

Procesamiento digital de señales de audio

Modelado espectral

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

- 1 Modelado espectral
- 2 Modelado sinusoidal
- 3 Modelado sinusoides + ruido
- 4 Modelado sinusoides + ruido + transitorios

Modelado espectral

Análisis-síntesis en modelado de señales

- extraer parámetros de modelo que representa bloque de forma de onda
- usar parámetros para reconstruir aproximación lo más cercana posible
- posibilidad de manipular los parámetros para generar transformaciones

Modelado espectral

típicamente se divide la señal en componentes de distinta naturaleza

- **sinusoides** de frecuencia, amplitud y fase variable [McAulay and Quatieri, 1986]
- **sinusoides + ruido** residuo modelado como ruido [Serra and Smith, 1990]
- **sinusoides + ruido + transitorios** agrega transientes [Verma and Meng, 2000]
permite manipulación independiente de cada componente

Modelado sinusoidal

modelo lineal de producción de VOZ: [Quatieri, 2001]

$$s(t) = \int_0^t h(t, t - \tau) u(\tau) d\tau$$

- $u(t)$: fuente de excitación
- $h(t, \tau)$: filtro lineal variante en el tiempo

convolución con respuesta al impulso diferente para cada tiempo t
se propone representar $u(t)$ como:

$$u(t) = \operatorname{Re} \sum_{k=0}^{K(t)} a_k(t) e^{j\Phi_k(t)}$$

- $a_k(t)$: amplitud
- $\Omega_k(t)$: frecuencia
- $\Phi_k(t)$: fase, $\Phi_k(t) = \int_0^t \Omega_k(\sigma) d\sigma + \Phi_k$

Modelado sinusoidal

transferencia del aparato vocal:

$$H(t, \Omega) = M(t, \Omega)e^{\Phi(t, \Omega)}$$

si $a_k(t)$ y $\Omega_k(t)$ ctes a lo largo de la respuesta al impulso del filtro,

$$s(t) = \text{Re} \sum_{k=0}^{K(t)} a_k(t) M(t, \Omega_k(t)) e^{j(\int_0^t \Omega_k(\sigma) d\sigma + \Phi_k + \Phi(t, \Omega_k(t)))}$$

modelo sinusoidal básico:

$$s(t) = \sum_{k=1}^{K(t)} A_k(t) e^{j\theta_k(t)}$$

- $A_k(t) = a_k(t) M(t, \Omega_k(t))$
- $\theta_k(t) = \Phi_k(t) + \Phi(t, \Omega_k(t)) = \int_0^t \Omega_k(\sigma) d\sigma + \Phi_k + \Phi(t, \Omega_k(t))$

Modelado sinusoidal

Ejemplo 1: armónicos de frecuencia constante

- $\Omega_k(t) = k\Omega_0$, $a_k(t) = 1$, $\theta_k = 0$
- $\theta_k(t) = \int_0^t \Omega_k(\sigma) d\sigma + \theta_k = k\Omega_0 t$
- $M(t, \Omega_k(t)) = M(k\Omega_0)$,
 $\Phi(t, \Omega_k(t)) = \Phi(k\Omega_0)$

$$s(t) = \sum_{k=1}^K M(k\Omega_0) e^{j(k\Omega_0 t + \Phi(k\Omega_0))}$$

Ejemplo 2: armónicos de frecuencia variable linealmente

- $\Omega_k(t) = k\Omega_0 ct$, $a_k(t) = 1$, $\theta_k = 0$
- $\theta_k(t) = \int_0^t \Omega_k(\sigma) d\sigma + \theta_k = kc\Omega_0 \frac{t^2}{2}$

$$s(t) = \sum_{k=1}^K M(kc\Omega_0 t) e^{j(kc\Omega_0 \frac{t^2}{2} + \Phi(kc\Omega_0 t))}$$

Modelado sinusoidal

Ejemplo 3: senoide discreta

$$x[n] = A \cos(\omega_0 n + \Phi)$$

- $\hat{A}^r, \hat{\omega}_0^r$, estimaciones a partir de la magnitud del bloque r de la STFT
- estimación de $x[n]$ como, $\hat{x}^r[n] = \hat{A}^r \cos(\hat{\omega}_0^r n) \quad rR \leq n \leq (r+1)R$
síntesis de bloques produce discontinuidad de forma de onda en bordes
- estimación de la fase como suma acumulada de la frecuencia

$$\hat{\theta}^r[n] = \sum_{rR}^n \hat{\omega}_0^r + \hat{\theta}^{r-1} = (n - rR) \hat{\omega}_0^r + \hat{\theta}^{r-1} \quad rR \leq n \leq (r+1)R$$

$$\hat{x}[n]^r = \hat{A}^r \cos(\hat{\theta}^r[n])$$

Modelado sinusoidal

Análisis y síntesis

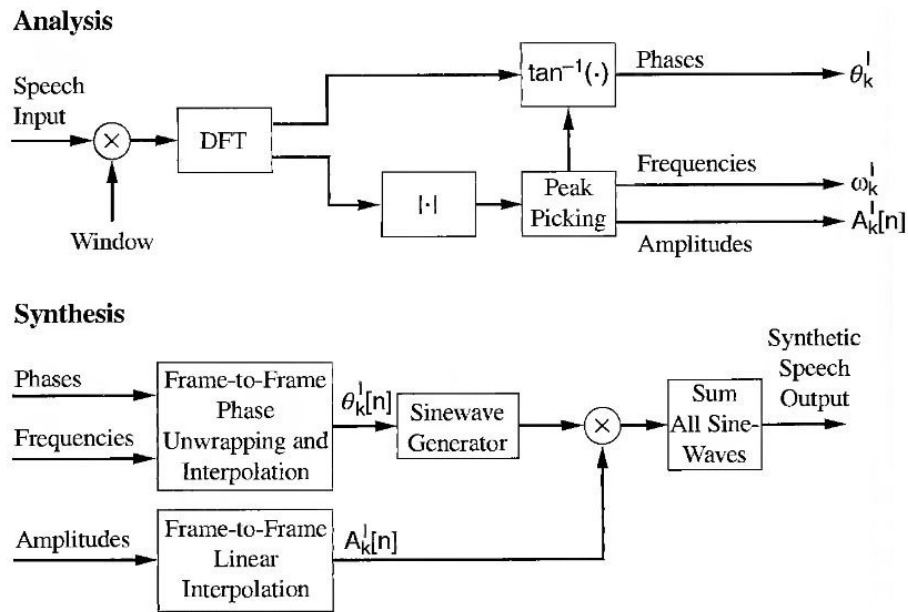
- desarrollar procedimientos robustos para extraer del análisis STFT las **amplitudes**, **frecuencias** y **fases** de los componentes sinusoidales
- la forma de onda se reconstruye interpolando a través de bloques sucesivos y modulando sinusoidales con las funciones resultantes
- estimación de parámetros puede mejorarse usando métodos de **interpolación**, **derivadas** de la señal y adecuado **enventanado**

Modelo discreto

fuerza de excitación y respuesta del tracto vocal constantes a lo largo de la duración de la ventana de análisis L

$$s[n] = \sum_{k=1}^{K^r} A_k^r e^{j\theta_k^r} e^{j\omega_k^r n}, \quad -\frac{L-1}{2} \leq n \leq \frac{L-1}{2}$$

Modelado sinusoidal

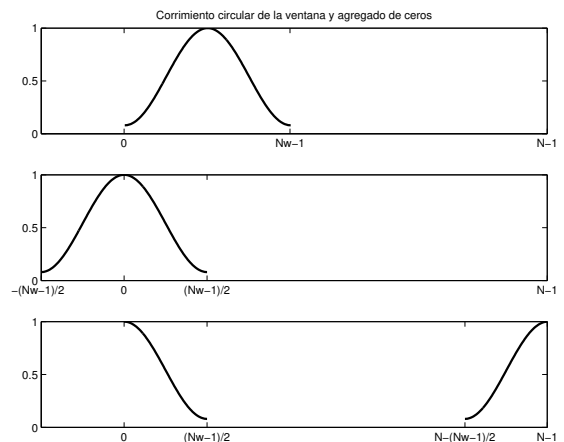


Modelado sinusoidal

Eventando y relleno de ceros

$$X_{rR}(e^{j\omega}) = \sum_{m=-\infty}^{\infty} w[rR - m]x[m]e^{-j\omega m}$$

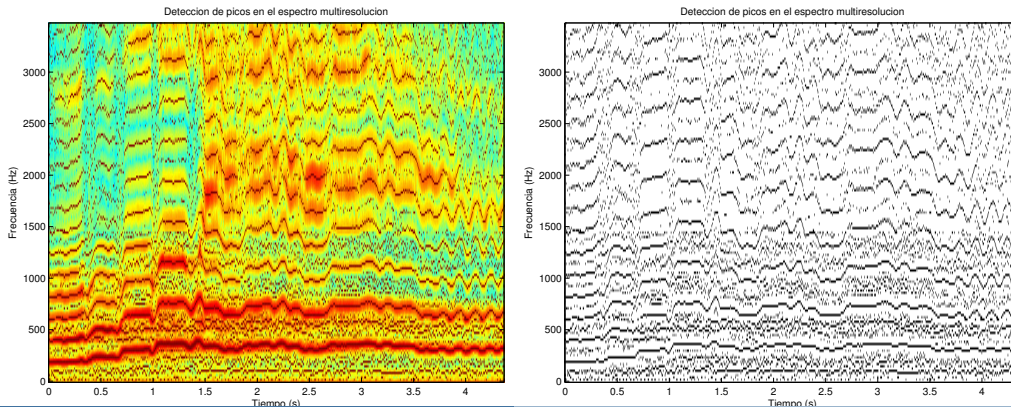
- eventado con corrimiento circular: disminuir error lineal de fase
 - ventana simétrica en $(L - 1)/2$ tiene fase lineal $-\omega(L - 1)/2$
- agregado de ceros para espectro interpolado



Modelado sinusoidal

Detección de picos

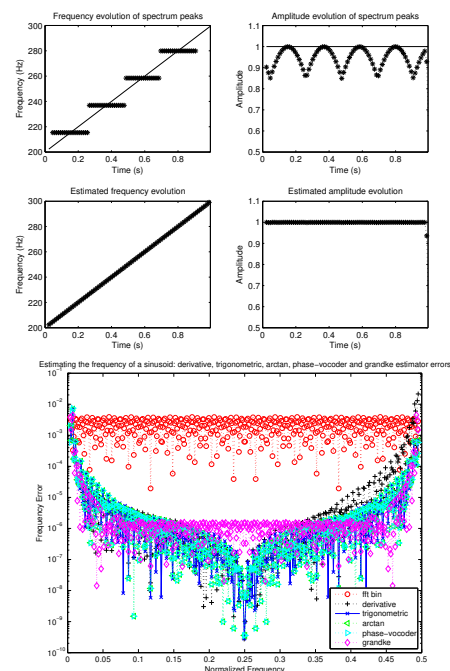
- máximos locales en la magnitud del espectro (resolución, umbrales)
- STFT: robusto frente a ruido y distorsión, funciona en espectros densos, muchos falsos positivos debido a picos espúreos
- MRFFT: produce menos picos espúreos (enmascaramiento debido al lóbulo principal más ancho) en alta frecuencia [Dressler, 2006]



Modelado sinusoidal

Estimación de amplitud y frecuencia instantánea (y fase)

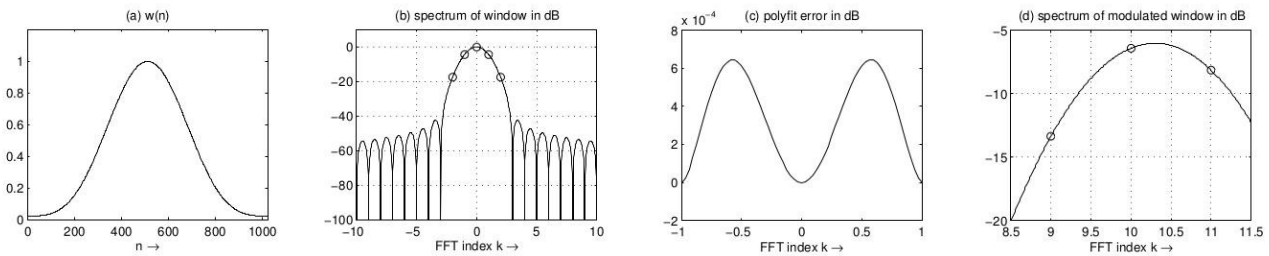
- DFT: $\forall k_m$ máximo local, $\omega_{k_m} = 2\pi k_m \frac{f_s}{N}$ y $A_{k_m} = 2 \frac{X_n(k_m)}{\sum_{n=0}^{N-1} w[n]}$
- se puede mejorar la resolución en frecuencia más allá de un bin (lo que a su vez puede usarse para mejorar la estimación de amplitud)
- diversos estimadores propuestos (e.g. ver [Keiler and Marchand, 2002])
 - basados en la derivada de la fase (phase-vocoder)
 - o en interpolación del espectro de magnitud o de fase



Modelado sinusoidal

Interpolación de magnitud del espectro

- se aproxima lóbulo principal de ventana en dB por una parábola $p(k)$
- $\forall k_m$ máximo local, $X_{dB}(k) = 20 \log_{10}(|X_n(k)|)$
 $A_1 = X_{dB}(k_m - 1)$, $A_2 = X_{dB}(k_m)$, $A_3 = X_{dB}(k_m + 1)$
 $d = \frac{1}{2} \frac{A_1 - A_2}{A_1 - 2A_2 + A_3}$ diferencia de frecuencia en bins
 $A_{k_m}(dB) = p(k_m + d) = A_2 - \frac{d}{4}(A_1 - A_3)$ amplitud estimada
- comparación con lóbulo ideal permite descartar picos espúreos



[Keiler and Marchand, 2002]

Modelado sinusoidal

Estimación usando el espectro de fase

- estimación de la frecuencia instantánea a partir de la diferencia de fase $\Delta\theta[k]$ de espectros sucesivos (método de phase-vocoder)
- $\forall k_m$ máximo local, $\omega_{k_m} = 2\pi(k_m + \mathcal{K}[k_m]) \frac{f_s}{N}$ con, [Dressler, 2006]

$$\mathcal{K}[k_m] = \frac{N}{2\pi R} \text{princarg} \left[\theta_{k_m}^r - \theta_{k_m}^{r-1} - \frac{2\pi R}{N} k \right]$$

donde \mathcal{K} es la desviación de frecuencia instantánea en unidad de bins

- A_{k_m} se estima usando la forma de la ventana, $|X_n(k_m)|$ y \mathcal{K}
- mejores resultados métodos basados en fase [Keiler and Marchand, 2002]

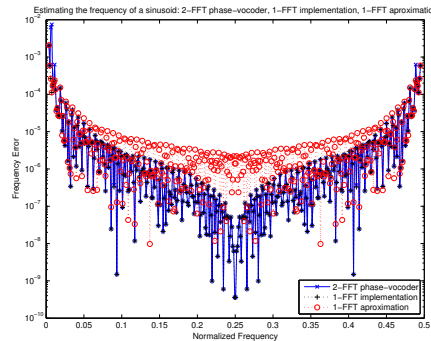
Modelado sinusoidal

Estimación usando el espectro de fase

- más precisa cuanto menor es el salto R y se suele usar una muestra
- requiere calcular 2 FFT, pero puede implementarse con 1 FFT

$$\begin{aligned}TF\{x[n+1]\}_{[k]} &= e^{j2\pi k/N} (TF\{x[n]\}_{[k]} + x[N] - x[0]) \\ &\approx e^{j2\pi k/N} TF\{x[n]\}_{[k]}\end{aligned}$$

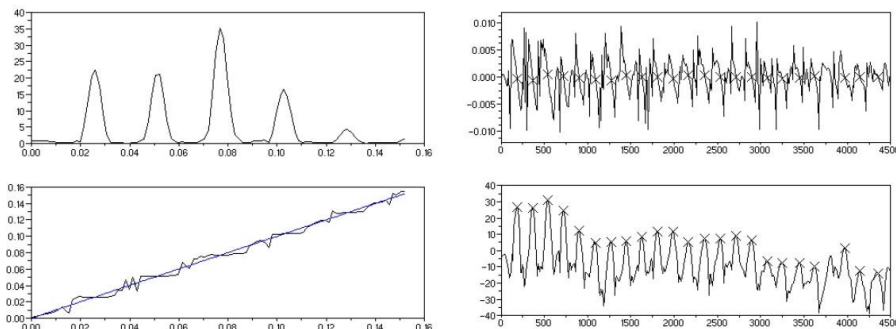
- válido para ventana rectangular, hay que enventanar en frecuencia



Modelado sinusoidal

Detección de picos (revisitado)

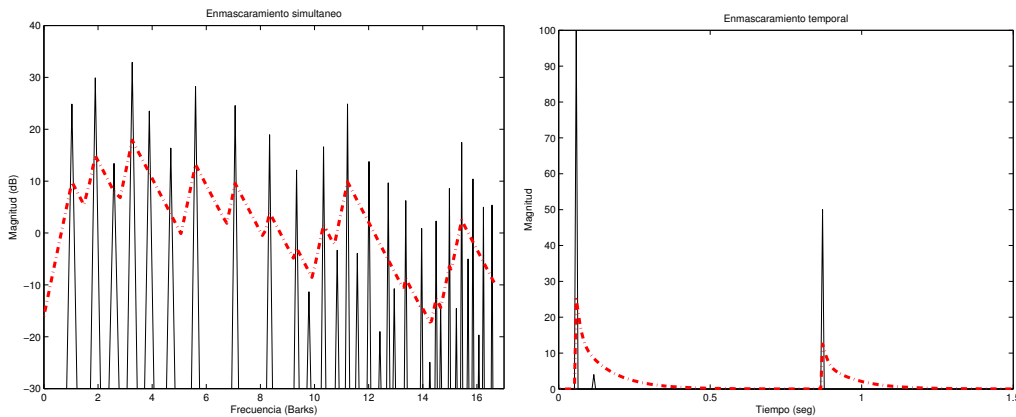
- detección de sinusoides en el espectro de fase [Charpentier, 1986]
- se valida un componente sinusoidal si cumple las condiciones,
 - frecuencia instantánea similar a la frecuencia del bin
 - frecuencia instantánea similar a la de los vecinos contiguos (ventana Hamming ó Hann, corresponden a 3 bins en frecuencia)



Modelado sinusoidal

Detección de picos (revisitado)

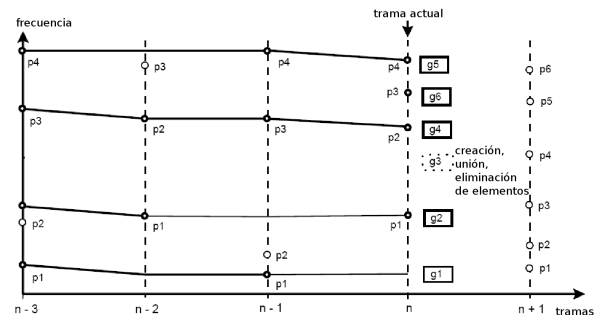
- eliminación de picos espúreos usando umbrales sobre magnitud
- se pueden aplicar criterios de enmascaramiento [Dressler, 2006]
 - simultáneo: umbrales en frecuencia
 - no simultáneo: umbrales en el tiempo



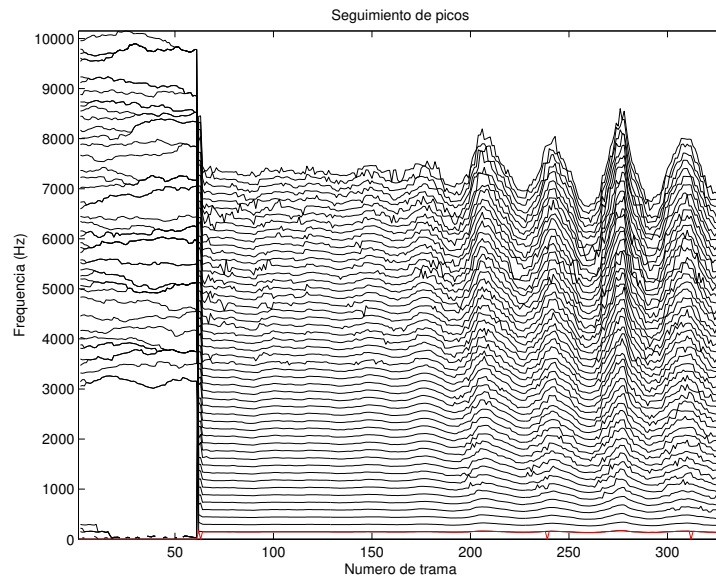
Modelado sinusoidal

Seguimiento de parciales

- unir picos trama a trama es un problema difícil
 - número de picos cambia al variar f_0
 - aparecen y desaparecen armónicos
- información relevante
 - cercanía de candidatos en frecuencia
 - usar además información de amplitud y fase
 - explotar armonicidad cuando existe (detección de f_0)
- diversas técnicas
 - agentes, HMM, filtro Kalman, etc.



Modelado sinusoidal



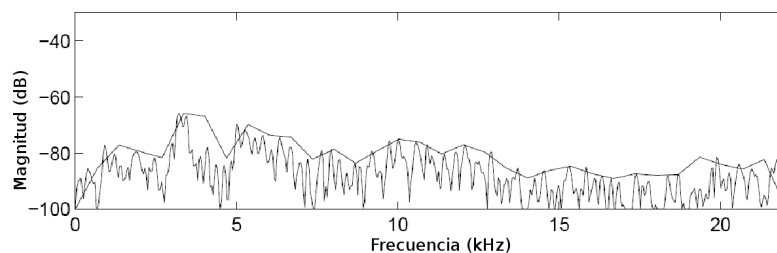
seguimiento usando SMS

Modelado sinusoides + ruido

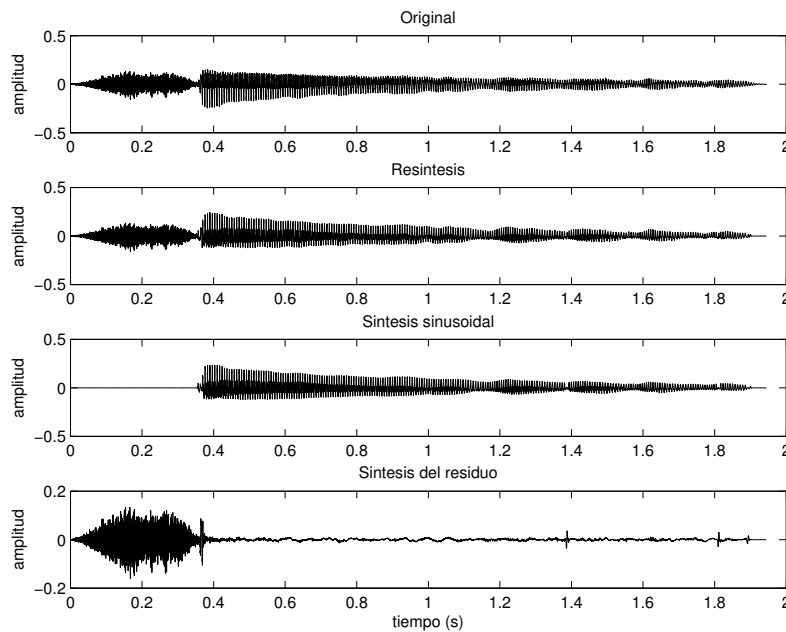
Modelado del residuo

- modelado sinusoidal básico falla en representar y transformar ruido y transitorios, los ataques son suavizados y el ruido suena artificial
- modelado **sinusoides + ruido** [Serra and Smith, 1990]
 - sustraer las sinusoides estimadas de la señal original
 - es necesario preservar la fase para cancelación correcta
 - el residuo puede modelarse como señal estocástica (ruido filtrado)

$$e[n] = s[n] - \sum_{k=1}^{K^r} A_k^r e^{j\theta_k^r} e^{j\omega_k^r n} \quad e[n] = h[n] * b[n]$$



Modelado sinusoides + ruido



síntesis usando SMS

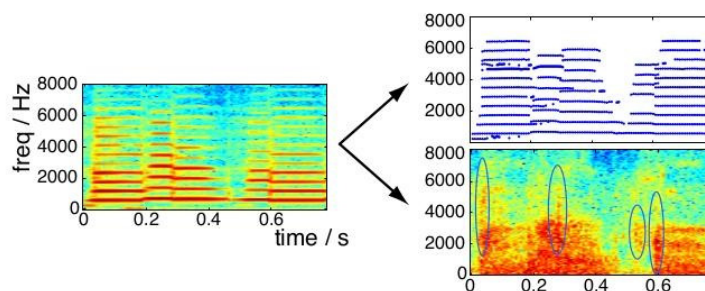
Modelado sinusoides + ruido + transitorios

Modelado de transitorios

- el modelado del ruido mejora el desempeño del modelo pero falla en manejar los transitorios, los ataques siguen siendo suavizados
- modelado **sinusoides + ruido + transitorios** [Verma and Meng, 2000]
 - se separan los transitorios abruptos del residuo

$$e[n] = \sum_k t_k[n] + h'[n] * b[n]$$

- los transientes se pueden reubicar en el tiempo y no se suavizan

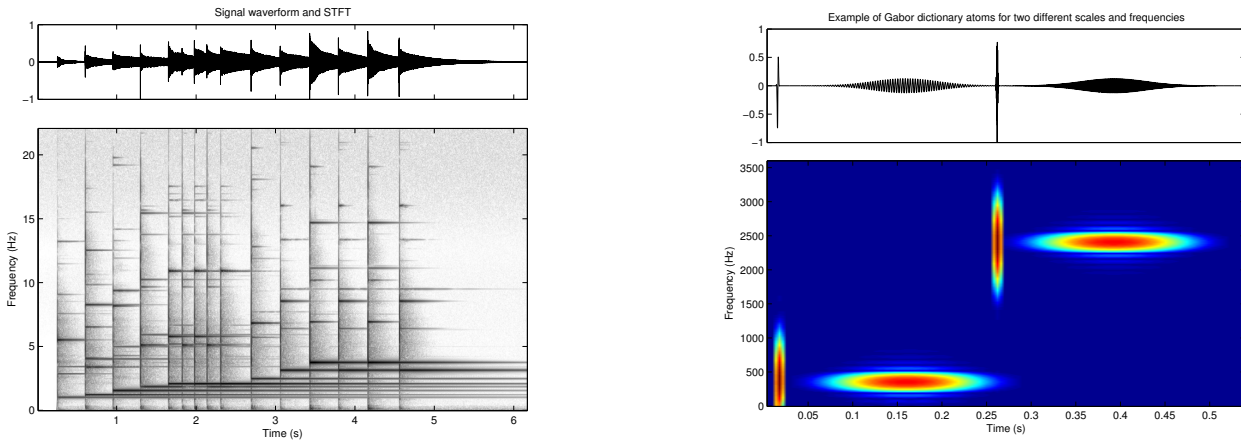


Modelado sinusoides + ruido + transitorios

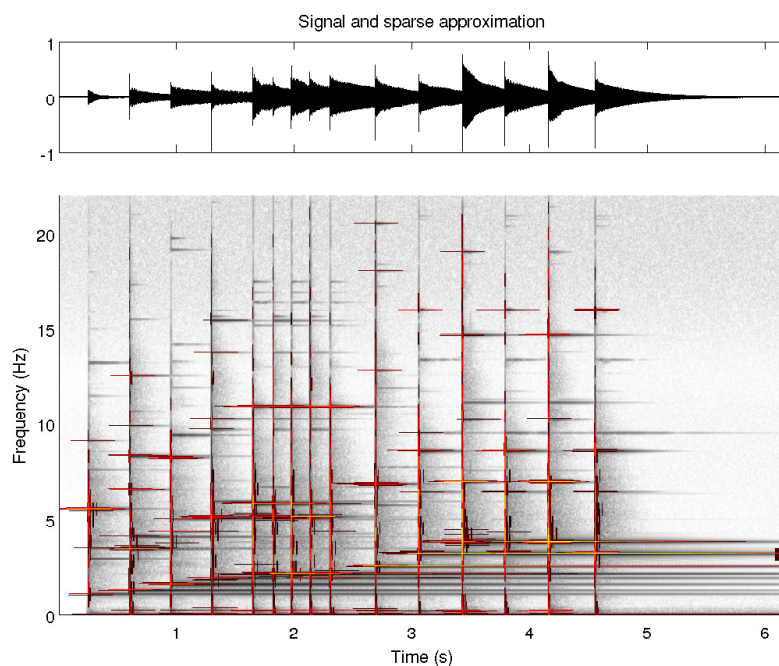
Modelado de transitorios

Matching Pursuit [Mallat and Zhang, 1993] sobre diccionarios redundantes:
enfoque de gradiente descendente para encontrar soluciones esparsas

Ejemplo: diccionario con átomos de Gabor de diferentes escalas temporales

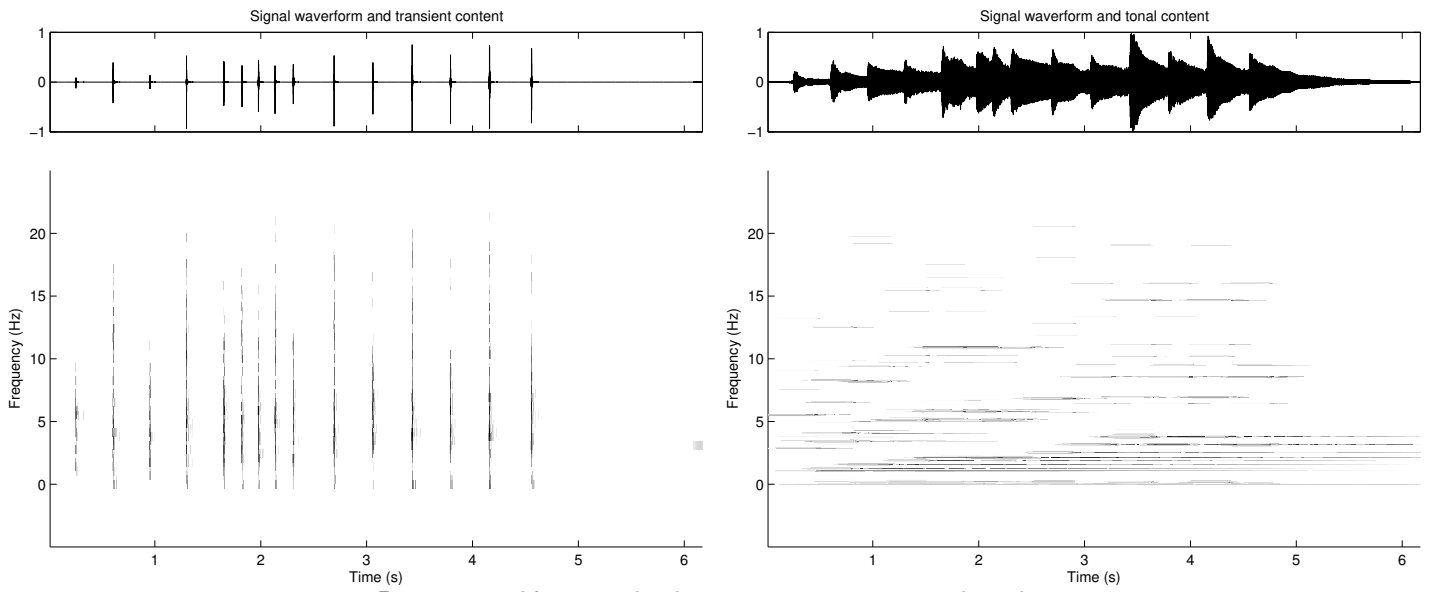


Ejemplo: Matching pursuit sobre diccionarios redundantes



Sonido original de glockenspiel y reconstrucción esparsa

Ejemplo: Matching pursuit sobre diccionarios redundantes

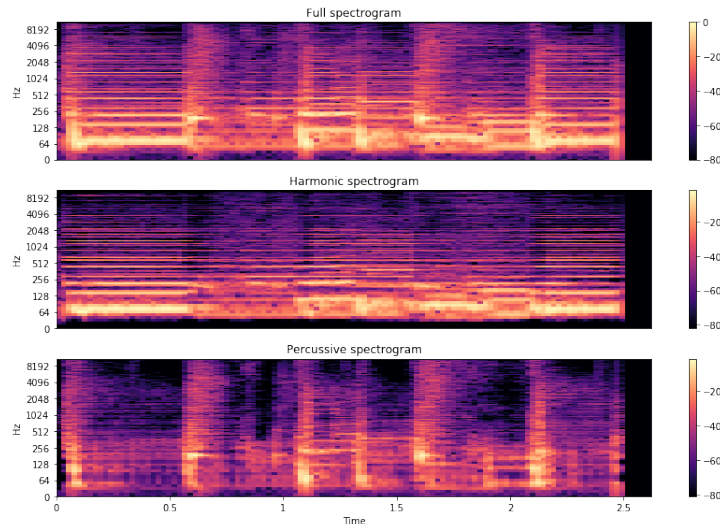


Reconstrucción: transitorios y componentes estacionarios

Modelado sinusoides + ruido + transitorios

Harmonic-percussive source separation





Filtrado de mediana del espectrograma en sentido horizontal y vertical para generar componentes armónicos y percusivos [FitzGerald, 2010].




Referencias I

-  Charpentier, F. (1986).
Pitch detection using the short-term phase spectrum.
In Proc. International Conf. Acoustics, Speech, Signal Processing, pages 113–116.
-  Dressler, K. (2006).
Sinusoidal Extraction Using and Efficient Implementation of a Multi-Resolution FFT.
In Proceedings of the DAFx-06, Montreal, Canada.
-  FitzGerald, D. (2010).
Harmonic/percussive separation using median filtering.
In Proceedings of the International Conference on Digital Audio Effects (DAFx), pages 246–253, Graz, Austria.
-  Keiler, F. and Marchand, S. (2002).
Survey on extraction of sinusoids in stationary sounds.
In Proceedings of the Digital Audio Effects (DAFx'02) Conference, pages 51–58.

Referencias II

-  Mallat, S. and Zhang, Z. (1993).
Matching pursuits with time-frequency dictionaries.
IEEE Transactions on Signal Processing, 41(12):3397–3415.
-  McAulay, R. and Quatieri, T. (1986).
Speech analysis/Synthesis based on a sinusoidal representation.
Acoustics, Speech, and Signal Processing, IEEE Transactions on, 34(4):744–754.
-  Quatieri, T. F. (2001).
Discrete-Time Speech Signal Processing: Principles and Practice.
Prentice Hall PTR, 1st edition.
-  Serra, X. and Smith, J. (1990).
Spectral Modeling Synthesis A Sound Analysis/Synthesis Based on a Deterministic plus Stochastic Decomposition.
Computer Music Journal, 14:12–24.

Referencias III

-  Verma, T. and Meng, T. (2000).
Extended spectral modeling synthesis with transient modeling synthesis.
Computer Music Journal, 24(2):47–59.