Análisis Espectral 1: Transformada Corta de Fourier y Ventanas

Juan-Pablo Cáceres
CCRMA
Stanford University

Agosto, 2007

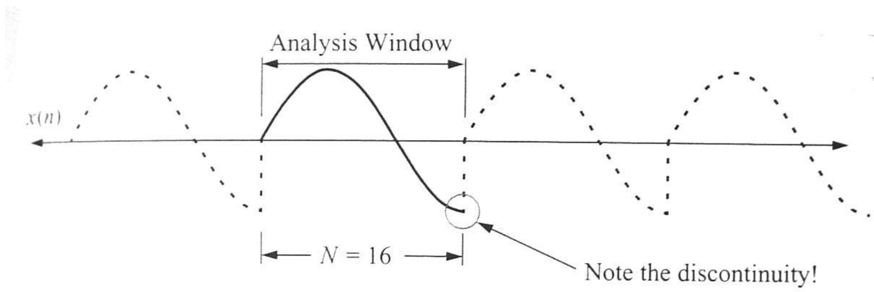
Contenidos

Análisis de Señales de la Vida Real

Recordemos que cuando se analizan señales de la vida real, en general sus componentes en frecuencia no estan alineadas con las frecuencias de análisis del DFT. Tomamos como ejemplo,

$$x[n] = \sin\left(f 2\pi \frac{n}{N}\right), \text{ con } N = 16,\ f = 3/4.$$

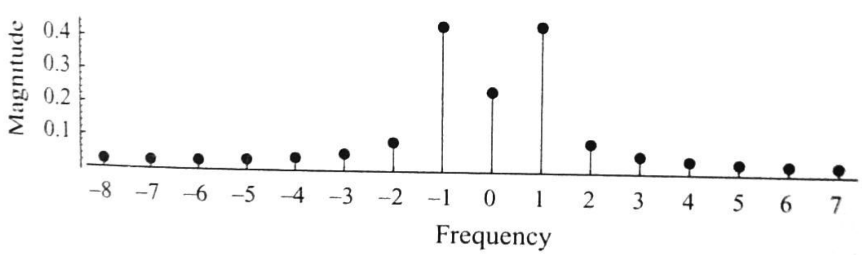
➡ **DFT interpreta la señal como 1 periodo de una señal periodica infinitamente larga.**



Discontinuidades en el Tiempo

Las **discontinuidades** periodicas en la señal producen un espectro con muchos armónicos en altas frecuencias.

➡ **La DFT "escucha" clicks en estas discontinuidades**
Vemos que el espectro contiene energía en todas las frecuenicas



Discontinuidades en la Señal

Sabemos por la ecuación:

$$x[n] = \sin\left(f 2\pi \frac{n}{N}\right), \text{ con } N = 16, \; f = 3/4$$

que nuestra señal sólo debiera contener una componente de frecuencia. Los 2 problemas son:

- *Resolución en Frecuencia*: Sólo vemos las frecuencias que son multiplos de nuestra frecuencia de análisis f_N
- *Escape (leakage)*: Discontinuidades en los bordes de la ventana de análisis producen esparcimiento de ruido en el resto del espectro.

Solución al problema de *Resolución en Frecuencia*
$\Rightarrow$ Zero-padding.

Como Resolver el Problema de Escape (Leakage)

Para resolver el problema de escape de frecuencias (leakage) podemos,

- Creamos una señal $w[n]$ de igual largo N que nuestra señal de análisis
- $w[n]$ gradualmente "fade in" y "fade out" en los bordes
- Multiplicamos $w[n]$ por $x[n]$ y así reducimos el efecto de las discontinuidades en los bordees

Sin embardo, estamos alterando la señal de análisis $x[n]$ por lo que la elección de $w[n]$ tendrá un efecto en el espectro resultante.

$\Rightarrow$ Este proceso se conoce como **"Windowing"**.

“Windowing” y Ventana Rectangular

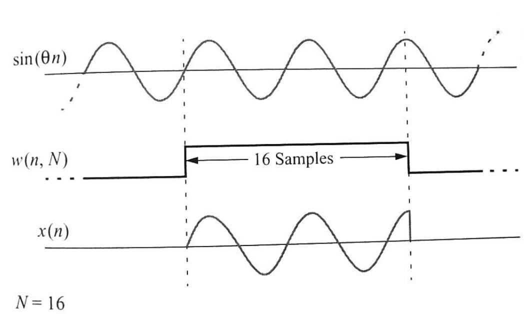
Cuando tomamos N puntos de una señal para analizar, estamos implícitamente multiplicando por una **Ventana Rectangular**:

$$w_R[n] \triangleq \begin{cases} 1 & 0 \leq n < N \\ 0 & \text{para otros } n \end{cases}$$

Por lo que la DFT es,

$$X[k] = \sum_{n=0}^{N-1} x[n] w_R[n] e^{-\frac{2\pi j}{N} kn}$$

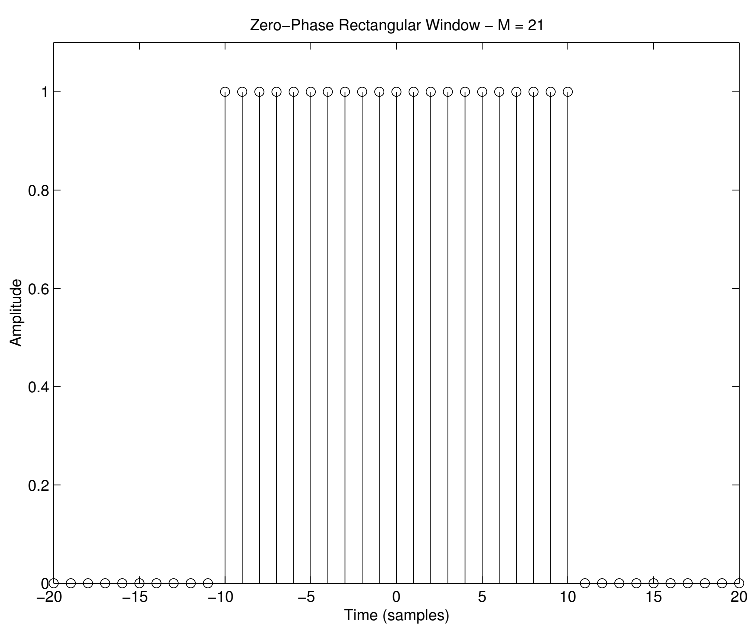
El problema con la Ventana Rectangular es que introduce discontinuidades



Análisis de la Ventana Rectangular

Podemos reesscribir la ventana rectangular de modo que sea símetrica con respecto a 0 para que tenga Fase-Cero,

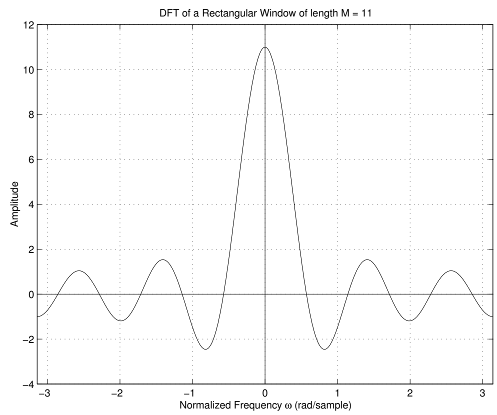
$$w_R[n] \triangleq \begin{cases} 1 & |n| \geq \frac{M-1}{2} \\ 0 & \text{para otros } n \end{cases}$$



Análisis de la Ventana Rectangular

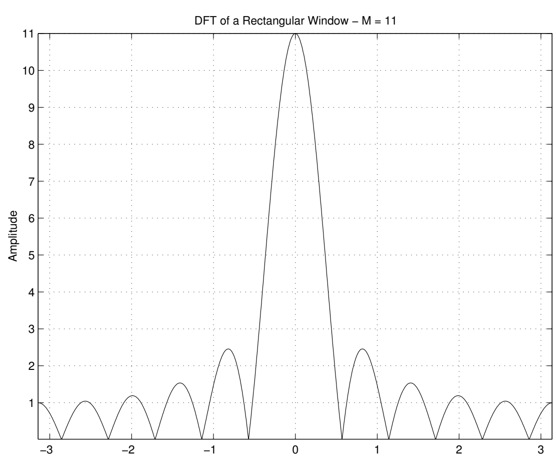
La tranformada de $w_R[n]$ es,

$$W_R(\omega) = \frac{\sin(M\omega/2)}{\sin(\omega/2)}$$

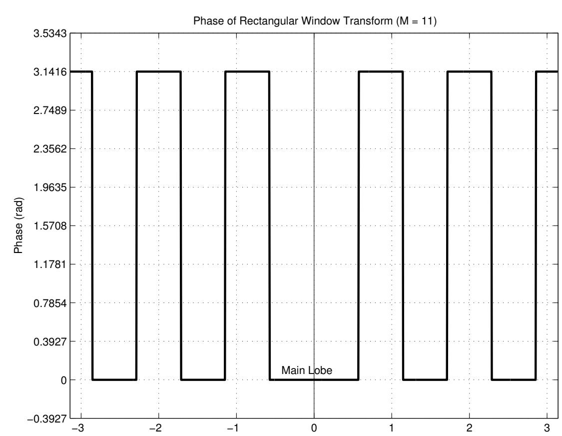


Notar que ésta es la **transformada completa (real)**, no sólo la magnitud. Sólo se obtienen estas transformadas para ventanas simétricas centradas en 0.

Espectro de la Ventana Rectangular

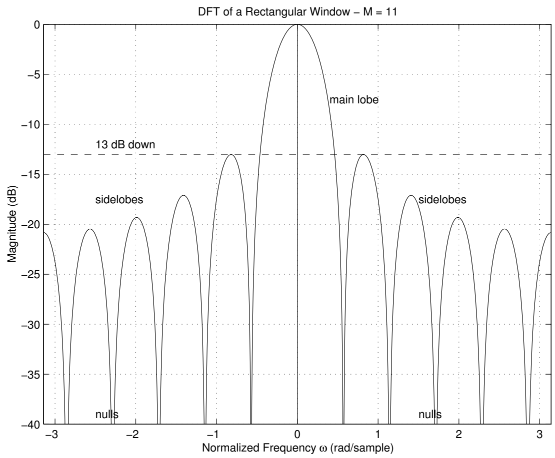


Respuesta de Amplitud

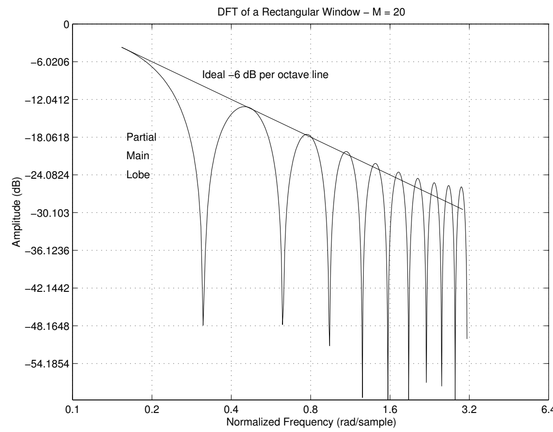


Respuesta de Fase

Interpretación en Audio



Respuesta de Amplitud en dB



“Rolloff” de 6 dB pot octava

Propiedades de Ventanas

En genral, las ventanas tendran 3 propiedades que son las que nos interesan y en las que basaremos nuenstra elección:

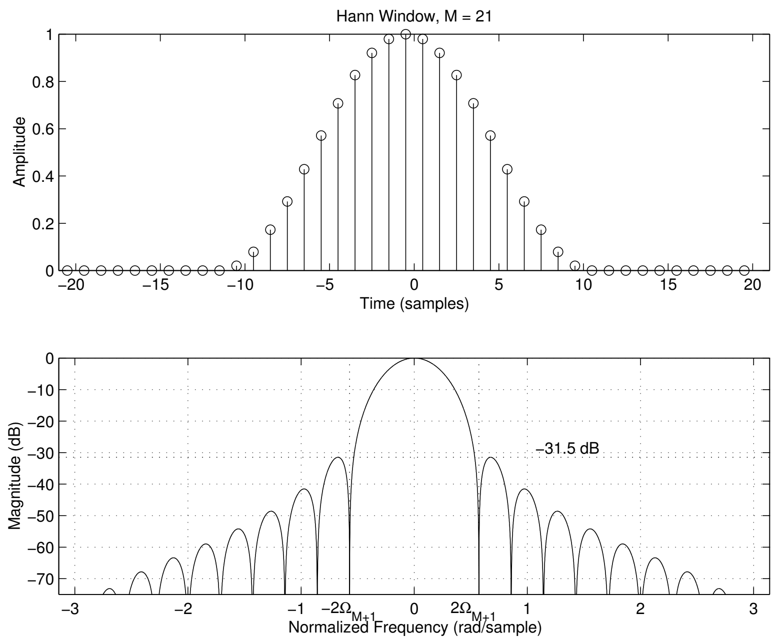
- Ancho del lóbulo principal (main lobe width)
- Nivel de lóbulos laterales (sidelobe level)
- Pendiente de caida (rolloff)

Lo que pasa en general es que ventanas con *ancho del lóbulo principal* **mayor** tienen **menores** *niveles de lóbulos laterales*.

Otras Ventanas

Hann o Hanning o “Raised Cosine”

$$w_H(n) = w_R(n)\left[\frac{1}{2} + \frac{1}{2}\cos(\Omega_M n)\right] = w_R(n)\cos^2\left(\frac{\Omega_M}{2} n\right)$$



Otras Ventanas

Propiedades de la ventana Hann:

- Lóbulo principal es $4\Omega_M$ de ancho
- El primer lóbulo es de -31 dB
- Lóbulos laterales decaen a una tasa de ≈ 18 dB / octava

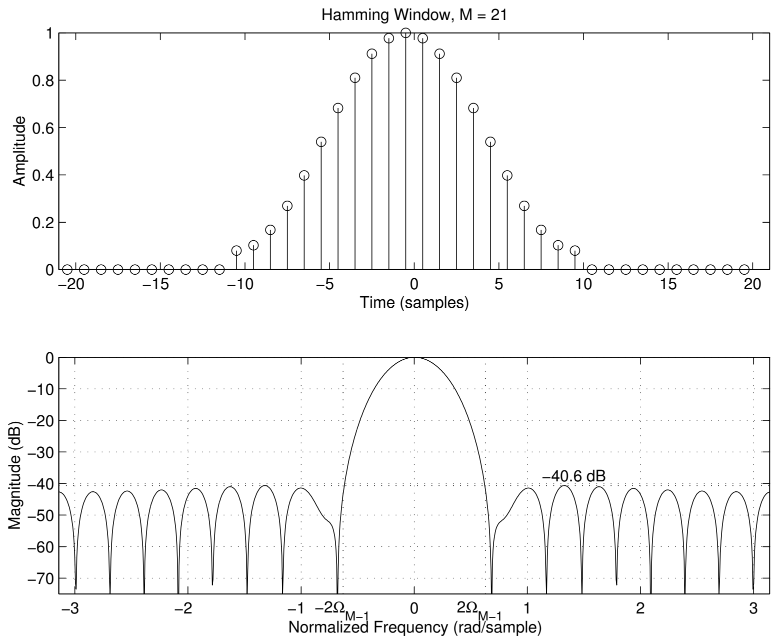
Comparado con la ventana Rectangular:

- Lóbulo principal es $2\Omega_M$ de ancho
- El primer lóbulo es de -13 dB
- Lóbulos laterales decaen a una tasa de ≈ 6 dB / octava

Nota: $\Omega_M \triangleq \frac{2\pi}{M}$

Otras Ventanas

Ventana de Hamming



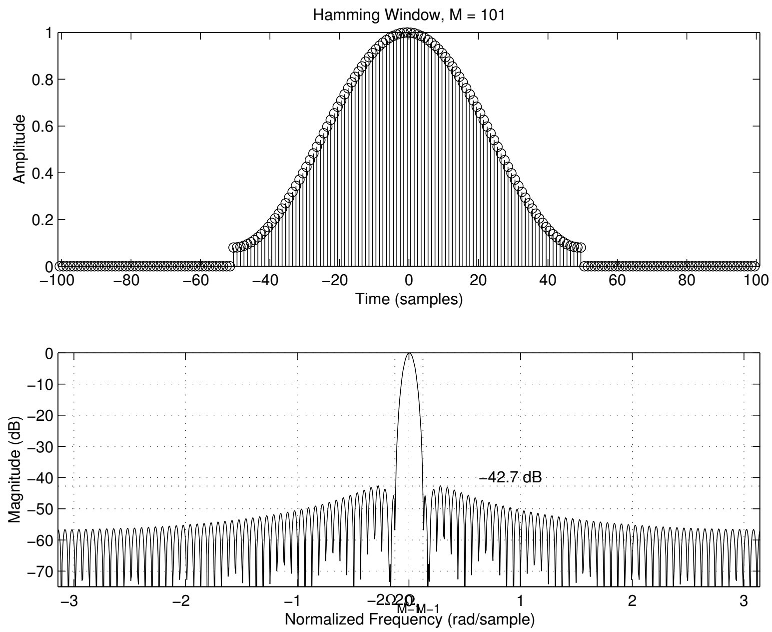
Otras Ventanas

Propiedades de la ventana Hamming

- Discontinuidades "golpean a cero" en los extremos
- Lóbulo principal es $4\Omega_M$ de ancho (como Hann)
- Lóbulos laterales decaen a una tasa de ≈ 6 dB / octava (pero "aliased")
- Primer lóbulo lateral mejorado con respecto a Hann
- Lóbulos laterales cercanos a ser "rizados igual" ("esequal ripple")

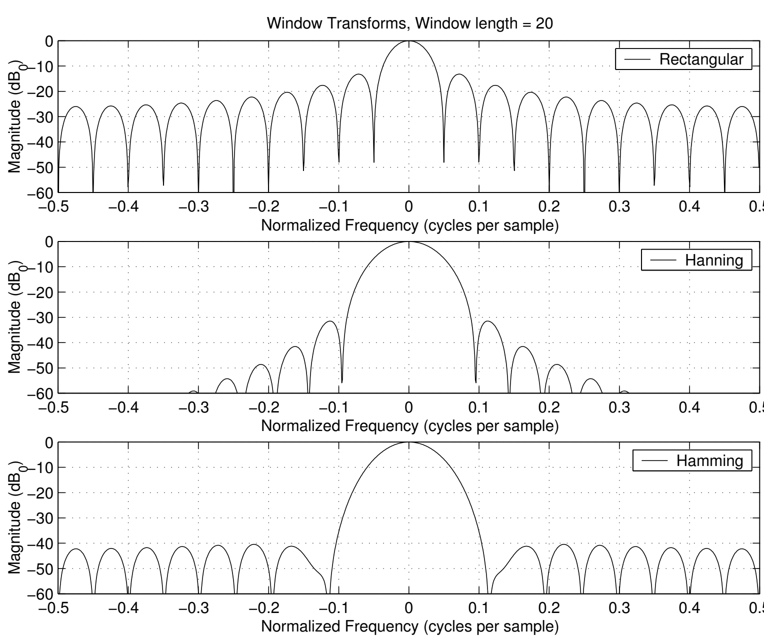
Otras Ventanas

Ventana de Hamming más larga



Otras Ventanas

Comparando las 3 ventanas



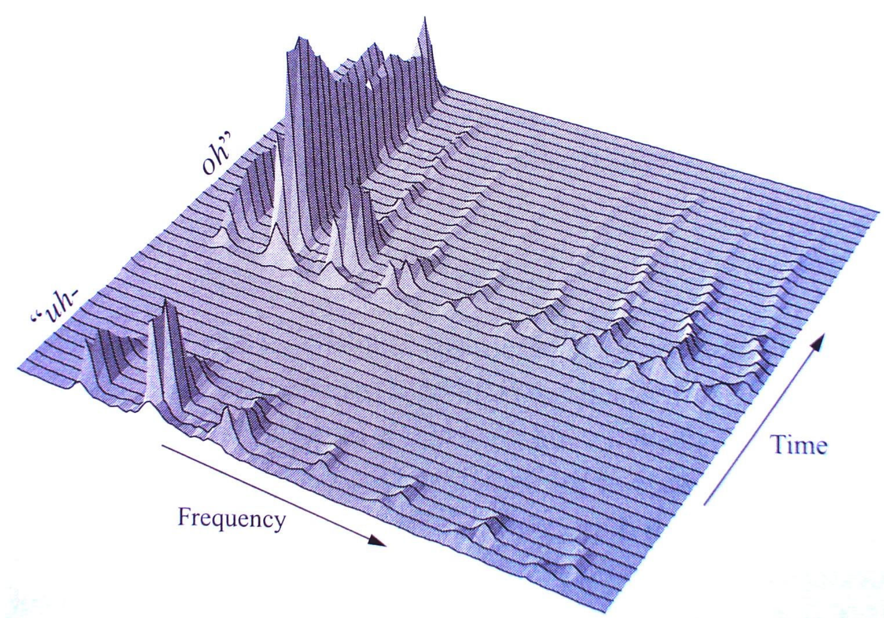
Otras Ventanas

Otras ventanas:

- Blackman-Harris
- Potencias-de-Coseno
- Bartlett ("Triangular")
- Poisson
- Hann-Poisson
- Kaiser-Bessel
- Dolph-Chebyshev
- Gaussian

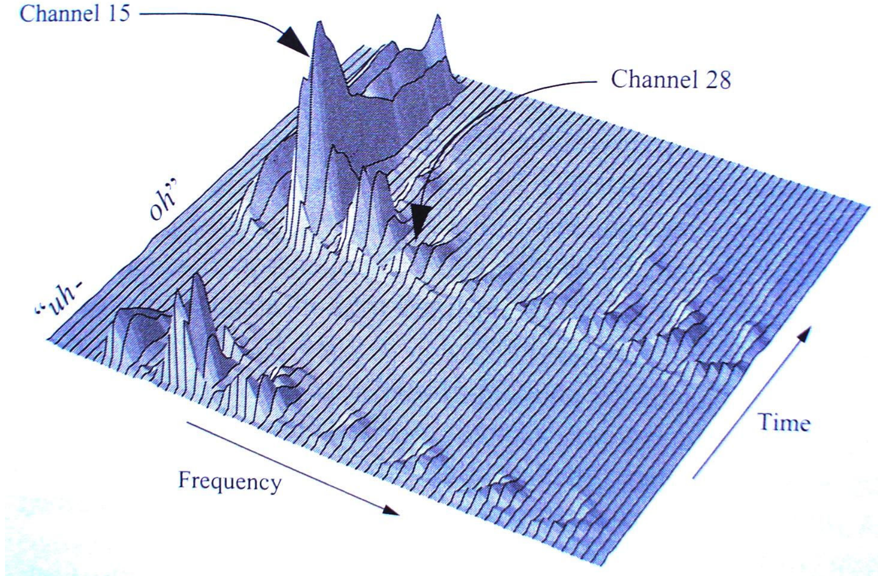
STFT: 2 Visiones

Existen dos visiones equivalentes de la **Tranformada Corta de Fourier (STFT)**



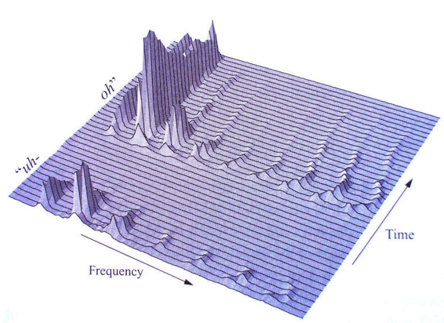
Transformada de Fourier por Ventanas

STFT: 2 Visiones

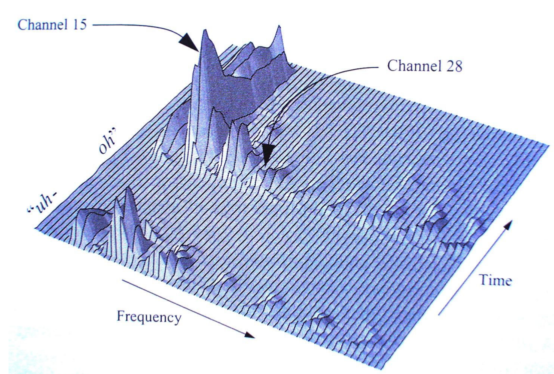


Banco de Filtros

STFT: 2 Visiones

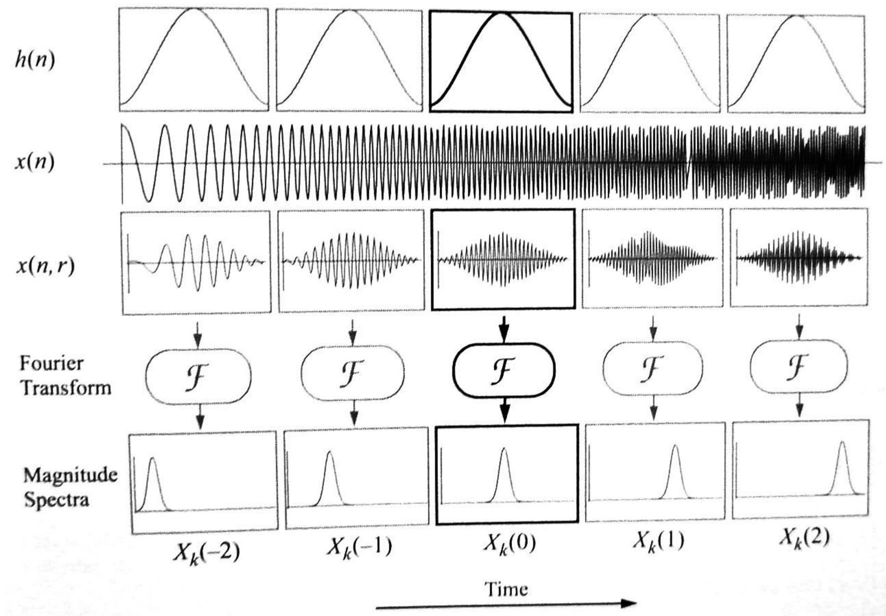


Transformada de Fourier por Ventanas



Banco de Filtros

Visión de Transformada de Fourier de la STFT



Interpretación Transformada de Fourier "Windowed" de la STFT.

La Transformada Corta de Fourier

La Transformada Corta de Fourier (STFT) se define como,

$$\boxed{X_m(\omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-mR]e^{-j\omega n}}$$

con

$$\begin{aligned} x[n] &= \text{señal de entrada en el tiempo } n \\ w[n] &= \text{ventana de largo } M \text{ (e.g., Hamming)} \\ X_m(\omega) &= \text{DTFT de señal multiplicada por ventana en tiempo } mR \\ R &= \text{largo de los saltos, en samplers, entre sucesivas DTFTs.} \end{aligned}$$

Inversa de La Transformada Corta de Fourier

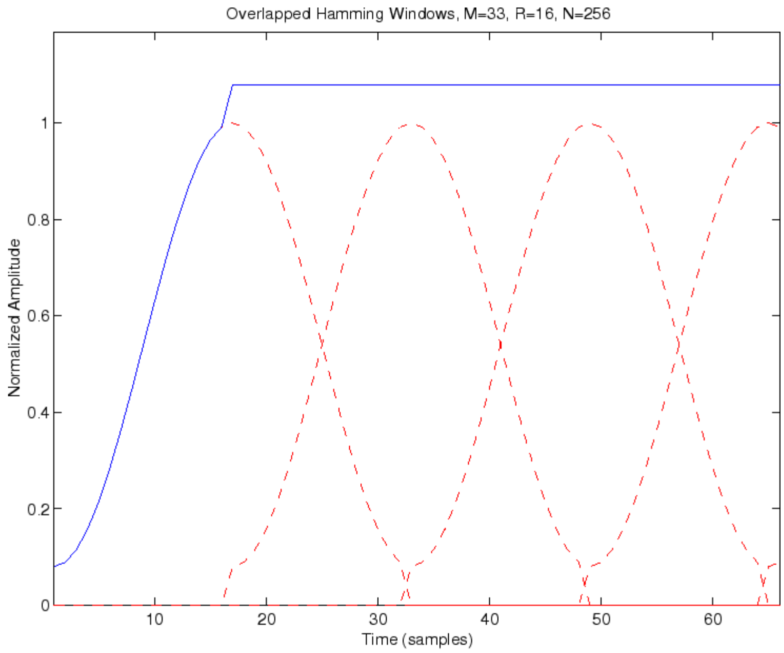
Si la ventana $w(n)$ tiene propiedad de *Constant OverLap-Add (COLA)* para saltos R:

$$\sum_{m=-\infty}^{\infty} w[n-mR] = 1,\ \forall n \in \mathbb{Z}\ (w \in \mathsf{COLA}(R))$$

entonces la suma de las DTFTs individuales será la DTFT de todo x:

$$\begin{aligned}
\sum_{m=-\infty}^{\infty} X_m(\omega) &\triangleq \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} x[n]w[n-mR]e^{-j\omega n} \\
&= \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n} \underbrace{\sum_{m=-\infty}^{\infty} w[n-mR]}_{1 \text{ si } w\in\mathsf{COLA}(R)} \\
&= \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n} \\
&\triangleq \mathsf{DTFT}_\omega(x) = X(\omega) \\
\longleftrightarrow \quad x[n] &= \frac{1}{2\pi}\int_{-\pi}^{\pi} X(\omega)e^{j\omega n}d\omega
\end{aligned}$$

Ejemplo de COLA

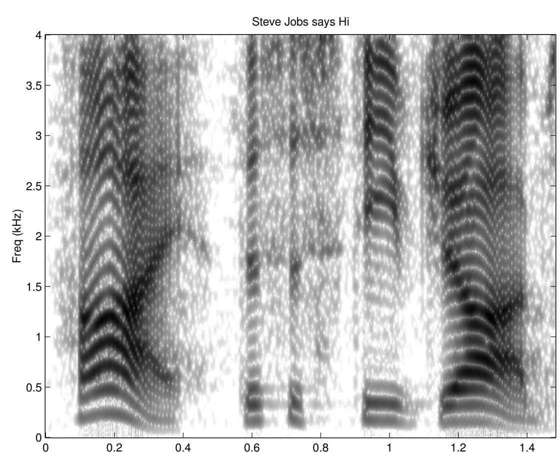


Ejemplo de COLA para Hamming

Espectrogramas (o Sonogramas)

- Plot de intensidad en dB de la STFT
- Comunmente llamado *sonograma* cuando es aplicado a audio
- Usado extensamente en análisis de voz, computer music y síntesis de sonido

- La audición humana funciona como y tipo de espectrograma en tiempo real con la cóclea del oído interno
- Modelamiento espectral está basado en la síntesis del espectro de tiempo corto.



Resolución en Frecuencia versus Temporal

Obtener **mayor resolución en frecuencia** requiere un largo de ventana M **mayor**.

Obtener **mayor resolución temporal** requiere un largo de ventana M **menor**.

Con esto en mente, los parámetros de un espectrograma son:

- Largo de ventana M
- Tipo de ventana (Hamming, Kaiser, etc.)
- Salto R
- Largo N de la FFT

Demo snd

Visión de Banco de Filtros de la SFTF

Si reagrupamos los términos de la definición de la STFT obtenemos la interpretación de *banco de filtros*:

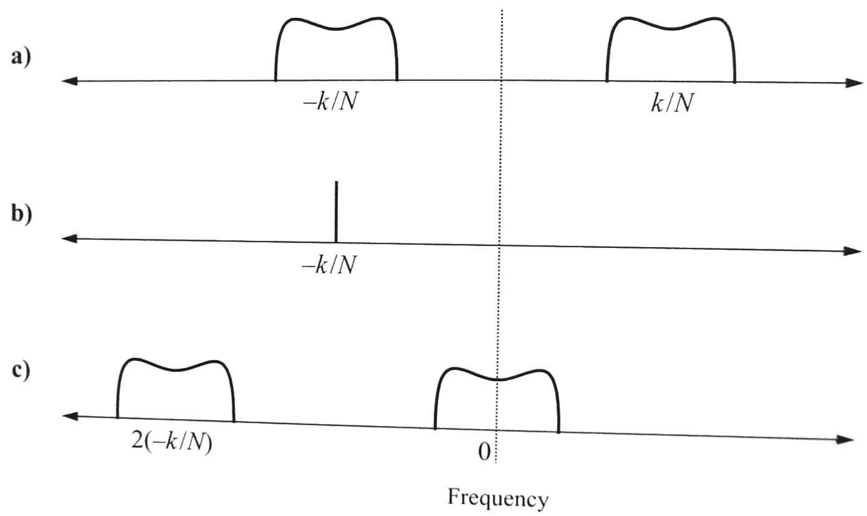
$$X_m(\omega_k) = \sum_{n=-\infty}^{\infty} \underbrace{[x[n]e^{-j\omega_k n}]}_{x_k[n]} w[n-m]$$
$$= [x_k * \textsc{Flip}(w)](m)$$

La operativa de esta interpretación es,

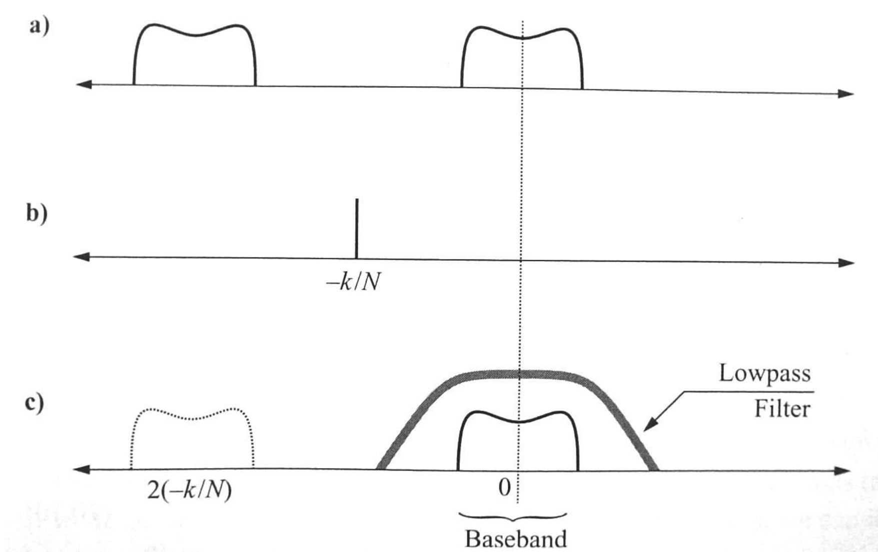
1. Se rota el espectro de x de modo de mover la frecuencia ω_k hacia 0 (se modula por $e^{-j\omega_k n}$ en el dominio del tiempo)
2. Se usa w como un filtro pasa-bajos en frecuencia 0

La STFT es interpretada como un set ordenado en frecuencia de señales en el tiempo de banda angosta.

Visión de Banco de Filtros de la SFTF

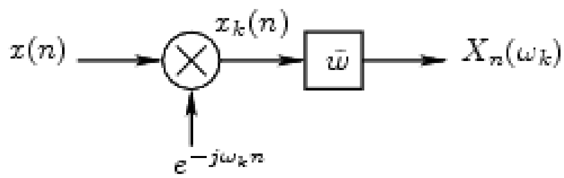


Visión de Banco de Filtros de la SFTF



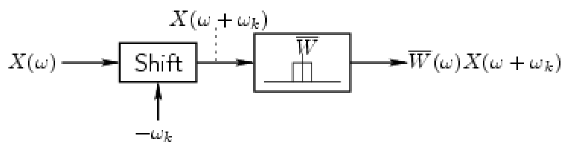
Visión de Banco de Filtros de la SFTF

Dominio del tiempo:



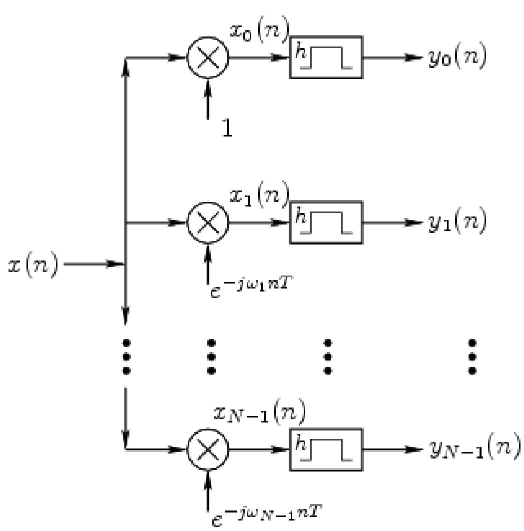
Un canal del banco de filtros de la STFT

Dominio de la Frecuencia:



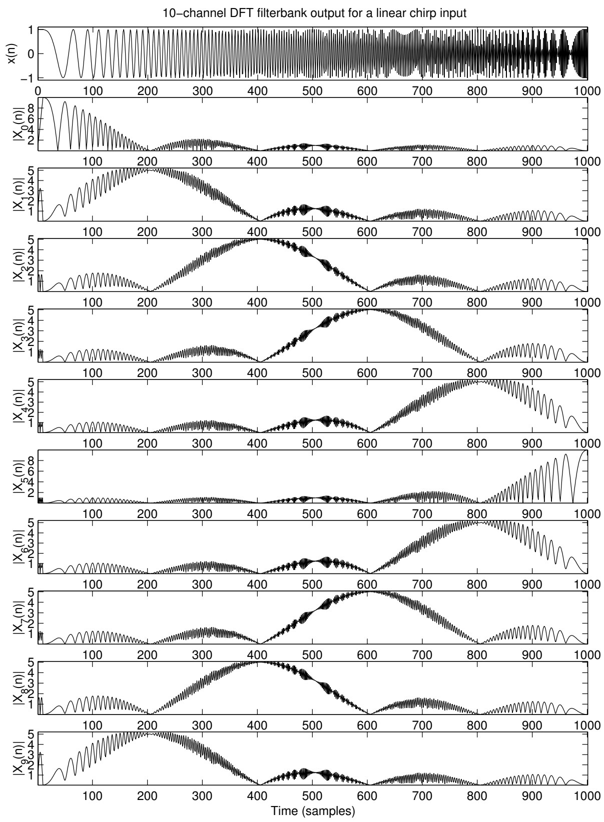
Un canal del banco de filtros de la STFT

Ejemplo con Ventana Cuandrada

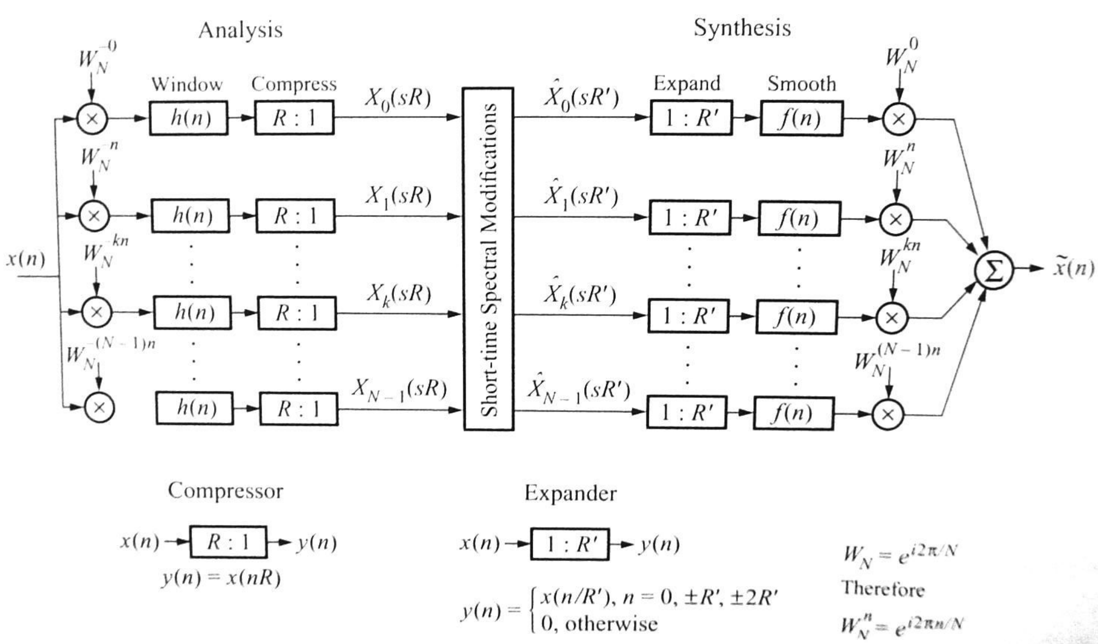


Banco de filtros de la STFT

Ejemplo de Output con Chirp



Análisis/Síntesis de la STFT



Conclusiones

La Transformada Corta de Fourier se aplica en

1. Espectrogramas
2. Peak Tracking en audio
3. Phase Vocoder
4. Sines+Noise modeling