

Introducción a la Teoría del Procesamiento Digital de Señales de Audio

Aplicaciones de la Transformada de Fourier al análisis y síntesis de sonido

Resumen

Técnicas de análisis y síntesis de sonido

Phase vocoder

- Orígenes, teoría y modelo

- Implementación usando banco de filtros

- Implementación con STFT

Aplicaciones

Modelado espectral

- Modelado sinusoidal

- Modelado del residuo

Técnicas de análisis y síntesis de sonido

Al generar sonido en una computadora existe siempre un **modelo** subyacente cuyos parámetros controlan la síntesis.

Varias de las diversas técnicas existentes pueden agruparse bajo la categoría de **análisis y síntesis**, por ejemplo el phase vocoder y el modelado espectral.

El análisis es la etapa en la que se determinan los parámetros del modelo, mientras que la síntesis consiste en la salida del modelo.

La síntesis puede ser virtualmente igual al sonido original, o pueden introducirse modificaciones alterando los parámetros del modelo.

La utilidad musical y el valor perceptivo del resultado dependen en gran medida en el grado de concordancia entre el modelo y la señal original.

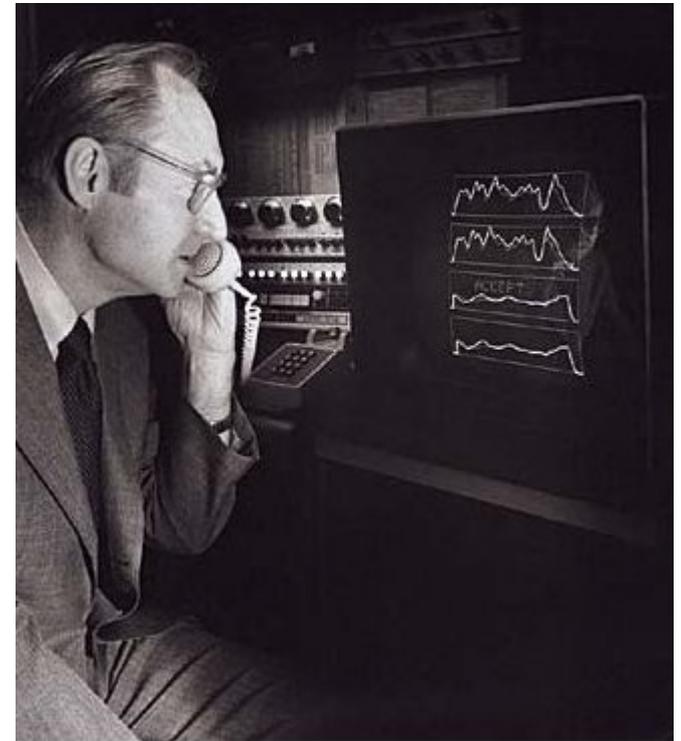
Phase vocoder

Origen: técnicas de codificación de voz
(vocoder – voice coder)

minimizar la cantidad de información
trasmitida, manteniendo la intelegibilidad

Flanagan y Golden, 1966
describen la técnica por primera vez
“Phase vocoder”, Bell System Technical Journal,,
1493-1509, 1966

Phase vocoder – se utiliza el término
phase para distinguir la técnica de los
codificadores tradicionales (channel
vocoder) y hacer énfasis en la
codificación del espectro de fase



Flanagan, Bell Labs, 1972

Phase vocoder

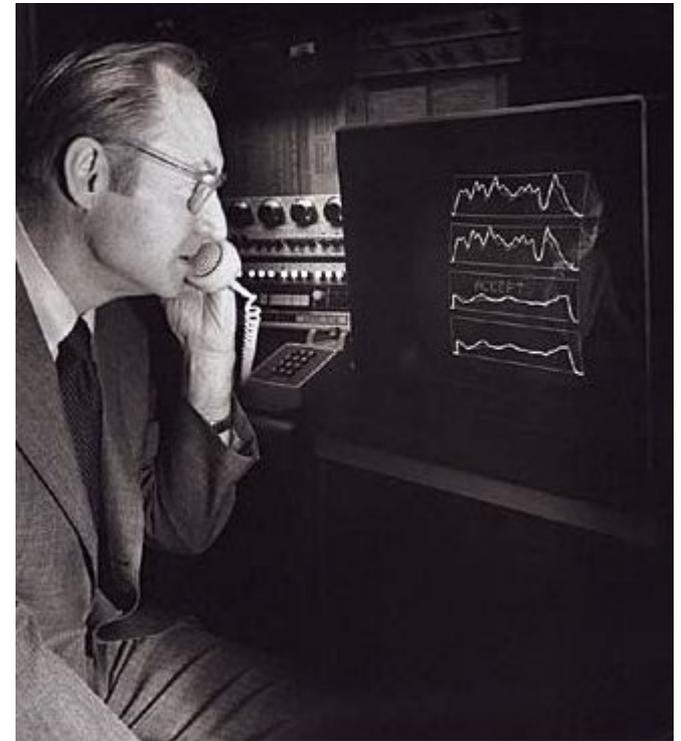
Origen: técnicas de codificación de voz
(vocoder – voice coder)

minimizar la cantidad de información
trasmitida, manteniendo la intelegibilidad

Flanagan y Golden, 1966
describen la técnica por primera vez
“Phase vocoder”, Bell System Technical Journal,,
1493-1509, 1966

Portnoff, "Implementation of the digital phase
vocoder using the fast fourier transform," IEEE
Trans. Acoustics, Speech, and Signal Processing,
Vol. ASSP-24, No. 3, June 1976.

Dolson, "The phase vocoder: a tutorial," Computer
Music Journal, Spring, Vol. 10, No. 4, 14-27, 1986.

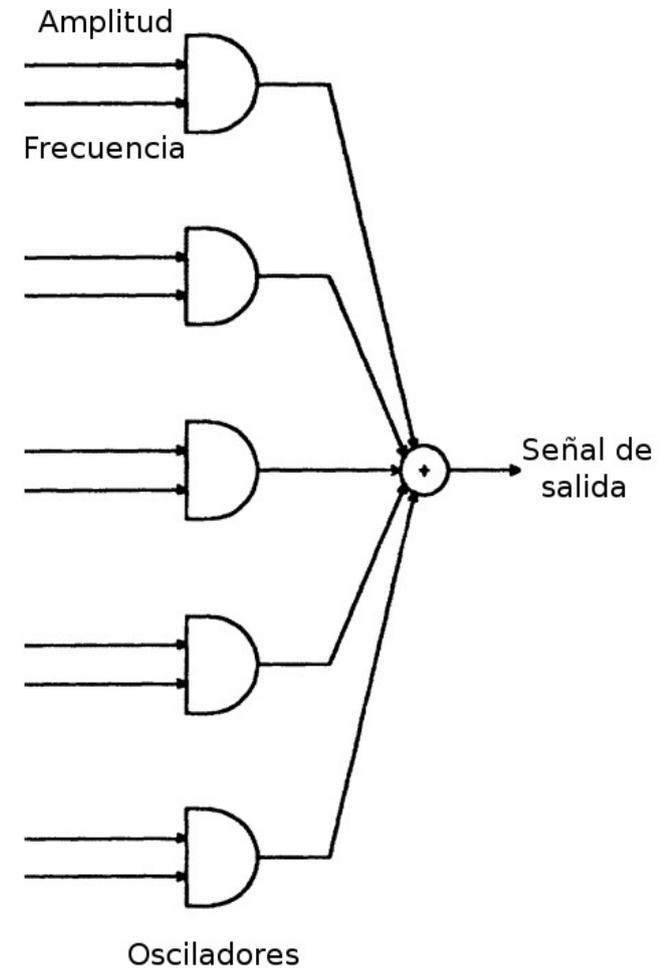


Flanagan, Bell Labs, 1972

Phase vocoder

Modelo: suma de sinusoides

Parámetros: amplitud y frecuencia de cada senoide (variables en el tiempo)



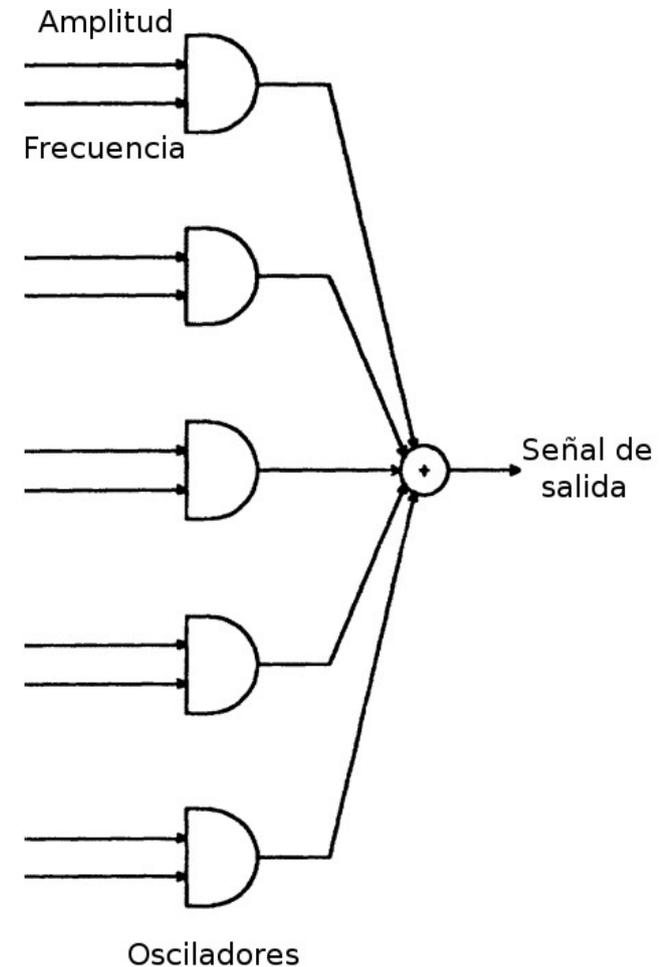
Phase vocoder

Modelo: suma de sinusoides

Parámetros: amplitud y frecuencia de cada senoide (variables en el tiempo)

En principio no se requiere que los parciales sean armónicos

Varios sonidos percutivos, o del tipo señal+ruido se escapan del modelo



Phase vocoder

Modelo: suma de sinusoides

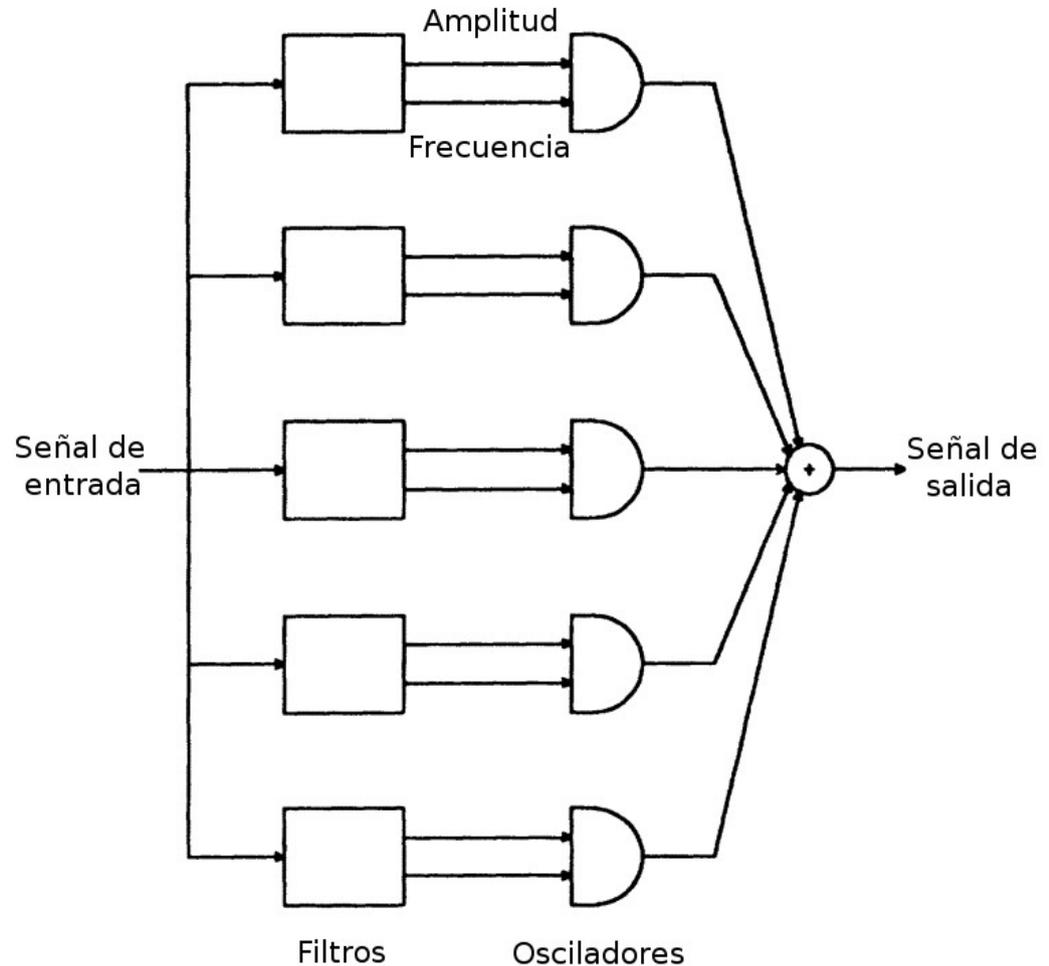
Parámetros: amplitud y frecuencia de cada senoide (variables en el tiempo)

En principio no se requiere que los parciales sean armónicos

Varios sonidos percutivos, o del tipo señal+ruido se escapan del modelo

El **análisis** puede interpretarse como un **banco de filtros** fijo

La salida de cada filtro es una amplitud y una frecuencia variables en el tiempo



Phase vocoder

Implementación del phase vocoder basada en banco de filtros

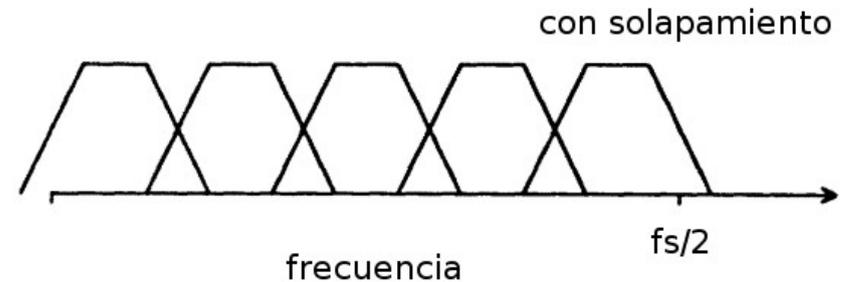
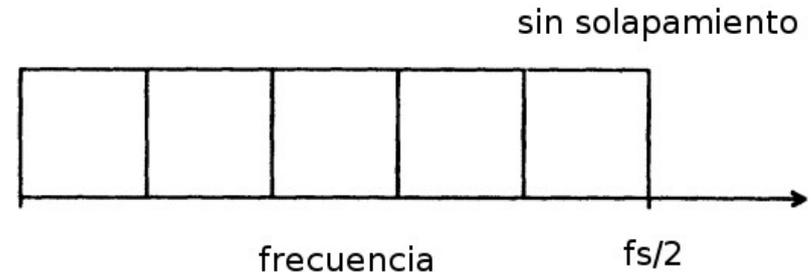
Cada filtro con la misma respuesta

Igualmente distribuidos en $0 - fs/2$

La respuesta combinada es plana

Suficiente número de filtros para asegurar un único parcial por filtro

Compromiso:
bajo solapamiento entre filtros implica respuesta al impulso larga

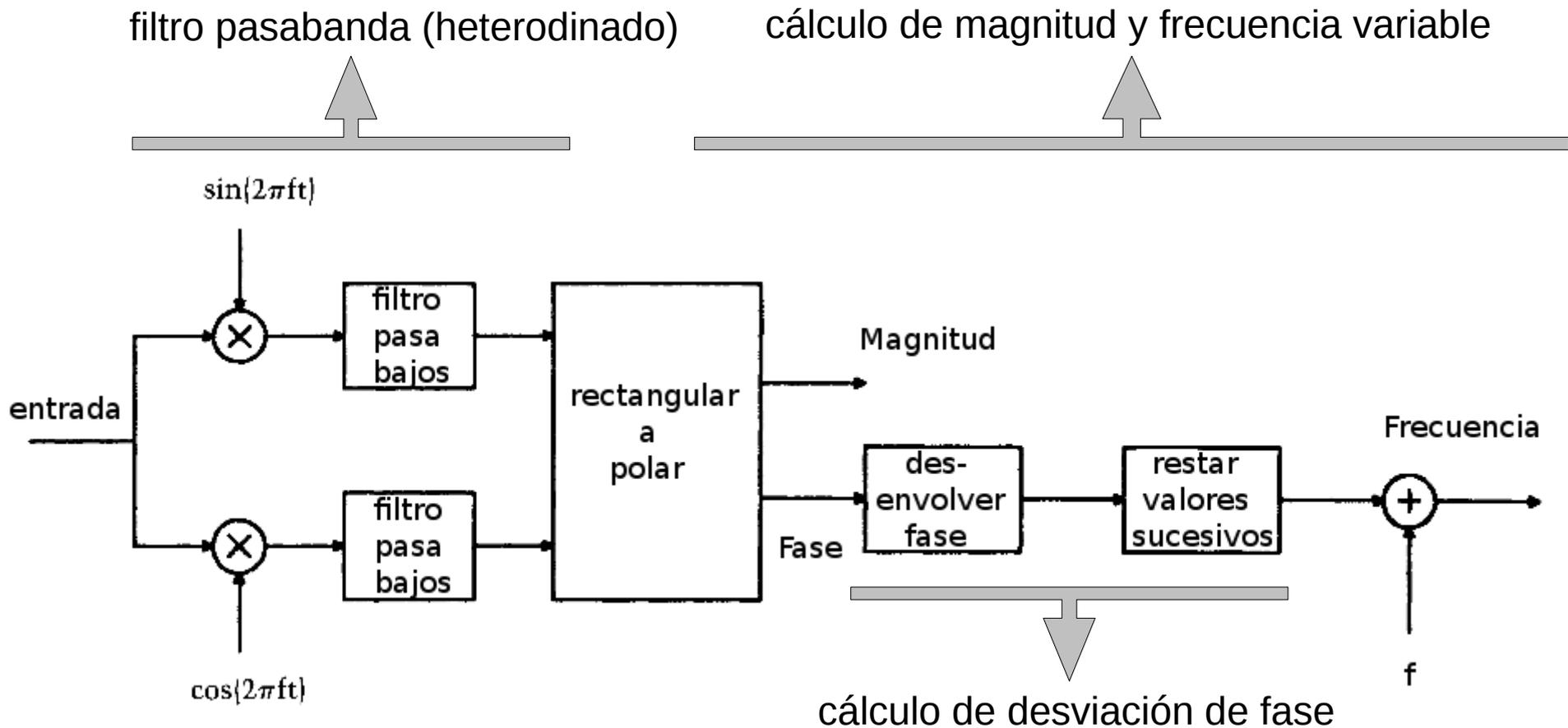


Sonidos de variación lenta pueden tolerar filtros de respuesta al impulso más larga, mientras que sonidos de variación rápida requieren filtros de respuesta también más rápida.

Phase vocoder

Implementación del phase vocoder basada en banco de filtros

La salida es una magnitud y una frecuencia variables en el tiempo



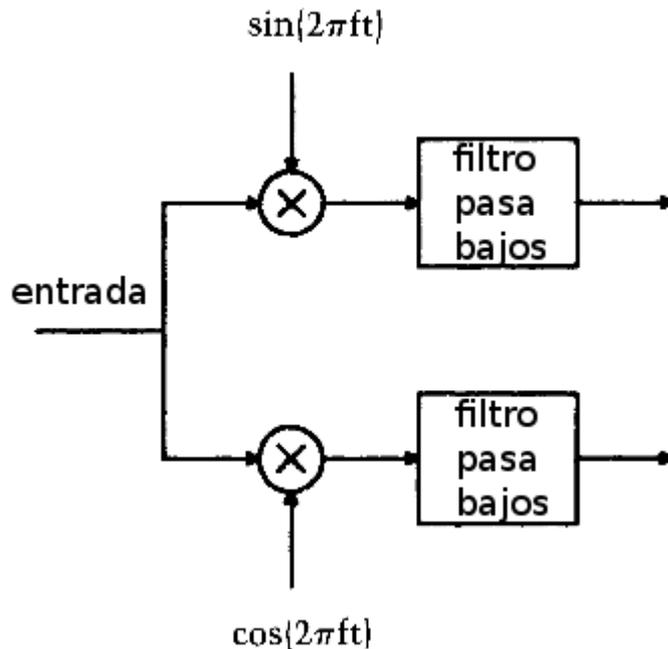
Phase vocoder

filtro pasabanda:

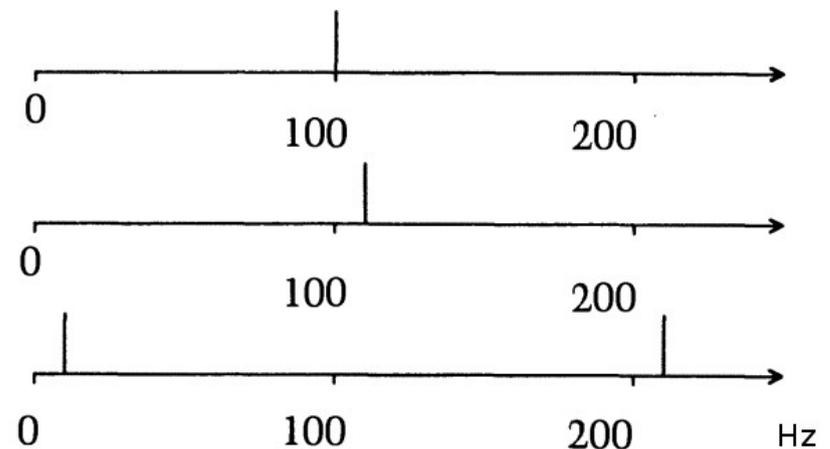
producto con sinusoides de frecuencia f en cuadratura

tiene el efecto de mover los componentes de frecuencia \pm la frecuencia de la senoide (modulación de amplitud)

el pasabajos retiene sólo las frecuencias de interés (heterodinado)



$$\begin{aligned} & \cos(2\pi 110 t)\cos(2\pi 100 t) \\ &= \cos(2\pi (110 - 100)t) + \cos(2\pi (110 + 100)t) \end{aligned}$$

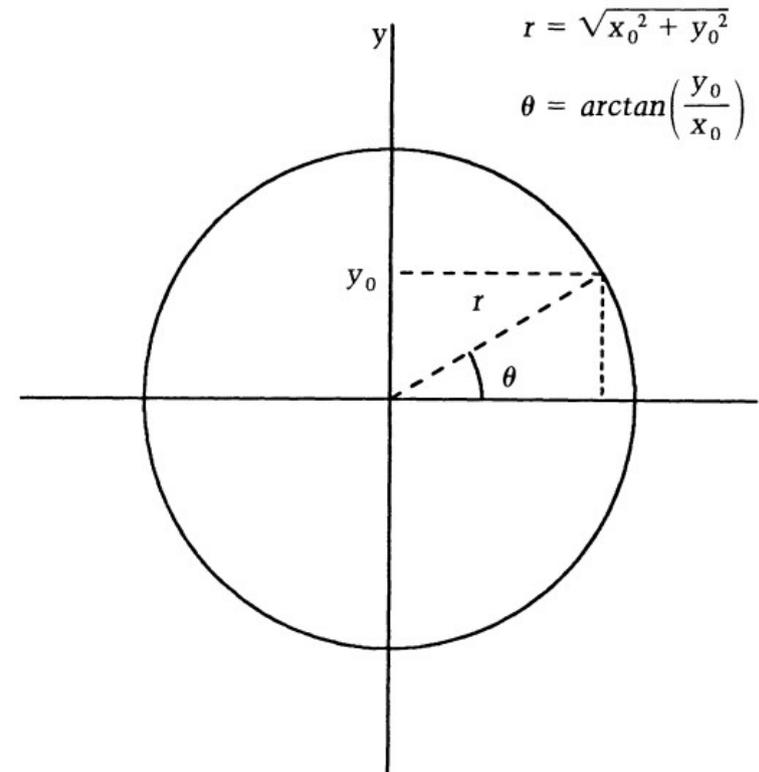
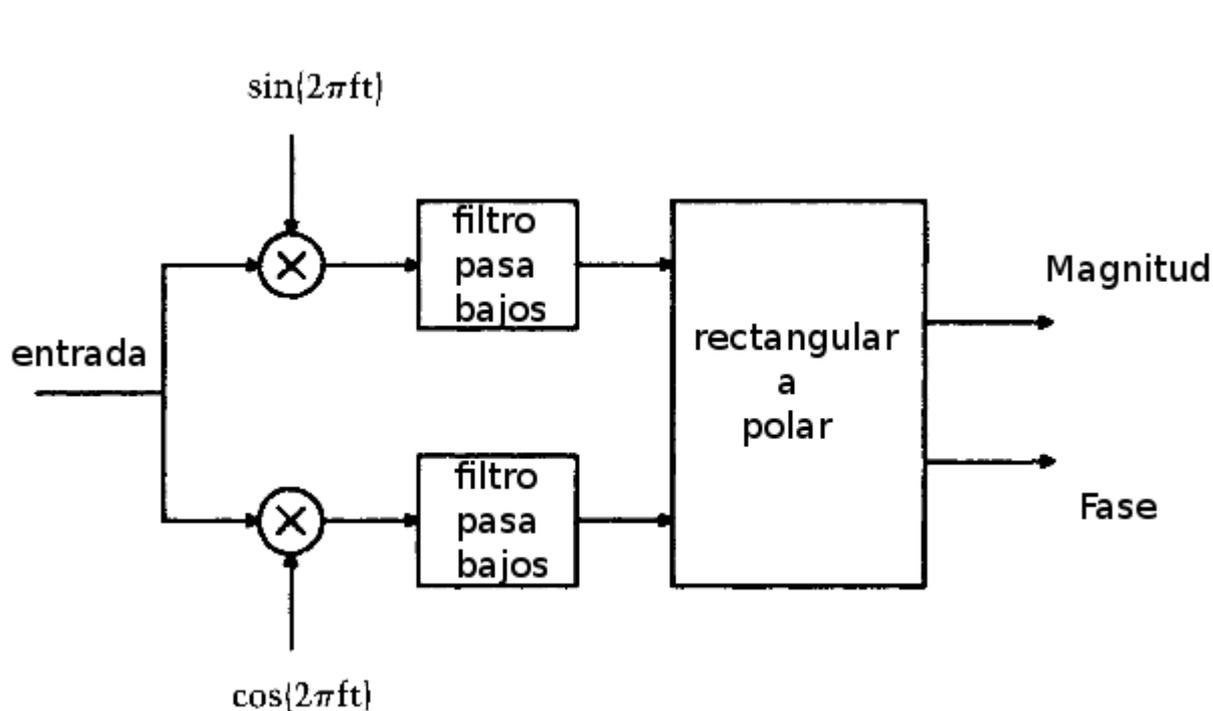


Phase vocoder

de coordenadas rectangulares a polares:

la salida del heterodinado son dos sinusoides desfasadas 90 grados

se pasa de esta representación rectangular a magnitud y fase

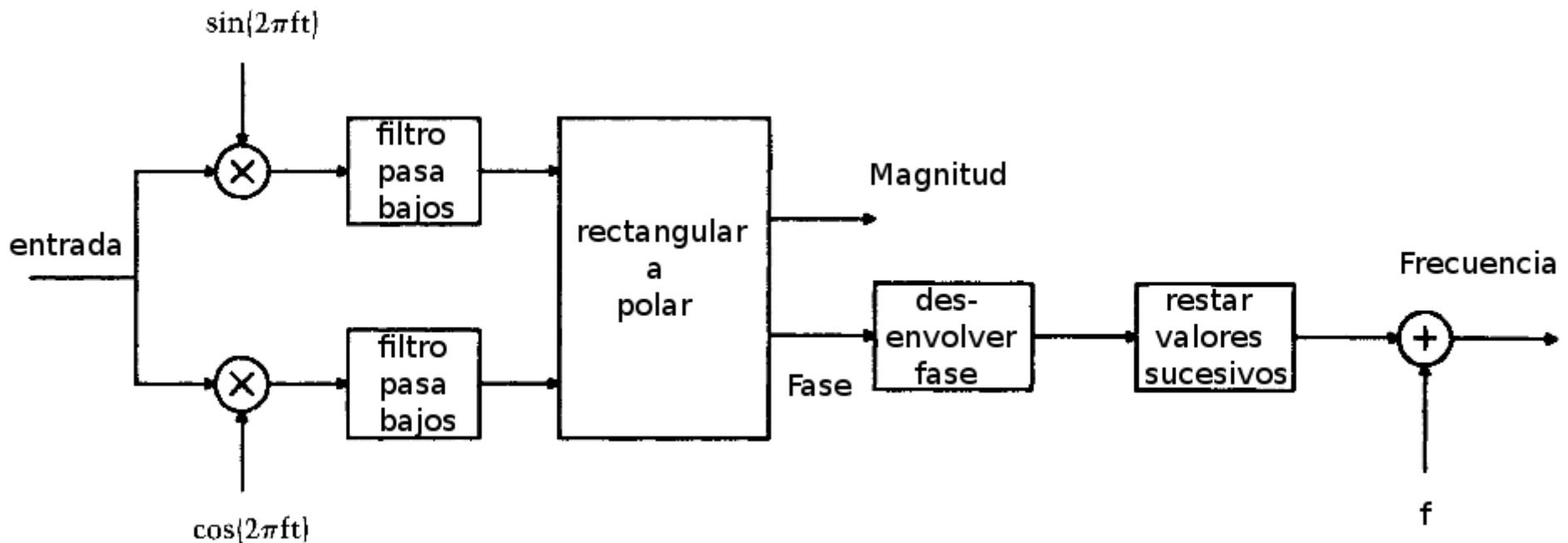


Phase vocoder

de fase a frecuencia:

la frecuencia es la variación de la fase por unidad de tiempo

se calcula el cambio de fase entre dos muestras sucesivas dividido el período de muestreo y se suma la frecuencia original

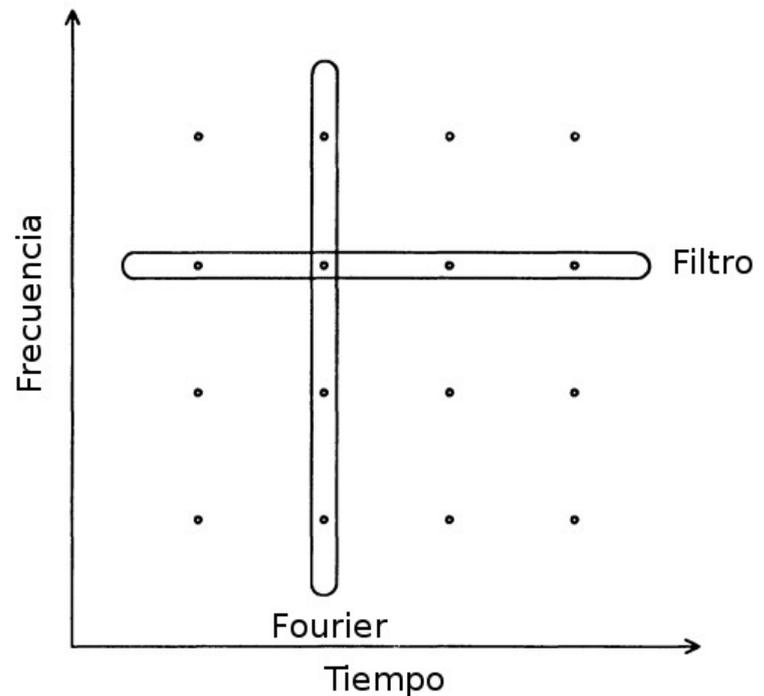
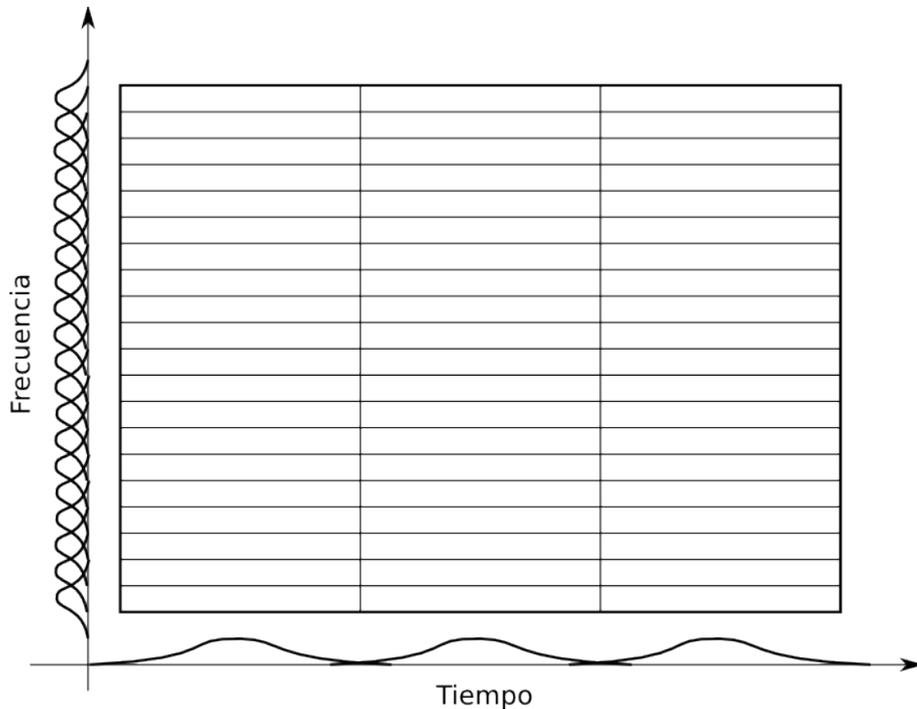


Phase vocoder

Implementación con la STFT

banco de filtros dado por la DFT y el tipo de ventana, descomposición en seno y coseno equivalente

cálculo de la magnitud y la fase, cambio de fase entre ventanas sucesivas y cálculo de frecuencia

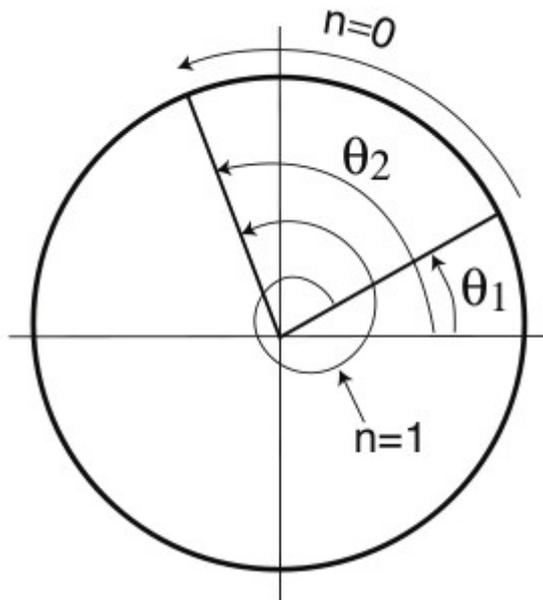


Phase vocoder

Implementación con la STFT

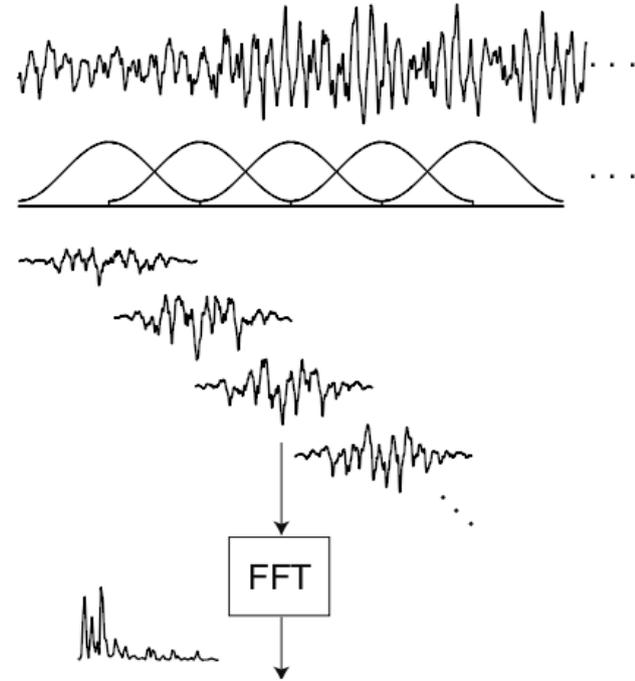
análisis, procesamiento y
síntesis en tramas

para cada bin de la DFT se
compara la fase entre ventanas
sucesivas



$$2\pi f (t_2 - t_1) = (\theta_2 - \theta_1) + 2\pi n$$

$$f = \frac{(\theta_2 - \theta_1) + 2\pi n}{2\pi (t_2 - t_1)}$$



Phase vocoder

Ejemplo

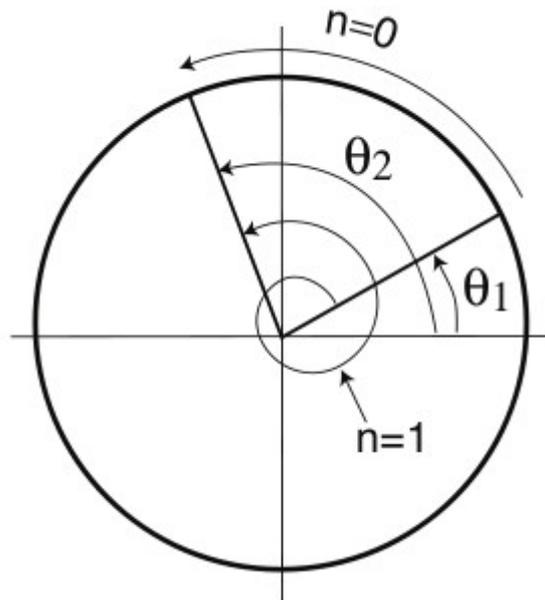
$f = 220 \text{ Hz}$, $f_s = 44.1 \text{ kHz}$, $N = 2048$

$f_s/N = 21.53 \text{ Hz}$, pico en $f_k = 215 \text{ Hz}$

$\Delta t = (t_2 - t_1) = 2048/2 * (1/f_s) = 0.023 \text{ s}$

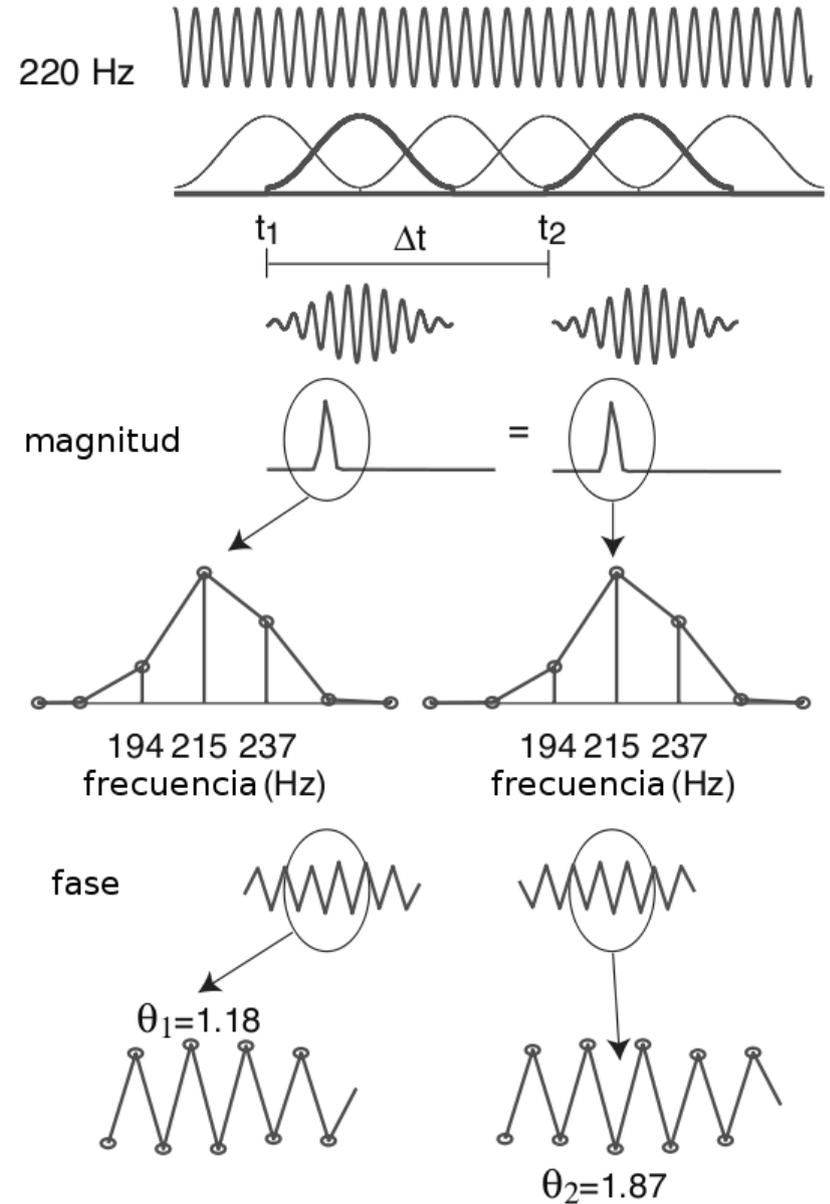
estimación usando los valores fase:

$f = 220.013 \text{ Hz}$



$$2\pi f (t_2 - t_1) = (\theta_2 - \theta_1)$$

$$f = \frac{(\theta_2 - \theta_1) + 2\pi n}{2\pi (t_2 - t_1)}$$



Phase vocoder

Resíntesis:

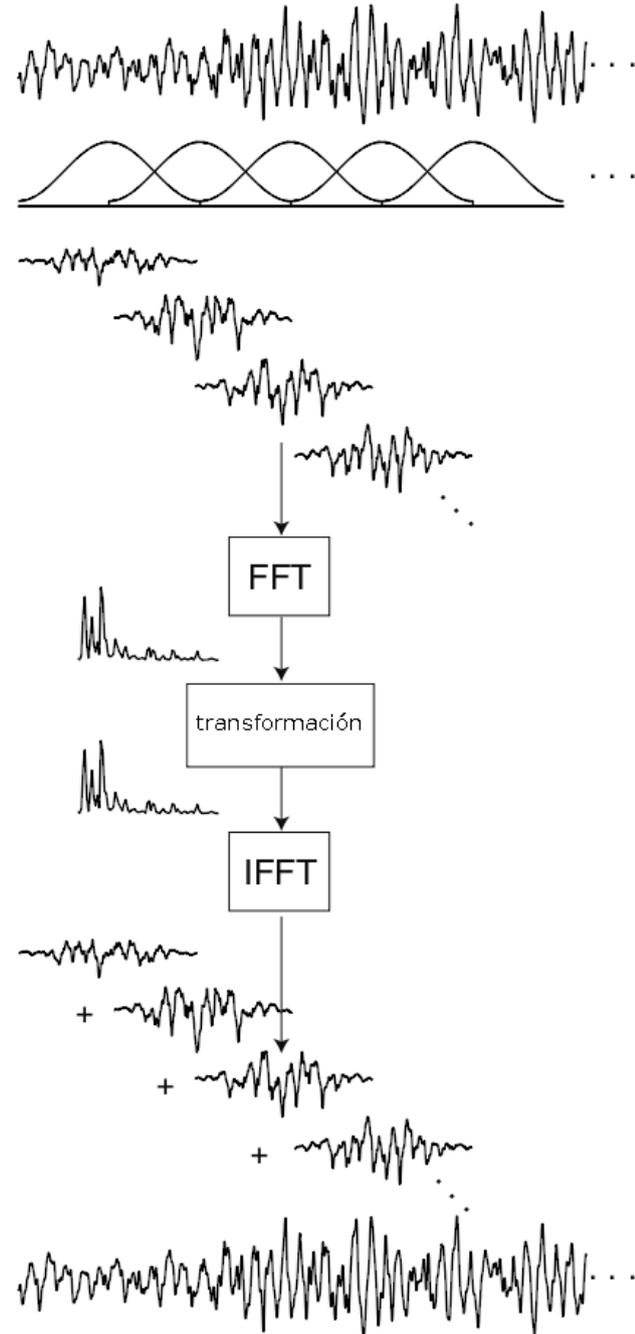
- síntesis aditiva de parciales (ó)
- IFFT y overlap-add

El análisis y síntesis usando la FFT son más rápidos

Se debe construir un espectro coherente, es decir con la magnitud y fase apropiadas derivadas del análisis.

$$\theta_2 = \theta_1 + 2\pi f (t_2 - t_1)$$

El enventanado solapado debe ser de magnitud unitaria (e.g. hop = N/4 con Hann).



Phase vocoder

Aplicaciones:

El objetivo del phase vocoder es separar la información espectral de la información temporal

La estrategia es dividir la señal en bandas de frecuencia y caracterizar las variaciones temporales de la señal en cada banda.

Esto falla cuando en cada banda no hay exclusivamente un único componente, o cuando sus variaciones son demasiado rápidas (en relación a la ventana de la FFT o a la respuesta del filtro pasabanda).

Cuando se cumple estas condiciones el phase vocoder da lugar a una gran cantidad de aplicaciones.

Estudios sobre percepción de timbre, Moorer y Grey 1970s

Transformaciones con fines musicales, dentro de las cuales el *escalamiento temporal* y la *trasposición de altura* son las más clásicas.

Phase vocoder

Escalamiento temporal (time stretching)

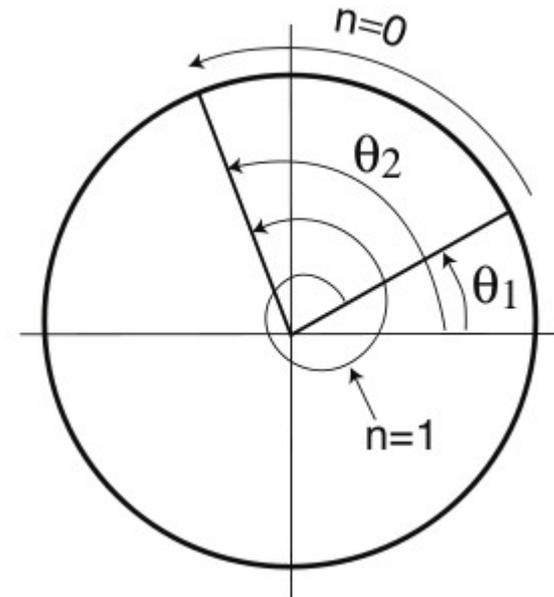
Dado que la información temporal y espectral están separadas es posible dilatar la evolución temporal manteniendo la información espectral.

En la implementación basada en filtros y osciladores, sólo se expande (o contrae) la duración de las señales de control (amplitud y frecuencia).

Usando la STFT es posible separar (o acercar) las tramas en la síntesis.

En este caso, un detalle crítico es que la fase debe cambiar en concordancia con el escalamiento temporal, para mantener la variación de frecuencia.

$$2\pi f (t_2 - t_1) = (\theta_2 - \theta_1) + 2\pi n$$



Phase vocoder

Trasposición de altura (pitch shifting)

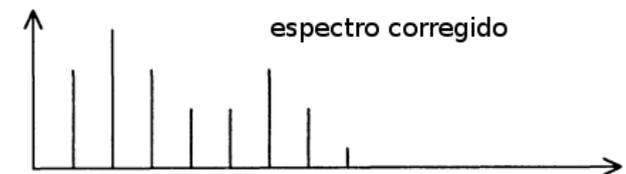
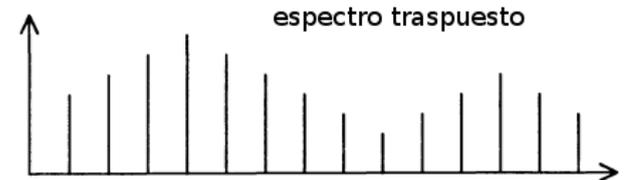
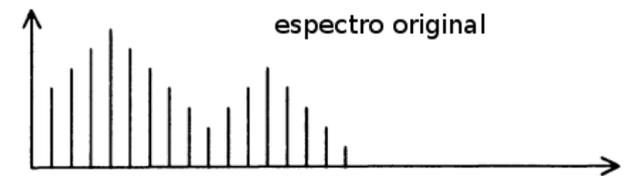
Trasladar el contenido espectral sin alterar la evolución temporal. Se escala en el tiempo con el factor de trasposición deseado y luego se reproduce (o muestrea) a una frecuencia correspondiente a dicho factor.

Por ejemplo, para traspasar una octava se escala al doble el tiempo y se reproduce a una frecuencia de muestreo del doble (misma duración con corrimiento del espectro). Lo último es equivalente a remuestrear.

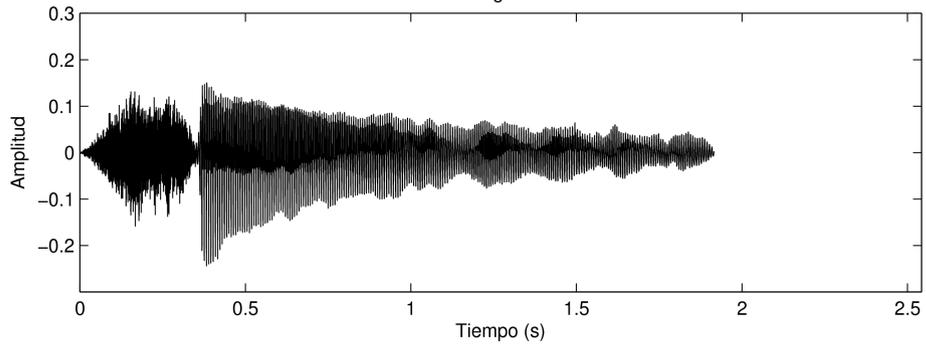
Un detalle es que el factor de escalamiento debe ser una relación entre enteros.

Otro problema es que no sólo se modifica la altura sino también la envolvente espectral.

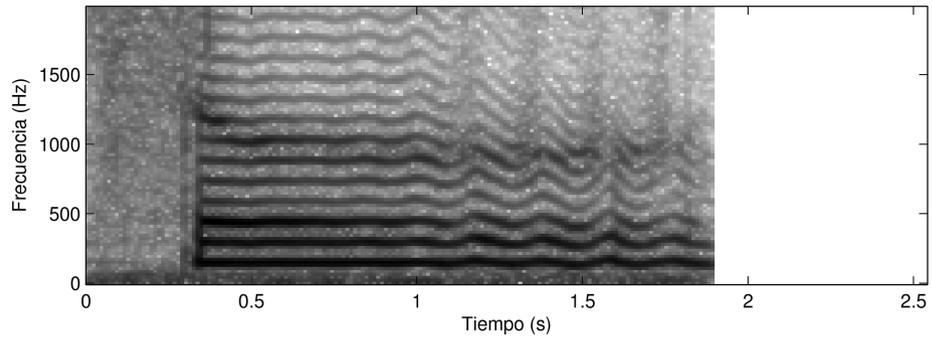
Puede corregirse calculando la envolvente original y aplicándola al espectro resultante.



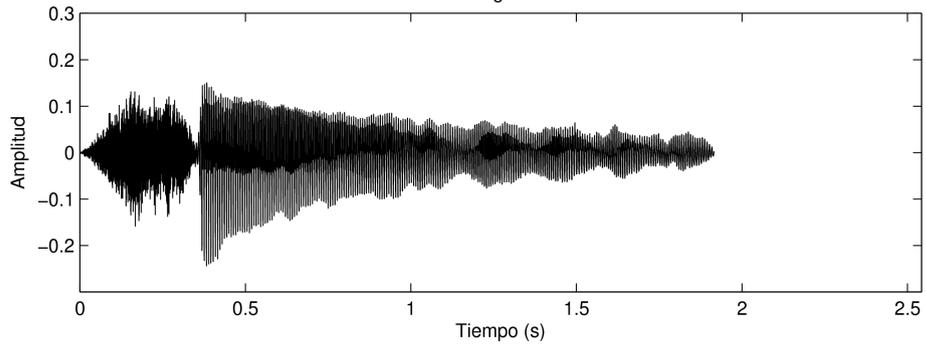
Original



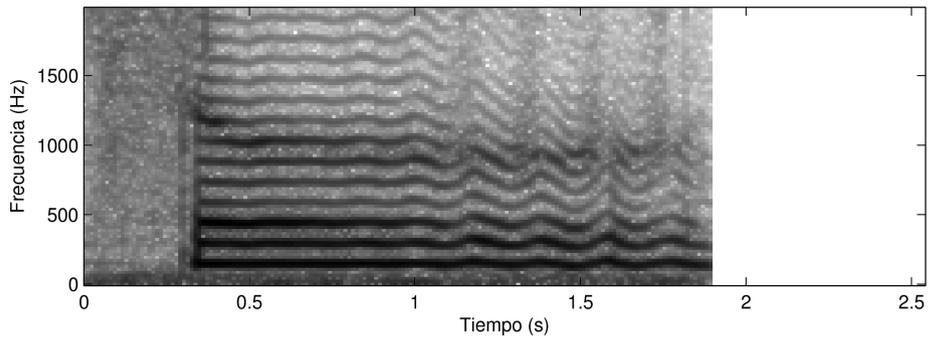
Espectrograma



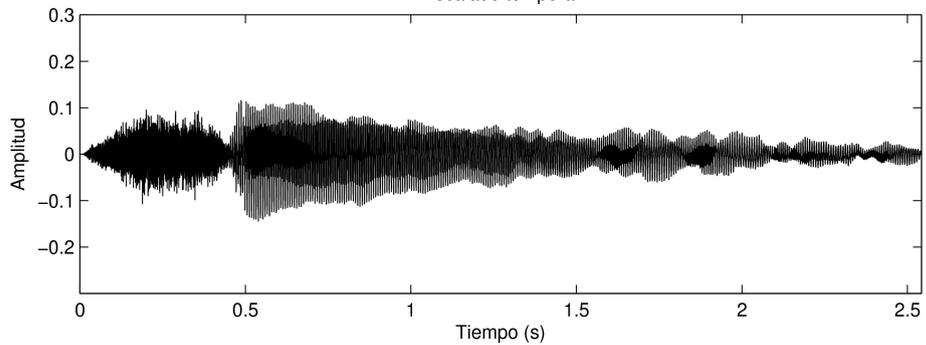
Original



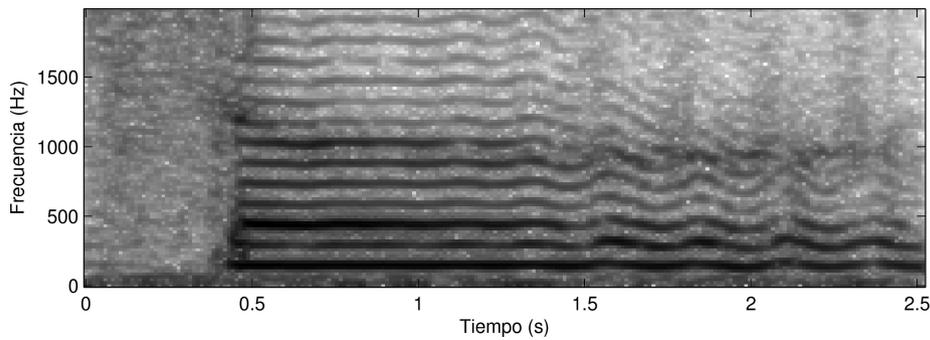
Espectrograma

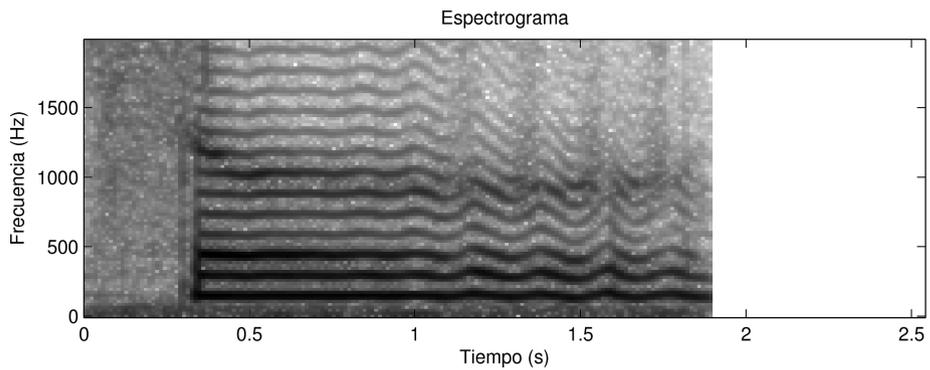
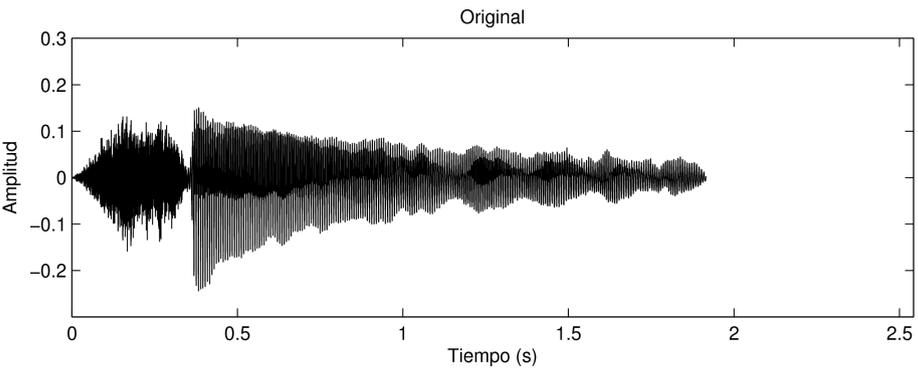


Escalado temporal

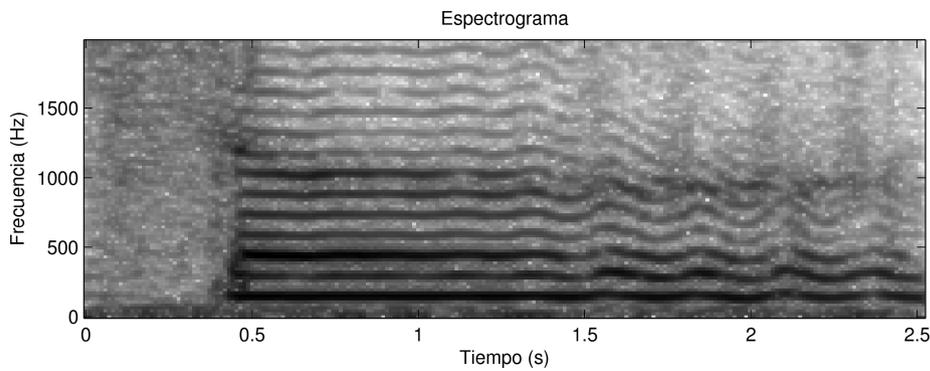
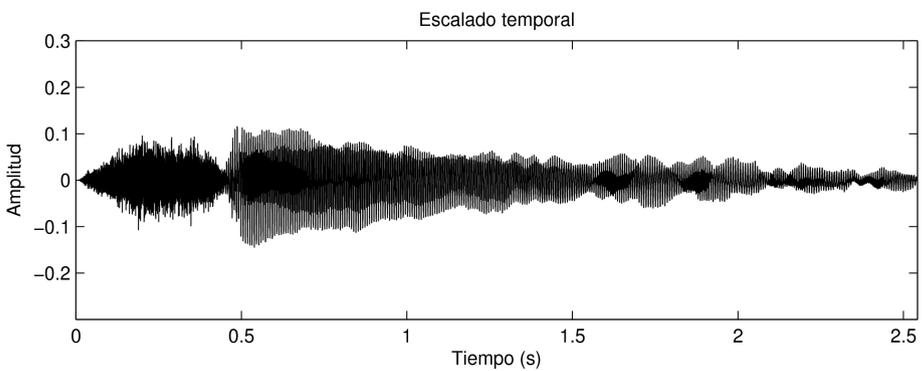
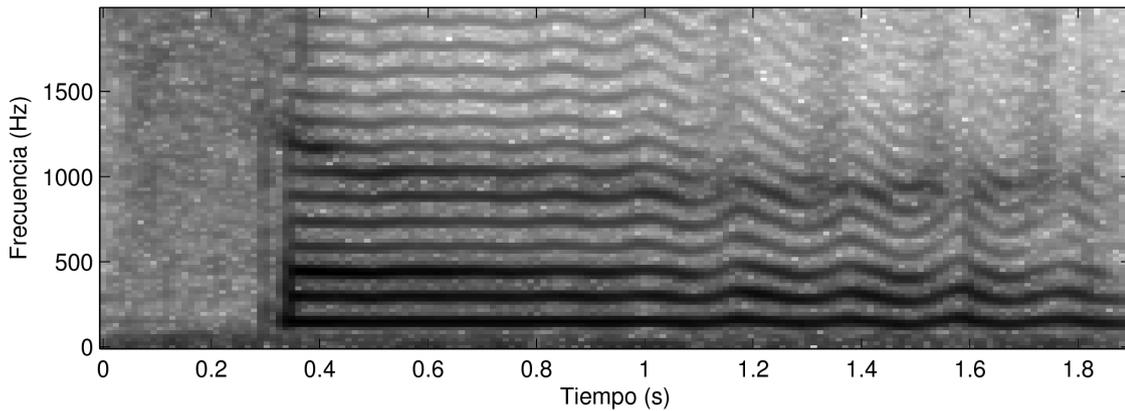


Espectrograma

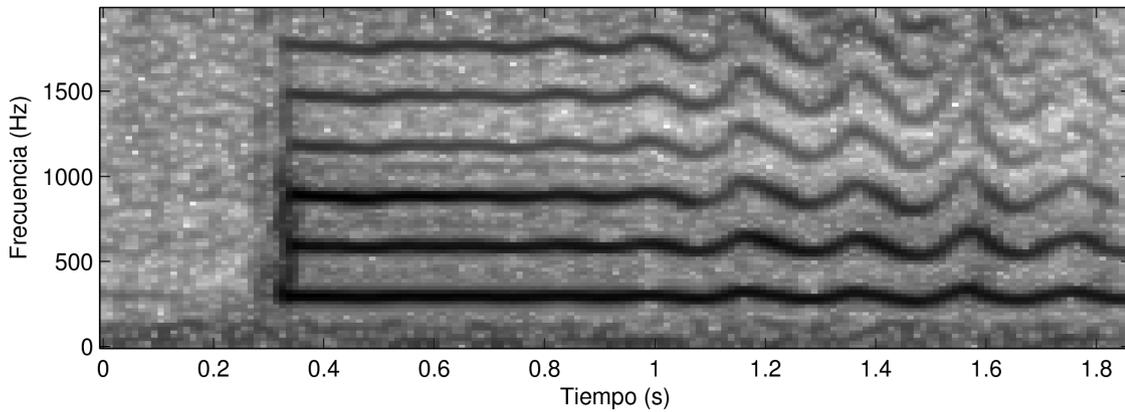




Espectrograma original



Espectrograma de trasposicion



Modelado sinusoidal

Trabajos pioneros

Julius Smith, PARSHL, 1987,

An Analysis/Synthesis Program for Non-Harmonic Sounds based on a Sinusoidal Representation, ICMC

Busca extender el phase-vocoder para manejar sonidos inarmónicos y con variaciones más rápidas. Aplica seguimiento de picos en la STFT. Se remplazan las diferencias de fase por la interpolación de picos espectrales.

Quatieri y McAulay, 1986,

Speech Analysis/Synthesis based on a Sinusoidal Representation, IEEE ASSP

Técnica similar basada en la STFT y modelado sinusoidal para el análisis de voz hablada.

Serra y Smith, 1990

Spectral Modeling Synthesis: A Sound Analysis/Synthesis System based on a Deterministic plus Stochastic Decomposition, Computer Music Journal

Se extiende el modelo incorporando la representación de componentes de ruido no asimilables a suma de sinusoides.

Modelado sinusoidal

Modelo: componente determinístico y estocástico

El sonidos producido por un instrumento musical puede modelarse como la suma de un conjunto de **sinusoides** más un **ruido residual**.

$$s(t) = \sum_{r=1}^R A_r(t) \cos(\theta_r(t)) + e(t)$$

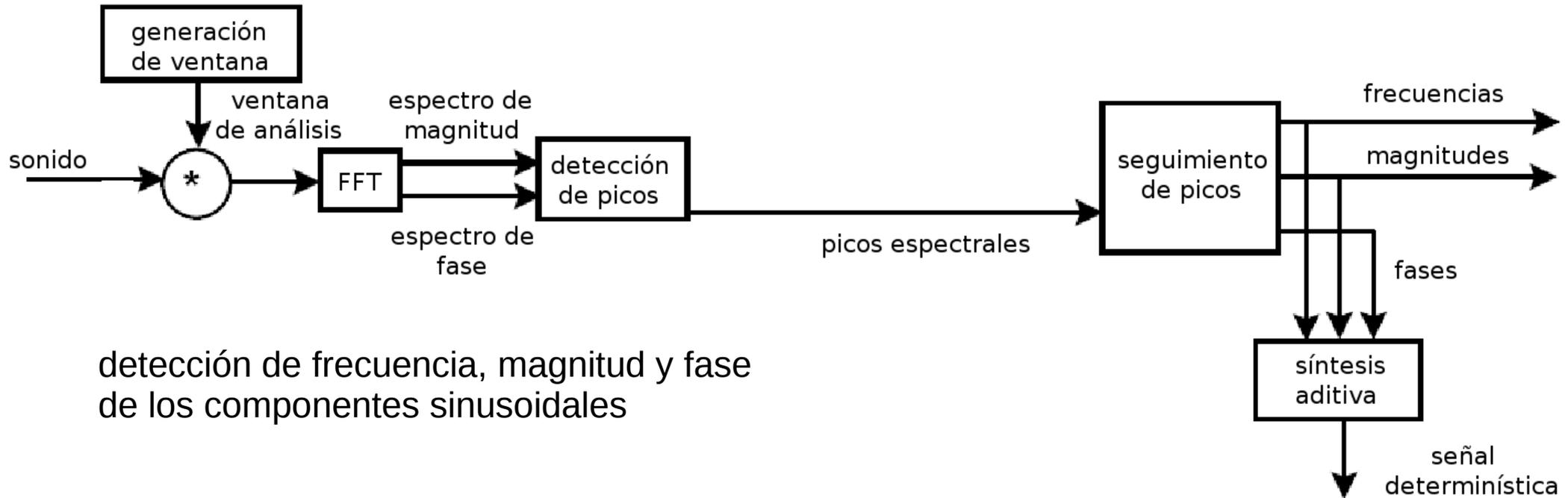
$A_r(t)$ y $\theta_r(t)$ amplitud y fase instantáneas de la senoide r-ésima
 $e(t)$ componente de ruido en el tiempo t

Las sinusoides (componente determinístico) corresponde a los modos principales de vibración del sistema. El residuo (componente estocástico) comprende la energía producida por el mecanismo de excitación que no se transforma en vibración estacionaria (transitorios) y otra componente de energía que es no sinusoidal por naturaleza.

Se asume:

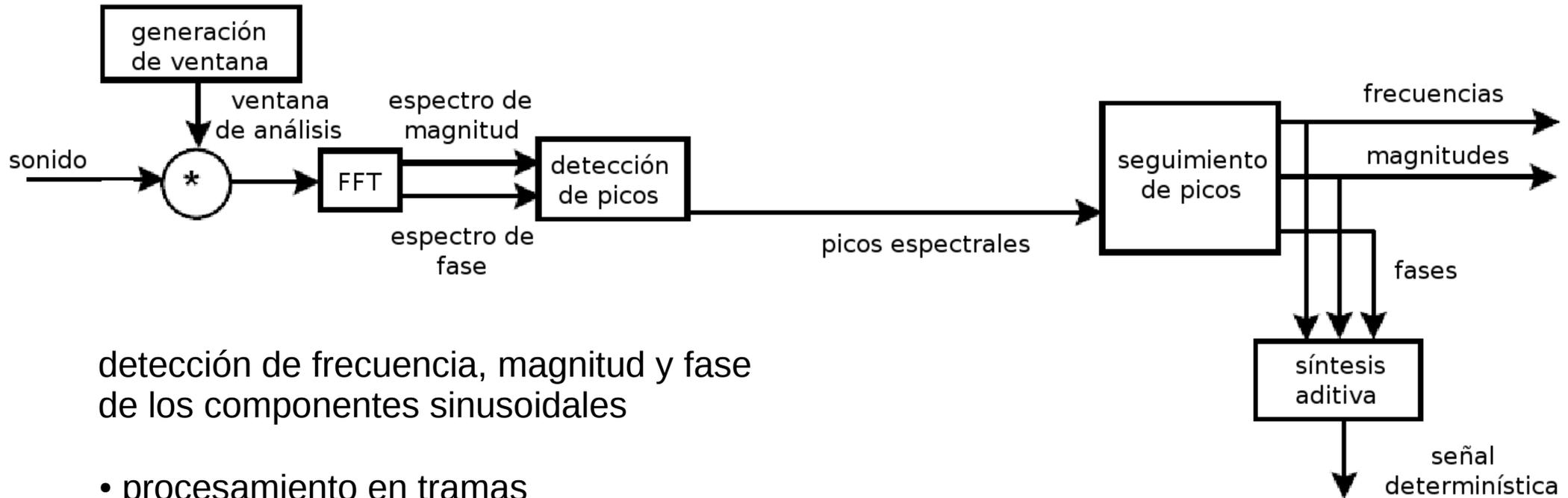
- sinusoides estables, variación lenta de amplitud y frecuencia
- componente estocástico puede describirse como ruido blanco filtrado

Modelado sinusoidal



detección de frecuencia, magnitud y fase de los componentes sinusoidales

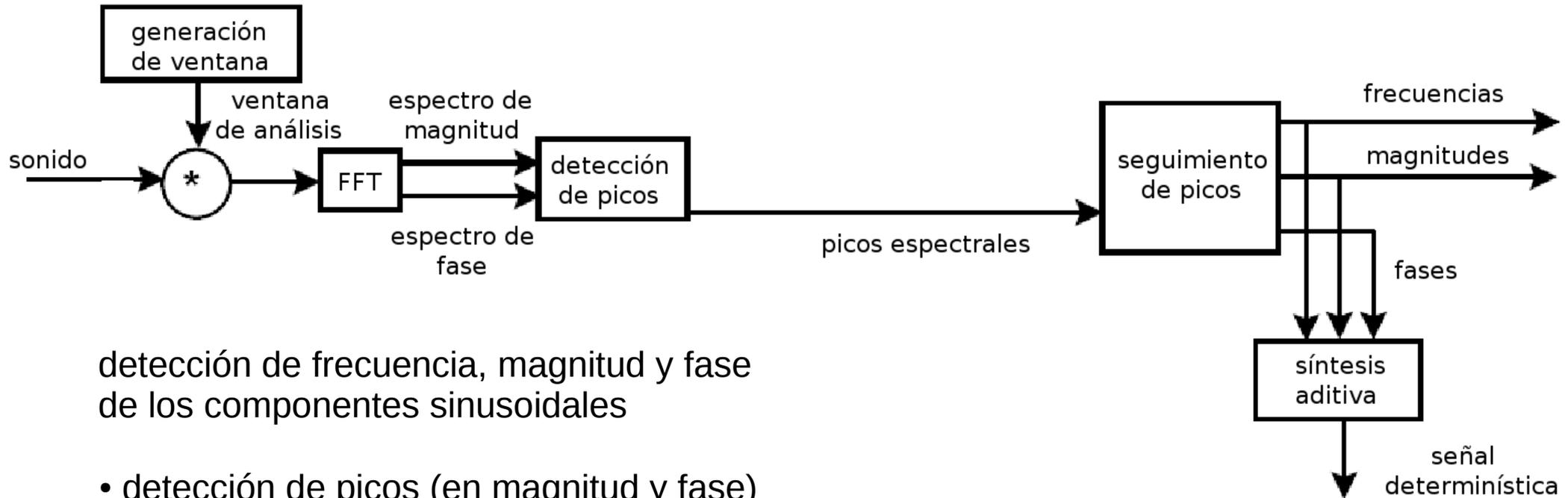
Modelado sinusoidal



detección de frecuencia, magnitud y fase de los componentes sinusoidales

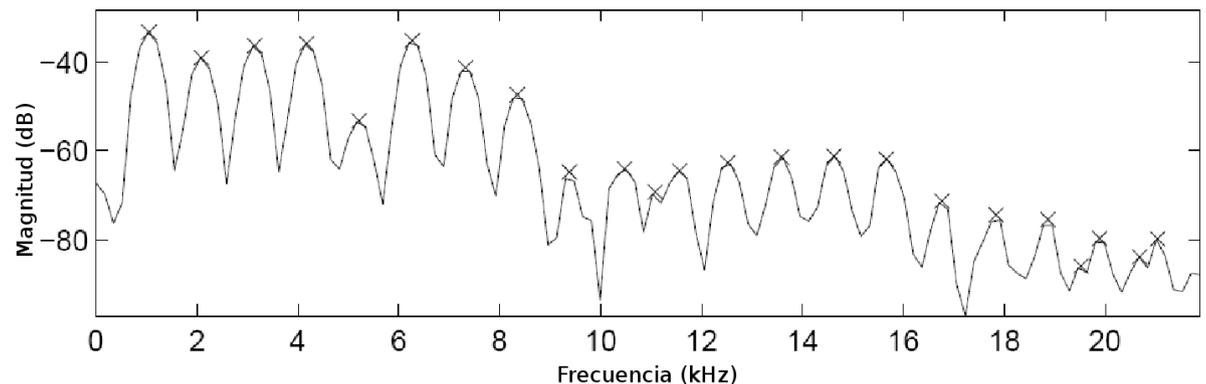
- procesamiento en tramas
- tamaño depende de distancia entre parciales
- ventana centrada para fase cero
- no necesario que sumen uno

Modelado sinusoidal

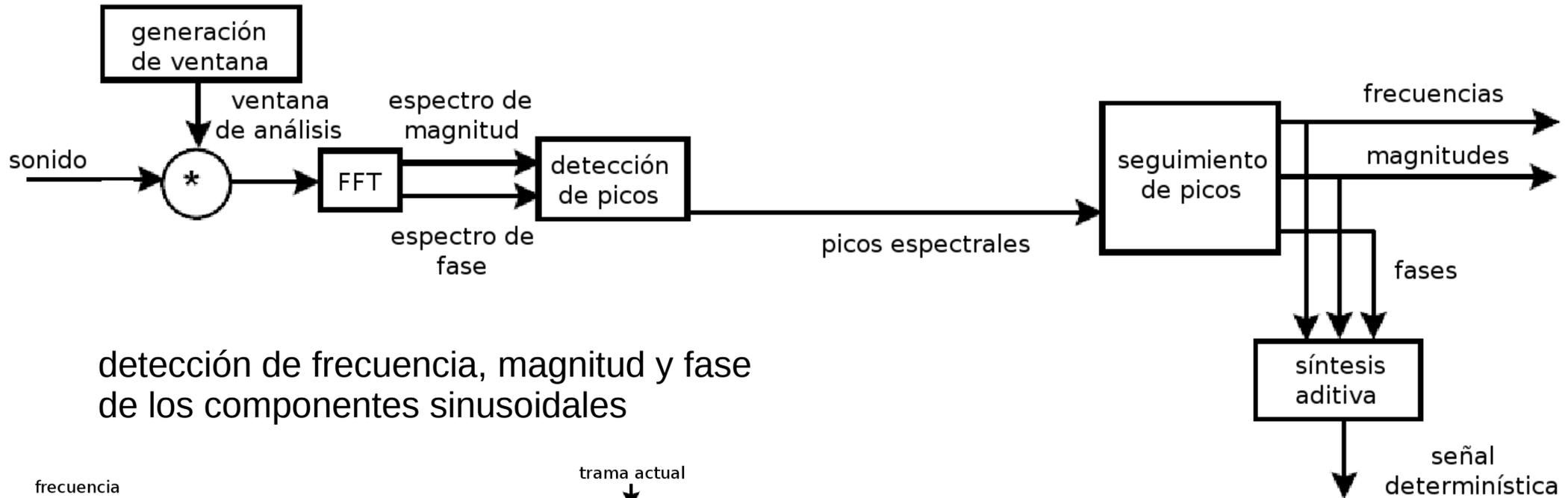


detección de frecuencia, magnitud y fase de los componentes sinusoidales

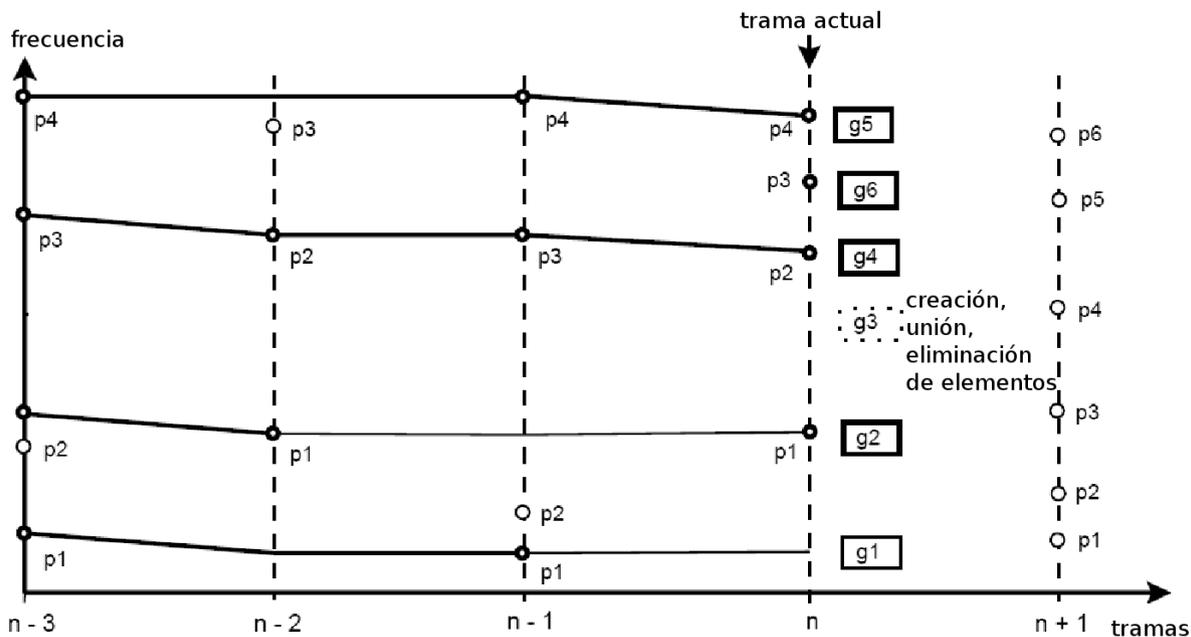
- detección de picos (en magnitud y fase)
- interpolación de frecuencia del pico usando magnitud de vecinos



Modelado sinusoidal

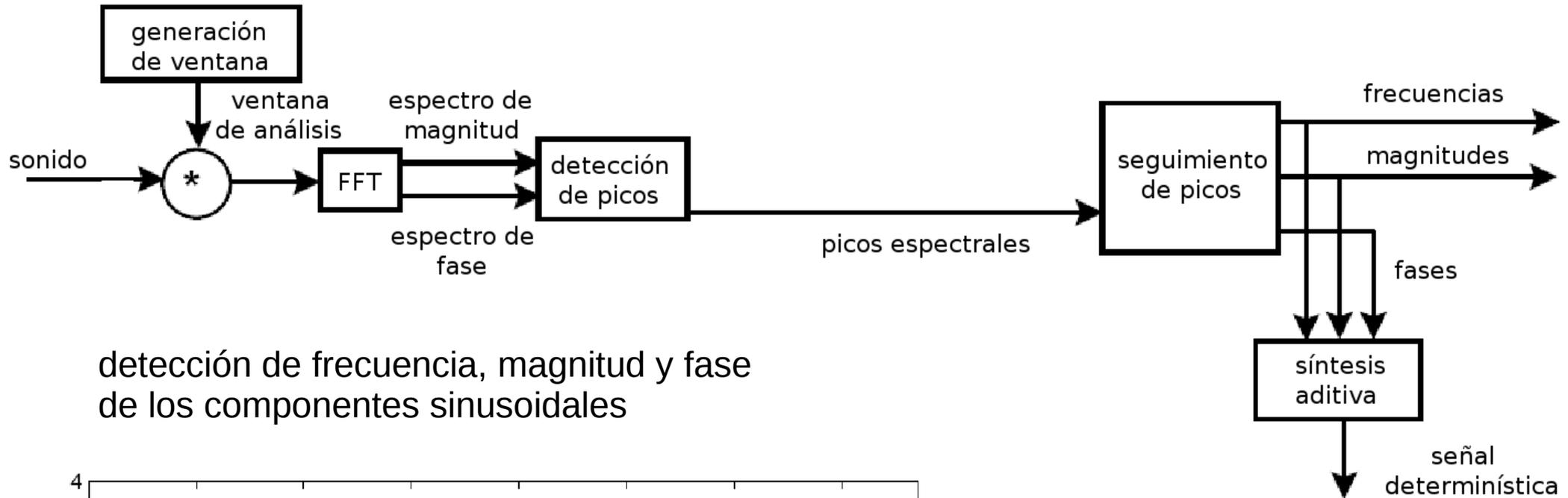


detección de frecuencia, magnitud y fase de los componentes sinusoidales

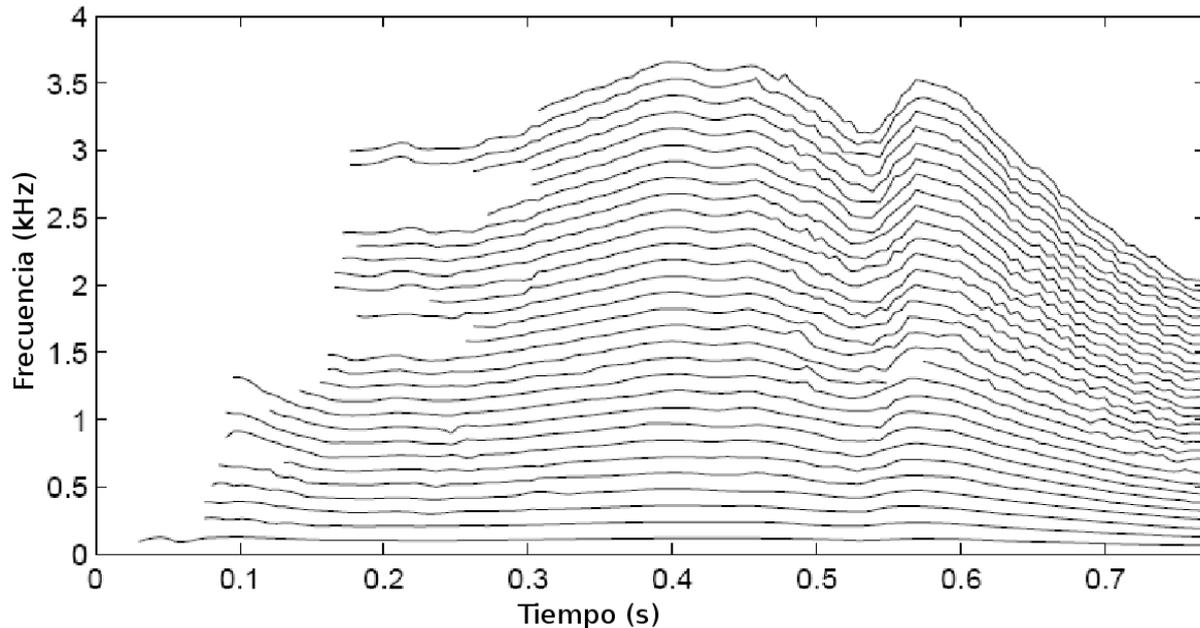


- seguimiento de picos usando heurística

Modelado sinusoidal

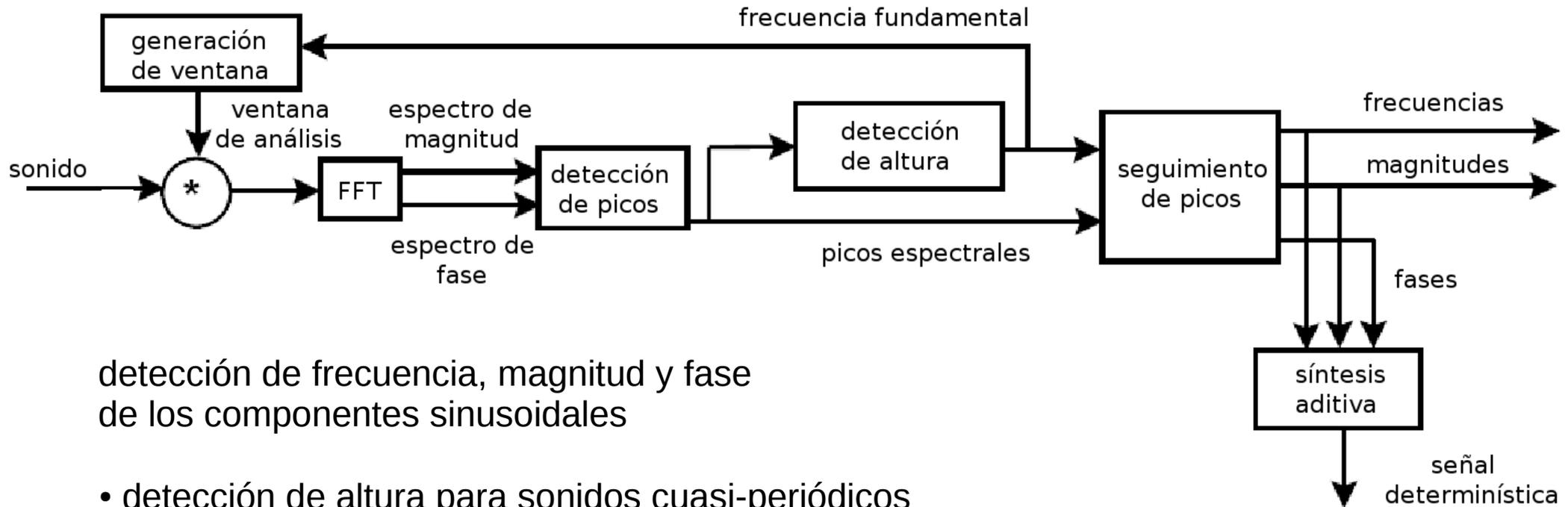


detección de frecuencia, magnitud y fase de los componentes sinusoidales



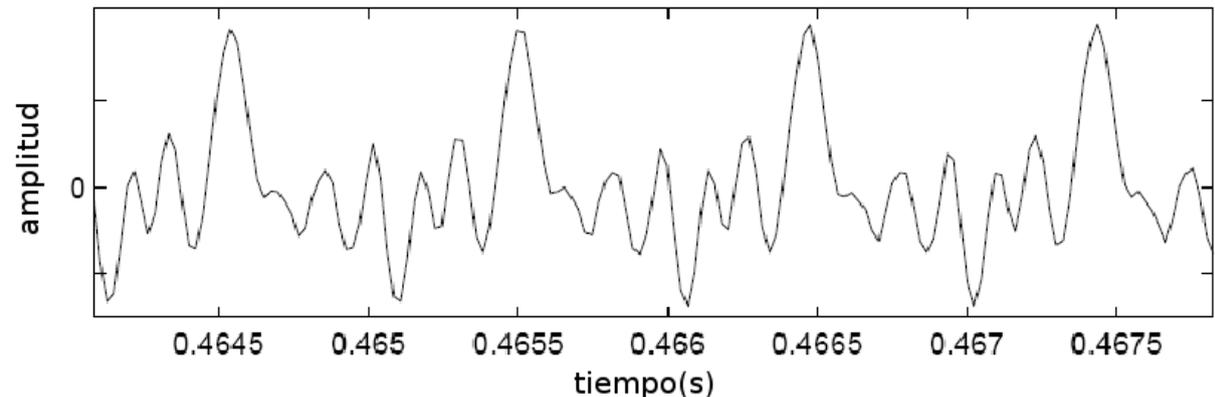
- seguimiento de picos usando heurística
- puede hacerse de atrás a adelante en el tiempo

Modelado sinusoidal

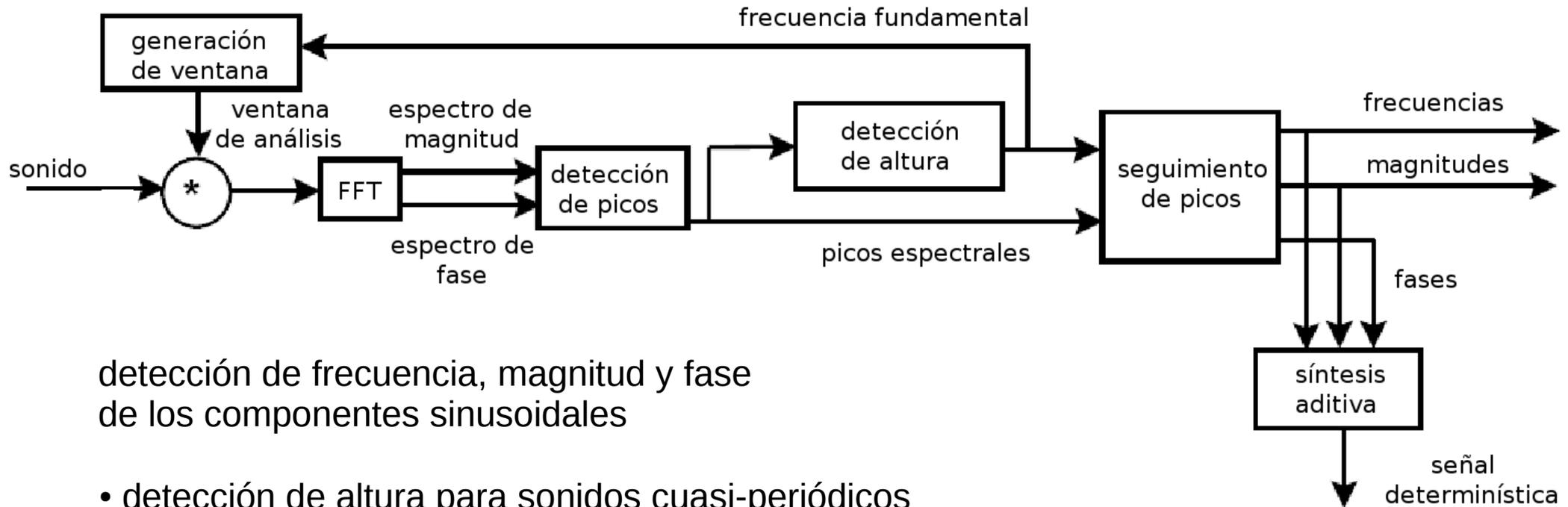


detección de frecuencia, magnitud y fase de los componentes sinusoidales

- detección de altura para sonidos cuasi-periódicos
- permite análisis sincrónico con altura (número justo de períodos de la señal en la ventana de análisis)
- facilita el seguimiento de picos

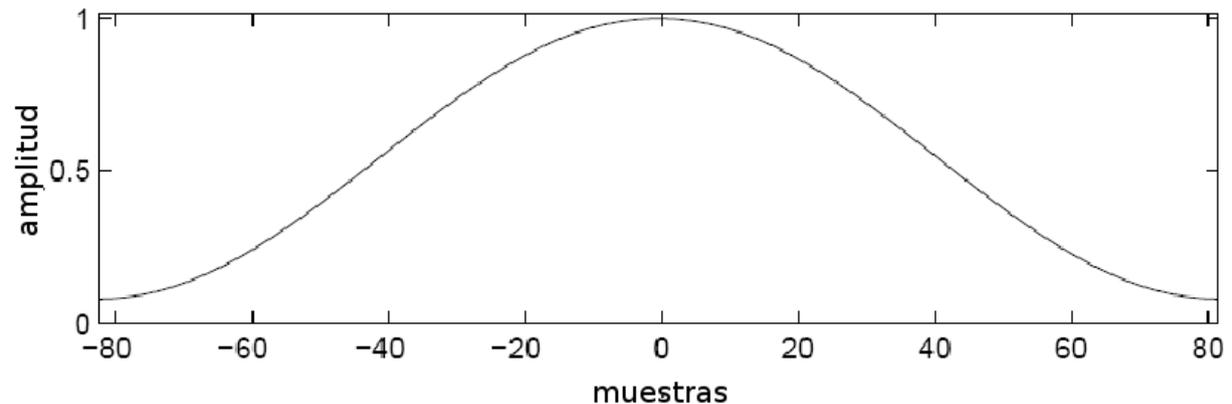


Modelado sinusoidal

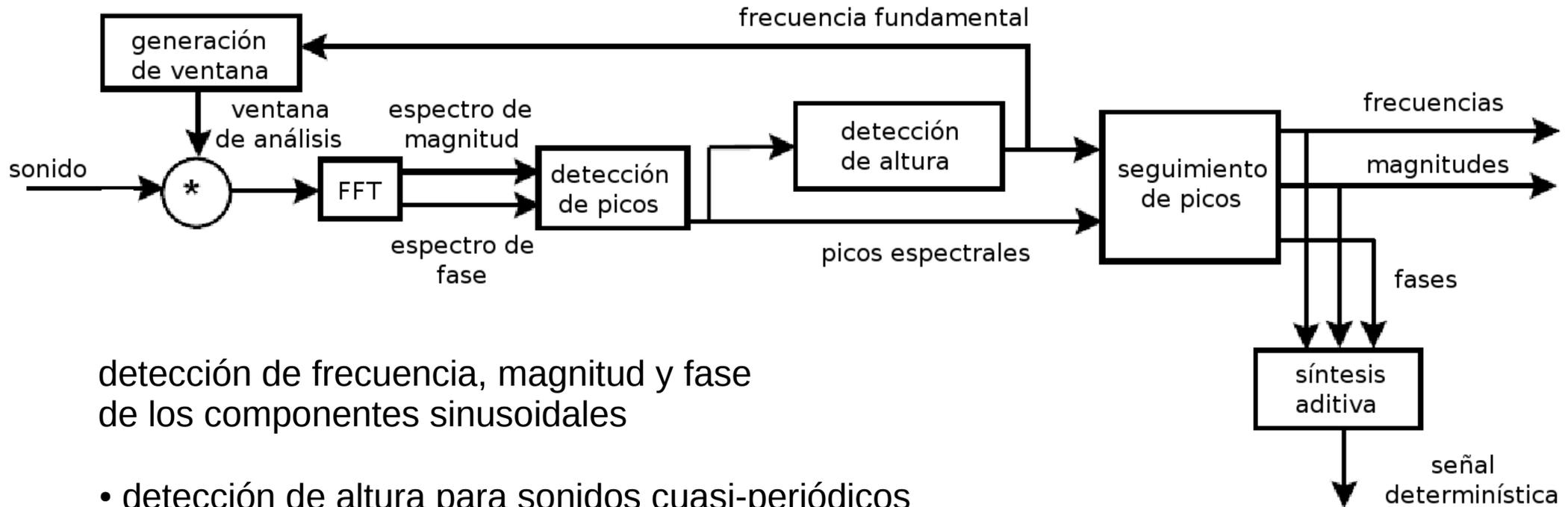


detección de frecuencia, magnitud y fase de los componentes sinusoidales

- detección de altura para sonidos cuasi-periódicos
- permite análisis sincrónico con altura (número justo de períodos de la señal en la ventana de análisis)
- facilita el seguimiento de picos

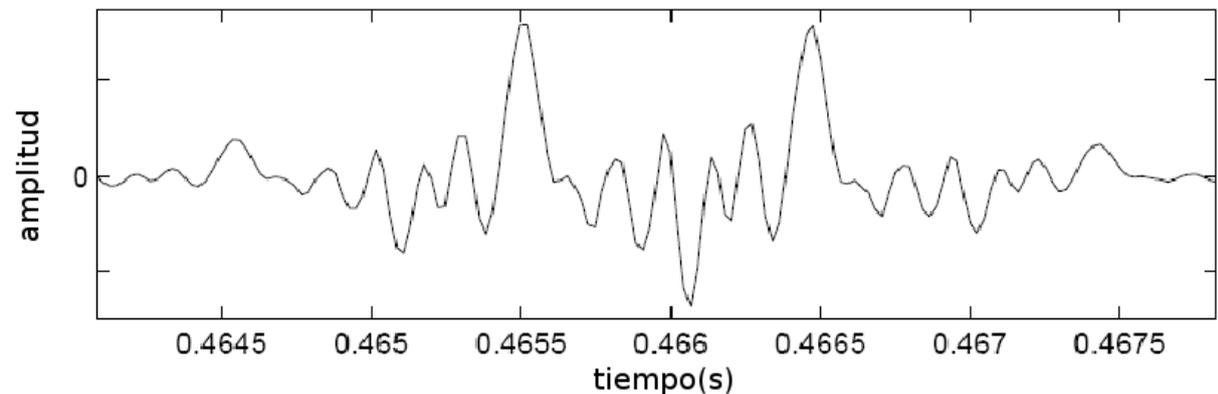


Modelado sinusoidal

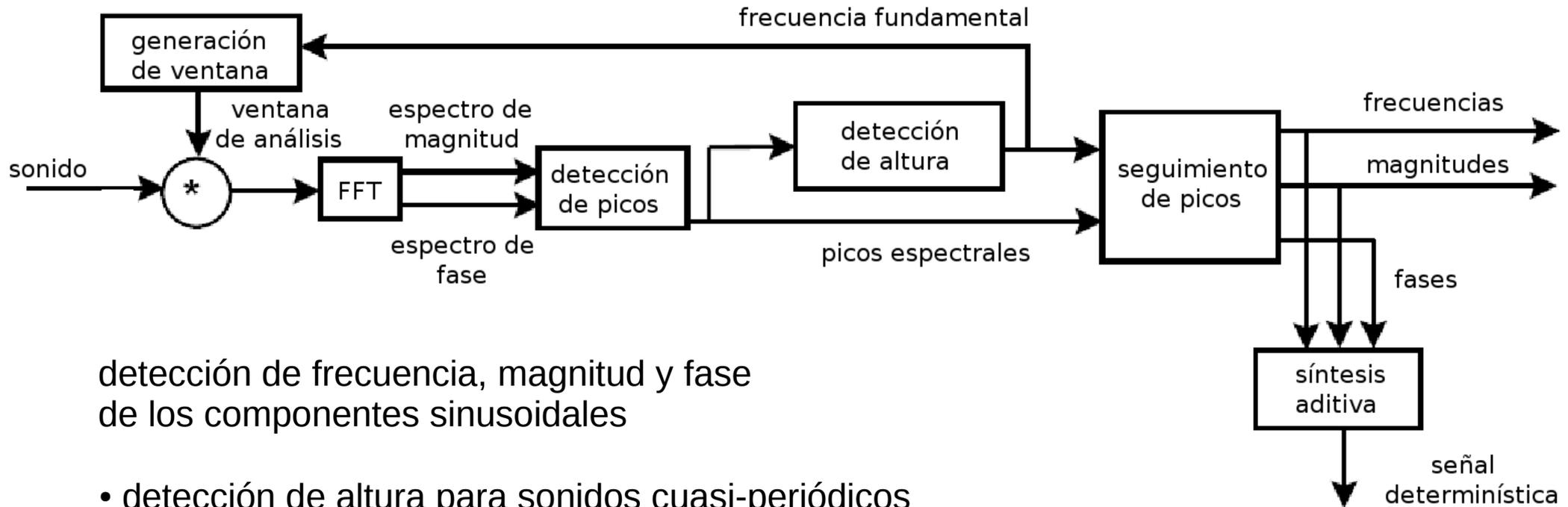


detección de frecuencia, magnitud y fase de los componentes sinusoidales

- detección de altura para sonidos cuasi-periódicos
- permite análisis sincrónico con altura (número justo de períodos de la señal en la ventana de análisis)
- facilita el seguimiento de picos

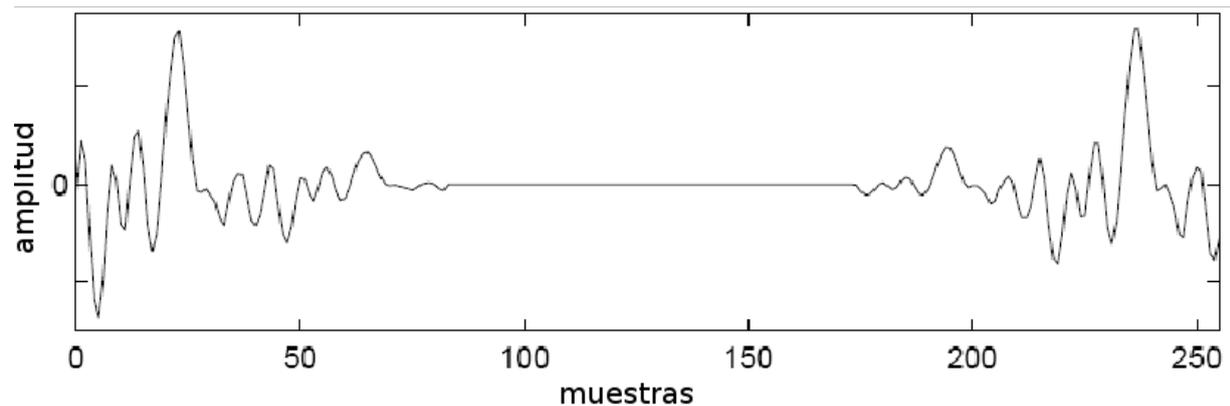


Modelado sinusoidal

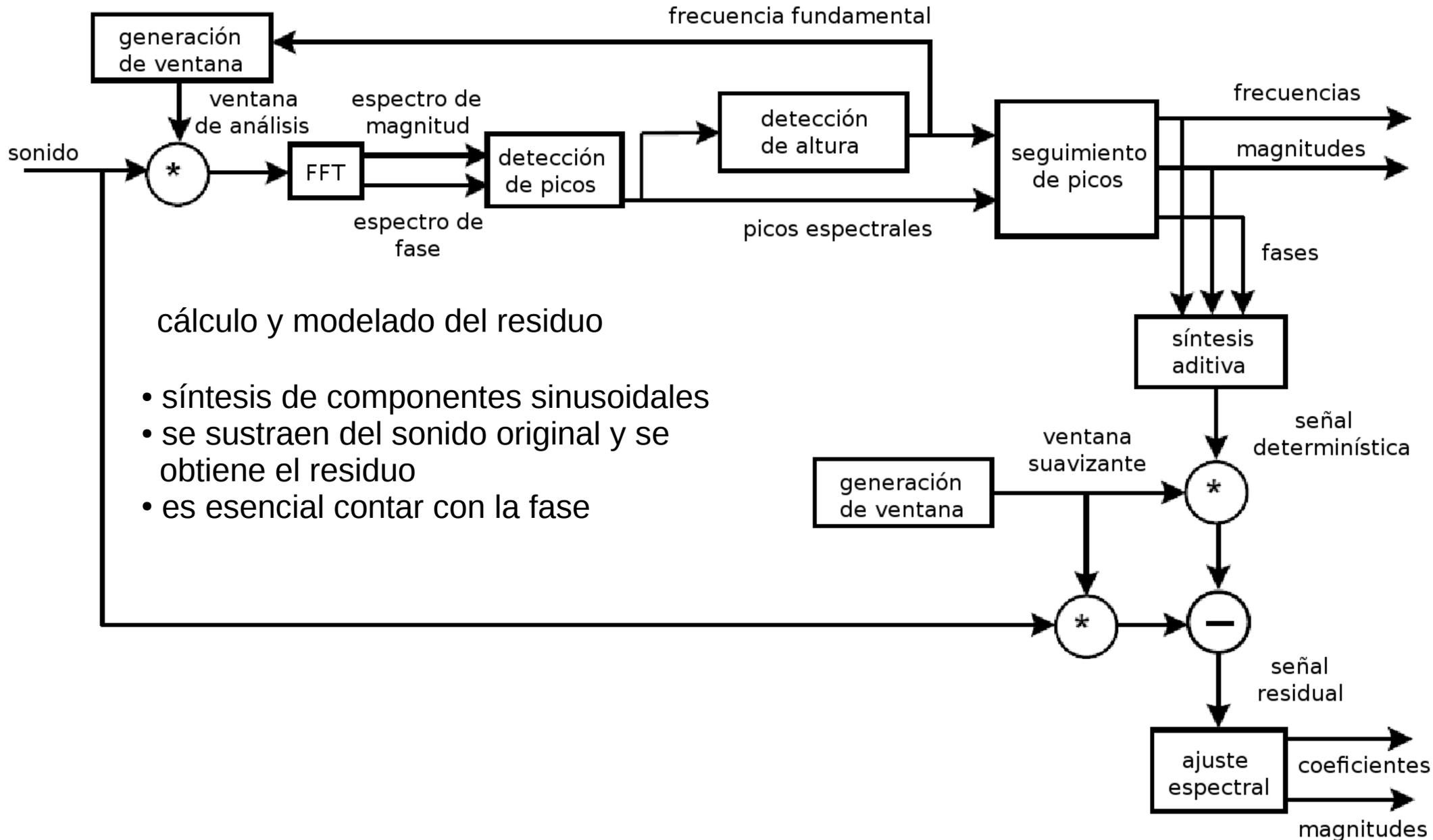


detección de frecuencia, magnitud y fase de los componentes sinusoidales

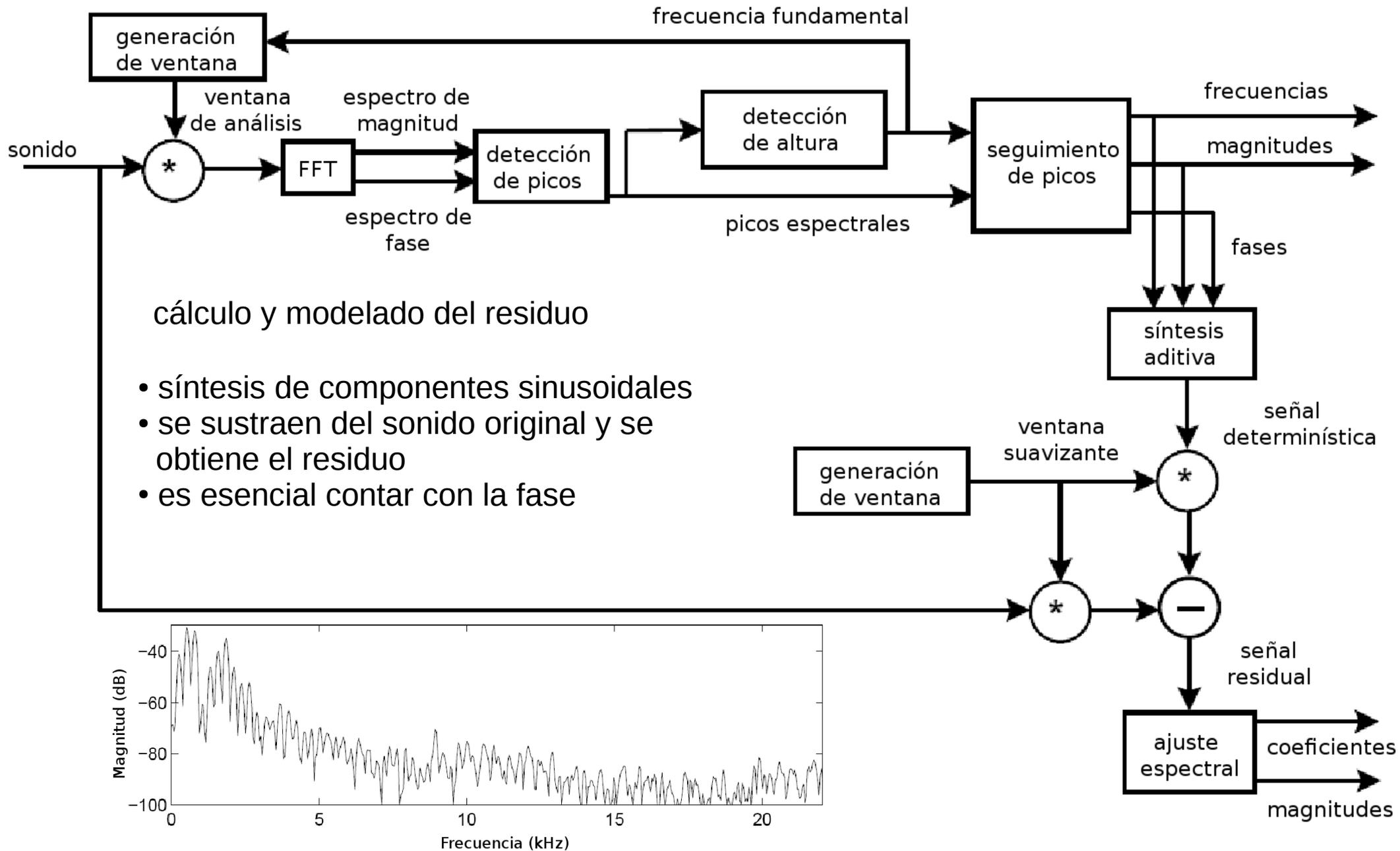
- detección de altura para sonidos cuasi-periódicos
- permite análisis sincrónico con altura (número justo de períodos de la señal en la ventana de análisis)
- facilita el seguimiento de picos



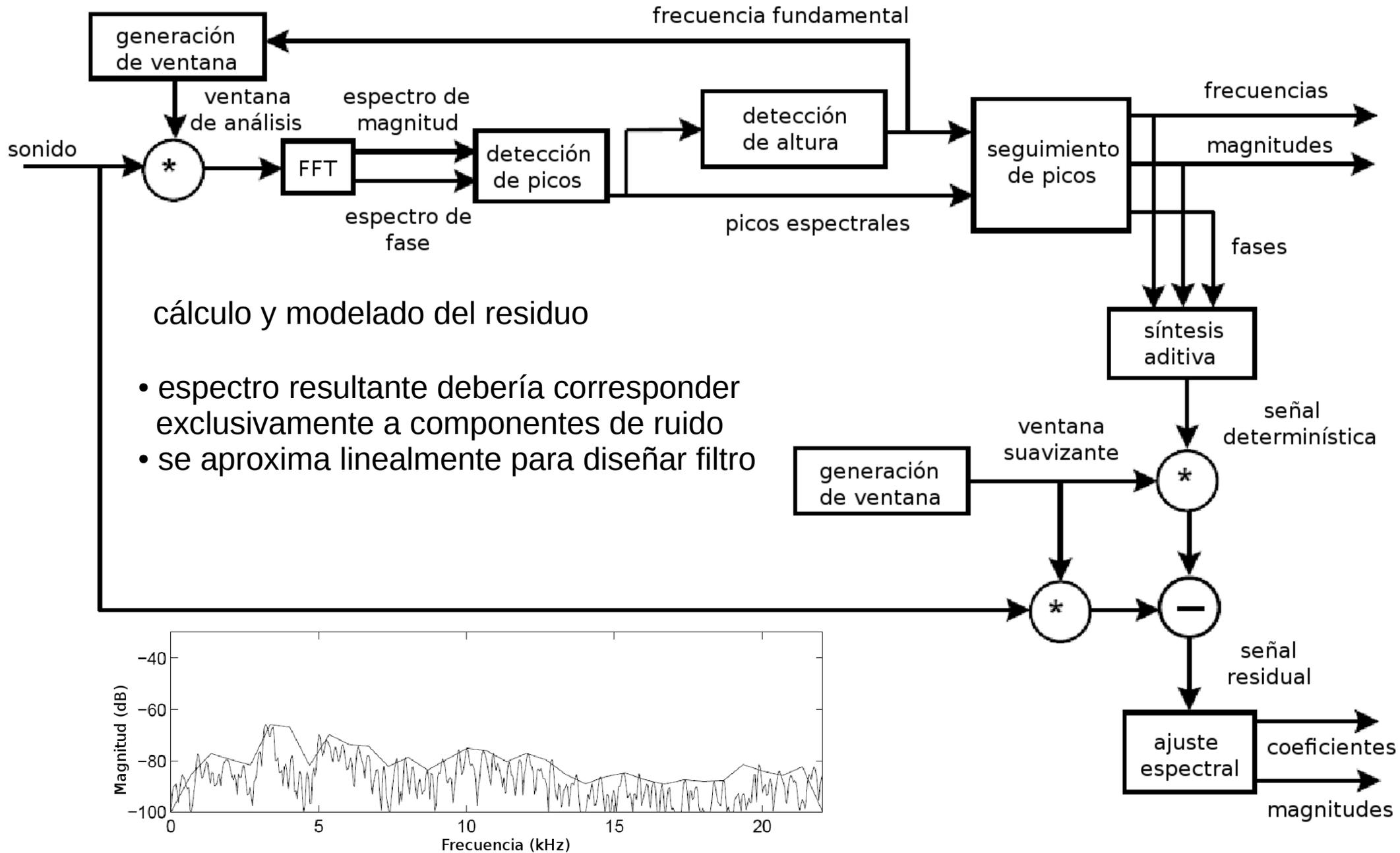
Modelado sinusoidal

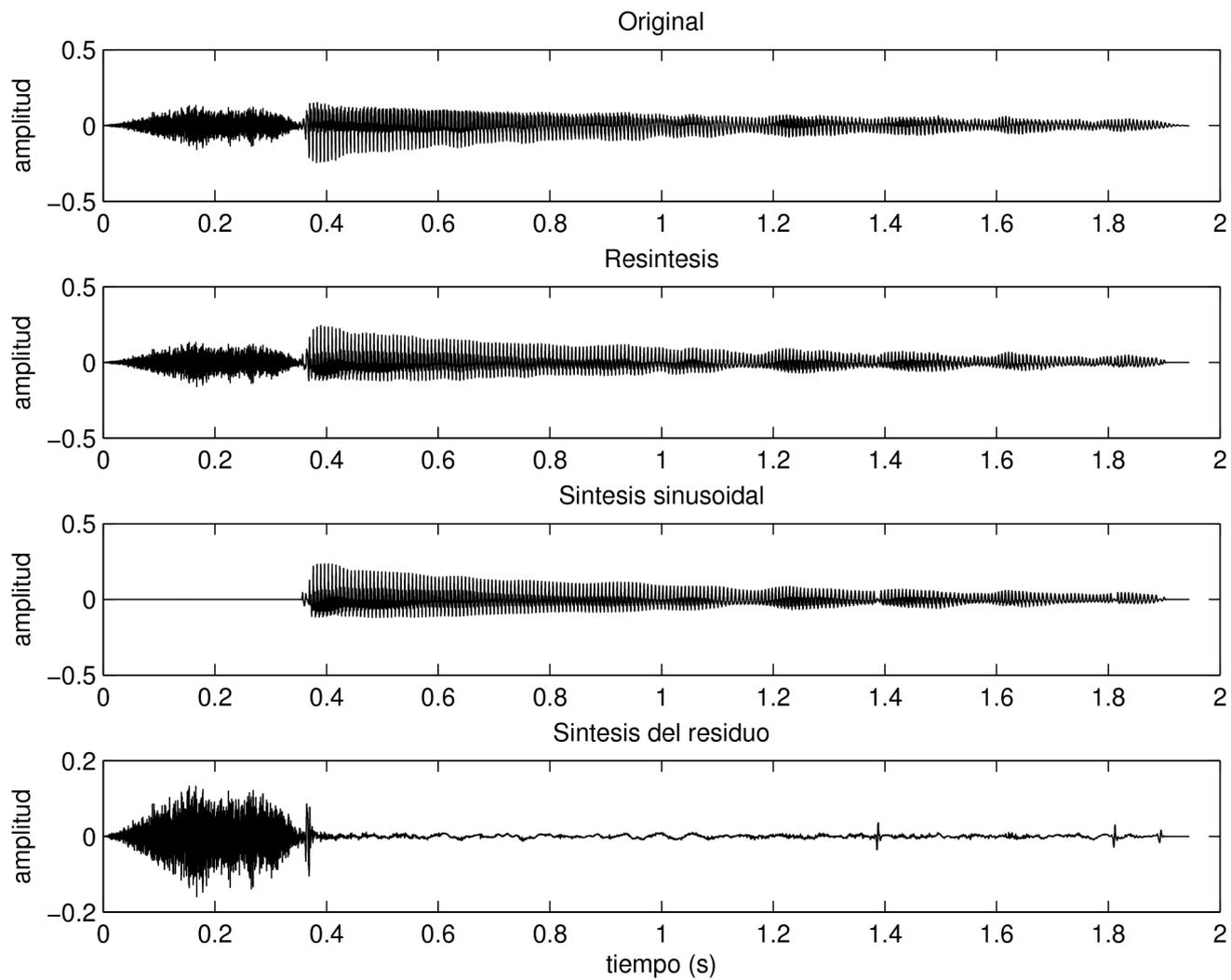


Modelado sinusoidal

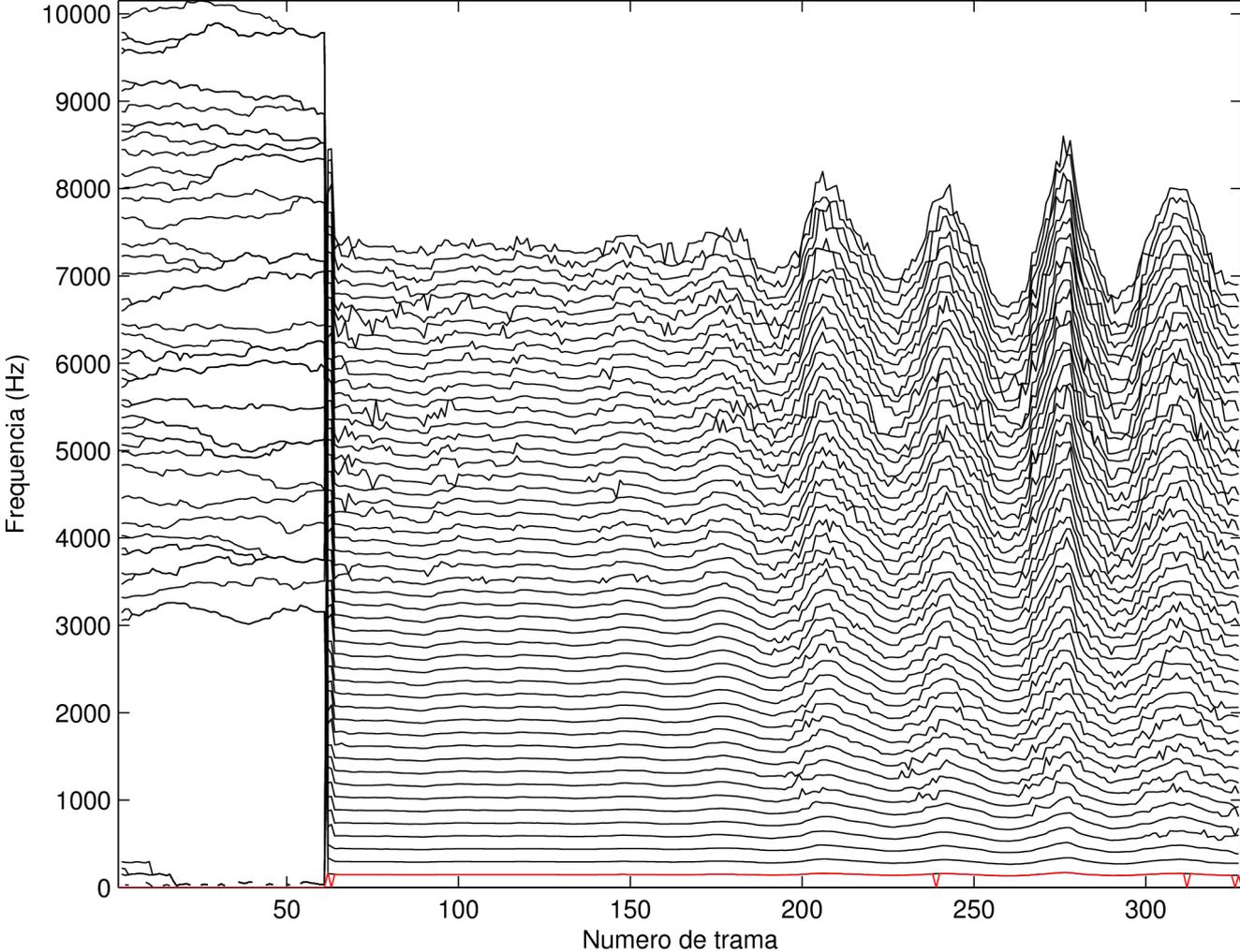


Modelado sinusoidal





Seguimiento de picos



Software

Phase vocoder

- Csound (pvanal, pvoc)

<http://www.csounds.com/>

Modelado sinusoidal

- SMS

<http://mtg.upf.edu/technologies/sms>

- ATSH

<http://www.dxarts.washington.edu/ats/>

- Spear

<http://www.klingbeil.com/spear/>