

LA TRANSFORMADA DE FOURIER. UNA VISIÓN PEDAGÓGICA

[1] Jesús Bobadilla, [2] Pedro Gómez y [1] Jesús Bernal

[1]

Departamento de Informática Aplicada
Escuela Universitaria de Informática
Ctra. de Valencia Km. 7, 28031 Madrid
Tel: +34.91.3367862, Fax: +34.91.3367527
e-mail: jbobi@eui.upm.es, jbernal@eui.upm.es

[2]

Departamento de Arquitectura y Tecnología de Sistemas Informáticos
Universidad Politécnica de Madrid, Campus de Montegancedo, s/n,
Boadilla del Monte, 28660 Madrid
Tel: +34.91.3367384, Fax: +34.91.3367412,
e-mail: pedro@pino.datsi.fi.upm.es

RESUMEN

En este artículo se presenta la Transformada de Fourier (TF) como un método para traspasar las señales en el dominio del tiempo al dominio de la frecuencia. De esta manera, las señales de voz pueden ser estudiadas en la forma habitual que presentan los sonogramas. El enfoque proporcionado en el artículo es claramente didáctico, con la intención de que aquellas personas no familiarizadas con los formalismos matemáticos puedan asimilar los conceptos fundamentales subyacentes en la TF. Las explicaciones se apoyan en gráficos variados y la utilización de un ejemplo práctico desarrollado de forma ordenada y detallada. Así mismo, al final del artículo se proporciona una demostración matemática que puede servir de ayuda a las personas interesadas en el formalismo de la TF.

ABSTRACT

This article presents the Fourier Transform (FT) as a method used to transfer the signals from the time domain to the frequency domain. In this way, speech signals can be studied using spectra graphics. The article uses a didactic point of view, with the intention of explaining the FT main concepts to the persons that are not familiarized with the necessary mathematical formalisms. Explanations are supported by a variety of graphics and the use of a practical example, developed in an ordered and detailed way. Likewise, a mathematical proof is provided at the end of the article.

1. INTRODUCCIÓN

En este artículo se pretende explicar de una forma sencilla y pedagógica los principios básicos en los que se basan la mayor parte de los espectrógrafos y aplicaciones informáticas para obtener los sonogramas de voz. Los sonogramas se consiguen traspasando las señales de voz desde el dominio del tiempo al dominio de la frecuencia, y de esta manera, emular la función principal que realiza el caracol en el oído interno.

La comprensión y utilización de métodos y algoritmos de traspaso de la señal en el dominio del tiempo al dominio de la frecuencia, resulta fundamental para obtener los parámetros básicos a partir de los cuales se podrán realizar caracterizaciones espectrales de los sonidos, reconocimiento del habla, etc.

Haciendo uso de las explicaciones y detalles proporcionados en el artículo, un lector con los conocimientos necesarios de informática podría realizar una aplicación sencilla que le permitiera obtener espectrogramas del habla, y de esta manera, realizar las ampliaciones y modificaciones necesarias para llevar a cabo pruebas empíricas con las que desarrollar estudios de fonética experimental.

El método escogido en este artículo para explicar el traspaso de las señales de voz desde el dominio del tiempo al de la frecuencia es la Transformada de Fourier (TF) [FOL92], [COC67], [RAB78], [RAB93]. Esta elección ha sido tomada debido a:

1. Su implantación en espectrógrafos y algoritmos de procesamiento de voz.
2. Su universalidad (utilización en diferentes campos de la ciencia).
3. El enfoque didáctico con el que la TF puede ser explicada.

Existen diversos métodos alternativos a la TF, como por ejemplo el de Predicción Lineal (LPC) [MAK75], [CAN74], [PAR86]. Cada uno de estos métodos presenta sus peculiaridades

propias, y resulta especialmente adecuado para la caracterización espectral de diversos parámetros de la voz

La TF ha sido abordada desde el formulismo de la señal discreta, más cercano a su utilización en métodos y algoritmos computables, con su vertiente práctica de creación de herramientas y aplicaciones en el campo tratado.

Este artículo ha sido estructurado cubriendo las siguientes secciones:

1. Presentación gráfica que muestra como una señal compleja se puede descomponer en un conjunto de señales simples.
2. Explicación de los conceptos fundamentales en los que se basa la TF, empleando enfoques que permiten realzar las ideas intuitivas que facilitan la comprensión del problema.
3. Desarrollo de ejemplos prácticos, explicados con detalle para afianzar los conceptos básicos explicados en el apartado anterior.
4. Formalización matemática de la TF mediante el uso de espacios de Hilbert. Este apartado ha sido incluido para dotar de rigor y completitud al artículo, así como para proporcionar un formalismo matemático elegante de la TF a aquellas personas especialmente interesadas en la vertiente teórica de este tema.

2. ANÁLISIS EN EL DOMINIO DE LA FRECUENCIA

Antes de comenzar describiendo ningún método concreto, resulta adecuado incidir en la idea básica subyacente en estos procesos, esto es, la descomposición de una señal compleja en sumatorio de señales simples. El oído humano, por medio del caracol, descompone las señales auditivas que le llegan en sus frecuencias fundamentales [MAR94], y ésta es la información básica a partir de la cual se elaboran las señales que le llegan al cerebro. Por tanto

podemos afirmar que el proceso de audición se fundamenta en la descomposición en frecuencias de la señal sonora. Para entender este concepto mejor nos basaremos en las tres figuras siguientes.

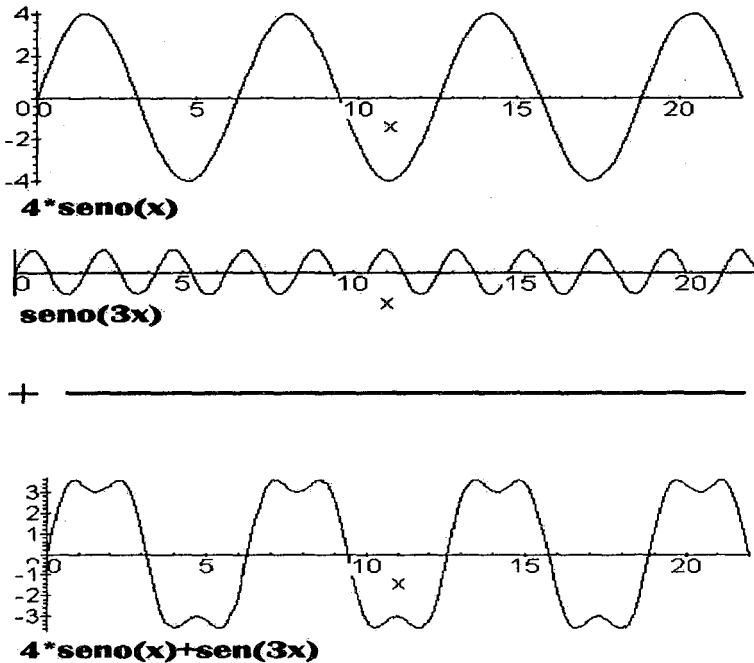
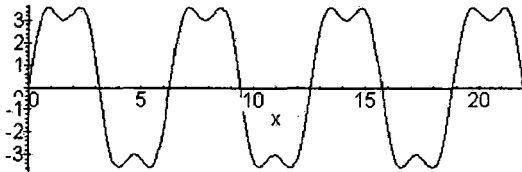


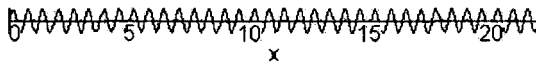
Figura 1
Ejemplo de descomposición de una señal compleja en sumatorio de señales simples.

La figura 1 nos muestra como la señal compleja representada en la parte inferior se puede descomponer en las dos señales simples de la parte superior, o desde otro punto de vista, la señal inferior puede ser creada sumando las dos funciones sinusoidales superiores. En este ejemplo, se podría reconocer a simple vista las frecuencias y amplitudes que dan lugar a la señal compleja, sin embargo,

complicando la señal un poco más, (en la figura 2) es difícil determinar la descomposición en senos (o cosenos) de la función final.

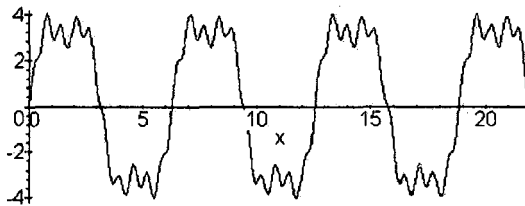


$$4 \cdot \text{seno}(x) + \text{seno}(3x)$$



$$0.5 \cdot \text{seno}(10x)$$

+



$$4 \cdot \text{seno}(x) + \text{seno}(3x) + 0.5 \cdot \text{seno}(10x)$$

Figura 2

Ejemplo de descomposición de una señal compleja en sumatorio de señales simples (continuación).

En la figura 3 se presenta una señal periódica que se puede descomponer en 5 señales sinusoidales de diferentes frecuencias y amplitudes. Llegados a este punto conviene remarcar que cuando analizamos una señal de voz en frecuencias, la información que consideramos relevante son las frecuencias de las señales simples con mayor amplitud (mayor peso en la determinación de la señal de voz) que se obtienen de la descomposición de la señal de voz original.

En las siguientes explicaciones de la TF conviene recordar los conceptos expuestos, puesto que el fundamento en el que se basa la TF estriba en la comparación de diversas ondas sinusoidales y cosenoidales simples con la señal de voz, cuanto más coincida una onda simple con la señal compleja, mayor importancia tiene su frecuencia en la determinación de la señal original.

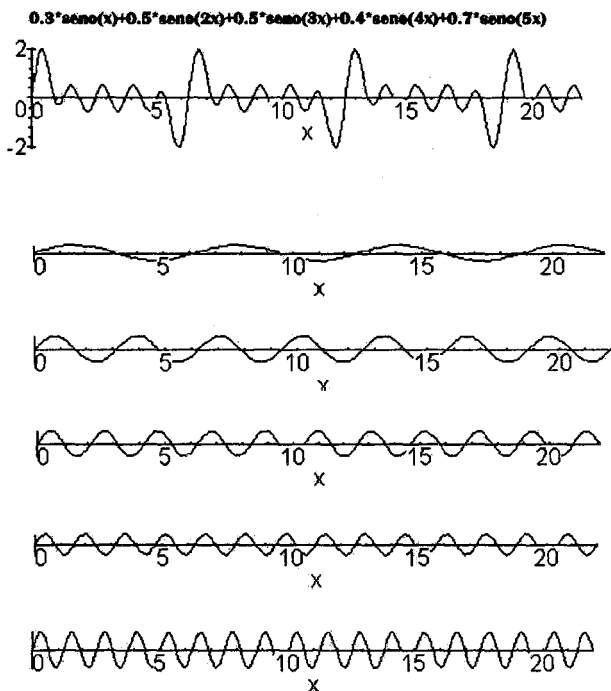


Figura 3
Ejemplo de descomposición de una señal compleja en sumatorio de señales simples (continuación).

3. TRANSFORMADA DE FOURIER. CONCEPTOS BÁSICOS

Para explicar el funcionamiento de la transformada de Fourier, partiremos de su formulación básica y explicaremos los conceptos fundamentales.

$$F\left(\frac{n}{NT}\right) = \frac{1}{N} \sum_{k=0}^{N-1} m(kT) e^{-j\frac{2\pi nk}{N}} \quad n = 0, 1, \dots, *N-1$$

FORMULACIÓN DE LA TRANSFORMADA DE FOURIER

Donde 'N' es el número de muestras de la ventana que se va a analizar, 'T' es el periodo de muestreo (inverso a la frecuencia de muestreo que denominaremos 'f'), 'n' es el índice de la frecuencia cuyo valor queremos obtener y 'm(kT)' indica la muestra tomada en el instante 'kT' (muestra Késima) de la ventana. Algunos de estos valores se ilustran en la figura 4.

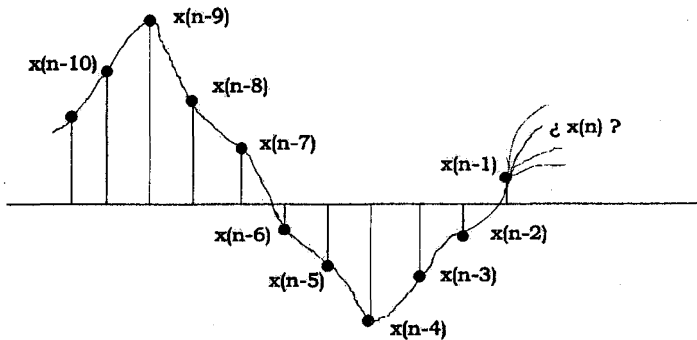


Figura 4

Señal muestreada, con indicaciones del significado de 'm(k)', 'T' y 'kT'.

El valor del parámetro 'n' determina la frecuencia concreta que se va a analizar, es decir, representa una de las frecuencias en las que se va a tratar de descomponer la señal de partida, de esta manera, para hacer el estudio con todas las frecuencias, usamos el rango completo de variación de 'n': $n=0, 1, 2, \dots, N-1$.

Desarrollando la fórmula de la transformada de Fourier para los distintos valores de 'n', tenemos:

$$n = 0 \Rightarrow F(0) = \frac{1}{N} \sum_{k=0}^{N-1} m(kT) * e^0$$

$$n = 1 \Rightarrow F\left(\frac{1}{N} f\right) = \frac{1}{N} \sum_{k=0}^{N-1} m(kT) * e^{-j2\pi \frac{k}{N}}$$

$$n = 2 \Rightarrow F\left(\frac{2}{N} f\right) = \frac{1}{N} \sum_{k=0}^{N-1} m(kT) * e^{-j4\pi \frac{k}{N}}$$

.....

$$n = N-1 \Rightarrow F\left(\frac{N-1}{N} f\right) = \frac{1}{N} \sum_{k=0}^{N-1} m(kT) * e^{-j2(N-1)\pi \frac{k}{N}}$$

Con las ecuaciones anteriores se obtienen las siguientes ideas:

■ La porción de la señal que se analiza se encuentra en el bloque de *muestras* $m(0), m(1), m(2), \dots, m(N-1)$, debido al sumatorio $\sum_{k=0}^{N-1} m(kT) * \dots$

■ La frecuencia $f=0\text{Hz}$. (correspondiente a $n=0$) se halla haciendo la media aritmética de los valores de las muestras, por lo que representa la componente continua de la señal.

■ El parámetro 'n' actúa de índice para obtener las distintas frecuencias de estudio, por ello nos encontramos con la secuencia:

$F\left(\frac{0}{N}f\right), F\left(\frac{1}{N}f\right), F\left(\frac{2}{N}f\right), \dots, F\left(\frac{N-1}{N}f\right)$, donde n/N es una proporción (en este caso de ' f ').

■ Los valores que se obtienen para $0 \leq n < N/2$ coinciden con los obtenidos en el intervalo $N/2 \leq n < N-1$, (con ' n ' par) por lo que es suficiente realizar los cálculos en una de las dos mitades. Según el criterio de Nyquist, el ancho de banda de la señal coincide con la mitad de la frecuencia de muestreo ' f ', correspondiente a $n=0, 1, 2, \dots, (N/2)-1$.

■ Los valores $n=1 \Rightarrow 2\pi, n=2 \Rightarrow 4\pi, \dots, n=N-1 \Rightarrow 2(N-1)\pi$, indican las frecuencias de las señales sinusoidales y cosenoidales con las que se comparará la señal original, este concepto se explicará a lo largo del apartado.

■ Si aumentamos el valor de ' N ', conseguimos hacer el análisis con un mayor número de frecuencias ($0 \leq n \leq N-1$), pero a costa de un mayor tiempo para calcular las operaciones del sumatorio ($0 \leq k \leq N-1$).

A modo de ejemplo, realizaremos el mismo desarrollo suponiendo una señal de voz muestreada a 10000 muestras/sg. y un bloque de 100 datos ($N=100$). Esto implica que se va a realizar el análisis de 10 ms. de tiempo.

$$n = 0 \quad \Rightarrow$$

$$F\left(\frac{0}{100}10000\right) = F(0\text{Hz.}) = \frac{1}{100} \sum_{k=0}^{99} m(kT) * e^0$$

$$n = 1 \quad \Rightarrow$$

$$F\left(\frac{1}{100}10000\right) = F(100\text{Hz.}) = \frac{1}{100} \sum_{k=0}^{99} m(kT) * e^{-j2\pi \frac{k}{100}}$$

$$n = 2 \Rightarrow$$

$$F\left(\frac{2}{100}10000\right) = F(200\text{Hz.}) = \frac{1}{100} \sum_{k=0}^{99} m(kT) * e^{-j4\pi \frac{k}{100}}$$

.....

$$n = 49 \Rightarrow$$

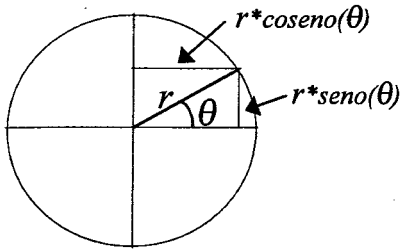
$$F\left(\frac{49}{100}10000\right) = f(4900\text{Hz.}) = \frac{1}{100} \sum_{k=0}^{99} m(kT) * e^{-j98\pi \frac{k}{100}}$$

Como se puede observar, el análisis se realiza sobre un ancho de banda de 5KHz. con una resolución espectral de 100 Hz.

Una vez explicados los conceptos básicos relativos a anchos de banda, resolución espectral, etc. del análisis, vamos a pasar a detallar el significado y funcionamiento del sumatorio de la fórmula de la transformada de Fourier.

Con un 'n' (frecuencia) y 'N' fijados, el sumatorio: $\sum_{k=0}^{N-1} m(kT) * e^{-j2\pi n \frac{k}{N}}$ depende únicamente del parámetro 'k'. En este sumatorio la operación primordial es la multiplicación de cada muestra $m(kT)$ del bloque de datos por el valor exponencial. Esta multiplicación, como veremos más tarde, hace las funciones de comparador entre ambos operandos ($m(kT)$ y el valor exponencial).

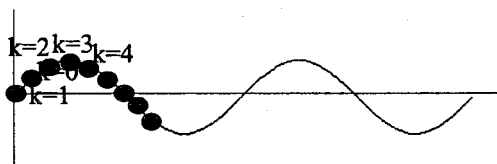
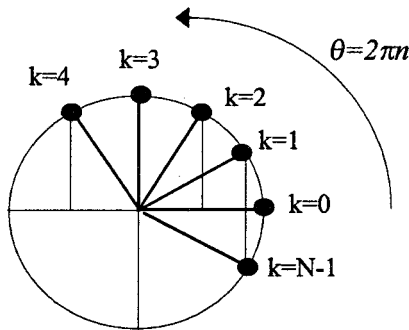
La expresión exponencial representa un número complejo en coordenadas polares ($r * e^{j\theta}$), donde 'r' es el módulo y 'θ' el argumento. Gráficamente:



$$r * e^{j\theta} = r(\cos\theta + j * \text{sen}\theta)$$

Por lo que: $e^{-j2\pi n \frac{k}{N}} = \cos\left(-2\pi n \frac{k}{N}\right) + j * \text{sen}\left(-2\pi n \frac{k}{N}\right)$

Puesto que el argumento del valor complejo de nuestra expresión depende de 'k', esto significa que nos encontramos ante un complejo que va girando en el círculo unidad con una velocidad angular $\theta=2\pi n$ y con saltos discretos de k/N radianes del recorrido total. La representación gráfica de este concepto se expresa a continuación:



La rotación del valor complejo en el círculo unidad se puede representar mediante sus correspondientes valores de senos y cosenos, de manera que el sumatorio:

$$\sum_{k=0}^{N-1} m(kT) * \left\{ \cos\left(-2\pi n \frac{k}{N}\right) + j * \text{sen}\left(-2\pi n \frac{k}{N}\right) \right\} \quad \text{se convierte}$$

en un recorrido en 'k' de cada muestra del bloque por cada valor de la función seno (y coseno) correspondiente $\{m(0)*\text{sen}(0)+m(1)*\text{sen}(-2\pi n/N)+m(2)*\text{sen}(-4\pi n/N)+\dots\}$ y análogamente con los cosenos.

Esta idea se representa en la figura 5, en la que aparecen tres senos con frecuencias 2π ($n=1$), 4π ($n=2$) y 6π ($n=3$). Estos senos se comparan con un bloque (de tamaño 'N') de la señal que se pretende descomponer. A simple vista se adivina que el seno que mejor encaja es el segundo (4π).

La razón por la que el sumatorio actúa de comparador (y este razonamiento es clave para entender el funcionamiento de la transformada de Fourier), es que cuando el bloque de la señal analizada es parecido al seno (o coseno) por el que se multiplican las muestras, el valor final del sumatorio será alto (alejado de cero), esto es así porque los valores positivos del seno (o coseno) se multiplicarán por valores positivos de la señal, y los valores negativos por los negativos de la señal. En el caso de que la señal no se parezca al seno (o coseno), los valores positivos y negativos se irán contrarrestando, y el resultado del sumatorio se aproximará más a cero.

Las figuras 6 y 7 muestran el resultado de comparar una señal compleja (función superior de la figura 6) por señales sinusoidales de diferentes frecuencias. Como se puede apreciar, cuando la frecuencia del seno forma parte de la composición de la señal original (en nuestro

caso $\text{seno}(x)$, $\text{seno}(3x)$ y $\text{seno}(10x)$), el resultado del sumatorio (área de la función) se hace bastante mayor que cero, mientras que si la frecuencia del seno no forma parte de la señal original (el resto de los casos), el valor del sumatorio se aproxima a cero.

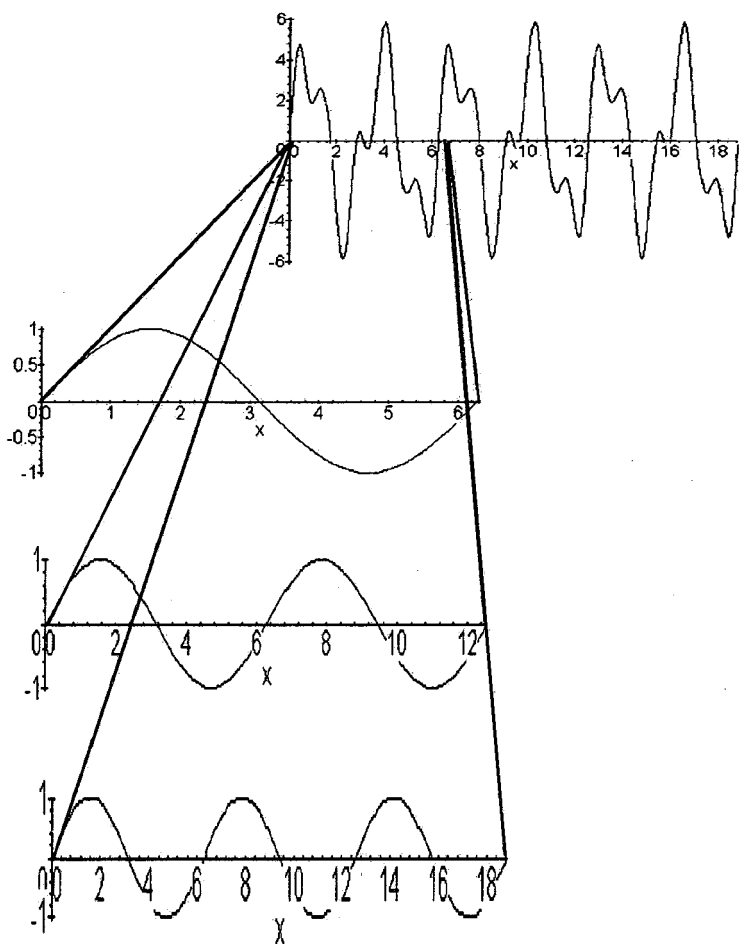


Figura 5

Comparación de funciones seno de diferentes frecuencias con un bloque de la señal que se pretende descomponer.

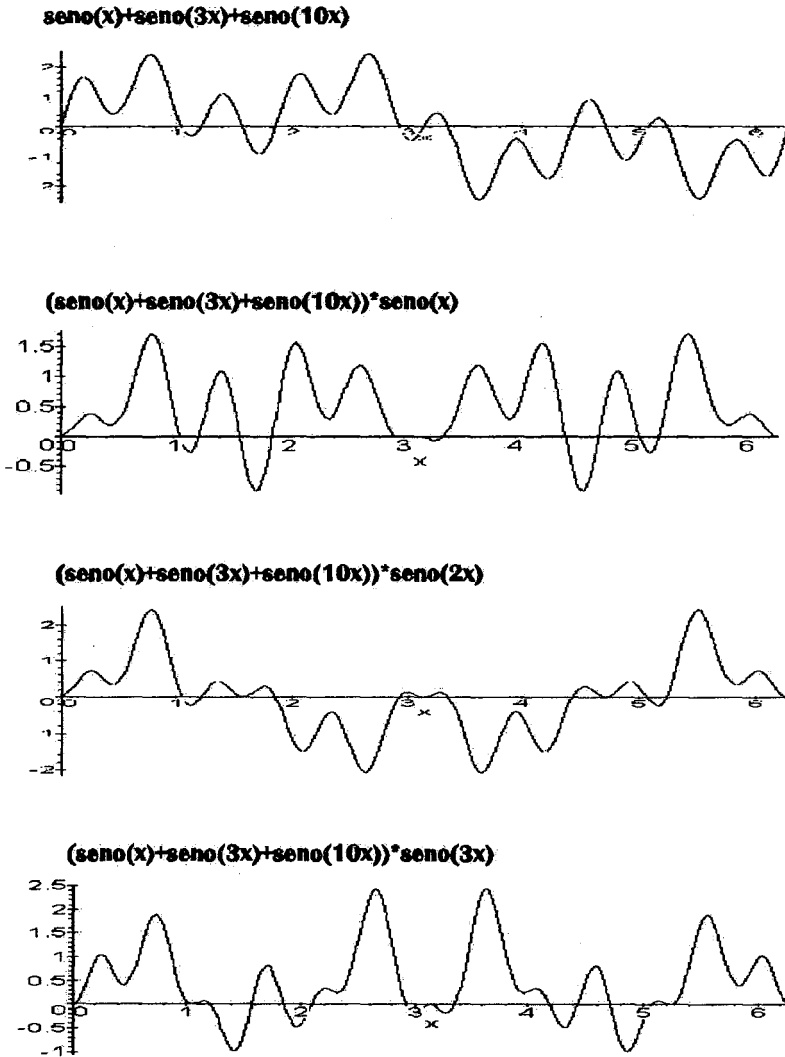


Figura 6
 Señal que se pretende descomponer y resultado de compararla con senos de diferentes frecuencias.

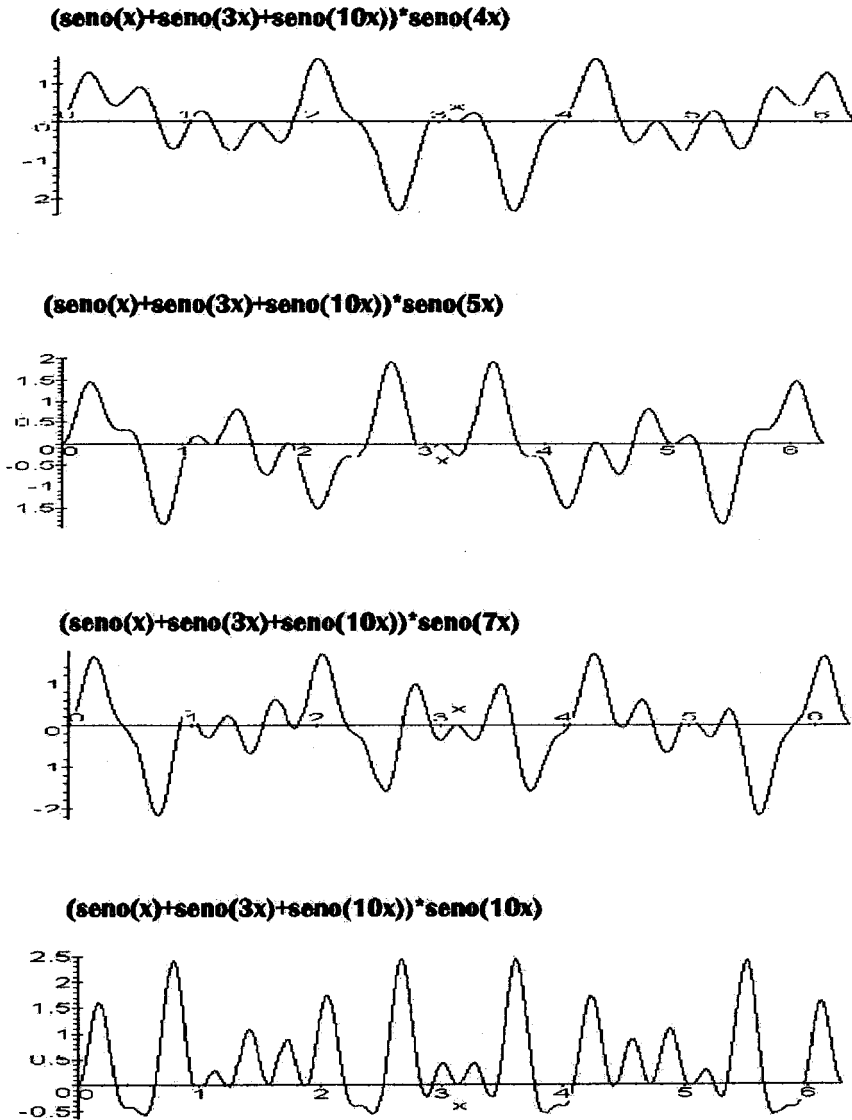
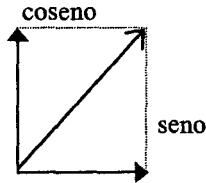


Figura 7
Señal que se pretende descomponer y resultado de compararla con senos de diferentes frecuencias (continuación).

El resultado del sumatorio (para un 'n' fijado) es un número complejo que indica la similitud de la señal analizada con el seno y coseno de la frecuencia dependiente de 'n'. La parte real del complejo representa la semejanza entre la señal y el coseno, la parte imaginaria se refiere al seno.



Obviamente un valor alto en el seno o en el coseno indica que la señal original se puede descomponer en un conjunto de señales simples entre las que se encontrará una de esta frecuencia. Para ponderar la importancia de ambos valores, habitualmente se utiliza la distancia euclídea:

$$\text{Modulo} = \sqrt{\text{Real}^2 + \text{Imag}^2}$$

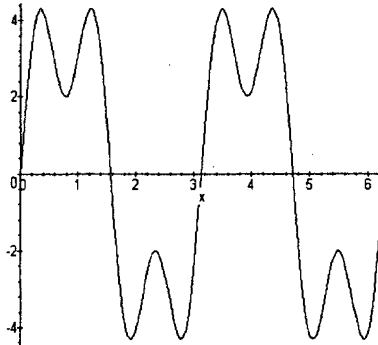
4. EJEMPLOS COMENTADOS

Como ejemplo de los conceptos explicados, vamos a realizar la descomposición en frecuencias de la señal discreta definida por las muestras:

$m(0) = 0$	$m(1) = 4.25$	$m(2) = 2.63$	$m(3) = 2.62$
$m(4) = 4.25$	$m(5) = 0$	$m(6) = -4.25$	$m(7) = -2.63$
$m(8) = -2.62$	$m(9) = -4.25$	$m(10) = 0$	$m(11) = 4.25$

$$\begin{aligned} m(12) &= 2.63 & m(13) &= 2.62 & m(14) &= 4.25 \\ m(15) &= 0 & m(16) &= -4.25 & m(17) &= -2.63 \text{ etc.} \end{aligned}$$

- Estas muestras han sido tomadas de la función: $4*\text{sen}(2x) + 2*\text{sen}(6x)$.
- La frecuencia de muestreo 'f' escogida es 20Hz. ($T=1/20\text{sg.}$).
- Puesto que $f=20\text{Hz}$. el máximo ancho de banda que se puede analizar es de 10Hz. (Teorema de Nyquist). La frecuencia máxima de la señal original es de 6Hz.



Aplicando la transformada de Fourier con $N=10$, y suponiendo que usamos una ventana rectangular que comienza en la muestra $m(7)$:

En primer lugar veremos las posibles frecuencias de estudio en función del parámetro 'n':

$$n=0 \Rightarrow F\left(\frac{0}{10} 20\right) = F(0\text{Hz.})$$

$n=1 \Rightarrow F\left(\frac{1}{10}20\right) = F(2\text{Hz.})$ (Frecuencia componente de la onda original)

$n=2 \Rightarrow F\left(\frac{2}{10}20\right) = F(4\text{Hz.})$

$n=3 \Rightarrow F\left(\frac{3}{10}20\right) = F(6\text{Hz.})$ (Frecuencia componente de la onda original)

$n=4 \Rightarrow F\left(\frac{4}{10}20\right) = F(8\text{Hz.})$

$[n=0..(N/2)-1]$

Vamos a realizar (por ejemplo) el desarrollo de la frecuencia 2Hz. ($n=1$), en donde esperamos encontrar un valor que indique que esta frecuencia es importante en la composición de la señal original.

$n=1$

$$F(2\text{Hz.}) = m(7) \left[\cos\left(-2\pi l \frac{0}{10}\right) + j \text{sen}\left(-2\pi l \frac{0}{10}\right) \right] +$$

$$m(8) \left[\cos\left(-2\pi l \frac{1}{10}\right) + j \text{sen}\left(-2\pi l \frac{1}{10}\right) \right] +$$

.....

$$m(16) \left[\cos\left(-2 \frac{9}{10}\right) + j \operatorname{sen}\left(-2 \frac{9}{10}\right) \right]$$

$$\begin{aligned} F(2\text{Hz.}) = & -2.63 [\cos(0)+j\operatorname{sen}(0)]+(-2.62)[\cos(-0.628)+j\operatorname{sen}(-0.628)]+ \\ & -4.25[\cos(-1.256)+j\operatorname{sen}(-1.256)]+(0)[\cos(-1.884)+j\operatorname{sen}(-1.884)] + \\ & 4.25 [\cos(-2.512)+j\operatorname{sen}(-2.512)] + 2.63[\cos(-3.14)+j\operatorname{sen}(-3.14)] + \\ & 2.62 [\cos(-3.768)+j\operatorname{sen}(-3.768)] + 4.25[\cos(-4.396)+j\operatorname{sen}(-4.396)] + \\ & 0 [\cos(-5.024)+j\operatorname{sen}(-5.024)] + (-4.25)[\cos(-5.652)+j\operatorname{sen}(-5.652)] \end{aligned}$$

$$F(2\text{Hz.}) = (-18.996+j6.231)=\sqrt{(-18.996)^2+(6.231)^2} = \boxed{20}$$

El valor obtenido es 20 (si lo ponderamos con el 1/N de la fórmula nos quedaría 2).

Ahora realizaremos el desarrollo para $n=2$ (4Hz.), donde esperamos un valor cercano a cero que nos indique que esa frecuencia no es constitutiva de la señal original.

$$\boxed{n=2}$$

$$\begin{aligned} F(4\text{Hz.}) = & \\ m(7) \left[\cos\left(-2\pi 2 \frac{0}{10}\right) + j \operatorname{sen}\left(-2\pi 2 \frac{0}{10}\right) \right] + & \end{aligned}$$

$$m(8) \left[\cos\left(-2\pi 2 \frac{1}{10}\right) + j \operatorname{sen}\left(-2\pi 2 \frac{1}{10}\right) \right] +$$

.....

$$m(16) \left[\cos\left(-2\pi 2 \frac{9}{10}\right) + j \operatorname{sen}\left(-2\pi 2 \frac{9}{10}\right) \right]$$

$$\begin{aligned} F(4\text{Hz.}) = & -2.63 [\cos(0) + j\operatorname{sen}(0)] + (-2.62) [\cos(-1.256) + j\operatorname{sen}(-1.256)] + \\ & -4.25 [\cos(-2.512) + j\operatorname{sen}(-2.512)] + (0) [\cos(-3.768) + j\operatorname{sen}(-3.768)] + \\ & 4.25 [\cos(-5.029) + j\operatorname{sen}(-5.029)] + 2.63 [\cos(-6.28) + j\operatorname{sen}(-6.28)] + \\ & 2.62 [\cos(-7.536) + j\operatorname{sen}(-7.536)] + 4.25 [\cos(-8.792) + j\operatorname{sen}(-8.792)] + \\ & 0 [\cos(-10.048) + j\operatorname{sen}(-10.08)] + (-4.25) [\cos(-11.304) + j\operatorname{sen}(-11.304)] \end{aligned}$$

$$F(4\text{Hz.}) = (0.018 + j0.011) = \sqrt{(0.018)^2 + (0.011)^2} = \boxed{0.011}$$

El valor obtenido es muy cercano a cero. Esto significa que al descomponer la señal original en sus frecuencias básicas, 4Hz. no aparece.

Realizando el mismo desarrollo para las demás frecuencias obtenemos:

$$n = 0 \rightarrow 0.020, \quad n = 3 \rightarrow 10, \quad n = 4 \rightarrow 0.025$$

Como se observa, las frecuencias integrantes de la señal original son:

$$n = 1 (f=2\text{Hz.}) \rightarrow \text{Importancia } 20 \quad (4*\text{sen}(2x))$$

$$n = 3 (f=6\text{Hz.}) \rightarrow \text{Importancia } 10 \quad (2*\text{sen}(6x))$$

Las demás frecuencias se pueden considerar inexistentes en la señal original.

Ahora vamos a modificar la señal de estudio para comprobar el funcionamiento de la transformada de Fourier en diversos casos básicos.

Función: $2*\text{sen}(2x) + 8*\text{sen}(6x)$

Resultados:

$$n = 0 (0\text{Hz.}) \rightarrow 0.023$$

$$n = 1 (2\text{Hz.}) \rightarrow 9.983 \quad [2*\text{sen}(2x)]$$

$$n = 2 (4\text{Hz.}) \rightarrow 0.006$$

$$n = 3 (6\text{Hz.}) \rightarrow 40 \quad [8*\text{sen}(6x)]$$

$$n = 4 (8\text{Hz.}) \rightarrow 0.023$$

Lo que muestra como se refleja la importancia de cada onda componente de la señal original.

Función: $4*\text{sen}(2x) + 2*\text{cos}(6x)$

Resultados:

$$n = 0 (0\text{Hz.}) \rightarrow 0.005$$

$$n = 1 (2\text{Hz.}) \rightarrow 20 \quad [4*\text{sen}(2x)]$$

$$n = 2 (4\text{Hz.}) \rightarrow 0.003$$

$$n = 3 (6\text{Hz.}) \rightarrow 10 \quad [2*\text{sen}(6x)]$$

$$n = 4 (8\text{Hz.}) \rightarrow 0.004$$

El método también funciona con funciones más complejas (por ejemplo formadas con senos y cosenos).

Función: $4 \cdot \text{sen}(2x) + 2 \cdot \text{sen}(6.5x)$

Resultados:

- $n = 0$ (0Hz.) $\rightarrow 1.66$
- $n = 1$ (2Hz.) $\rightarrow 18.554 [4 \cdot \text{sen}(2x)]$
- $n = 2$ (4Hz.) $\rightarrow 2.557$
- $n = 3$ (6Hz.) $\rightarrow 9.778 [2 \cdot \text{sen}(6.5x)]$
- $n = 4$ (8Hz.) $\rightarrow 2.115$

En este caso, al no coincidir la frecuencia base de 6.5Hz. con los valores buscados para cada 'n', existe un fenómeno de traspaso de los resultados hacia las frecuencias adyacentes a 6.5Hz.

Función: $4 \cdot \text{sen}(2x) + 2 \cdot \text{sen}(7x)$

Resultados:

- $n = 0$ (0Hz.) $\rightarrow 0.306$
- $n = 1$ (2Hz.) $\rightarrow 20.513 [4 \cdot \text{sen}(2x)]$
- $n = 2$ (4Hz.) $\rightarrow 1.260$
- $n = 3$ (6Hz.) $\rightarrow 5.327 [2 \cdot \text{sen}(7x)]$
- $n = 4$ (8Hz.) $\rightarrow 7.759$

Sería necesario más definición espectral para detectar la frecuencia de 7Hz. entre la de 6Hz. ($n=3$) y 8Hz. ($n=4$).

Función: $4 \cdot \text{sen}(2x) + 2 \cdot \text{sen}(6x) + 6 \cdot \text{sen}(8x)$

Resultados:

- $n = 0$ (0Hz.) $\rightarrow 0$
- $n = 1$ (2Hz.) $\rightarrow 19.97 [4 \cdot \text{sen}(2x)]$
- $n = 2$ (4Hz.) $\rightarrow 0.036$
- $n = 3$ (6Hz.) $\rightarrow 9.969 [2 \cdot \text{sen}(6x)]$
- $n = 4$ (8Hz.) $\rightarrow 30 [6 \cdot \text{sen}(8x)]$

Nueva comprobación del funcionamiento de la transformada de Fourier.

Una vez estudiado el funcionamiento de la transformación de Fourier y realizado un ejemplo numérico, resulta muy sencillo comprender una plantilla de código que implemente la funcionalidad buscada. Las siguientes líneas de código de alto nivel, realizan la TF de una serie de ' N ' muestras ' $m(k)$ ', con ' N ' par.

```

pi = 3.14
BUCLE n = 0 TO (N/2) - 1
COMIENZO
  Sumac ← 0
  Sumas ← 0
  BUCLE k = 0 TO N-1
  Sumac ← Sumac + m(k) * COS(-2 * pi * n * k / N)
  Sumas ← Sumas + m(k) * SIN(-2 * pi * n * k / N)
  FIN (* k (bucle del sumatorio) *)
  Modulo[n] ← SQR(Sumac * Sumac + Sumas * Sumas)
  FIN (* n (frecuencias) *)

```

Código para realizar la transformada de Fourier

El resultado queda almacenado en la matriz Modulo[N], sus valores habitualmente son aplanados por una función logaritmo decimal, para evitar que el rango de soluciones sea muy amplio.

Resulta necesario aclarar que el coste computacional requerido para realizar la transformada de Fourier tal y como aquí se ha descrito es muy alto. Cuando se programan implementaciones sobre computadoras, se emplea habitualmente la transformada rápida de Fourier (FFT). Este método se basa en la minimización de los cálculos realizados, evitándose la repetición de los cálculos que se emplean en las distintas etapas de la transformada tradicional.

Una vez explicados los conceptos fundamentales del funcionamiento de la transformada de Fourier, se pasará a desarrollar la teoría que nos dirige a la expresión matemática tomada como punto de partida en el apartado anterior.

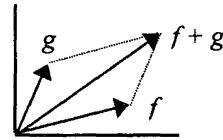
5. TRANSFORMADA DE FOURIER. FORMALISMO

Una vez aclarados los conceptos básicos y el funcionamiento de la transformada de Fourier, vamos a realizar la formalización matemática que da origen a la ecuación de partida del apartado anterior. Para ello nos basaremos en los espacios de Hilbert (espacios infinito-dimensionales) [ALE76], que nos brindan un camino simple y elegante para obtener la expresión básica de traspaso de señales en el dominio del tiempo al dominio de la frecuencia mediante la transformación de Fourier.

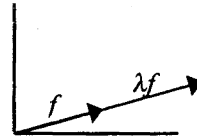
Para comenzar, repasaremos algunas propiedades de los espacios n-dimensionales que nos servirán como referencia para entender y emplear sus equivalentes en los espacios de Hilbert. Cada propiedad se ilustra con un gráfico bidimensional.

Sean $f = [f_1, f_2, \dots, f_n]$ y $g = [g_1, g_2, \dots, g_n]$ dos vectores n-dimensionales:

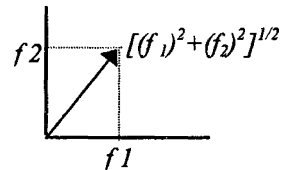
$$f + g = [f_1 + g_1, f_2 + g_2, \dots, f_n + g_n] \quad (1)$$



$$\lambda f = [\lambda f_1, \lambda f_2, \dots, \lambda f_n], \lambda \in \mathfrak{R} \quad (2)$$



$$\text{Longitud de } f = \sqrt{f_1^2 + f_2^2 + \dots + f_n^2} \quad (3)$$

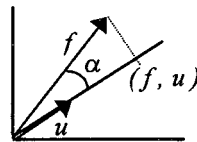
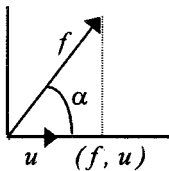


Producto escalar: $(f, g) = f_1 * g_1 + f_2 * g_2 + \dots + f_n * g_n$ (4)

Producto escalar: $(f, g) =$

$$= \sqrt{f_1^2 + f_2^2 + \dots + f_n^2} \sqrt{g_1^2 + g_2^2 + \dots + g_n^2} \cos(\alpha)$$
 (5)

En el caso de que un vector 'u' tenga como longitud la unidad, se puede observar con claridad como el producto escalar de un vector 'f' por 'u': (f, u) se corresponde con la proyección de 'f' sobre el eje determinado por 'u'.



Como caso particular de suma importancia se encuentra la condición de ortogonalidad (perpendicularidad) entre vectores. Dados dos vectores no nulos 'f' y 'g', se dice que son ortogonales si su producto escalar es nulo: $(f, g) = 0 \Rightarrow \cos(\alpha) = 0 \Rightarrow \alpha = 90^\circ$.

Partiendo de (4) se pueden establecer las siguientes propiedades del producto escalar:

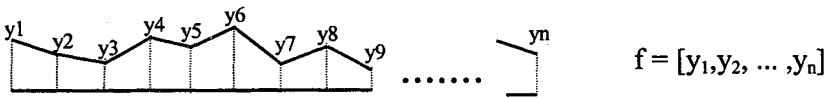
$$(f, g) = (g, f)$$
 (6)

$$(\lambda f, g) = \lambda (f, g)$$
 (7)

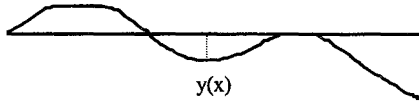
$$(f, g_1 + g_2) = (f, g_1) + (f, g_2)$$
 (8)

$$(f, f) \geq 0 \quad (9)$$

Podemos utilizar las características de los espacios n -dimensionales para trabajar con funciones discretas, para ello basta con considerar las distintas muestras de la función como componentes de un vector multidimensional, tal y como se representa en el siguiente gráfico:



Los espacios de Hilbert son infinito-dimensionales, por lo que nos encontramos con valores de ' n ' tendiendo a infinito y distancia entre componentes tendiendo a cero. En este caso estamos preparados para trabajar con funciones continuas, de hecho, un vector en un espacio de Hilbert se define como una función $f(x) \mid x \in [a, b]$.



En espacios infinito-dimensionales, la suma de vectores y la multiplicación de un vector por un número se definen como la adición de funciones y la multiplicación de una función por un número.

En (3) se define la longitud de un vector n -dimensional como $\sqrt{\sum_{i=1}^n f_i^2}$. En un espacio de Hilbert donde ' f ' es una función tenemos:

$$\text{Longitud } f = \sqrt{\int_a^b f^2(x) dx} \quad (10)$$

En (5) se define el producto escalar de dos vectores n-dimensionales 'f' y 'g' como:

$$(f, g) = \sqrt{\sum_{i=1}^n f_i^2} \sqrt{\sum_{i=1}^n g_i^2} \cos(\alpha)$$

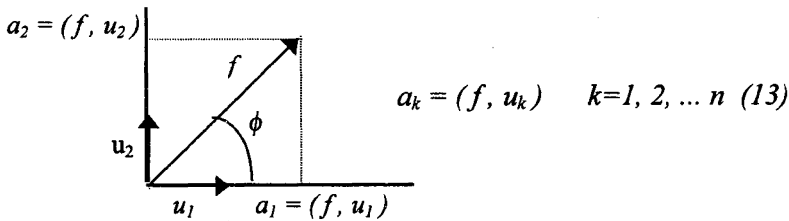
Traspassando la ecuación anterior al espacio infinito-dimensional :

$$(f, g) = \sqrt{\int_a^b f^2(x) dx} \sqrt{\int_a^b g^2(x) dx} \cos(\alpha) \quad (11)$$

Análogamente, la expresión (4) $\left((f, g) = \sum_{i=1}^n f_i g_i \right)$ se convierte a su

equivalente en espacios de Hilbert: $(f, g) = \int_a^b f(x)g(x)dx \quad (12)$

De forma similar a lo que ocurría en los espacios n-dimensionales, en un espacio de Hilbert las funciones cuyo producto escalar es cero son ortogonales. En los espacios n-dimensionales también se cumple la siguiente propiedad: si tomamos 'n' vectores arbitrarios 'u_i', perpendiculares entre sí y de longitud unidad, todo vector n-dimensional se puede caracterizar mediante sus componentes halladas realizando la proyección del vector sobre cada uno de los 'n' ejes determinados por los vectores 'u_i'.



En espacios de Hilbert, un sistema de funciones $\phi_1(x)$, $\phi_2(x)$, ... $\phi_n(x)$

es ortogonal si se cumple: $\int_a^b \phi_j(x)\phi_k(x)dx = 0 \quad \forall j \neq k$ (14).

La ortonormalidad del sistema se consigue cuando se cumple:

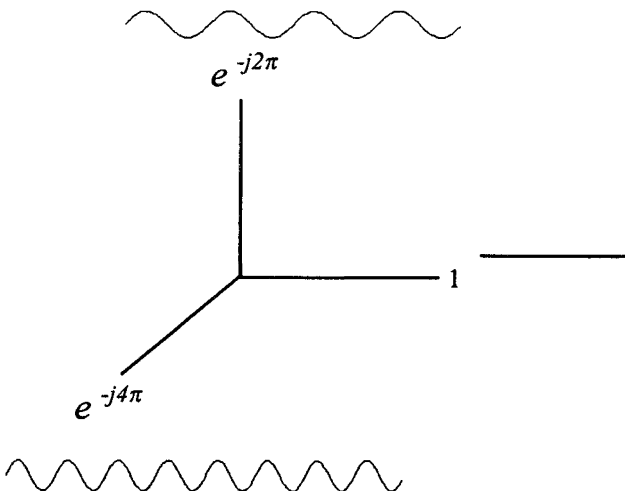
$$\int_a^b \phi_i^2(x)dx = 1 \quad i=1..n \quad (15)$$

Tomando la sucesión de funciones:

$$1, e^{-j2\pi}, e^{-j4\pi}, e^{-j6\pi}, \dots, e^{-j2n\pi}$$

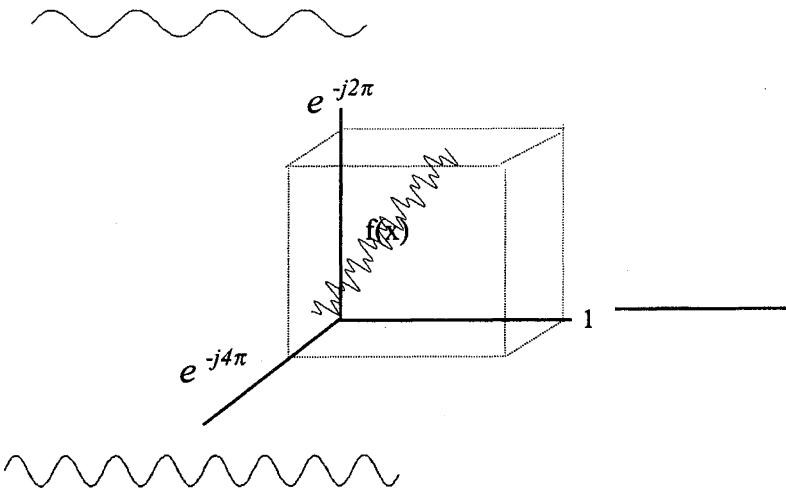
obtenemos la base

$$\phi_n = e^{-j2n\pi}, \{ e^{-j2n\pi} = \cos(2n\pi) - j*\sen(2n\pi) \} \quad (16)$$

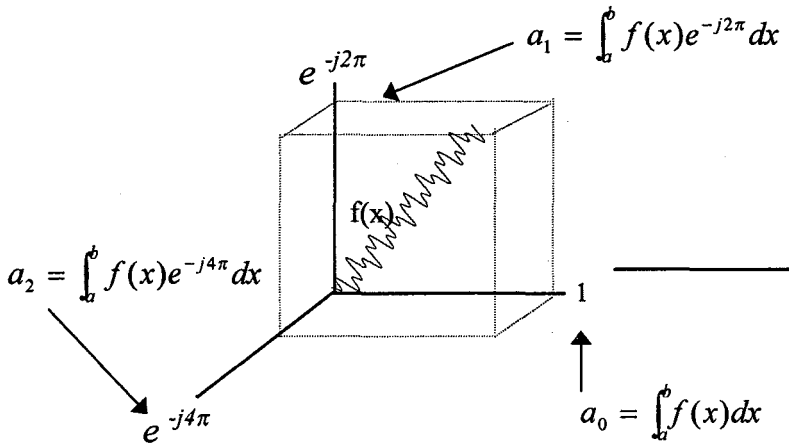


Una función $f(x)$ se puede expresar como:

$$f(x) = a_1\phi_1(x) + a_2\phi_2(x) + \dots + a_n\phi_n(x) \quad (17)$$



Aplicando (12) y (13) obtenemos:
$$a_n = \int_a^b f(x)\phi_n(x)dx \quad (18)$$



El gráfico anterior ilustra la descomposición de una señal compleja empleando una base de senos y cosenos de diferentes frecuencias. El producto escalar de la función analizada por cada elemento de la base nos indica la proporción en la que cada frecuencia ($2n\pi$) participa en la composición de la señal original.

Cuando existe una marcada ortogonalidad entre la señal analizada y uno de los elementos de la base (ϕ_k), significa que la frecuencia determinada por ' ϕ_k ' no interviene de forma significativa en la constitución de la señal original. Analíticamente, esta característica se refleja en el término $\cos(\alpha)$ que existe en la ecuación (11). Gráficamente, los términos ' a_n ' se corresponden con las proyecciones de la onda analizada sobre cada uno de los ejes determinados por los elementos de la base.

Combinando los resultados obtenidos en (16) y (18), los valores ' a_n ' se pueden expresar como:

$$a_n = \int_a^b f(x)e^{-j2n\pi} dx \quad (19)$$

Trabajando con 'N' señales discretas $m(kT)$ cuyo recorrido se abarca desde un índice $k=0$ hasta $k=N-1$, la ecuación (19) se convierte en la utilizada en el apartado anterior para ilustrar el funcionamiento de la transformada de Fourier:

$$F\left(\frac{n}{NT}\right) = \frac{1}{N} \sum_{k=0}^{N-1} m(kT) e^{-j\frac{2\pi nk}{N}} \quad (20)$$

Empleando el formalismo que nos brindan los espacios de Hilbert, se ha llegado con sencillez a la expresión de la transformada de Fourier, cuya versión discreta nos permite realizar un análisis intuitivo de su funcionamiento y propiedades básicas.

6. CONCLUSIONES

La TF nos permite traspasar señales de voz desde el dominio del tiempo hasta el dominio de la frecuencia. Una vez que disponemos de la información frecuencial, podemos crear sonogramas con facilidad.

Enfocando las explicaciones desde el formulismo de la señal discreta, obtenemos dos ventajas:

1. Resulta fácil transmitir el funcionamiento interno de la TF como suma de muestras de ondas con diferentes frecuencias.
2. Partimos de una base más cercana a los algoritmos informáticos que se utilizan para la obtención de espectros de voz.

Para entender los conceptos fundamentales de la TF no es necesario poseer unos profundos conocimientos matemáticos, puesto que las explicaciones pueden apoyarse en diversos gráficos

representativos que nos ayudan a crearnos una visión intuitiva de su funcionamiento.

Por último, cabe indicar que una persona con los suficientes conocimientos informáticos puede utilizar con facilidad los conocimientos que se aportan en este artículo para desarrollar una aplicación de creación de espectrogramas de voz.

7. REFERENCIAS BIBLIOGRÁFICAS

- [CAN74] S. McCandless, "An algorithm for automatic formant extraction using linear prediction spectra", *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 22(2), 1974, pp. 135-141
- [COC67] W.T. Cochran, *What is the Fast Fourier Transform?*, IEEE Trans. Audio Electroacoust., Vol. AU-15, 1967, pp. 45-55
- [FOL92] G.B. Folland, *Fourier analysis and its applications*, Wadsworth & Brook, 1992
- [RAB78] L.R. Rabiner, R.W. Schafer, *Digital processing of speech signals*, Prentice Hall, 1978
- [RAB93] L.R. Rabiner, *Fundamentals of speech recognition*, Biing-Hwang Juang, Prentice Hall, 1993
- [MAK75] J. Makhoul, "Linear prediction: a tutorial review", *Proceedings of the IEEE*, Vol. 63(4), 1975, pp. 561-580
- [PAR86] T. Parsons, *Voice and speech processing*, Mc Graw Hill, 1986