral

Procesamiento en el tiempo

- las propiedades de la señal varían con el tiempo (e.g. 20 sonidos/seg)
- **solución de compromiso**: bloques solapados de tiempo-corto



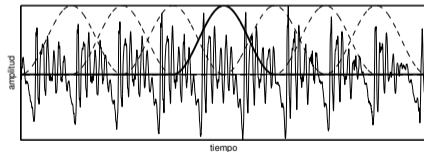
Procesamiento en el dominio del tiempo

Introducción

- procesamiento para estimar parámetros de análisis y realizar diversas tareas (voz/silencio, sonoro/sordo, formantes, fonemas, f_0 , etc.)
- **procesamiento en el tiempo**: operaciones sobre la forma de onda
- **procesamiento en frecuencia**: sobre alguna representación espectral

Procesamiento en el tiempo

- las propiedades de la señal varían con el tiempo (e.g. 20 sonidos/seg)
- **solución de compromiso**: bloques solapados de tiempo-corto
 - incertidumbre con estimaciones de **tiempo-corto**



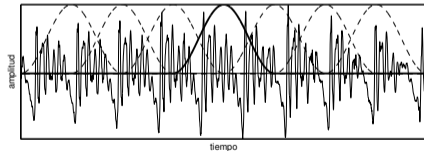
Procesamiento en el dominio del tiempo

Introducción

- procesamiento para estimar parámetros de análisis y realizar diversas tareas (voz/silencio, sonoro/sordo, formantes, fonemas, f_0 , etc.)
- **procesamiento en el tiempo**: operaciones sobre la forma de onda
- **procesamiento en frecuencia**: sobre alguna representación espectral

Procesamiento en el tiempo

- las propiedades de la señal varían con el tiempo (e.g. 20 sonidos/seg)
- **solución de compromiso**: bloques solapados de tiempo-corto
 - incertidumbre con estimaciones de **tiempo-corto**
5-20 ms: escasez de datos,



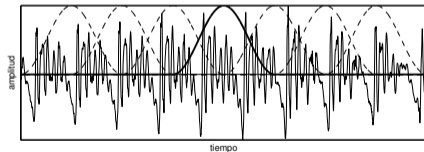
Procesamiento en el dominio del tiempo

Introducción

- procesamiento para estimar parámetros de análisis y realizar diversas tareas (voz/silencio, sonoro/sordo, formantes, fonemas, f_0 , etc.)
- **procesamiento en el tiempo**: operaciones sobre la forma de onda
- **procesamiento en frecuencia**: sobre alguna representación espectral

Procesamiento en el tiempo

- las propiedades de la señal varían con el tiempo (e.g. 20 sonidos/seg)
- **solución de compromiso**: bloques solapados de tiempo-corto
 - incertidumbre con estimaciones de **tiempo-corto**
 - 5-20 ms: escasez de datos,
 - 20-100 ms: cambios rápidos (transitorios),



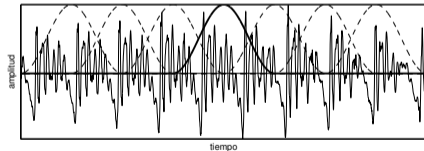
Procesamiento en el dominio del tiempo

Introducción

- procesamiento para estimar parámetros de análisis y realizar diversas tareas (voz/silencio, sonoro/sordo, formantes, fonemas, f_0 , etc.)
- **procesamiento en el tiempo**: operaciones sobre la forma de onda
- **procesamiento en frecuencia**: sobre alguna representación espectral

Procesamiento en el tiempo

- las propiedades de la señal varían con el tiempo (e.g. 20 sonidos/seg)
- **solución de compromiso**: bloques solapados de tiempo-corto
 - incertidumbre con estimaciones de **tiempo-corto**
 - 5-20 ms: escasez de datos,
 - 20-100 ms: cambios rápidos (transitorios),
 - 100-500 ms: grandes cambios (f_0 , formantes, etc)



Procesamiento en el dominio del tiempo

Esquema general de procesamiento de tiempo-corto

- la mayoría de las técnicas pueden expresarse como:

$$Q_{\hat{n}} = \left(\sum_{m=-\infty}^{\infty} T\{x[m]\}w[n-m] \right) \Big|_{n=\hat{n}}$$

Procesamiento en el dominio del tiempo

Esquema general de procesamiento de tiempo-corto

- la mayoría de las técnicas pueden expresarse como:

$$Q_{\hat{n}} = \left(\sum_{m=-\infty}^{\infty} T\{x[m]\}w[n-m] \right) \Big|_{n=\hat{n}}$$

con $T\{\}$ transformación lineal o no-lineal, y $w[n]$ ventana de análisis

Procesamiento en el dominio del tiempo

Esquema general de procesamiento de tiempo-corto

- la mayoría de las técnicas pueden expresarse como:

$$Q_{\hat{n}} = \left(\sum_{m=-\infty}^{\infty} T\{x[m]\} w[n-m] \right) \Big|_{n=\hat{n}}$$

con $T\{\}$ transformación lineal o no-lineal, y $w[n]$ ventana de análisis

- $Q_{\hat{n}}$: promedio local ponderado de $T\{x[n]\}$ en tiempo \hat{n}

Procesamiento en el dominio del tiempo

Esquema general de procesamiento de tiempo-corto

- la mayoría de las técnicas pueden expresarse como:

$$Q_{\hat{n}} = \left(\sum_{m=-\infty}^{\infty} T\{x[m]\} w[n-m] \right) \Big|_{n=\hat{n}}$$

con $T\{\}$ transformación lineal o no-lineal, y $w[n]$ ventana de análisis

- $Q_{\hat{n}}$: promedio local ponderado de $T\{x[n]\}$ en tiempo \hat{n}
- efecto del eventanado:

Procesamiento en el dominio del tiempo

Esquema general de procesamiento de tiempo-corto

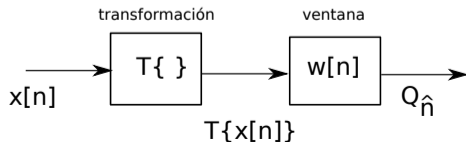
- la mayoría de las técnicas pueden expresarse como:

$$Q_{\hat{n}} = \left(\sum_{m=-\infty}^{\infty} T\{x[m]\}w[n-m] \right) \Big|_{n=\hat{n}}$$

con $T\{\}$ transformación lineal o no-lineal, y $w[n]$ ventana de análisis

- $Q_{\hat{n}}$: promedio local ponderado de $T\{x[n]\}$ en tiempo \hat{n}
- efecto del eventanado:

$$Q_{\hat{n}} = T\{x[n]\} * w[n] \Big|_{n=\hat{n}}$$



Procesamiento en el dominio del tiempo

Esquema general de procesamiento de tiempo-corto

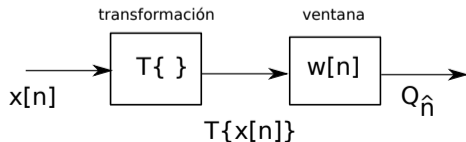
- la mayoría de las técnicas pueden expresarse como:

$$Q_{\hat{n}} = \left(\sum_{m=-\infty}^{\infty} T\{x[m]\}w[n-m] \right) \Big|_{n=\hat{n}}$$

con $T\{\}$ transformación lineal o no-lineal, y $w[n]$ ventana de análisis

- $Q_{\hat{n}}$: promedio local ponderado de $T\{x[n]\}$ en tiempo \hat{n}
- efecto del eventanado: $w[n]$ funciona como pasabajos sobre $T\{x[n]\}$

$$Q_{\hat{n}} = T\{x[n]\} * w[n] \Big|_{n=\hat{n}}$$



① Introducción

② Energía y magnitud de tiempo corto

③ Tasa de cruces por cero

④ Estimación de frecuencia fundamental

Autocorrelación

Función diferencia

Procesamiento en el dominio del tiempo

Energía de tiempo corto

- representación de las variaciones de amplitud:

Procesamiento en el dominio del tiempo

Energía de tiempo corto

- representación de las variaciones de amplitud:

$$E_n = \sum_{m=-\infty}^{\infty} (x[m]w[n-m])^2$$

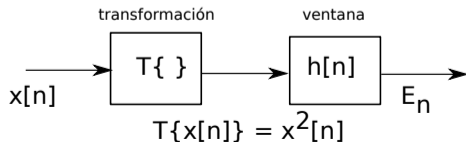
Procesamiento en el dominio del tiempo

Energía de tiempo corto

- representación de las variaciones de amplitud:

$$E_n = \sum_{m=-\infty}^{\infty} (x[m]w[n-m])^2$$

$$E_n = \sum_{m=-\infty}^{\infty} x^2[m]h[n-m]$$



Procesamiento en el dominio del tiempo

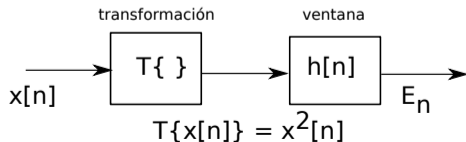
Energía de tiempo corto

- representación de las variaciones de amplitud:

$$E_n = \sum_{m=-\infty}^{\infty} (x[m]w[n-m])^2$$

$$E_n = \sum_{m=-\infty}^{\infty} x^2[m]h[n-m]$$

donde $T\{x[n]\} = x^2[n]$ y $h[n] = w^2[n]$



Procesamiento en el dominio del tiempo

Efecto de la ventana

$$E_n = T\{x[n]\} * h[n]|_{n=\hat{n}} = x^2[n] * h[n]|_{n=\hat{n}}$$

Procesamiento en el dominio del tiempo

Efecto de la ventana

$$E_n = T\{x[n]\} * h[n]|_{n=\hat{n}} = x^2[n] * h[n]|_{n=\hat{n}}$$

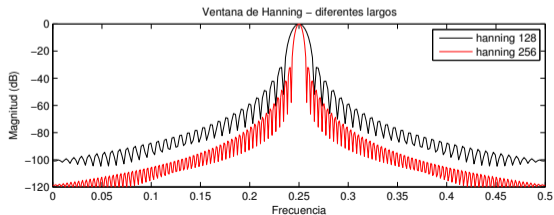
- ventana larga y de amplitud constante E_n cambia poco en el tiempo

Procesamiento en el dominio del tiempo

Efecto de la ventana

$$E_n = T\{x[n]\} * h[n]|_{n=\hat{n}} = x^2[n] * h[n]|_{n=\hat{n}}$$

- ventana larga y de amplitud constante E_n cambia poco en el tiempo
- **compromiso** habitual de representación de **tiempo-corto**

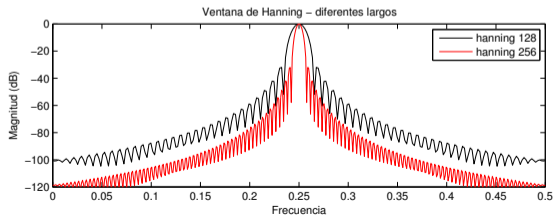


Procesamiento en el dominio del tiempo

Efecto de la ventana

$$E_n = T\{x[n]\} * h[n]|_{n=\hat{n}} = x^2[n] * h[n]|_{n=\hat{n}}$$

- ventana larga y de amplitud constante E_n cambia poco en el tiempo
- **compromiso** habitual de representación de **tiempo-corto**
 - ventana corta para responder a variaciones rápidas

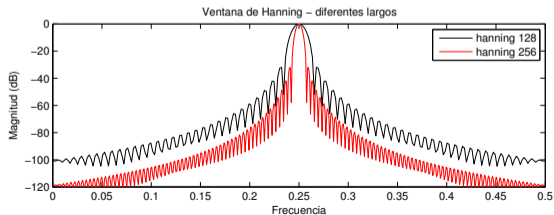


Procesamiento en el dominio del tiempo

Efecto de la ventana

$$E_n = T\{x[n]\} * h[n]|_{n=\hat{n}} = x^2[n] * h[n]|_{n=\hat{n}}$$

- ventana larga y de amplitud constante E_n cambia poco en el tiempo
- **compromiso** habitual de representación de **tiempo-corto**
 - ventana corta para responder a variaciones rápidas
 - ventana larga para un promedio adecuado y función suave

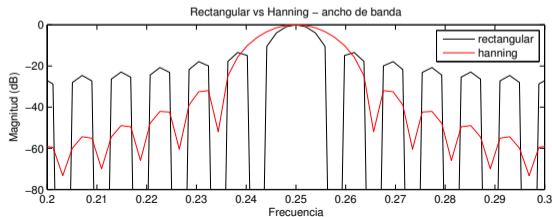
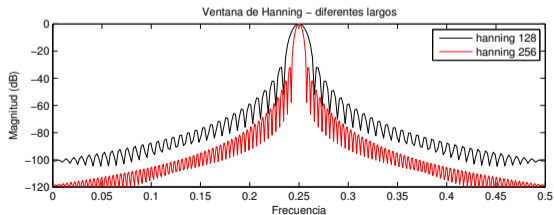


Procesamiento en el dominio del tiempo

Efecto de la ventana

$$E_n = T\{x[n]\} * h[n]|_{n=\hat{n}} = x^2[n] * h[n]|_{n=\hat{n}}$$

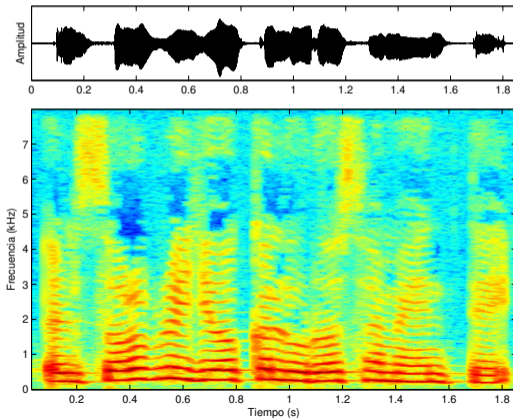
- ventana larga y de amplitud constante E_n cambia poco en el tiempo
- **compromiso** habitual de representación de **tiempo-corto**
 - ventana corta para responder a variaciones rápidas
 - ventana larga para un promedio adecuado y función suave
- para un mismo largo, tipo de ventana determina ancho de banda



Procesamiento en el dominio del tiempo

Energía de tiempo corto

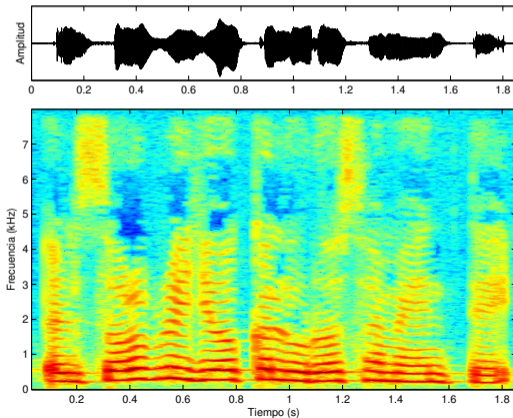
- distinguir voz de silencio (sólo para SNR alta)



Procesamiento en el dominio del tiempo

Energía de tiempo corto

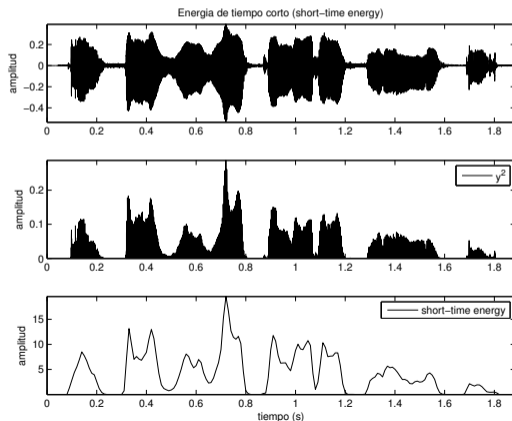
- distinguir voz de silencio (sólo para SNR alta)
- discriminar fonemas sonoros de sordos (principalmente a $f_s \leq 16$ kHz)



Procesamiento en el dominio del tiempo

Energía de tiempo corto

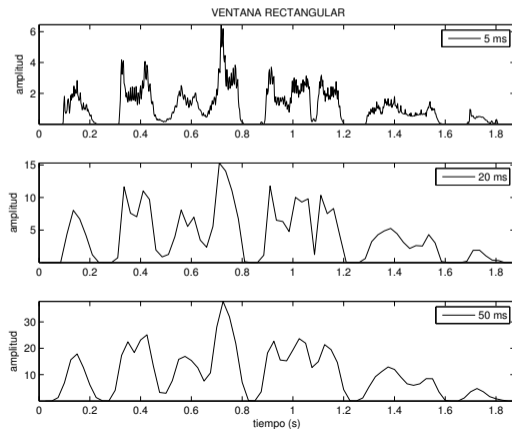
- distinguir voz de silencio (sólo para SNR alta)
- discriminar fonemas sonoros de sordos (principalmente a $f_s \leq 16$ kHz)



Procesamiento en el dominio del tiempo

Energía de tiempo corto

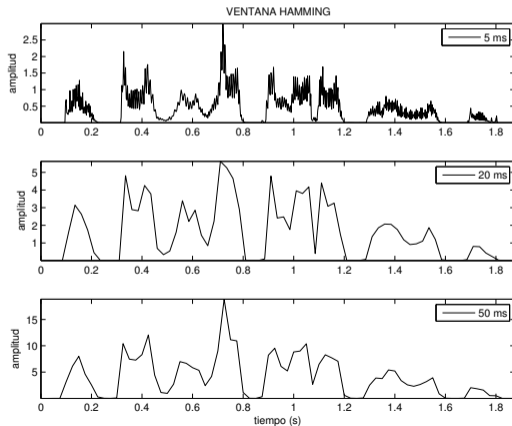
- distinguir voz de silencio (sólo para SNR alta)
- discriminar fonemas sonoros de sordos (principalmente a $f_s \leq 16$ kHz)



Procesamiento en el dominio del tiempo

Energía de tiempo corto

- distinguir voz de silencio (sólo para SNR alta)
- discriminar fonemas sonoros de sordos (principalmente a $f_s \leq 16$ kHz)



Procesamiento en el dominio del tiempo

Magnitud de tiempo corto

- E_n muy sensible a variaciones grandes de amplitud

Procesamiento en el dominio del tiempo

Magnitud de tiempo corto

- E_n muy sensible a variaciones grandes de amplitud
- alternativa: magnitud de tiempo corto

$$M_n = \sum_{m=-\infty}^{\infty} |x[m]| w[n - m]$$

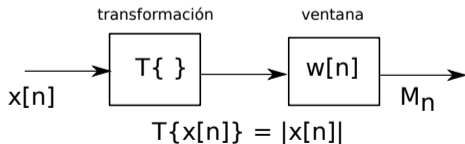
Procesamiento en el dominio del tiempo

Magnitud de tiempo corto

- E_n muy sensible a variaciones grandes de amplitud
- alternativa: magnitud de tiempo corto

$$M_n = \sum_{m=-\infty}^{\infty} |x[m]| w[n - m]$$

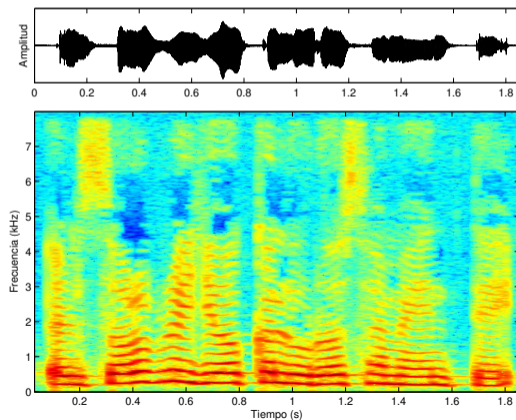
donde $T\{x[n]\} = |x[n]|$ y $h[n] = w[n]$



Procesamiento en el dominio del tiempo

Magnitud de tiempo corto

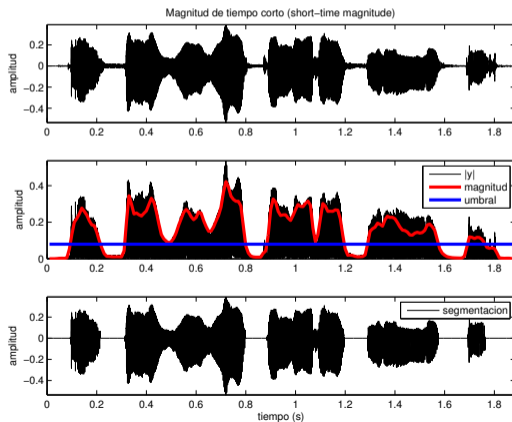
- ejemplo de segmentación de fonemas sonoros y sordos



Procesamiento en el dominio del tiempo

Magnitud de tiempo corto

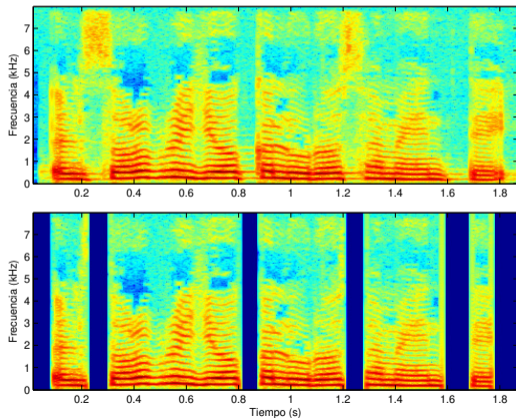
- ejemplo de segmentación de fonemas sonoros y sordos



Procesamiento en el dominio del tiempo

Magnitud de tiempo corto

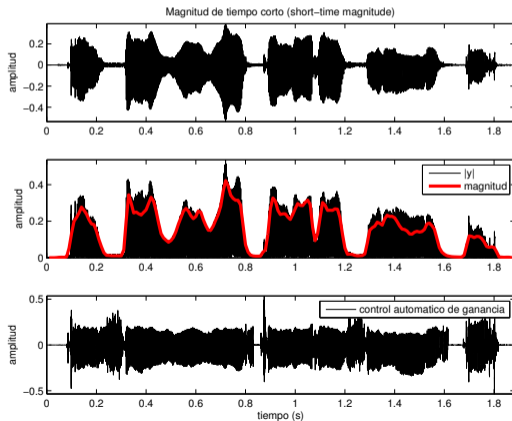
- ejemplo de segmentación de fonemas sonoros y sordos



Procesamiento en el dominio del tiempo

Magnitud de tiempo corto

- ejemplo de segmentación de fonemas sonoros y sordos
- ejemplo de control automático de ganancia (AGC)

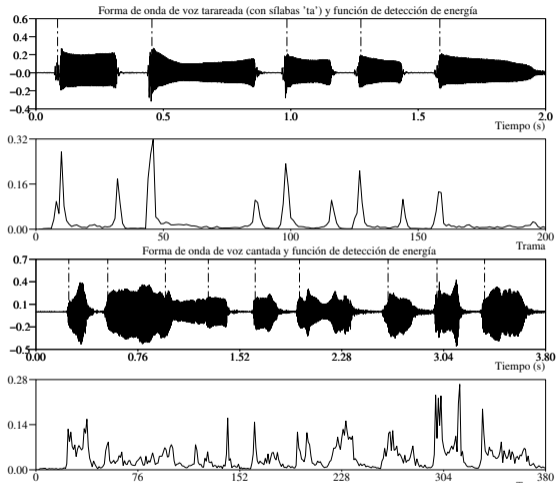


Procesamiento en el dominio del tiempo

Magnitud de tiempo corto

aplicación: segmentación
de audio en notas

- derivada de la magnitud de tiempo corto: $\frac{d}{dt} M_n$ [Dixon, 2001]

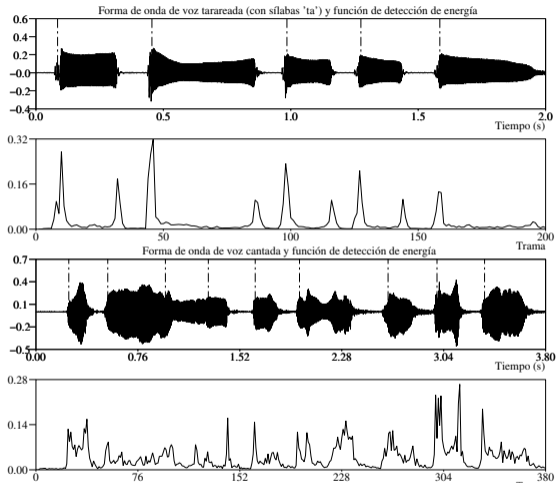


Procesamiento en el dominio del tiempo

Magnitud de tiempo corto

aplicación: segmentación
de audio en notas

- derivada de la magnitud de tiempo corto: $\frac{d}{dt} M_n$ [Dixon, 2001]
- refinamiento: [Klapuri, 1999]

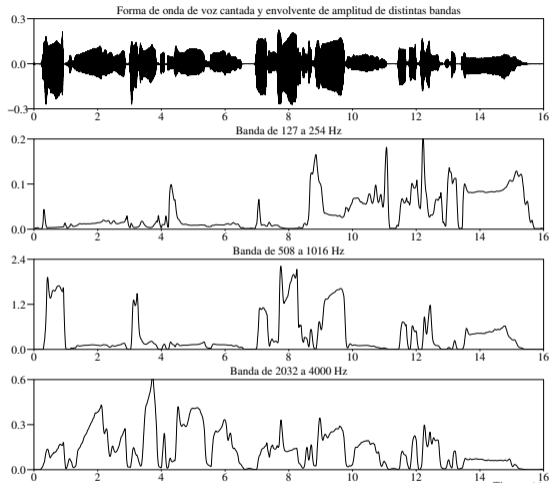


Procesamiento en el dominio del tiempo

Magnitud de tiempo corto

aplicación: segmentación
de audio en notas

- derivada de la magnitud de tiempo corto: $\frac{d}{dt} M_n$ [Dixon, 2001]
 - refinamiento: [Klapuri, 1999]
- procesamiento en bandas de frecuencia



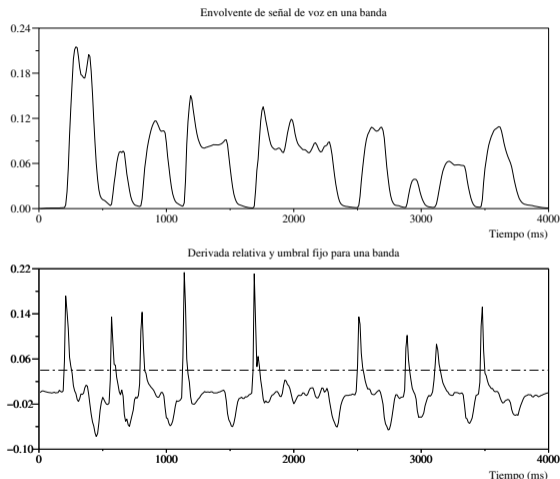
Procesamiento en el dominio del tiempo

Magnitud de tiempo corto

aplicación: segmentación
de audio en notas

- derivada de la magnitud de tiempo corto: $\frac{d}{dt} M_n$ [Dixon, 2001]
 - refinamiento: [Klapuri, 1999]
- procesamiento en bandas de frecuencia

derivada relativa $\frac{\frac{d}{dt} M_n}{M_n}$

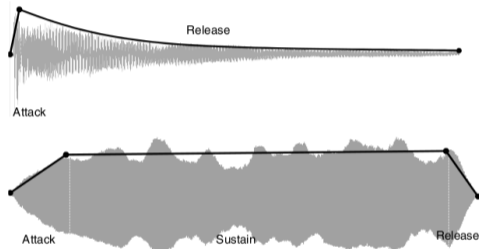


Procesamiento en el dominio del tiempo

Magnitud de tiempo corto

aplicación: clasificación de instrumentos musicales

- características derivadas de la magnitud de tiempo corto



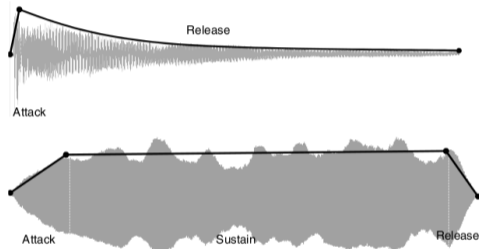
notas de guitarra y violín [Klapuri and Davy, 2006]

Procesamiento en el dominio del tiempo

Magnitud de tiempo corto

aplicación: clasificación de instrumentos musicales

- características derivadas de la magnitud de tiempo corto
- e.g. tiempo de ataque:



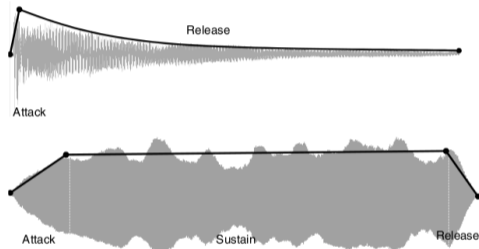
notas de guitarra y violín [Klapuri and Davy, 2006]

Procesamiento en el dominio del tiempo

Magnitud de tiempo corto

aplicación: clasificación de instrumentos musicales

- características derivadas de la magnitud de tiempo corto
- e.g. tiempo de ataque: entre 20 y 80 % de amplitud máxima



notas de guitarra y violín [Klapuri and Davy, 2006]

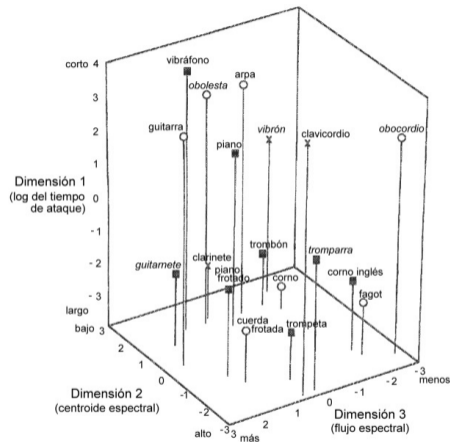
Procesamiento en el dominio del tiempo

Magnitud de tiempo corto

aplicación: clasificación de instrumentos musicales

- características derivadas de la magnitud de tiempo corto
- e.g. tiempo de ataque: entre 20 y 80 % de amplitud máxima

$$\text{LAT} = \log_{10}(t_{80} - t_{20})$$



espacio tímbrico tridimensional [MacAdams, 1999]

- 1 Introducción
- 2 Energía y magnitud de tiempo corto
- 3 Tasa de cruces por cero
- 4 Estimación de frecuencia fundamental
 - Autocorrelación
 - Función diferencia

Procesamiento en el dominio del tiempo

Tasa de cruces por cero

medida simple del contenido espectral:

Procesamiento en el dominio del tiempo

Tasa de cruces por cero

medida simple del contenido espectral:

$$Z_n = \sum_{m=-\infty}^{\infty} |\operatorname{sgn}\{x[m]\} - \operatorname{sgn}\{x[m-1]\}| w[n-m]$$

Procesamiento en el dominio del tiempo

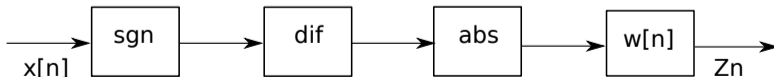
Tasa de cruces por cero

medida simple del contenido espectral:

$$Z_n = \sum_{m=-\infty}^{\infty} |\operatorname{sgn}\{x[m]\} - \operatorname{sgn}\{x[m-1]\}| w[n-m]$$

$$\begin{aligned} \operatorname{sgn}\{x[n]\} &= 1 & x[n] \geq 0 \\ &= -1 & x[n] < 0 \end{aligned}$$

$$\begin{aligned} w[n] &= \frac{1}{2N} & 0 \leq n \leq N-1 \\ &= 0 & \text{en otro caso} \end{aligned}$$



Procesamiento en el dominio del tiempo

Tasa de cruces por cero

- e.g. senoide de frecuencia f_0 , a f_s :
 $Z_n = 2f_0/f_s$ cruces/muestra

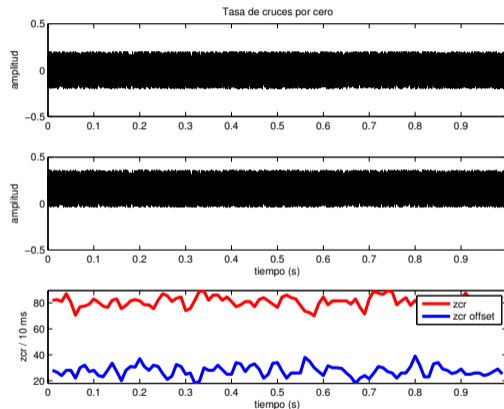
Tasa de cruces por cero

- e.g. senoide de frecuencia f_0 , a f_s :
 $Z_n = 2f_0/f_s$ cruces/muestra
- normalización para un cierto intervalo de tiempo τ (e.g. $\tau = 10\text{ms}$):
 $Z_M = Z_n \times M$ con $M = \tau \times f_s$
(resulta independiente de f_s)

Procesamiento en el dominio del tiempo

Tasa de cruces por cero

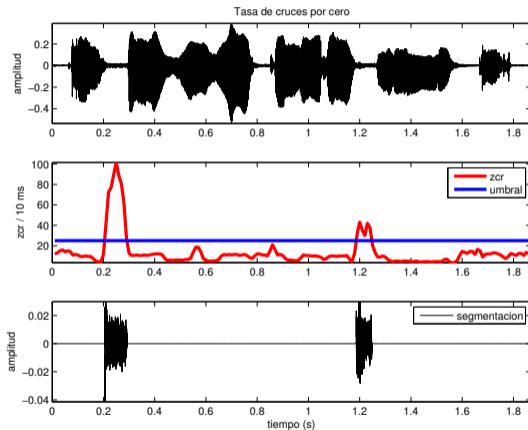
- e.g. senoide de frecuencia f_0 , a f_s :
 $Z_n = 2f_0/f_s$ cruces/muestra
- normalización para un cierto intervalo de tiempo τ (e.g. $\tau = 10\text{ms}$):
 $Z_M = Z_n \times M$ con $M = \tau \times f_s$
(resulta independiente de f_s)
- sensible a valor de DC y zumbido de baja frecuencia (filtro pasa-altos)



Procesamiento en el dominio del tiempo

Tasa de cruces por cero

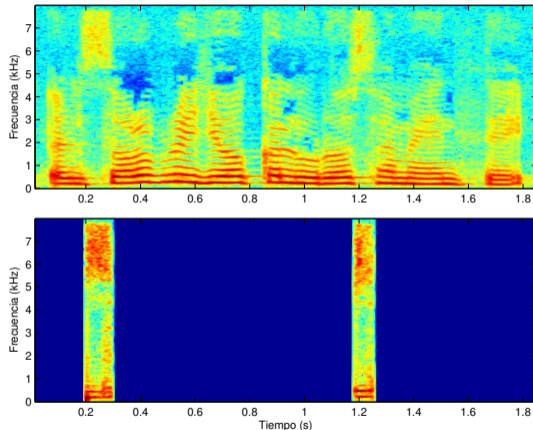
- ejemplo de detección de fonemas sordos



Procesamiento en el dominio del tiempo

Tasa de cruces por cero

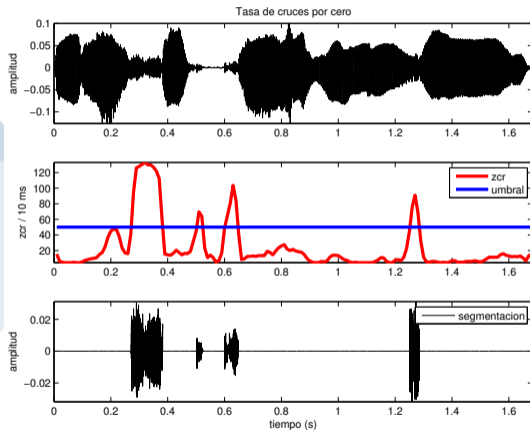
- ejemplo de detección de fonemas sordos



Procesamiento en el dominio del tiempo

Tasa de cruces por cero

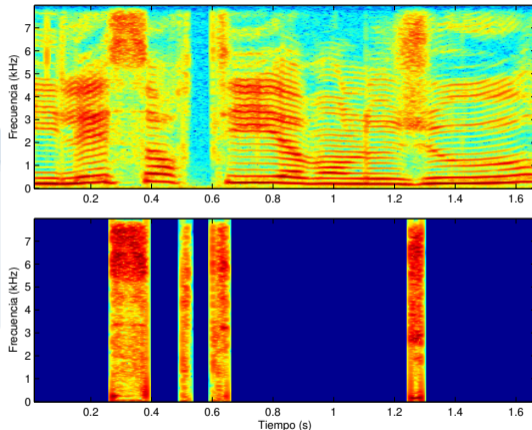
- ejemplo de detección de fonemas sordos



Procesamiento en el dominio del tiempo

Tasa de cruces por cero

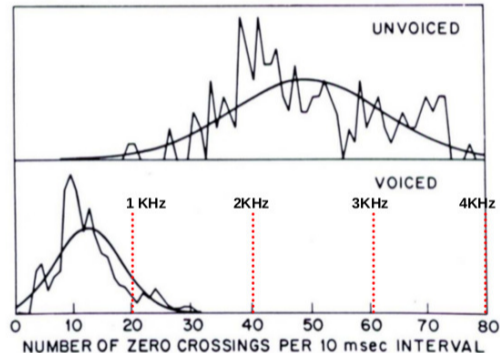
- ejemplo de detección de fonemas sordos



Procesamiento en el dominio del tiempo

Tasa de cruces por cero

- ejemplo de detección de fonemas sordos
- dificultad en establecer valor de umbral

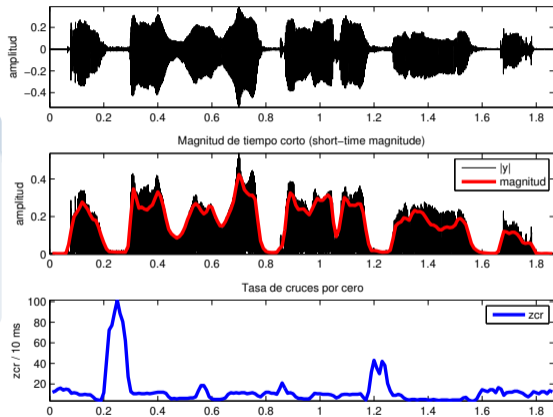


[Rabiner and Schafer, 2011]

Procesamiento en el dominio del tiempo

Tasa de cruces por cero

- ejemplo de detección de fonemas sordos
- dificultad en establecer valor de umbral
- combinación con magnitud de tiempo corto (M_n) para segmentación de voz



- 1 Introducción
- 2 Energía y magnitud de tiempo corto
- 3 Tasa de cruces por cero
- 4 Estimación de frecuencia fundamental**
 - Autocorrelación
 - Función diferencia

Procesamiento en el dominio del tiempo

Estimación de frecuencia fundamental

- uno de los problemas fundamentales del procesamiento de audio
(décadas de desarrollo desde 1960s y cientos de algoritmos propuestos)

Procesamiento en el dominio del tiempo

Estimación de frecuencia fundamental

- uno de los problemas fundamentales del procesamiento de audio
(décadas de desarrollo desde 1960s y cientos de algoritmos propuestos)

voicing determinar tipo de excitación: sonoro-sordo

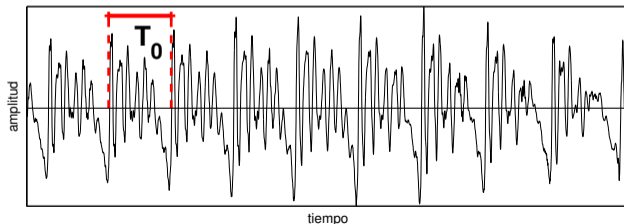
Procesamiento en el dominio del tiempo

Estimación de frecuencia fundamental

- uno de los problemas fundamentales del procesamiento de audio
(décadas de desarrollo desde 1960s y cientos de algoritmos propuestos)

voicing determinar tipo de excitación: sonoro-sordo

f_0 determinar frecuencia fundamental (también altura ó pitch)



Procesamiento en el dominio del tiempo

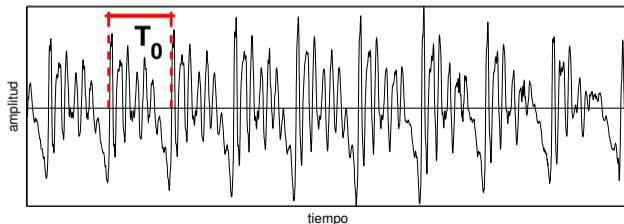
Estimación de frecuencia fundamental

- uno de los problemas fundamentales del procesamiento de audio
(décadas de desarrollo desde 1960s y cientos de algoritmos propuestos)

voicing determinar tipo de excitación: sonoro-sordo

f_0 determinar frecuencia fundamental (también altura ó pitch)

- $f_0 = 1/T_0$: período fundamental, duración promedio de varios ciclos



Procesamiento en el dominio del tiempo

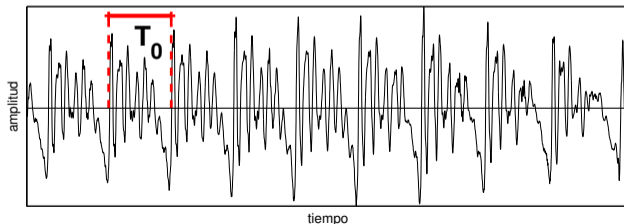
Estimación de frecuencia fundamental

- uno de los problemas fundamentales del procesamiento de audio
(décadas de desarrollo desde 1960s y cientos de algoritmos propuestos)

voicing determinar tipo de excitación: sonoro-sordo

f_0 determinar frecuencia fundamental (también altura ó pitch)

- $f_0 = 1/T_0$: período fundamental, duración promedio de varios ciclos
- f_0 : frecuencia de senoide que evoca la misma sensación de altura



Procesamiento en el dominio del tiempo

Dificultades para la estimación de f_0

- señal no-estacionaria y cuasi-periódica

Procesamiento en el dominio del tiempo

Dificultades para la estimación de f_0

- señal no-estacionaria y cuasi-periódica
- generalmente unos pocos períodos de f_0

Procesamiento en el dominio del tiempo

Dificultades para la estimación de f_0

- señal no-estacionaria y cuasi-periódica
- generalmente unos pocos períodos de f_0
- intervalo de variación de 4 octavas o más (e.g. 50-800 Hz en voz)

Procesamiento en el dominio del tiempo

Dificultades para la estimación de f_0

- señal no-estacionaria y cuasi-periódica
- generalmente unos pocos períodos de f_0
- intervalo de variación de 4 octavas o más (e.g. 50-800 Hz en voz)
- variación de formantes, interferencia de formantes de baja frecuencia

Procesamiento en el dominio del tiempo

Dificultades para la estimación de f_0

- señal no-estacionaria y cuasi-periódica
- generalmente unos pocos períodos de f_0
- intervalo de variación de 4 octavas o más (e.g. 50-800 Hz en voz)
- variación de formantes, interferencia de formantes de baja frecuencia
- errores típicos de detección de armónicos o sub-armónicos

Procesamiento en el dominio del tiempo

Dificultades para la estimación de f_0

- señal no-estacionaria y cuasi-periódica
- generalmente unos pocos períodos de f_0
- intervalo de variación de 4 octavas o más (e.g. 50-800 Hz en voz)
- variación de formantes, interferencia de formantes de baja frecuencia
- errores típicos de detección de armónicos o sub-armónicos
- varias alturas simultáneas, interferencia de otros sonidos

Procesamiento en el dominio del tiempo

Dificultades para la estimación de f_0

- señal no-estacionaria y cuasi-periódica
- generalmente unos pocos períodos de f_0
- intervalo de variación de 4 octavas o más (e.g. 50-800 Hz en voz)
- variación de formantes, interferencia de formantes de baja frecuencia
- errores típicos de detección de armónicos o sub-armónicos
- varias alturas simultáneas, interferencia de otros sonidos

Algoritmos de *pitch tracking*

- tres etapas: pre-procesado, estimación y post-procesado

Procesamiento en el dominio del tiempo

Dificultades para la estimación de f_0

- señal no-estacionaria y cuasi-periódica
- generalmente unos pocos períodos de f_0
- intervalo de variación de 4 octavas o más (e.g. 50-800 Hz en voz)
- variación de formantes, interferencia de formantes de baja frecuencia
- errores típicos de detección de armónicos o sub-armónicos
- varias alturas simultáneas, interferencia de otros sonidos

Algoritmos de *pitch tracking*

- tres etapas: pre-procesado, estimación y post-procesado
- en el dominio del tiempo (determinación de T_0)

Procesamiento en el dominio del tiempo

Dificultades para la estimación de f_0

- señal no-estacionaria y cuasi-periódica
- generalmente unos pocos períodos de f_0
- intervalo de variación de 4 octavas o más (e.g. 50-800 Hz en voz)
- variación de formantes, interferencia de formantes de baja frecuencia
- errores típicos de detección de armónicos o sub-armónicos
- varias alturas simultáneas, interferencia de otros sonidos

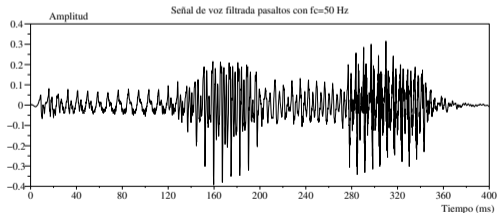
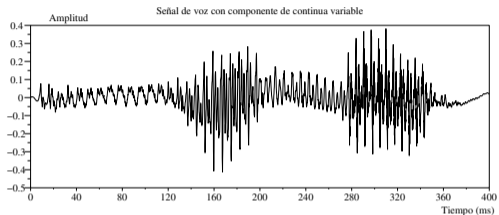
Algoritmos de *pitch tracking*

- tres etapas: pre-procesado, estimación y post-procesado
- en el dominio del tiempo (determinación de T_0)
- en el dominio de la frecuencia (estimación de f_0)

Procesamiento en el dominio del tiempo

Pre-procesado

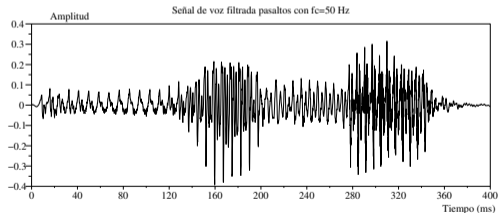
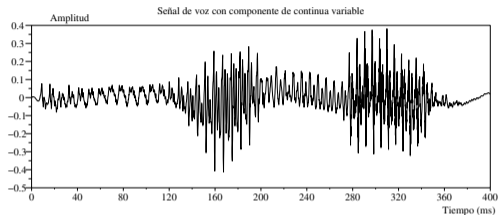
- filtrar componentes de muy baja frecuencia



Procesamiento en el dominio del tiempo

Pre-procesado

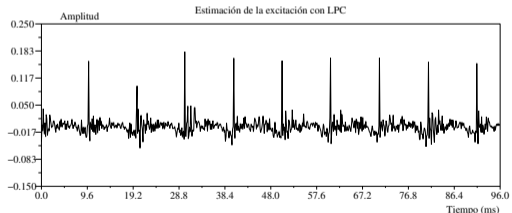
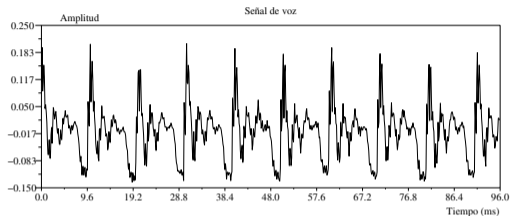
- filtrar componentes de muy baja frecuencia
- filtro pasabajos, mejora periodicidad y atenúa efecto de formantes



Procesamiento en el dominio del tiempo

Pre-procesado

- filtrar componentes de muy baja frecuencia
- filtro pasabajos, mejora periodicidad y atenúa efecto de formantes
- blanqueo espectral, estimar excitación glotal



Procesamiento en el dominio del tiempo

Autocorrelación de tiempo corto

- función de autocorrelación:

$$\phi[k] = \sum_{m=-\infty}^{\infty} x[m]x[m+k]$$

Procesamiento en el dominio del tiempo

Autocorrelación de tiempo corto

- función de autocorrelación:

$$\phi[k] = \sum_{m=-\infty}^{\infty} x[m]x[m+k]$$

- estimación de tiempo corto:

$$R_n[k] = \sum_{m=-\infty}^{\infty} x[m]w[n-m]x[m+k]w[n-k-m]$$

Procesamiento en el dominio del tiempo

Autocorrelación de tiempo corto

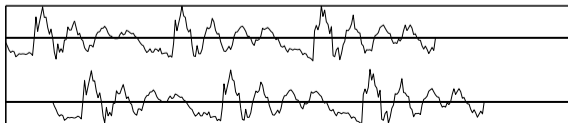
- función de autocorrelación:

$$\phi[k] = \sum_{m=-\infty}^{\infty} x[m]x[m+k]$$

- estimación de tiempo corto:

$$R_n[k] = \sum_{m=-\infty}^{\infty} x[m]w[n-m]x[m+k]w[n-k-m]$$

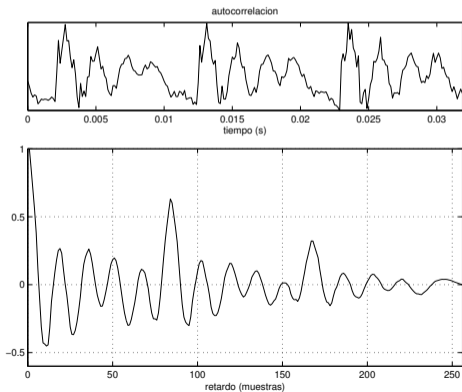
- $R_n[k] = R_n[-k]$, $R_n[k] \leq R_n[0] \forall k$, $R_n[0]$ energía



Procesamiento en el dominio del tiempo

Autocorrelación de tiempo corto

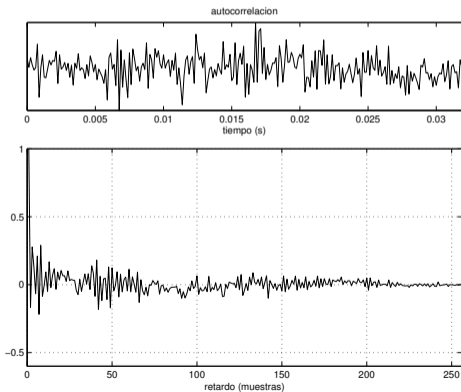
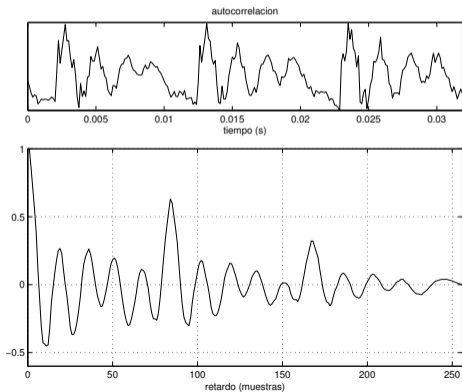
- sonoro picos marcados en retardos múltiplos del período



Procesamiento en el dominio del tiempo

Autocorrelación de tiempo corto

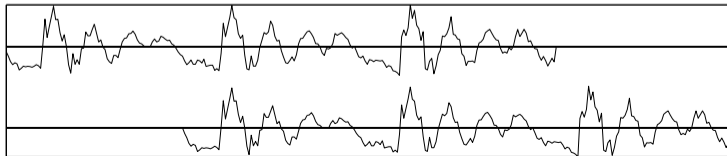
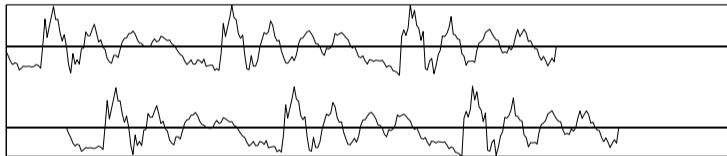
- **sonoro** picos marcados en retardos múltiplos del período
- **sordo** la falta de picos marcados indica ausencia de periodicidad



Procesamiento en el dominio del tiempo

Autocorrelación de tiempo corto

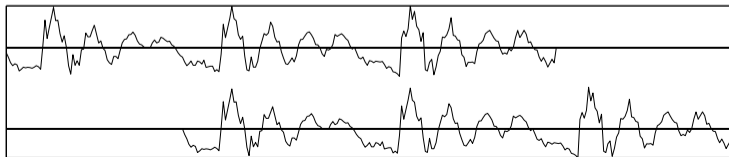
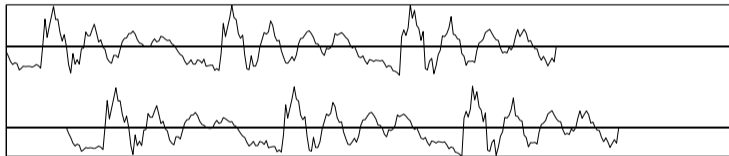
- largo de la ventana N de al menos dos períodos



Procesamiento en el dominio del tiempo

Autocorrelación de tiempo corto

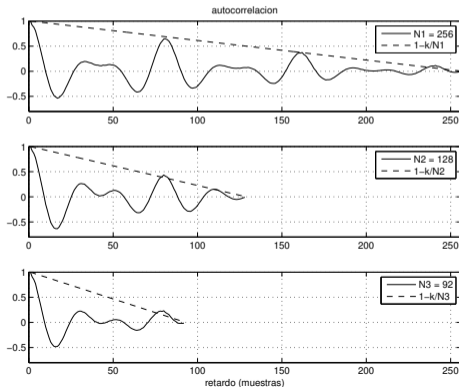
- largo de la ventana N de al menos dos períodos
- al crecer k intervienen menos muestras ($R_n[0]:N$, $R_n[k]:N - k$)



Procesamiento en el dominio del tiempo

Autocorrelación de tiempo corto

- largo de la ventana N de al menos dos períodos
- al crecer k intervienen menos muestras ($R_n[0]:N$, $R_n[k]:N - k$)
- reducción de la amplitud de $R_n[k]$ al crecer k



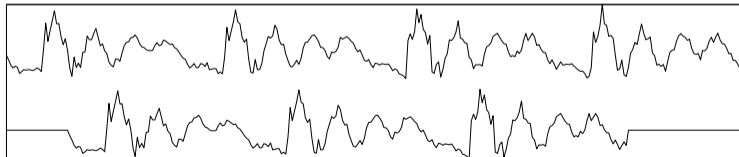
Procesamiento en el dominio del tiempo

Autocorrelación modificada

- autocorrelación modificada:

$$\hat{R}_n[k] = \sum_{m=-\infty}^{\infty} x[m]w_1[n-m]x[m+k]w_2[n-k-m]$$

$w_1[n]$ largo N , $w_2[n]$ largo $N + K$, con K mayor retardo deseado



Procesamiento en el dominio del tiempo

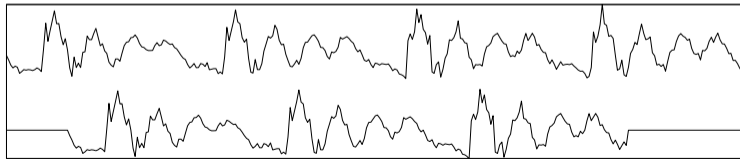
Autocorrelación modificada

- autocorrelación modificada:

$$\hat{R}_n[k] = \sum_{m=-\infty}^{\infty} x[m]w_1[n-m]x[m+k]w_2[n-k-m]$$

$w_1[n]$ largo N , $w_2[n]$ largo $N + K$, con K mayor retardo deseado

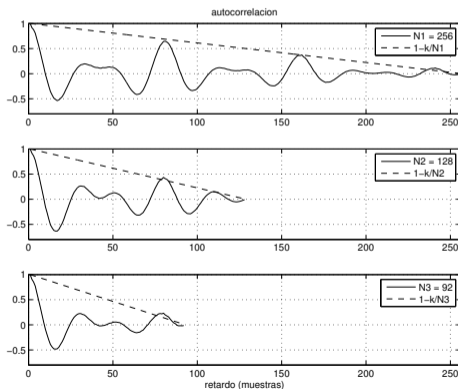
- $\hat{R}_n[k] \neq \hat{R}_n[-k]$, es una *correlación cruzada*



Procesamiento en el dominio del tiempo

Autocorrelación modificada

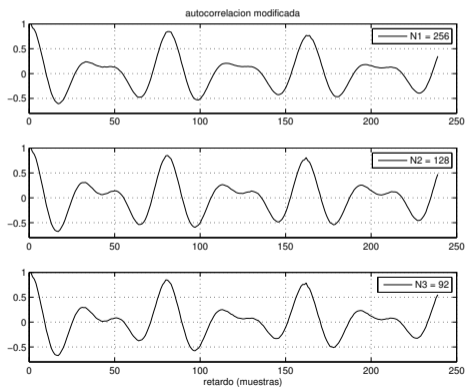
- no presenta el decaimiento al aumentar k



Procesamiento en el dominio del tiempo

Autocorrelación modificada

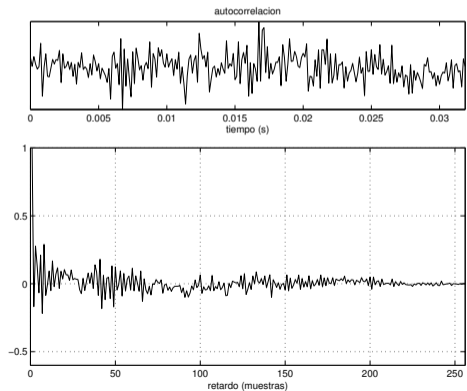
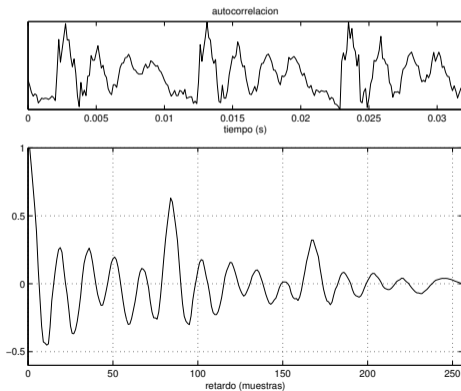
- no presenta el decaimiento al aumentar k



Procesamiento en el dominio del tiempo

Autocorrelación modificada

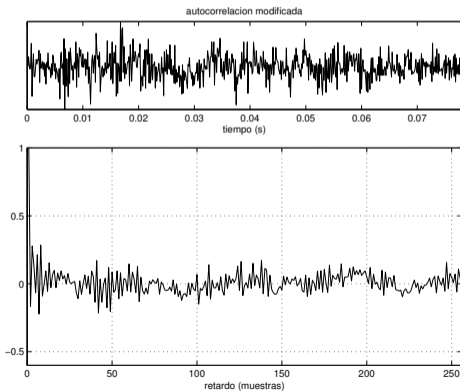
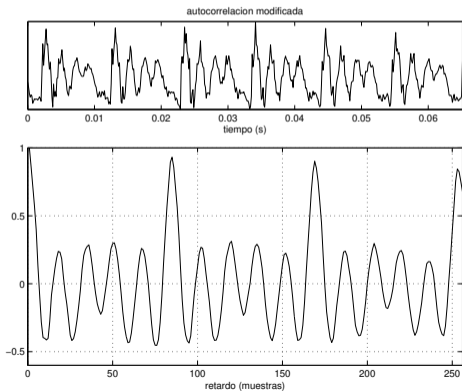
- no presenta el decaimiento al aumentar k
- período correspondiente al mayor pico de menor retardo



Procesamiento en el dominio del tiempo

Autocorrelación modificada

- no presenta el decaimiento al aumentar k
- período correspondiente al mayor pico de menor retardo



Procesamiento en el dominio del tiempo

Función diferencia

- motivación: $d[n] = x[n] - x[n - k] \approx 0 \quad \forall k$ múltiplo del período

Procesamiento en el dominio del tiempo

Función diferencia

- motivación: $d[n] = x[n] - x[n - k] \approx 0 \quad \forall k$ múltiplo del período
- función diferencia (Average Magnitude Difference Function, AMDF):

$$\hat{D}_n[k] = \sum_{m=-\infty}^{\infty} |x[m]w_1[n - m] - x[m + k]w_2[n - k - m]|$$

Procesamiento en el dominio del tiempo

Función diferencia

- motivación: $d[n] = x[n] - x[n - k] \approx 0 \quad \forall k$ múltiplo del período
- función diferencia (Average Magnitude Difference Function, AMDF):

$$\hat{D}_n[k] = \sum_{m=-\infty}^{\infty} |x[m]w_1[n - m] - x[m + k]w_2[n - k - m]|$$

- presenta mínimos donde $\hat{R}_n[k]$ tiene máximos

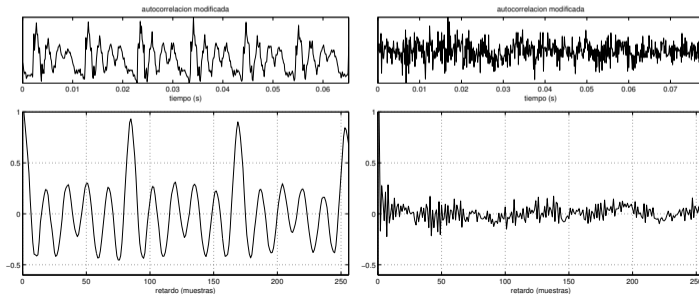
Procesamiento en el dominio del tiempo

Función diferencia

- motivación: $d[n] = x[n] - x[n - k] \approx 0 \quad \forall k$ múltiplo del período
- función diferencia (Average Magnitude Difference Function, AMDF):

$$\hat{D}_n[k] = \sum_{m=-\infty}^{\infty} |x[m]w_1[n - m] - x[m + k]w_2[n - k - m]|$$

- presenta mínimos donde $\hat{R}_n[k]$ tiene máximos



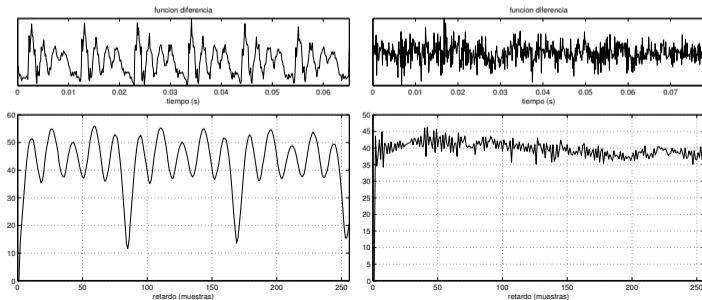
Procesamiento en el dominio del tiempo

Función diferencia

- motivación: $d[n] = x[n] - x[n - k] \approx 0 \quad \forall k$ múltiplo del período
- función diferencia (Average Magnitude Difference Function, AMDF):

$$\hat{D}_n[k] = \sum_{m=-\infty}^{\infty} |x[m]w_1[n - m] - x[m + k]w_2[n - k - m]|$$

- presenta mínimos donde $\hat{R}_n[k]$ tiene máximos

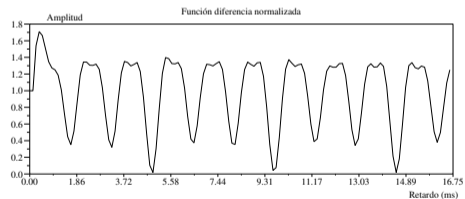
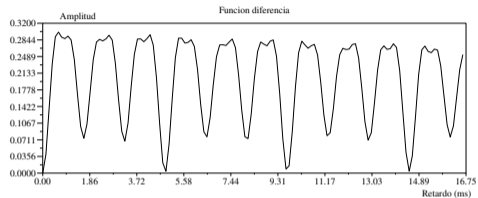


Procesamiento en el dominio del tiempo

YIN [de Cheveigné and Kawahara, 2002]

- distancia cuadrática

$$\hat{D}_n[k] = \sum_{m=0}^{N-1} (x[m] - x[m+k])^2$$



Procesamiento en el dominio del tiempo

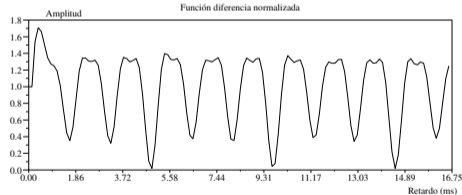
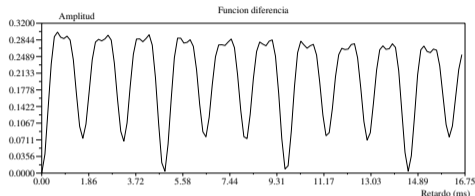
YIN [de Cheveigné and Kawahara, 2002]

- distancia cuadrática

$$\hat{D}_n[k] = \sum_{m=0}^{N-1} (x[m] - x[m+k])^2$$

- normalización

$$\bar{D}_n[k] = \begin{cases} 1, & k = 0 \\ \frac{\hat{D}_n[k]}{\frac{1}{k} \sum_{j=1}^k \hat{D}_n[j]}, & k \neq 0 \end{cases}$$



Procesamiento en el dominio del tiempo

YIN [de Cheveigné and Kawahara, 2002]

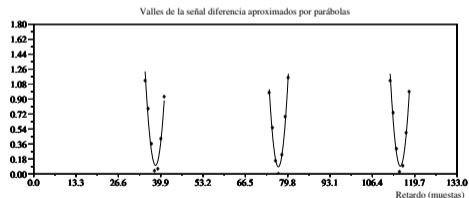
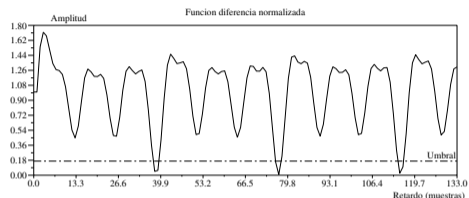
- distancia cuadrática

$$\hat{D}_n[k] = \sum_{m=0}^{N-1} (x[m] - x[m+k])^2$$

- normalización

$$\bar{D}_n[k] = \begin{cases} 1, & k = 0 \\ \frac{\hat{D}_n[k]}{\frac{1}{k} \sum_{j=1}^k \hat{D}_n[j]}, & k \neq 0 \end{cases}$$

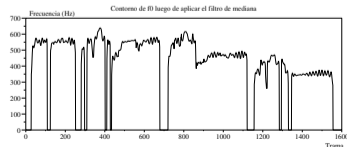
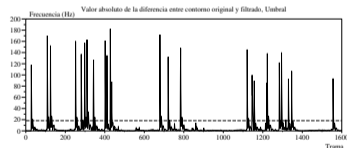
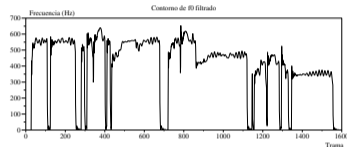
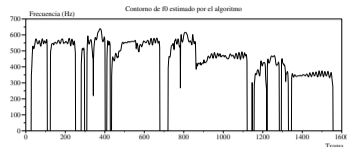
- aproximación por parábolas, para aumentar resolución y evitar errores al buscar mínimo



Procesamiento en el dominio del tiempo

Post-procesado

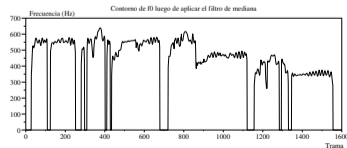
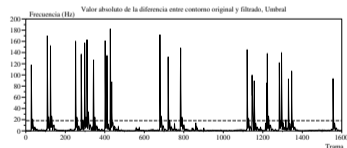
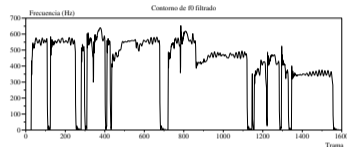
- contorno de f_0 ruidoso



Procesamiento en el dominio del tiempo

Post-procesado

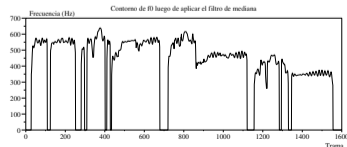
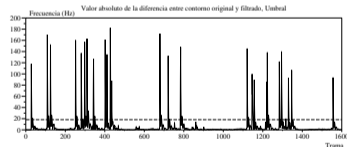
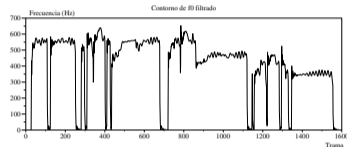
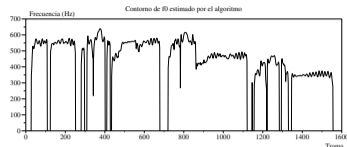
- contorno de f_0 ruidoso
- corregir errores más nocivos (e.g. armónico)



Procesamiento en el dominio del tiempo

Post-procesado

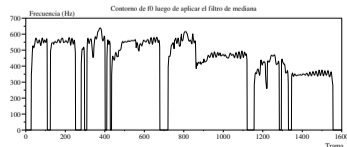
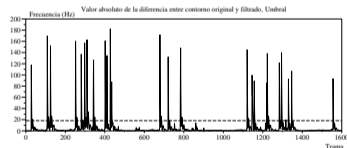
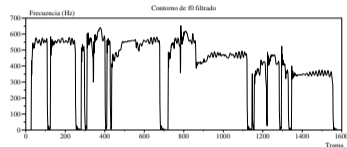
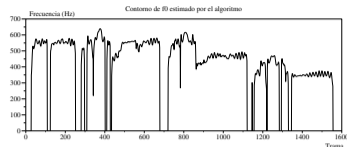
- contorno de f_0 ruidoso
- corregir errores más nocivos (e.g. armónico)
- pasabajos afecta discontinuidades



Procesamiento en el dominio del tiempo

Post-procesado

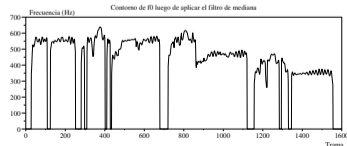
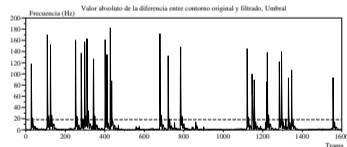
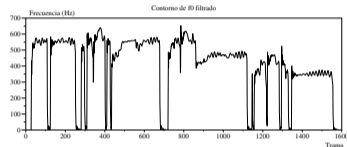
- contorno de f_0 ruidoso
- corregir errores más nocivos (e.g. armónico)
- pasabajos afecta discontinuidades
- filtro de mediana
 $\bar{x}[\hat{n}] = \text{median}\{x[n]\}$



Procesamiento en el dominio del tiempo

Post-procesado

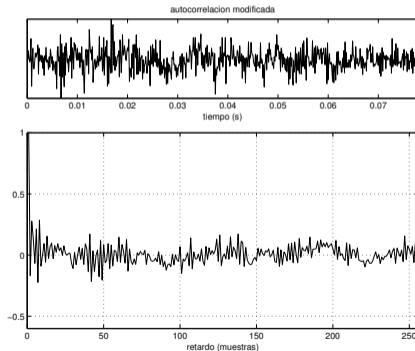
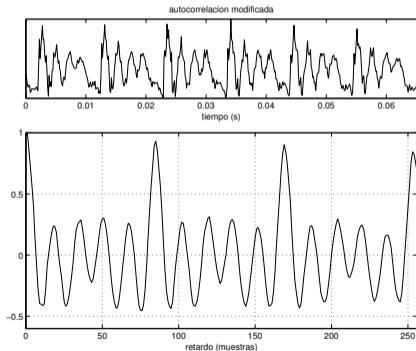
- contorno de f_0 ruidoso
- corregir errores más nocivos (e.g. armónico)
- pasabajos afecta discontinuidades
- filtro de mediana
 $\bar{x}[\hat{n}] = \text{median}\{x[n]\}$
- detección de errores gruesos con filtro pasabajos y corrección local con filtro de mediana



Procesamiento en el dominio del tiempo

Voicing

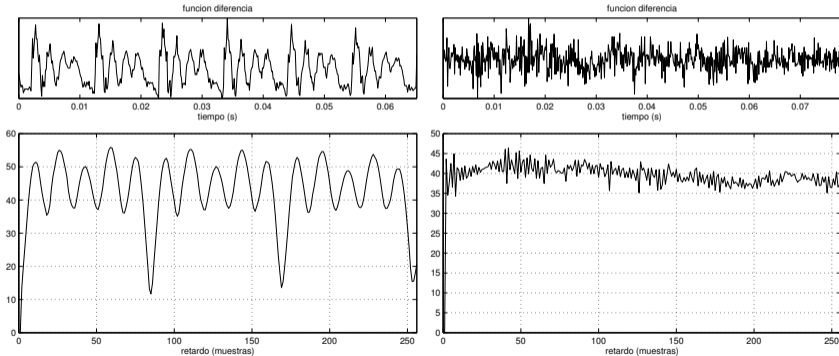
- tanto en $\hat{R}_n[k]$ como en $\hat{D}_n[k]$ el valor del máximo/mínimo detectado es un indicador del tipo de excitación sonoro-sordo



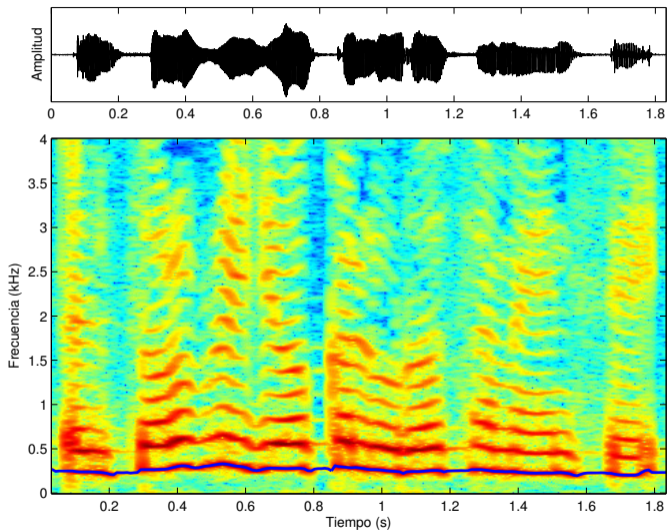
Procesamiento en el dominio del tiempo

Voicing

- tanto en $\hat{R}_n[k]$ como en $\hat{D}_n[k]$ el valor del máximo/mínimo detectado es un indicador del tipo de excitación sonoro-sordo

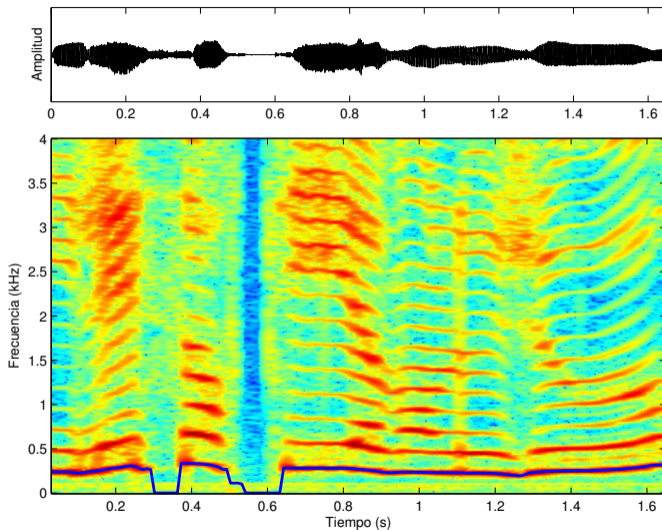


Procesamiento en el dominio del tiempo



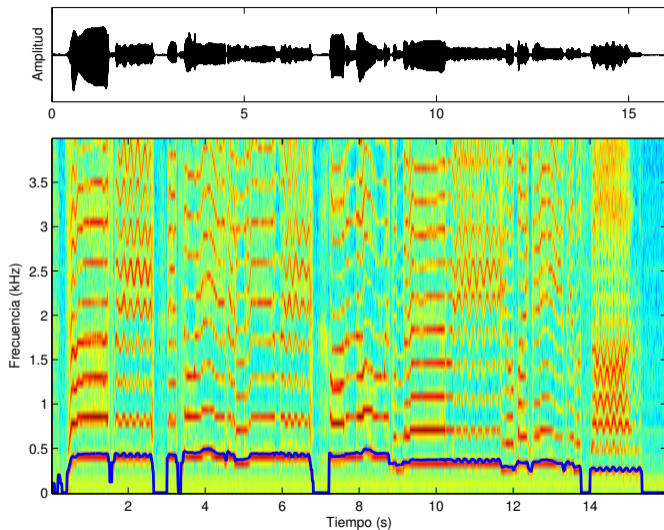
ejemplo pitch tracking voz hablada

Procesamiento en el dominio del tiempo






ejemplo pitch tracking voz hablada

Procesamiento en el dominio del tiempo



ejemplo pitch tracking voz cantada

Referencias I

-  de Cheveigné, A. and Kawahara, H. (2002).
Yin, a fundamental frequency estimator for speech and music.
JASA, 111:1917–1930.
-  Dixon, S. (2001).
Learning to detect onsets of acoustic piano tones.
Proceedings of the Workshop on Current Directions in Computer Music Research.
-  Klapuri, A. and Davy, M. (2006).
Signal Processing Methods for Music Transcription.
Springer, 1st edition.
-  Klapuri, A. P. (1999).
Sound onset detection by applying psychoacoustic knowledge.
Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing.

Referencias II



Rabiner, L. R. and Schafer, R. W. (2011).

Theory and Applications of Digital Speech Processing.

Prentice Hall, 1st edition.

Chapter 6 - Time-domain methods for speech signals.