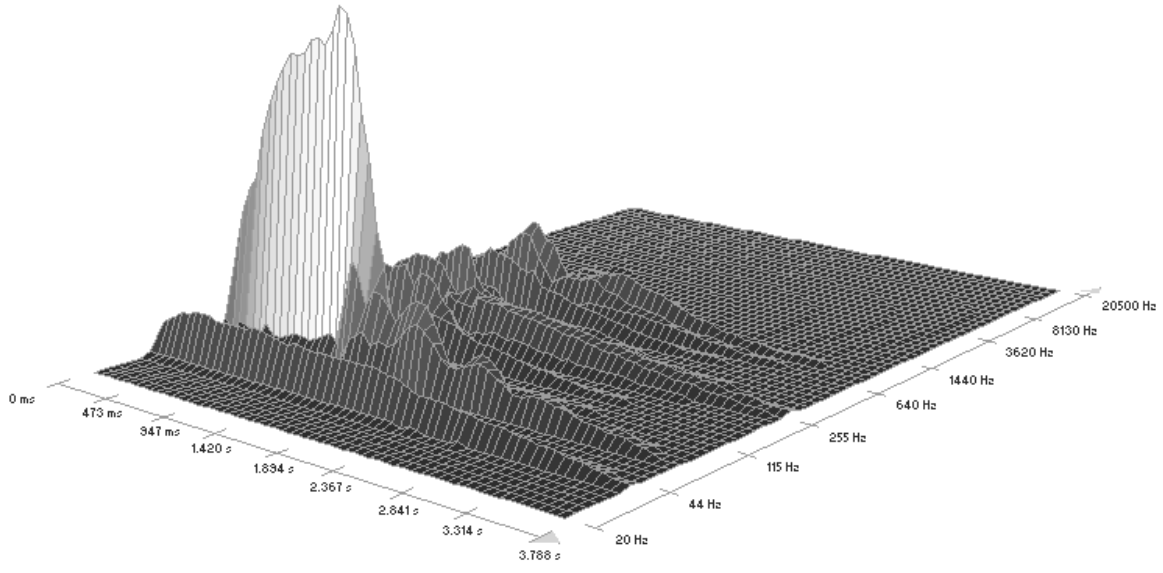
UNIVERSIDAD CATÓLICA ARGENTINA
"Santa María de los Buenos Aires"

Facultad de Artes y Ciencias Musicales

PABLO CETTA

ANÁLISIS ESPECTRAL
APLICADO A LA MÚSICA



TRANSFORMADA DISCRETA DE FOURIER

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-2\pi jkn/N}$$

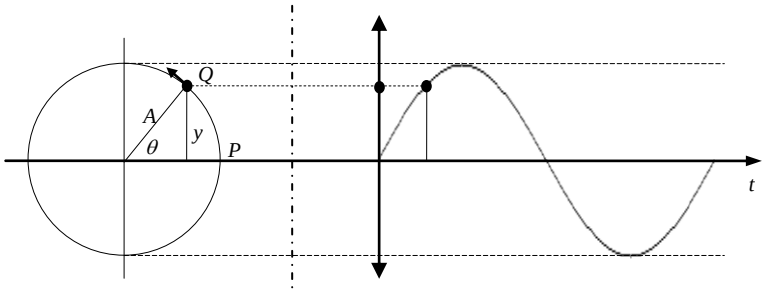
para $k = 0, 1, 2, \ldots, N\text{-}1$

La TDF es una función de variable discreta, describiremos cómo se obtiene esta expresión matemática aplicada al análisis de señales digitales y qué representa.

Movimiento oscilatorio armónico y sonidos puros

Los movimientos oscilatorios conforman el estímulo en la percepción de la sensación sonora, y la complejidad de ese movimiento determina la complejidad del sonido percibido. Vamos a definir matemáticamente el movimiento oscilatorio más simple –denominado movimiento oscilatorio armónico- que sirve de estímulo a la percepción de los sonidos puros.

Partimos de un sistema análogo, un objeto que se desplaza con movimiento circular uniforme. La proyección sobre un eje de un punto con movimiento circular uniforme describe un movimiento oscilatorio armónico.



Proyección del movimiento circular uniforme en oscilatorio armónico.

Un objeto realiza un movimiento circular uniforme si su trayectoria es una circunferencia y si su velocidad es constante, por lo cual, recorre longitudes de arco iguales en tiempos iguales.

Si el arco PQ es el espacio recorrido y t es el tiempo empleado en recorrerlo, su velocidad es

$$v = \frac{PQ}{t}$$

A la velocidad también podemos definirla en función del ángulo barrido o ángulo de fase (θ). A esta forma de expresar la velocidad la denominamos *angular* (ω).

$$\omega = \frac{\theta}{t}$$

Y desde aquí expresar el ángulo en función de la velocidad angular:

$$\theta = \omega t$$

Los ángulos pueden expresarse en distintos sistemas de medida, los más habituales suelen ser el sistema sexagesimal, que mide los ángulos en grados sexagesimales, o el sistema natural que los mide en radianes.

El sistema sexagesimal parte de un ángulo recto, que mide 90°, y define la unidad, 1°, como la noventa ava parte del ángulo recto.

El sistema natural expresa la medida de un ángulo a través de un número abstracto, y es el que se utiliza en matemática, dado que las variables que se utilizan son numéricas. La unidad en este sistema, 1 radian, se define a partir de una circunferencia de radio *r*, y es el ángulo que barre un arco de circunferencia cuya longitud es igual al radio.

$$\alpha = \frac{longitud\ del\ arco}{longitud\ del\ radio}$$

Como la longitud de la circunferencia es $2\pi r$, el ángulo que corresponde a una vuelta completa es

$$\alpha = \frac{2\pi r}{r} = 2\pi$$

El ángulo mide 2π radianes. Los radianes suelen abreviarse *rad*, pero en general no se escribe.

Podemos relacionar, entonces, ambos sistemas de medida recordando que a una vuelta completa le corresponde un ángulo de 360°, cuatro ángulos rectos, y equivale a 2π radianes. Luego, 180° es π radianes y 90° es π/2 radianes.

Nos interesa determinar la elongación (*y*), la distancia vertical del objeto al eje horizontal, que es igual a la proyección del radio de la circunferencia sobre el eje vertical.

$$y = A\sin\theta$$

Dado un triángulo rectángulo, se define al seno del ángulo (sin α) como el número que resulta del cociente entre la longitud del cateto opuesto y la longitud de la hipotenusa.

Así, expresamos *y* en función del ángulo del sistema análogo, o ángulo de fase, y la amplitud *A*, o radio. Si deseamos la elongación expresada en función del tiempo, podemos reemplazar el valor del ángulo por ω*t*, recurriendo a la velocidad angular.

$$y = A\sin(\omega t)$$

En este caso el movimiento oscilatorio comienza con fase inicial θ = 0. Si el movimiento comienza en cualquier otra posición deberíamos sumar el valor del ángulo de fase inicial al inicio del movimiento (*t* = 0).

$$y = A\sin(\omega t + \phi)$$

Para independizarnos del movimiento circular, podemos expresar la velocidad angular ω en función de la frecuencia (*f*) que es la cantidad de ciclos por segundo.

En el movimiento circular uniforme un giro de 360° se desarrolla en el tiempo que dura un ciclo, (T). Por lo tanto, si

$$\omega = \frac{\theta}{t} = \frac{2\pi}{T}$$

Como la frecuencia y el período se vinculan por $f = 1/T$, nos queda que $\omega = 2\pi f$ .Reemplazando la velocidad angular resulta:

$$y = A\sin(2\pi\ ft + \phi)$$

El valor de la elongación para una frecuencia dada está ahora en función del tiempo. Esto podríamos anotarlo como $y = y(t)$.

La descripción matemática de la señal digital es similar. La variable tiempo ya no transcurre de manera continua, sino que asume valores discretos.

$$y(n) = A\sin(2\pi\ fnT + \phi)$$

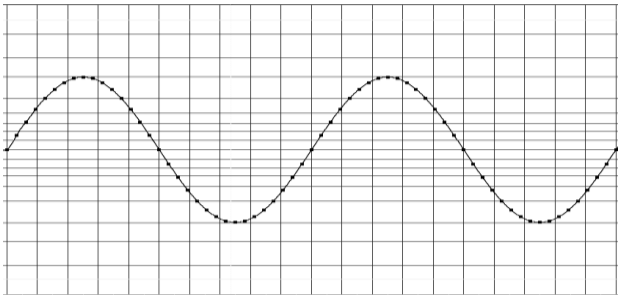
donde *n* es el número de muestra considerado y *T* es el período de muestreo (el tiempo que transcurre entre una muestra y la siguiente). El tiempo ahora se mide por cantidad de períodos de muestreo (*nT*).

Si *R* es la frecuencia de muestreo (o cantidad de muestras tomadas por segundo), podemos relacionarla con el período de muestreo *T*, del mismo modo que la frecuencia y el período en el movimiento oscilatorio, $T = 1/R$.

$$y(n) = A\sin(2\pi fn / R + \phi)$$

Dado que *R* es una constante que depende del problema, para simplificar los cálculos en la práctica se escribe:

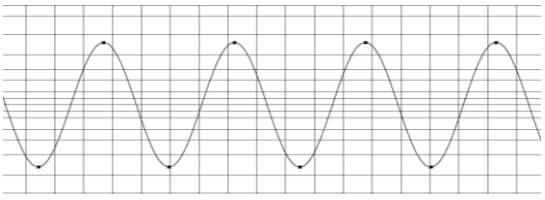
$$y(n) = A\sin(2\pi\ fn + \phi) = A\sin(\omega n + \phi)$$



Dos ciclos de una señal sinudoidal de 1000Hz, muestreada a 32kHz, y 32 muestras por ciclo.

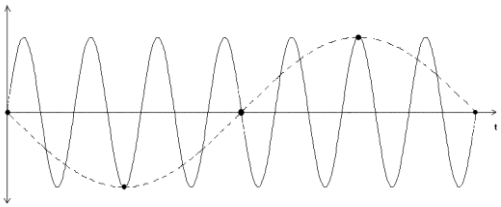
Recordemos la ley fundamental que se emplea en audio digital, conocida como *Teorema de Nyquist* (1928), expresa que para muestrear una componente de frecuencia *X*, la frecuencia de muestreo debe ser por lo menos de 2*X*. Por ejemplo, si deseamos registrar un sonido puro de 10.000 Hz, *R* debe ser de al menos de 20.000 Hz. Una frecuencia de muestreo de 44.100 Hz (de un disco compacto, por ejemplo) debería, al menos en la teoría, permitirnos registrar digitalmente sonidos cuya componente de frecuencia más alta sea de 22.050 Hz.

Si la frecuencia de muestreo es exactamente igual al doble de la frecuencia del sonido a digitalizar, cada ciclo estará representado por sólo dos valores (críticamente muestreado). Esto se observa en la figura siguiente.



Sinusoide críticamente muestreada, con dos muestras por ciclo.

Si la frecuencia de muestreo es menor al doble de la frecuencia del sonido se produce una indeterminación denominada *aliasing*, que produce distorsión. Puede verse en el siguiente gráfico que la lectura de las muestras obtenidas reproduce una componente de menor frecuencia que no existía en la señal original.



Una sinusoide muestreada por menos de dos muestras por ciclo, que produce *aliasing* .

Para evitar el *aliasing* se emplea, en los conversores, un filtro pasa-bajos en la entrada. Su frecuencia de corte se encuentra a la mitad de la frecuencia de muestreo, e impide que cualquier componente que supere en frecuencia la mitad de *R* pase al conversor analógico-digital. Recordemos que un filtro pasa-bajos deja pasar todas las componentes que se encuentran por debajo de su frecuencia de corte y retiene al resto. Otro filtro pasa-bajos a la salida, suaviza el contorno de la forma de la señal generada por el conversor digital-analógico a partir de la secuencia de números almacenada. Una señal críticamente muestreada sólo puede ser reproducida como una señal sinusoidal, dado que la frecuencia de corte del filtro no permite el paso de armónicos.

Retomando las cuestiones matemáticas que nos ocupan, veamos ahora otra forma de expresar la función $y(n) = A\sin(\omega n + \phi)$

Nuestra función es una constante *A* multiplicada por el seno de la suma de dos ángulos. La trigonometría enseña que

$$\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$$

Entonces, podemos escribir

$$\begin{aligned} y(n) = A\sin(\omega n + \phi) &= A[\sin(\omega n)\cos\phi + \cos(\omega n)\sin\phi] \\ &= A\sin\phi\cos(\omega n) + A\cos\phi\sin(\omega n) \\ &= a\cos(\omega n) + b\sin(\omega n) \end{aligned}$$

donde $a = A\sin\phi$ y $b = A\cos\phi$

Si deseamos obtener el valor de *A*, utilizamos la relación trigonométrica fundamental

$$\sin^2\phi + \cos^2\phi = 1$$

Sumando los cuadrados de a y b

$$\begin{aligned} a^2 + b^2 &= (A\sin\phi)^2 + (A\cos\phi)^2 \\ &= A^2\sin^2\phi + A^2\cos^2\phi \\ &= A^2(\sin^2\phi + \cos^2\phi) \\ &= A^2 \end{aligned}$$

Por lo tanto, la amplitud es

$$A = \sqrt{a^2 + b^2}$$

Para conocer el ángulo de fase inicial planteamos el siguiente cociente

$$\frac{a}{b} = \frac{A\sin\phi}{A\cos\phi} = \tan\phi$$

Dado un triángulo rectángulo, se define al coseno del ángulo (cos α) como el número que resulta del cociente entre la longitud del cateto adyacente y la longitud de la hipotenusa. Dado un triángulo rectángulo, se define a la tangente del ángulo (tan α) como el número que resulta del cociente entre la longitud del cateto opuesto y la longitud del cateto adyacente, resultando igual al cociente entre el seno y el coseno del ángulo.

Conocida la tangente del ángulo, podemos conocer el ángulo a través de su inversa, ¿cuál es el ángulo cuya tangente vale $\frac{a}{b}$?, esta relación se denomina arco tangente y se escribe así:

$$\phi = \tan^{-1}\frac{a}{b}$$

Identidad de Euler

La identidad de Euler, una fórmula muy utilizada en el procesamiento digital de señales de audio, nos dice que

$$e^{j\theta} = \cos\theta + j\sin\theta$$

e , al igual que π, es un número irracional cuyo valor es 2.718281828459…, y j representa a la unidad imaginaria.

El conjunto de números reales, ***R***, puede representarse mediante una recta, que llamamos recta real, donde a cada punto que la forma le corresponde un número real. Las operaciones que realizamos con estos números cumplen con los axiomas y leyes que definen a ***R***.

Pero en el caso de $\sqrt{-1}$ no encontramos solución en ***R***, y en consecuencia se define un nuevo conjunto de números. Estos son los números imaginarios (***I***), a cuya unidad llamaremos j, y que también pueden ser representados en una recta, denominada recta imaginaria.

$$j = \sqrt{-1}$$

Ambos conjuntos, ***R*** e ***I***, definen un nuevo conjunto de números, los números complejos, ***C***.

Un número complejo se expresa así: $z = x + jy$, donde x es la parte real del complejo, e y la parte imaginaria.

Consideremos un par de ejes cartesianos ortogonales, donde el horizontal es la recta real y el vertical la recta imaginaria. Ambos ejes determinan un plano, llamado el plano complejo, en el cual cada punto de ese plano tiene una componente real y una imaginaria. Por esto podemos expresar el número complejo z en forma polar del siguiente modo:

$$z = r\cos\theta + jr\sin\theta$$

donde r es el vector que posiciona al punto, y θ el ángulo que forma r con el semieje positivo de la recta real.

Haciendo uso de la identidad de Euler podemos escribir a z de la siguiente forma:

$$z = re^{j\theta} = r(\cos\theta + j\sin\theta) = r\cos\theta + jr\sin\theta$$

Esta forma exponencial reduce considerablemente la dificultad de las operaciones con números complejos. En la multiplicación, por ejemplo, la operación resulta más simple:

$$z_1 z_2 = (r_1 e^{j\theta_1})(r_2 e^{j\theta_2}) = (r_1 r_2)(e^{j\theta_1} e^{j\theta_2}) = r_1 r_2 e^{j(\theta_1+\theta_2)}$$

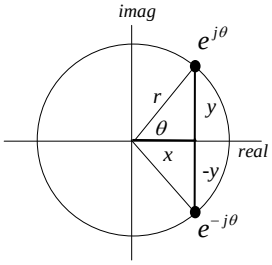
Solo basta recordar las propiedades del producto de potencias de igual base, para comprender lo anterior.

Por otra parte, es interesante observar que haciendo $\theta = \pi$ en la relación de Euler, surge la fórmula considerada de singular belleza

$$e^{j\pi} = \cos\pi + j\sin\pi$$
$$e^{j\pi} = -1 + 0$$
$$e^{j\pi} + 1 = 0$$

dado que reúne de forma simple a las constantes *e*, π, *j*, 0 y 1 con las operaciones de suma, multiplicación, potenciación e igualdad.

Otra derivación importante surge de la suma de dos complejos conjugados, aquellos que tienen su parte imaginaria con distinto signo. Si $z_1 = \cos\theta + j\sin\theta$, $z2 = \cos\theta - j\sin\theta$ es el complejo conjugado, con un ángulo igual a –θ.



Complejos conjugados

En esta situación se verifica que $\cos(-\theta) = \cos\theta$, y por otro lado que $\sin(-\theta) = -\sin\theta$

$$z_1 + z_2 = e^{j\theta} + e^{-j\theta} = \cos\theta + j\sin\theta + \cos\theta - j\sin\theta = 2\cos\theta$$

Lo que nos permite determinar la componente real como:

$$\cos\theta = \frac{e^{j\theta} + e^{-j\theta}}{2}$$

Y si planteamos la diferencia entre estos complejos conjugados

$$z_1 - z_2 = e^{j\theta} - e^{-j\theta} = \cos\theta + j\sin\theta - \cos\theta + j\sin\theta = 2j\sin\theta$$

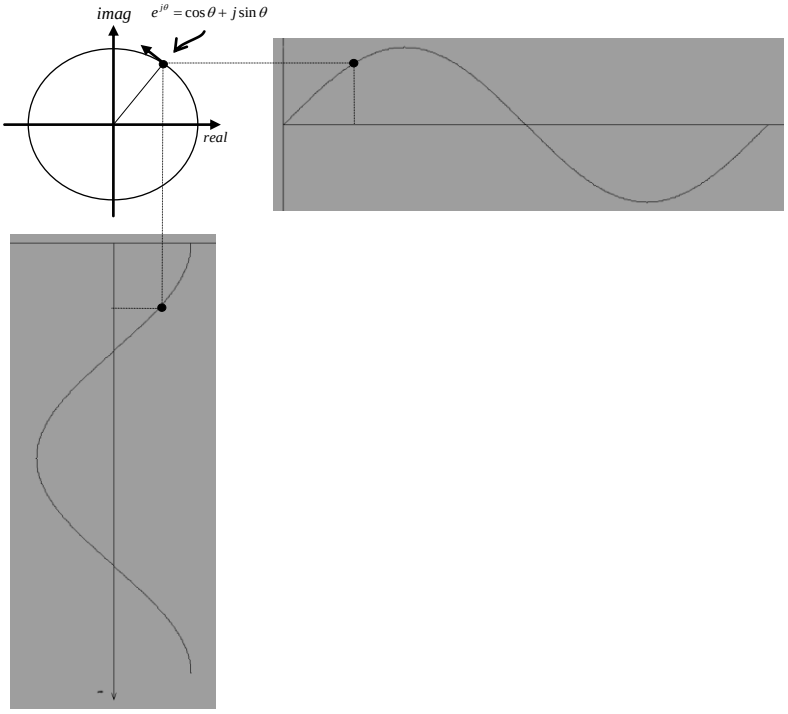
podemos determinar la componente imaginaria del complejo:

$$\sin\theta = \frac{e^{j\theta} - e^{-j\theta}}{2}$$

Observemos que expresamos seno y coseno del ángulo en función de θ y - θ.

La sinusoide compleja

En el procesamiento de señales digitales suele usarse el término sinusoidal para una señal definida tanto por la función seno como por la función coseno. De hecho, una función coseno con fase inicial 0 no es sino una función seno con fase inicial $-\pi/2$. Su utilización es común cuando se trata de señales representadas de forma compleja.



Movimiento circular uniforme en el plano complejo.

Consideremos el movimiento circular uniforme de un punto sobre la circunferencia unitaria. La proyección sobre el eje real describe un movimiento oscilatorio armónico que es posible expresar en términos del coseno del ángulo de fase, y otro movimiento oscilatorio sobre el eje imaginario, expresable como el seno del mismo ángulo. Si recurrimos a la identidad de Euler, podemos definirlo del siguiente modo:

$$e^{j\theta} = \cos\theta + j\sin\theta$$

O, ya que $\theta = \omega t$

$$e^{j\omega t} = \cos\omega t + j\sin\omega t$$

Consideramos nuevamente

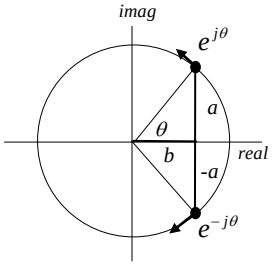
$$\cos(\omega t) = \frac{e^{j(\omega t)} + e^{-j(\omega t)}}{2}$$

notamos que en realidad cos(ωt) es la suma promediada de dos movimientos circulares de sentidos opuestos, y por lo tanto contiene ambas frecuencias,

Del mismo modo,

$$\sin \omega t = \frac{e^{j\omega t} - e^{-j\omega t}}{2}$$

también contiene ambas frecuencias.



Movimientos circulares de igual velocidad y sentido opuesto.

Transformada Discreta de Fourier

La Transformada Discreta de Fourier (TDF) nos permite obtener el espectro de una forma de onda. Este espectro está representado por componentes sinusoidales cuyas frecuencias son armónicos de una fundamental, denominada frecuencia fundamental de análisis.

Si partimos de una forma de onda arbitraria, *x(n)* y enviamos para su análisis una cantidad *N* de muestras sucesivas, representativas de un único ciclo de una onda infinitamente periódica. El resultado obtenido contiene los armónicos presentes en el período considerado de esa onda supuestamente periódica.

Consideremos un caso simple, *x(n)= A sin(ωn),* cuyo período está representado por *N* muestras.

$$x(n) = A\sin(\omega n) = A\sin(2\pi\ fn) = A\sin(2\pi \frac{1}{N} n)$$

Si *N* = 8, los valores de amplitud para cada muestra del ciclo estarán dados por:

$$x(n) = A\sin(2\pi \frac{1}{8} n) = A\sin(\frac{\pi}{4} n)$$

y son:

$$x(n) = 0, \frac{A}{\sqrt{2}}, A, \frac{A}{\sqrt{2}}, 0, -\frac{A}{\sqrt{2}}, -A, -\frac{A}{\sqrt{2}}$$

Con n = 0, 1, 2, 3, 4, 5, 6, 7 = 0, …, N-1

La suma de todos los valores es igual a cero, porque la función seno es simétrica respecto al eje horizontal. La pregunta es entonces cómo extraer de forma matemática la amplitud máxima de esa sucesión de muestras.

Dado que $x(n) = A\sin(\omega n)$, si multiplicamos a la función por $\sin(\omega n)$ nos queda:

$$A\sin(\omega n)\sin(\omega n) = A\sin^2(\omega n)$$

Al elevar los valores de la función seno al cuadrado, la amplitud de cada muestra se vuelve positiva en todos los casos. Si sumamos ahora los valores de todas las muestras, obtenemos lo siguiente:

$$\sum_{n=0}^{N-1} x(n)\sin(\omega n) = 0 + \frac{A}{2} + A + \frac{A}{2} + 0 + \frac{A}{2} + A + \frac{A}{2} = 4A = N\frac{A}{2}$$

Hemos extraído, de este modo, la amplitud de $x(n)$ a través de un procedimiento puramente matemático. Del mismo modo, si $x(n)$ representa a una forma de onda compleja, podríamos utilizar el mismo procedimiento para extraer la amplitud de todas sus componentes.

En este caso multiplicaríamos a la función $x(n)$ por $\sin(\omega n)$, $\sin(2\omega n)$, $\sin(3\omega n)$, y así sucesivamente.

A fin de comprobar el procedimiento podemos multiplicar a la señal $x(n) = \sin(\omega n)$ por $\sin(2\omega n)$ a fin de intentar extraer un segundo armónico. El resultado debe dar 0, pues sabemos de antemano que la función a analizar es una única sinusoide.

La solución al problema planteado la encontramos a partir de la utilización de la siguiente identidad trigonométrica:

$$\sin(\alpha)\sin(\beta) = \frac{1}{2}[\cos(\alpha - \beta) - \cos(\alpha + \beta)]$$

Nos queda, entonces

$$\sum_{n=0}^{N-1} A\sin(\omega n)\sin(2\omega n) = \sum_{n=0}^{N-1} \frac{A}{2}[\cos(-\omega n) - \cos(3\omega n)]$$

$$= \frac{A}{2}\sum_{n=0}^{N-1}\cos(-\omega n) - \frac{A}{2}\sum_{n=0}^{N-1}\cos(3\omega n) = 0$$

Consideremos ahora el caso de una forma de onda compleja cualquiera, ya que por el Teorema de Fourier sabemos que siempre podrá descomponerse en un número finito de ondas simples de frecuencia y amplitud determinadas. La expresión de cada componente puede tener esta forma:

$$x(n) = A\sin(\omega n + \phi) = a\cos(\omega n) + b\sin(\omega n)$$

Nos interesa determinar a y b para conocer A y ϕ.

Podemos multiplicar las muestras por cos(ωn) y sumar los resultados para extraer *a*

$$x(n)\cos(\omega n) = a\cos^2(\omega n) + b\sin(\omega n)\cos(\omega n)$$

y multiplicar por sin(ωn) para extraer *b*

$$x(n)\sin(\omega n) = a\cos(\omega n)\sin(\omega n) + b\sin^2(\omega n)$$

Dado que

$$\sum_{n=0}^{N-1} \sin(\omega n)\cos(\omega n) = 0$$

nos queda entonces

$$\sum_{n=0}^{N-1} x(n)\cos(\omega n) = N\frac{a}{2} \qquad \text{y} \qquad \sum_{n=0}^{N-1} x(n)\sin(\omega n) = N\frac{b}{2}$$

Luego por cos($2\omega n$) y sin($2\omega n$) para el segundo armónico, y así sucesivamente. De modo general podemos escribirlo como sigue.

$$a_k = \frac{1}{N}\sum_{n=0}^{N-1} x(n)\cos(k\omega n) \qquad \text{y} \qquad b_k = \frac{1}{N}\sum_{n=0}^{N-1} x(n)\sin(k\omega n)$$

Posteriormente, con los valores de a_k y b_k podemos calcular la amplitud (A_k) y la fase (ϕ_k) correspondientes a cada armónico, teniendo presente que la amplitud obtenida corresponde **a la mitad de su valor real**, pues se halla repartida entre el valor de frecuencia positivo (ωn) y el negativo ($-\omega n$). El espectro calculado es, por lo tanto, simétrico alrededor de 0 Hz.

Tratemos de determinar ahora cuáles son las frecuencias de las componentes que resultan del análisis. La variable *k* representa el índice de frecuencia, y la frecuencia fundamental de análisis ocurre para *k* = 1

$$f_k = \frac{kR}{N} \qquad \text{para } k = 0, 1, 2, \ldots, N\text{-}1$$

f_k es la frecuencia del armónico *k*, *R* es la frecuencia de muestreo (*sampling rate*), y *N* es el número de muestras analizadas. Para *k* = 1, entonces, obtenemos la frecuencia de la fundamental, para *k* = 2 la del segundo armónico, y así sucesivamente.

Resumiendo podemos decir que mediante productos y sumas, determinamos la amplitud y el ángulo de fase de cada componente de la señal analizada.

Otro modo de representar la función es valiéndonos de la relación de Euler. En nuestro caso podemos adaptar la identidad a las variables que nos interesan particularmente:

$$e^{j2\pi ft} = \cos(2\pi\ ft) + j\sin(2\pi\ ft)$$

y dado que cos(-*x*) = cos(*x*) y sin(-*x*) = -sin(*x*), podemos escribir

$$e^{-j2\pi ft} = \cos(-2\pi\ ft) + j\sin(-2\pi\ ft) = \cos(2\pi\ ft) - j\sin(2\pi\ ft)$$

Como la frecuencia está dada por kR/N, y el tiempo por nT, $2\pi ft$ se convierte en $2\pi \frac{kR}{N} nT$

Y como $R = 1/T$ queda, simplificando, $2\pi\ ft = 2\pi kn/N$

Por lo tanto,

$$e^{-j2\pi kn/N} = \cos(2\pi kn/N) - j\sin(2\pi kn/N)$$

Si llamamos *a* a la parte real y *b* a la parte imaginaria, según vimos, podemos calcular *A* y ϕ para cada componente.

Finalmente, la función queda expresada del siguiente modo:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-2\pi\ jkn/N} \quad \text{para } k = 0, 1, 2, \ldots, N\text{-}1$$

$$TDF\left[x(n)\right] = X(k)$$

$$\left|X(k)\right| = \sqrt{a_k{}^2 + b_k{}^2} = A_k$$

$$\arg\left[X(k)\right] = \tan^{-1}\frac{b_k}{a_k} = \phi_k$$

Cada una de las componentes del espectro *X*(*k*) en que puede descomponerse una forma de onda compleja *x*(*n*), de la que tenemos *N* muestras, puede obtenerse calculando su amplitud y su ángulo de fase. Esto se logra sumando los productos entre cada muestra y una función sinusoidal compleja.

El siguiente fragmento de programa en lenguaje C describe la aplicación práctica de la fórmula.

```
for(k = 0; k < N; k++){                          //k de 0 a N-1
       real[k] = imag[k] = 0;                    //se inicializan a 0
       for(n = 0; n < N; n++){                   //n de 0 a N-1
              real[k] += x[n] * cos(2Pi * k * n / N) ;
              imag[k] -= x[n] * sin (2Pi * k * n / N) ;
              }
       real[k] /= N ;                            //se escala por 1/N
       imag[k] /= N ;                            //para A entre 0 y 1
       }
```

Los datos almacenados en las matrices *real* e *imag* son los coeficientes *a* y *b*, desde donde se obtiene *A* y la fase. La frecuencia representada por el índice *k* es, como vimos antes, $f_k = kR / N$

Si analizamos 16 muestras de un sonido muestreado a 16.000 Hz, la frecuencia fundamental de análisis corresponde a 1000 Hz (*R/N*), y las demás componentes (armónicos) tienen frecuencias múltiplos de esa fundamental.

La cantidad de valores que se obtienen equivale a *N* (el número de muestras). Sabiendo que hay 16 muestras obtenemos 16 valores de *a* y *b*, y el resultado es en espejo, dado que al superar *R*/2 se produce *aliasing*. Algo similar ocurre en el cine, cuando se filma una rueda con rayos que acelera su velocidad (una carreta, por ejemplo). En determinado momento, vemos que la rueda

parece detenerse y comienza a girar en sentido opuesto. Recordemos que una película también es discreta, y está formada por 24 cuadros por segundo.

Los primeros 9 (0 a 8) valores de k (hasta $R/2$, frecuencia de Nyquist) corresponden a las frecuencias positivas del espectro, los siguientes, a las negativas en orden decreciente. Ejemplo:

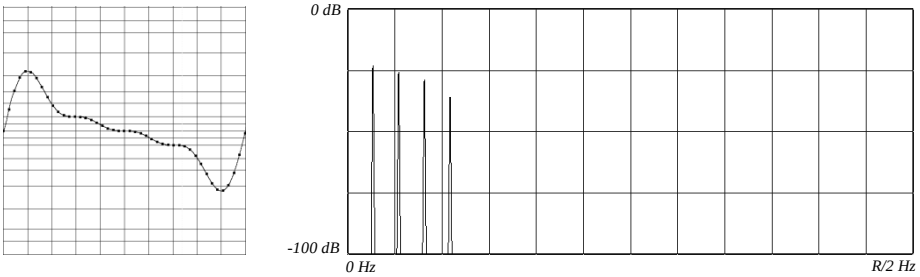
$k = 0$	0 R/N Hz
$k = 1$	1 R/N Hz
$k = 2$	2 R/N Hz
…	
$k = 8$	8 R/N Hz = $R/2$
$k = 9$	-7 R/N Hz
$k = 10$	-6 R/N Hz
…	
$k = 15$	-1 R/N Hz

Los valores de amplitud para cada índice de frecuencia son, según vimos, $N\frac{A}{2}$, a excepción del armónico de 0 Hz ($k = 0$), también llamado DC por *direct current*, y el armónico con frecuencia $R/2$ o frecuencia de Nyquist ($k = 8$ en el ejemplo anterior), que equivalen a N veces A. Esto se produce porque tanto la componente de 0 Hz como la componente cuya frecuencia es la mitad de la frecuencia de muestreo poseen su parte imaginaria igual a 0, sin que pueda distinguirse el signo de la frecuencia. Una componente críticamente muestreada (dos muestras por ciclo) sólo puede ser una función coseno, con valores A y $-A$. Y algo similar ocurre con el armónico a 0 Hz, pues con $k = 0$ su parte imaginaria es cero, y la real representa N veces su amplitud.

La TDF inversa nos permite convertir los datos de una señal discreta representada en el dominio de la frecuencia (espectro) al dominio del tiempo (forma de onda).

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k)e^{+2\pi\, jkn/N} \qquad \text{para } n = 0, 1, 2, \ldots, N\text{-}1$$

Por otra parte, la Transformada Rápida de Fourier (FFT, por *Fast Fourier Transfom*) es un algoritmo de computación que implementa de manera rápida y eficaz la Transformada Discreta. Requiere que N sea potencia de 2.



Un ciclo de una señal periódica con 4 armónicos, y su espectro calculado con la FFT. Como es común en la práctica, sólo se representa la parte positiva.

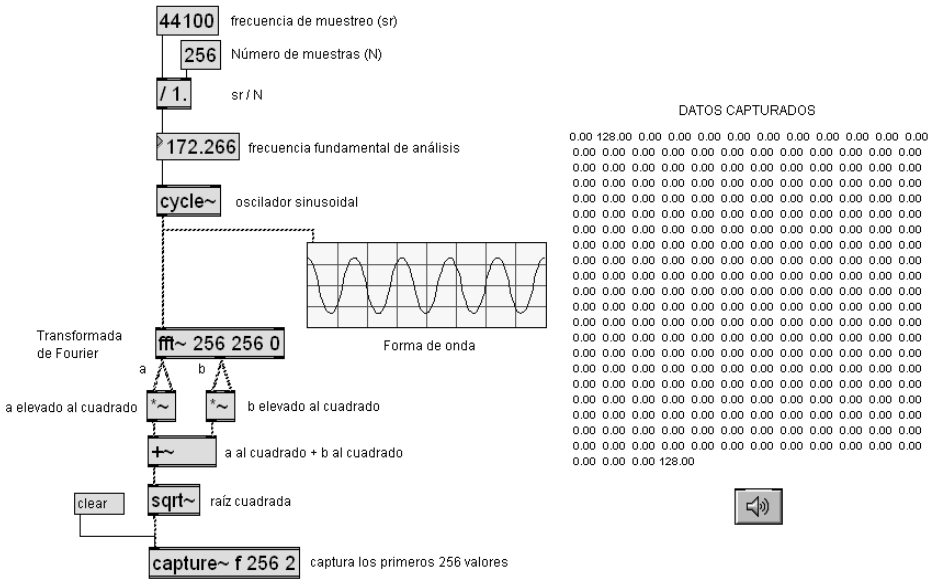
EJEMPLOS Y APLICACIONES DE LA TRANSFORMADA DISCRETA DE FOURIER

A fin de ilustrar los conceptos estudiados anteriormente, veremos algunos ejemplos sobre tratamiento digital de señales. Otras aplicaciones interesantes ocurren en la resíntesis de sonidos complejos, utilizando instrumentos acústicos convencionales; cada instrumento imita el comportamiento de los parciales de un sonido a emular. Esta técnica es común en el movimiento francés denominado *espectralismo*, que plantea en sus obras la resíntesis de espectros obtenidos de sonidos instrumentales grabados, o del cálculo de espectros a partir de técnicas de síntesis del sonido (síntesis aditiva, frecuencia modulada y amplitud modulada).

Para ello, utilizaremos el lenguaje de programación para el procesamiento de sonido y música en tiempo real Max-MSP[1], y el entorno de programación para composición asistida Open Music[2].

FFT

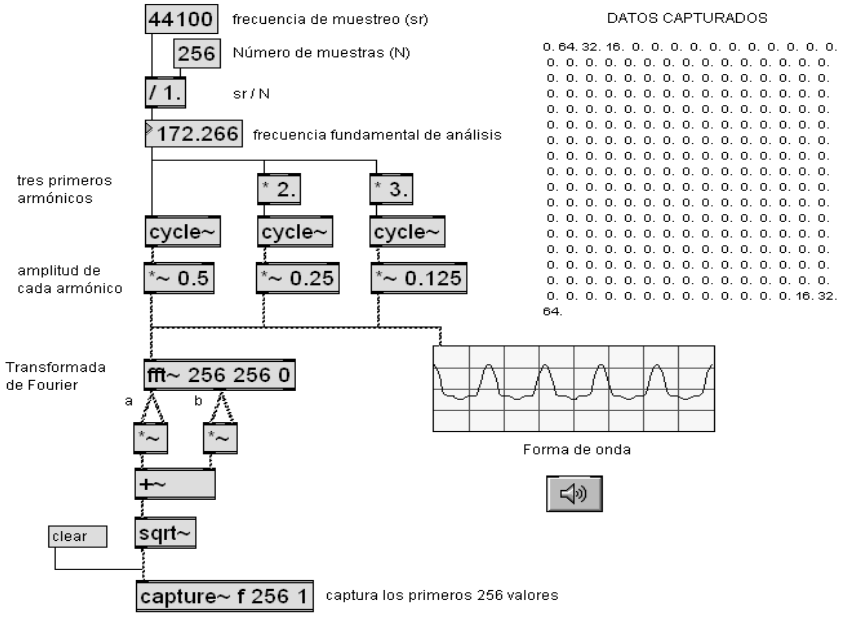
En el ejemplo 1, el objeto *cycle~* produce una sinusoide que oscila entre +1 y -1, y su frecuencia es calculada para que coincida exactamente con la frecuencia fundamental de análisis *R*/*N*. La frecuencia de muestreo es 44.100 Hz y el número de muestras de la ventana de análisis es 256, lo que da por resultado 172,266 Hz. Como se trata de una señal sinusoidal, la TDF sólo debería detectar energía para el primer armónico. Se observa en la tabla obtenida que el primer valor es 0, y corresponde a la componente de 0 Hz (k = 0). El segundo valor es 128 y corresponde a $N\frac{A}{2}$, para k = 1. Esto implica que la amplitud detectada es 0,5, o sea la mitad que corresponde a las frecuencias positivas. El último valor de la tabla también es 128 y representa a la frecuencia negativa de nuestra sinusoide (k = 255). Según se observa en el programa, los valores de *a* y *b* los obtuvimos con el objeto *fft~*, y la amplitud fue calculada con $\sqrt{a^2+b^2}$.



Ejemplo 1.

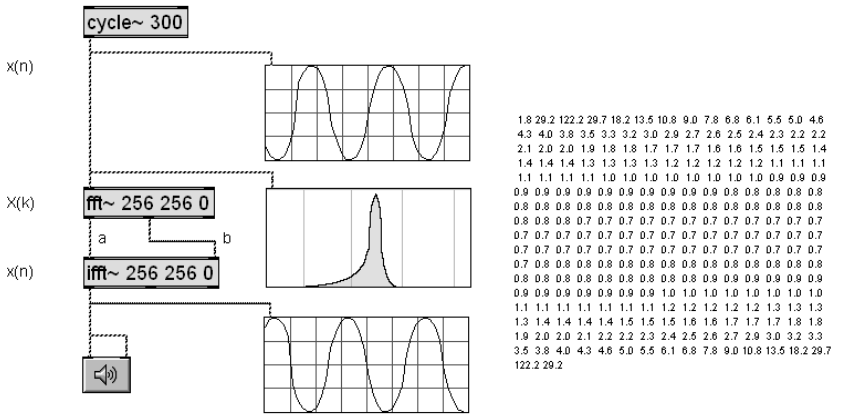
[1] http://www.cycling74.com
[2] http://recherche.ircam.fr/equipes/repmus/OpenMusic

El ejemplo 2 es similar al anterior, pero en este caso ingresamos una señal compuesta por los primeros tres armónicos, cuyas amplitudes son 0,5; 0,25 y 0,125 respectivamente. Se observan nuevamente los resultados en espejo, con valores 64, 32 y 16 (*N.A*/2).



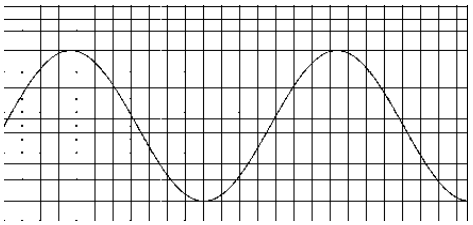
Ejemplo 2.

Los programas anteriores muestran el análisis de una señal cuya fundamental, o primer armónico, coincide exactamente con la frecuencia fundamental de análisis de la TDF, pero esto jamás ocurre en la práctica. Para ello, es preciso tener en cuenta que las muestras a analizar (*N*) siempre serán consideradas como un ciclo de una onda infinitamente periódica, aunque esto no sea cierto. Así como la onda está muestreada en el tiempo, el espectro está muestreado en frecuencia, y los valores de frecuencia a detectar serán siempre los armónicos de la frecuencia fundamental de análisis. En el primer ejemplo vimos que para $R = 44100$ y $N = 256$, su valor era 172,266 Hz. Pero qué ocurre si analizamos una señal sinusoidal de 300 Hz que no coincide en frecuencia con ningún armónico de esa fundamental de análisis. Dado que *N* no abarca exactamente un ciclo de la señal, la forma de onda a analizar ya no es más sinusoidal, y por lo tanto, posee armónicos, los más prominentes en las frecuencias más cercanas a los 300 Hz. La información irrelevante de los resultados suele denominarse *artefactos* de análisis, y un modo de minimizarla es a través de técnicas de modificación de la forma de la ventana.



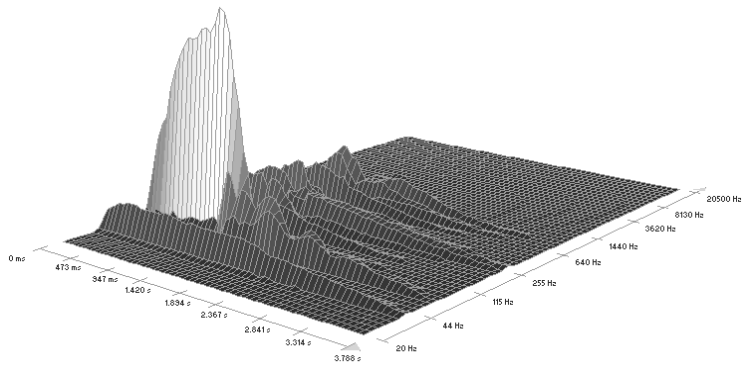
Ejemplo 3.

Según se observa en la lista de datos capturados, el pico más prominente aparece en el segundo armónico de la fundamental de análisis (344,56 Hz), que es la frecuencia más cercana a los 300 Hz. Su amplitud es 0,95 (A. N / 2 = 122,2).



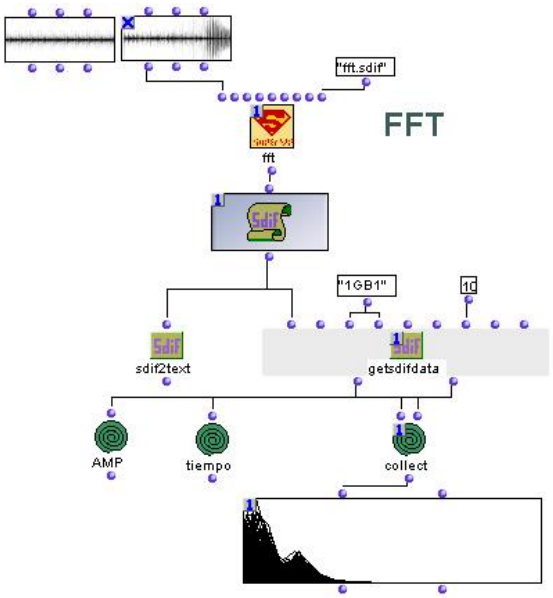
256 muestras de una sinusoide de 300 Hz, muestreada a 44100 Hz. Se observa que no representa un número entero de ciclos de una sinusoide, sino un ciclo de una forma compleja.

Un modo de aumentar la resolución en frecuencia es aumentar el tamaño de la ventana. Si R = 32000 Hz, y N = 32, la frecuencia fundamental de análisis es 1000 Hz y sus armónicos 2000, 3000, etc. Pero si aumentamos N a 3200 muestras, crece la resolución en frecuencia, pues se buscan componentes cada 10 Hz. En este caso, es mucho más probable que la frecuencia a detectar coincida con los armónicos de la frecuencia fundamental de análisis. Al mismo tiempo, ocurre que la mayoría de los sonidos que nos interesa analizar distan mucho de ser periódicos, son cambiantes en el tiempo. Como contrapartida, al aumentar la resolución en frecuencia disminuye la resolución temporal, esto significa que si el sonido analizado varía, perdemos información sobre sus estados intermedios al ir calculando cómo se transforma el espectro en el tiempo. Una solución a este problema es que la ventana no avance N muestras, sino que se superponga a la anterior. A modo de ejemplo, si N = 1024, podemos correr la ventana cada 256 muestras, mejorando así la resolución temporal.



Representación tridimensional de la sucesión de espectros de un sonido impulsivo.

En el entorno Open Music se realiza la FFT a través de un objeto de la librería SuperVP. La información obtenida puede ser representada de diversas formas. Los datos que aporta el objeto se almacenan en un archivo con el formato SDIF.



Ejemplo 4.

Síntesis Cruzada

La TDF nos permite realizar operaciones de transformación de las señales representadas en el dominio de la frecuencia, que no serían posibles en el dominio del tiempo. La aplicación que veremos a continuación se denomina *síntesis cruzada*, y consiste en multiplicar los valores de amplitud de las componentes de un sonido con los de otro. A modo de ejemplo, una señal constituida por golpes sobre un instrumento de percusión, puede articular la aparición de una música, que a la vez es filtrada por la envolvente espectral del instrumento percutido a causa de la multiplicación de los espectros. Para comprender mejor esta técnica es necesario tener en cuenta algunas consideraciones previas.

La TDF de dos formas de onda sumadas ($f(n)$ y $g(n)$) equivale a la suma de sus espectros.

$$DFT(f(n)+g(n)) = F(k)+G(k)$$

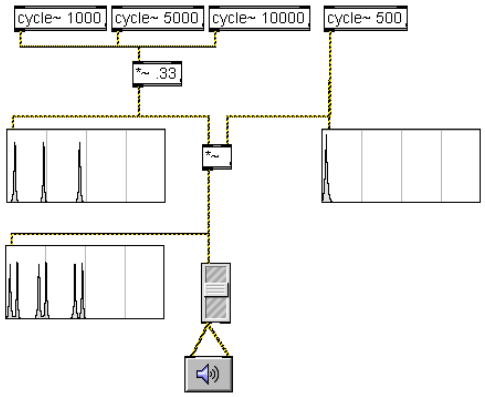
Del mismo modo, la TDF de una forma de onda multiplicada por una constante (a), equivale a la multiplicación de las muestras de amplitud del espectro por esa constante.

$$DFT(a.f(n)) = a.F(k)$$

En cambio, la TDF aplicada a la multiplicación de dos formas de onda, no corresponde a la multiplicación de sus respectivos espectros, sino a la *convolución* de los mismos.

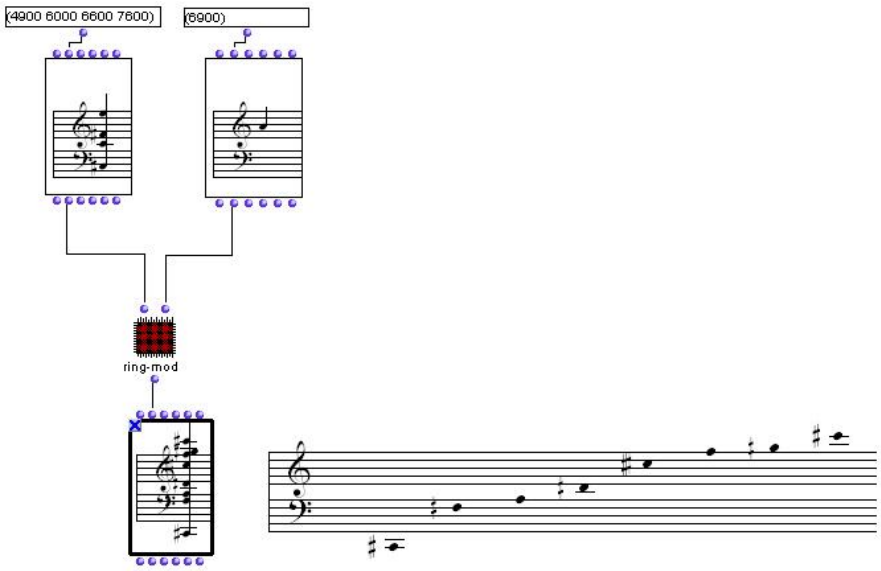
$$DFT(f(n).g(n)) = F(k)*G(k)$$

En el siguiente ejemplo, se observa la multiplicación de dos señales, una compleja (formada por componentes sinusoidales de 1000, 5000 y 10000 Hz) con otra simple de 500 Hz. Vemos que a nivel espectral se produce la suma y la resta de las frecuencias de cada una de las componentes de la señal compleja, con la sinusoide de 500 Hz (1000 + 500, 1000 – 500, 5000 + 500, 5000 – 500, 10000 + 500 y 10000 – 500). Al resultado de la convolución de espectros se lo suele denominar *modulación en anillo*, debido al nombre del circuito que realizaba esta operación en los sintetizadores analógicos, y según se ve en la figura, se obtiene multiplicando las muestras de amplitud de las dos formas de onda en los medios digitales.

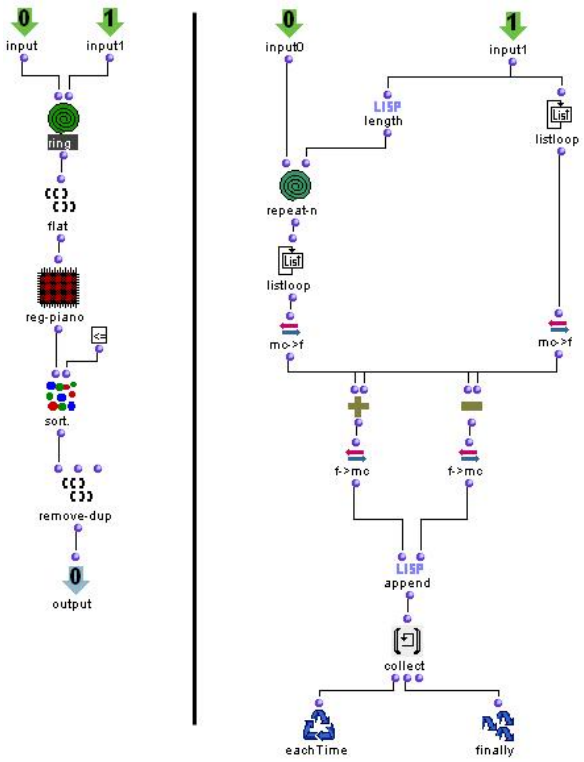


Ejemplo 5.

A las componentes resultantes de la suma y la resta de las componentes de las dos señales de entrada se las denomina también *sonidos de combinación* (adicionales y diferenciales). En el ejemplo que sigue, realizado en OM, se obtienen los sonidos de combinación de las notas de dos acordes. Cada nota de un acorde representa a una componente del "espectro". El programa realiza la suma y la resta de las frecuencias de las notas, y convierte los resultados de esas operaciones a las notas más cercanas (por 1/4 ó 1/8 de tono). Esta técnica es también común en el espectralismo.



Ejemplo 6.

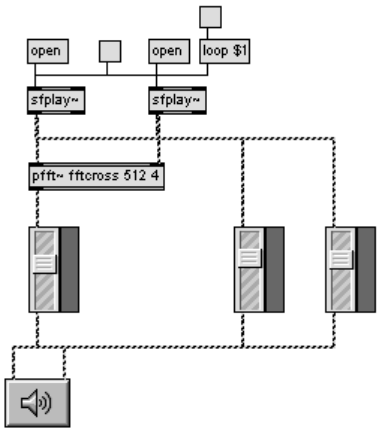


Subpatches del Ejemplo 6.

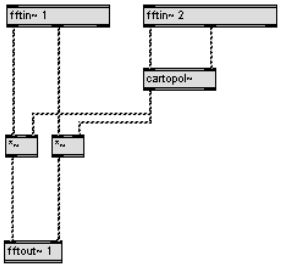
Finalmente, la multiplicación de dos espectros no equivale a la multiplicación de sus formas de onda, sino a la convolución de las mismas.

$$DFT(f(n) * g(n)) = F(k).G(k)$$

Al realizar la multiplicación de dos espectros se produce un escalamiento de las componentes de ambos. A modo de ejemplo, si una componente de frecuencia del primer espectro posee amplitud 0, anula a la del otro espectro, y viceversa.



Ejemplo 7.

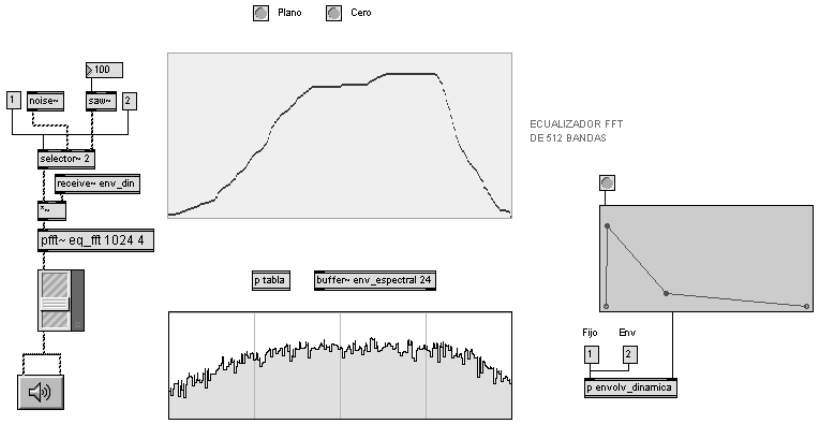


Subpatch *pfft~*

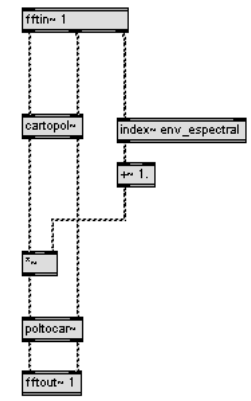
Los objetos *fftin~* realizan la FFT de cada una de las señales. El objeto *cartopol~* (cartesianas a polares) convierte los valores *a* y *b* de la segunda señal en valores de amplitud, mediante $\sqrt{a^2+b^2}$. Finalmente, esos valores de amplitud son multiplicados por *a* y *b* de la primera señal, y se realiza la transformada inversa con *fftout~* para convertir al dominio del tiempo.

Filtro FFT

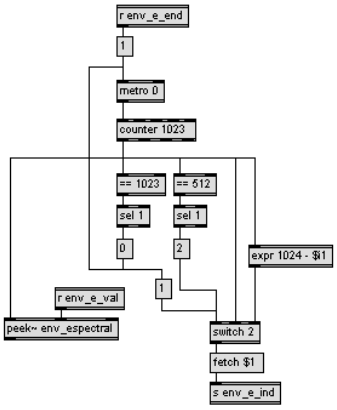
Otra aplicación posible es crear un filtro espectral a partir de una curva determinada, en nuestro caso, dibujada con el *mouse*. El objeto *multislider*, sobre el cual se describe la curva, presenta 512 puntos que pueden variar entre -1 y 1. El subpatch *tabla* duplica esa curva en espejo y almacena los valores en un *buffer*. Cada uno de los puntos de la función almacenada se multiplica por los 1024 (*N*) pares de valores *a* y *b* calculados por la FFT, modificando así la amplitud de cada componente múltiplo de la frecuencia fundamental de análisis de modo muy preciso.



Ejemplo 8.



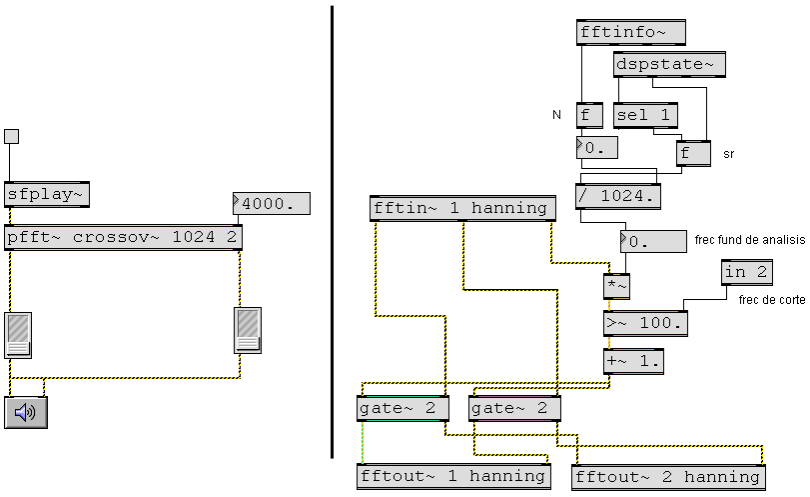
Subpatch *pfft~*



Subpatch *tabla*

Crossover FFT

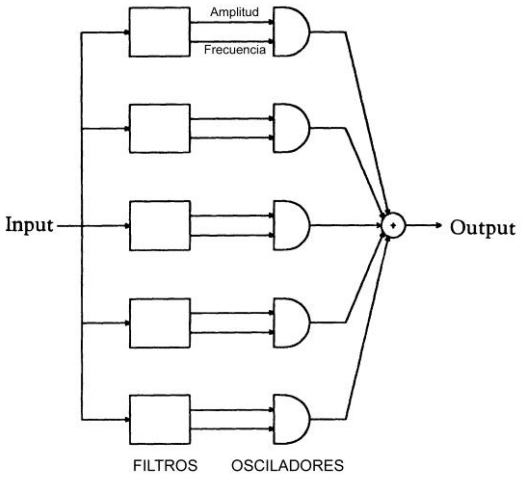
Otra aplicación posible consiste en separar las componentes espectrales en un número determinado de bandas, a fin de obtener una señal de audio para cada banda. Cada una de esas señales puede ser destinada a un canal de audio diferente, a fin de espacializarlas con trayectorias diferenciadas.



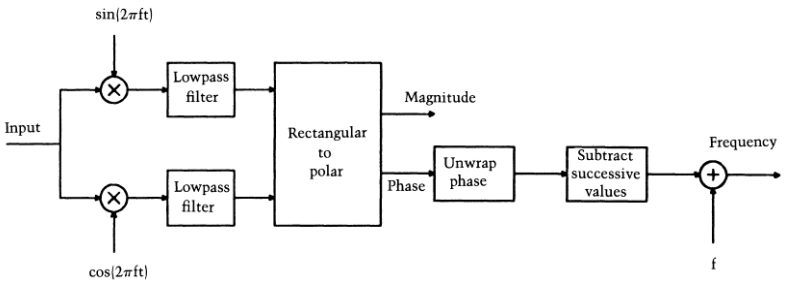
Ejemplo 9.

Phase Vocoder

El *phase vocoder* se basa en la TDF, y comprende dos etapas, una de análisis y otra de resíntesis. Entre una etapa y la otra, es posible modificar los datos obtenidos, a fin de realizar operaciones de cambio de altura sin afectar la duración del sonido, o cambio de duración sin afectar la altura. La etapa de resíntesis consta de un banco de osciladores que reproducen los valores de frecuencia y amplitud de cada componente, obtenidos durante el análisis.



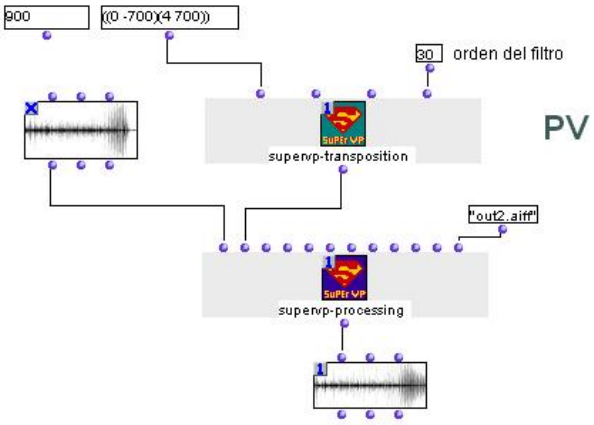
Análisis y resíntesis en un *phase vocoder*.



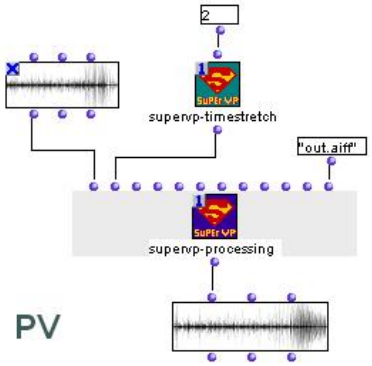
Filtro de un *phase vocoder*.

En cada filtro se realiza la multiplicación de la señal de entrada por un armónico de la frecuencia fundamental de análisis. Si esa componente está presente en la señal original, se produce un adicional (cercano o igual al doble de su frecuencia) y un diferencial (cercano o igual a 0 Hz). El adicional es eliminado mediante un filtro pasa-bajos. Luego se extrae la frecuencia y la amplitud, pasando de coordenadas cartesianas (a y b) a polares (A y ϕ). La diferencia de fase de cada componente, obtenida entre el análisis de un bloque de muestras y el siguiente, se suma al valor de frecuencia obtenido, a fin de corregir la diferencia que pueda existir entre la frecuencia original y la frecuencia del armónico de análisis.

En los siguientes ejemplos, se muestra la implementación de un *phase vocoder* en el entorno OM, tanto en transposición sin alterar la duración, como en modificación de la duración sin afectar la altura.



Ejemplo 10.



Ejemplo 11.

Convolución

Señales discretas

Sabemos que en la teoría de los sistemas discretos las señales son representadas mediante secuencias. A la secuencia completa *x*, en la cual el enésimo término se representa por *x(n)*, se la expresa así:

$$x(.) = \{x(n)\} \ , \quad -\infty < n < \infty$$

La secuencia unitaria o impulso $\delta(n)$, por su parte, se define como:

$$\delta(n) = \begin{array}{l} 0, \ n \neq 0 \\ 1, \ n = 0 \end{array}$$

Y la secuencia de escalón unitario $u(n)$:

$$u(n) = \begin{array}{l} 0, \ n < 0 \\ 1, \ n \geq 0 \end{array}$$

Ambas secuencias se relacionan a partir de:

$$u(n) = \sum_{k=-\infty}^{n} \delta(k)$$

$$\delta(n) = u(n) - u(n-1)$$

Definimos como *energía* de una señal a:

$$\varepsilon = \sum_{n=-\infty}^{\infty} |x(n)|^2$$

Las señales discretas son susceptibles de diversas transformaciones. Las operaciones de transformación básicas, como la suma, el producto y la multiplicación de la secuencia por un número se representan,

$$x(.) + y(.) = \{x(n) + y(n)\}$$

$$x(.).y(.) = \{x(n)y(n)\}$$

$$a.x(.) = \{ax(n)\}$$

y la señal retrasada en *k* muestras:

$$y(n) = x(n-k)$$

donde *k*, al igual que *n*, son enteros.

Una señal cualquiera puede expresarse como una suma de impulsos escalados:

$$a_{-3}\delta(\mathrm{n}+3)+a_{-2}\delta(\mathrm{n}+2)+a_{-1}\delta(\mathrm{n}+1)+a_0\delta(\mathrm{n})+a_1\delta(\mathrm{n}-1)+a_2\delta(\mathrm{n}-2)...$$

$\delta(0)=1$, por lo tanto, sólo cuando *n* vale -3, esa muestra tiene a_{-3} de amplitud. Esto puede generalizarse mediante:

$$x(n)=\sum_{k=-\infty}^{\infty}x(k).\delta(\mathrm{n}-\mathrm{k})$$

Por último, una transformación cualquiera puede anotarse:

$$y(n)=T[x(n)]$$

Sistemas lineales e invariantes

Los sistemas lineales cumplen con el principio de superposición

$$T[ax_1(n)+bx_2(n)]=aT[x_1(n)]+b.T[x_2(n)]=ay_1(n)+by_2(n)$$

a y *b* son constantes arbitrarias, las *x* son señales de entrada, y las *y* de salida.

Vimos que una secuencia puede ser representada como la suma secuencias unitarias escaladas y retardadas.

$$y(n)=\sum_{k=-\infty}^{\infty}x(k).\delta(\mathrm{n}-\mathrm{k})$$

Aplicando una transformación nos queda:

$$y(n)=T\left[\sum_{k=-\infty}^{\infty}x(k).\delta(\mathrm{n}-\mathrm{k})\right]$$

y si es lineal,

$$y(n)=\sum_{k=-\infty}^{\infty}x(k).T[\delta(\mathrm{n}-\mathrm{k})]=\sum_{k=-\infty}^{\infty}x(k).h_k(n)$$

siendo $h_k(n)$ la respuesta a $\delta(\mathrm{n}-\mathrm{k})$.

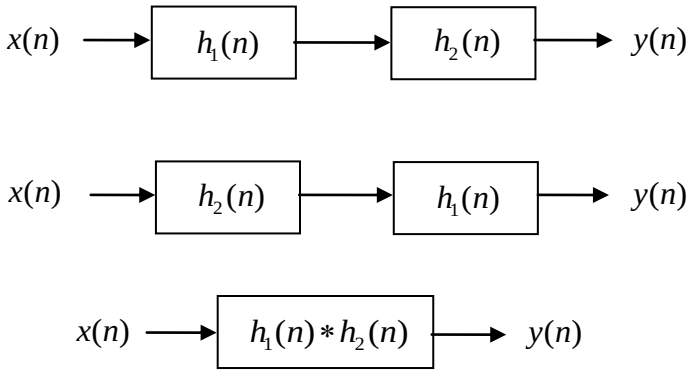
Por otra parte, si $y(n)$ es la repuesta de un sistema a la señal $x(n)$, y se cumple que $y(n-k)$ es la respuesta a $x(n-k)$, entonces decimos que el sistema es invariante en el tiempo. Del mismo modo, si $h(n)$ es la respuesta al impulso $\delta(n)$, entones la respuesta a $\delta(n-k)$ es $h(n-k)$. La ecuación anterior se convierte, entones en

$$y(n)=\sum_{k=-\infty}^{\infty}x(k).h(\mathrm{n}-\mathrm{k})$$

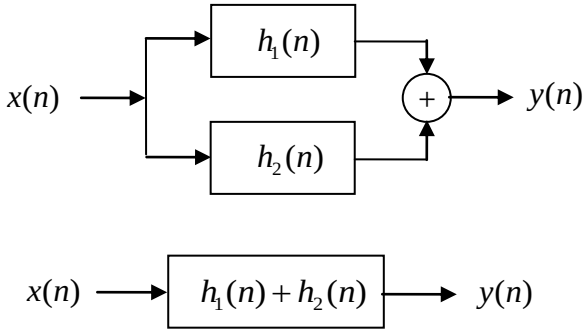
ecuación a la que se denomina de *convolución*.

$$y(n) = \sum_{k=-\infty}^{\infty} x(k).h(\mathrm{n}-\mathrm{k}) = h(n) * x(n)$$

Los siguientes sistemas son lineales e invariantes en el tiempo, y tienen idéntica repuesta al impulso (el paso por la caja significa convolución $x(n) * h(n)$)



Lo mismo ocurre con esta combinación en paralelo.



Resumiendo, si la transformación es la DFT:

$$DFT\left[f(n)+g(n)\right]=F(k)+G(k)$$

$$DFT\left[a.f(n)\right]=a.F(k)$$

Combinando estas propiedades podemos caracterizar a la DFT como una transformación lineal.

$$DFT\left[a.f(n)+b.g(n)\right]=a.F(k)+b.G(k)$$

Sin embargo,

$$DFT\left[f(n).g(n)\right]=F(k)*G(k)$$ (convolución de los espectros)

$$y(n) = x_1(n).x_2(n) = A_1 \cos(2\pi\ f_1 nT).A_2 \cos(2\pi\ f_2 nT)$$

El espectro puede calcularse a partir de:

$$\cos A \cos B = \frac{1}{2}\left[\cos(A+B) + \cos(A-B)\right]$$

De forma equivalente:

$$DFT\left[f(n) * g(n)\right] = F(k).G(k) \qquad \text{(convolución de secuencias)}$$

La convolución lineal o aperiódica de dos secuencias *f(n)* y *g(n)* se define por:

$$y(n) = f(n) * g(n) = \sum_{k=0}^{n} f(k)g(n-k)$$

siendo $Ny = Nf + Ng - 1$

La respuesta a impulso tiene ceros a la derecha y a la izquierda, caso contrario es circular o periódica.

Ejemplo:

$f(n) = \{1,2,1,3\}$ $g(n) = \{4,5\}$ *Ny* = 4 + 2 – 1 = 5

	***n*=0**	***n*=1**	***n*=2**	***n*=3**	***n*=4**
***k*=0** **(*n* - *k*)=**	1x4 0	1x5 1	1x0 2	1x0 3	1x0 4
***k*=1** **(*n* - *k*)=**	2x0 -1	2x4 0	2x5 1	2x0 2	2x0 3
***k*=2** **(*n* - *k*)=**	1x0 -2	1x0 -1	1x4 0	1x5 1	1x0 2
***k*=3** **(*n* - *k*)=**	3x0 -3	3x0 -2	3x0 -1	3x4 0	3x5 1
	4	5			
		8	10		
			4	5	
				12	15
y(n)	**4**	**13**	**14**	**17**	**15**

Propiedades:

$f(n) * \delta(n) = f(n)$ $\delta(n)$ = función de muestra unitaria

$f(n) * \delta(n-d) = f(n-d)$

$$f(n) * a.\delta(n) = a.f(n)$$

$$f(n) * a.\delta(n-d) = a.f(n-d)$$

Ejemplo *n-d*:
n = 0, 1, 2, 3, 4, 5
d = 4
n-d = -4, -3, -2, -1, 0, 1
$\delta(n-d)$ = 0 0 0 0 1 0, pues $\delta(n)$ tiene ceros a la izquierda

Ventanas

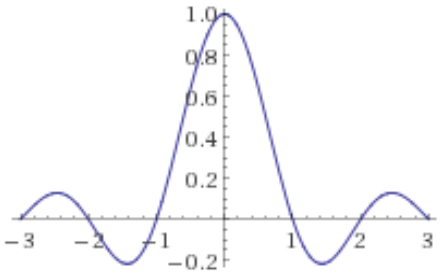
Los artefactos propios de analizar una señal cuyas componentes no son múltiplos de la frecuencia fundamental de análisis (R/N), o bien de R, pueden ser minimizados mediante el uso de diversas funciones de ventana. Hasta ahora, y sin habernos detenido en ello, al aplicar la DFT a una señal cualquiera, utilizamos una ventana rectangular de amplitud unitaria.

La DFT de una ventana rectangular es la función denominada sinc x (seno cardinal de x, también llamada función de muestreo).

$$\text{sinc}\,(x) = \frac{\sin x}{x} \text{ ó } 1 \text{ si } x = 0$$

En procesamiento de señales se emplea la función seno cardinal normalizada, que se expresa:

$$\text{sinc}\,(x) = \frac{\sin \pi x}{\pi x}$$

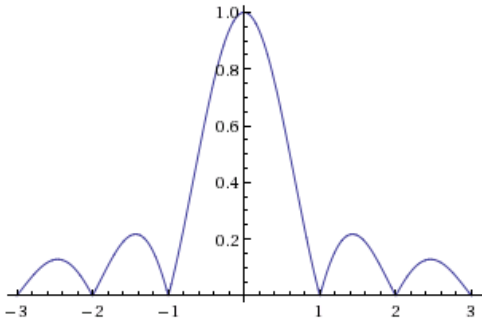


Multiplicar la señal a analizar por la ventana rectangular equivale a convolucionar la función sinc(x) con cada armónico de la frecuencia fundamental de análisis. A cada armónico se le superpone una copia centrada de sinc(x), cuyos cruces por cero coinciden con los armónicos vecinos.

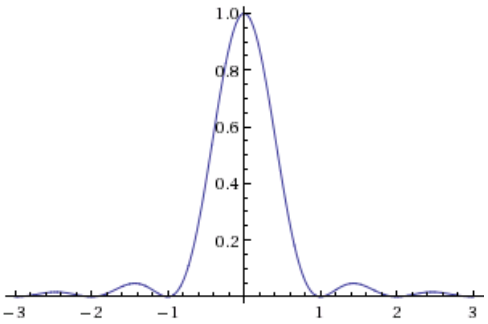
La DFT puede entonces interpretarse como un banco de filtros donde la respuesta de cada filtro está dada por la función *sinc*. Pongamos como ejemplo que la frecuencia de una componente no es múltiplo de la fundamental de análisis, sino 1.5 veces ella, su amplitud entonces estará determinada por:

$$\text{sinc}\ (x) = 20.\log\left|\frac{\sin(1.5\pi)}{1.5\pi}\right| = -13.46\ dB$$

Podemos considerar el valor absoluto de sinc(x) para observar mejor cómo se verá afectada la magnitud de las componentes que no sean múltiplos exactos de la fundamental de análisis, pero sin olvidar que los valores negativos producen un cambio de fase en la resultante.



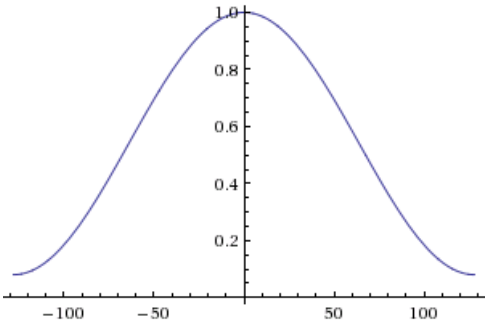
Es posible diseñar otras ventanas diferentes a la rectangular, que generen lóbulos menos pronunciados, controlando así la amplitud de los artefactos de análisis. La función $\text{sinc}^2(x)$, por ejemplo, es el resultado de aplicar una ventana de forma triangular.

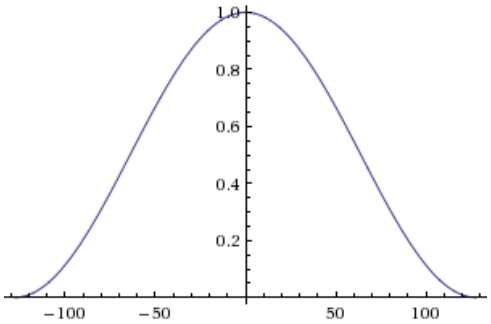


Otro tipo ampliamente usado es la ventana Hamming generalizada:

$$w_H(n) = \alpha + (1-\alpha)\cos\frac{2\pi n}{N} \quad \text{con} \quad -\frac{N-1}{2} \le n \le \frac{N-1}{2}$$

Si a vale 0.54, la ventana se denomina *Hamming*, y si vale 0.5, se denomina *hanning*. El gráfico de la izquierda muestra una ventana *Hamming* para un $N = 256$, y la de la derecha la *hanning* para igual N. Observar que la primera llega a tocar el eje horizontal.





Agradecimientos

A la Prof. Constanza Galdo por sus valiosos aportes.

Bibliografía

Chowning, J. M. "The synthesis of complex audio spectra by means of frequency modulation". *Journal of the Audio Engineering Society*, 21(7):526-534. Reprinted in Curtis Roads and John Strawn, eds. *Foundations of Computer Music*, Cambridge, MA: MIT Press, 1985.

Moore, F. Richard. *Elements of Computer Music*. MIT Press. Massachussets. 1991.

Moore, F. R. "An introduction to the mathematics of digital signal processing, Part I". *Computer Music Journal*, 2(1):38-47. Reprinted in John Strawn, ed. *Digital Signal Processing: An Anthology*, Los Altos, CA: William Kaufmann, Inc. 1985.

Moore, F. R. "An introduction to the mathematics of digital signal processing, Part II". *Computer Music Journal*, 2(2):38-60. Reprinted in John Strawn, ed. *Digital Signal Processing: An Anthology*, Los Altos, CA: William Kaufmann, Inc. 1985

Smih, Julius."Introduction to complex numbers". Adapted from Music 320 Course Reader. CCRMA. Stanford University. 1999.

Smith, Julius O. *Introduction to Digital Filters, September 2005 Draft*, Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, Web published at http://ccrma.stanford.edu/~jos/filters05/.

Smith, Julius. *Mathematics of the Discrete Fourier Transform (DFT) with Audio Applications, Second Edition*. CCRMA online book. http://ccrma.stanford.edu/~jos/mdft/mdft.html