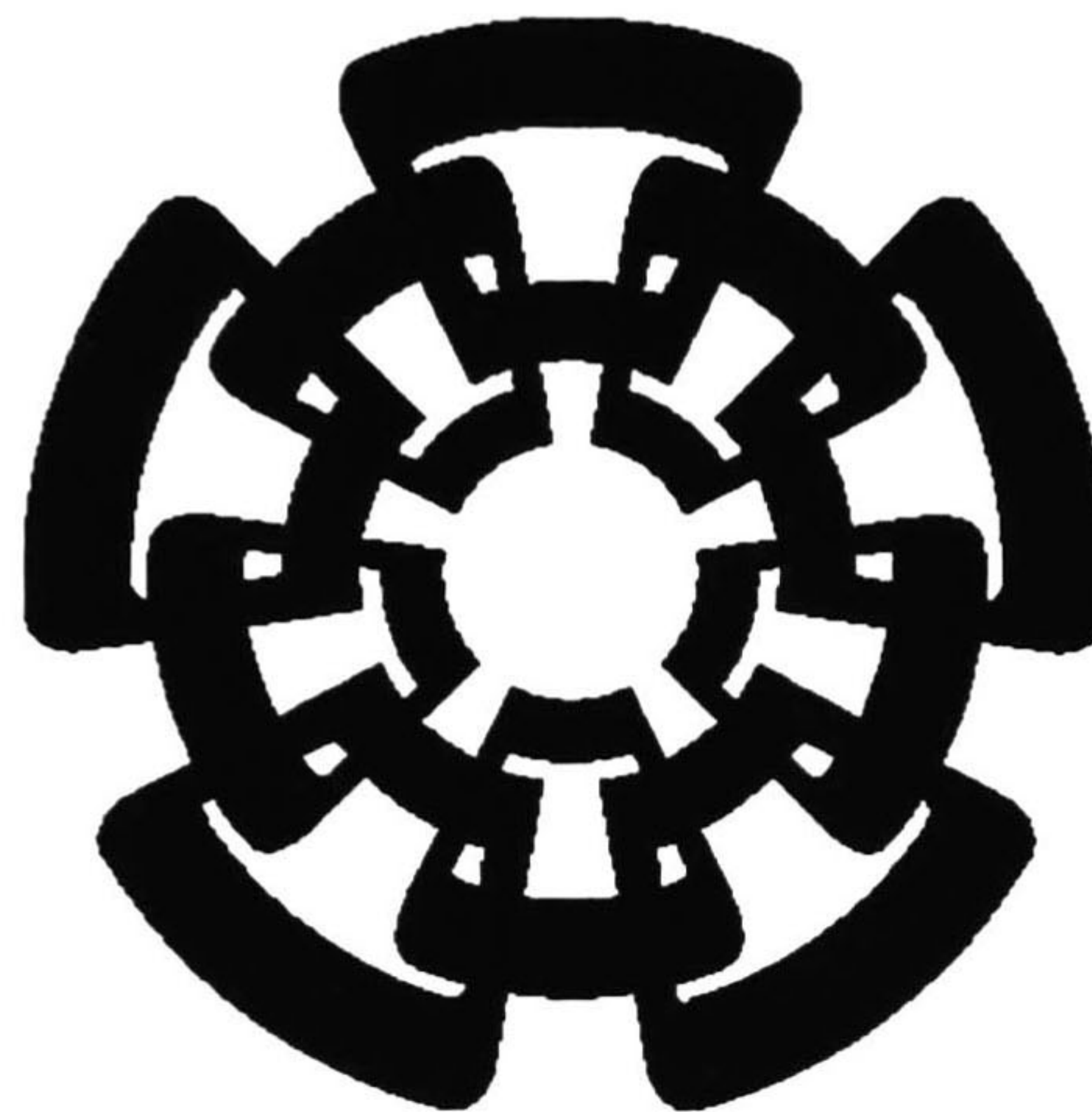

CENTRO DE INVESTIGACION Y DE ESTUDIOS AVANZADOS DEL I.P.N

Laboratorio de Ingeniería Eléctrica y Ciencias de la Computación - Guadalajara



“Compresión de la Señal de Electrocardiografía basada en el Análisis de Ondeletas e Implementación en un Procesador de Señales”

**TESIS PARA LA OBTENCION DEL GRADO DE:
MAESTRO EN CIENCIAS EN LA ESPECIALIDAD DE INGENIERIA
ELECTRICA**

P R E S E N T A :

JOSE DE JESUS PEREZ SEVILLA

***Becario de CONACyT**

Director de Tesis: Dr. Arturo Veloz Guerrero.

Guadalajara, Jal.



Septiembre de 1997

CLASIF.:
ADQUIS.: DEPTO. SERV. BIBL
FECHA: 18-V-98
PROCED.: TESIS - 1998
»

I. AGRADECIMIENTOS

Quisiera agradecer primeramente a Dios, no sólo por haber me permitido llegar al final de mis estudios de maestría, sino también por haber motivado mi desarrollo como persona al mismo tiempo que mi desarrollo como ingeniero. El conocimiento del amor que tiene por mí y sus cuidados, son para mí la seguridad de que mi vida, de la cual mi preparación forma parte, va mejorando.

Quiero agradecer a mis padres, que además de darme la vida me han dado su ejemplo de superación constante y su apoyo incondicional en cada momento, definitivamente no sería lo que soy sin los padres tan maravillosos que tengo.

También agradezco a mi esposa su paciencia y apoyo en el tiempo que no pude estar con ella por cumplir con los estudios.

Es necesario hacer notar que esta tesis no se hubiera realizado sin la constante participación de mi maestro y asesor el Dr. Arturo Veloz, así como las constantes revisiones de mi compañera Claudia Feregrino, también a mi hermana María Concepción por la revisión final y los comentarios de redacción.

A mis maestros por su ayuda durante mis estudios y a mis compañeros que compartieron conmigo los buenos y malos momentos de estos dos años, siempre dispuestos a trabajar en equipo y a enseñarme lo que no entendía o lo que creía entender y estaba mal. También por estar dispuestos a escuchar los avances y retrocesos en esta tesis, permitiéndome organizar mis ideas a costa de su tiempo.

También mi más profundo agradecimiento al CINVESTAV y al CONACYT que como organizaciones han permitido que mis estudios lleguen a su fin.

II CONTENIDO

Capítulo

I. Agradecimientos.

II. Contenido.....	1
1 Introducción.....	2
2 Transformada de Ondeleta.....	3
2.1 Desarrollo Histórico.....	3
2.2 Transformada de Ondeleta Continua.....	6
2.3 Transformada de Ondeleta Discreta y Multirresolución.....	12
3 Ondeletas y Bancos de Filtros.....	19
3.1 Preliminares.....	19
3.2 Dominio del Tiempo.....	23
3.3 Dominio de la Frecuencia.....	25
3.4 Dominio de Polifase.....	25
3.5 Diseño de Filtros.....	27
4 Compresión basada en Análisis de Ondeletas.....	29
4.1 Codificación por Transformada.....	29
4.2 Transformación Optima.....	31
4.3 Cuantización y Distribución Optima de los Bits.....	32
4.4 Codificación por Entropía.....	33
5 Sistema de Electrocardiografía Ambulatorio.....	35
5.1 Electrocardiografía.....	35
5.2 El Sistema HOLTER Digital.....	37
5.3 Simulación y Resultados.....	38
5.4 Detalles de Implementación.....	45
6 Conclusiones y Trabajo Futuro.....	49
7 Referencias.....	50
8 Apéndices.....	51
8.1 Apéndice A.....	51
8.2 Apéndice B.....	52
8.3 Apéndice C.....	54

1. INTRODUCCION

En los últimos 15 años se ha desarrollado un marco de trabajo común para algunas técnicas que se utilizaban en distintos campos de la Física, Matemáticas e Ingeniería. Esta teoría común se conoce como ondeletas o wavelets, el nombre en español se deriva de su nombre en francés “ondelettes” puesto por su creador J. Morlet en 1976.[5]

La transformada de ondeleta en sus diferentes presentaciones : continua, de escala discreta, de tiempo discreto, ortogonales, no ortogonales y biortogonales, posee propiedades útiles en el tratamiento de procesos o señales no estacionarias, en los que se representan de forma eficiente las variaciones rápidas que suceden sólo durante instantes de tiempo pequeños [13].

En el presente trabajo se aplica la transformada de ondeleta en la compresión de señales de electrocardiografía (ECG) con la finalidad de almacenarlas en un menor espacio. Pensando en la aplicación a un equipo de ECG digital portátil, la meta del diseño fue lograr una compresión mayor de 7 :1 con una distorsión que en forma estándar se sabe que no afecta la información útil de la señal de ECG.

La utilidad de la transformada de ondeleta (T.O.) en la compresión, se deriva de que al aplicarla a la señal de ECG, más del 90% de los coeficientes son casi cero (menores al 5% del valor máximo de la señal), lo que indica que la información se encuentra concentrada en unos cuantos coeficientes, los cuales pueden ser codificados con mayor precisión y los demás con menos bits ; además, al realizar la transformada los valores pequeños tienen una ocurrencia mayor, lo que permite utilizar la codificación por entropía y obtener buenos resultados.

Es importante notar que la compresión no sólo se logra con la utilización de las ondeletas, sino que es necesaria una distribución apropiada de los bits en los coeficientes, mas cuantización adecuada y un algoritmo de codificación, en este caso el de Huffman, que aproveche la estadística de la señal, [9].

Una de las grandes ventajas de la T.O. es que posee un algoritmo de implementación eficiente, el algoritmo de Mallat, el cual se basa en bancos de filtros con razón de muestreo múltiple que actúan en cascada (Iterated Multirate Filter Banks), lo que reduce el número de operaciones necesarias para obtener la transformada [11].

Se realizaron simulaciones en Matlab con señales de ECG, realizando la descomposición, cuantización y codificación para determinar la relación de compresión para las diferentes ondeletas, incluyendo la de Daubechies [4], las simétricas [4], y algunas biortogonales [19], con la intención de encontrar la óptima para la aplicación.

Por último, se utilizó un sistema basado en un DSP de TI para probar los algoritmos en tiempo real. Este sistema consta de un ECG analógico, un convertidor analógico digital, el DSP para realizar la transformada, y una PC para realizar la cuantización y codificación también en la PC se simula un banco de memoria .

2. TRANSFORMADA DE ONDELETA

2.1 ONDELETAS Y SUS ORIGENES. (HISTORIA Y DESARROLLO)

El desarrollo de las ondeletas, que se ha dado en los últimos 15 años está conectado a muchas áreas, incluidas las matemáticas puras y aplicadas, física, computación, e ingeniería [5]. La historia de las ondeletas se puede comparar con un árbol cuyas raíces se adentran profundamente en muchas direcciones, pero sin duda, la raíz principal lo forma el análisis de Fourier, que comienza con el trabajo de un matemático francés, Jean Baptiste Joseph Fourier sobre el análisis de la conducción del calor, en el siglo XIX, sobre el cual el físico James Clerk Maxwell escribió “es un poema matemático grandioso” [2]. Las consecuencias de este poema han transformado el mundo como hoy lo conocemos. Sus ideas han dominado el análisis matemático por más de 100 años, apareciendo en la ingeniería y la ciencia en casi todos los modelos, como diría el matemático T.W. Korner : “el análisis de Fourier forma ya parte del sentido común de nuestra sociedad” [2]. Existen dos conceptos innovadores en el análisis de Fourier : el primero una afirmación matemática, probada después por Dirichlet , y el otro una explicación de por qué esta afirmación es valiosa; la afirmación dice que cualquier función periódica puede representarse en una suma de funciones de senos y cosenos, pero aun más trascendente es que la función y la suma de senos y cosenos son dos representaciones del mismo objeto en diferentes lenguajes.

La Transformada de Fourier es un prisma matemático que separa una función en las frecuencias que la componen, de la misma forma como el prisma descompone la luz en colores. A pesar de los grandes avances de Fourier, esta transformada tiene el inconveniente de ocultar la información del tiempo o espacio en las fases de los coeficientes, es decir, si se analiza una melodía con las técnicas de Fourier se puede saber qué notas la componen, pero es imposible saber cuándo fueron tocadas.

Para superar este inconveniente, tratando de imitar el comportamiento de los sentidos, los cuales presentan información en tiempo y frecuencia al mismo tiempo, se utiliza el análisis de Fourier con ventanas de tiempo, de esta manera se puede limitar el período de tiempo en el que ocurren los sucesos, deslizar la ventana y analizar otro segmento. Este método propuesto por Gabor en 1946, impone compromisos en la elección de la ventana, ya que mientras más pequeña sea ésta mejor se pueden localizar los cambios súbitos, pero se ocultan los cambios en las frecuencias bajas que componen la señal.

A finales de los 70's un ingeniero geofísico francés, de nombre Jean Morlet, propuso un método alternativo al análisis de Fourier de ventanas de tiempo. El modo estándar para analizar el subsuelo en busca de petróleo es mandar vibraciones o impulsos bajo tierra y analizar sus ecos, es decir, la señales que se obtienen de regreso. Este análisis se supone que es capaz de decir qué tan profundo, qué tan anchas y de qué están hechas las capas del subsuelo, así de una forma muy general, las frecuencias de los ecos están relacionadas con el grosor de las capas, correspondiendo las altas frecuencias a las capas delgadas. Sin embargo, este análisis no es sencillo, en las propias palabras de Morlet “ hay cientos de capas, y las reflexiones de cada una de ellas interfieren entre sí, lo que se convierte en un desastre, y por eso uno trata de separarlas” Para separarlas Morlet revisó las técnicas de Gabor de 30 años atrás, pero se dio cuenta del

compromiso en la elección de la ventana, entonces Morlet le dio otro enfoque : en lugar de dejar fijo el tamaño de la ventana y llenarlo de oscilaciones, dejó fijo el número de oscilaciones y varía el tamaño de la ventana, expandiéndola o comprimiéndola como si fuera un acordeón. Cuando la ventana se expande, las oscilaciones dentro de ella se separan, disminuyendo su frecuencia, cuando se comprime, éstas se comprimen también aumentando su frecuencia.

Como las nuevas funciones tienen la misma forma, Morlet las llamó ondeletas de forma constante, (la palabra ondeletas se deriva de el término en francés "Ondelettes" que significa ondas pequeñas) más tarde el adjetivo " de forma constante" se suprimió y ahora se les llama sólo ondeletas, que es el término general que utilizan los geofísicos para las señales que envían al subsuelo, tal vez algunos ingenieros geofísicos todavía se sorprendan al encontrar la palabra en otro contexto.

Morlet pasó tiempo tratando de convencer a sus colegas de que esta transformada era una herramienta que podría tener un uso más general, pero, en palabras de su audiencia "si eso fuera cierto, entonces estaría en los libros de matemáticas, como no está ahí, probablemente no tiene importancia" [2]. Tratando de darle un fundamento matemático más sólido, Morlet consultó a un físico amigo suyo de la escuela, el cual lo mandó con Alex Grossmann, quien trabajaba en mecánica cuántica de estados de fase. Entre los dos tardaron un año transformando la receta numérica de Morlet en una transformada con una fórmula de inversión, es decir, demostraron que es posible reconstruir exactamente la señal a partir de los coeficientes de la transformada de ondeleta y se dieron cuenta que la transformada de Morlet era equivalente a un formalismo matemático llamado "estados coherentes" desarrollado en mecánica cuántica por E. Aslaksen y J. Klauder para un modelo simplificado de gravedad cuántica.

Otro de los protagonistas en el desarrollo de la teoría de las ondeletas, Yves Meyer, se vio involucrado casi por accidente, cuando esperaba para poder hacer uso de la fotocopidora. Leyendo los papeles que su coordinador estaba fotocopiando descubrió que un artículo de Grossmann utilizaba matemáticas familiares para él, que fueron descubiertas en su especialidad el análisis armónico por el matemático A. Calderón en los 60's. Meyer, al ser un especialista en el campo tuvo la reacción, muy común en los seres humanos, de decir, "pero si esto nosotros lo sabemos desde hace mucho tiempo"[2], y motivado por la nueva aplicación y enfoque que Morlet y Grossmann le daban a la fórmula de Calderón, se puso en contacto con Grossmann. Este encuentro de dos de las raíces de la transformada de ondeleta une los campos de aplicación y de teoría, beneficiando a ambos.

Cuando Meyer fue a buscar a Grossmann en 1985, el cálculo de la transformada de ondeleta era lento porque se necesitaba gran cantidad de operaciones, debidas a la redundancia implícita en la forma continua de la transformada, es decir, la información que está codificada en un coeficiente de la transformada existe también en los coeficientes vecinos. Meyer intentó demostrar que no existen ondeletas que eviten esta redundancia, pero en lugar de encontrar esta demostración, encontró un conjunto de ondeletas "ortogonales" : en éstas, la información que cada una codifica no puede ser codificada por las demás , y con ellas se evitan repeticiones de la información.

Estando Meyer dando un curso de ondeletas en los Estados Unidos recibió unas llamadas de un estudiante de 23 años de la maestría en visión por computadora de la Universidad de Pennsylvania, Filadelfia, Stéphane Mallat quien le hace ver que lo que los matemáticos, físicos y otros ingenieros desarrollaron se utiliza también en visión por computadora con otros nombres : los algoritmos de multirresolución de Burt y Adelson y la pirámide de Laplace, y que se necesita

ver el problema desde un punto de vista más amplio. Después de tres días, ambos resolvieron los detalles matemáticos. Como resultado, Mallat escribe su artículo “ Multiresolution approximation and wavelets”, donde demuestra que todas esas áreas y algunas otras como los filtros QMF y los bancos de filtros Q, tienen un fundamento en el análisis armónico y en la teoría de multirresolución ; además, propone un algoritmo rápido y eficiente para calcular la transformada de ondeleta a partir de un banco de filtros.

Cuando Mallat propuso su algoritmo utilizó una familia de ondeletas construidas por Battle-Lamarié de longitud infinita, y la implementación se realizó truncando con cierto error. Ingrid Daubechies, en ese tiempo estudiante de doctorado de Grossmann, se dedicó a estudiar las series de ondeletas, haciendo énfasis en el concepto de marco y en la cuestión de la regularidad, es decir, qué tan suavemente cambia la señal, y logró encontrar una familia de ondeletas con soporte finito que, por lo tanto, NO necesitan ser truncadas [4]. Estas señales no son derivadas de fórmulas analíticas, sino que son hijas de la era de la computación, siendo el resultado de las iteraciones de filtros de respuesta finita (FIR).

De esta forma, la construcción de nuevas bases de ondeletas adquiere un método alternativo, que es el de proponer un filtro FIR con ciertas características y observar el tipo de ondeletas que genera. La utilización de estos bancos de filtros es más antigua en la ingeniería, desde Croisier-Esteban-Galland con los filtros de espejo o QMF. Con la aportación de Daubechies se une por completo esta parte de la raíz , aportando nuevas formas de construir ondeletas con las formas biortogonales, o también llamadas filtros en cuadratura conjugados (CQF), y a las diferentes implementaciones de codificación en subbandas y los paquetes de ondeletas.

Con esto se ponen las bases para los grandes desarrollos en aplicaciones que utilizan las ondeletas como son : compresión [6][10][15], detección de singularidades [4], modelado de sistemas [11][17][20], etc.

2.2 TRANSFORMADA DE ONDELETA CONTINUA.

PRELIMINARES

Se conoce como teoría de ondeletas al marco de trabajo, que comprende teorías desarrolladas independientemente, pero que están relacionadas íntimamente como :

- Transformada de ondeleta.
- Expansión en series de ondeletas.
- Análisis de multirresolución.
- Codificación por subbandas.

La teoría de ondeletas es útil en el análisis de señales *no estacionarias*[13], como alternativa a la STFT, debido a su mayor facilidad de caracterizar sólo los cambios de las señales.

La noción de los fenómenos estacionarios viene de probabilidad, e indica que el fenómeno no cambia en el tiempo sus características estadísticas como su media, su distribución, y su varianza, así como que su autocorrelación sólo depende de t_2-t_1 .

Cuando se asume que el proceso es estacionario, se puede utilizar la transformada de Fourier, porque la información en el tiempo no es relevante, esta transformada considera la información en el tiempo desde menos infinito hasta mas infinito.

$$F(f) = \int_{-\infty}^{\infty} f(t)e^{-2j\pi ft} dt$$

Para el caso no estacionario, se adapta tomando solo una parte de la señal a la vez, esto se hace multiplicando la señal por una función de duración finita, centrada a un tiempo t , para obtener $F(w, t)$. A la función de duración finita se le llama ventana $v(t)$, esta es la transformada de Gabor o STFT [13].

$$F_G(f, \tau) = \int_{-\infty}^{\infty} v(t - \tau) f(t) e^{-2j\pi ft} dt$$

Sin embargo, esto tiene el inconveniente de que las características de la ventana $v(t)$, son las mismas para todo t , lo que limita la resolución en tiempo o en frecuencia.

Para clarificar esto es necesario cuantificar la resolución, por ejemplo, en frecuencia la resolución

está dada por el ancho de banda definido como [13]: $\Delta f^2 = \frac{\int_{-\infty}^{\infty} f^2 |V(f)|^2 df}{\int_{-\infty}^{\infty} |V(f)|^2 df}$

Tomando como resolución, el que dos senoides con diferente frecuencia pueden discriminarse sólo si estan separadas una distancia mayor que Δf .

En el tiempo la resolución se estima a partir del intervalo de autocorrelación definido como [13]:

$$\Delta t^2 = \frac{\int_{-\infty}^{\infty} t^2 |v(t)|^2 dt}{\int_{-\infty}^{\infty} |v(t)|^2 dt},$$

de igual manera con esta definición se tiene que dos impulsos pueden

diferenciarse sólo si están separados por más de Δt .

Utilizando estas definiciones es posible probar que para toda función $v(t)$: $\Delta f \Delta t \geq \frac{1}{4\pi}$, esto se conoce como el principio de incertidumbre, o desigualdad de Heisenberg (su demostración se encuentra en el apéndice A). Es por esto que al elegir la función de ventana en la Transformada de Gabor se establece un compromiso entre la resolución en tiempo y la resolución en frecuencia, y solo se puede intercambiar una por otra.

En la mayoría de las señales los fenómenos de altas frecuencias normalmente duran poco tiempo ; por el contrario, en los fenómenos de bajas frecuencias los eventos duran un tiempo mayor. En el caso del análisis con la Transformada de Fourier de Tiempo Finito (STFT) el querer localizar con mayor precisión el instante de tiempo en el que ocurren los cambios de alta frecuencia, implica hacer la ventana más corta con lo que su resolución en frecuencia es pobre afectando mucho en las frecuencias bajas; por el contrario si se quieren discriminar las bajas frecuencias entre si, es necesario hacer la ventana muy grande con lo que es imposible conocer en qué tiempo ocurrieron los cambios de altas frecuencias, es decir, en la STFT la división del espacio tiempo-frecuencia es igual para las frecuencias altas, medianas y bajas. (Figura 2.1) Para aprovechar las características de las señales no estacionarias, y poder representarlas mas convenientemente, se puede hacer una división de las frecuencias en bandas cuya amplitud esté relacionado con su frecuencia central.

En esta forma se tiene , $\frac{\Delta f}{f} = c$

Si $c=2$, entonces el plano tiempo-frecuencia queda dividido como el de la Figura 2.1, otorgando buena resolución en el tiempo a las altas frecuencias y buena resolución en frecuencia para las bajas frecuencias.

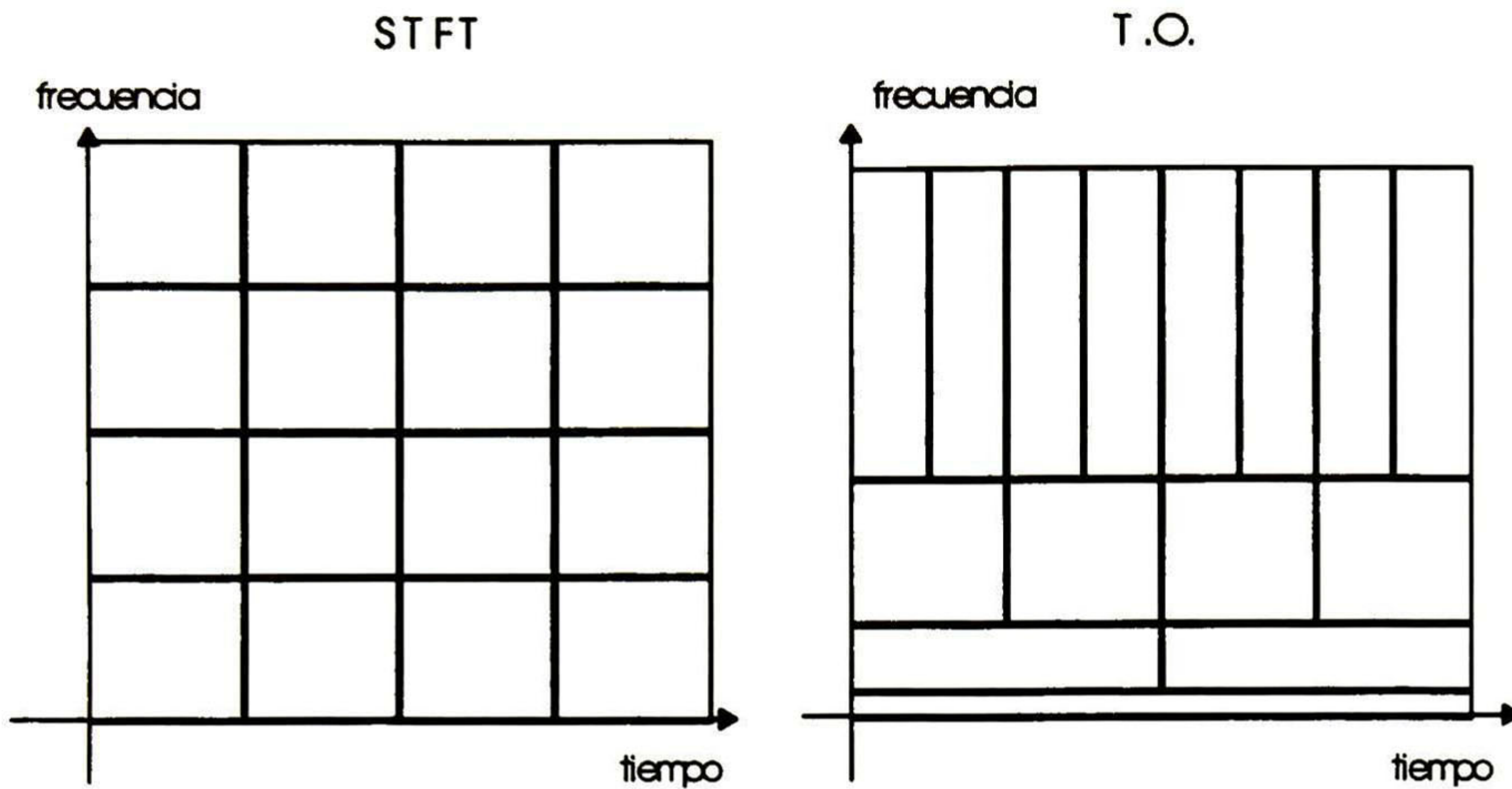


Figura 2.1. Diversas formas de dividir el plano tiempo - frecuencia

La transformada de ondeleta puede verse como un banco de filtros que realiza una división del plano tiempo-frecuencia como la indicada en el lado derecho de la figura 2.1; adaptando el intervalo de correlación de su ventana y el ancho de banda de su filtro dependiendo de la frecuencia que se está analizando [13]. Además, todas las respuestas al impulso de los filtros son versiones comprimidas y expandidas de una función prototipo. A la operación de comprimir o expandir una señal en el tiempo se le llama escalamiento.

Para entender mejor el concepto de escala y cómo está relacionado con la frecuencia, se recuerda que al multiplicar la variable independiente por una constante se altera la forma de la señal, como se ve en la figura 2.2.

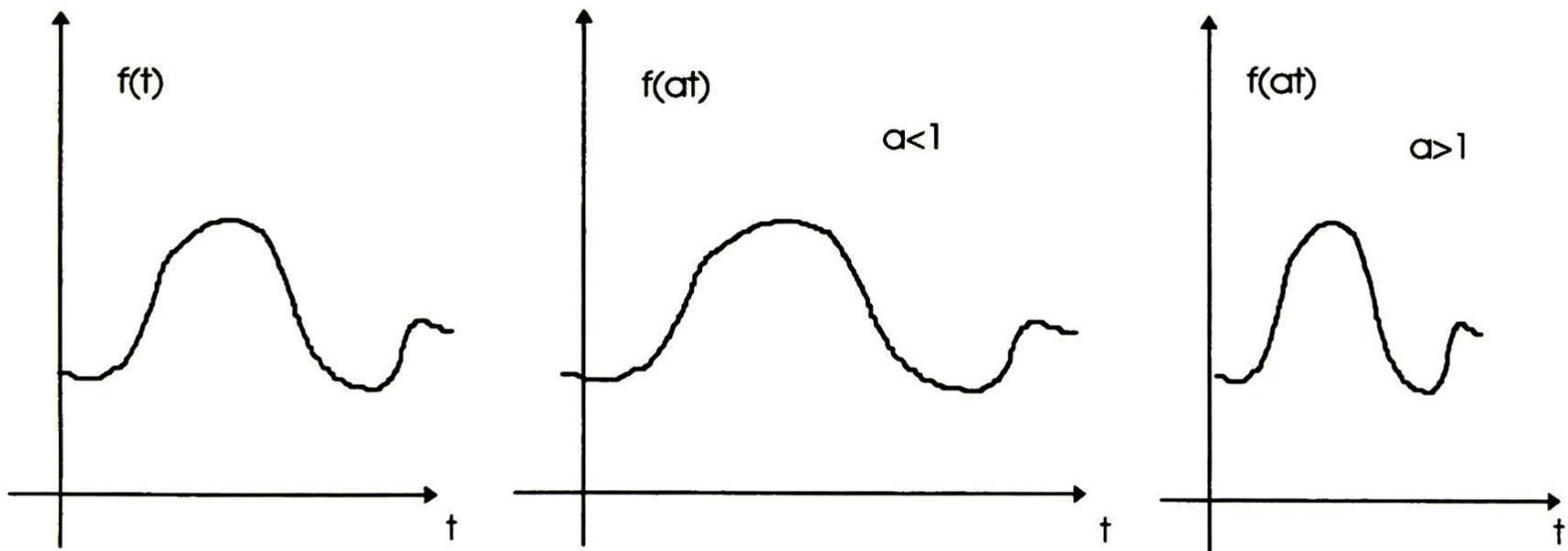


Figura 2.2 Escalamiento de una función en el tiempo.

Al realizar la operación $f(t) \Rightarrow f(at)$, la función se expande si $a < 1$ y se contrae si $a > 1$. El parámetro 'a' se conoce como la escala.

Pensando en que la respuesta del filtro del que se deriva la descomposición en subbandas es $h(t)$, se define la transformada de ondeleta como[13]:

$$F_o(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} h\left(\frac{t-\tau}{a}\right) f(t) dt$$

En esta definición se ve la transformada como un filtro variable, que realiza una descomposición en diferentes anchos de banda .

O con un cambio de variable:

$$F_o(a, \tau) = \sqrt{a} \int_{-\infty}^{\infty} h\left(t - \frac{\tau}{a}\right) f(at) dt$$

Se puede ver la misma señal comprimida para observar los fenómenos que duran mucho tiempo o expandida para observar los detalles en instantes de tiempo específicos, así se ve que la transformada de ondeleta es como un microscopio matemático que permite observar una señal más al detalle restringiendo la amplitud de lo que se enfoca, o tener una vista global, perdiendo los detalles de la función.

La escala en esta transformada tiene la misma interpretación que en los mapas, es decir, permite reconocer objetos de diferentes dimensiones :

Escalas grandes = vista global.

Escalas pequeñas = vista detallada.

La definición anterior de la transformada de ondeleta con $h(t) = v(t)e^{-2j\pi f_0 t}$ puede relacionarse con STFT, así la escala queda relacionada como : $f = f_0/a$, donde se observa la relación $1/a \leftrightarrow f$, y como escala y frecuencia son términos que están relacionados, la palabra “escala” se prefiere a “frecuencia”, en la literatura de las ondeletas, la palabra “frecuencia” se reserva para el análisis de Fourier.

Así como en el análisis de Fourier existe el espectrograma, que es la representación gráfica de la magnitud al cuadrado de $F(w, t)$, la representación gráfica de la magnitud al cuadrado de $F(a, \tau)$ en un mapa de tiempo-escala se denomina Escalograma, y tiene la cualidad de mostrar la distribución de energía en un tiempo de terminado para cada escala. Esta representación es muy útil, pero tiene la desventaja de que tanto el escalograma como el espectrograma no son invertibles en general, porque la información de la fase es necesaria para reconstruir la señal.

Algunas diferencias de la representación de escalogramas y espectrogramas se muestra en la figura 2.3, donde se representa la gráfica idealizada de un escalograma y un espectrograma de una función impulso ; como el impulso tiene igual energía en todas las frecuencias, en esta figura se puede apreciar la variación en la resolución en tiempo que tiene la transformada de ondeleta para diferentes escalas.

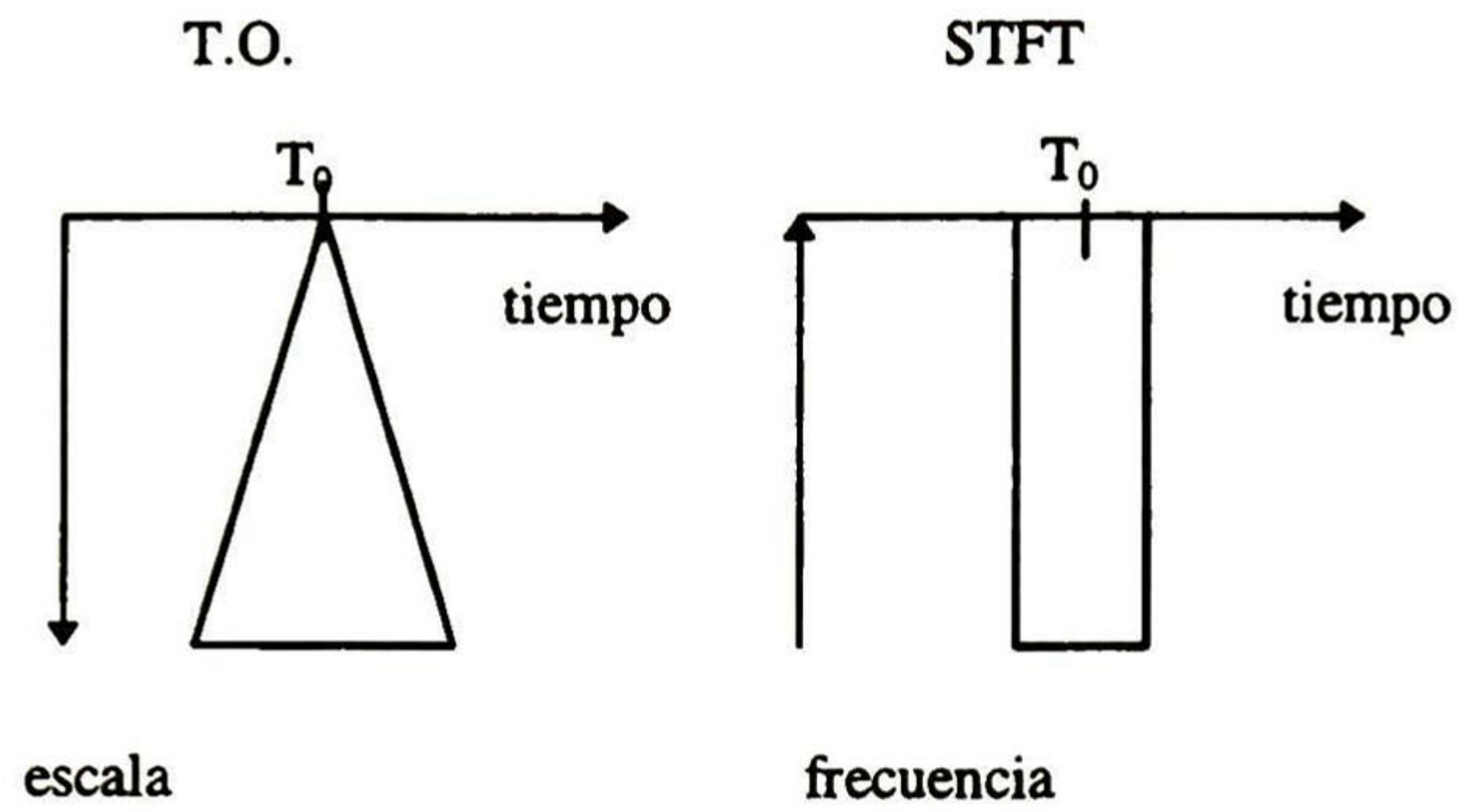


Figura 2.3 Escalograma y espectrograma de un impulso en el instante T_0 , en la gráfica se nota que cuando la escala decrece la frecuencia aumenta.

ANALISIS Y SINTESIS

Las respuestas al impulso $h_{a,\tau}(t) = \frac{1}{\sqrt{a}}h\left(\frac{t-\tau}{a}\right)$ para diferentes valores de a son funciones que oscilan en el tiempo y que además son de soporte finito o decaen rápidamente en el tiempo, por estas características se les llaman ondeletas. Se puede pensar en las ondeletas como una base de funciones para el espacio de señales $L_2(\mathbf{R})$ [4][13](para mayor referencia sobre espacios de señales y bases refiérase al apéndice B). En lo siguiente se considera que los conceptos de álgebra lineal, espacios y subespacios son familiares al lector.

La transformada es un producto interno, es decir, mide la similitud de la función $f(t)$ con la función $h_{a,\tau}(t) = \frac{1}{\sqrt{a}}h\left(\frac{t-\tau}{a}\right)$, con esto indica la aportación de esa banda a la señal original en la ventana de tiempo dada.

$$F_O(a, \tau) = \int_{-\infty}^{\infty} h_{a,\tau}(t) f(t) dt$$

Para la síntesis, se utiliza la fórmula de reconstrucción,

$$f(t) = c \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_O(a, \tau) h_{a,\tau}(t) \frac{dad\tau}{a^2}$$

Algunas consideraciones son necesarias para que la reconstrucción sea exitosa, como que:

$$\int_{-\infty}^{\infty} h(t) dt = 0, \text{ o lo que es equivalente [3], } \int_{-\infty}^{\infty} \frac{|H(w)|^2}{|w|} dw = \frac{1}{c} < +\infty$$

esto se deriva de que $h(t)$ debe tener energía finita y al menos su primer momento igual a cero, asegurando que la integral de la reconstrucción converge en L_2 . Al tener $h(t)$ energía finita la integral de reconstrucción puede sumar su energía en todas las frecuencias y la constante c permanece finita. (ver apéndice C)

Las funciones $h_{a,\tau}(t)$ no son ortogonales, porque al tener las escalas y los desplazamientos como variables continuas el producto interno de funciones cuyos parámetros a y τ estén muy cercanos no puede ser cero.

Esto significa que, la transformada de ondeleta con un valor de escala y desplazamiento representa una parte de la señal que el siguiente valor de escala y desplazamiento también están representando, o en otras palabras, el hecho de que la transformada sea continua provoca que exista mucha redundancia, lo que no sucede en el caso de Fourier. Sin embargo, con la transformada de ondeleta continua la reconstrucción de la señal original con error cero es posible (Apéndice C).

2.3 TRANSFORMADA DE ONDELETA DISCRETA Y MULTIRRESOLUCIÓN.

Las versiones continuas de la transformada de ondeleta se comportan como una base ortonormal sin serlo, por que la señal se puede reconstruir a partir de ellas y además la transformada conserva la energía, es decir, cumple con el teorema de Parseval [4], lo que nos hace pensar que es posible tomar solo algunos valores de los parámetros a y τ para formar una base ortogonal.

Una forma natural de hacerlo es en base a su relación con las frecuencias $\frac{\Delta f}{f} = c = \frac{a_1}{a_0}$ [13], tomando escalas que sigan esa relación se obtiene una discretización del espacio tiempo-frecuencia como se muestra en el lado derecho de la figura 2.4, solo para comparación se incluye la posible división del plano tiempo-frecuencia dado por la discretización de la STFT

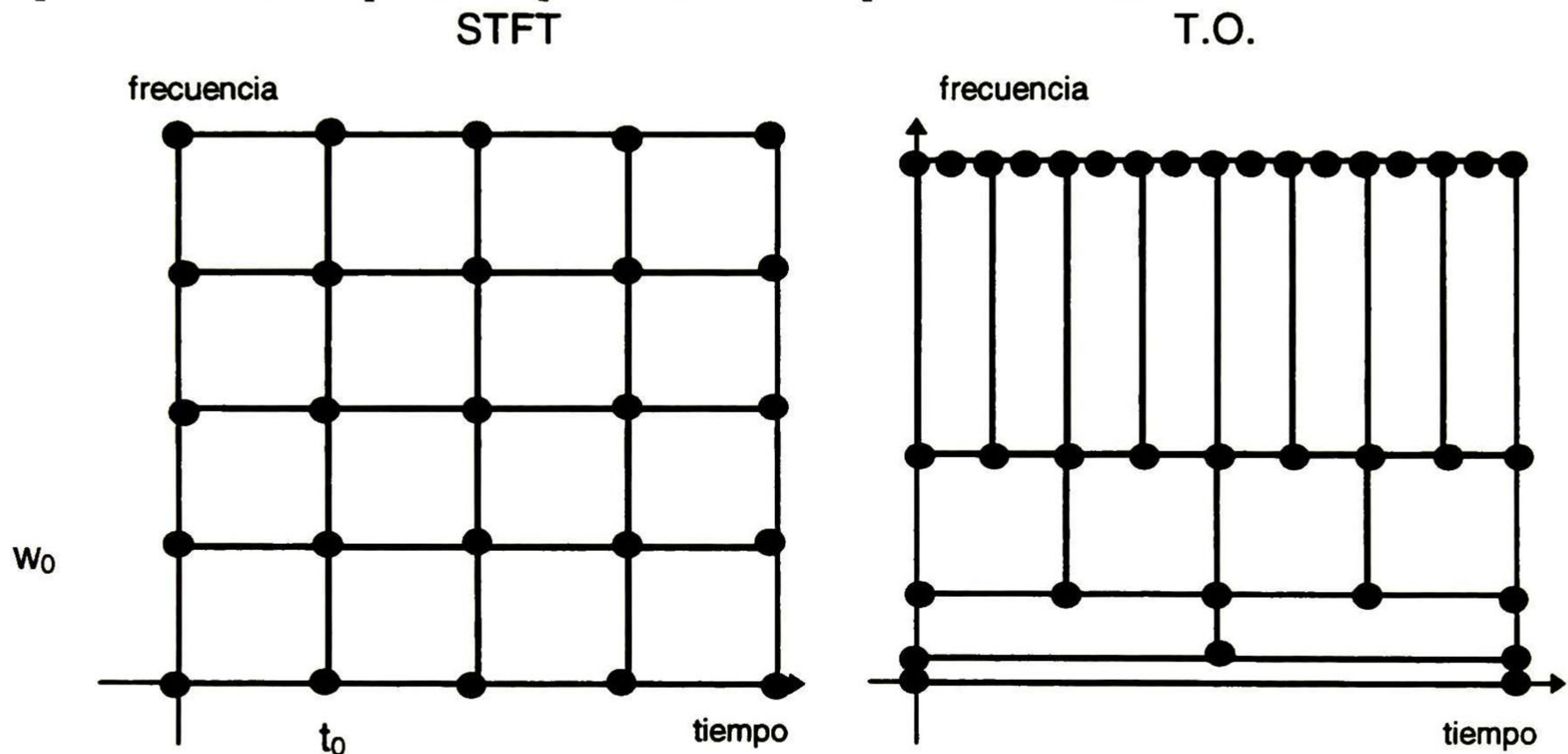


Figura 2.4 Discretización de los parámetros de tiempo y frecuencia para la STFT y de tiempo y escala para la transformada de ondeleta.

Con esta selección de escalas y desplazamientos resulta una base como:

$$h_{j,k}(t) = a_0^{-\frac{j}{2}} h(a_0^{-j} t - kT), \text{ donde resultan los coeficientes: } c_{j,k} = \int f(t) h_{j,k}(t) dt$$

Cada coeficiente $c_{j,k}$ representa la aportación de la función $h_{j,k}(t)$ en la señal que se está analizando $f(t)$.

Y la fórmula de reconstrucción: $f(t) \cong C \sum_j \sum_k c_{j,k} h_{j,k}(t)$. recolecta las contribuciones de cada coeficiente. En esta fórmula la constante C depende solo de la elección de $h(t)$ como en el caso de la transformada de ondeleta continua.

Si a_0 se acerca a 1, a_0^{-j} es casi igual a a_0^{-j-1} y la escala varía muy lentamente, por lo que tenemos más escalas de las que se necesitan para representar la señal, y se dice que la

representación es redundante. Como resultado sólo se ponen mínimas restricciones a la función $h(t)$.

Si a_0 aumenta (por ejemplo $a_0 = 2$), se obtiene $f(t)$ sólo para algunos casos de $h(t)$, pero la redundancia es menor o nula.

La teoría de marcos de ondeletas proporciona un ambiente de trabajo para representar ambos casos [4].

Si existe un conjunto de $h_{j,k}$ que cumple : $AE_x \leq \sum |c_{j,k}|^2 \leq BE_x$, donde E_x es la energía de la señal $f(t)$, entonces, este conjunto constituye un marco de ondeletas [4], donde las constantes A y B son los límites del marco y definen la perfección de la reconstrucción. En especial cuando $A=B$ se tiene una reconstrucción perfecta y las funciones $h_{j,k}$, a pesar de tener sólo un conjunto discreto de escalas forman una base ortogonal para el espacio L_2 , de esta forma $\{h_{j,k}\}$ es un conjunto no únicamente ortogonal sino también discreto y derivado de una sola función prototipo [4].

El siguiente paso en la implementación de la transformada de ondeleta de forma eficiente es la discretización en el tiempo, es decir, la entrada a la transformada es una secuencia de números que representa la realización en el tiempo de una señal. Las frecuencias de la señal están limitadas por la razón de muestreo según el Teorema de Nyquist.

Para poder comprender la transformada de ondeleta discreta es necesario introducir el concepto de escalamiento de secuencias de tiempo discreto [14]. Este concepto de escala es similar al anteriormente utilizado para la ondeleta continua, sin embargo, tiene características propias.

Si tenemos una secuencia $x[n]$ se puede obtener una secuencia 'dilatada' en tiempo discreto si se realiza un sobremuestreo seguido de un filtrado pasa bajas que interpole la señal.

Si se desea obtener una secuencia 'comprimida' se necesita pasar la secuencia por un filtro pasa bajas para después hacer un submuestreo.

Esto se muestra en forma gráfica en la figura 2.5, la señal x , con una escala $S(x)$ después de pasar por un 'dilatador' tiene una escala $2 \cdot S(x)$, la misma señal después de pasar por un 'compresor' tiene una escala $S(x)/2$.

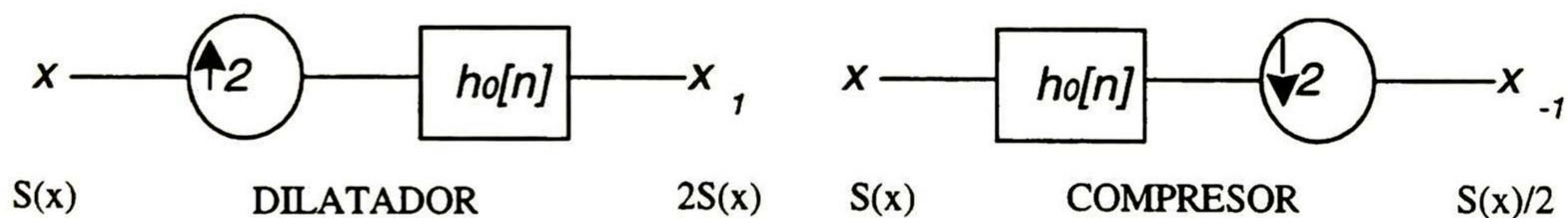


Figura 2.5 Operadores para cambiar de escala en tiempo discreto.

Al escalar una señal de tiempo discreto se deben de cumplir las siguientes características :

Las operaciones de dilatación y compresión son lineales.

- Los desplazamientos de tiempo se escalan de la misma manera, es decir, si $x(n)$ se dilata en escala 2 : 1 a $y(n)$, entonces, $x(n-k)$ se dilata a $y(n-2k)$

Las versiones escaladas tienen una evolución en el tiempo similar a la secuencia original, es decir, conservan la forma. Para que esto se cumpla es necesario que la señal que disminuye su escala tenga cierto grado de regularidad, [14].

La figura 2.6 muestra la dilatación de una señal progresivamente obteniendo señales que tienen una evolución en el tiempo similar a la original, pero con mayor número de muestras.

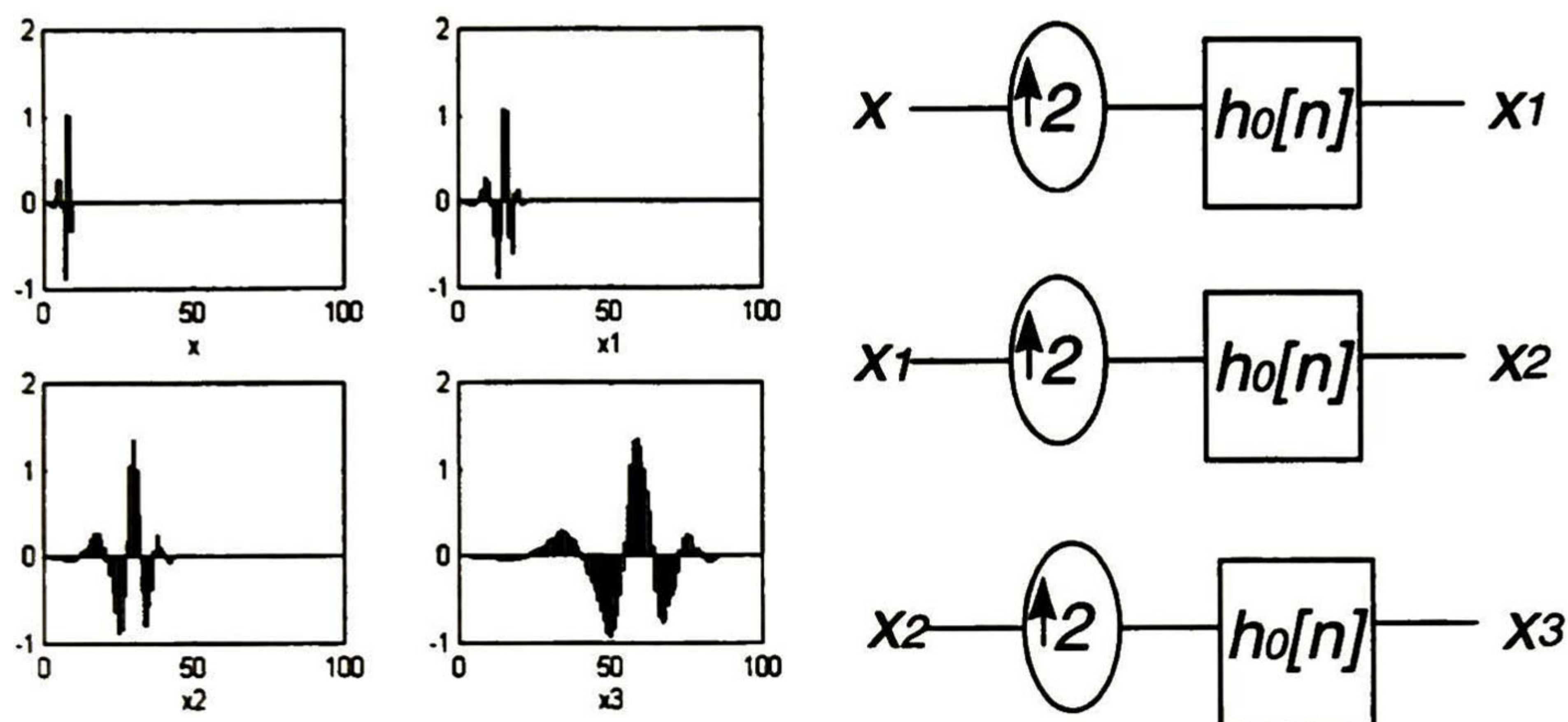


Figura 2.6 A la izquierda gráfica de una secuencia y sus versiones escaladas, a la derecha las operaciones necesarias para lograr el escalamiento de la secuencia.

Cuando se tiene una función $f(t)$ y se obtienen sus muestras en instantes de tiempo discretos $f(n)$, dada la secuencia original se puede utilizar la transformada de ondeleta discreta, para encontrar su descomposición en términos de una base equivalente de ondeletas discretas formadas por las secuencias $w_j(n-2^j m)$. Las funciones $w_j(n)$ son el producto del $j-1$ escalamiento en tiempo discreto de una secuencia base $h_1(n)$, con esto se tiene que $w_1(n) = h_1(n)$.

En caso de que no se efectúen un número infinito de dilataciones, las ondeletas no cubren todo el espectro de frecuencias. Esto se puede apreciar en la figura 2.7, si se considera que cada ondeleta representa un filtro pasabanda, al escalar el filtro se reduce su frecuencia central junto con su ancho de banda, pero en cada escalamiento siempre quedan un conjunto de frecuencias bajas que no pueden ser cubiertas por los filtros pasabandas.

Para representar estas frecuencias se introducen las secuencias de escalamiento $\phi_j(n-2^j m)$ que son respuestas al impulso de unos filtros pasa bajas obtenidas también como secuencias escaladas del filtro $g(n)$ [14].

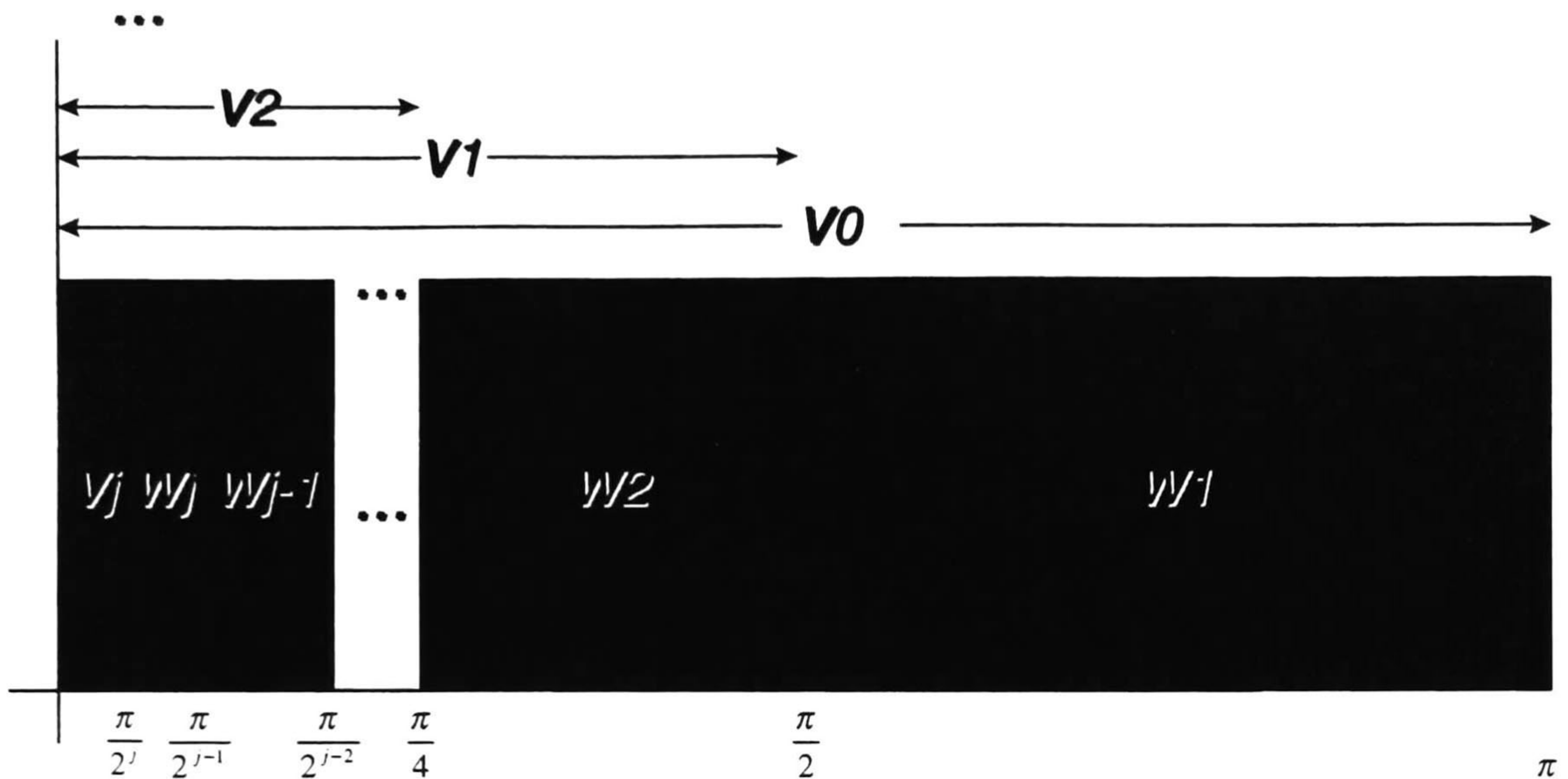


Figura 2.7 División ideal del espacio de frecuencias por las ondeletas.

Desde el punto de vista del análisis de multirresolución, cada conjunto de secuencias de escalamiento $\phi_j(n-2^j m)$ son una base para el subespacio de las señales pasa bajas con frecuencias menores a $2^j \pi$, así como las secuencias $w_j(n-2^j m)$ son una base para las señales pasa banda, y al realizar la transformada de ondeleta se está efectuando un análisis de diferentes resoluciones de la señal al descomponerla en varios subespacios de señales pasa banda y pasa bajas

Al tener un número finito de escalas aparece un conjunto de coeficientes $a(j,m)$, que representa las frecuencias bajas de la señal, además de los conjuntos $d(j,m)$ que representan las frecuencias en las j -ésima banda, entonces las fórmulas de análisis son [14]:

$$a(j,m) = \sum_n f(n) \phi_j(2^j n - m)$$

$$d(j,m) = \sum_n f(n) w_j(2^j n - m) \quad y$$

Y la sumatoria de la reconstrucción queda [14]:

$$f(n) = \sum_{j=1}^{L-1} \sum_m d(j,m) w'_j(n - 2^j m) + \sum_m a(L,m) \phi'_L(n - 2^L m)$$

Donde las funciones $\phi'_j(n-2^j m)$ y las funciones $w'_j(n-2^j m)$ son también bases de los espacios de señales pasa bajas y pasa banda, aunque no tienen que ser iguales a las bases de análisis. En caso de que no sean iguales las bases de análisis y reconstrucción se requiere que las bases de análisis sean ortogonales a las de reconstrucción formando un conjunto biortogonal [19][20].

Una expansión de este tipo proporciona un tipo de microscopio matemático o un análisis de múltiples resoluciones de la misma señal, cada vez que ' j ' se incrementa, la escala de la

ondeleta se duplica, la sumatoria se encarga de medir la similitud entre la señal y la ondeleta cada vez más dilatada, entonces los coeficientes $a(j,m)$ y $d(j,m)$ dan información de cierta escala de la señal alrededor de un intervalo de tiempo, por ejemplo : los cambios lentos sólo podrán verse para valores de 'j' grandes, mientras que los cambios rápidos son magnificados en los valores de 'j' pequeños.

El convolucionar la señal de escala $S(f)$ con un filtro que tiene una escala $2*s(w)$ es equivalente a convolucionar una señal de escala $S(f)/2$ con un filtro de escala $s(w)$, [14]; por esto en lugar de calcular los coeficientes dilatando la respuesta al impulso de los filtros, se calcula disminuyendo la escala de la señal de entrada.

Así en la práctica los resultados de la transformada de ondeleta se calculan de forma recursiva a partir de los coeficientes de la escala anterior, utilizando un eficiente algoritmo recursivo propuesto por Mallat, [11].

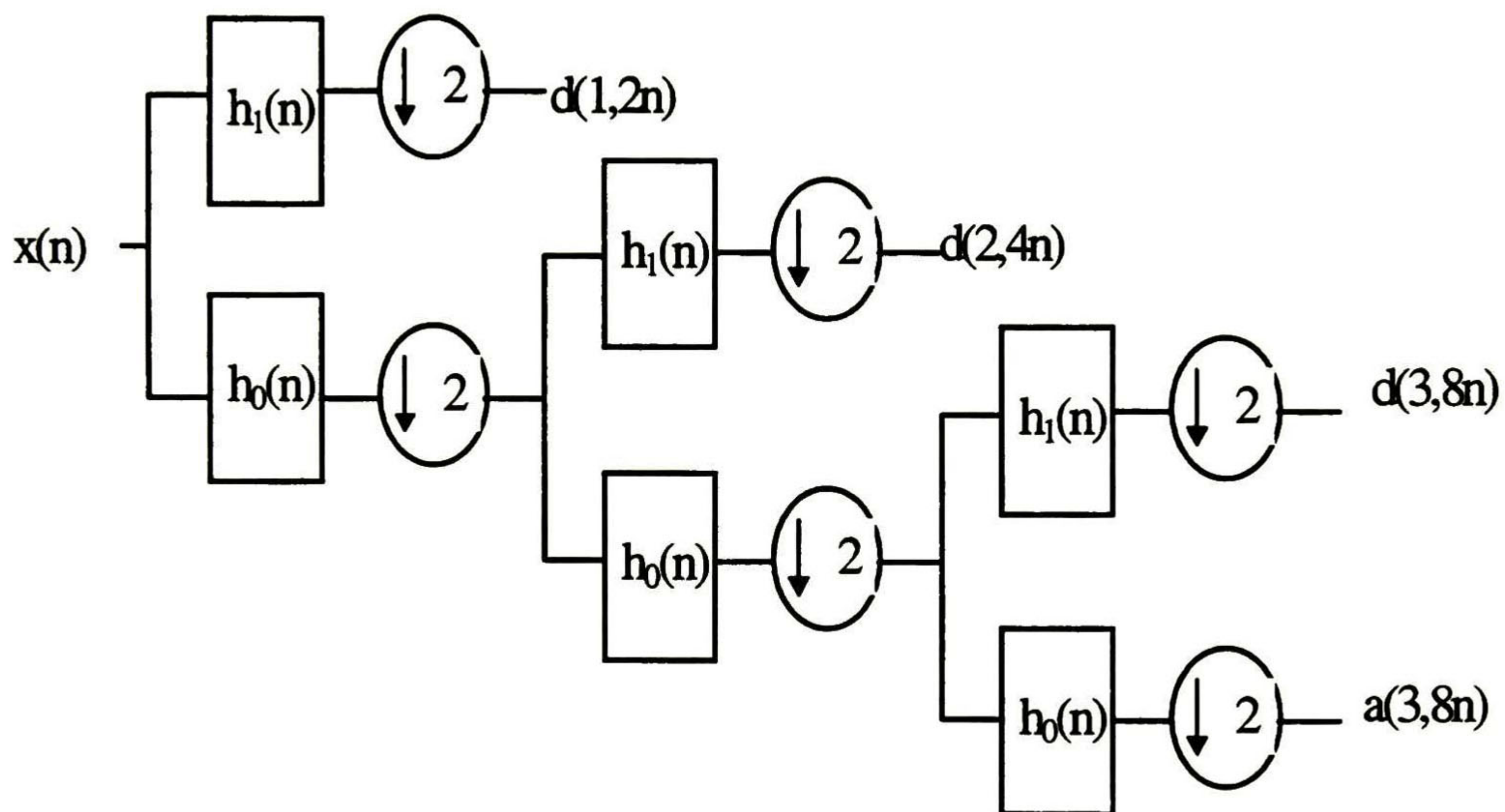


Figura 2.8 Banco de Filtros Iterado utilizado en el algoritmo de Mallat.

El número de operaciones necesarias para realizar la transformada de ondeleta discreta depende linealmente de L , donde L es el número de coeficientes en los filtros. Esto se debe a que cada filtro realiza la mitad de las operaciones que el filtro de nivel superior debido a la operación de decimado o submuestreo, entonces, si C_0 es la complejidad de cada filtro se tiene que la complejidad C es :

$$C = C_0 + \frac{C_0}{2} + \frac{C_0}{4} + \frac{C_0}{8} + \dots < 2C_0 \quad \text{y depende linealmente de } L \text{ porque } C_0 \propto L.$$

Por causa de los operadores de submuestreo, a la salida de la primera escala solo existen coeficientes con desplazamiento par, a la salida de la segunda escala solo existen coeficientes con desplazamiento múltiplo de 4, y así sucesivamente.

Se observa en la figura 2.8 que la implementación de la transformada de ondeleta es un banco de filtros, por lo que cada una de la escalas se relaciona con una parte del contenido de frecuencias de la señal original. Desde este punto de vista se ven las ondeletas como las respuestas al impulso de un banco de filtros y cada coeficiente en diferente escala representa información de un conjunto de frecuencias diferentes. La secuencia $w_j(n-2^j m)$ representa un filtro con una frecuencia central y un ancho de banda relacionado con su escala. De este modo la función básica $\phi_\varphi(n-2^j m)$ es un filtro pasa baja y las funciones $w_j(n-2^j m)$ son filtros pasa banda, (Figura 2.9).

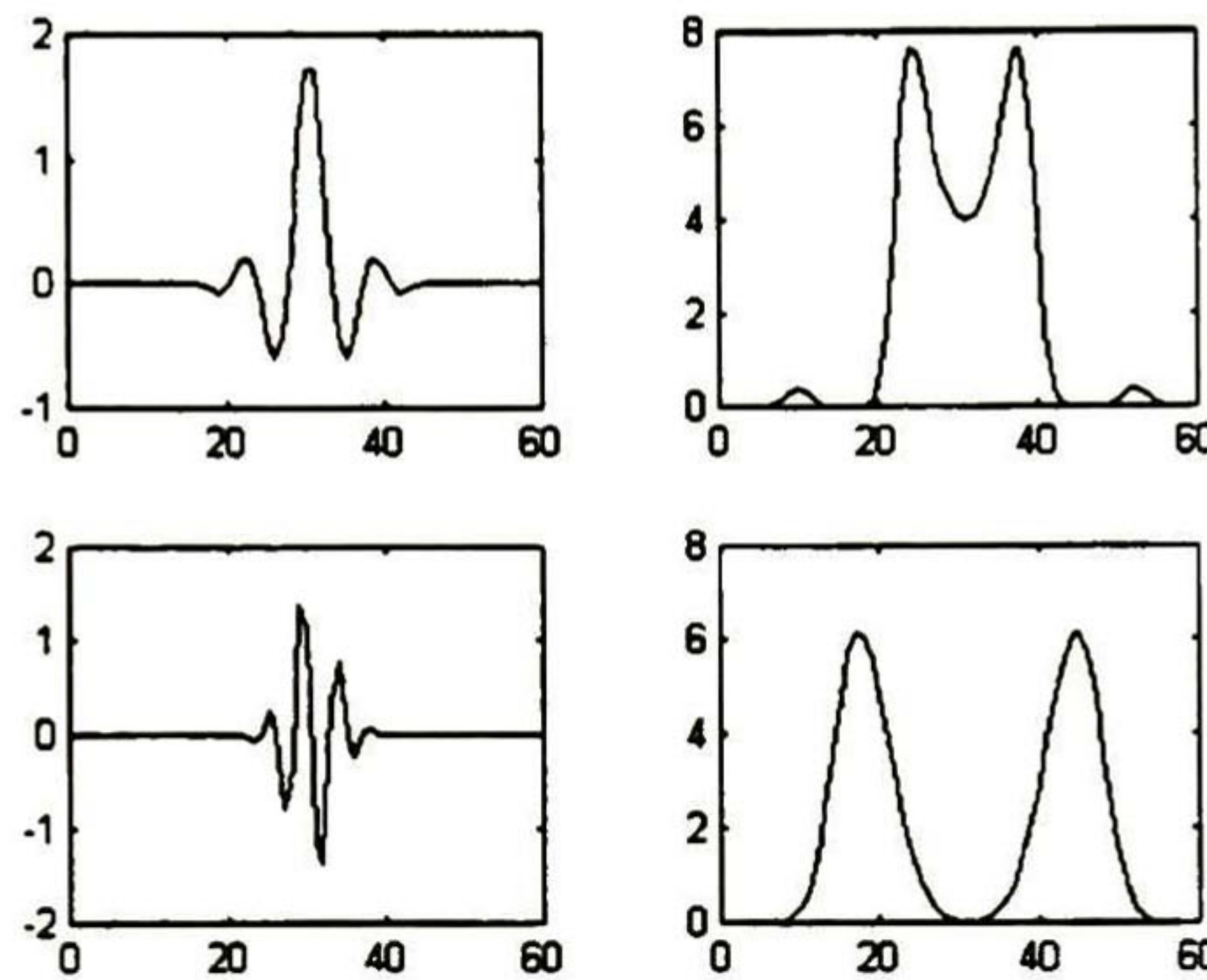


Figura 2.9. Función Básica $\phi_j(n)$, la ondeleta asociada $w_j(n)$ y sus Transformadas de Fourier.

Entonces es evidente que $d(j,m)$ da información acerca de $f(n)$ alrededor del intervalo de frecuencias entre $2^j \pi$ y $2^{j+1} \pi$ y el instante de tiempo $2^j m$, dicho de otra manera, proporciona información de los detalles de la señal a escala 2^j , por eso se les llama coeficientes de detalle. La información de los coeficientes $a(j,m)$, por estar asociada con la función $\phi(n)$, da información de las frecuencias bajas de $f(n)$ de 0 a $2^j \pi$ por lo que se puede decir que es una aproximación burda de la señal, y a los coeficientes $a(j,m)$ se les llama coeficientes de aproximación a la escala 2^j de la señal.

Si se restringe la atención a las ondeletas de soporte finito, entonces los coeficientes de los filtros que determinan tanto la función básica como la ondeleta asociada sólo son diferentes de cero para $0 < k < K-1$.

Los filtros que generan las ondeletas de tiempo discreto, de la misma forma que los de tiempo continuo, deben satisfacer ciertas condiciones que aseguren que la expansión se mantenga para cualquier función que pertenezca a $l_2(\mathbf{R})$, entonces los filtros cumplen con generar la familia de ondeletas como:

$$w_j(n) = \sum_{k=0}^{K-1} h_1[k] \phi_{j-1}[2n - k] \quad \text{y} \quad \phi_j[n] = \sum_{k=0}^{K-1} h_0[k] \phi_{j-1}[2n - k]$$

Para la reconstrucción se tienen las mismas condiciones, aunque la base de reconstrucción no tiene que ser igual a la de análisis, cuando son iguales se dice que las ondeletas forman una base ortogonal. Esta base de reconstrucción también se genera a partir de una secuencia básica y sus escalamientos.

$$w'_j(n) = \sum_{k=0}^{K-1} g_1[k] \phi'_{j-1}[2n-k] \quad \text{y} \quad \phi'_j[n] = \sum_{k=0}^{K-1} g_0[k] \phi'_{j-1}[2n-k]$$

Para que la reconstrucción sea sin error, los coeficientes de los filtros de análisis y reconstrucción deben de satisfacer un número de restricciones, que son equivalentes a que las ondeletas generadas sean biortogonales, [19].

$$\begin{aligned} \sum_{n=0}^{K-1} h_1(n) g_1(n-2l-2k) &= \delta(k-l) \\ \sum_{n=0}^{K-1} h_0(n) g_0(n-2l-2k) &= \delta(k-l) \\ \sum_{n=0}^{K-1} h_1(n) g_0(n-2l-2k) &= 0 \\ \sum_{n=0}^{K-1} h_0(n) g_1(n-2l-2k) &= 0 \end{aligned}$$

Note que las ecuaciones anteriores no aseguran que las funciones $w_j(n-2^j m)$ son ortogonales entre sí, pero asegura que son ortogonales a la base de reconstrucción $w'_j(n-2^j m)$ y lo mismo para las funciones básicas $\phi_j(n-2^j m)$ y $\phi'_j(n-2^j m)$, permitiendo que se tenga reconstrucción perfecta.

3. ONDELETAS Y BANCOS DE FILTROS.

3.1 PRELIMINARES.

La razón de ser de este capítulo es la de establecer la relación entre los bancos de filtros y las ondeletas, así como la de conocer las diferentes formas de calcular los coeficientes de los filtros que se utilizan en el algoritmo de Mallat.

La transformada de ondeleta y los bancos de filtros son dos campos íntimamente relacionados pero no son subconjuntos uno del otro, es decir, algunas transformadas de ondeleta se pueden implementar como bancos de filtros iterados con razones de muestreo múltiples, sin embargo no todos los bancos de filtros iterados multirate resultan en una descomposición en ondeletas.

Para esclarecer la relación entre los bancos de filtros y la transformada de ondeletas, se utilizan dos bases que en resolución en tiempo y resolución en frecuencia son casos extremos. Una de ellas es la expansión en funciones de Haar y la otra la expansión en términos de los senos cardinales o expansión de Sincs.

Expansión de Haar de Señales de Tiempo Discreto

La base de Haar, a pesar de ser muy simple, no deja de resaltar las características principales de las implementaciones con bancos de filtros, como la de invariancia periódica en el tiempo y el soporte finito. La unidad básica es una operación de diferencia y promedio de sólo dos puntos. Aunque es una operación muy sencilla, cuando se realiza resulta en una descomposición ortogonal de tiempo discreto. Una ventaja adicional es que esta descomposición tiene muy buena resolución en el tiempo ya que sus filtros son FIR de sólo dos coeficientes, entonces, es posible relacionar el tiempo de la salida con el de la entrada con una buena aproximación.

Este conjunto de funciones forman un conjunto que es capaz de representar una señal como una suma ponderada de las mismas funciones :

$$x[n] = \sum_k X[k]w_k[n],$$

Las funciones de la base de Haar están dadas por

$$w_{2k}[n] = \begin{cases} \frac{1}{\sqrt{2}} & n = 2k, 2k+1 \\ 0 & \text{para cualquier otro valor de } n \end{cases} \quad \text{y} \quad w_{2k+1}[n] = \begin{cases} \frac{1}{\sqrt{2}} & n = 2k, \\ \frac{-1}{\sqrt{2}} & n = 2k+1 \\ 0 & \text{para cualquier otro valor de } n \end{cases}$$

Esto es equivalente a definir las funciones de la base con índice impar como versiones trasladadas de ellas mismas, y lo mismo para las funciones con índice par.

$$w_{2k} = w_0(n-2k), \quad \text{y} \quad w_{2k+1} = w_1(n-2k)$$

Tomando la notación anterior, la descomposición se calcula al realizar el producto interno de estas funciones con las señal de entrada,

$$X[2k] = \sum x[n]w_{2k}[n] = \frac{1}{\sqrt{2}}(x[2k] + x[2k + 1]) \quad y$$

$$X[2k + 1] = \sum x[n]w_{2k+1}[n] = \frac{1}{\sqrt{2}}(x[2k] - x[2k + 1])$$

Debido a la relación entre la convolución y el producto punto, esta descomposición se puede implementar utilizando convoluciones, y por lo tanto lleva a un banco de filtros. Considerando el filtro $h_0[n]$ con la siguiente respuesta al impulso :

$$h_0[n] = \begin{cases} \frac{1}{\sqrt{2}} & n = -1, 0, \\ 0 & \text{para otros} \\ & \text{valores de } n. \end{cases}$$

y efectuando una convolución con la señal de entrada, se obtiene $X[2k]$ como resultado en el instante $2k$ porque,

$$h_0[n] * x[n] \Big|_{n=2k} = \sum_l h_0[2k - l]x[l] = \frac{1}{\sqrt{2}}(x[2k] + x[2k + 1]) = X[2k]$$

De manera similar, se puede definir un filtro $h_1[n]$ con respuesta al impulso de forma que su convolución entregue los coeficientes impares:

$$h_1[n] = \begin{cases} \frac{1}{\sqrt{2}} & n = 0, \\ -\frac{1}{\sqrt{2}} & n = -1, \\ 0 & \text{para otros} \\ & \text{valores de } n. \end{cases}$$

utilizando este filtro se obtiene $X[2k+1]$,

$$h_1[n] * x[n] \Big|_{n=2k} = \sum_l h_1[2k - l]x[l] = \frac{1}{\sqrt{2}}(x[2k] - x[2k + 1]) = X[2k + 1]$$

Evaluar la convolución en los índices pares es equivalente a filtrar la señal seguido por un submuestreo tomando una de cada dos muestras, entonces, $X[2k]$ y $X[2k+1]$ se pueden obtener a partir de un banco de filtros de 2 canales, con filtros $h_0[n]$ y $h_1[n]$. A esto se le llama un banco de filtros de análisis, como el que se presenta en la Figura 3.1.

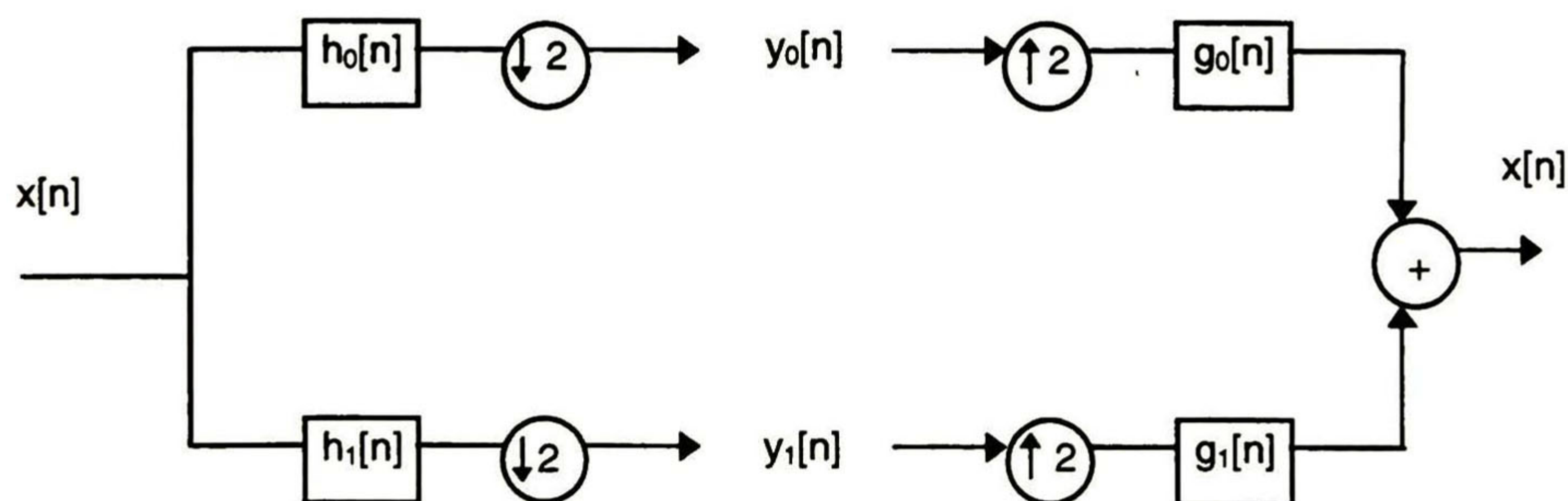


Figura 3.1 de los bancos de Filtros de dos canales

A las salidas de los decimadores por 2 se les denomina también $y_0[k]$ y $y_1[k]$ respectivamente en la nomenclatura de los bancos de filtros.

Es importante notar que las repuestas al impulso del filtro de análisis, son las mismas funciones de la base pero invertidas en el tiempo, es decir, $h_0[n] = w_0[-n]$, y $h_1[n] = w_1[-n]$, porque la convolución es un producto interno que involucra una inversión de la función en el tiempo.

Además en este ejemplo se puede verificar la invariancia periódica respecto a los desplazamientos pares. Si $x'[n] = x[n-2l]$, entonces

$$X'[2k] = \frac{1}{\sqrt{2}}(x'[2k] + x'[2k+1]) = \frac{1}{\sqrt{2}}(x[2k-2l] + x[2k+1-2l]) = X[2k-2l]$$

de la misma forma se puede probar que esto no se cumple con respecto a los desplazamientos impares, de donde se afirma que esta descomposición no es invariante en el tiempo.

Para reconstruir la señal se puede hacer exactamente lo mismo, sólo que los filtros en este caso son iguales a las funciones de la base.

$$g_0[n] = w_0[n], \quad g_1[n] = w_1[n], \quad \text{entonces}$$

$$x[n] = \sum_k y_0[k]g_0[n-2k] + \sum_k y_1[k]g_1[n-2k]$$

Se debe comentar una propiedad más de esta descomposición. Como se mencionó anteriormente, un canal calcula el promedio y otro la diferencia entre dos muestras consecutivas, y a pesar de que estas operaciones no son los mejores filtros pasa bajas y pasa altas, llevan a una importante interpretación: la reconstrucción a partir de $y_0[k]$ y $y_1[k]$, se forma sumado una versión burda de la señal o de menos ancho de banda y una versión de detalle que cuantifica los cambios entre una muestra y otra, por lo que contiene las componentes de mayor frecuencia de la señal, haciendo posible recuperar la señal perfectamente.

Expansión de Sincs de Señales de Tiempo Discreto

Aunque notablemente simple, la base de Haar sufre de un importante inconveniente, su resolución en frecuencia no es muy buena, porque su respuesta al impulso es un rectángulo y su transformada de Fourier contiene lóbulos en todas las frecuencias. En su lugar se puede usar una base que tenga una respuesta en frecuencia ideal, es decir, que tenga cero como respuesta en la banda de rechazo aunque es claro que la resolución en tiempo se va a deteriorar, como se verá, la respuesta al impulso es infinita y decae sólo proporcionalmente a $1/n$.

Esta base, se puede construir comenzando con un filtro pasa bajas $g_0[n]$ definido por su transformada discreta de Fourier $G_0(e^{j\omega}) = \sqrt{2}, \omega \in [-\pi/2, \pi/2]$ y 0 para $\omega \in [\pi/2, 3\pi/2]$.

Utilizando la transformada inversa da,

$$g_0[n] = \frac{\sqrt{2}}{2\pi} \int_{-\pi/2}^{\pi/2} e^{j\omega n} d\omega = \frac{1}{\sqrt{2}} \frac{\sin \pi n / 2}{\pi n / 2}$$

Para el filtro pasa altas, se puede tomar una versión modulada de éste, invertida en el tiempo y desplazada en uno para lograr incluir la frecuencia $\omega = \pi/2$, de este modo $g_1[n] = (-1)^n g_0[-n+1]$.

Como en el caso de la base de Haar, la base de funciones se obtiene de las respuestas al impulso y sus desplazamientos pares, $g_0[n-2k]=w_{2k}[n]$, $g_1[n-2k]=w_{2k+1}[n]$. Los coeficientes de la expansión se pueden encontrar con un par de filtros $h_i[n]=g_i[-n]$ seguidos de la operación de submuestreo por 2.

Desde un punto de vista de multirresolución, la operación de filtrado con $h_0[n]$, seguido por submuestreo, después el sobremuestreo y la interpolación con $g_0[n]$ da como resultado la proyección de la señal en un subespacio de señales limitadas en banda en el intervalo $\omega \in [-\pi/2, \pi/2]$ dada por x_0 . De forma similar, la otra ruta de la señal da como resultado la proyección de la señal en el subespacio de las señales pasa altas, resultando x_1 . Los dos subespacios son ortogonales y su suma es $l_2(\mathbf{Z})$. También es claro que x_0 es una señal burda o una aproximación de la señal x , mientras que x_1 contiene las frecuencias adicionales para reconstruir x a partir de x_0 .

La descomposición de una señal vista en forma gráfica se puede observar en la Figura 3.2, así como la forma en que se da la reconstrucción. Este proceso se puede observar en las fórmulas utilizando el análisis de Fourier.

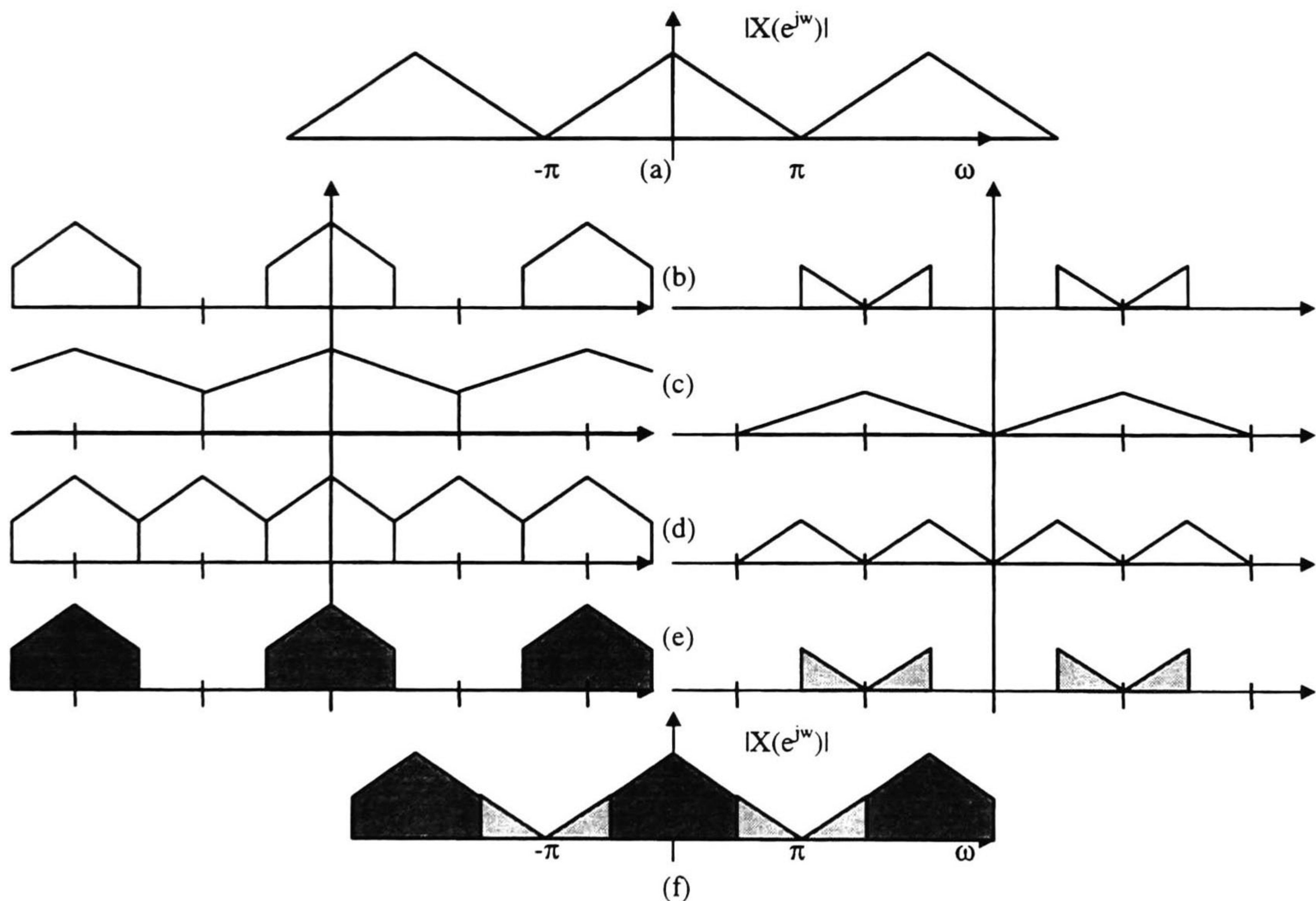


Figura 3.2 Descomposición de dos canales de una señal utilizando filtros ideales. A la izquierda la ruta del pasa bajas, a la derecha la ruta del pasa altas. (a) Espectro Original. (b) Espectro después del filtrado. (c) Espectro después de submuestreo. (d) Espectro después de sobremuestreo. (e) Espectro después de la interpolación. (f) Espectro reconstruido.

Ambas descomposiciones llevan a filtros que son variantes en el tiempo. De hecho, el ser variantes en el tiempo es una cualidad inherente de las expansiones ortogonales.

En estos casos se ha visto casos extremos en la resolución de tiempo y de frecuencia basados en los filtros de Haar y sinc respectivamente. Existen casos más interesantes entre estos dos casos extremos implementados con bancos de filtros como los de la figura 3.1, pero, estos filtros deben cumplir con algunas restricciones. Para encontrar los coeficientes de los filtros que llevan a tales descomposiciones se utilizan varios tipos de análisis como :

Análisis en el dominio del tiempo.

Análisis en el dominio de la frecuencia o modulación.

Análisis en el dominio de la transformada de polifase.

A continuación se menciona someramente las bases de éstos. Para mayor referencia consultar el libro de Vetterli y Kovacevic, [20]

ANALISIS EN EL DOMINIO DEL TIEMPO.

Los bancos de filtros se pueden representar de varias maneras, una de ellas es como una matriz infinita que realiza una transformación sobre un vector también infinito, que representa la señal de entrada.

$$\begin{pmatrix} y_0[0] \\ y_1[0] \\ y_0[1] \\ y_1[1] \end{pmatrix} = \begin{pmatrix} X[0] \\ X[1] \\ X[2] \\ X[3] \end{pmatrix} = T_a \cdot \begin{pmatrix} x[0] \\ x[1] \\ x[2] \\ x[3] \end{pmatrix}$$

$$y = X = T_a x$$

Por esto, los coeficientes de la transformada se representan también como un vector infinito. La matriz de análisis T_a es la que contiene los coeficientes de los filtros, y puede expresarse como :

$$T_a = \begin{pmatrix} h_0[L-1] & h_0[L-2] & h_0[L-3] & \dots & h_0[0] & 0 & 0 \\ h_1[L-1] & h_1[L-2] & h_1[L-3] & \dots & h_1[0] & 0 & 0 \\ 0 & 0 & h_0[L-1] & \dots & h_0[2] & h_0[1] & h_0[0] \\ 0 & 0 & h_1[L-1] & \dots & h_1[2] & h_1[1] & h_1[0] \end{pmatrix}$$

Donde h_x son los coeficientes de los filtros y L es la longitud de los mismos.

Para sintetizar la señal se utiliza otra matriz de síntesis T_s , que cumpla con

$$x = T_s y = T_s X = T_s T_a x = x.$$

Definida similarmente a T_a como :

$$T_s^T = \begin{pmatrix} g_0[0] & g_0[1] & g_0[2] & \dots & g_0[L'-1] & 0 & 0 \\ g_1[0] & g_1[1] & g_1[2] & \dots & g_1[L'-1] & 0 & 0 \\ 0 & 0 & g_0[0] & \dots & g_0[L'-3] & g_0[L'-2] & g_0[L'-1] \\ 0 & 0 & g_1[0] & \dots & g_1[L'-3] & g_1[L'-2] & g_1[L'-1] \end{pmatrix}$$

Como en el caso anterior, g_x son los coeficientes de los filtros y L' es su longitud .

Entonces desde este punto de vista, como la base de funciones está formada por los coeficientes de los filtros, que son las filas de las matrices, la condición de reconstrucción expresada anteriormente implica que :

$$T_a T_s = T_s T_a = I$$

En término de las respuestas al impulso de los filtros se pueden construir unas matrices de forma que expresen lo que ocurre por cada una de las rutas de la señal, es decir,

$$\begin{pmatrix} y_i[0] \\ y_i[1] \end{pmatrix} = \begin{pmatrix} \dots h_i[L-1] & h_i[L-2] & h_i[L-3] \dots \\ \dots & 0 & 0 & h_i[L-1] \dots \end{pmatrix} \begin{pmatrix} x[0] \\ x[1] \end{pmatrix}$$

$y_i \qquad \qquad H_i \qquad \qquad x$

Y definiendo a G_i^T de forma similar pero con los coeficientes de los filtros sin inversión en el tiempo, la salida del sistema puede escribirse como :

$$(G_0 H_0 + G_1 H_1) x$$

y para que la señal sintetizada sea igual a la original (condición de reconstrucción perfecta) es necesario que :

$$G_0 H_0 + G_1 H_1 = I.$$

Por supuesto intercalando los renglones de H_0 y H_1 se forma la matriz T_a e intercalando las columnas de G_0 y G_1 se forma T_s

3.3 ANALISIS EN EL DOMINIO DE LA MODULACION.

Otra forma de representar una señal es por medio de su transformada de Fourier, o en forma mas general, de su transformada Z. Utilizando las propiedades de la transformada Z, la señal después de pasar por el submuestreo, el filtrado y el sobremuestreo, da :

$$X_0(z) = \frac{1}{2} G_0(z) [H_0(z) X(z) + H_0(-z) X(-z)]$$

$$X_1(z) = \frac{1}{2} G_1(z) [H_1(z) X(z) + H_1(-z) X(-z)]$$

para las señales de cada una de las bandas, donde la suma de ambas da la señal reconstruida. Expresando esto en forma de matriz y utilizando la matriz de modulación de análisis $H_m(z)$, se tiene la siguiente relación.

$$X_R(z) =$$

$$X_0(z) + X_1(z) = \frac{1}{2} \begin{pmatrix} G_0(z) & G_1(z) \end{pmatrix} \begin{pmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{pmatrix} \begin{pmatrix} X(z) \\ X(-z) \end{pmatrix}$$

$$\mathbf{H}_m(z) \quad \mathbf{x}_m(z)$$

Para que la señal reconstruida sea igual a la original es necesario que se cumpla :

$$G_0(z)H_0(z) + G_1(z)H_1(z) = 2$$

$$G_0(z)H_0(-z) + G_1(z)H_1(-z) = 0$$

En otras palabras, que los filtros de reconstrucción complementen en frecuencia a los filtros de análisis y además eliminen los componentes de alias en frecuencia introducidos en el submuestreo.

De las ecuaciones anteriores denominando $P(z)$ al producto de $G_0(z)$ y $H_0(z)$ se tiene que,

$$G_0(-z) H_0(-z) = G_1(z) H_1(z) = P(-z)$$

y entonces :

$$P(z) + P(-z) = 2,$$

esta ecuación tiene una importancia especial en el diseño de filtros que lleva a la descomposición en ondeletas. A los polinomios $P(z)$ que cumplen con la ecuación anterior se les llama *válidos* y tienen la característica de que sus coeficientes pares son cero excepto para $p[0]$ que es igual a 1.

3.4 ANALISIS EN EL DOMINIO DE POLIFASE.

El método de análisis anterior aunque proporciona ciertas herramientas para el diseño de los filtros que llevan a la descomposición en ondeletas y es una representación natural del proceso, tiene el inconveniente de ser redundante, ya que los coeficientes de los filtros están repetidos en sus versiones moduladas. Un punto de vista más compacto es el que se tiene con la representación de polifase, la cual consiste en separar la señal en componentes submuestreados que poseen un retraso con respecto a los demás, a esto se le llama la transformada de polifase :

$$x[n] = x_0[n] + z^{-1} x_1[n] + \dots \quad \text{donde : } x_i[n] = x[2n+i].$$

En el caso de dos componentes, la transformada Z de la transformación de polifase da :

$$X(z) = X_0(z^2) + z^{-1} X_1(z^2).$$

Entonces por conveniencia se introduce en notación de matrices la obtención de las dos señales intermedias como :

$$\begin{pmatrix} Y_0(z) \\ Y_1(z) \end{pmatrix} = \begin{pmatrix} H_{00}(z) & H_{01}(z) \\ H_{10}(z) & H_{11}(z) \end{pmatrix} \begin{pmatrix} X_0(z) \\ X_1(z) \end{pmatrix}$$

$$\mathbf{y}_p(z) \quad \mathbf{H}_p(z) \quad \mathbf{x}_p(z)$$

donde $H_{ij}(z)$ es el componente j -ésimo de la descomposición de polifase de H_i , es decir,
 $H_i(z) = H_{i_0}(z^2) + z^{-1}H_{i_1}(z^2)$

La matriz $\mathbf{H}_p(z)$ se denomina matriz de análisis de polifase.

En la parte de la síntesis, el sistema se puede analizar en forma similar. Puede implementarse con una transformada de polifase inversa precedida por una matriz de polifase $\mathbf{G}_p(z)$ definida por :

$$\mathbf{G}_p(z) = \begin{pmatrix} G_{00}(z) & G_{01}(z) \\ G_{10}(z) & G_{11}(z) \end{pmatrix}$$

donde,

$$G_i(z) = G_{i_0}(z^2) + z^{-1}G_{i_1}(z^2)$$

Utilizando esta notación, la señal reconstruida queda como :

$$\mathbf{X}_R(z) = \begin{pmatrix} 1 & z^{-1} \end{pmatrix} \begin{pmatrix} G_{00}(z^2) & G_{10}(z^2) \\ G_{01}(z^2) & G_{11}(z^2) \end{pmatrix} \begin{pmatrix} Y_0(z^2) \\ Y_1(z^2) \end{pmatrix}$$

sustituyendo $y(z)$:

$$\mathbf{X}_R(z) = \begin{pmatrix} 1 & z^{-1} \end{pmatrix} \begin{pmatrix} G_{00}(z^2) & G_{10}(z^2) \\ G_{01}(z^2) & G_{11}(z^2) \end{pmatrix} \begin{pmatrix} H_{00}(z^2) & H_{01}(z^2) \\ H_{10}(z^2) & H_{11}(z^2) \end{pmatrix} \begin{pmatrix} X_0(z^2) \\ X_1(z^2) \end{pmatrix}$$

por lo que si ,

$$\begin{pmatrix} G_{00}(z) & G_{10}(z) \\ G_{01}(z) & G_{11}(z) \end{pmatrix} \begin{pmatrix} H_{00}(z) & H_{01}(z) \\ H_{10}(z) & H_{11}(z) \end{pmatrix} = \mathbf{I}$$

entonces la señal reconstruida es igual a la señal original.

3.5. DISEÑO DE FILTROS

En esta sección se ven los procedimientos basados en estas herramientas para el diseño de filtros que lleven a transformadas de ondeleta .

Filtros de Espejo en Cuadratura (QMF).

Tomando la relación del análisis de modulación que asegura que se cancelan los componentes alias,

$$G_0(z)H_0(-z) + G_1(z)H_1(-z) = 0$$

si se utiliza un sólo filtro modulado para H_0 y H_1 entonces se puede conseguir eliminar los alias utilizando los mismos filtros en la reconstrucción, también con su versión modulada, pero con signo contrario, es decir :

$$H_0(-z) = H_1(z),$$

$$G_0(z) = H_0(z),$$

$$G_1(z) = -H_1(z) = -H_0(-z).$$

y el diseño se reduce a encontrar un sólo filtro. Pero para que la reconstrucción sea perfecta, considerando un retraso en la señal se debe de cumplir que :

$$H_0(z)G_0(z) + H_1(z)G_1(z) = 2z^{-l}$$

Considerando la solución de los QMF se tiene que :

$$H_0^2(z) - H_0^2(-z) = 2z^{-l}$$

Es de esta ecuación de donde les viene el nombre ya que en el círculo unitario $H(-z)$ es la versión reflejada de $H(z)$, y ambos están al cuadrado. Este tipo de filtros no es posible implementarlos con fase lineal excepto para el caso trivial de los filtros de Haar, esto es porque para filtros con fase lineal los componentes en polifase H_{00} y H_{01} de H_0 están relacionados como $H_{00}(z) = z^{-L/2+1}H_{01}(z^{-1})$, donde L es la longitud del filtro, y esta ecuación junto con la solución para filtros QMF nos dan $H_{00} = 1/2z^{-n}$ que no puede cumplirse si H_0 es FIR y sus coeficientes son reales mas que para los filtros Haar [20]. Para filtros de dimensión mayor sólo se tienen aproximaciones. (ver la tabla 3.1)

Filtros de Daubechies :

Daubechies [4] propuso que los filtros que efectuaran una descomposición de este tipo deberían llevar a una base de ondeletas continuas, por lo que propuso los filtros con mayor regularidad, es decir, que posean el mayor número de ceros en $\omega = \pi$, para diseñarlos propuso filtros de forma que la función de autocorrelación fuese igual a $P(z) = (1+z^{-1})^k(1+z)^kR(z)$ y encontrar las condiciones para que $P(z) + P(-z) = 2$. Así encontró toda una familia de filtros de respuesta al impulso finita (FIR) que producen descomposición en ondeletas. (ver la tabla 3.1)

Filtros Biortogonales :

Aunque no es posible encontrar filtros FIR con reconstrucción perfecta, que tengan fase lineal, y que den una descomposición en una base ortogonal, relajando la condición de que se forme una base ortogonal, se puede obtener una descomposición con diferentes filtros para la reconstrucción y análisis que posean fase lineal, y que sean filtros FIR.

Vetterli [19] propuso tomar como referencia un filtro H_0 FIR tal que sus componentes de la descomposición de polifase o el mismo con su versión modulada cumplieran con ser polinomios que no tuvieran factores comunes excepto z^{-1} , es decir, que su máximo común divisor fuera z^{-1}

Esto garantiza que existe un filtro H_1 , que cumple con la condición de reconstrucción. Se puede elegir un filtro inicial, $F(z) = (1+z^{-1})^{2k}$ que cumple con el requisito para toda k , y buscar el filtro $R(z)$ de fase lineal, tal que $P(z) = F(z)R(z)$ cumple con que $P(z) + P(-z) = 2$. Y resolver la ecuación para los componentes pares de $P(z)$. Otro dato conocido es la longitud de $R(z)$, éste está restringido a ser de longitud par, ya que $P(z)$ es par y $F(z)$ también es par. Una vez definidos los filtros de análisis los de síntesis quedan especificados por las relaciones del análisis en modulación o en polifase. (ver la tabla 3.1)

Otros tipos de filtros :

Otros tipos de ondeletas se calculan utilizando directamente la convolución sobre señales discretizadas que provienen de ondeletas continuas, tal es el caso de los filtros de Meyer, Morlet y otros.

Smith and Barnwell	Daubechies	Biort. 3.3 pasa bajas	Biort 3.3 pasa altas
0.04935260	0.23037781	0.06629126	0
-0.01553230	0.71484657	-0.1988737	0
-0.08890390	0.63088076	-0.1546796	-0.17677669
0.31665300	-0.02798376	0.9943689	0.53033008
0.78751500	-0.18703481	0.9943689	-0.53033008
0.50625500	0.03084138	-0.1546796	0.17677669
-0.03380010	0.03288301	-0.1988737	0
-0.10739700	-0.01059740	0.0662912	0

Tabla 3.1 Coeficientes de la respuesta al impulso de algunos filtros que generan descomposición en ondeletas.[20]

NOTACION CON MATRICES Y VECTORES

Siendo los datos de entrada x y salida \bullet considerados como un vector columna para señales de una dimensión (1-D), se tiene la siguiente relación :

$$\bullet = Ax \quad x = B \bullet ; \quad B = A^{-1}$$

Donde A es la matriz de la transformada y B es la matriz de la Transformada Inversa.

Tomando un caso sencillo : $A = B = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ donde los vectores que forman la base de la transformada (base de Haar) son las columnas de la matriz de transformación.

Al realizar la transformada inversa se tiene,

$$x(0) = b_{00}\Theta(0) + b_{01}\Theta(1); \quad y \quad x(1) = b_{10}\Theta(0) + b_{11}\Theta(1);$$

entonces la transformada inversa, $x = B \bullet$, tiene una interpretación explícita o representa a x en término de la base de vectores que forma la matriz B . En forma vectorial esta relación queda :

$$\mathbf{x} = \mathbf{b}_0\Theta(0) + \mathbf{b}_1\Theta(1);$$

De esta forma queda claro que el vector x es la suma ponderada de los dos vectores base y que los pesos de esta suma son los coeficientes de la transformada.

Sin embargo, la codificación por entropía necesita de una versión cuantizada de los resultados de la transformada, $u(i) = Q[\Theta(i)]$, y entonces la salida queda como :

$$\mathbf{y} = \mathbf{b}_0Q[\Theta(0)] + \mathbf{b}_1Q[\Theta(1)] = u(0)\mathbf{b}_0 + u(1)\mathbf{b}_1.$$

Es de esperarse que x sea diferente de y , por lo que se espera un error de reconstrucción : $e = x - y$, que se puede utilizar para medir la calidad de la reconstrucción.

Los bloques de codificación y decodificación por entropía son transparentes en el proceso, y son utilizados para disminuir al máximo la tasa de transmisión.

Obviamente estos resultados se pueden extender fácilmente a N dimensiones, por ejemplo, en 1-D se puede definir un vector y dado por la multiplicación :

$$\Theta = A x$$

donde el vector x lo forman los datos de entrada $x = [x(0), x(1), x(2), \dots, x(N-1)]^T$, entonces la transformación, está dada como , $\Theta(k) = \sum_{n=0}^{N-1} x(n)a(k, n)$ para $k = 0, 1, \dots, N - 1$.

Los coeficientes $a(k,n)$ son los elementos de la k -ésima columna y el n -ésimo renglón de la matriz \mathbf{A} .

La transformada inversa se realiza como :

$$x(n) = \sum_{k=0}^{N-1} \Theta(k)b(k,n) \quad \text{para } n = 0,1,\dots, N-1.$$

Los coeficientes $b(k,n)$ son los elementos de la k -ésima columna y el n -ésimo renglón de la matriz \mathbf{B} .

Si la matriz \mathbf{A} satisface que $\mathbf{A}^T = \mathbf{A}^{-1}$, es decir, es una transformación ortogonal, entonces, $\mathbf{B}=\mathbf{A}^T$. Esta transformación es de mucha utilidad porque cumple con muchas propiedades, por ejemplo: se conserva la energía en el dominio de la transformada y el error cuadrático medio también se conserva en el dominio de la transformada, lo que significa que si se eliminan algunos coeficientes o se cuantizan el error que esto causa no se magnifica cuando se aplica la transformada inversa.

Para datos en dos dimensiones (2-D) como las imágenes, los datos de entrada y salida serían matrices de $N \times N$ y así se pueden formar tensores asociados a datos de mayores dimensiones.

4.2 TRANSFORMACION OPTIMA.

La codificación inicia con un cambio en la representación de los datos, es decir, con una transformación de los mismos, que tiene como finalidad obtener coeficientes que concentren la mayor cantidad de la energía para que la cuantización se realice de forma más eficiente.

Pero existen muchas transformaciones que se pueden aplicar a los datos de entrada. La transformada que se busca debe disminuir la redundancia de los datos, es decir, debe hacer cero la autocorrelación entre las diferentes muestras, y al mismo tiempo, debe concentrar la mayor información en el menor número de coeficientes. El efecto deseado se logra seleccionando la transformada que logre diagonalizar la matriz de autocorrelación de los datos, a esta transformación se le llama de Karhunen-Loeve de tiempo discreto, y se define como la matriz \mathbf{T} formada por los vectores propios de la matriz de autocorrelación, ordenados de tal forma que sus valores propios asociados estén ordenados de mayor a menor, es decir,

$$\lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N \leq 0 \quad \text{y} \quad \mathbf{T} = [\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N]$$

A pesar de ser óptima en varios sentidos, la transformada de Karhunen-Loeve tiene el inconveniente de que su cálculo equivale a encontrar los valores y vectores propios de la señal de entrada, y esto es muy costoso computacionalmente hablando, en su lugar se utilizan otras transformadas que aproximan su funcionamiento pero son de menor complejidad [20][10]. Tal es el caso de la transformada de ondeleta presentada en los capítulos 2 y 3.

El resultado de la transformación no puede codificarse eficazmente por estar en forma continua, entonces, para que se pueda dar compresión es necesario pasar del conjunto continuo de valores resultados de la transformación a un alfabeto discreto. Esta es la razón de ser del siguiente bloque, la cuantización.

4.3 CUANTIZACION Y DISTRIBUCION OPTIMA DE LOS BITS

La forma más sencilla de lograr este objetivo es la de dividir el intervalo entre los valores máximo y mínimo que toman los coeficientes y dividirlo en M subintervalos de igual longitud, después tomar cada muestra, encontrar en qué intervalo se encuentra y asignar en lugar de ella el valor medio del intervalo.

El procedimiento anteriormente descrito se conoce como cuantización escalar y es el más sencillo de los métodos para relacionar los datos continuos con un alfabeto discreto, pero puede no ser el mejor, porque considera de igual importancia a todos los subintervalos y eso en general no es cierto. En realidad la importancia de cada intervalo depende de la estadística de los datos pero para valorar la importancia de cada subintervalo se necesita una medida de calidad de la cuantización; para esto se utiliza el error de cuantización.

Cuando se realiza la cuantización se sustituyen valores de las muestras por valores representativos “cercaños”, lo que implica que la señal reconstruida a partir de los coeficientes cuantizados no va a ser la misma. A la diferencia entre la señal original y la señal reconstruida se le conoce como “error de cuantización”.

Se puede ver el proceso más a detalle para encontrar en qué condiciones la distorsión es mínima.

En cualquier sistema real el número de niveles de cuantización N es finito, y para esto es necesario elegir los valores en los que termina cada intervalo y asignar los valores y_k que el receptor asigna a todas las entradas que caen en el intervalo $[x_k, x_{k+1}]$. Entonces, si la fuente tiene una función de distribución de probabilidad $p(x)$ de las amplitudes x_k , el esperado del error de cuantización se puede poner como una función D de los valores $\{x_k\}$, $\{y_k\}$ y $p(x)$, así como de una función $f(x-y)$, que se utiliza para calcular la distorsión en términos de la diferencia entre las muestras.

Si $D = E[f(x-y)]$, entonces la distorsión queda como,

Por ejemplo, si la función $f(x-y) = (x-y)^2$ entonces la medida de distorsión es el error cuadrático medio.

$$D = \sum_{i=1}^N \int_{x_i}^{x_{i+1}} f(x-y_i) p(x) dx$$

donde $x_{N+1} = \infty$, $x_1 = -\infty$. Si se quiere minimizar D, se obtienen las condiciones necesarias de derivar D con respecto a las x_k y a las y_k e igualando a 0.

$$\frac{\partial D}{\partial x_i} = f(x_i - y_{i-1}) p(x_i) - f(x_i - y_i) p(x_i) = 0$$

$$\frac{\partial D}{\partial y_i} = - \int_{x_i}^{x_{i+1}} f'(x - y_i) p(x) dx = 0$$

Dado que las relaciones funcionales introducidas por $p(x)$ son muy complicadas se utilizan métodos numéricos seleccionando y_i y calculando los siguientes x 's y y 's a través de las siguientes formulas [12]:

$$x_i = (y_i + y_{i-1})/2 \quad \text{o} \quad y_i = 2x_i - y_{i-1}$$

$$\int_{x_i}^{x_{i+1}} (x - y_i) p(x) = 0$$

derivadas de las anteriores considerando que la función $f(x)$ es una función cuadrática.

El método anteriormente descrito es propuesto por Lloyd a partir de los resultados de Lloyd, por lo que se conoce como el Método de Lloyd-Max. Su funcionamiento se basa en encontrar los centroides de los intervalos para asegurar que la distorsión es mínima para esas muestras.

Además, en la cuantización se incluye un filtrado no lineal de los coeficientes de la transformada de ondeleta, este se realiza asignando cero a todos los coeficientes cuyo valor sea menor al de un umbral dado. La forma como este umbral afecta la distorsión en la reconstrucción no está bien definida. Sin embargo, se observa que el aplicar un umbral a los coeficientes tiene el efecto de rechazar el ruido gaussiano de la señal, y aumenta considerablemente la relación de compresión.

Independientemente de como se elijan los niveles de cuantización es necesario determinar también el número de niveles en el cuantizador, cuando se tiene diferentes salidas cuantizadas independientemente, este problema es equivalente a minimizar una función de costo, con un número finito de recursos, en este caso los recursos son los bits que en total se asignan a los cuantizadores y el número de niveles en cada uno es el resultado de la forma en que se distribuyan los bits.

Para solucionar este problema se puede utilizar un algoritmo conocido como GREEDY [7], que asigna un bit a la vez colocándolo donde la distorsión es menor, dicho de otra manera, el algoritmo se asegura de que cada bit se coloca en donde su rendimiento es mayor. Este algoritmo no es óptimo porque no prueba todas las combinaciones para la distribución de los bits, pero lleva a buenos resultados.

A continuación el algoritmo se resume en la tabla 4.1.

1. Inicializar el número de bits en cada cuantizador $Nb[n]=1$.
2. Asignar a un bit más al cuantizador $n \quad Nb[n]=Nb[n]+1$.
3. Calcular la distorsión con esta asignación.
4. Repetir los pasos 2 y 3 para todos los cuantizadores, y conservar el de menor distorsión.
5. Repetir 2, 3 y 4 hasta que todos los bits sean asignados.

Tabla 4.1 Algoritmo de distribución de los bits.

4.4 CODIFICACION POR ENTROPIA

El último paso en la codificación por transformada es la codificación por entropía, la cual es un proceso reversible al igual que la transformada, lo que significa que no contribuye a la distorsión de la señal.

A la salida del cuantizador se tiene un número finito de valores, que llamaremos símbolos, dependiendo de la estadística de la entrada cada símbolo a_i tienen una probabilidad de ocurrencia $p(i)$.

La codificación en general se basa en transmitir secuencias de símbolos en lugar del mensaje original. Cada símbolo puede estar formado por una secuencia de dígitos, que por simplicidad se consideran binarios, '1' ó '0'. Si el número de símbolos disponibles para codificar el mensaje es N y cada uno tiene una probabilidad asociada $p(i)$ de aparecer en el mensaje de forma que:

$$\sum_{i=1}^N p(i) = 1,$$

Si se define la longitud de un símbolo $L(i)$, como el número de bits que representan un símbolo, entonces la longitud promedio es :

$$L_m = \sum_{i=1}^N p(i)L(i),$$

La entropía de una secuencia de símbolos, es el número mínimo de bits por símbolo, y depende de la estadística de la entrada, se define como :

$$H = \sum_i p(i) \log_2(p(i))$$

La idea principal de la codificación es encontrar un mapeo reversible del conjunto de valores a la salida del cuantizador $\{a_i\}$ a un nuevo conjunto de valores $\{b_i\}$ de longitud variable, de modo que la longitud promedio se minimize, es decir se aproxime a la entropía.

Para conseguirlo se siguen algunas restricciones básicas impuestas al conjunto de símbolos, Huffman propuso un algoritmo basado en estas suposiciones por lo a este conjunto se le conoce como códigos de Huffman, [8].

Las restricciones son:

- a) No es posible que dos símbolos tengan arreglos de bits idénticos.
 - b) Una vez que se sabe que la secuencia de símbolos comenzó, los códigos no deben de necesitar información adicional de dónde comienza o termina cada uno.
- Esto produce una tercera restricción, c) Ningún código puede ser prefijo de otro.

El diseño del código de Huffman, es un algoritmo iterativo que comienza por asignar los últimos bits de los símbolos menos probables, considerando que los otros bits van a ser iguales para ese par de símbolos, se les asigna 1 y 0, y se suma su probabilidad para considerarlos en adelante como un sólo símbolo. Después, se acomodan otra vez por su probabilidad y se repite el algoritmo hasta que sólo quedan dos símbolos y la suma de su probabilidad es uno.

De esta manera se asignan códigos de longitud menor a los símbolos con mayor probabilidad de ocurrencia, minimizando la longitud promedio de los símbolos

5 SISTEMA DE ELECTROCARDIOGRAFIA AMBULATORIO

5.1 ELECTROCARDIOGRAFIA

La electrocardiografía es el trazo, ya sea sobre una tira de papel o en una pantalla, que representa los potenciales en la superficie del corazón, estos voltajes en realidad se estiman a partir de los potenciales en varios puntos en la superficie del cuerpo y en base a ellos se infiere el funcionamiento eléctrico y mecánico del corazón.

Una de las principales técnicas para el diagnóstico de las enfermedades del corazón está basada en la interpretación de las características eléctricas del corazón.

Con un electrocardiograma (ECG) es posible determinar varias características del corazón como el ritmo cardiaco, la posible presencia de infartos, etc. Existen básicamente tres tipos de técnicas utilizadas en Electrocardiografía [16], y son :

- 1.-El ECG estándar.(12 derivaciones).
- 2.-El Vectocardiograma, (3 derivaciones ortogonales)
- 3.-El ECG ambulatorio o de monitoreo.(1 ó 2 derivaciones)

La más familiar de ellas es el electrocardiograma estándar, en donde se toman 12 diferencias de voltaje, llamadas derivaciones cardiacas, de la superficie del cuerpo de un paciente en reposo. El segundo, toma otro conjunto de potenciales como entrada a un modelo de la excitación cardiaca vectorial de tres dimensiones, esto produce una salida gráfica de la excitación del corazón llamada vectocardiograma. Finalmente, para observaciones de mucho tiempo, en la unidad de cuidados intensivos, o para pacientes que necesitan llevar puesto el electrocardiógrafo fuera de la sala de consultas para desarrollar sus actividades cotidianas, y al mismo tiempo cuidar y almacenar la actividad eléctrica de su corazón, se les proporciona un aparato que observa y/o guarda 1 ó 2 derivaciones principalmente para buscar trastornos del ritmo cardiaco. Este se conoce como análisis de arritmias, ECG ambulatorio o HOLTER.

Sin embargo, todos los enfoques tienen en común el hecho de ser mediciones no invasivas por medio de las cuales se trata de inferir el funcionamiento mecánico y eléctrico del corazón. Esta inferencia es totalmente empírica, es decir, que un cardiólogo aprende el significado de las diferentes partes del ECG de otros expertos, que a su vez lo aprendieron de otros expertos y de la práctica continua.

En un principio, cuando Einthoven inventó el método, las señales se tomaban sumergiendo las extremidades del paciente en cubetas con solución salina, y se utilizaba un galvanómetro para registrar el ECG. El método consistía en medir el voltaje en la superficie del cuerpo y tomar la diferencia entre dos de ellos, tomando esta cantidad como la magnitud de un vector cardiaco en esa dirección. Este vector representa el impulso eléctrico en esa dirección que causa las contracciones de los músculos cardiacos. De esta manera, se forman tres derivaciones o vectores conocidos como el triángulo de Einthoven, que se muestra en la figura 5.1. Se tiene como referencia de tierra la pierna izquierda para todos los voltajes, la diferencia entre el voltaje del brazo izquierdo (LA) y del brazo derecho (RA) forma la primera derivación (I), así mismo la diferencia entre el voltaje del brazo izquierdo y la pierna derecha forma la segunda derivación ($II = LA - RL$) y el del brazo derecho menos el de la pierna derecha forman la tercera. ($III = RA - RL$).

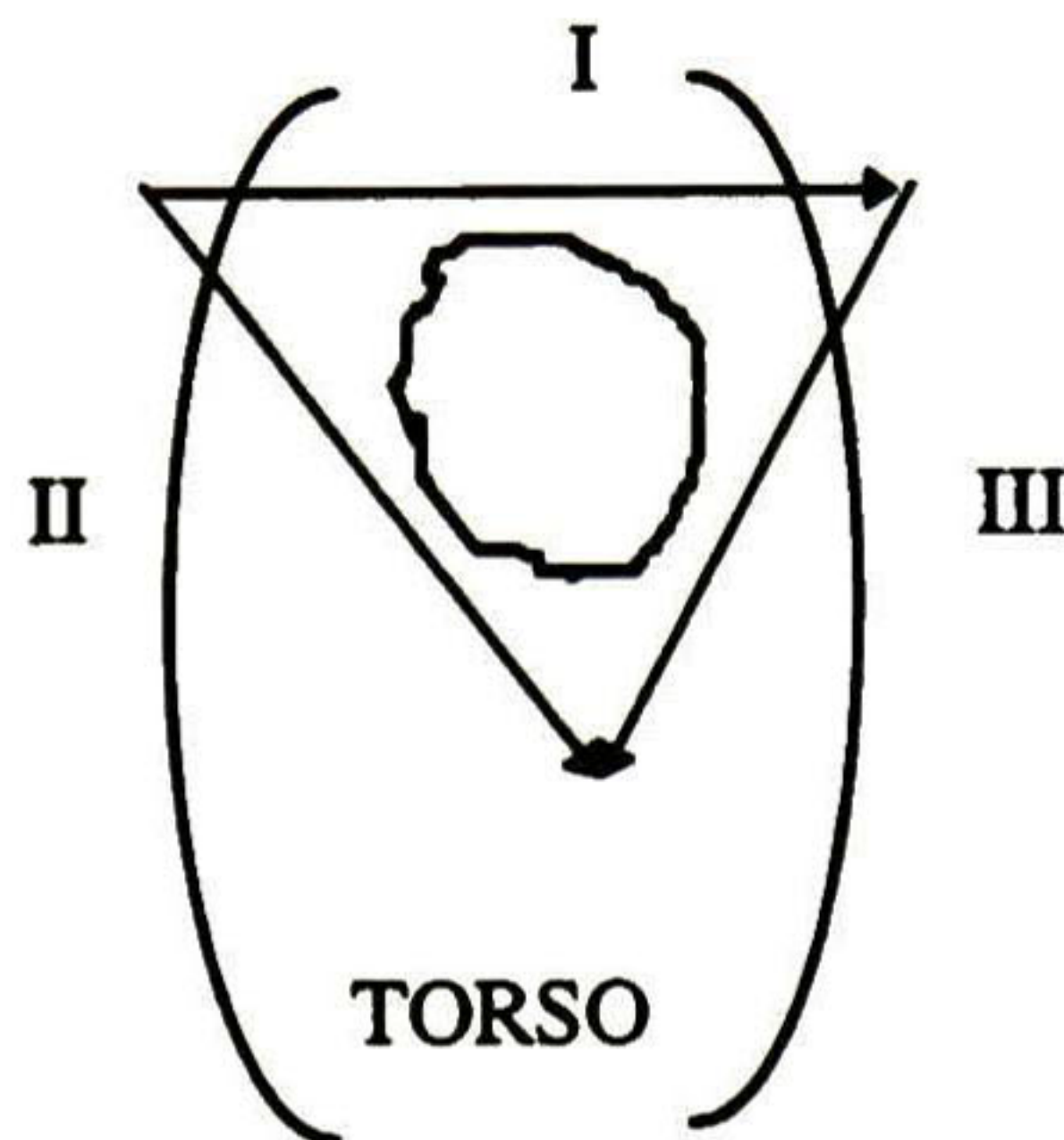


Figura 5.1. Triángulo de Einthoven.

Con la llegada de los amplificadores, la señal eléctrica pudo ser aumentada lo suficiente para imprimirse en una tira de papel continuo, pero el método no ha cambiado con el tiempo, sólo se han agregado más derivaciones (hasta formar 12) tomando los voltajes en más puntos del tórax, pero el fundamento es el mismo. Estos voltajes tienen características comunes que tienen que ver con el funcionamiento normal del corazón.

El corazón puede verse como un conjunto de músculos que cumple con la función de impulsar la sangre a las diferentes partes del cuerpo. El corazón tiene 4 cavidades que tienen la posibilidad de contraerse y dilatarse para expulsar y recibir la sangre, dos aurículas encargadas de recibir la sangre y los ventrículos encargadas de expulsar la sangre. Para realizar su función correctamente las distintas partes del corazón deben trabajar en sincronía, ésta es dictada por el estímulo eléctrico que controla las contracciones del corazón, y se registra en el ECG.

El electrocardiograma representa los cambios en los potenciales eléctricos durante el ciclo cardíaco. La forma característica de esta señal se debe a un potencial de acción que se propaga dentro del corazón y causa la contracción de las diferentes porciones del músculo cardíaco. Esta excitación interna comienza en el nodo sinusoidal y se propaga a la aurícula; esto genera la onda P característica en el ECG (figura 5.2). La excitación cardíaca alcanza entonces el ventrículo depolarizándolo y dando como resultado el complejo QRS típico de la onda de ECG. Una vez que los ventrículos han sido estimulados completamente (que corresponde al segmento ST de el ECG), éstos se repolarizan correspondiendo a la onda T del ECG. Los voltajes y tiempos de estos puntos y segmentos son características importantes para el diagnóstico.

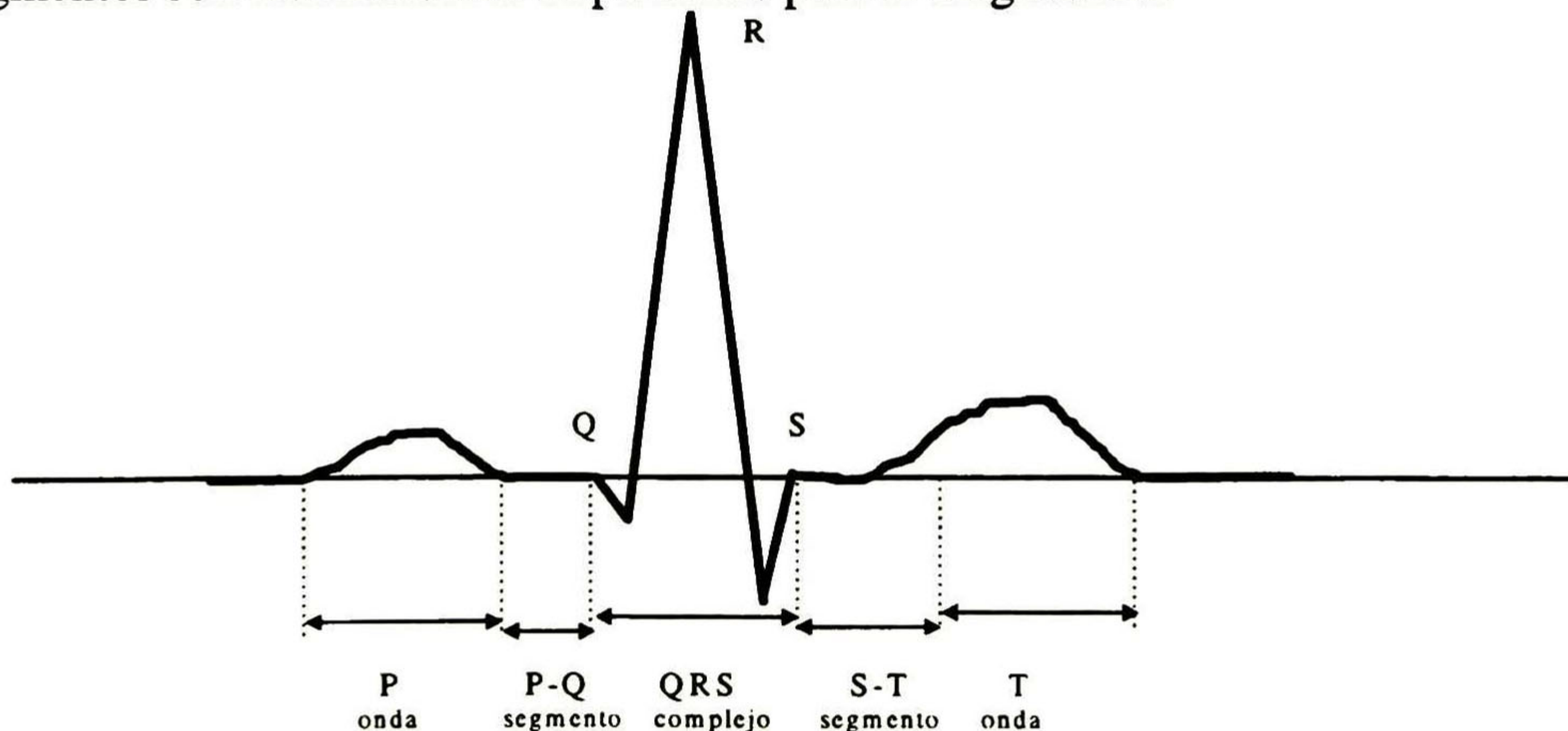


Figura 5.2 Representación esquemática de un ciclo de la señal de electrocardiografía.

5.2 EL SISTEMA HOLTER DIGITAL.

Como se mencionó anteriormente, uno de los usos de la Electrocardiografía es el monitoreo de los pacientes durante su vida cotidiana, esto es útil porque existen afecciones que no son detectables cuando el paciente está en reposo. Con la finalidad de examinar la actividad cardiaca de un sujeto por largos periodos de tiempo, existe el ECG ambulatorio o HOLTER, que es un equipo de una o dos derivaciones que se almacenan de algún modo durante 24 horas para su posterior análisis e interpretación.

Esta tarea de almacenar señales para su posterior reproducción se hacía anteriormente de forma analógica, mediante el uso de cintas magnéticas, pero el desarrollo y la consecuente reducción de los costos de la tecnología digital así como las ventajas que ésta ofrece, han provocado que también los sistemas de ECG sean digitales.

El sistema consta de un ECG analógico, que es el encargado de tomar las señales por medio de electrodos conectados al pecho del paciente, y amplificarlas aproximadamente unas 1000 veces, el amplificador utilizado es diferencial, es decir, amplifica sólo la diferencia entre las señales de modo que el ruido en el sistema que es común a las dos entradas disminuye.

Esta señal se pasa por un filtro que la limita en banda para eliminar los componentes alias que pueden aparecer cuando se convierte a señal digital, eliminando además la componente de ruido de 60 Hz producida por las líneas de energía.

La señal de salida de los filtros se convierte a digital con una frecuencia de muestreo de 200 Hz, ya que la señal de ECG tiene solo componentes de frecuencias menores a 100 Hz útiles en este tipo de análisis, y se utilizan 12 bits de resolución principalmente por el rango dinámico del voltaje de corriente directa sobre el que está montado el trazo de electrocardiografía.

Como resultado de muestrear a 200 Hz con 12 bits de resolución durante 24 horas se tiene una señal que necesita 207.360 Mbits, es decir, 25.920 Mbytes para su almacenamiento.

Sin embargo, la señal digital tiene aún mucha redundancia, por lo que se comprime codificándola por medio de la transformada de ondeleta, utilizando el procedimiento que se describe en el capítulo 4, con la finalidad de ahorrar memoria (que es el componente más caro del sistema).

Después, la información ya en forma compacta se almacena en una memoria no volátil para poder reproducirla y analizarla posteriormente.

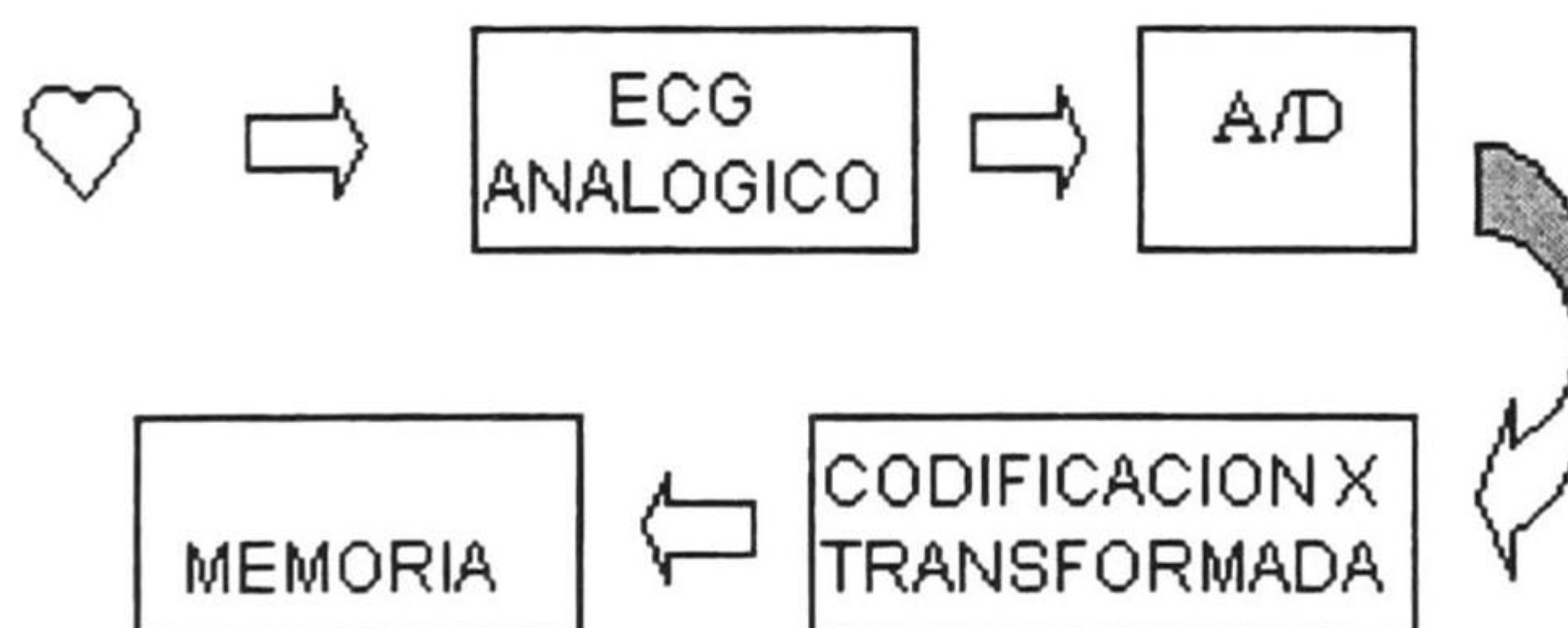


Figura 5.3. Diagrama a bloques del sistema .

Una vez que el paciente ha portado el sistema durante 24 horas, se presenta en el consultorio en donde se extraen los datos a una computadora y se reconstruye la señal, analizando la señal en las partes más significativas.

5.3 SIMULACION Y RESULTADOS.

Con la finalidad de establecer los parámetros del diseño del sistema, como los filtros, el número de niveles de descomposición y el tamaño de la memoria, se implementó en la computadora el bloque principal de la codificación por transformada.

La simulación se realizó en MATLAB para aprovechar las herramientas existentes, como la dedicada a ondeletas (Wavelet Toolbox) y la dedicada a Comunicaciones (Communications Toolbox), de estas herramientas se extrajeron los procedimientos principales.

La simulación se corrió basándose en señales de ECG digitales extraídas de INTERNET y otras recolectadas en el laboratorio procedentes de algunos voluntarios muestreadas con un convertidor analógico-digital de 12 bits a 200 Hz.

En las simulaciones se implementó el bloque de codificación por transformada que se conforma de tres sub-bloques, 1) Transformada de Ondeleta, 2) Cuantización, y 3) Codificación por Entropía. (ver Capítulo 4)

En la transformada de ondeleta la principal tarea fue definir la ondeleta que se iba a utilizar. Este bloque tiene como finalidad compactar la mayor cantidad de energía en el menor número de coeficientes posible, de manera que los bloques restantes hagan su trabajo de forma eficiente. Como la implementación se basa en filtros de muestreo multiple iterados las ondeletas simuladas forman parte de aquellas que tienen soporte finito.(ver capítulo 3)

Para poder comparar los resultados de la compresión con diferentes ondeletas todos los resultados deben referirse a un nivel de distorsión fijo de la señal, o con un nivel de compresión fijo comparando la distorsión producida, esta distorsión se mide utilizando una diferencia cuadrática porcentual (PRD= Percent Root-mean-squared Difference) dada por :

$$PRD = \left\{ \frac{\sum_{i=1}^n [x(i) - y(i)]^2}{\sum_{i=1}^n [x(i)]^2} \right\}^{1/2} \times 100\%$$

Donde x es la señal original, y y la señal reconstruida, con esta medida un método de compresión sin pérdida, dará un PRD=0% y entre menor sea la PRD el porcentaje de energía en el error será menor.

Esta forma de medir la distorsión se ha utilizado en trabajos previos de compresión de ECG lo cual permite comparar los resultados. Utilizando la codificación por transformada se logran resultados hasta de 7 :1 con un PRD de 3.9% [6], en contraste con los algoritmos como el AZTEC que logra relaciones de compresión de 7 :1 pero con PRD de 10%, y el FAN que logra compresión de 7 :1 con un PRD de 8.7% [16].

También se ha utilizado cuantización vectorial para la compresión de las señales de ECG, pero no se reporta con las mismas medidas de distorsión lo que dificulta el establecer comparaciones. [21]

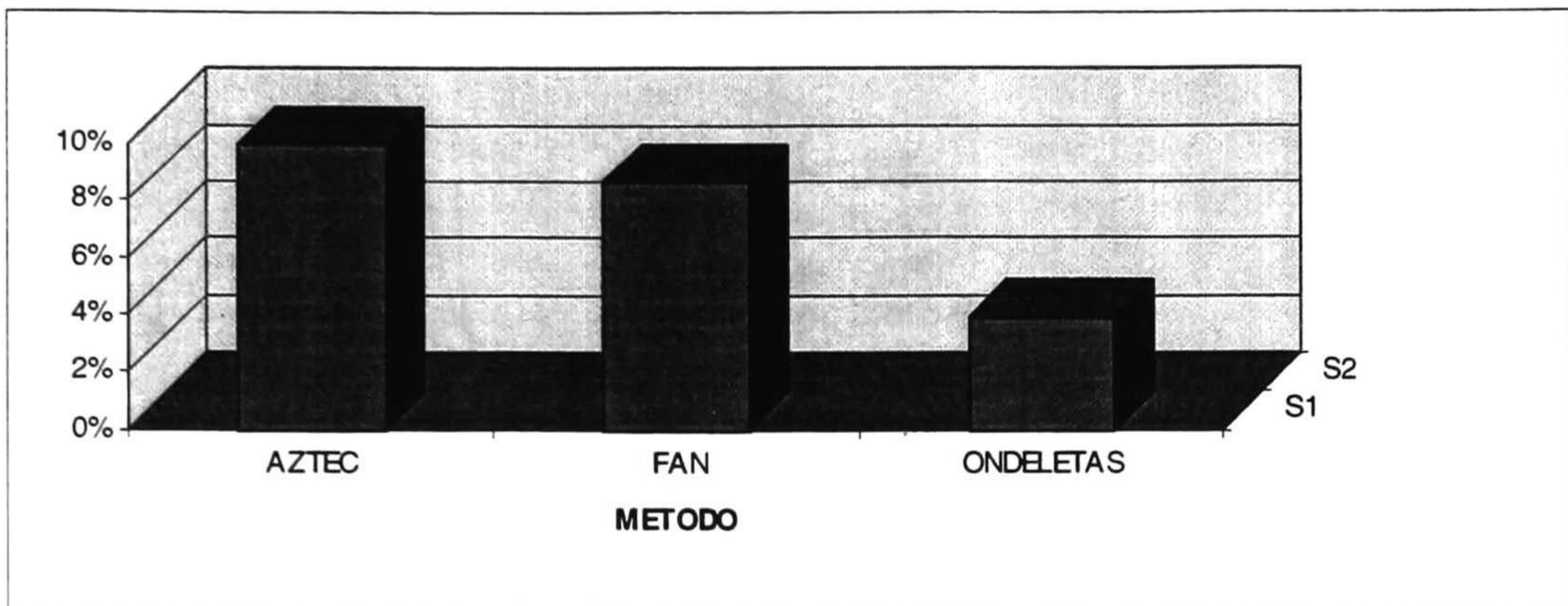


Figura 5.4. Niveles de distorsión con una compresión de 7 :1 para diferentes métodos.

Para realizar las gráficas de las figuras 5.5 y 5.6 se hizo una estimación de la entropía a la salida de los cuantizadores. Con el conjunto de señales de ECG disponibles, se calculó su transformada de ondeleta y con un determinado número de bits en los cuantizadores se discretizan los coeficientes y se estima sus probabilidades. Con estas probabilidades se calcula una entropía para cada banda y se calcula la entropía general como un promedio ponderado de las entropías en cada banda.

$$H_e = \frac{2^{K-1}h_{d1} + 2^{K-2}h_{d2} + \dots + h_{dK} + h_a}{2^K}$$

Donde H_e es la estimación de la entropía y h_{d1} es un valor aproximado de la entropía para la primera escala y así sucesivamente, h_{dK} es la entropía de la última escala y h_a es la entropía de la aproximación. Después se incrementa el número de bits en los cuantizadores y se repite el proceso, por supuesto entre mayor es el número de bits en los cuantizadores menor es la distorsión a la salida, pero el número de bits por muestra es mayor.

Para elegir la ondeleta prototipo se evaluaron todas las ondeletas con que cuenta la herramienta de ondeletas de Matlab, considerando que se tienen todas las familias de ondeletas de soporte finito, las cuales incluyen las de Daubechies [4], las biortogonales [19], las de Haar [20], las de Coifman [4], y las simétricas, [4].

La ondeleta que obtiene mejores resultados es aquella que logra concentrar la mayor energía en el menor número de coeficientes, y es la generada por el par de filtros biortogonales de longitud 3,9 de la herramienta para ondeletas de Matlab, como se ve en la Figura 5.5.

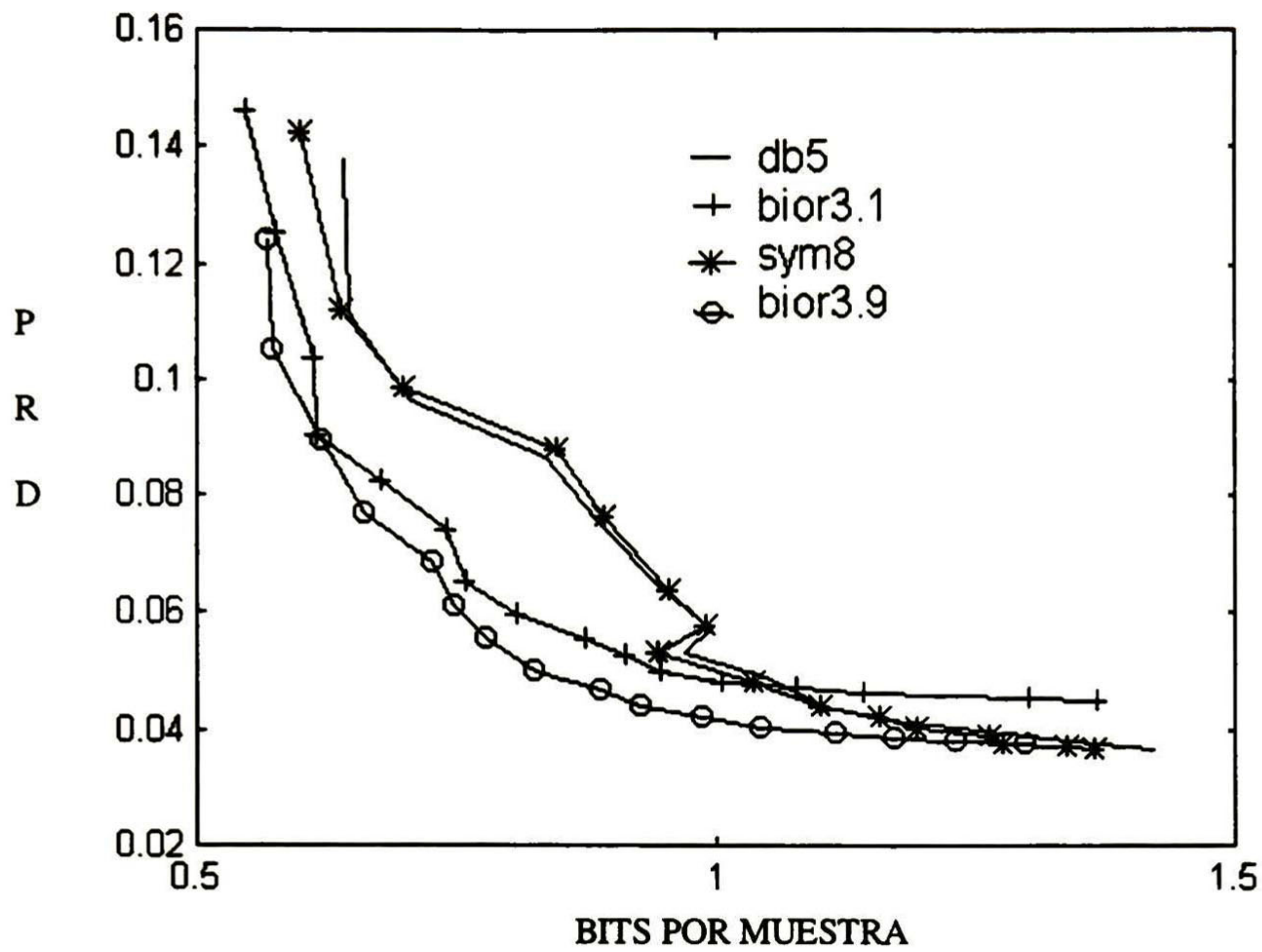


Figura 5.5. Niveles de Compresión para diferentes ondeletas, PRD Vs Entropía.

Además de la ondeleta prototipo, es decir aquella que se expande y comprime para formar la base, en un sistema práctico también se debe seleccionar el número de niveles de descomposición. En la Figura 5.6 se puede apreciar que la compresión varía de acuerdo a este parámetro, encontrando que el sistema se desempeña eficientemente con 4 niveles. Hay que recordar que existe un compromiso entre el número de niveles y el costo computacional de la transformada.

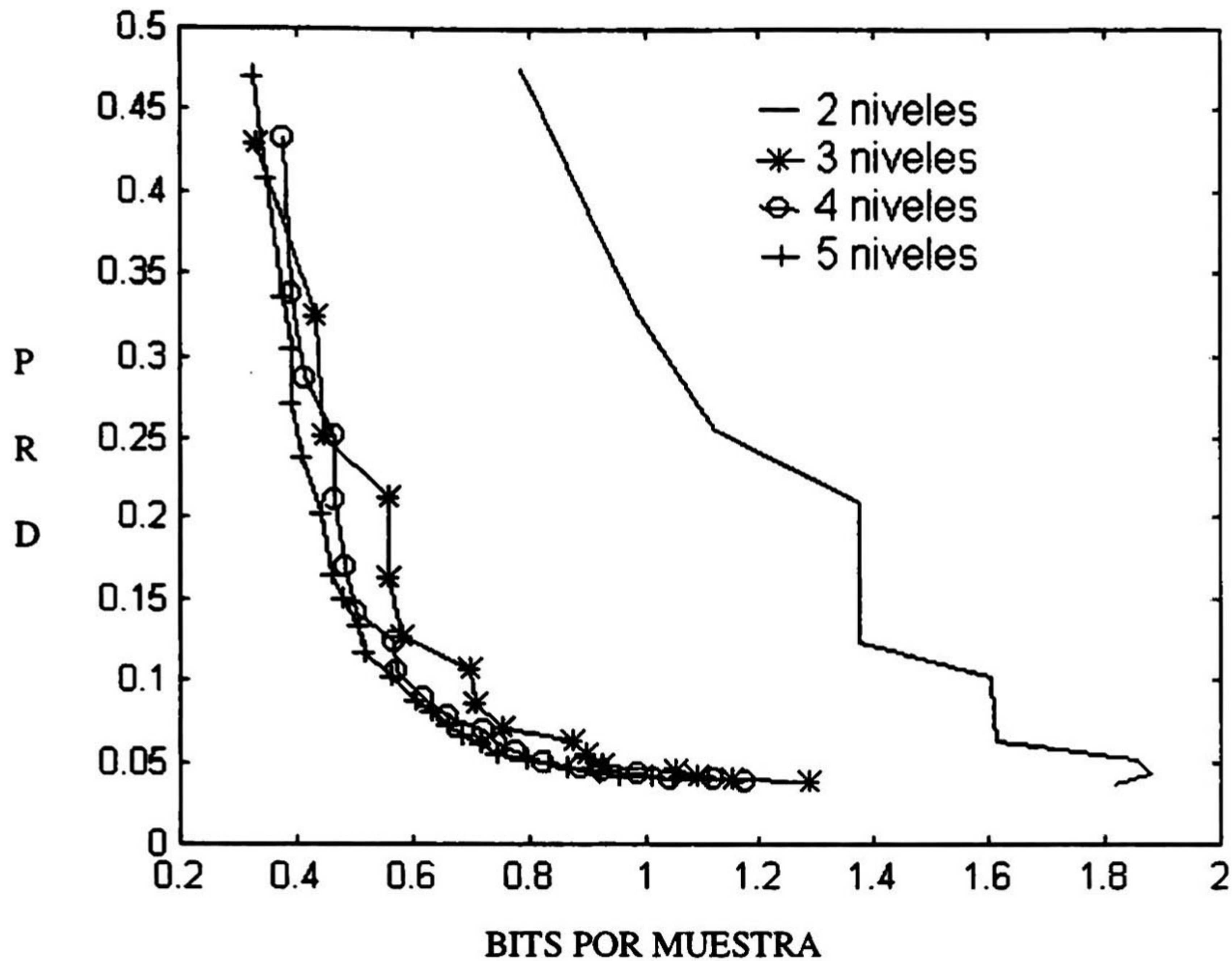


Figura 5.6. Compresión con Diferentes Números de Subbandas.

Para cada uno de éstos niveles en la descomposición se necesita un cuantizador independiente, y fijando el número de bits que utiliza, se determina el número de niveles de cuantización y por lo tanto también el error que introduce al sistema. Si todas las salidas de la transformada contribuyeran igual en la reconstrucción entonces todas tendrían el mismo número de niveles de cuantización pero esto no es cierto y existen varios métodos para encontrar el número de niveles óptimo para cada cuantizador.

Se implementaron 2 métodos para encontrar la distribución de los bits : el primero es teórico, y toma en cuenta la estadística de los coeficientes [7] pero falla porque algunas suposiciones del método no se cumplen, como el que la distorsión total es sólo una suma de las distorsiones de cada salida (lo cual únicamente es cierto para las transformaciones ortogonales). El otro método es iterativo, asignando un bit a la vez calcula en cuál de todas las bandas se disminuye más la distorsión total, claro que el resultado es un subóptimo por no considerar todos los casos posibles, pero mejora el desempeño del sistema considerablemente.

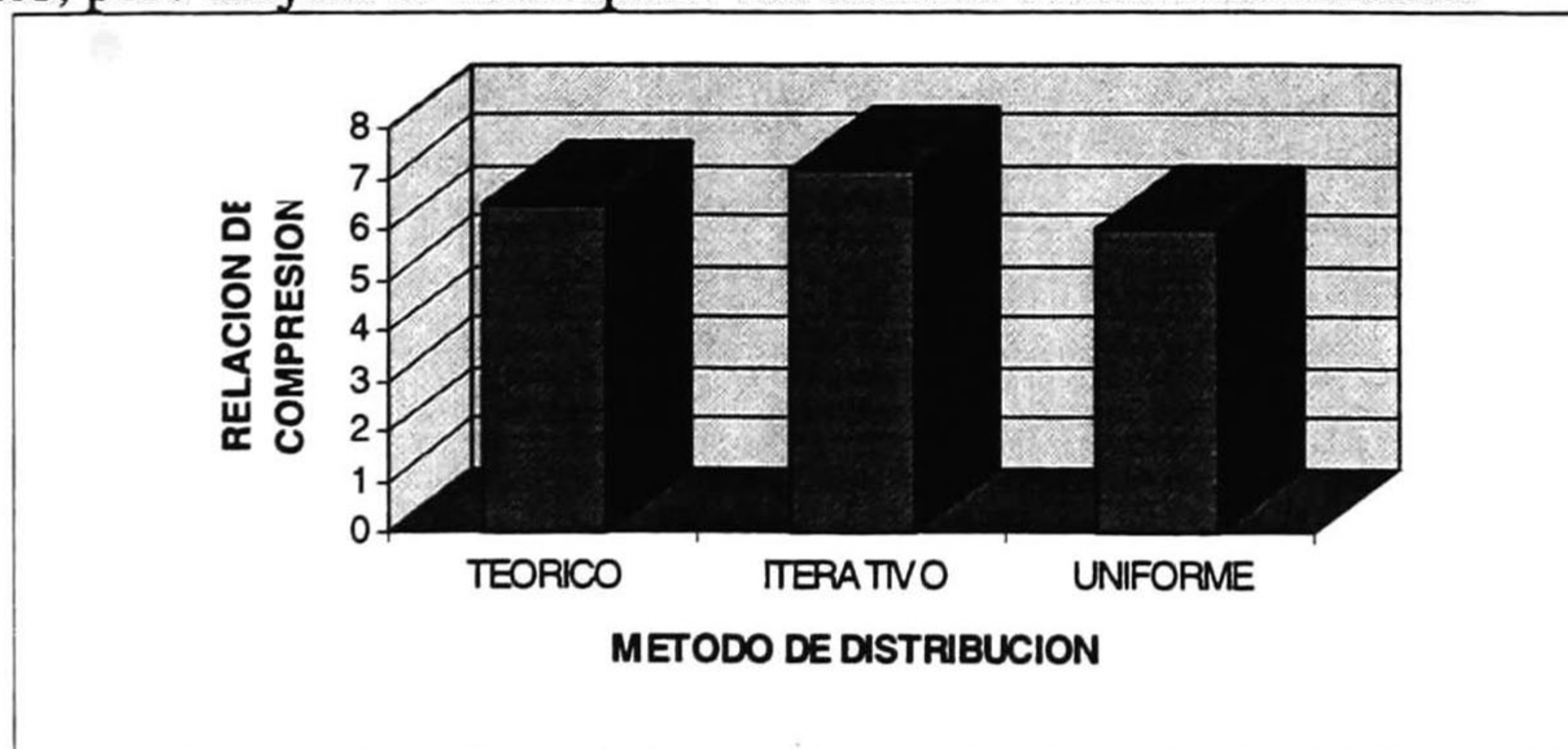


Figura 5.7. Resultados en la Distribución de los Bits.

Como se mostró en el capítulo 4, la cuantización uniforme introduce mayor distorsión que la no uniforme para el mismo número de niveles si se calculan los valores del cuantizador no uniforme con el algoritmo de Lloyd. En la Figura 5.8 se muestra la compresión de una señal con cuantización uniforme y con los valores obtenidos con el algoritmo de Lloyd.

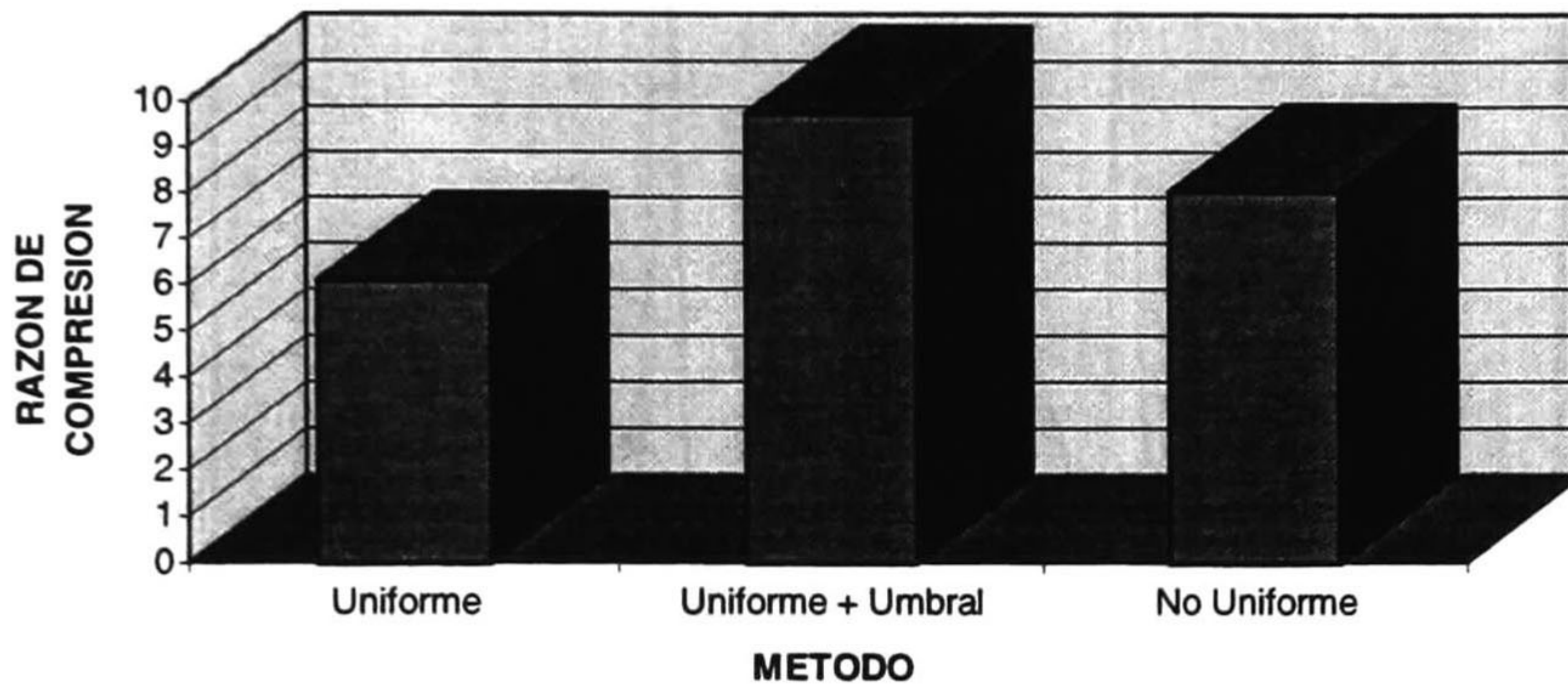


Figura 5.8. Compresión con Cuantización Uniforme y No Uniforme.

Como resultado de la transformada, la información relevante se concentra en unos pocos coeficientes, de hecho, más del 93% de los coeficientes tienen una amplitud menor al 5% de la amplitud máxima. Si se eliminan los coeficientes que no alcanzan una amplitud mínima, entonces la razón de compresión aumenta, porque se utiliza una codificación estadística que asigna la palabra de menor longitud a la muestra más frecuente y ésta última es invariablemente el cero. Su efecto en la distorsión total no es muy grande debido a que los coeficientes que llevan mayor información son los que tienen mayor energía.

Si se eliminan los coeficientes menores al 5% de la amplitud máxima, apenas se introduce en la señal una distorsión con $PRD=9.4\%$, a pesar de ser el 93% de los coeficientes. Además, cuando se introduce en el sistema, el ruido se reparte entre los coeficientes de todas las escalas, generando muchos coeficientes con pequeña amplitud, entonces el resultado de eliminar los coeficientes con poca amplitud es equivalente a separar la señal del ruido.

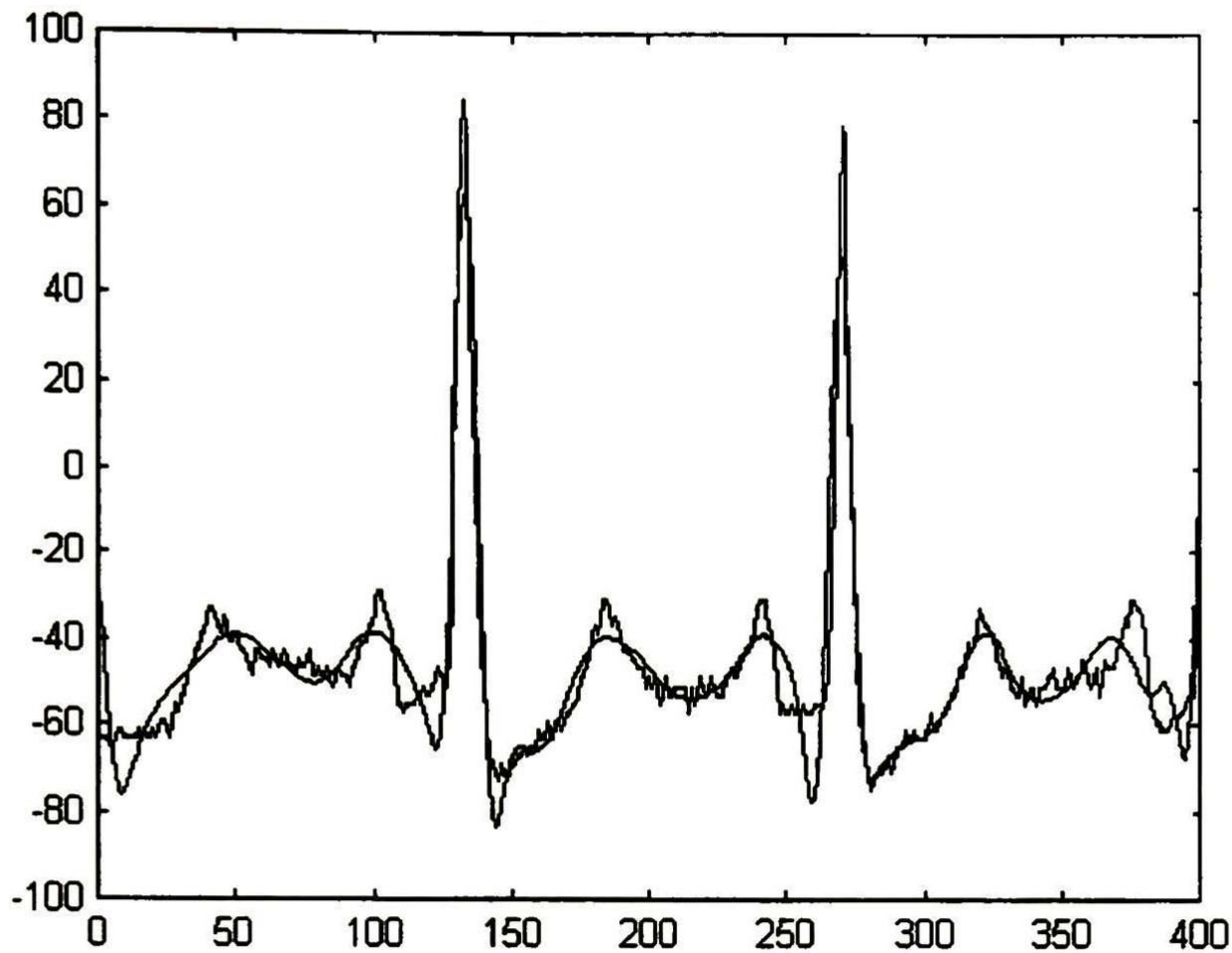


Figura 5.9. Señal de ECG y la señal resultante de aplicar un umbral a los coeficientes menores del 5% del valor máximo de la transformada antes de la reconstrucción.

En la parte de la codificación por entropía se implementó el algoritmo de Huffman en MATLAB, utilizando matrices para almacenar los árboles generados por los códigos y sus distribuciones asociadas. Un inconveniente de esta implementación es que la memoria se utiliza de una forma poco eficiente, haciendo más lenta la simulación. Sin embargo, se consideró importante tener toda la simulación en un sólo programa.

En la figura 5.10 se presenta una tabla de los códigos de Huffman para los valores de la salida de frecuencias más alta.

VALOR	PROBABILIDAD	CODIGO DE HUFFMAN
-261.2775	3.3179e-005	0 0 0 0 0 0 0
277.4315	3.3179e-005	0 0 0 0 0 0 1
-64.2898	5.5740e-003	0 0 0 0 0 1
45.9760	1.0153e-002	0 0 0 0 1
-18.6424	3.6662e-002	0 0 0 1
12.7358	6.6921e-002	0 0 1
-3.6111	3.6805e-001	0 1
1.9926	5.1257e-001	1

Figura 5.10. Valores de los códigos de Huffman.

En la simulación general intervienen los 3 bloques anteriores, mostrando la relación de compresión final y el tamaño de memoria necesaria para almacenar 24 horas de señal de ECG, así como las señales reconstruidas, después de la compresión y decompresión.

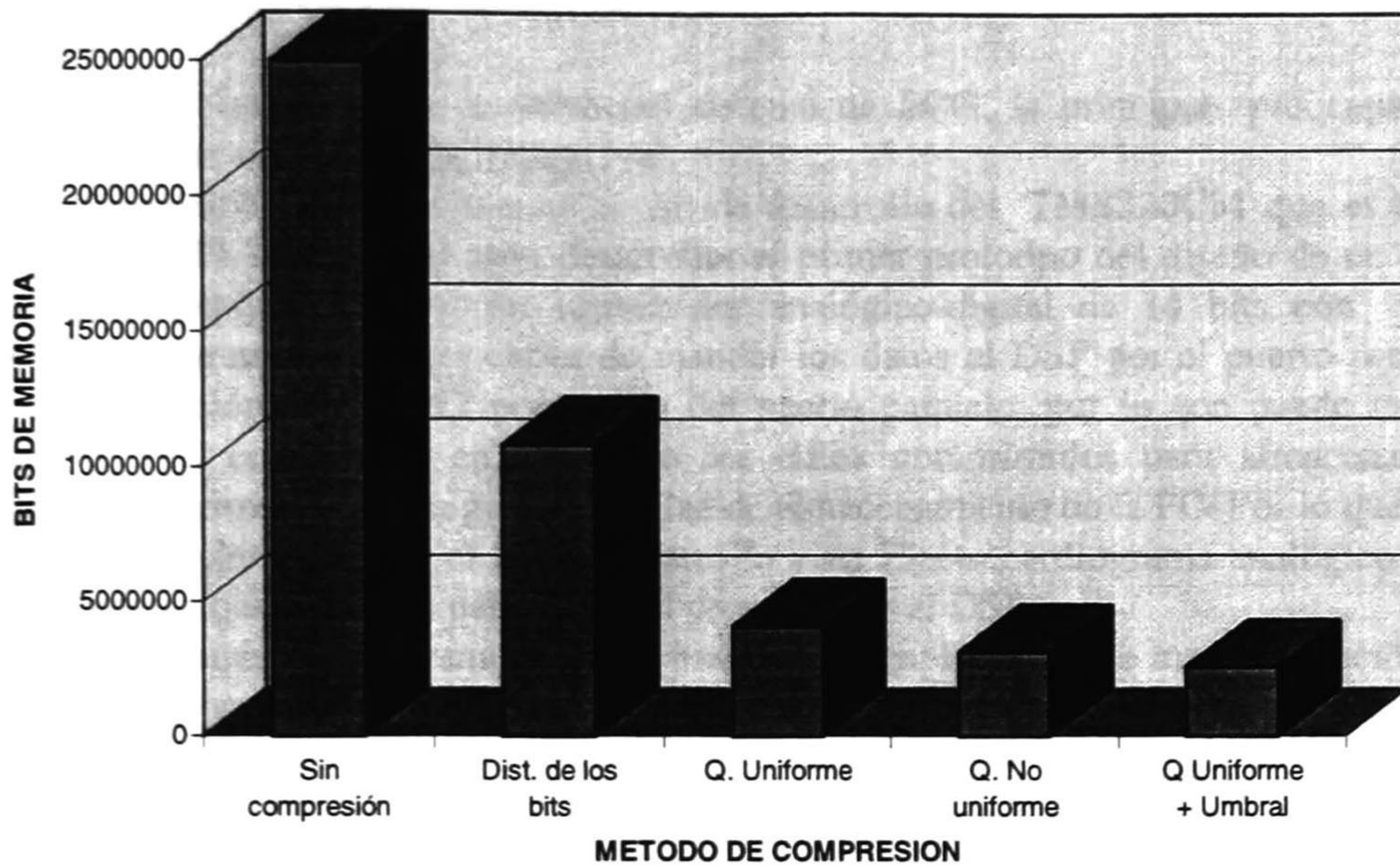


Figura 5.11. Relación de Compresión final y el tamaño de la memoria.

Como puede observarse en la figura 5.12 la calidad de la reconstrucción es aceptable desde el punto de vista médico porque preserva las características clínicamente importantes de la señal, como las distancias entre los puntos PQ, y QS, o la distancia entre R's , así como la forma del segmento ST y el valor máximo de la señal en el complejo QRS.

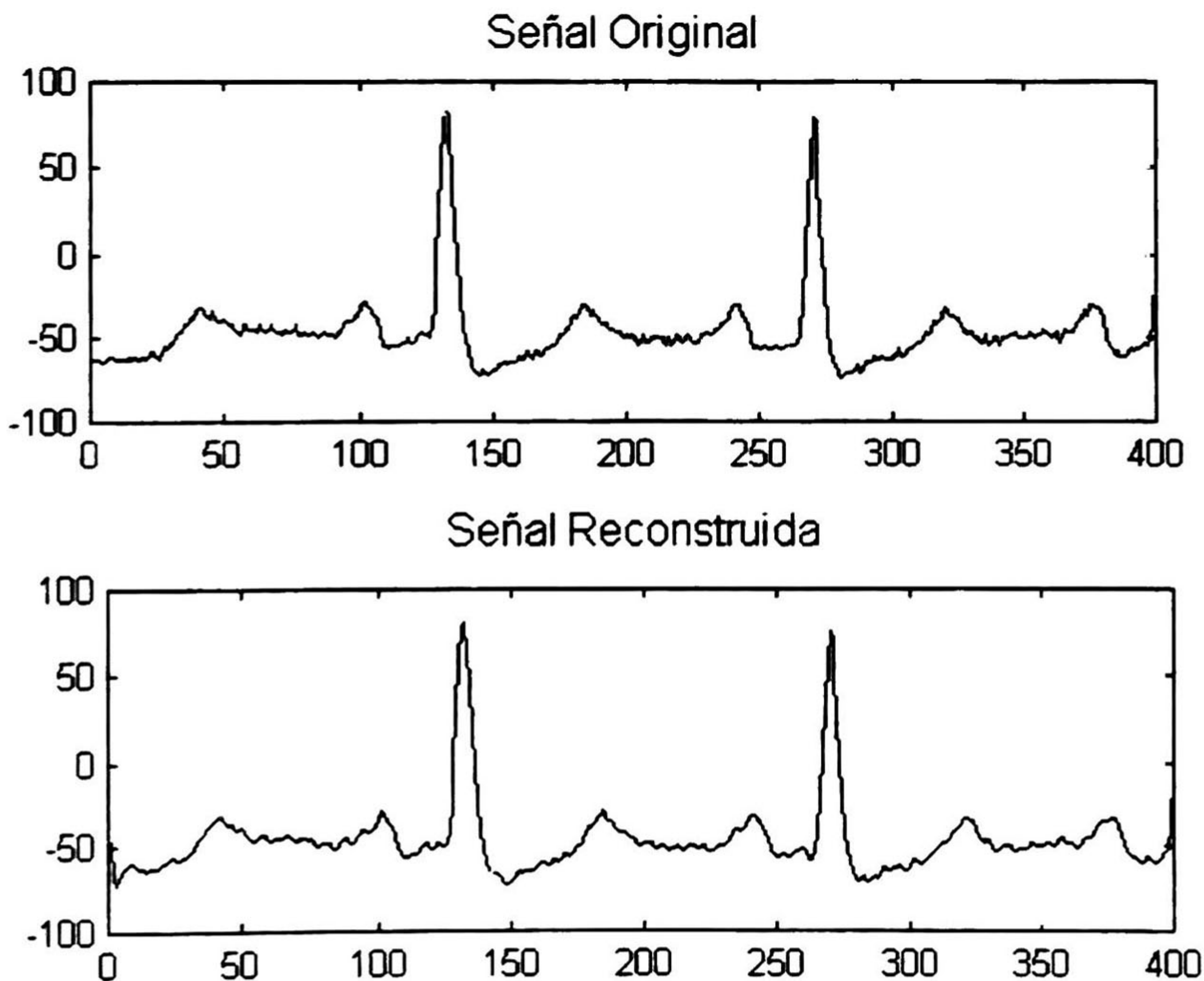


Figura 5.12. Señales antes y después de la compresión .

5.4 DETALLES DE LA IMPLEMENTACION.

Sobre la implementación física del sistema de ECG, la principal preocupación es la evaluación de los algoritmos en tiempo real.

La implementación se basa en el kit de desarrollo del TMS320C31 que es un DSP de punto flotante. El kit se presta para desarrollar el primer prototipo del diseño de un sistema de ECG digital, porque contiene un convertidor analógico-digital de 14 bits con tiempo de conversión programable que es capaz de mandar los datos al DSP por el puerto serie, además tiene comunicación con la PC, por medio del puerto paralelo, por lo que puede intercambiar datos a la PC, como sería en este caso los datos comprimidos para almacenarlos en la computadora, aprovechando la gran capacidad de almacenamiento de la PC. Por lo que el sistema completo se implementó con el kit de desarrollo y un Electrocardiograma analógico y una PC, que es la misma que se utiliza para cargar el programa en el DSP.

El diagrama del ECG analógico, se basa en un amplificador de instrumentación, con alta ganancia, y otros amplificadores operacionales para realizar el filtrado especialmente de la componente de 60 Hz..

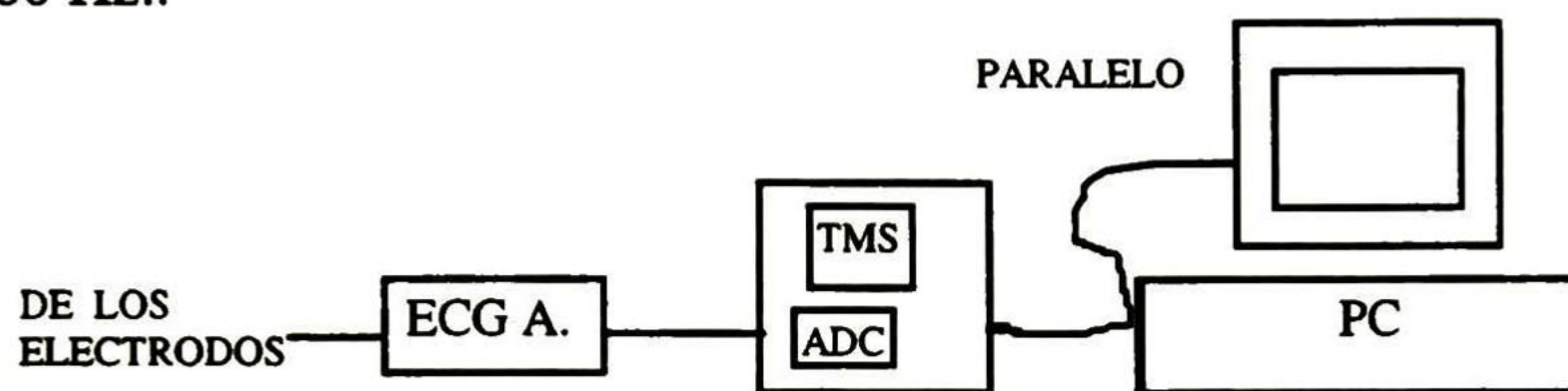


Figura 5.13. Diagrama de la implementación de un sistema de compresión de la señal ECG basado en el Starter Kit de Texas Instruments.

A continuación se explican los programas cargados en la PC y en el DSP que realizan las funciones de almacenamiento y compresión respectivamente.

El programa en la PC es más sencillo, su función es tomar los datos que pone el DSP en una localidad de memoria asignada de antemano, realizar la codificación de Huffman y guardarlos en el disco duro. La comunicación entre el DSP y la PC tiene un protocolo de Maestro-Esclavo, en este caso la PC es el maestro y determina cuándo y qué le manda o recibe del DSP, y el DSP solo deshabilita la interrupción del maestro en las operaciones que necesitan sincronización. Así la PC lee los datos comprimidos de una localidad predeterminada en la memoria de el DSP cuando éste está listo.

El programa en la PC se auxilia de un kernel de comunicaciones cargado con anterioridad en el DSP, en donde están las rutinas de comunicación provistas con el DSP starter kit.

El programa en el DSP está encargado de fijar la velocidad de adquisición de los datos, configurar el puerto serie del DSP y realizar la transformada, después los escribe a una localidad donde la PC los pueda leer, y habilita la interrupción de la PC.

A continuación se anexa la subrutina en el DSP que realiza la transformada de ondeleta.

```

;a-l----b          ;estructura de el banco de filtros
; |___c___d       ;utilizada en el programa, despues
; |              ;de cada senal existe un filtrado
; |---e---|-----f      ;y un submuestreo, y los filtros
; |___g___h      ;están en cascada.
; |              ;
; |-----i       ;
;
; inicializa los arreglos circulares A_DATA(32), C_DATA(32), E_DATA(32), G_DATA(32),
; y sus apuntadores.
fmain  ldf 0.0,R0
      ldi @A_DATA,AR2
      rpts 128
      stf R0,*AR2++
      ldi @A_DATA,AR2 ; dir_a en AR2
      ldi @C_DATA,AR5 ; dir_c en AR5
      ldi @E_DATA,AR6 ; dir_e en AR6
      ldi @G_DATA,AR7 ; dir_g en AR7

main   ; ldi 0x30,IE      ; Ignore the host while collecting data
      ldi @S0_rdata,R0 ; Clear SP under/overflow
      ldi 0,R0          ;
      sti R0,@S0_xdata ;
      ldi @S0_rdata,R0 ;
      ldi 0,R0          ;
      sti R0,@S0_xdata ;

*****
; Inicializa los apuntadores de los arreglos donde se guardan los
; coeficientes de la transformada de ondeleta en cada bloque
*****

      ldi @I_DATA,AR4 ; dir_i en Memoria
      sti AR4,@I_DAT ;
      ldi @H_DATA,AR4 ; dir_h en Memoria
      sti AR4,@H_DAT ;
      ldi @F_DATA,AR4 ; dir_f en Memoria
      sti AR4,@F_DAT ;
      ldi @D_DATA,AR4 ; dir_d en Memoria
      sti AR4,@D_DAT ;
      ldi @B_DATA,AR4 ; dir_b en Memoria
      sti AR4,@B_DAT ;
      ldi 10,R4        ; Para generar la rampa
      ldi 1,R7         ; Pone el numero de muestra a uno
NMUEST call GETADC

```



```

float R0,R0      ;convierte a flotante para las operaciones
stf R0,*AR2++%   ;a(muestra)=dato
tstb 0x01,R7     ;R7 es el contador de muestras R7='muestra'
bnz endwav      ;prueba si el numero de muestra es par
; si es par ejecuta el siguiente nivel de descomposicion.

```

```

LDI @H0_ADDR,AR0 ;pone el apuntador a h0
ldi AR2,AR1      ;pasa la direccion de a al FIR
call FIR         ;c(n+1)=conv(a,h0)
stf R2,*AR5++%   ;guarda c(n+1) en un arreglo circular

```

```

ldi @H1_ADDR,AR0 ;pone el apuntador a h1
ldi AR2,AR1      ;pasa la direccion de a al FIR
call FIR         ;b(n+1)=conv(a,h1)
fix R2,R2        ;convierte a entero para guardarlo
ldi @B_DAT,AR4   ;saca de la memoria la direccion de b

```

```

tstb 0x02,R7     ;prueba si el numero de 'muestra'
bnz endwave      ;es multiplo de 4 ?
and 0xffff,R2    ;limpia la parte superior de R2
ldi R3,*AR4      ;carga el dato anterior
or R3,R2         ;empaqueta los dos en uno
sti R2,*AR4++    ;guarda el dato
sti AR4,@B_DAT   ;actualiza el valor de la direccion en memoria
bu S1

```

```

endwave lsh R2,16 ;pone el dato en la parte mas significativa del word
sti R2,*AR4      ;guarda el dato
bu endwav

```

; si es multiplo de 4 ejecuta el siguiente nivel

```

S1 LDI @H1_ADDR,AR0 ;pone el apuntador a h1
ldi AR5,AR1      ;pasa la direccion de c al FIR
call FIR         ;d(n+1)=conv(c,h1)
fix R2,R2        ;convierte a entero para guardarlo
ldi @D_DAT,AR4   ;saca de la memoria la direccion de d
sti R2,*AR4++    ;guarda el dato
Sti AR4,@D_DAT   ;actualiza el valor de la direccion en memoria

```

```

LDI @H0_ADDR,AR0 ;pone el apuntador a h0
ldi AR5,AR1      ;pasa la direccion de c al FIR
call FIR         ;e(n+1)=conv(c,h0)
stf R2,*AR6++%   ;guarda e(n+1) en un arreglo circular

```

```

tstb 0x04,R7     ;prueba si el numero de 'muestra'
bnz endwav      ;es multiplo de 8

```

; si es multiplo de 8 ejecuta el siguiente nivel


```

LDI @H1_ADDR,AR0 ;pone el apuntador a h1
ldi AR6,AR1 ;pasa la direccion de e al FIR
call FIR ;f(n+1)=conv(e,h1)
fix R2,R2 ;convierte a entero para guardarlo
ldi @F_DAT,AR4 ;saca de la memoria la direccion de f
sti R2,*AR4++ ;guarda el dato
Sti AR4,@F_DAT ;actualiza el valor de la direccion en memoria

```

```

LDI @H0_ADDR,AR0 ;pone el apuntador a h0
ldi AR6,AR1 ;pasa la direccion de e al FIR
call FIR ;g(n+1)=conv(e,h0)
stf R2,*AR7++% ;guarda g(n+1) en un arreglo circular

```

```

tstb 0x08,R7 ;prueba si el numero de 'muestra'
bnz endwav ;es multiplo de 16

```

; si es multiplo de 16 ejecuta el siguiente nivel

```

LDI @H1_ADDR,AR0 ;pone el apuntador a h1
ldi AR7,AR1 ;pasa la direccion de g al FIR
call FIR ;h(n+1)=conv(g,h1)
fix R2,R2 ;convierte a entero para guardarlo
ldi @H_DAT,AR4 ;saca de la memoria la direccion de h
sti R2,*AR4++ ;guarda el dato
Sti AR4,@H_DAT ;actualiza el valor de la direccion en memoria

```

```

LDI @H0_ADDR,AR0 ;pone el apuntador a h0
ldi AR7,AR1 ;pasa la direccion de g al FIR
call FIR ;i(n+1)=conv(g,h0)
fix R2,R2 ;convierte a entero para guardarlo
ldi @I_DAT,AR4 ;saca de la memoria la direccion de i
sti R2,*AR4++ ;guarda el dato
Sti AR4,@I_DAT ;actualiza el valor de la direccion en memoria

```

```

endwav addi 1,R7 ; incrementa 'muestra'
cmpi 32,R7 ; compara con 32 para completar el bloque si es menor
ble NMUEST ; salta a la siguiente muestra.
;-----

```

Nota : Los programas tanto de simulación en Matlab, como los de implementacion en C y en ensamblador se anexan en un disquette.

6 CONCLUSIONES Y TRABAJO FUTURO

La transformada de ondeleta es útil en la compresión de señales de Electrocardiografía porque proporciona una representación alternativa de la información en la cual unos cuantos coeficientes son relevantes.

La transformada de ondeleta discreta no es una versión discretizada de la transformada de ondeleta continua, pero conserva la característica de representar eficientemente los cambios rápidos en el tiempo.

La compresión por medio de ondeletas necesita de otros bloques de procesamiento, cuyo desempeño repercute en forma significativa en la relación de compresión total, se sugiere que se evalúe el método de compresión utilizando otros cuantizadores y codificadores por ejemplo un esquema de cuantización vectorial o uno utilizando el algoritmo de compresión de Lempel Ziv.

Los coeficientes de la transformada de ondeleta discreta tienen un rango dinámico que se incrementa con cada nivel de descomposición, un trabajo posterior puede explorar las propiedades estadísticas de la transformada de ondeleta para hacer uso de ellas en este esquema de compresión.

El efectuar un filtrado no lineal de los coeficientes aplicando un umbral, aumenta en forma considerable la relación de compresión total aumentando moderadamente la distorsión, sin embargo, no es muy claro cuál es el umbral óptimo para un grado de distorsión deseado.

La elección de la ondeleta prototipo repercute en la relación de distorsión por lo que es importante determinar la ondeleta óptima para cada señal, el trabajo futuro podría incluir el diseñar una ondeleta especial adaptada a la forma de la señal de electrocardiografía.

La elección del número de niveles de descomposición, o escalas calculadas en la transformada de ondeleta discreta, repercute en el desempeño del sistema pero no de forma significativa, ya que la complejidad tiende a ser lineal con respecto al número de retardos en los filtros de descomposición.

La transformada de ondeleta puede implementarse eficientemente en un procesador digital de señales de punto flotante como es el TMS320C31 de Texas Instruments, sería interesante observar los efectos de realizar la transformada en un procesador de punto fijo y de menor tamaño de palabra.

Por medio de la compresión de la señal de ECG se puede implementar un sistema de electrocardiografía ambulatorio que guarde 24 horas de señal con solo 3 Mbytes de memoria en el sistema.

Se podría evaluar un esquema adaptativo utilizando paquetes de ondeletas.

7 REFERENCIAS

- [1] M. Akay, "Wavelets in Medicine", IEEE Spectrum, Vol 34, No 5, Mayo 1997, pp 50-56
- [2] B. Burke Hubbard, "The World According to Wavelets", Ed. A. K. Peter, 1996.
- [3] A. Cohen y J. Kovacevic, "Wavelets : The Matematical Background", Proceedings of the IEEE, Vol. 84, No. 4, Abril 1996, pp 514-521.
- [4] I. Daubechies, "Ten Lectures on Wavelets", CBMS Conference SIAM, Philadelphia, 1992.
- [5] I. Daubechies, "Where Do Wavelets Come From- A Personal Point of View ", Proceedings of the IEEE, Vol. 84, No. 4, Abril 1996, pp 510-513.
- [6] A. Djohan, T. Nguyen y W. Tompkins, "ECG Compression Using Symetric Wavelet Transform", Proceedings of EMBC 95.
- [7] A. Gersho y R. M. Gray, "Vector Quantization and Signal Compression", Kluwer Academic Publisher, Norwell Massachusetts, 1992.
- [8] D. A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes", Proceedings of IRE, Vol. 40, Septiembre 1952, pp 1098-1101.
- [9] N. S. Jayant y P. Noll "Digital Coding of Waveforms. Principles and Application to Speech and Video", Prentice Hall, Englewood Cliffs, New Jersey, 1984.
- [10] W. Kinser y A. Langi, "Speech and Image Signal Compression with Wavelets", Signal Processing Technology and Applications, IEEE Technical Activities Board, 1995.
- [11] S. Mallat, "Multifrequency Channel Decompositions of Images and Wavelets Models", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 37, No. 12, Diciembre 1989, pp 2091-2110.
- [12] J. Max, "Quantizing for Minimum Distortion", IRE Trans. Inform. Theory, Vol IT-6, Marzo 1960, pp 7-12.
- [13] O. Rioul y M. Vetterli, "Wavelets and Signal Procesing", IEEE Signal Processing Magazine, Octubre 1991, pp 14-37.
- [14] O. Rioul, "A Discrete-Time Multiresolution Theory", Transactions on Signal Processing, Vol. 41, No. 8, Agosto 1993, pp 2591-2606.
- [15] D. Sinha y A. H. Tewfik, "Low Bit Rate Transparent Audio Compression using Adapted Wavelets", IEEE Transactions on Signal Processing, Vol. 41, No. 12, Diciembre 1993, pp 3463-3479.
- [16] W. Tompkins, "Biomedical Digital Signal Processing", Prentice Hall, Englewood Cliffs, New Jersey, 1993.
- [17] M. Unser y A. Aldroubi, "A Review of Wavelets in Biomedical Applications", Proceedings of the IEEE, Vol. 84, No. 4, Abril 1996, pp 629-638.
- [18] P.P. Vaidyanathan, "Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications : A Tutorial", Proceedings of the IEEE, Vol. 78, No. 1, Enero 1990, pp 56-242.
- [19] M. Vetterli y C. Herley , "Wavelets and Filter Banks : Theory and Design" IEEE Transactions on Signal Processing, Vol 40, No 9, Septiembre 1992, pp 2207-2232.
- [20] M. Vetterly y J. Kovacevic, "Wavelets and Subband Coding", Prentice Hall, Englewood Cliffs, New Jersey, 1995.
- [21] B. Wang y W. Liao, " Compression of ECG data by Vector Quantization", IEEE Engineering in Medicine and Biology Magazine, Vol 16, No. 4, Julio-Agosto 1997, pp 23-26.

8 APENDICES

APENDICE A

PRINCIPIO DE INCERTIDUMBRE

Partiendo de el teorema de Cauchy-Schwartz

$$\left| \int tf(t)f'(t)dt \right|^2 \leq \int |tf(t)|^2 dt \int |f'(t)|^2 dt \quad \text{nombrando cada parte de la desigualdad}$$

$$|k|^2 \leq I$$

Aplicando la definición de Δt se tiene que I es igual a :

$$I = \Delta t^2 \int |f'(t)|^2 dt = \Delta t^2 \frac{1}{2\pi} \int \Delta \omega^2 |F(\omega)|^2 d\omega, \text{ por el teorema de Parseval,}$$

$$\text{entonces, } I = \frac{1}{2\pi} \Delta t^2 \Delta \omega^2$$

Para la integral k tenemos que al integrar por partes :

$$k = \int tf(t)f'(t)dt = tf(t)\Big|_{-\infty}^{\infty} - \int f(t)^2 dt - \int tf(t)f'(t)dt$$

Pasando el termino que es igual a k al lado derecho y despejando tenemos :

$$k = \int tf(t)f'(t)dt = \frac{1}{2}tf(t)\Big|_{-\infty}^{\infty} - \frac{1}{2} \int f(t)^2 dt$$

Si $f(t)$ tiende a cero mas rápido que $\frac{1}{\sqrt{t}}$, y además $\int f(t)^2 dt = 1$, entonces $k = -1/2$

$$\text{y } \frac{1}{4} \leq \frac{1}{2\pi} \Delta t^2 \Delta \omega^2, \text{ entonces, } \frac{\pi}{2} \leq \Delta t^2 \Delta \omega^2$$

conociendo que $\omega = 2\pi f$, sustituyendo en $\Delta \omega^2$ y simplificando :

$$\Delta \omega^2 = \int 4\pi^2 f^2 |F(2\pi f)|^2 2\pi df = 8\pi^3 \int f^2 |F(2\pi f)|^2 df = 8\pi^3 \Delta f^2$$

entonces,

$$\Delta f \Delta t \geq \frac{1}{4\pi}$$

que es lo que se quería demostrar.

APENDICE B

FUNDAMENTOS DE LA DESCOMPOSICION DE SEÑALES.

En el presente apéndice se hace referencia a los fundamentos para entender las transformaciones lineales como la representación de un vector o señal en términos de otro conjunto de vectores o señales.

NOTACION

Es necesario introducir un poco de notación, los conjuntos que se van a manejar son :
Conjuntos.

- C** : Los números complejos,
- R** : Los números reales,
- Z** : Los números enteros,
- N** : Los números naturales,
- Cⁿ** : Los vectores de *n* números complejos,
- Rⁿ** : Los vectores de *n* números reales,
- W_N=e^{-j2πN}** : Números complejos de magnitud 1.

También es necesario introducir la notación para las operaciones en estos conjuntos
Operaciones.

- a*** : conjugado.
- a^t** : transpuesto
- Im(a)**: parte imaginaria
- Re(a)**: parte real.

Las funciones también pueden ser elementos de un conjunto, por ejemplo, todas las funciones que sean de variable continua, o discreta, o que estén limitadas por una cota, etc.

La notación para distinguir las funciones de tiempo continuo de las funciones de tiempo discreto o secuencias es la siguiente.

Funciones:

- f(t)** : continuas.
- x[n]** : discretas.
- d(t)** : delta de Dirac.
- d[n]** : delta de Dirac discreta.

ESPACIOS Y PRODUCTOS.

Se dice que un conjunto de elementos es un espacio, si cumple una serie de propiedades en relación a dos operaciones. Nótese que la definición no especifica ni el conjunto ni las operaciones, por lo que los elementos pueden ser escalares enteros, reales o complejos o aun vectores o señales.

El espacio vectorial **W** debe cumplir las siguientes características, si **F, G, H, 0** y **1** pertenecen al espacio **W**, y **a** y **b** son escalares, entonces :

1) Conmutabilidad y Cerradura		$F+G=G+F=H$
2) y 3) Asociatividad	$\bullet, +$	$F+(G+H)=(F+G)+H$ $F\bullet(G\bullet H)=(F\bullet G)\bullet H$
4) y 5) Distributibilidad	$+ \rightarrow \bullet$, escalar	$F\bullet(G+H)=F\bullet G+F\bullet H$ $a(F+H)=aF+aH$
6) y 7) Inversos	$\bullet, +$	$F+(-F)=0_+$ $F\bullet F^{-1}=1$
8) y 9) Neutros	$\bullet, +$	$G+0=G$ $G\bullet 1=G$

Cuando se tiene un subconjunto de W que también cumple con todas las propiedades de el espacio vectorial, se le llama Subespacio.

El conocer si un conjunto es o no un espacio vectorial, bajo ciertas operaciones es muy útil porque, todos los espacios vectoriales se pueden representar de diferentes formas como una combinación lineal de algunos de sus elementos.

Una combinación lineal se define como la sumatoria de los productos de los elementos con un conjunto de escalares, es decir,

$$y = \sum \alpha_i x_i$$

Al subconjunto de vectores que se pueden formar como una combinación lineal de los vectores x_i se le llama $\text{Span} \{x_i\}$.

Si un vector puede ser representado por una combinación lineal de un conjunto de vectores dado, se dice que ese vector es 'dependiente' con respecto a los vectores del conjunto, si no, entonces es 'independiente'

Otra de las operaciones que se puede definir entre los vectores es la de el producto interno, este se define como : $\langle x, y \rangle = \sum y^*_i x_i$, para vectores, y $\langle x, y \rangle = \int x(t) y^*(t) dt$ para funciones continuas, pero en cualquier caso el resultado es un escalar que nos da idea de la similitud entre las secuencias o funciones.

El producto interno, también llamado producto punto, tiene las siguientes características

Propiedades del Producto Interno.

$$\begin{aligned} \langle x + y, z \rangle &= \langle x, z \rangle + \langle y, z \rangle \\ \langle x, \alpha y \rangle &= \alpha \langle x, y \rangle \\ \langle x, y \rangle^* &= \langle y^*, x^* \rangle \\ \langle x, x \rangle &> 0, \text{ y } \langle x, x \rangle = 0 \text{ ssi } x = 0 \\ \langle x, y \rangle &= 0 \text{ Ortogonalidad} \end{aligned}$$

En base al producto interno se puede definir una norma o medida de los vectores en el espacio vectorial, que por las propiedades del producto interno, cumple con las siguientes propiedades.

Definición y Propiedades de la Norma.

$$\|x\| = \sqrt{\langle x, x \rangle}$$

Definición

$$\|\langle x, y \rangle\| \leq \|x\| \|y\|$$

Desigualdad de Cauchy-Schwarz

$$\|x + y\| \leq \|x\| + \|y\|$$

Desigualdad del triángulo.

$$\|x + y\|^2 + \|x - y\|^2 \leq 2(\|x\|^2 + \|y\|^2)$$

Regla del trapecio.

Ejemplos de espacios vectoriales con normas bien definidas son : l_2, L_2, C^n, R^n

Si un espacio vectorial, tiene un conjunto de vectores independientes que son capaces de generar todo el espacio con combinaciones lineales de ellos mismos, entonces a ese conjunto se le llama **base** del espacio vectorial.

Cuando los vectores que forman la base son ortogonales entre si, es decir, que : $\langle x_i, x_j \rangle = 0$ para todo $i \neq j$ y $\langle x_i, x_i \rangle = c$ para $i=j$, se dice que la base es ortogonal

Descomposición y Síntesis.

Si se tiene una base ortogonal $\{x_i\}$ de vectores de un espacio vectorial dado W , entonces todo vector que y pertenece a W puede ser representado como :

$y = \sum c_i x_i$, donde los valores c_i son escalares calculados como el producto interno entre x_i y y , es decir, $c_i = \langle x_i, y \rangle$.

APENDICE C

DEMOSTRACION DE LA TRANSFORMADA DE ONDELETA INVERSA.

En este apéndice se demuestra que la transformada de ondeleta inversa reconstruye una señal $f(t)$ a partir de su Transformada de ondeleta $F_O(a, \tau)$.

Dada la transformada de ondeleta continua $F_O(a, \tau)$ de una función $f(t)$ que pertenece a L_2 la función puede ser recuperada como:

$$f(t) = c \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_O(a, \tau) h_{a, \tau}(t) \frac{dad\tau}{a^2}$$

Si llamamos $I(a)$ a la integral sobre τ , tenemos que :

$$I(a) = \int_{-\infty}^{\infty} F_O(a, \tau) h_{a, \tau}(t) d\tau$$

Tomando la transformada de ondeleta en términos de la transformada de Fourier de la señal y la ondeleta.

$$F_O(a, \tau) = \int_{-\infty}^{\infty} h_{a, \tau}(t) f(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} H_{a, \tau}(w) F(w) dw$$

Como las ondeletas son todas generadas con los desplazamientos y escalamientos de una ondeleta prototipo, la ecuación anterior queda como :

$$F_o(a, \tau) = \int_{-\infty}^{\infty} h_{a,\tau}(t) f(t) dt = \frac{\sqrt{a}}{2\pi} \int_{-\infty}^{\infty} H(aw) F(w) e^{j\tau w} dw$$

Sustituyendo en $I(a)$, se tiene :

$$I(a) = \frac{\sqrt{a}}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} H(aw) F(w) e^{j\tau w} dw \right) h_{a,\tau}(t) d\tau = \int_{-\infty}^{\infty} H(aw) F(w) \int_{-\infty}^{\infty} h_{a,\tau}(t) e^{j\tau w} d\tau dw$$

En la última integral sustituyendo la definición de $h_{a,\tau}(t)$

$$\int_{-\infty}^{\infty} h_{a,\tau}(t) e^{j\tau w} d\tau = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} h\left(\frac{t-\tau}{a}\right) e^{j\tau w} d\tau$$

Haciendo un cambio de variable como $T = \frac{(t-\tau)}{a}$.

$$\sqrt{a} e^{j\tau w} \int_{-\infty}^{\infty} h(T) e^{-jaTw} dT = \sqrt{a} e^{j\tau w} H(aw)$$

Substituyendo este resultado en $I(a)$ tenemos

$$I(a) = \frac{|a|}{2\pi} \int_{-\infty}^{\infty} |H(aw)|^2 F(w) e^{j\tau w} dw, \text{ sustituyendo } I(a) \text{ en la fórmula de reconstrucción :}$$

$$f(t) = c \int_{-\infty}^{\infty} \frac{I(a)}{a^2} da = c \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{|H(aw)|^2}{|a|} F(w) e^{j\tau w} dw da$$

Tomando la integral sobre a y sustituyendo $A=aw$, entonces $da=dA/w$ y $1/a = w/A$

$$\int_{-\infty}^{\infty} \frac{|H(aw)|^2}{|a|} da = \int_{-\infty}^{\infty} \frac{|H(A)|^2}{|A|} dA = \frac{1}{c}, \text{ esa integral es independiente de } w, \text{ utilizando el resultado}$$

en la fórmula de reconstrucción,

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(w) e^{j\tau w} dw.$$

Lo cual sin duda es cierto por ser la transformada inversa de Fourier, completando así la demostración.



**CENTRO DE INVESTIGACION Y DE ESTUDIOS AVANZADOS DEL IPN
UNIDAD GUADALAJARA**

El jurado designado por el *Laboratorio de Ingeniería Eléctrica y Ciencias de la Computación del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional*, aprueba la tesis que presenta para su examen final de Maestría en Ciencias en la Especialidad de Ingeniería Eléctrica, José de Jesús Pérez Sevilla, el día 29 de septiembre de 1997.

Dr. Manuel Edgardo Guzmán Rentería

Dr. Manuel Mauricio Lara Barrón

Dr. Rogelio Alcántara Silva



CINVESTAV
BIBLIOTECA CENTRAL



SSIT000003813