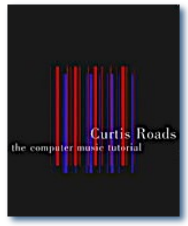
Tema 1: Representación del sonido

Contenidos

- **Naturaleza y representación del sonido**
  - **Representación en el tiempo: la señal**
  - **Representación en la frecuencia: el espectro**
- **Representación digital**
- **Evolución temporal del espectro y FFT**
- **Caracterización del espectro:**
  - **Frecuencia fundamental y armonicidad**
  - **Envolvente espectral y formantes**
- **El ruido**

Computer Music
C. Dodge & T.A. Pierce



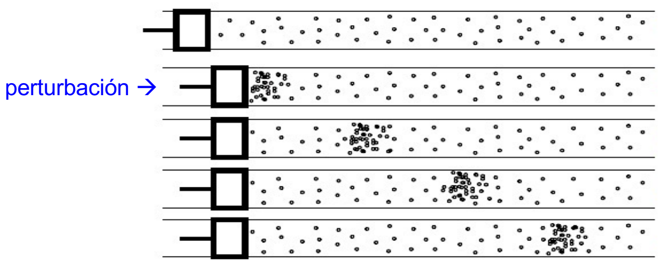
The Computer Music Tutorial
Curtis Roads

John Watkinson

El sonido en el dominio temporal

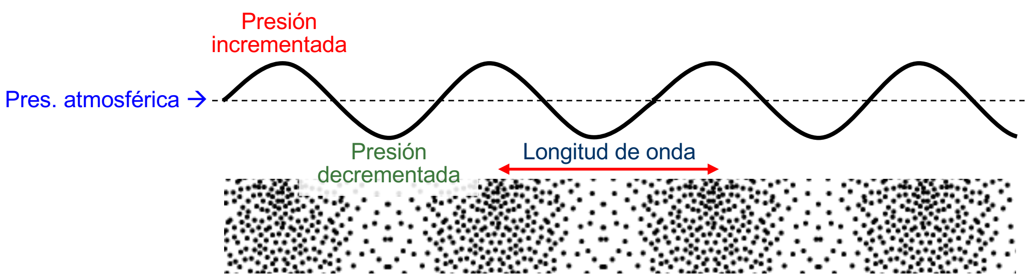
Naturaleza del sonido

- Las perturbaciones del aire que nos rodea las percibimos como **sonido** cuando llegan a nuestros oídos.
- Las perturbaciones se propagan desde los objetos que las producen hasta nuestros oídos como ondas de presión.



- Se trata de una onda **longitudinal**.
- En contraste con una onda en una cuerda, que sería *transversal*.

- Señales sonoras: Son la representación de las variaciones de presión.



El sonido en el dominio temporal

Naturaleza del sonido

- ¿A qué **velocidad** se propagan las ondas sonoras en el aire?

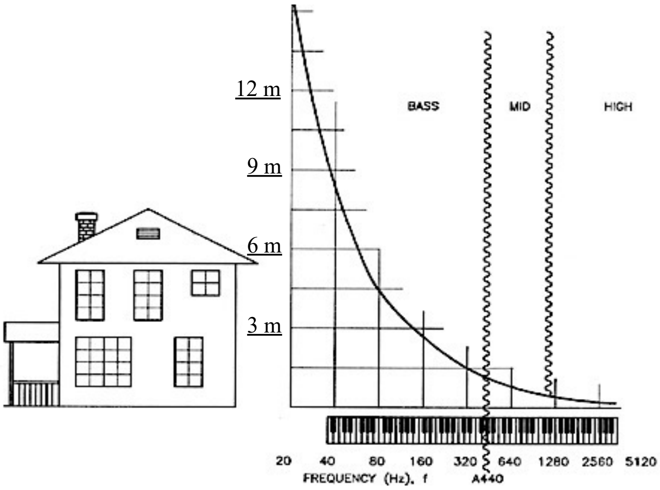
  v = 340 m/s a 15 ºC al nivel del mar (P_{atm} = 1013 mbar)

  El sonido cruza una sala de 10 m en 29 ms (va y vuelve en 59 ms)

- ¿Cuál es el *tamaño* del sonido? (su **longitud de onda**)

$\lambda = v / f$ → λ_{1kHz} = 340 / 1000 = 0,34 metros = 34 cm

1 kHz ~ 1 cabeza



Frecuencia (Hz)	Longitud de onda
20.000	1,7 cm
1.000	34 cm
440 (LA_4)	77 cm
340 (~FA_4)	1 m
20	17 m

El sonido en el dominio temporal

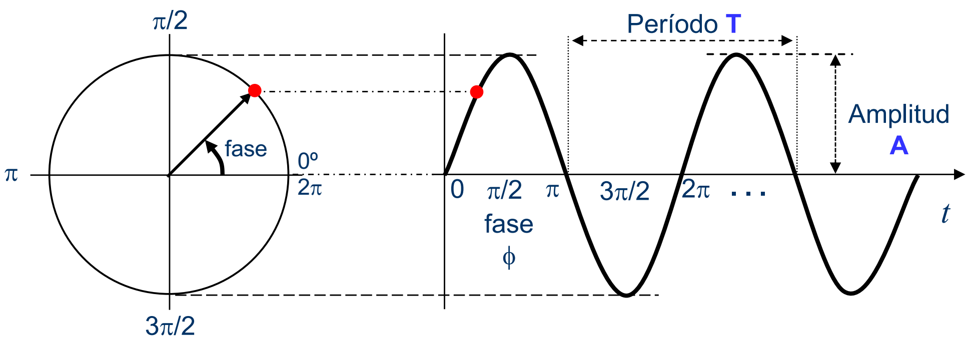
Representación del sonido

- La señal más *simple* posible es la sinusoidal.
- Se representa mediante la función seno.
- Es *simple* porque contiene **una única frecuencia** (f):

$$s(t) = A \ \text{sen} \ \overbrace{(2 \pi f t + \underbrace{\phi_o}_{\text{FASE INICIAL}})}^{\text{FASE}}$$

Frecuencia $f = 1 / T$ (Hz) (s)

1 Hz = 1 ciclo/s



Definiciones:

- El **periodo** T es el tiempo que tarda la onda en pasar 2 veces por la misma fase.
- La **amplitud** A es el valor máximo de la señal en valor absoluto.
- La **longitud de onda** λ es la distancia recorrida por la onda en un periodo.

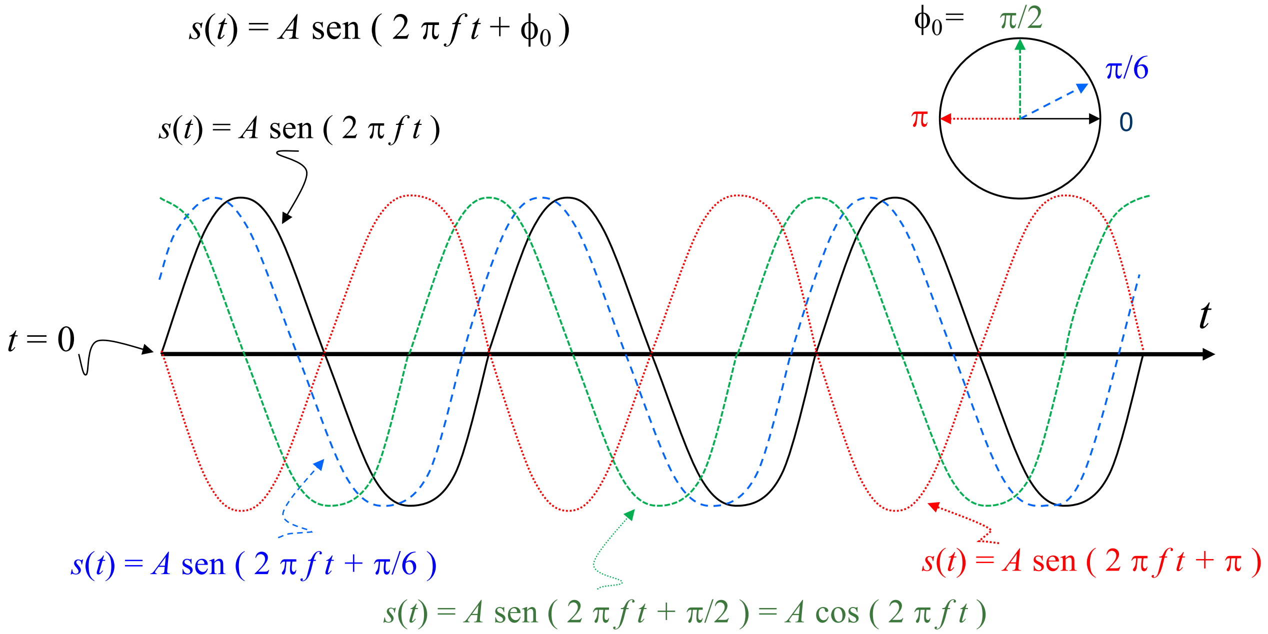
El sonido en el dominio temporal

Representación del sonido

- **La influencia de la fase inicial** (ϕ_0)

  Permite describir cuál es el valor de la onda en el inicio del tiempo (t = 0), pero su influencia es mínima en la percepción del sonido (nula si es estacionario).

$$s(t) = A \text{ sen } (2\pi f t + \phi_0)$$

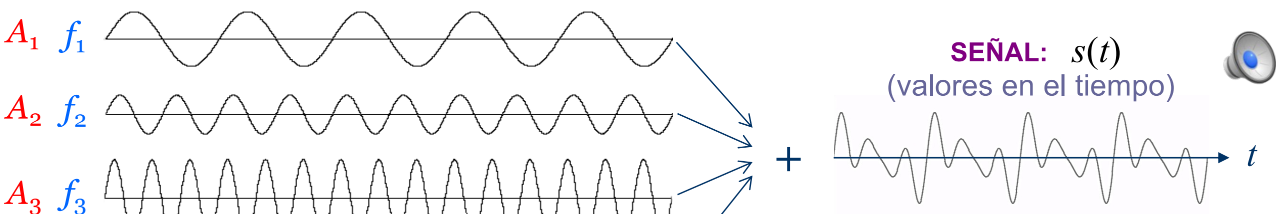


Las señales y sus frecuencias

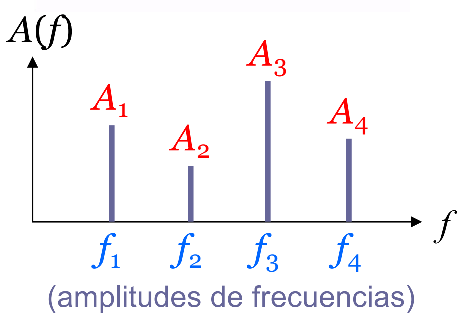
Sonidos compuestos

- **Cualquier sonido "real"** se compone de múltiples frecuencias (**frecuencias parciales**) y es una combinación lineal de funciones sinusoidales.
- Cada sinusoide aporta una frecuencia y su amplitud es su "peso" en la combinación (en la composición del sonido).

$$s(t) = A_1 \text{ sen } (2\pi f_1 t) + A_2 \text{ sen } (2\pi f_2 t) + A_3 \text{ sen } (2\pi f_3 t) + A_4 \text{ sen } (2\pi f_4 t)$$



Si representamos qué amplitud tiene cada una de las frecuencias que componen la señal, entonces tendremos el **ESPECTRO**:



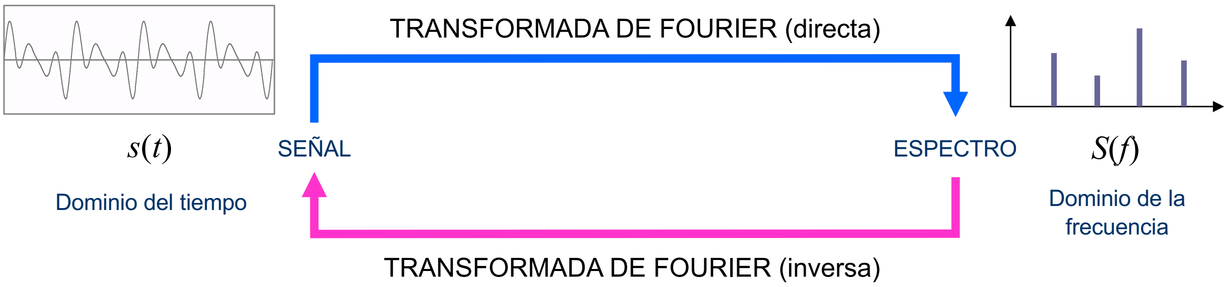
ver: 4armonicos-w.csd

Las señales y sus frecuencias

Señal y espectro son representaciones *duales*

- Representan la misma cosa en dos espacios distintos y pueden obtenerse una desde la otra.
- La conexión entre ambas representaciones es la **transformada de Fourier** que se puede calcular en ambas direcciones:

(una señal periódica puede descomponerse en un conjunto de ondas simples)



(las amplitudes del espectro multiplicadas por ondas simples recomponen una señal periódica)

Nota: estas $S(f)$ son lo mismo que las anteriores A_i

Representación digital del sonido

Señales y espectros son diferentes dentro de un ordenador.

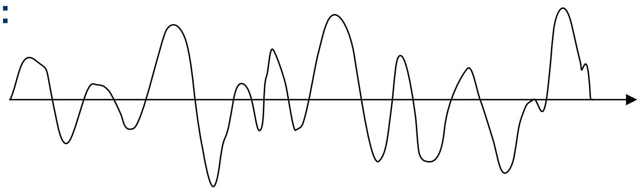
Representación del sonido digital

En digital

Para representar el sonido en un ordenador hay que convertirlo en números.

Señal analógica:

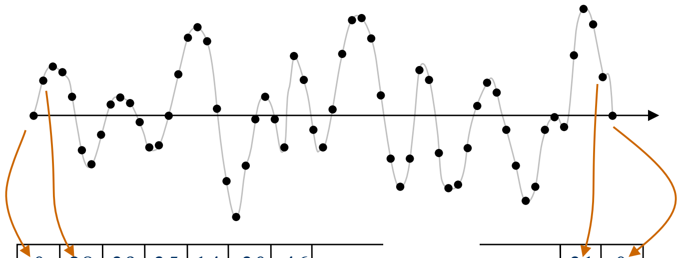
Presión o voltaje $s(t)$



t Tiempo "analógico" (en segundos)

Señal digital:

Números $s[n]$



Tomo N muestras

n Tiempo "digital" ~~(cuenta de muestras)~~

$t = n\ T_s$
t en enteros

$s[n] =$ | 0 | 28 | 39 | 35 | 14 | -30 | -46 | . . . | 31 | 0 |

n=0 n=1 n=2 . . . n=N–1

Array con N valores

T_s

Tiempo de espera antes de tomar la sig. muestra — **Periodo de muestreo** ⟹ $f_s = 1 \,/\, T_s$ **frecuencia de muestreo** — número de muestras que se toma en cada segundo

Transformada de Fourier discreta (TFD)

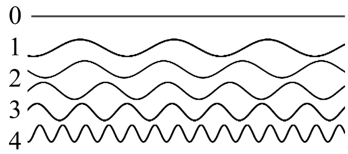
- Es la forma de calcular el **espectro de una señal digital** (por tanto, discreta).
- De una señal con N muestras, la TFD calcula N frecuencias de su espectro entre 0 y la frecuencia de muestreo, f_s

Señal $s[n]$ = N muestras

Ondas sinusoidales para N diferentes valores de frecuencia

$$S[k] = \sum_{n=0}^{N-1} s[n] \operatorname{sen}\left(-\frac{2\pi k n}{N}\right)$$

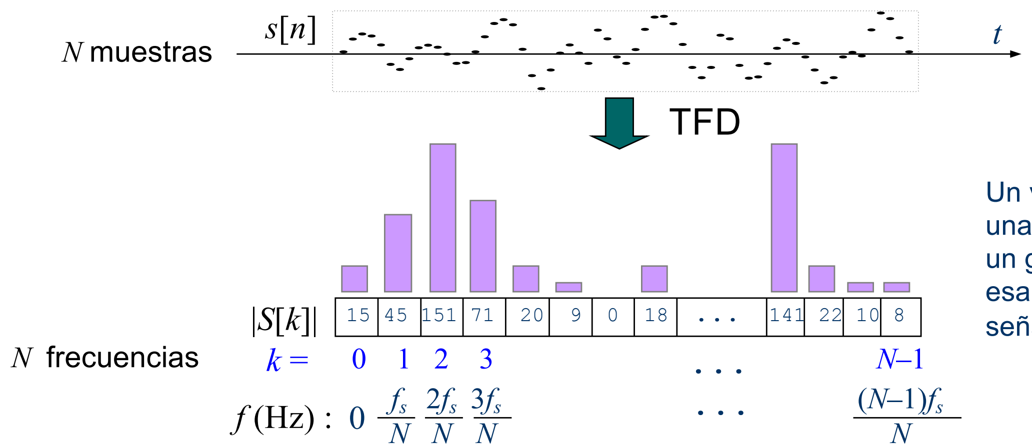
k representa el número de frecuencia que se calcula



- Lo que hace la TFD es evaluar lo que se parece la señal a cada una de esas ondas sinusoidales.
- Las N frecuencias a las que se calcula el espectro están equiespaciadas entre 0 y f_s →
  - **Resolución frecuencial:** $\Delta f = f_s \,/\, N$

Transformada de Fourier discreta (TFD)

Gráficamente:



Un valor grande para una frecuencia = hay un gran contenido de esa frecuencia en la señal (y al contrario).

Cada valor de frecuencia $= k\,\Delta f = k\,f_s / N$ (en Hz)
(entre 0 y f_s)

- En la TFD no todas las frecuencias están disponibles. Sólo los múltiplos de f_s / N.
- (!) Sólo las frecuencias entre 0 y $f_s/2$ son útiles (N / 2 frecuencias). Los valores entre $f_s/2$ y f_s son duplicados de esas. No se usan.

Evolución temporal del espectro

Para sonidos que evolucionan en el tiempo debemos calcular muchos espectros y ponerlos todos unos junto a otros.

La TFD solo funciona en la practica para sonidos estacionarios

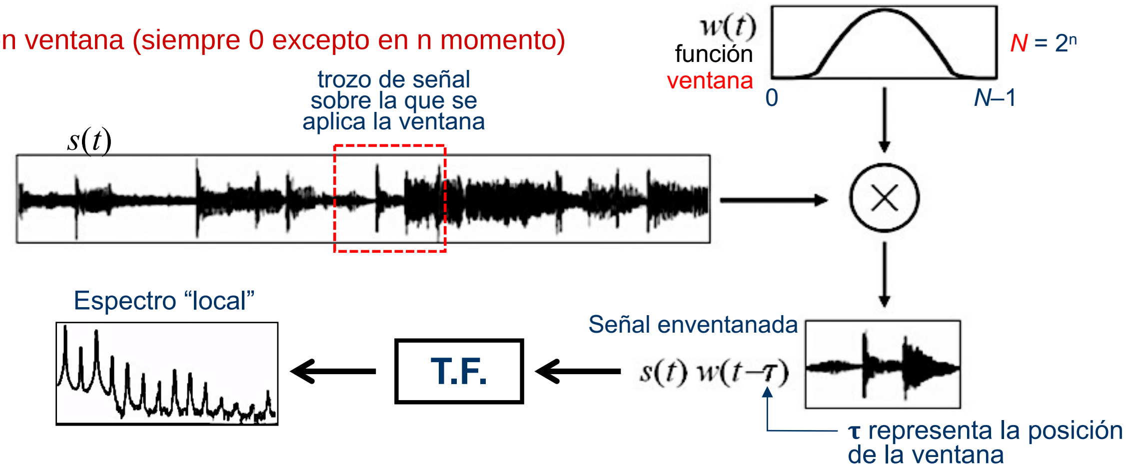
Evolución temporal del espectro

La T.F. es indefinida en el tiempo, pero en la práctica los sonidos evolucionan en el tiempo → Hacer la T.F. de la señal, pero "por trozos".

- **Solución matemática:**

  Multiplicar señal por una **"ventana"** y hacer la transformada del resultado = **Transformada de Fourier a corto plazo** (*Short Time Fourier Transform*, STFT)

multiplicar onda * funcion ventana (siempre 0 excepto en n momento)



- Así podemos calcular el espectro de la señal donde está la ventana.
- Luego hacemos avanzar la ventana y repetimos la operación, y así sucesivamente.

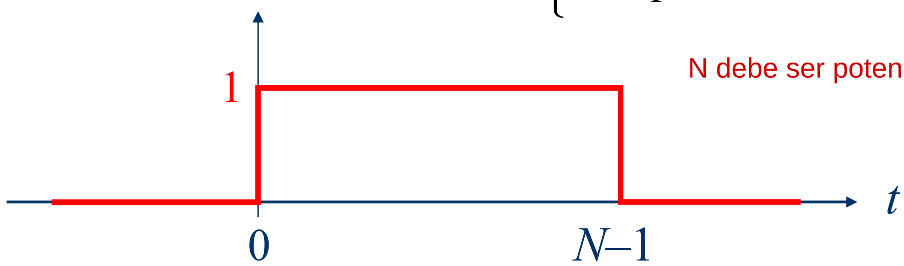
Evolución temporal del espectro

Funciones ventana:

Son funciones que se anulan en todos los puntos salvo en un entorno de cero.

P.ej.: VENTANA RECTANGULAR: $$w[m] = \begin{cases} 1 & \text{si } 0 \le m \le N-1 \\ 0 & \text{para otro valor de } m \end{cases}$$

N debe ser potencia de 2



- La función ventana **se multiplica** por la señal, dejando sólo N muestras ≠ 0
- El tamaño de la ventana debe ser **potencia de 2**, por razones de eficiencia:
  - → **Transformada Rápida de Fourier** (FFT, *Fast Fourier Transform*).
  - Tamaños habituales: …, 256, 512, 1024, 2048, 4096, 8192, …
- La presencia de la ventana **modifica el espectro** calculado:
  - El espectro de una señal *enventanada* ≠ espectro señal teórica infinita.

la ventana modifica la señal y por ende el espectro

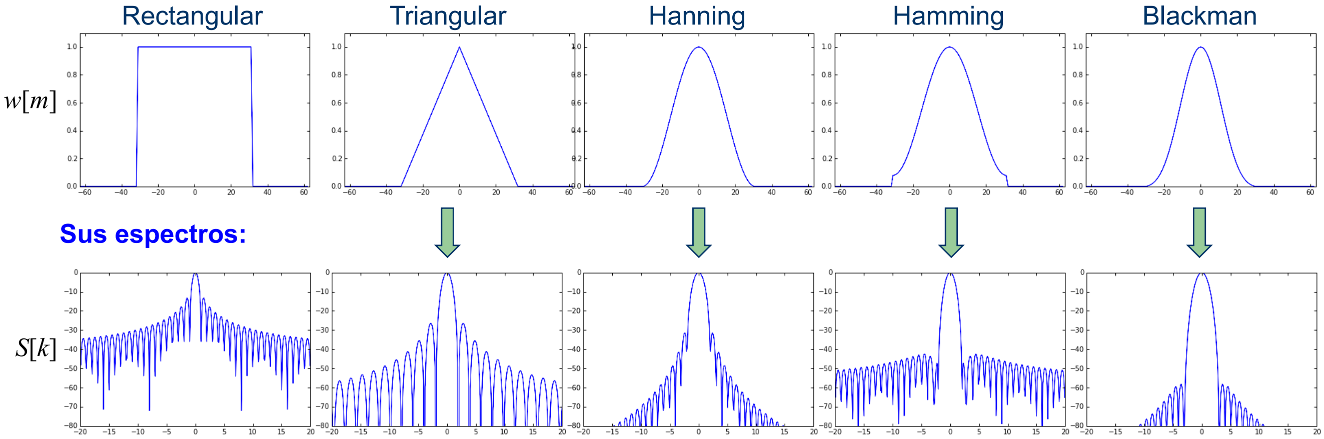
Evolución temporal del espectro

Tipos de ventanas:

Diferentes formas de ventanas modifican el espectro ideal de diferentes maneras.

T.F de las funciones inferiores

Ventanas:



Esas formas aparecerán en el espectro <u>para cada parcial detectado</u>.

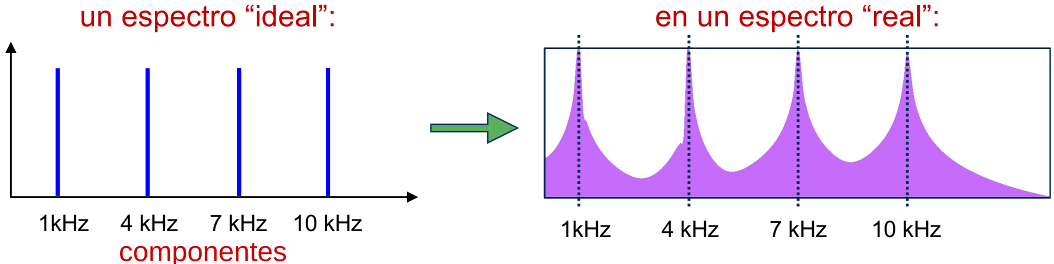
Evolución temporal del espectro

Parciales y espectros "reales"

- Los <u>parciales</u> como **líneas**, sólo aparecen en los espectros **teóricos**.
- Cualquier espectro calculado a partir de un sonido real tendrá el efecto de la aplicación de la ventana:

Parciales teóricos en un espectro "ideal":

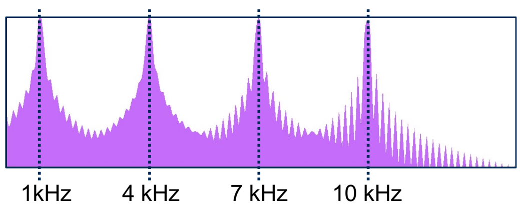
Parciales calculados en un espectro "real":



En los espectros **teóricos**, los parciales son **líneas**

En los espectros "**reales**", los parciales son **picos** (máximos locales)

Al aplicar otro tipo de ventana, cambia <u>la forma</u> del espectro pero no la posición de los parciales:



ver: audacity

Evolución temporal del espectro

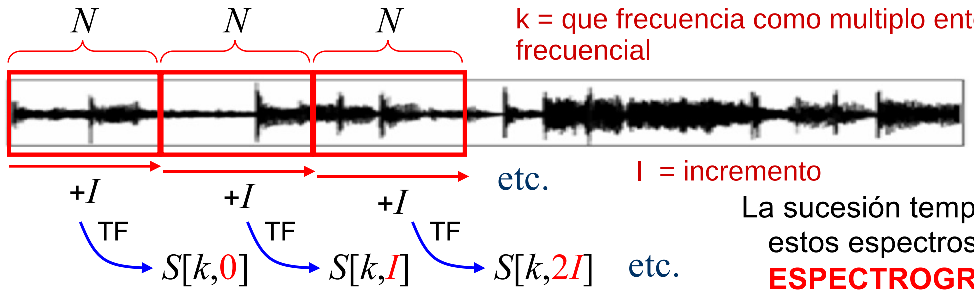
Funcionamiento de la STFT

short term fuorier transform

- La ventana w (de tamaño N) va avanzando cada vez I muestras sobre la señal.
  - → Para cada posición se obtiene un nuevo espectro $S[k,r]$ (en la posición r)
  - → El número de ventanas de análisis para una señal = Tamaño_señal / N

inicio r=0

k = que frecuencia como multiplo entero de la resolucion frecuencial



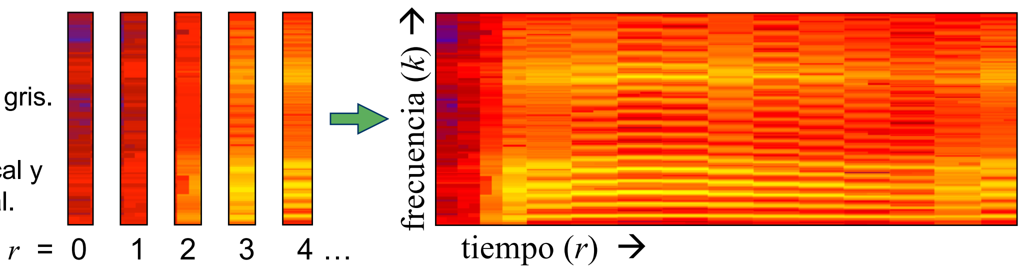
I = incremento

La sucesión temporal de todos estos espectros se denomina **ESPECTROGRAMA** .

- El espectrograma se representa como $S[k,rI]$; k : frecuencias calculadas del espectro

Para representar la **amplitud** se usan colores o niveles de gris.

Normalmente: **frecuencia** en vertical y **tiempo** en horizontal.



Evolución temporal del espectro

Resolución en tiempo y frecuencia:

Resolución frecuencial: Δf resolucion frencuencial = fs/N

Antes hemos dicho que si la TFD usa *N* valores para su cálculo:

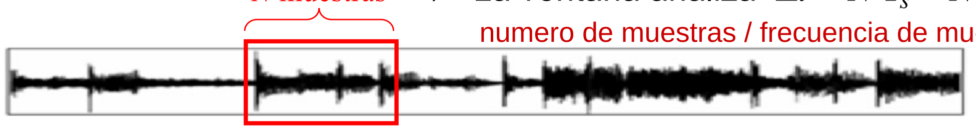
obtengo N frecuencias entre 0 y f_s separadas f_s / N (Hz) → $\Delta f = f_s / N$ (Hz)

Resolución temporal: Δt cada cuanto tiempo se analiza la señal

Se refiere al tamaño y avance de la ventana de análisis: ¿cada cuánto analizo?

N muestras ⇒ La ventana analiza $\Delta t = N\,T_s = N / f_s$ (seg.)

numero de muestras / frecuencia de muestreo



Ts = 1/fs

Por lo tanto:

⇒ Resolución frecuencial: $\Delta f = f_s / N$ (Hz)

⇒ Resolución temporal: $\Delta t = N / f_s$ (s)   inversamente proporcionales

Resolución en frecuencia y en tiempo se condicionan mutuamente: $\Delta f = 1 / \Delta t$

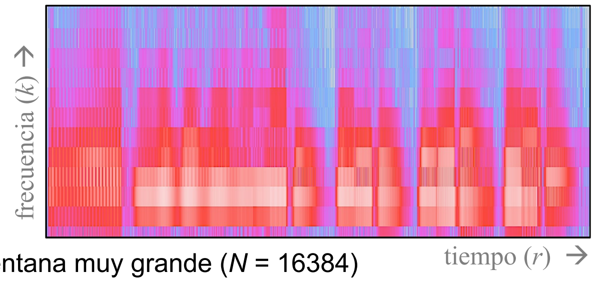
Evolución temporal del espectro

Resolución en tiempo y frecuencia:

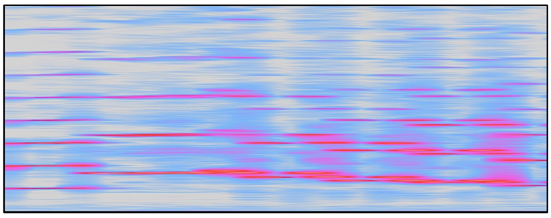
- No podemos conocer <u>con toda precisión</u> qué frecuencias ocurren y cuándo.
- Si aumento la longitud de la ventana:
  - → mejor cálculo de las frecuencias (más detalle, $\Delta f \downarrow$)
  - → pero empeora la localización ($\Delta t \uparrow$)

(y viceversa)

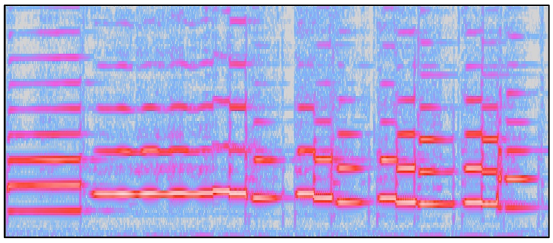
Ventana muy pequeña (N = 128)



Ventana muy grande (N = 16384)



Solución de compromiso (N = 1024)



Valores más usados: entre 512 y 4096 muestras.

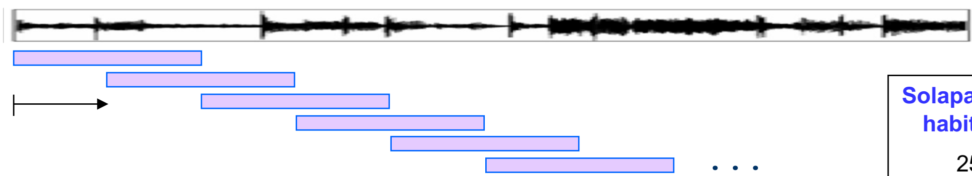
Evolución temporal del espectro

Mejora de la resolución: incrementar tiempo sin causar problemas en la frecuencia

- Se puede **paliar** el problema de la resolución temporal usando información redundante: **SOLAPAMIENTO** ($I < N$) incremento menor a tamaño de ventana
- Se expresa con el % de solapamiento en el avance de la ventana respecto a N

S = 50%

$I = N / 2$



Solapamientos habituales:

25 %
50 %
75 %
87,5 %
90 %
93,75 %

Si hay solapamiento = $S\ \%$ $\Rightarrow$ $\Delta t = \dfrac{N(1-(S/100))}{f_s}$

(y la Δf no cambia)

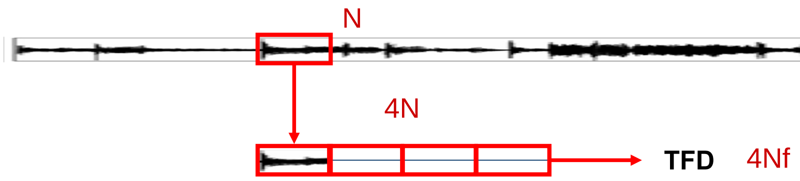
cambia

- No es gratis: el precio a pagar es la repetición de cálculos en la zona solapada.

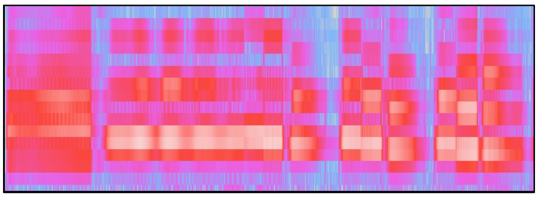
Evolución temporal del espectro

Mejora de la resolución:

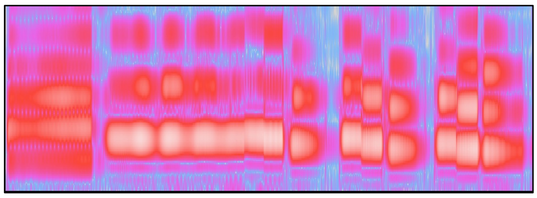
- La resolución frecuencial también se puede mejorar usando el **RELLENO POR CEROS** (zero padding).
- Se trata de añadir nN ceros al final de la ventana a analizar antes de calcular la TFD.



N = 256 muestras → 256 frecuencias



N = 256 + 7N ceros → 2048 frecuencias



Los valores calculados son interpolados respecto a los sin relleno y Δt no cambia.

Evolución temporal del espectro

El problema de la eficiencia:

- Las soluciones para obtener una buena resolución espectral pasan por hacer muchos cálculos (ventanas grandes y solapadas).
- Esto puede ser **inaceptable** en dispositivos poco potentes.

La Transformada de Fourier usa todas las muestras (N) para calcular cada una de las frecuencias (N) → Complejidad $O(N^2)$

- Por ejemplo: si la ventana es del orden de 10^3 muestras hará del orden de 10^6 operaciones!

  … para cada posición de la ventana!!

En la práctica se usa la **Transformada Rápida de Fourier** (FFT) cuya complejidad es $O(N \log_2 N)$

- Para la misma ventana anterior sólo hará $10^3 \times \log_2 10^3 \cong 10^3 \times 10 = 10^4$ operaciones.

La **condición** para que la FFT sea realmente así de rápida es que el tamaño de la ventana sea $N = 2^n$

Caracterización del espectro

Elementos relevantes que podemos encontrar en el espectro.

Caracterización del espectro

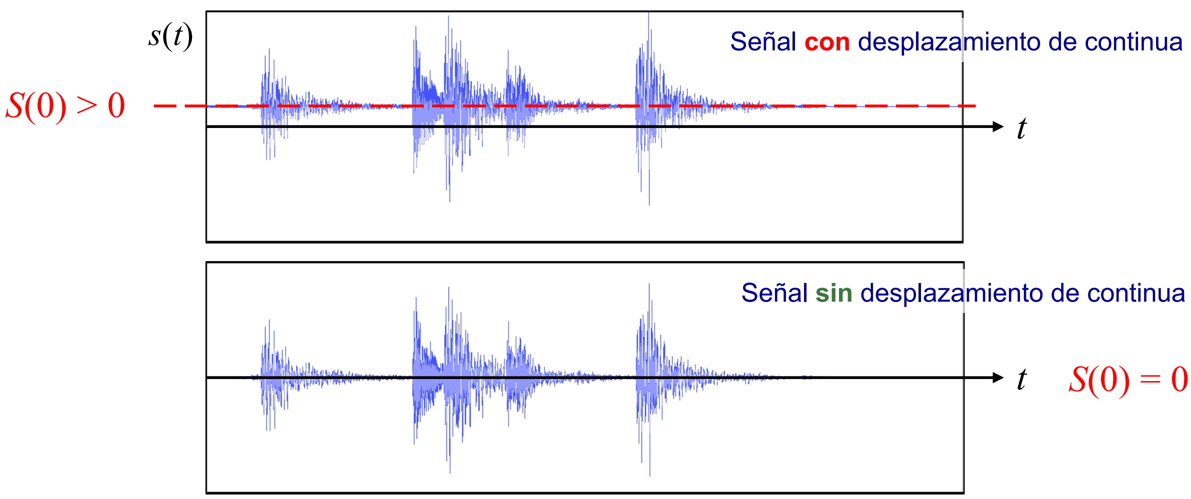
Frecuencia nula

f = cuantas veces por unidad de tiempo sucede n

T = cada cuanto tiempo pasa algo

- **Frecuencia = 0:** $S(0)$ es el **desplazamiento de continua** (*DC offset*)

$S(0)$ se puede calcular como el valor promedio de la señal.

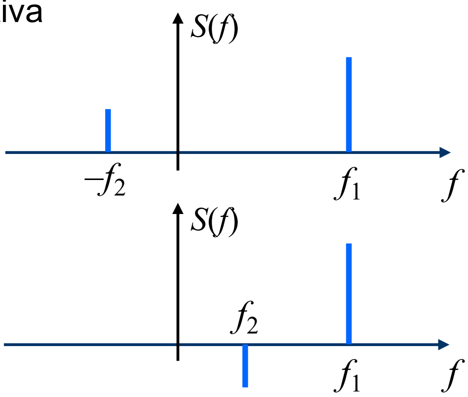


La presencia de un desplazamiento de continua reduce la amplitud que puede tener la señal al trabajar con ella: la señal de arriba "*toca*" el máximo (ya no puede aumentar su amplitud), la de abajo (que es la de arriba corregida) todavía tiene **margen** para aumentar su amplitud.

Caracterización del espectro

Frecuencias negativas

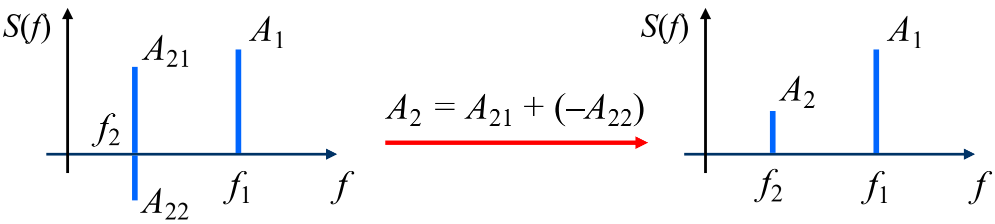
Son la misma frecuencia que una positiva pero con la fase invertida



sen(−α) = sen(α+180º)

la f negativa es lo mismo que una positiva pero con la fase invertida = 180º

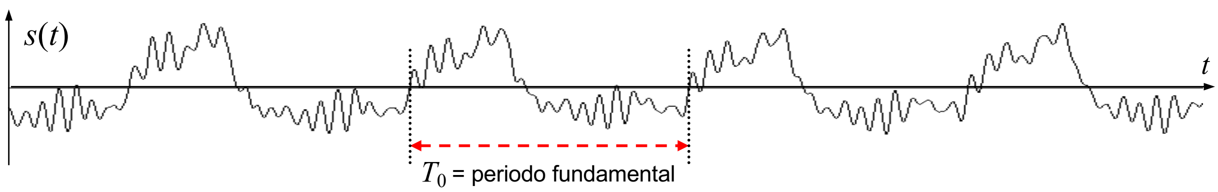
Si hay **dos componentes en una misma frecuencia** sumarán sus amplitudes si están en fase y se restarán si están en contrafase.



Frecuencia fundamental

Definición:

- Las **ondas periódicas** tienen frecuencia fundamental (f_0)
  - La f_0 es un parámetro psicoacústico muy importante.
  - **En la onda** se calcula como la inversa del periodo fundamental ($f_0 = 1 / T_0$).



El menor que verifica para toda la señal: $s(t) = s(t + T_0)$

  - Para que se comporte como periodo fundamental, el valor de T_0 debe ser < 50 ms

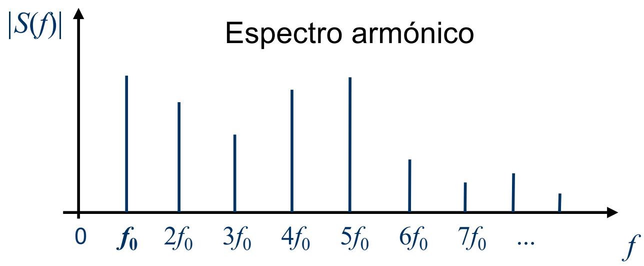
- Si nos fijamos **en el espectro** para conocer la fundamental:
  - Se calcula como el máximo común divisor de las frecuencias parciales.
    - **Sólo** las f_0 > 20 Hz pueden serlo (la frecuencia fundamental es algo que se oye)
    - Si el M.C.D. es < 20 Hz → se considera F_0 = primer parcial audible (> 20 Hz)
    - Los parciales < 20 Hz no intervienen en su cálculo.

no oimos por debajo de 20 Hz

Armonicidad

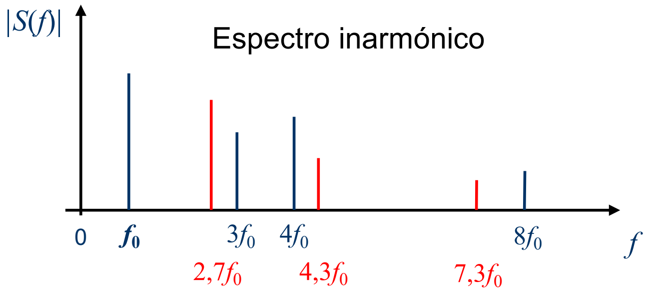
Definiciones:

– (Parciales) armónicos
- Los múltiplos enteros de la fundamental
  - Espectros armónicos
  - Sonidos armónicos



Ej.

– (Parciales) inarmónicos
- Los múltiplos no enteros de la fundamental
  - Espectros inarmónicos
  - Sonidos inarmónicos



Ej.

- Los sonidos reales no suelen ser perfectamente armónicos o inarmónicos:

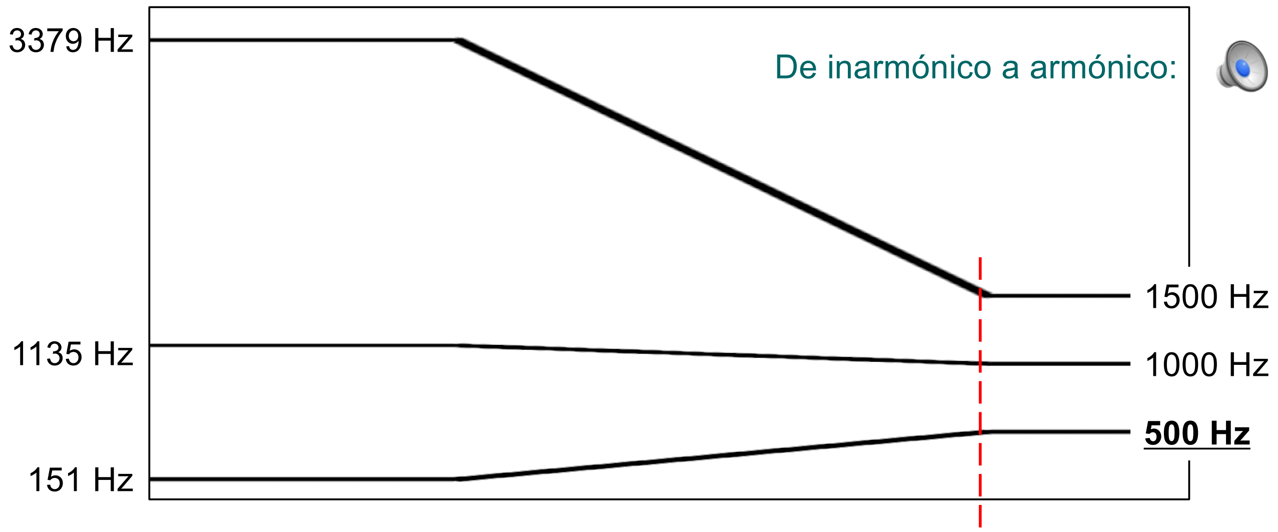
→ Grado de armonicidad (entre 0 y 1) = NP_{ARM} / NP_{TOT}

las señales armonicas son perdiocas, las señales inarmonicas no

Armonicidad

Fusión espectral:

- Es la diferencia más importante entre los sonidos armónicos e inarmónicos.
- En los sonidos armónicos sólo se percibe la fundamental y los parciales no se oyen por separado sino como el "timbre" del sonido complejo.



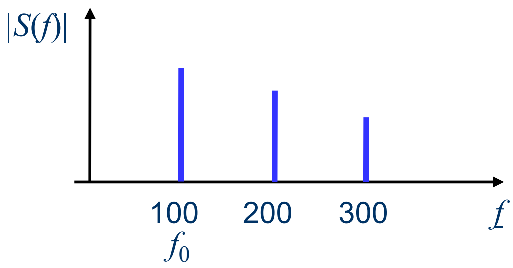
Aquí se deben percibir 3 sonidos simples independientes.

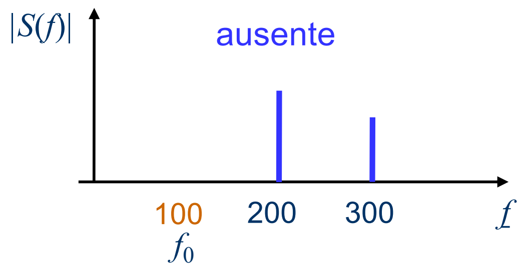
Aquí se debe percibir 1 sonido único compuesto

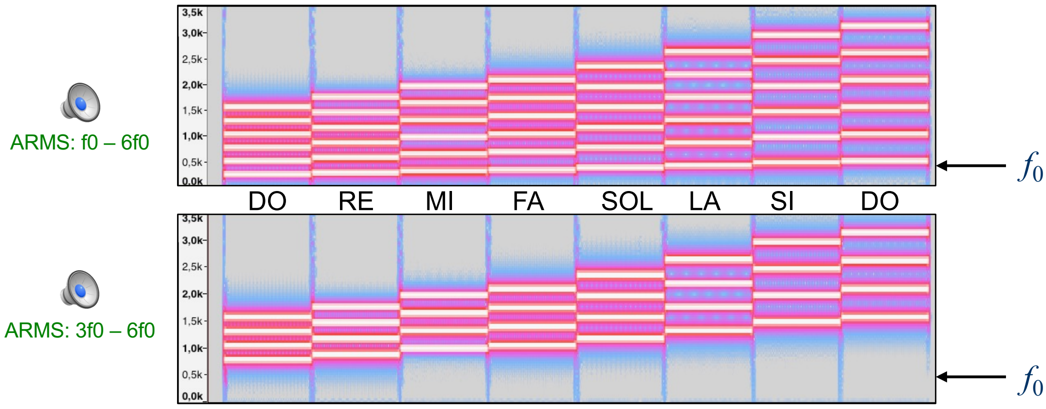
Armonicidad

Fusión espectral:

La frecuencia fundamental (f_0) puede no ser un parcial del espectro: <u>fundamental ausente</u>. No está pero *se oye*.



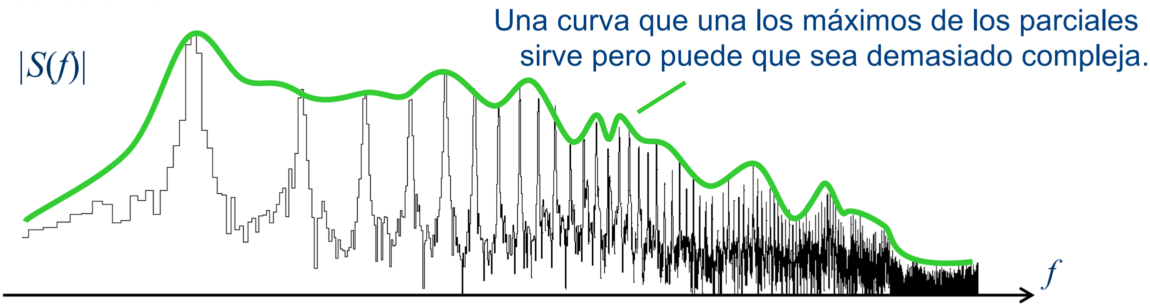


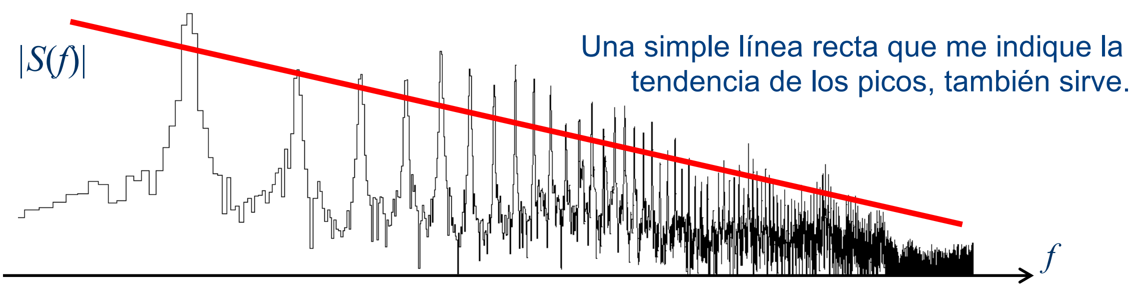


Caracterización del espectro

Envolvente espectral:

- Línea imaginaria suave que sigue las amplitudes de los parciales.
- Sirve para describir de una manera sencilla el espectro.
- Por ej., los sonidos naturales tienden a tener parciales de amplitud decreciente.





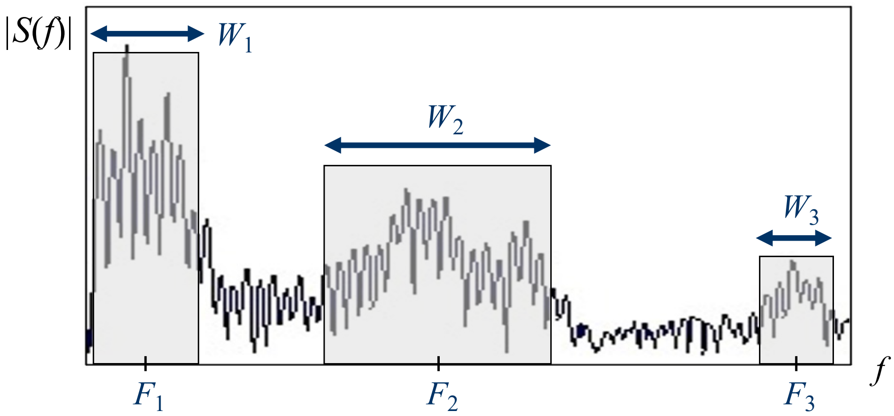
Caracterización del espectro

Formantes:

Bandas de frecuencia en las que se dan las mayores amplitudes de un espectro.

- Se deben a <u>resonancias del generador</u> del sonido → a su geometría.
- Son <u>independientes de la frecuencia fundamental</u> del sonido.
- **Se especifican** por su frecuencia central, ancho de banda y amplitud.

F_i (Hz) W_i (Hz) A_i (dB)



tienen que ver con las posibles resonancias con la fuente del sonido, dependen de la geometría de las cavidades resonantes de la fuente

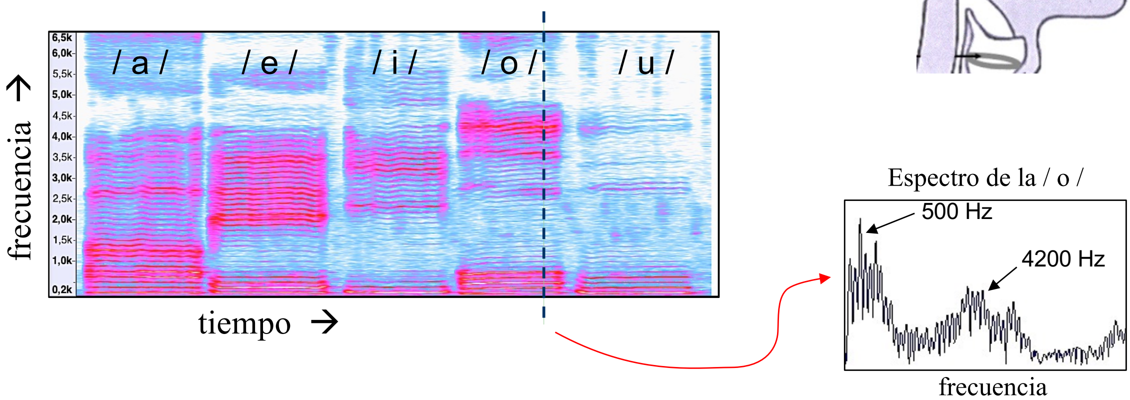
Caracterización del espectro

Formantes:

- Son determinantes para caracterizar sonidos.
- Por ejemplo, en el habla nos permiten distinguir las vocales.
- Al hablar cambiamos la forma del tracto vocal y, por tanto, sus resonancias.

Ejemplo:

Las vocales del habla → (Las vocales son la parte armónica del habla, con f_0 ~ 100 Hz hombres, 200 Hz mujeres)

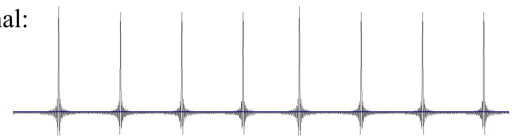


Caracterización del espectro

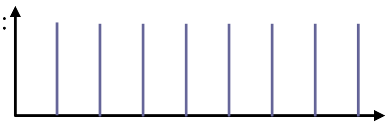
Formantes:

- Síntesis de las vocales usando la caracterización de sus formantes:
  - Las cuerdas vocales se comportan como un emisor de pulsos a la frecuencia fundamental.

Señal:



Espectro:



  - A esta señal le aplicamos un filtro pasa-banda por cada formante, centrado en la frecuencia del formante y con el mismo ancho de la banda pasante:

Vocal	Formante(s) en castellano
/a/	1000 ± 200 Hz
/e/	500 ± 100 y 2400 ± 200 Hz
/i/	300 ± 100 y 3250 ± 250 Hz
/o/	500 ± 100 y 4200 ± 250 Hz
/u/	300 ± 100 Hz

El ruido

Las perturbaciones no periódicas son percibidas como ruidos.

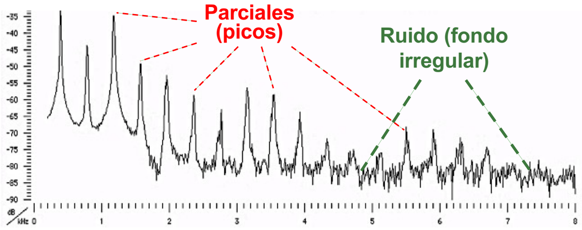
Ruido

el ruido no tiene parciales sino masas de energía difusa

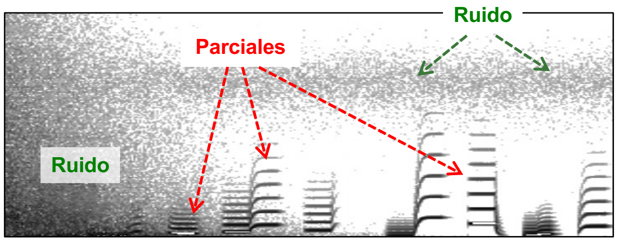
Definición

- El ruido es un componente no periódico del sonido, caracterizado en el espectro por tener parciales <u>en todas las frecuencias</u>.
- Mientras que los parciales son líneas o picos bien definidos, el ruido siempre es una "mancha" en el espectrograma:

EN EL ESPECTRO



EN EL ESPECTROGRAMA



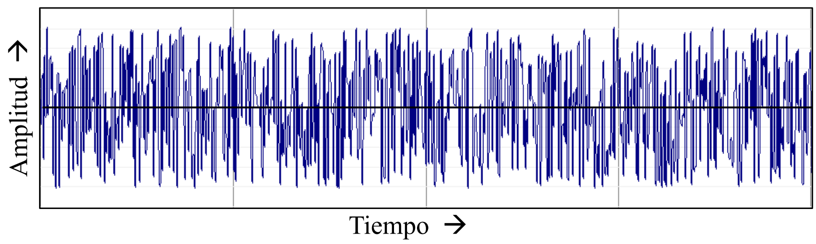
- Existen diferentes tipos de ruidos **estacionarios** en función de la distribución de la energía en su espectro.
  - Se les atribuyen **colores** en analogía a cómo se "verían" si fueran luz en el espectro visible.

Ruido

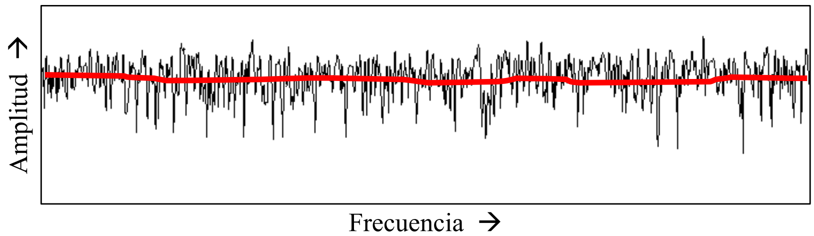
Tipos

- **Ruido Blanco**
  - Distribución de frecuencias = **constante**.
  - En el espectro, todos sus parciales tienen **la misma amplitud** promedio.
  - En digital: la señal se construye como una sucesión **de números aleatorios**.
  - Es un sonido **artificial**: inexistente en la naturaleza.
  - Lo producen dispositivos eléctricos o electrónicos (radios no sintonizadas, p.ej.)

Señal



Espectro



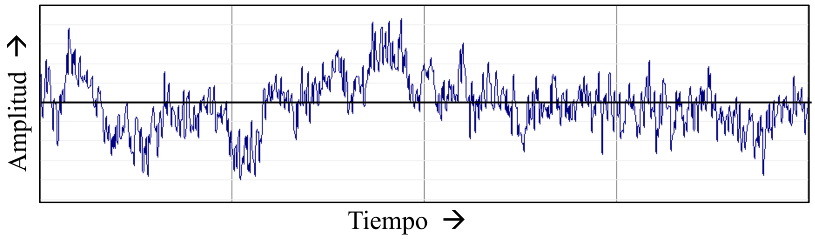
Ruido

Tipos

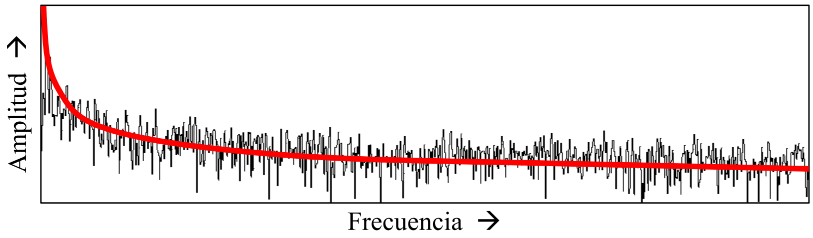
la mitad de amplitud a 2Khz que a 1Khz

- **Ruido Rosa**
  - Distribución de frecuencias del tipo **1 / *f*** → a mayor frecuencia, menos amplitud.
  - En concreto: **mitad** de amplitud **cada doble** de frecuencia.
  - Es un sonido muy frecuente **en la naturaleza**.
  - Muy utilizado en acústica porque tiene la misma energía proporcionalmente en todo el espectro → P.ej.: entre 100 y 200 Hz tiene la misma que entre 1000 y 2000 Hz.

Señal



Espectro



Sonido de una catarata de río:

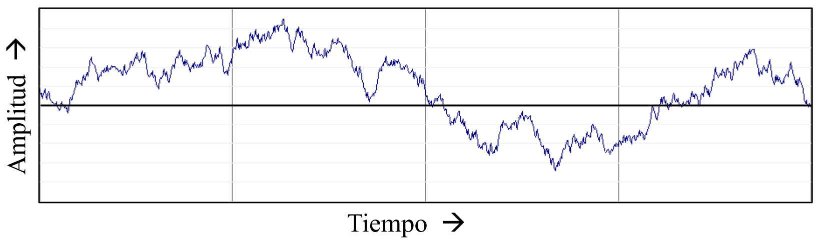
Ruido

Tipos

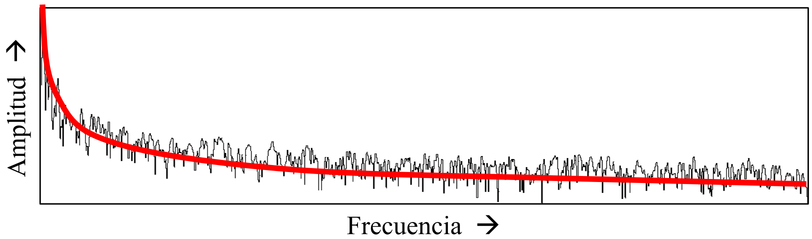
1/4 de amplitud cada doble de frecuencia

- **Ruido Marrón o Rojo**
  - Distribución de frecuencias **1 / f^2** → a mayor frecuencia, mucha menos amplitud.
  - ej.: **1/4** de amplitud **cada doble** de frecuencia. La energía se concentra en los **graves**.
  - Es un ruido existente en la **naturaleza**. Los ruidos distantes se perciben así.
  - El término *marrón* proviene de la traducción de "Brown", físico que descubrió el movimiento y los procesos naturales (brownianos) involucrados en este ruido.

Señal



Espectro

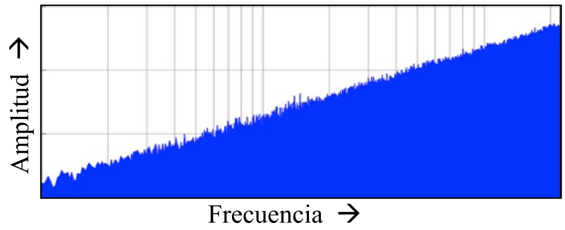


Ruido

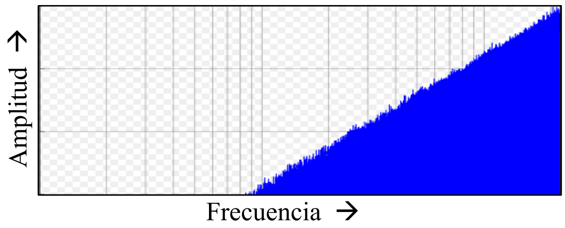
amplitud proporcional al cuadrado

Otros tipos de ruidos

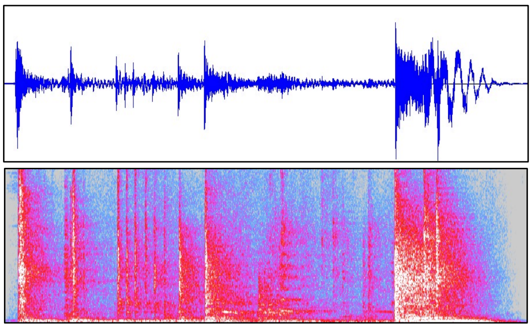
- **Ruido Azul**
  - Distribución de frecuencias = f
  - La amplitud se duplica cada doble de frec.



- **Ruido Violeta**
  - Distribución de frecuencias = f^2
  - La amplitud x 4 cada doble de frecuencia.
  - Se obtiene restando cada 2 valores consecutivos del ruido blanco .



- **Ruidos no estacionarios**
  - Cualquier movimiento no periódico de un objeto produce una alteración en el aire que percibimos como un ruido.
  - Por ejemplo, las consonantes del habla son fundamentalmente ruidos.
  - En este espectrograma no hay "líneas" horizontales (parciales), son todo "manchas".



Tema 1: Representación del sonido

Contenidos

- **Naturaleza y representación del sonido**
  - **Representación en el tiempo: la señal**
  - **Representación en la frecuencia: el espectro**
- **Representación digital**
- **Evolución temporal del espectro y FFT**
- **Caracterización del espectro:**
  - **Frecuencia fundamental y armonicidad**
  - **Envolvente espectral y formantes**
- **El ruido**