



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería Mecánica Eléctrica

LA TRANSFORMACIÓN WAVELET EN EL PROCESADO DIGITAL DE AUDIO

Manuel Alejandro Ríos Rivas

Asesorado por el Ing. Enrique Edmundo Ruiz Carballo

Guatemala, octubre de 2009

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**LA TRANSFORMACIÓN WAVELET EN EL PROCESADO DIGITAL DE
AUDIO**

TRABAJO DE GRADUACIÓN

PRESENTADO A JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA

POR:

MANUEL ALEJANDRO RÍOS RIVAS

ASESORADO POR EL ING. ENRIQUE EDMUNDO RUIZ CARBALLO

AL CONFERÍRSELE EL TÍTULO DE
INGENIERO ELECTRÓNICO

GUATEMALA, OCTUBRE DE 2009

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Murphy Olympo Paiz Recinos
VOCAL I	Inga. Glenda Patricia García Soria
VOCAL II	Inga. Alba Maritza Guerrero de López
VOCAL III	Ing. Miguel Ángel Dávila Calderón
VOCAL IV	Br. José Milton De León Bran
VOCAL V	Br. Isaac Sultán Mejía
SECRETARIA	Inga. Marcia Ivónne Véliz Vargas

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

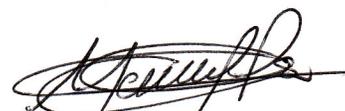
DECANO	Ing. Murphy Olympo Paiz Recinos
EXAMINADOR	Ing. Edmundo Enrique Ruiz Carballo
EXAMINADOR	Ing. Ingrid Salomé García Rodríguez de Loukota
EXAMINADOR	Ing. Otto Fernando Andrino Gonzáles
SECRETARIA	Inga. Marcia Ivónne Véliz Vargas

HONORABLE TRIBUNAL EXAMINADOR

Cumpliendo con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

**LA TRANSFORMACIÓN WAVELET EN EL PROCESADO DIGITAL DE
AUDIO,**

tema que me fuera asignado por la Dirección de la Escuela de Ingeniería Mecánica Eléctrica, con fecha 2 de septiembre de 2008.



Manuel Alejandro Ríos Rivas

Guatemala, 22 de julio de 2009

Ingeniero

Julio Cesar Solares Peñate

Coordinador Área de Electrónica

Escuela de Ingeniería Mecánica Eléctrica

Facultad de Ingeniería

Universidad de San Carlos

Guatemala

Ingeniero Solares:

Por medio de la presente, le informo que he asesorado y revisado el trabajo de graduación con título: **LA TRANSFORMACIÓN WAVELET EN EL PROCESADO DIGITAL DE AUDIO**, desarrollado por el estudiante **Manuel Alejandro Ríos Rivas**, con carné 200412789. Después de revisar su contenido final considero que satisface el objetivo del mismo y lo apruebo como trabajo de graduación.

Al agradecer su amable atención y colaboración, me suscribo atentamente.


PhD. Enrique Edmundo Ruiz Carballo

Colegiado No. 2225





FACULTAD DE INGENIERIA

Escuelas de Ingeniería Civil, Ingeniería
Mecánica Industrial, Ingeniería Química,
Ingeniería Mecánica Eléctrica, Técnica
y Regional de Post-grado de Ingeniería
Sanitaria.

Ciudad Universitaria, zona 12
Guatemala, Centroamérica

Guatemala, 03 de septiembre de 2009

Señor Director
Ing. Mario Renato Escobedo Martínez
Escuela de Ingeniería Mecánica Eléctrica
Facultad de Ingeniería, USAC.

Señor Director:

Por este medio me permito dar aprobación al Trabajo de Graduación titulado: **"LA TRANSFORMACIÓN WAVELET EN EL PROCESADO DIGITAL DE AUDIO"**, desarrollado por el estudiante **Manuel Alejandro Ríos Rivas**, ya que considero que cumple con los requisitos establecidos.

Sin otro particular, aprovecho la oportunidad para saludarlo.

Atentamente,

ID Y ENSEÑAD A TODOS


Ing. Julio César Solares Peñate
Coordinador de Electrónica





REF. EIME 64. 2009.

El Director de la Escuela de Ingeniería Mecánica Eléctrica, después de conocer el dictamen del Asesor, con el Visto Bueno del Coordinador de Área, al trabajo de Graduación del estudiante; Manuel Alejandro Ríos Rivas titulado: "LA TRANSFORMACIÓN WAVELET EN EL PROCESADO DIGITAL DE AUDIO", procede a la autorización del mismo.


Ing. Mario Renato Escobedo Martínez



GUATEMALA, 07 DE OCTUBRE 2,009.



El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería Mecánica Eléctrica, al trabajo de graduación titulado: **LA TRANSFORMACIÓN WAVELET EN EL PROCESADO DIGITAL DE AUDIO**, presentado por el estudiante universitario **Manuel Alejandro Ríos Rivas**, autoriza la impresión del mismo.

IMPRÍMASE.



Ing. Murphy Olympo Paiz Recinos
DECANO

Guatemala, octubre de 2009



/gdech

ACTO QUE DEDICO A

Dios

Por ser el pilar que sostiene la unión de mi conocimiento y fe; donde su grandeza se ve reflejada en el sentimiento más universal de todos los hombres que es el amor.

Mis padres

Marta Elvira Rivas Guevara y Carlos Aníbal Ríos Álvarez, por el amor incondicional que me han brindado durante todos estos años; así como también el regalo del conocimiento que es lo más grande que me pudieron brindar.

Mis hermanos

Carlos Aníbal, Juan Ricardo y Luis Pedro, por su bondad, tenacidad y nobleza que los ha caracterizado y me ha permitido crecer junto a ellos como persona.

Mi tía Carmen Dolores Rivas Guevara

Por ser mi otra madre, quien cuidó de mí desde pequeño y donde todos sus sacrificios permitieron que mucho de lo que hoy he podido alcanzar se deba a ella.

Mi tía Rosaura Rivas Guevara y mi primo Carlos Rivas Guevara

Por brindar su apoyo y cariño en todo momento mostrando que las cosas más pequeñas pueden tener un gran significado.

Keyla y Rubén Barahona:

Por ser mis otros hermanos en los que su apoyo me ha mostrado que Dios nos puede bendecir a través de los que uno quiere.

AGRADECIMIENTOS A:

Todos mis amigos conocidos durante el transcurso de la carrera, con los cuales compartimos buenos y malos momentos, en especial a: Sol María Girón, Marcos Morales, German Chew, Karla Orozco, Mario Alarcón, Sergio Zepeda, Werner Arriola, David Ramírez, Marlon Pérez y Juan Carlos González.

El Ing. Enrique Ruiz, por su apoyo tanto académico como personal.

El departamento de matemática, en especial al Ing. Arturo Samayoa, Inga. Glenda García, Inga. Vera Marroquín, Ing. Alfredo Gonzales y al Ing. Carlos Bolaños.

Todos aquellos que de una u otra manera han estado a mi lado y han sido partícipes de lo que ahora he alcanzado.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES	IX
LISTA DE SÍMBOLOS	XV
GLOSARIO	XIX
RESUMEN	XXVII
OBJETIVOS	XXIX
INTRODUCCIÓN	XXXI
1. FUNDAMENTOS EN EL PROCESAMIENTO DIGITAL DE SEÑALES	1
1.1 Sistemas lineales	1
1.1.1 Señales y sistemas	1
1.1.2 Requerimientos de linealidad	3
1.1.3 Linealidad estática y fidelidad sinusoidal	6
1.1.4 Propiedades especiales de la linealidad	7
1.1.5 Superposición	10
1.2 Estadística, probabilidad y ruido	11
1.2.1 Media y desviación estándar	12
1.2.2 Estadística y procesos subyacentes	14
1.2.3 Histograma, función masa de probabilidad y función densidad de probabilidad	16
1.2.4 La distribución normal	22
2. TRANSFORMACIONES DE SEÑALES	25
2.1 Transformaciones continuas	25
2.1.1 La transformada de Fourier	26

2.1.1.1	Convolución	28
2.1.1.2	Teorema de Parseval	29
2.1.2	La transformada de Fourier de corta duración	30
2.1.3	La transformada wavelet	32
2.2	Transformaciones discretas	35
2.2.1	La transformada Z	35
2.2.2	La transformada discreta en el tiempo de Fourier	39
2.2.2.1	Teorema de Parseval	40
2.2.2.2	Relación entre la transformada Z y la transformada en tiempo discreto de Fourier	40
2.2.3	La transformada discreta de Fourier	42
2.2.4	La transformada rápida de Fourier	46
2.2.4.1	Radix-2 decimación en el tiempo FFT	47
2.2.4.2	Radix-4 FFT	49
2.2.4.3	Split-radix FFT	49
3.	ESTUDIO DE LA PSICOACÚSTICA	51
3.1	Sistema auditivo humano	51
3.2	Timbre	57
3.3	Umbral absoluto y umbral de enmascaramiento	63
3.4	Fenómenos psicofísicos	65
3.4.1	Bandas críticas y filtros auditivos	65
3.4.1.1	Experimento de Fletcher: ensanchamiento de banda	66
3.4.1.2	Experimento de Zwicker: ruido recortado	67
3.4.1.3	Mediciones de Moor y Glasberg en la forma de los filtros auditivos	69
3.4.2	Patrones de enmascaramiento y patrones de excitación	71

3.4.3	Asimetría del enmascaramiento	73
3.4.4	Enmascaramiento temporal	76
4.	PROCESADO DIGITAL DE AUDIO	77
4.1	ACD DAC	77
4.1.1	Cuantización	78
4.1.2	Teorema de muestreo	83
4.1.3	Conversión digital a analógico	90
4.1.4	Filtros Analógicos	92
4.2	Filtros Digitales	95
4.2.1	Filtros FIR	99
4.2.1.1	Filtro de desplazamiento de promedio	99
4.2.1.2	Filtro de ventana <i>sinc</i>	99
4.2.1.3	Filtro modificados	100
4.2.1.4	Convolución FFT	100
4.2.2	Filtros IIR	100
4.2.2.1	Filtros de un solo polo	100
4.2.2.2	Filtros de banda estrecha	101
4.2.2.3	Filtros de respuesta de fase	101
4.2.2.4	Filtros Chebyshev	102
4.3	Multi-resolución	103
4.3.1	Operaciones básicas de la multi-resolución	104
4.3.1.1	Decimación e interpolación	104
4.3.1.2	Descomposición polifásica	108
4.3.2	Aplicación de la multi-resolución.	110
4.4	Desventajas del tratamiento clásico	112

5. WAVELETS	115
5.1 Introducción	115
5.2 Funciones wavelet	116
5.2.1 Wavelet Haar	117
5.2.2 Wavelet Daubechies	118
5.2.3 Otras wavelet ortogonales	120
5.2.4 Wavelets crudas	121
5.2.5 Wavelets biortogonales	122
5.3 Transformada wavelet continua	123
5.1.1 Wavelets para el análisis en escala de tiempo	128
5.1.2 Reconstrucción integral y semi-discreta	129
5.4 Series wavelet	131
5.4.1 Muestreo diádico	131
5.4.2 Descomposición por octavas	133
5.5 Transformada wavelet discreta	134
5.6 Análisis de bancos de filtros con Q constante	136
5.7 Bancos de filtros	137
5.7.1 Reconstrucción perfecta	139
5.7.2 Descomposición en paquetes	143
5.7.3 Resíntesis wavelet	146
5.8 Relación entre los bancos de filtros y la transformada wavelet	147
5.9 Propiedades de la transformada wavelet	151
5.10 Aplicaciones de las wavelet	152
5.10.1 La transformada wavelet en el procesamiento de señales de audio	153

6.	WAVELET HAAR	155
6.1	Introducción	155
6.2	Función Haar de escala	155
6.2.1	Relación de ortogonalidad y normalidad	165
6.3	Función wavelet Haar	167
6.3.1	Relación de ortogonalidad y normalidad	175
6.4	Análisis multi-resolución	177
6.5	Transformada wavelet discreta (DWT)	179
6.6	Transformada rápida wavelet (FWT) y banco de filtros	186
6.6.1	Análisis o descomposición de señales	186
6.6.2	Síntesis o reconstrucción de señales	189
7.	WAVELETS EN EL PROCESADO DE SEÑALES DE AUDIO	199
7.1	Introducción	199
7.2	Requerimientos y consideraciones	200
7.3	Reducción de ruido de una señal de audio	203
7.3.1	Umbral universal	205
7.3.2	Ejemplo de reducción de ruido estacionario	206
7.3.3	Ejemplo de reducción de ruido no estacionario	211
7.3.3.1	Ruido con cambios en la varianza	211
7.3.3.2	Ruido con estructura desconocida	216
7.3.4	Aplicación: reducción de distintos niveles de ruido a una señal de audio	216
7.4	Compresión de señales de audio	223
7.4.1	Métodos de compresión	224
7.4.1.1	Umbral global	224
7.4.1.2	Umbral por nivel	226
7.4.2	Comparación de los umbrales de compresión	227

8.	WAVELETS EN LA CODIFICACIÓN PERCEPTUAL DE AUDIO	231
8.1	Introducción	231
8.2	Estructura de un codificador perceptual de audio genérico	232
8.2.1	Consideraciones del banco de filtros	234
8.2.2	Modelo psicoacústico	237
8.2.2.1	Modelo fisiológico	237
8.2.2.2	Modelo en base a la excitación de patrones	238
8.2.2.3	Modelo en base a patrones de enmascaramiento	239
8.2.3	Etapas de cuantización y codificación	240
8.2.4	Codificación sin ruido	242
8.3	Resolución tiempo-frecuencia	242
8.3.1	Estructura del árbol wavelet	243
8.3.2	Base del filtro wavelet	244
8.3.3	Conclusión	245
8.4	Representación de señales de audio usando WFB	246
8.4.1	Representación de las sub bandas en el dominio de la frecuencia	246
8.4.2	Orden de las sub bandas	249
8.4.3	Localización en el dominio del tiempo	251
8.5	Optimización del WFB	254
8.5.1	Minimización del traslape entre bandas por umbral	254
8.5.2	Algoritmo modificado de Remez para filtros ortogonales en cuadratura	255
8.5.3	Filtro de 32 bandas modificado	257

CONCLUSIONES	261
RECOMENDACIONES	263
REFERENCIAS BIBLIOGRÁFICAS	265
BIBLIOGRAFÍA	267
ANEXOS	271

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Sistema continuo y discreto.	2
2.	Homogeneidad de un sistema discreto.	3
3.	Aditividad de un sistema discreto.	4
4.	Invarianza al desplazamiento en un sistema discreto.	5
5.	Propiedad conmutativa de un sistema lineal.	8
6.	Sistema con múltiples entradas y múltiples salidas.	9
7.	Linealidad de la multiplicación.	9
8.	Síntesis y descomposición.	10
9.	Relación entre la desviación estándar y el valor pico a pico.	14
10.	Procesos no estacionarios.	15
11.	Proceso estacionario	15
12.	Histograma de señal luego de conversión analógica a digital.	17
13.	Histograma y función masa de probabilidad.	18
14.	Función densidad de probabilidad.	19
15.	Función densidad de probabilidad de algunas funciones	21
16.	Distribución normal o Gaussiana	23
17.	Ventana cuadrada para $\omega = \pi$ y $t = 8$.	31
18.	Ventana cuadrada para $\omega = 6\pi$ y $t = 8$.	32
19.	Transformada wavelet: señal y función de análisis para $a=1/2$	34

20.	Transformada wavelet: señal y función de análisis para $a=1/4$	34
21.	Diagramas de flujo equivalentes.	47
22.	Diagrama de flujo para una FFT de 8 puntos.	48
23.	Diagrama funcional del oído humano.	52
24.	Detección de fase del oído humano.	59
25.	Forma de onda generada por un violín.	60
26.	El teclado de un piano.	62
27.	Umbral absoluto de audición.	64
28.	Experimento de Fletcher.	66
29.	Experimento de Zwicker.	68
30.	Experimento de Moor y Glasberg.	69
31.	Bandas críticas de Zwicker y Moor.	70
32.	Patrón de enmascaramiento de ruido de banda estrecha centrado en 410 Hz.	71
33.	Asimetría del enmascarado.	75
34.	Enmascaramiento no simultáneo.	76
35.	Formas de onda de la digitalización de una señal.	79
36.	Error de cuantización.	80
37.	Muestreo de señales.	83
38.	Relación de frecuencia y fase en el muestreo de señales.	86
39.	El teorema de muestreo en el dominio del tiempo y la frecuencia.	88
40.	Espectro de frecuencia de filtro rectangular y señal reconstruida.	91
41.	Diagrama de bloques de un sistema de DSP.	93
42.	Banco de filtros de M-canales.	104
43.	Componentes típicos de un banco de filtros.	105

44.	Espectro para las señales decimadas e interpoladas sin efectos de aliasing.	107
45.	Espectro para las señales decimadas e interpoladas bajo los efectos de aliasing.	108
46.	Función wavelet Haar.	117
47.	Familia de funciones wavelet Daubechies.	119
48.	Respuesta en frecuencia de funciones wavelet Daubechies.	120
49.	Comparación de los núcleos de análisis de la STFT y la WT.	124
50.	Comparación de la resolución de la STFT y la WT.	127
51.	Escalogramas para wavelet real y analítica.	130
52.	Muestreo diádico de la transformada wavelet.	132
53.	Banco de filtros simple.	138
54.	Banco de filtros de 2 canales y 4 salidas.	139
55.	Banco de filtros de análisis y resíntesis.	140
56.	Traslape de frecuencias de filtros paso alto y paso bajo.	141
57.	Patrón alternante de signos.	141
58.	Patrón de giro alternativo.	142
59.	Paquetes de arboles wavelet	143
60.	Árbol wavelet.	144
61.	Comparación STFT y WT.	146
62.	Resíntesis wavelet de 3 niveles.	147
63.	Wavelet tipo Daubechies 2.	149
64.	Coeficientes de filtro paso bajo y paso alto.	150
65.	Respuesta en frecuencia de los filtros de análisis.	150
66.	Función de escala Haar.	156
67.	Escalado de función $f(t) = t^2$.	161

68.	Escalado de función $f(t) = t^2$ usando la función de escalamiento en v_1 .	164
69.	Función wavelet Haar.	169
70.	Aproximación de la función $f(t) = t^2$ mediante la función wavelet en el espacio W_0 .	172
71.	Aproximación de la función $f(t) = t^2$ mediante la función wavelet en el espacio W_1 .	175
72.	Espacios escalados y wavelet	178
73.	Función de escalado y wavelet para espacio W_2	182
74.	Algoritmo de análisis, DWT	193
75.	Algoritmo de síntesis, IDWT	184
76.	Reconstrucción a partir de aproximación y detalles	195
77.	Análisis multi-resolución para determinar vector DWT	196
78.	Reconstrucción de señal a partir del vector DWT	197
79.	Umbral duro y suave.	204
80.	Señal producida por efecto Doppler con ruido aditivo.	206
81.	Análisis y detalle de señal Doppler.	208
82.	Aplicación del umbral suave sobre los coeficientes de detalle.	209
83.	Reconstrucción de la señal a Doppler.	209
84.	Residuos de la señal a Doppler.	211
85.	Señal con ruido de múltiples varianzas.	213
86.	Reducción de ruido usando múltiples umbrales.	214
87.	Aplicación del umbral suave y duro.	215
88.	Comparación de residuos.	215
89.	Diagrama de reducción de ruido usando Simulink.	217
90.	Relación S/N vs. potencia umbral.	219

91.	Relación S/N vs. energía retenida.	220
92.	Nivel de potencia del umbral vs. energía retenida.	221
93.	Curva de compresibilidad de la señal Doppler.	225
94.	Curva de compresibilidad para distintas wavelets.	226
95.	Compresión aplicando el umbral global.	228
96.	Compresión aplicando el umbral por nivel.	229
97.	Codificador y decodificador perceptual de audio genérico.	233
98.	Estructura de tres WFB.	244
99.	Filtro de 32 bandas Daubechies.	247
100.	Filtro para la sub banda no.18.	248
101.	Efectos de filtrado y sub muestreo.	249
102.	Secuencia natural y secuencia relacionada a cada banda en un WFB de 32 bandas.	251
103.	Coeficientes necesarios al convolucionar con un filtro.	252
104.	Filtros obtenidos a partir del algoritmo modificado de Remez.	256
105.	Filtro de 32 bandas con $L = 32$, $K = 2$ y $B = 0.05$.	258
106.	Diagrama de simulación de reducción de ruido por medio de Simulink.	
107.	Diagrama de interno del bloque umbral para el caso de múltiples umbrales.	
108.	Diagrama de interno del bloque umbrales.	
109.	Diagrama de interno del bloque cálculo umbrales.	
110.	Diagrama de interno del bloque umbrales dinámicos.	
111.	Diagrama de interno del bloque umbral para el caso de un único umbral.	

TABLAS

I.	Propiedades importantes de la transformada de Fourier.	28
II.	Propiedades importantes de la transformada Z.	38
III.	Propiedades importantes de la transformada de Fourier en tiempo discreto.	39
IV.	Propiedades importantes de la transformada discreta de Fourier.	45
V.	Unidades de la intensidad del sonido.	55
VI.	Coeficientes de reconstrucción para la función $f(t) = t^2$ usando la función de escalamiento en v_0 .	161
VII.	Coeficientes de reconstrucción para la función $f(t) = t^2$ usando la función de escalamiento en v_1 .	164
VIII.	Coeficientes para la función $f(t) = t^2$ usando la función wavelet Haar en W_0 .	172
IX.	Coeficientes para la función $f(t) = t^2$ usando la función wavelet Haar en W_1 .	174
X.	Intervalos de tiempo para $f_{j,k}(t)$.	181
XI.	Energía y potencia al aplicar distintos criterios de umbral.	218
XII.	Valores de M.	227
XIII.	Soporte en el tiempo para distintos filtros.	253

LISTA DE SÍMBOLOS

ADC	Convertidor analógico a digital
a_j	Aproximación de nivel j .
<i>Biorn.m</i>	Wavelet biortogonal de análisis de grado n y síntesis de grado m .
DAC	Convertidor digital a analógico
Db	Decibel
DFT	Transformada discreta de Fourier
d_j	Detalle de nivel j .
<i>Dn</i>	Wavelet Daubechies de grado n .
DSP	Procesamiento digital de señales.
DWT	Transformada wavelet discreta.

FFT	Transformada rápida de Fourier.
FIR	Respuesta finita al impulso.
Hz	Hertz
IIR	Respuesta infinita al impulso.
JND	Mínima distorsión notable.
MATLAB	Matrix Laboratory, Laboratorio matriz
MPEG	Moving Picture Expert Group, Grupo de expertos en imágenes en movimiento
OSHA	<i>Occupational Safety and Health Administration</i> , Seguridad ocupacional y administración de salud (Entidad gubernamental estadounidense)
PER	Porcentaje de Energía Retenida
QMF	Filtro espejo en cuadratura.
RMS	Valor cuadrático medio o valor eficaz.

SNR	Relación señal a ruido
SPL	Nivel de presión del sonido
Symn	Wavelet symmlet de grado n .
V	Volts
V_j	Sub espacio de escala j .
WFB	Banco de filtros wavelet
W_j	Sub espacio wavelet j .
WT	Transformada wavelet
μ	Media
ϕ	Función de escala.
σ^2	Varianza
ψ	Función wavelet

GLOSARIO

Algoritmo	Es una lista bien definida, ordenada y finita de operaciones que permite hallar la solución a un problema dado un estado inicial y una entrada.
Aliasing	Es el efecto que causa que señales continuas distintas se tornen indistinguibles cuando se les muestrea digitalmente.
Análisis	Se refiere al proceso de descomposición de una señal por medio de un banco de filtros en el análisis multiresolución.
Análisis multiresolución	Consiste en la observación de señales a distintas escalas de resolución y desarrollar representaciones de una señal complicada en términos de sus bases ortonormales.
Aproximación	Es el resultado de realizar una descomposición de una señal con la función de análisis wavelet.
Banco de filtros	Se refiere al arreglo de filtros pasa banda que separa la señal de entrada en distintas componentes, cada una posee una sola sub banda de frecuencias de la señal original.
Banda crítica	Son rangos de frecuencia dentro de los cuales un tono bloquea la percepción de otro tono.

Biortogonal	Se denomina biortogonal a la pareja de espacios vectoriales que su sub conjunto de espacios indexados son duales el uno con otro.
Bit	Es la unidad mínima de información empleada en cualquier dispositivo digital, sus estados pueden ser 0 (apagado) o 1 (encendido).
Codificación perceptual	Una técnica para comprimir aún más el sonido digital mediante la eliminación de las frecuencias que no pueden ser percibidas por el oído humano.
Compansión	Es un proceso que involucra la compresión y expansión del rango dinámico de una señal para mejorar la transmisión de la misma sobre un canal limitado.
Convolución	Es una operación matemática de dos funciones que produce una tercera función, que es vista como una versión trasladada de una de las funciones originales.
Cuantización	Se dice del proceso de aproximar un rango continuo de valores por una parte relativamente pequeña y finita de valores discretos o valores enteros.
Decimación	Es un proceso donde se disminuye la frecuencia de muestreo de una secuencia a una tasa más baja. El proceso consiste de un filtrado paso bajo de los datos de entrada para posteriormente ser sub muestreados a una tasa menor a la de entrada.

Densidad espectral	Es una función matemática que informa cómo se encuentra distribuida la potencia o la energía de una señal sobre las distintas frecuencias de las que está formada.
Detalle	Se refiere al resultado de la descomposición de una señal con una función wavelet.
Enmascaramiento frecuencial	Es la disminución de la sonoridad de un tono a una cierta frecuencia, en presencia de otro tono simultáneo a una frecuencia diferente.
Estadística	Es una ciencia con base matemática referente a la recolección, análisis e interpretación de datos, que busca explicar condiciones regulares en fenómenos de tipo aleatorio.
Función de escala	Se refiere a la función que por medio de dilataciones y traslaciones da origen a la función wavelet madre. En el análisis multiresolución permite encontrar los coeficientes de análisis.
Filtro digital	Es un sistema que realiza operaciones matemáticas a una señal discreta en el tiempo para reducir o mejorar determinados aspectos de la señal.
Filtro electrónico	Es un elemento que discrimina una determinada frecuencia o gama de frecuencias de una señal eléctrica que pasa a través de él, pudiendo modificar tanto su amplitud como su fase.

Homogeneidad

Propiedad de un sistema lineal donde un cambio en la amplitud de la señal de entrada resulta en un cambio correspondiente en la amplitud de la señal de salida.

Interpolación

Es un proceso donde se aumenta la frecuencia de muestreo de una secuencia a una tasa más elevada. Se realiza mediante la inserción de ceros en la secuencia original y luego aplicar un filtro especial paso bajo.

Muestreo

Consiste en tomar muestras periódicas de la amplitud de una señal analógica, sobre intervalos constantes. El ritmo de este muestreo determina el número de muestras que se toman en un intervalo de tiempo.

Momento de desvanecimiento

Cuando una función es representada por medio de la transformada Z un momento de desvanecimiento se refiere al número de ceros que se encuentran situados en π .

Octava

Es el intervalo que separa dos sonidos cuyas frecuencias fundamentales tienen una relación de dos a uno.

Ortogonalidad

Se refiere a la propiedad de dos vectores de que su proyección escalar sea cero.

Ortonormal	Se refiere al conjunto de vectores que pertenecen a un espacio vectorial cuyo producto interno es cero (ortogonal) y la norma de cada uno de sus vectores es igual a uno (normal).
Principio de lugar	Establece como el sistema nervioso detecta que lugar de la membrana basilar ha sido estimulada en base a su resonancia, así como también la intensidad a partir de la cantidad de células ciliares estimuladas y la magnitud de su deformación.
Principio de Volley	Consiste en el trabajo conjunto de distintas neuronas, con potenciales de acción desfasados, para representar frecuencias más altas de las que normalmente una sola neurona podría representar.
Probabilidad	Se encarga de medir la frecuencia con la que se obtiene un resultado (o conjunto de resultados) al llevar a cabo un experimento aleatorio, del que se conocen todos los resultados posibles, bajo condiciones suficientemente estables.
Psicoacústica	Estudia la percepción subjetiva de las características del sonido: intensidad, tono y timbre. Además estudia la relación entre la sensación percibida por un sonido determinado o la forma en la que se escucha.

Psicofísica

Es una rama de la psicología que estudia la relación entre la magnitud de un estímulo físico y la intensidad con la que éste es percibido por parte de un observador.

Ruido blanco

Es una señal aleatoria que se caracteriza en que sus valores en dos instantes de tiempo distintos no guardan correlación estadística. Su densidad espectral de potencia es constante.

**Ruido blanco
Gaussiano**

Posee las mismas propiedades que el ruido blanco, pero su distribución de probabilidad tiene una distribución Gaussiana.

Señal

Cualquier cantidad física cuyas variaciones pueden ser medidas en el tiempo y/o el espacio.

Señal continua

Consiste en una señal que puede expresarse como una función cuyo dominio se encuentra en el conjunto de los números reales, y normalmente es el tiempo.

Señal discreta

Consiste en una serie temporal que puede expresarse como una función cuyo dominio consiste de valores discretos enteros, donde cada valor es denominado muestra.

Síntesis

Se refiere al proceso inverso al análisis en un banco de filtros y consiste en la reconstrucción de una señal por medio de un banco de filtros en un análisis multiresolución.

Sistema lineal	Es un sistema que obedece las propiedades de escalado (homogeneidad) y de superposición (aditividad).
Timbre	Se determina por el contenido armónico de una señal y permite distinguir la misma nota producida por dos instrumentos musicales diferentes.
Tono	Es la frecuencia de la componente fundamental en el sonido que lo permite caracterizar como más agudo o más grave.
Ventana	Consiste en una función que decrece suficientemente rápido hacia cero fuera de un intervalo elegido. Son usadas con frecuencia en el análisis y el procesamiento de señales para evitar las discontinuidades al principio y al final de los bloques analizados.
Volumen	Es la percepción subjetiva que el ser humano tiene de la potencia de un determinado sonido.
Wavelet	Es una onda con una amplitud de oscilación, que comienza en cero, aumenta, y luego disminuye de nuevo a cero. Se usa para dividir una determinada función o señal en tiempo continuo en diferentes componentes de escala

RESUMEN

En el análisis de señales comúnmente existe el concepto de transformación de las mismas, y cuando este análisis se convierte en un procesado muchas veces involucra también un tratamiento de la señal descompuesta. Clásicamente este análisis se realiza en el dominio de la frecuencia con la transformada de Fourier, sin embargo esta falla cuando se trata de una localización en el tiempo o en la misma frecuencia. Es bajo estas circunstancias que la transformación wavelet ha obtenido gran aceptación como una herramienta alternativa o complementaria en el análisis de señales.

Mucho del procesamiento digital se realiza exclusivamente a sistemas lineales y se emplea en gran medida métodos estadísticos para poder describirlos. El capítulo 1 presenta una descripción de los fundamentos en el procesamiento digital de señales y la relación que posee con los sistemas lineales y la estadística.

Inherente al procesamiento digital de señales se encuentra la transformación y manipulación de señales, el capítulo 2 aborda una descripción desde el punto de vista de las transformaciones clásicas y ya bien establecidas, para entender cuál es la fortaleza y debilidad de cada una de ellas.

El entendimiento del sistema auditivo humano ha permitido que se desarrollen esquemas cada vez más complejos que se asemejen al mismo. En el capítulo 3 se presenta una descripción del sistema auditivo humano que pretende describir los procesos básicos y el funcionamiento del mismo. El capítulo 4 pretende fundir los tres primeros capítulos y que el lector pueda tener una visión global en el tratamiento digital de señales de audio.

El concepto formal de wavelet se presenta en el capítulo 5, se presenta la transformación continua y discreta para posteriormente abordar el tema de los bancos de filtros; siendo los bancos de filtros la piedra angular en tratamiento basado en wavelets. La wavelet Haar ha demostrado ser la wavelet más sencilla de construir, de ahí que en el capítulo 6 se presente una serie de aplicaciones usando la misma, cuyo objetivo es el de entender el proceso que involucra la transformación con wavelets para posteriormente usar wavelets más complejas.

La aplicación de la transformación wavelet por medio de bancos de filtros es abordada en el capítulo 7. Se parte con la reducción de ruido de una señal de audio donde se presentan los resultados de una simulación en Simulink y se muestra la comparación entre distintos criterios para abordarla. Posteriormente se presenta la capacidad intrínseca que poseen las wavelets como métodos de compresión para las señales de audio. Finalmente el capítulo 8 es un resumen de cómo los bancos de filtros basados en wavelet pueden ser aplicados y mejorados para su utilización en la codificación perceptual de audio.

OBJETIVOS

General

Presentar y estudiar la transformada wavelet mostrando su aplicación al procesamiento digital de señales de audio.

Específicos:

1. Establecer los fundamentos del procesamiento de señales, así como también los requisitos que implica el procesado digital de audio.
2. Comparar la transformada wavelet frente a otros tipos de transformadas en el tratamiento de señales.
3. Estudiar la wavelet Haar, dada su sencillez, para comprender el concepto de función de la escala y el análisis multiresolución para el posterior uso de wavelets más complejas.
4. Mostrar la aplicación de la transformada wavelet en la reducción de ruido de una señal de audio por medio de simulaciones en MATLAB, así como también la inherente capacidad que presenta en la compresión de señales.
5. Mostrar la factibilidad del uso de bancos de filtros wavelet para la codificación perceptual de audio en base al entendimiento de los fenómenos psicoacústicos del sistema auditivo humano.

INTRODUCCIÓN

En el análisis de señales existe una gran variedad de transformaciones que pueden ser aplicadas a una señal, de manera que se ajuste a los distintos campos de estudio. La transformación más popular ha sido la transformada de Fourier, debido a que convierte una señal en el dominio del tiempo a una señal en el dominio de la frecuencia. Su utilidad ha sido comprobada en una variedad de aplicaciones, sin embargo no se encuentra basada en el tiempo. Posteriormente para combatir este problema se desarrolló la transformada de Fourier de corta duración. Desafortunadamente, esta transformada también falla cuando se trabaja sobre resoluciones de alta y baja frecuencia al mismo tiempo.

El concepto como tal de las wavelets surgió en 1984 con Alex Grossmann y Jean Morlet, aunque ya en 1930 las técnicas descritas por *Littlewood* y *Paley* esbozaban un futuro prometedor para este nuevo análisis. Esencialmente el análisis por medio de wavelets consiste en descomponer una señal en distintos niveles de escala, lo que permite trabajar sobre distintas resoluciones de tiempo y frecuencia, comúnmente conocidos como análisis y detalles. Luego de que en 1987 Ingrid Daubechies construyera la wavelet que lleva su nombre y junto con la revolución computacional de la época, cada día se establece la transformación por wavelets como una de las preferidas para el procesamiento de señales.

El creciente interés por las wavelets ha demostrado que es una herramienta que aún tiene mucho por ofrecer y en comparación con otras transformaciones de señales, que llevan muchos años en desarrollo, ha llegado a alcanzar el éxito en muy poco tiempo.

Muestra de los especiales beneficios de la transformada wavelet es su capacidad inherente de descomponer, por medio de bancos de filtros con selectividad constante, la cual la hace muy útil en el procesamiento digital de señales de audio. Esta capacidad permite que sea de especial interés, como un método complementario o alternativo al tratamiento de Fourier, en la reducción de ruido de señales, detección de patrones, detección de transientes e incluso en la codificación perceptual de audio.

El funcionamiento adecuado de la transformación wavelet depende en gran medida de la función madre que se utilice para representar una señal y de ahí que sus resultados sean exitosos o no. Entender las propiedades de la señal a estudiar, así como también lo que se desea llegar a alcanzar con el análisis permitirá alcanzar resultados satisfactorios. De ahí que muchos autores consideren que el uso de las wavelets no es solamente la aplicación de una ciencia sino más bien el desarrollo de un arte.

1. FUNDAMENTOS EN EL PROCESAMIENTO DIGITAL DE SEÑALES

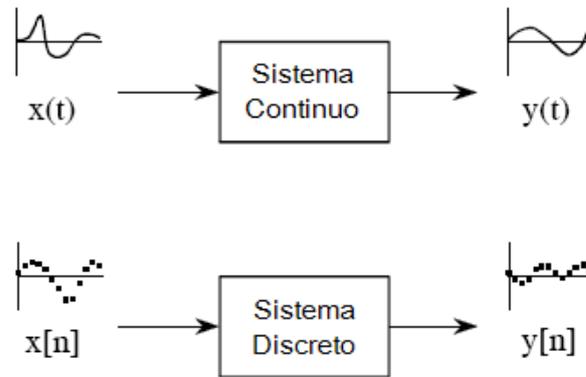
1.1 Sistemas lineales

Muchas técnicas de procesamiento digital de señales, DSP, se basan en la estrategia de dividir y conquistar (*divide-and-conquer*) llamada superposición. En donde, la señal procesada es dividida en componentes simples, cada componente es procesado, y los resultados son reunidos. Esto tiene la ventaja de convertir una tarea muy difícil en varias más sencillas. La superposición solo puede ser usada con sistemas lineales, con los cuales ciertas reglas matemáticas son aplicables.

1.1.1 Señales y sistemas

Una señal es una descripción de cómo un parámetro varía con otro parámetro. Por ejemplo, el voltaje a través de un circuito electrónico. Un sistema es un proceso que produce una señal de salida en respuesta a una señal de entrada. La figura 1 muestra un diagrama de bloques de un sistema, tanto continuo como discreto; en este caso una entrada continua produce una salida continua, y una entrada discreta produce una salida discreta.

Figura 1. Sistema continuo y discreto.



Pueden existir muchas razones por las cuales se quiera entender un sistema. Por ejemplo, se puede querer remover el ruido en una señal de audio, se puede desear remover eco en una grabación de audio e incluso agregar efectos a una grabación musical. En otros casos, el sistema puede tener una distorsión o un efecto de interferencia que se desea medir. Por ejemplo, en una transmisión telefónica la señal de voz de salida es muy distinta a la señal de entrada. Si se entiende como la línea de transmisión, en este caso el sistema, cambia la señal se podría compensar este efecto. Existen además otros casos, donde el sistema representa un proceso físico que se desea analizar o estudiar. Como es el caso del radar y el sonar, en los cuales estos métodos comparan la señal transmitida y reflejada para determinar la posición de un objeto remoto. En términos de la teoría de sistemas, el problema es encontrar un sistema que cambie la señal transmitida en la señal recibida.

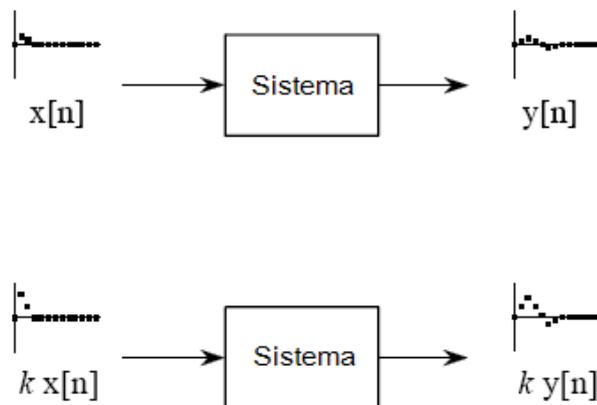
Entender todos los posibles sistemas en el mundo sería una tarea exuberante, sin embargo, la mayoría de los sistemas útiles caen en la categoría de sistemas lineales. Sin el concepto de sistema lineal, nos veríamos forzados a examinar las características de cada sistema no relacionado. Con esta aproximación, nos podemos enfocar en los rasgos de la categoría de un sistema lineal como un todo. Con lo cual se debe establecer cuáles son las propiedades que hacen de un sistema lineal.

1.1.2 Requerimientos de linealidad

Un sistema es llamado lineal si posee dos propiedades matemáticas: homogeneidad y aditividad. Si se puede demostrar que un sistema posee estas dos propiedades, entonces se ha comprobado que un sistema es lineal. De igual manera, si un sistema no posee una o ninguna de estas propiedades el sistema no es lineal. Una tercera propiedad, invarianza al desplazamiento, no es un requerimiento estricto para la linealidad, pero es propiedad mandataria para muchas técnicas de DSP. Cuando se utiliza el término de sistema lineal en DSP, se debe de asumir que incluye la invarianza al desplazamiento a menos que se tenga razón alguna para creer lo contrario. Estas tres propiedades forman la matemática de cómo la teoría de sistemas lineales es definida y usada.

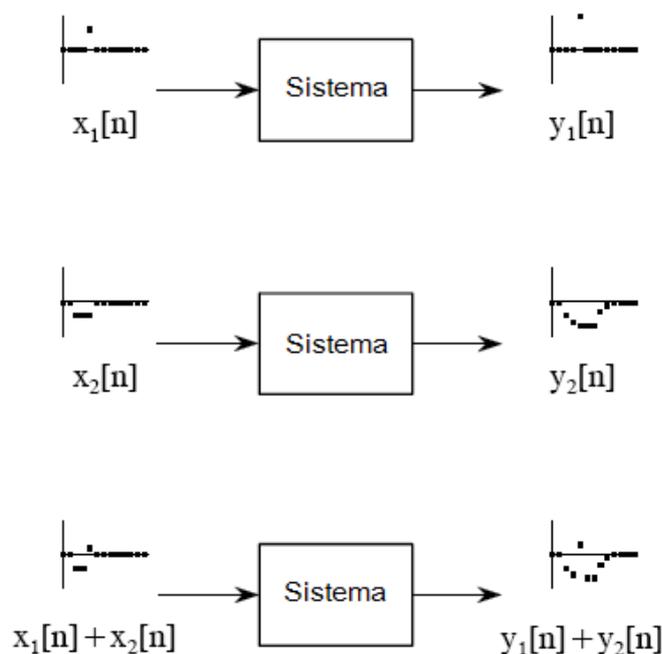
Homogeneidad significa que un cambio en la amplitud de la señal de entrada resulta en un cambio correspondiente en la amplitud de la señal de salida. Si a una señal de entrada $x[n]$ resulta en una señal de salida $y[n]$, una entrada de $kx[n]$ resulta en una salida de $ky[n]$, para cualquier señal de entrada y constante k . La figura 2 muestra este concepto.

Figura 2. Homogeneidad de un sistema discreto.



La propiedad aditiva se muestra en la figura 3. Considerando un sistema donde una entrada $x_1[n]$ produce una salida $y_1[n]$ y una entrada $x_2[n]$ produce una salida $y_2[n]$. Se dice que el sistema es aditivo si la entrada $x_1[n] + x_2[n]$ resulta en la salida $y_1[n] + y_2[n]$, para todas las posibles señales de entrada. Un sistema es aditivo si las señales sumadas en la entrada producen señales que son sumadas a la salida. El punto de importancia aquí es, que las señales sumadas pasen a través del sistema sin interactuar entre ellas.

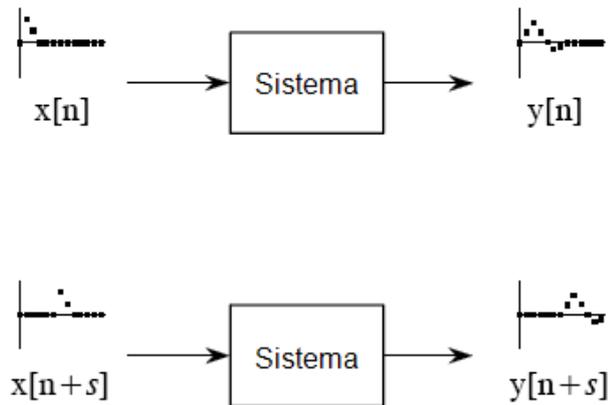
Figura 3. Aditividad de un sistema discreto.



Invarianza al desplazamiento significa que un desplazamiento en la señal de entrada resultará en nada más que un desplazamiento idéntico en la señal de salida. De manera más formal, si una señal de entrada $x[n]$ resulta en una señal de salida $y[n]$, una señal de entrada $x[n - s]$ resulta en una señal de salida $y[n - s]$, para cualquier señal de entrada y cualquier constante s . Al agregar una constante s , a la variable independiente n , la forma de onda puede ser adelantada o retardada en la dirección horizontal.

Por ejemplo, cuando $s = 3$, la señal es desplazada a la izquierda tres muestras; cuando $s = -3$, la señal es desplazada tres muestras a la derecha. La figura 4 muestra el concepto de invarianza al desplazamiento.

Figura 4. Invarianza al desplazamiento en un sistema discreto.



La invarianza es importante porque significa que las características del sistema no cambian con el tiempo (o cualquiera que sea la variable independiente). La mayoría de los sistemas encontrados y analizados son invariantes al desplazamiento. Esto es afortunado, porque es difícil tratar con sistemas que cambian sus características cuando se encuentran operando.

Por ejemplo, si se diseña un filtro digital para compensar los efectos de degradación de una línea de transmisión telefónica. El filtro se encarga de hacer la voz más natural y fácil de entender. Sin embargo, cuando llega el invierno se observa que las características de la línea telefónica han cambiado con la temperatura. El filtro antes creado ahora ya no funciona bien. Esta situación requerirá un algoritmo más sofisticado que se pueda adaptar a las condiciones cambiantes.

La linealidad es un concepto muy amplio, que incluye mucho más que señales y sistemas, por lo que muchas veces el concepto de invarianza al desplazamiento no tiene sentido alguno. Por lo que la invarianza al desplazamiento puede ser pensada como un aspecto adicional para la linealidad necesaria cuando señales y sistemas se encuentran involucrados.

1.1.3 Linealidad estática y fidelidad sinusoidal

Homogeneidad, aditividad e invarianza son importantes porque ellas proveen las bases matemáticas para definir los sistemas lineales. Desafortunadamente, estas propiedades solas no proveen a los científicos e ingenieros con un criterio de que son los sistemas lineales. Las propiedades de linealidad estática y fidelidad sinusoidal son de bastante ayuda aquí. Estas no son de especial importancia desde un punto de vista matemático, pero se relacionan a cómo los humanos piensan acerca de los sistemas lineales.

La linealidad estática define como un sistema lineal reacciona cuando las señales no se encuentran cambiando, por ejemplo, cuando son DC o estáticas. La respuesta estática de un sistema lineal es muy simple: la salida es la entrada multiplicada por una constante. Todos los sistemas lineales tienen la propiedad de linealidad estática. Lo opuesto usualmente es cierto, pero no siempre. Existen sistemas que muestran linealidad estática solamente. En estos sistemas no importa si la señal de entrada es estática o cambiante. Estos son llamados sistemas sin memoria, porque su salida solo depende en el estado presente de su entrada, y no de su historia. Por ejemplo, la corriente instantánea en un resistor depende solamente en el voltaje aplicado a él, y no en como los valores de la señal llegaron a ser los valores que en ese momento son. Si un sistema posee linealidad estática, y es sin memoria, entonces el sistema debe ser lineal. Esto provee una manera de entender (y de probar) la linealidad de estos sistemas.

Una característica importante de los sistemas lineales es cómo se comportan con las sinusoidales o fidelidad sinusoidal. Si la entrada a un sistema lineal es una onda sinusoidal, la salida también será una onda sinusoidal, y con la misma frecuencia que la entrada. Sinusoides son las únicas formas de onda que poseen esta propiedad. Aunque una sinusoidal en la entrada garantiza una sinusoidal a la salida, las dos pueden diferir en amplitud y fase.

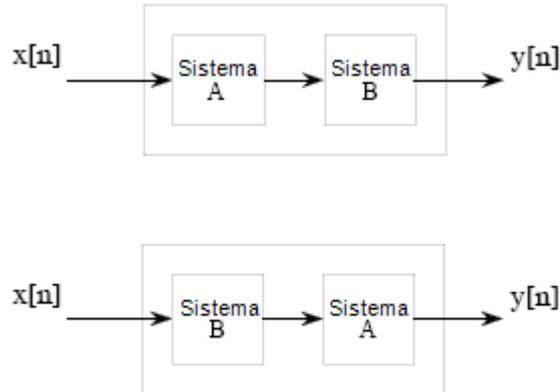
Un sistema que siempre produce una señal sinusoidal en respuesta a una entrada sinusoidal no necesariamente es lineal.

1.1.4 Propiedades especiales de la linealidad

La linealidad es conmutativa, una propiedad que envuelve la combinación de dos o más sistemas. En el caso de dos sistemas combinados en cascada, como los de la figura 5, la salida de un sistema es la entrada del siguiente. Si cada sistema es lineal, entonces la combinación total es lineal. La propiedad conmutativa establece que el orden de los sistemas en cascada puede ser reordenado sin afectar las características totales del sistema.

Otro tipo de propiedades es el de los sistemas con múltiples entradas y/o múltiples salidas, que serán lineales si está compuesto por subsistemas lineales y adiciones de señales. La complejidad no importa, solo que nada no lineal es permitido adentro del sistema.

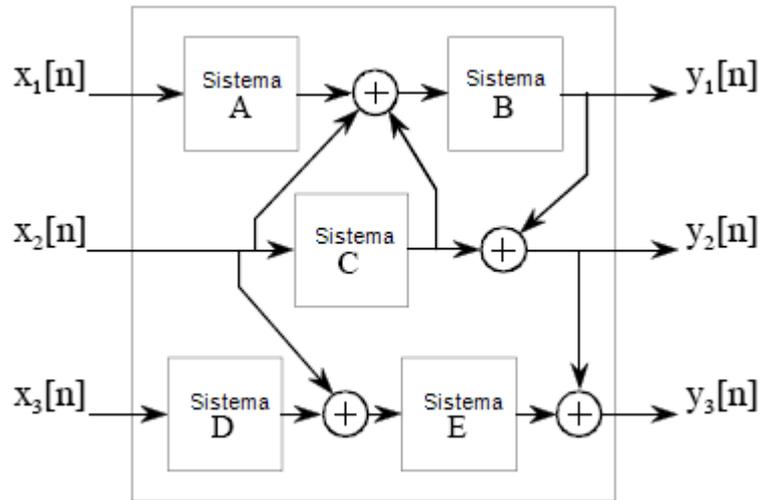
Figura 5. Propiedad conmutativa de un sistema lineal.



Para comprender lo que significa la linealidad en sistemas con múltiples entradas y/o múltiples salidas, se puede considerar lo siguiente. Si se hace que en una entrada exista una señal mientras que las otras entradas son llevadas a cero. Esto causará que las múltiples salidas respondan con algún patrón de señales. Ahora, si se repite este procedimiento con otra señal en otra entrada, manteniendo las otras entradas a cero. Esta segunda entrada resultará en otro patrón de señales en las múltiples salidas. Si ahora se situasen ambas señales en sus respectivas entradas simultáneamente. Las señales que aparecerán en las salidas serán simplemente la superposición de las señales de salida producidas cuando las señales de entrada fueron aplicadas por separado. La figura 6 presenta un sistema con múltiples entradas y múltiples salidas.

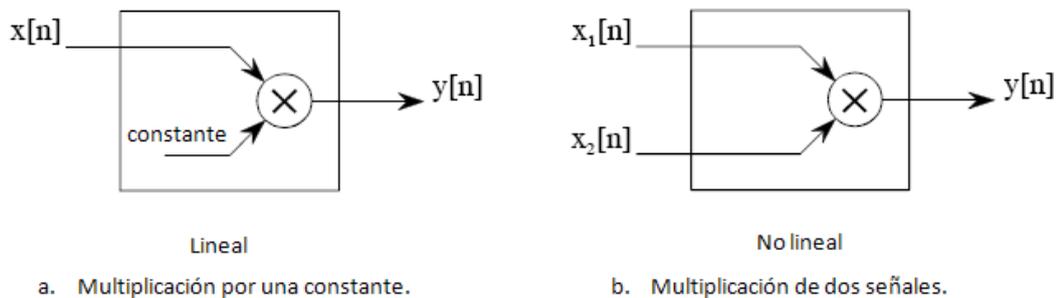
El uso de la multiplicación en sistemas lineales es frecuentemente mal entendido. Esto es porque la multiplicación puede ser lineal o no lineal, dependiendo por lo que la señal es multiplicada. La figura 7 muestra estos dos casos. Un sistema que multiplica la señal de entrada por una constante, es lineal. Este sistema puede ser un amplificador o un atenuador, dependiendo si la constante es más grande o más pequeña que uno, respectivamente. Por otro lado, multiplicando una señal por otra señal no es lineal.

Figura 6. Sistema con múltiples entradas y múltiples salidas.



Otra fuente clara de error es el caso de señales parásitas añadidas a la parte electrónica, como los desfases de DC o el ruido térmico. Si estas señales son vistas como que viniesen internamente del sistema, el proceso no es lineal. Por el contrario, si la señal extraña puede ser vista de procedencia externa al sistema, en una entrada separada a un sistema de múltiples entradas. Esto convierte al proceso en lineal, y entonces, solo una suma de señales es requerida en el sistema.

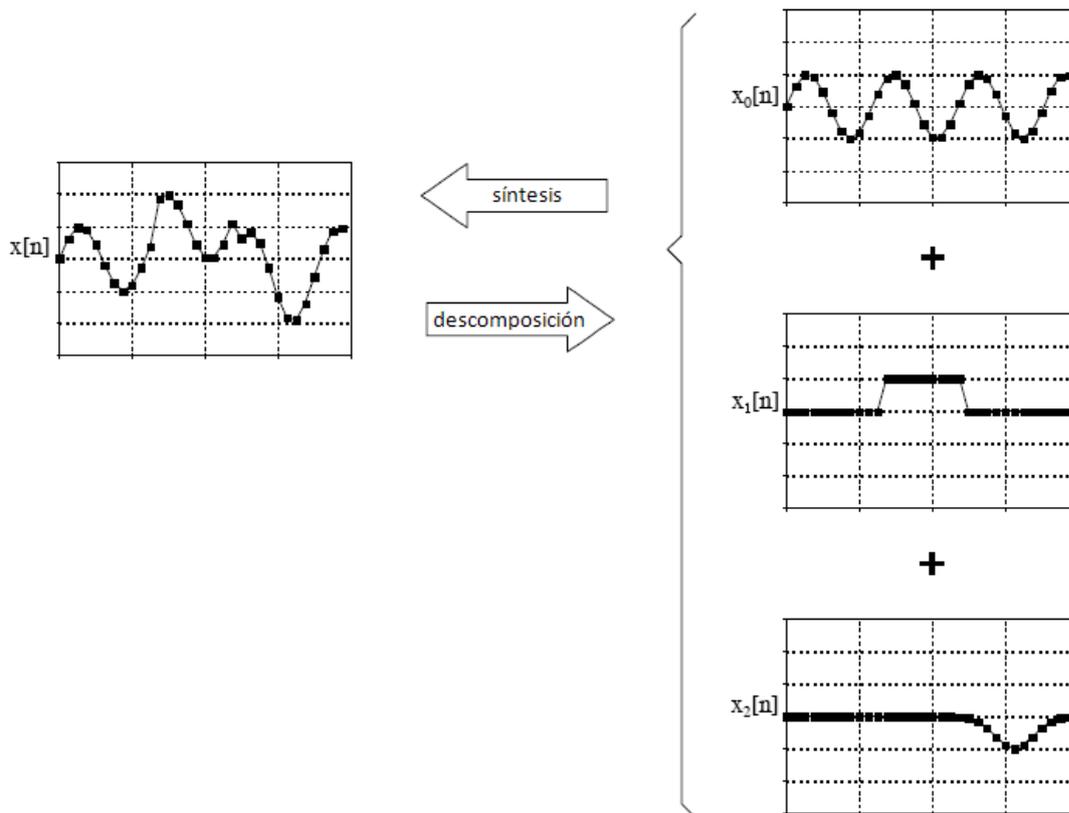
Figura 7. Linealidad de la multiplicación.



1.1.5 Superposición

Cuando se trata con sistemas lineales, la única manera que las señales pueden ser combinadas es por escalado, multiplicación de las señales por constantes, seguido de la adición. Por ejemplo la figura 8 muestra tres señales $x_0[n]$, $x_1[n]$ y $x_2[n]$ son sumadas para formar una cuarta señal $x[n]$. Este proceso de combinar señales por escalado y adición se denomina síntesis.

Figura 8. Síntesis y descomposición.



La descomposición es la operación inversa de la síntesis, donde una sola señal es partida en dos o más componentes aditivas. Esto es más envolvente que la síntesis, porque existen infinitas posibilidades de descomposición para una señal dada.

En el corazón de DSP se encuentra la superposición, la estrategia general para entender como las señales y los sistemas pueden ser analizados. Considerando el caso de una señal de entrada, llamada $x[n]$, pasada a través de un sistema lineal, cuya salida es $y[n]$. Si la señal $x[n]$ puede ser descompuesta en un grupo de señales más simples: $x_0[n]$, $x_1[n]$, $x_2[n]$, etc. Estas componentes de la señal de salida son sintetizadas en la señal de salida $y[n]$.

El punto principal de este método es que la señal de salida obtenida es idéntica a la producida al pasar directamente la señal de entrada a través del sistema. Entonces, en lugar de tratar de entender cómo señales complicadas son cambiadas por el sistema, solo necesitamos saber cómo señales simples son modificadas por el sistema. Las señales de entrada y salida son vistas como la superposición de simples formas de onda, lo cual es la base de casi todas las técnicas de procesamiento de señales,

1.2 Estadística, probabilidad y ruido

La estadística y la probabilidad son usadas en el procesamiento digital de señales, DSP, para caracterizar señales y los procesos para generarlas. Por ejemplo, un uso primario de DSP sería la reducción de interferencia, ruido, y otros componentes indeseables en los datos adquiridos. Estos podrían ser una parte inherente de la señal medida, debida a imperfecciones en el sistema de adquisición de datos, o introducidos como un subproducto inevitable por la operación de DSP.

La estadística y la probabilidad permiten que estas características disruptivas sean medidas y clasificadas, siendo el primer paso para el desarrollo de estrategias que permitan remover estas componentes ofensivas.

1.2.1 Media y desviación estándar

La media, μ , en la jerga estadística es el valor promedio de una señal. Se encuentra sumando todas las muestras y dividiéndolas por el número N de muestras tomadas, expresado de forma matemática es:

$$\mu = \frac{1}{N} \sum_{i=0}^{N-1} x_i \quad (1-1)$$

En la electrónica, la media es comúnmente llamada el valor DC o de corriente directa. Así como, AC (corriente alterna) se refiere a cómo la señal fluctúa entorno a un valor medio. Si la señal es una forma de onda simple y repetitiva, como una sinusoidal o una onda cuadrada, sus variaciones pueden ser descritas por su amplitud pico-a-pico. Desafortunadamente no todas las señales muestran una amplitud pico-a-pico bien definida, es más casi todas las señales adquiridas no presentan esta característica.

Sin embargo, poseen una naturaleza más bien aleatoria por lo que un método más generalizado debe de ser usado en estos casos, llamado desviación estándar.

Como punto de partida, la expresión $|x_i - \mu|$ describe cuánto varía o se desvía la muestra i-ésima de la media. La desviación promedio de una señal se encuentra sumando las desviaciones de las muestras individuales, y luego dividiéndolas por el número de muestras, N. Es de notar que se toma el valor absoluto de cada desviación antes de sumarlo; de lo contrario los términos positivos y negativos llevarían a un valor promedio de cero.

La desviación promedio provee de un único número que representa la distancia típica que las muestras se encuentran de la media. La desviación promedio casi nunca es usada en la estadística, esto es porque no se ajusta de una manera adecuada con la física de las señales y la forma en que estas operan. En muchos casos, el parámetro importante no es la desviación de la media, si no la potencia representada por la desviación de la media. Un ejemplo es cuando señales de ruido se combinan en un circuito eléctrico, el ruido resultante es igual a la potencia combinada de las señales individuales y no a la combinación de su amplitud.

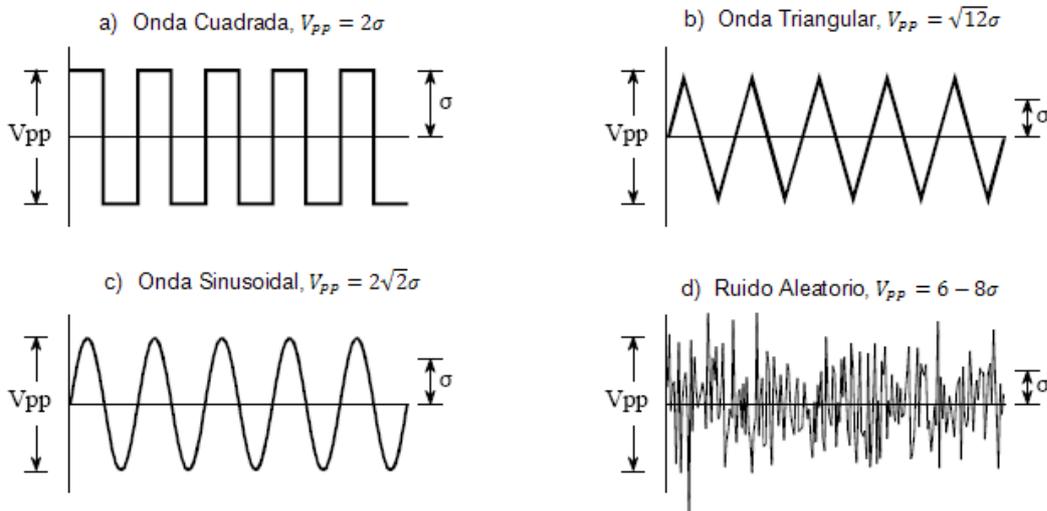
La desviación estándar es similar a la desviación promedio, excepto que el promedio se hace con la potencia en lugar de la amplitud. Esto se alcanza al elevar al cuadrado cada desviación antes de tomar el promedio. Posteriormente la raíz cuadrada es tomada para compensar el proceso de elevar al cuadrado. En forma de ecuación, la desviación es calculada:

$$\sigma^2 = \frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - \mu)^2 \quad (1-2)$$

El término σ^2 , ocurre frecuentemente en estadísticas y se le da el nombre de varianza. La desviación estándar es una medida de cuán lejos fluctúa una señal de la media. La varianza representa la potencia de esta fluctuación.

Entre la terminología usada comúnmente en la electrónica se destaca el concepto de valor RMS (*root-mean-square*) o valor cuadrático medio. Por definición, la desviación estándar solo mide la porción AC de la señal, mientras que el valor RMS mide tanto las componentes AC y DC. Si una señal no posee componente de DC, su valor RMS es idéntico a la desviación estándar. La figura 9 muestra la relación entre la desviación estándar y el valor pico-a-pico de algunas formas de onda comunes.

Figura 9. Relación entre la desviación estándar y el valor pico a pico.

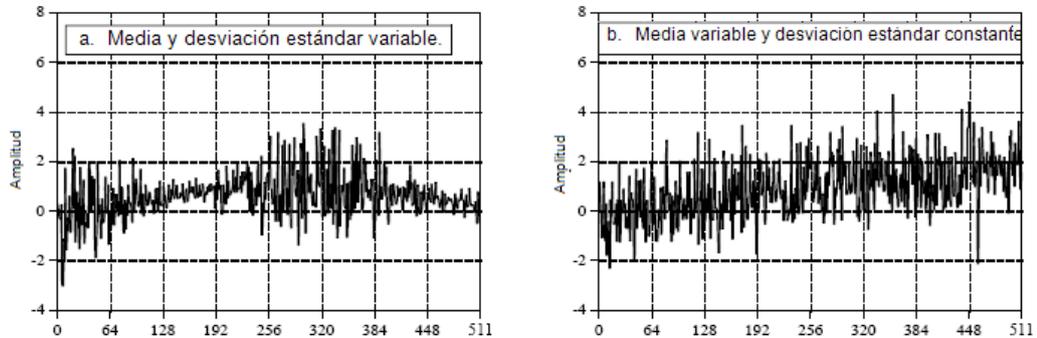


En algunos casos, la media describe lo que se está midiendo, mientras que la desviación estándar representa el ruido u otras interferencias. Para estos casos, la desviación estándar no es importante en sí, pero si la comparación de ésta con la media. Lo que conlleva al concepto de relación señal a ruido o SNR, *signal-to-noise ratio*, el cual es el cociente entre la media y la desviación estándar. Mejor calidad de los datos implica valores altos de SNR, por lo que estos son deseados.

1.2.2 Estadística y procesos subyacentes

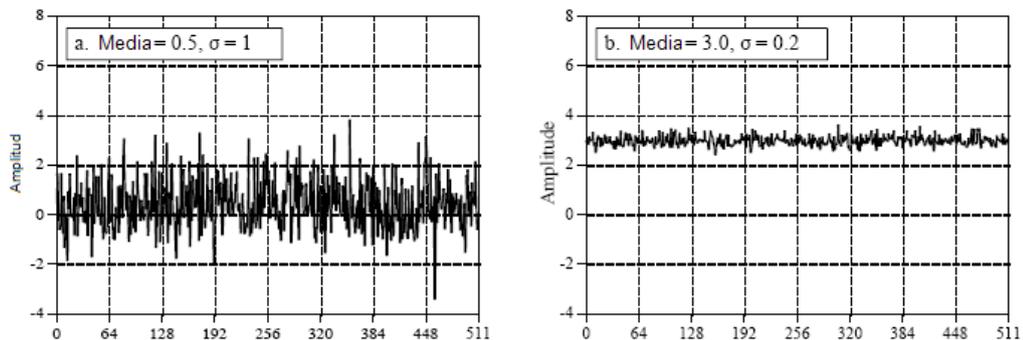
La estadística es la ciencia que se encarga de interpretar los datos numéricos, como los adquiridos por señales. En comparación, la probabilidad se utiliza en DSP para entender los procesos que generan estas señales. La clave a muchas técnicas de DSP se sitúa en la distinción entre la señal adquirida y los procesos subyacentes. Las probabilidades de un proceso subyacente son constantes, pero las estadísticas de una señal adquirida cambian cada vez que un experimento es repetido. Esta irregularidad adquirida se denomina usualmente como: variación estadística, fluctuación estadística, y ruido estadístico.

Figura 10. Procesos no estacionarios.



La figura 10 muestra una señal que a primera vista pareciera ser afectada por ruido estadístico, pero se puede apreciar que de ser este ruido las variaciones no serían tan abruptas, sino más bien tienen una mayor relación al proceso subyacente. Este tipo de procesos que cambian sus características de esta manera reciben el nombre de procesos no-estacionarios. Un proceso estacionario tiene la forma de la figura 11, donde todas las variaciones tienen origen al ruido estadístico.

Figura 11. Proceso estacionario



De nuevo al observar la figura 10 se puede ver que la media, que cambia de forma lenta, interfiere con el cálculo de la desviación estándar de la señal, sobre un intervalo pequeño se ve que es uno. Sin embargo, la desviación estándar de toda la señal es 1.16. Este error se puede eliminar si se divide la señal en secciones pequeñas. Si fuese necesario, la desviación estándar de cada sección podría ser promediada para tener único valor.

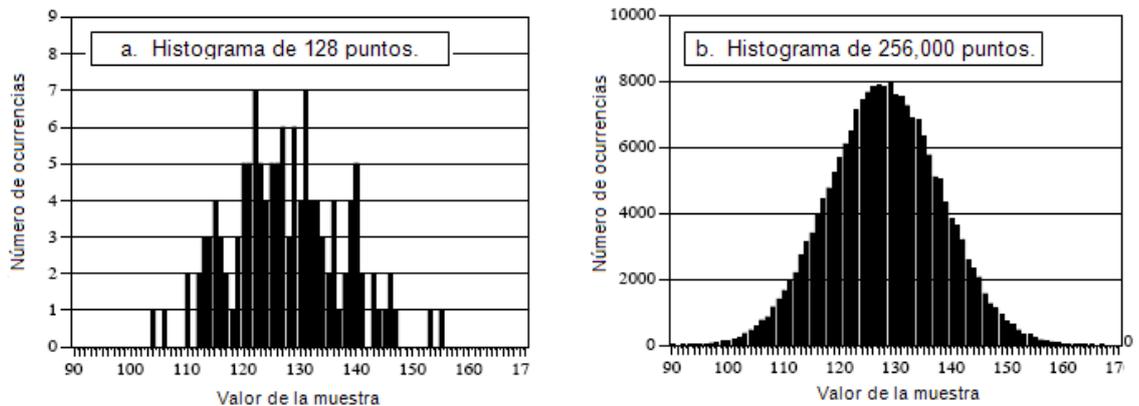
1.2.3 Histograma, función masa de probabilidad y función densidad de probabilidad

Supongamos que se le adapta un convertidor analógico a digital de 8 bits a una computadora, y se adquieren 256,000 muestras. El valor de cada muestra estará entre uno y 256 posibilidades, desde 0 a 255. Con un histograma se puede observar el número de muestras que se encuentran en una señal que poseen estas distintas posibilidades. La figura 12.a muestra el histograma de una señal con las primeras 128 muestras tabuladas. Conforme el número de muestras que se incluye en un histograma incrementa, la suavidad que se presenta el histograma incrementa, la figura 12.b muestra un histograma con todas las muestras tabuladas.

Se presentará el histograma como H_i , dónde i es un índice que va desde 0 hasta $M-1$, y M es número de posibles valores que cada muestra puede tener. Por la forma en que se encuentra definida, la suma de todos los valores en el histograma debe ser igual a la suma de todos los puntos en la señal:

$$N = \sum_{i=0}^{M-1} H_i \quad (1-3)$$

Figura 12. Histograma de señal luego de conversión analógica a digital.



Un histograma puede ser usado para calcular de forma eficiente la media y la desviación estándar de un grupo grande de datos. Un histograma agrupa muestras de un mismo valor, lo cual permite que las estadísticas sean calculadas con un menor grupo, en lugar de una gran cantidad de muestras individuales. Usando este método, la media y la desviación estándar pueden ser calculadas a partir del histograma por las ecuaciones:

$$\mu = \frac{1}{N} \sum_{i=0}^{M-1} iH_i \quad (1-4)$$

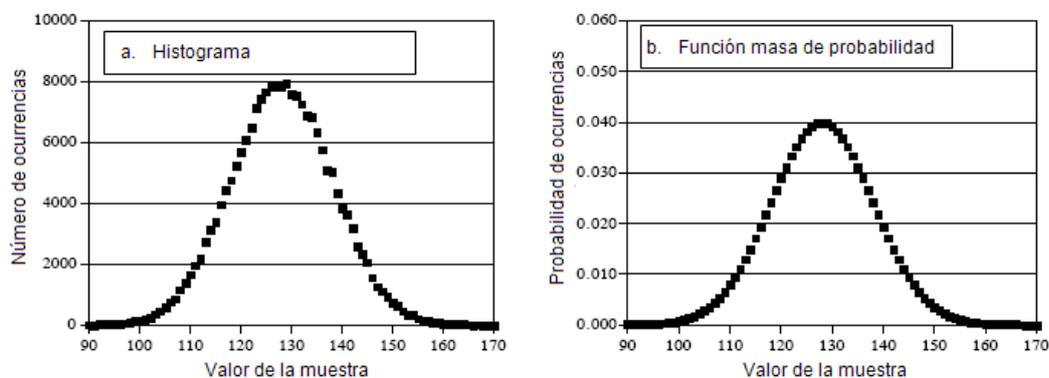
$$\sigma^2 = \frac{1}{N-1} \sum_{i=0}^{N-1} (i - \mu)^2 H_i \quad (1-5)$$

La noción de que la señal adquirida es una versión con ruido del proceso subyacente es bastante importante; tan importante que algunos conceptos se les dan distintos nombres.

Un histograma es de lo que se encuentra formada la señal adquirida. La correspondiente curva del proceso subyacente se denomina función masa de probabilidad, PMF (*probability mass function*). Un histograma siempre es calculado usando un número finito de muestras, mientras que la PMF puede ser estimada o inferida del histograma o deducida por alguna técnica matemática.

La figura 13 muestra un ejemplo de una PMF, y un posible histograma que se encuentre asociada a ella. El concepto para entender estos conceptos se encuentra en las unidades del eje vertical. El eje vertical en un histograma es el número de veces que un valor particular ocurre en la señal. El eje vertical en una PMF contiene información similar, excepto que es expresado en una base fraccional. Cada valor del histograma es dividido por el total número de muestras para aproximar la PMF. Esto significa que cada valor en la PMF debe encontrarse entre cero y uno, y que la suma de todos los valores en la PMF debe ser igual a uno. La PMF es importante porque describe la probabilidad de que cierto valor sea generado.

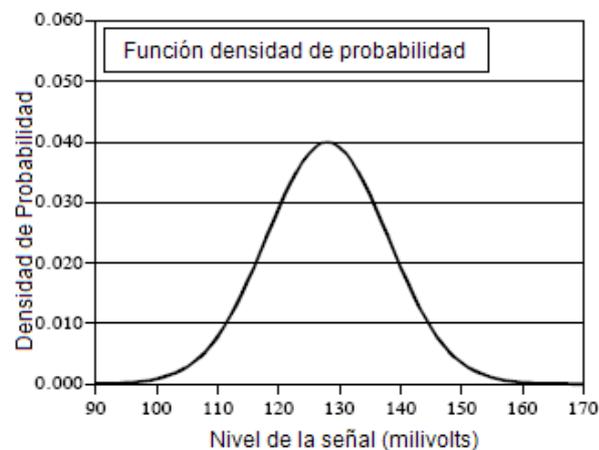
Figura 13. Histograma y función masa de probabilidad.



Un histograma y la PMF solo pueden ser usadas para datos discretos, como los datos digitalizados de una computadora. Un concepto similar aplica a las señales digitales. La función de densidad de probabilidad, *PDF* (*probability density function*), o también llamada función de distribución de probabilidad, es a las señales continuas lo que la función masa de probabilidad a las señales discretas.

La figura 14 muestra la PDF de una señal analógica, la línea continua muestra que la señal puede tener un rango continuo de valores, como el voltaje en un circuito eléctrico.

Figura 14. Función densidad de probabilidad.



El área debajo de la función de probabilidad sobre cierto intervalo proporciona la probabilidad de ocurrencia de dicha señal en este intervalo. Del hecho de que la señal siempre debe significar algo, el área total debajo de la curva PDF, la integral de $-\infty$ a $+\infty$, debe de ser igual a uno. Esto es análogo a que la suma de todos los valores PMF sea igual a uno, y que la suma de todos los valores del histograma sean igual a N.

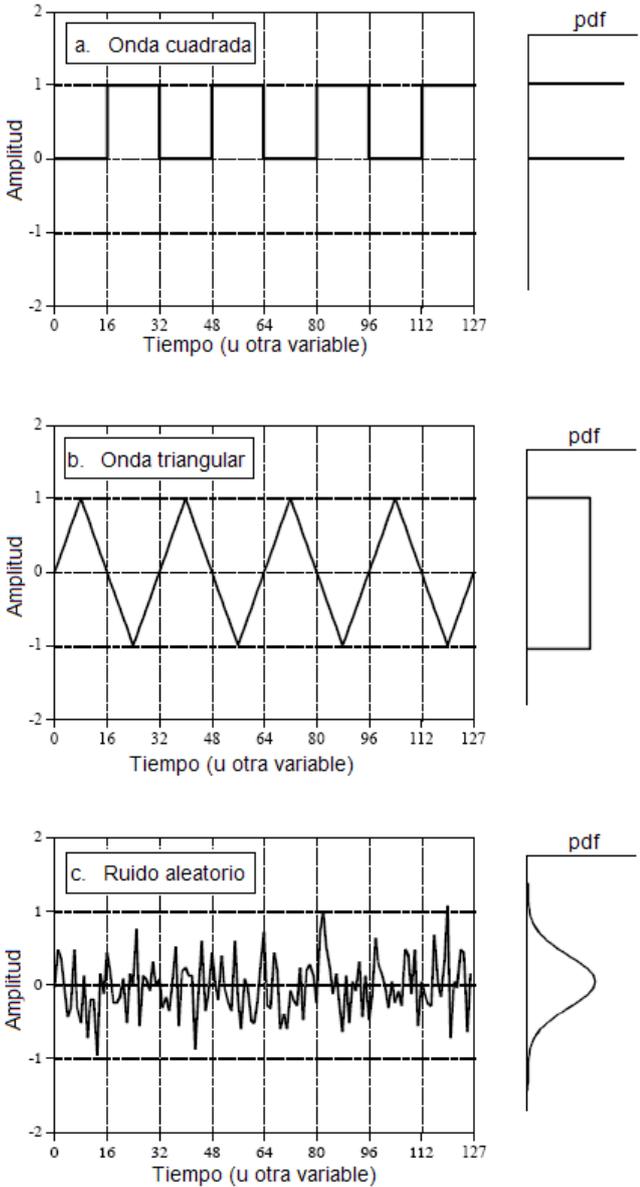
La figura 15 muestra tres formas de onda continuas y sus PDFs. Si estas fueran señales discretas, significativamente al cambiar el nombre del eje horizontal por “número de muestra”, funciones de masa de probabilidad serían usadas.

Un problema surge en el cálculo de un histograma cuando el número de niveles que cada muestra puede tener es mucho más grande que el número de muestras de la señal. Esto siempre es verdadero para señales representadas en notación de punto flotante, donde cada muestra es almacenada como un valor fraccionario. En el caso de la representación entera el valor de una muestra requerirá estar entre dos valores enteros, mientras que el punto flotante permite millones de posibles valores fraccionales entre estos dos valores enteros.

Esta aproximación anteriormente descrita para calcular el histograma envuelve el conteo del número de muestras que posee cada posible nivel de cuantización. Esto no es posible para los datos de punto flotante ya que existen miles de millones de posibles niveles que deberían de ser tomados en cuenta. Lo que es peor, casi todos estos posibles niveles no tendrán ninguna muestra correspondiente a ellos.

Existe una solución a este problema por medio de la técnica conocida como *binning*. Esto se realiza por medio de la selección arbitraria de la longitud del histograma a un valor conveniente, por ejemplo 1000 puntos; usualmente llamados *bins*. El valor de cada *bin* representa el número total de muestras en la señal que se contienen dentro de cierto rango. Muchos *bins* hacen difícil de estimar la amplitud del proceso subyacente. Esto porque solamente algunas muestras caen en cada bin, haciendo que el ruido estadístico sea muy alto. En el otro caso, pocos *bins* hacen que sea muy difícil de estimar el proceso subyacente PMF en la dirección horizontal. En resumen, el número de *bins* controla un compromiso entre la resolución a lo largo del eje y, y la resolución a lo largo del eje x.

Figura 15. Función densidad de probabilidad de algunas funciones comunes.



1.2.4 La distribución normal

Señales formadas por procesos aleatorios usualmente tienen una función de distribución de probabilidad, PDF, con la forma de una campana. Esta es llamada distribución normal, distribución de Gauss, o Gaussiana. Esto tiene su fundamento en el teorema del límite central, uno de los conceptos más importantes en la probabilidad. En su forma más simple, el teorema del límite central enuncia que la suma de números aleatorios se distribuye normalmente cuando más y más números son sumados juntos. El teorema del límite central no implica que los números individuales sean de una distribución en particular, e incluso de la misma distribución. Esta es la razón del porque señales normalmente distribuidas se encuentran distribuidas en la naturaleza. En cualquier caso cuando distintas fuerzas aleatorias están interactuando, la PDF resultante se convierte Gaussiana.

La forma básica de la curva es generada por un exponente cuadrático negativo:

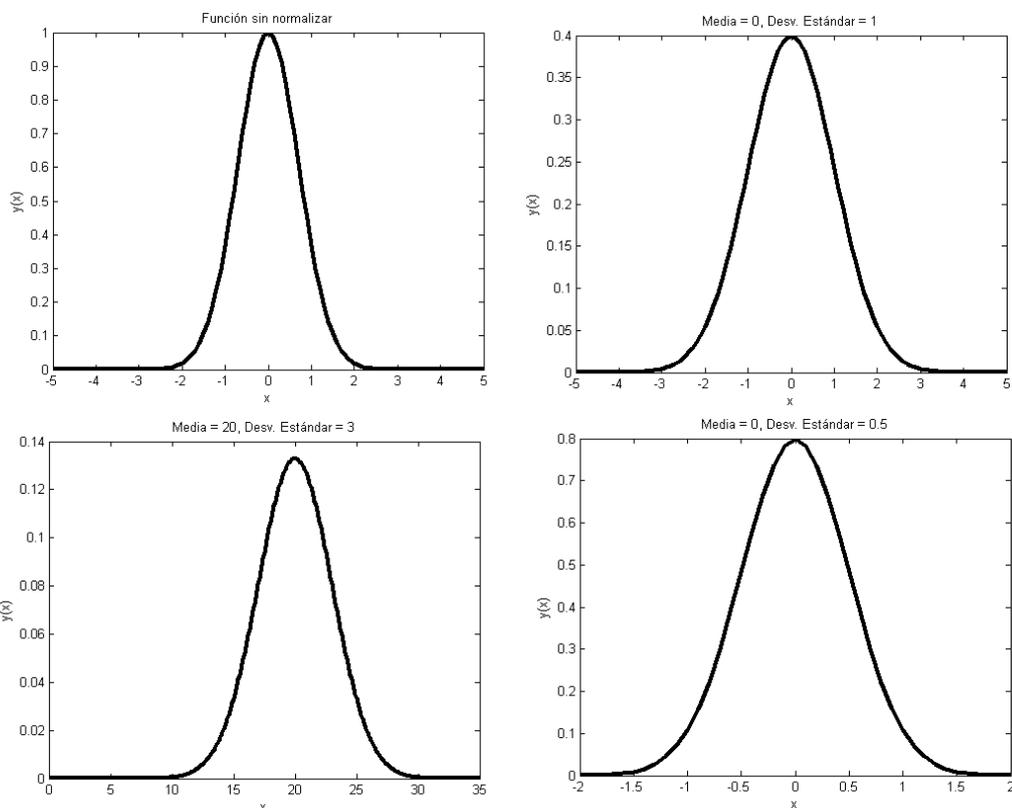
$$y(x) = e^{-x^2} \quad (1-6)$$

Esta curva, de alguna manera cruda, puede ser convertida en una función Gaussiana al añadir una media ajustable, μ , y una desviación estándar, σ . Además, esta ecuación debe estar normalizada para que el área total bajo la curva sea igual a uno, un requerimiento de todas las distribuciones de probabilidad. Esto resulta en la forma general de la distribución normal, una de las más importantes relaciones en la estadística y la probabilidad:

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad (1-7)$$

La figura 16 muestra varios ejemplos de curvas Gaussianas con distintas medias y desviaciones estándar. La media centra la curva sobre un valor en particular, mientras que la desviación estándar controla el ancho de la forma de campana. Una característica de la función Gaussiana es que la cola cae a cero muy rápidamente, mucho más rápido que otras funciones comunes. Es el porqué de las señales normalmente distribuidas parecen tener un valor pico a pico. En principio, señales de este tipo pueden experimentar excursiones de amplitud ilimitada. En práctica, la caída abrupta de la PDF Gaussiana dicta que estos extremos casi nunca ocurran. Esto resulta en una forma de onda que tiene una apariencia relativamente limitada, con una amplitud pico a pico aparente de $6 - 8\sigma$.

Figura 16. Distribución normal o Gaussiana



La integral de la PDF es usada para encontrar la probabilidad de que una señal se encuentre en determinado rango de valores. Esto hace la integral de la PDF lo suficientemente importante, tanto que tiene su propio nombre, conocida como la función de distribución acumulativa, CDF. Esta integral no puede ser calculada usando métodos convencionales, por lo que se utiliza integración numérica para resolverla. Lo cual involucra el muestreo de la curva Gaussiana continua de forma muy fina. Las muestras en esta señal discreta son sumadas para simular una integración. El cálculo de estos valores de la integral usualmente se encuentra en muchos libros de estadística.

2. TRANSFORMACIONES DE SEÑALES

Cualquier fenómeno que pueda ser representado de manera cuantitativa mediante una función continua o discreta se puede considerar una señal. Una señal continua es una señal “suave” que está definida para todos los puntos de un intervalo determinado del conjunto de los números reales. Una señal discreta es una señal discontinua que está definida para todos los puntos de un intervalo determinado del conjunto de los números enteros. Una transformación es un operador aplicado a una función que bajo dicha transformación ciertas operaciones son simplificadas. Existen transformaciones para funciones continuas y discretas, las cuales pueden ser transformaciones integrales y transformaciones discretas respectivamente.

2.1 Transformaciones continuas

En esta sección trataremos con señales que serán representadas por una función $f(t)$ dependiendo del tiempo t . Se asumirá que t es un parámetro continuamente variante, por lo que $f(t)$ se le conoce como una señal continua en el tiempo. Se pretende transformar $f(t)$ en una representación que incorpore la información deseada de la señal, de la manera más compacta posible. La transformada de Fourier presenta información acerca de la contribución de ciertas frecuencias a la señal, la transformación wavelet indica si los detalles de un determinado tamaño están presentes en una señal y cuantifica su respectiva contribución.

Ambas transformaciones son llamadas locales si no solo miden globalmente frecuencias y tamaños de detalles, respectivamente, sino si también indican donde se encuentran localizadas en la señal.

Todas las transformadas de la señal $f(t)$ descritas comparten un principio computacional común: la señal es multiplicada con cierta función de análisis e integrada en el dominio del tiempo. Este tipo de transformación en una notación más adecuada se puede ver como:

$$f(t) \xrightarrow{\text{transformación}} \int_{-\infty}^{+\infty} f(u)\overline{g(u)}du \quad (2-1)$$

La función de análisis $g(u)$ caracteriza la transformada escogida. En general puede tratarse de una función compleja, la línea sobre la función indica el complejo conjugado. La función $g(u)$ de cierta manera depende de los parámetros, por ejemplo frecuencias o tamaños de detalle, que deben medirse. La transformada dependerá de estos mismos parámetros por lo que la transformada nuevamente será una función. Estas funciones son nombradas como la transformada o la señal transformada.

Otro aspecto común de las transformadas discutidas es la invertibilidad que poseen, de la señal transformada la señal original $f(t)$ puede ser reconstruida.

2.1.1 La transformada de Fourier

Una función periódica puede ser expresada como la suma de componentes espectrales. Estas componentes poseen amplitudes finitas y se encuentran separadas a intervalos de frecuencia finitos. La potencia normalizada de la forma de onda es finita, al igual que la energía normalizada de la señal en un período de tiempo finito.

Si el período de una señal se aumenta a un valor infinito, señal no periódica, el espaciado discreto de componentes espectrales se vuelve infinitesimal convirtiéndose en una función continua. La energía normalizada de la forma de onda no periódica se mantiene finita, pero como la señal ya no es repetida, su potencia normalizada se vuelve infinitesimal.

El parámetro relevante para la transformada de Fourier es la frecuencia circular ω , la función de análisis es $g_\omega(u) = e^{j\omega u}$. Entonces la señal transformada es una función dependiente de ω y es denotada por $F(\omega)$.

$$F(\omega) = \int_{-\infty}^{\infty} f(u) \overline{g_\omega(u)} du$$

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

Aquí, $\omega = 2\pi f$, y f es la frecuencia en Hz.

La transformada de Fourier $F(\omega)$ de la señal $f(t)$ posee las siguientes propiedades:

1. $F \in L_\infty(\mathbb{R})$ con $\|F\|_\infty \leq \|f\|_1$.
2. F es continua.
3. Si la derivada $f'(t)$ existe y es absolutamente integrable, entonces

$$\int_{-\infty}^{\infty} f'(t) e^{-j\omega t} dt = j\omega F(\omega). \quad (2-3)$$

4. Para $\omega \rightarrow \infty$ y $\omega \rightarrow -\infty$ se tiene que $F(\omega) \rightarrow 0$.

Si $F(\omega)$ es absolutamente integrable, $f(t)$ puede ser reconstruida a partir de $F(\omega)$ a través de la transformada inversa de Fourier

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega. \quad (2-4)$$

para todos los t donde $f(t)$ es continua.

De ahora en adelante se utilizará la notación $f(t) \leftrightarrow F(\omega)$ para indicar la pareja de transformación de Fourier.

La tabla I presenta las propiedades más importantes de la transformada de Fourier.

Tabla I. Propiedades importantes de la transformada de Fourier.

Propiedad	Relación
Linealidad	$\alpha x(t) + \beta y(t) \leftrightarrow \alpha X(\omega) + \beta Y(\omega)$
Simetría	$X(t) \leftrightarrow 2\pi x(-\omega)$
Escalado	$x(\alpha t) \leftrightarrow \frac{1}{ \alpha } X\left(\frac{\omega}{\alpha}\right)$
Desplazamiento	$x(t - t_0) \leftrightarrow e^{-j\omega t_0} X(\omega)$ $e^{j\omega t_0} x(t) \leftrightarrow X(\omega - \omega_0)$
Modulación	$\cos(\omega_0 t)x(t) \leftrightarrow \frac{1}{2} X(\omega - \omega_0) + \frac{1}{2} X(\omega + \omega_0)$
Conjugación	$x^*(t) \leftrightarrow X^*(-\omega)$
Derivadas	$\frac{d^n}{dt^n} x(t) \leftrightarrow (j\omega)^n X(\omega)$ $(-jt)^n x(t) \leftrightarrow \frac{d^n}{d\omega^n} X(\omega)$
Convolución	$x(t) * y(t) \leftrightarrow X(\omega)Y(\omega)$ $x(t)y(t) \leftrightarrow \frac{1}{2\pi} X(\omega) * Y(\omega)$

2.1.1.1 Convolución

Supongamos ahora que $v_1(t)$ tiene la transformada de Fourier $V_1(\omega)$ y $v_2(t)$ tiene la transformada $V_2(\omega)$. Al producto $V_1(\omega)V_2(\omega)$ y a la transformación inversa $v(t) \leftrightarrow V_1(\omega)V_2(\omega)$ se le da mucha importancia en el análisis espectral, y se puede presentar por medio del teorema de convolución, que se puede escribir como:

$$v(t) = \int_{-\infty}^{\infty} v_1(t)v_2(t - \tau)d\tau \quad (2-5)$$

o de forma equivalente

$$v(t) = \int_{-\infty}^{\infty} v_2(t)v_1(t - \tau)d\tau. \quad (2-6)$$

Las integrales en las ecuaciones (2-5) y (2-6) son llamadas integrales de convolución, y al proceso de evaluar $v(t)$ a través de estas integrales es conocido como “tomar la convolución de las funciones $v_1(t)$ y $v_2(t)$ ”.

Un caso especial de aplicación del teorema de convolución, y uno de gran utilidad, es el siguiente bajo estas consideraciones. Suponiendo que una forma de onda $x(t)$ cuya transformada es $X(\omega)$ es aplicada a un sistema lineal con una función de transferencia $H(\omega)$. La transformada a la salida es $X(\omega) H(\omega)$. La salida $y(t) = x_o(t)$, en base al teorema de convolución, es:

$$y(t) = \int_{-\infty}^{\infty} x(t)h(t - \tau)d\tau. \quad (2-7)$$

2.1.1.2 Teorema de Parseval

Una de las mayores contribuciones de la transformada de Fourier es la aplicación que tiene en el teorema de *Parseval*. La relación de *Parseval*, establece que el producto interno de dos señales puede ser calculado en el dominio del tiempo así como también en el dominio de la frecuencia. Para dos señales $x(t)$ y $y(t)$ y sus transformadas de Fourier $X(\omega)$ y $Y(\omega)$, respectivamente, se tiene:

$$\int_{-\infty}^{\infty} x(t)y^*(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)Y^*(\omega)d\omega. \quad (2-8)$$

De la relación de *Parseval* se desprende el teorema que hace honor al mismo, el cual se establece que la energía de una señal puede ser calculada tanto en el dominio del tiempo como en el dominio de la frecuencia.

Esto se realiza al hacer $x(t) = y(t)$ por lo que se tiene:

$$\begin{aligned}
 E_x &= \int_{-\infty}^{\infty} |x(t)|^2 dt \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega.
 \end{aligned}
 \tag{2-9}$$

La densidad de energía puede ser encontrada como:

$$G_E(\omega) \equiv \frac{dE}{d\omega} = \frac{1}{2\pi} |X(\omega)|^2.
 \tag{2-10}$$

La validez de estos resultados depende en la ortogonalidad de las componentes espectrales. En este caso el intervalo de ortogonalidad se extiende desde $-\infty$ hasta $+\infty$, que en el caso de una señal periódica se extendería en el período de la misma.

2.1.2 La transformada de Fourier de corta duración

La transformada de Fourier de corta duración, STFT, (*Short Time Fourier Transform*) es una transformada relacionada con la transformada de Fourier; la cual es usada para determinar el contenido de frecuencia sinusoidal y la fase de una sección local de una señal que cambia respecto al tiempo. La STFT busca la aparición de la frecuencia circular ω a cierto tiempo t . La correspondiente función de análisis es: $g_{(\omega,t)}(u) = e^{ju\omega} w(u - t)$. Donde $w(u)$ es una función de ventana, que se encuentra generalmente centrada en el origen. En la expresión $w(u - t)$ esta ventana se encuentra desplazada el tiempo t deseado.

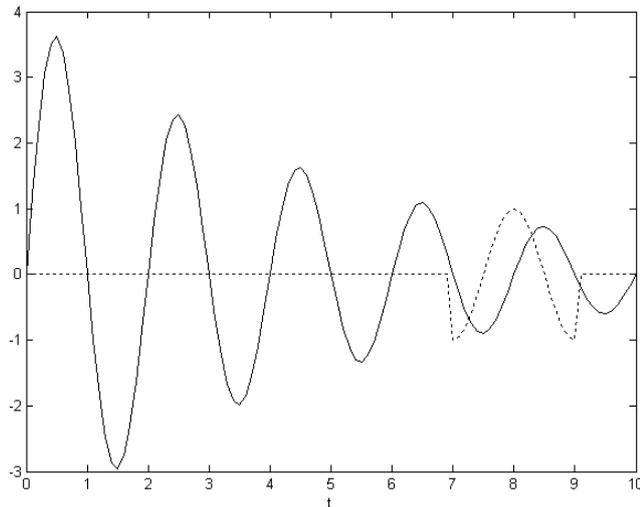
Ahora la señal transformada depende tanto de ω como de t . Como ahora depende de la forma de función de ventana, es denotada con $F_\omega(\omega, t)$.

$$F_\omega(\omega, t) = \int_{-\infty}^{\infty} f(u) \overline{g_{(\omega, t)}(u)} du$$

$$F_\omega(\omega, t) = \int_{-\infty}^{\infty} f(t) w(u - t) e^{-ju\omega} du \quad (2-11)$$

La figura 17 muestra una ventana cuadrada de ancho 2, centrada simétricamente en 0, $\omega = \pi$ y $t = 8$. La línea discontinua muestra la parte real de la función de análisis $g_{(\pi, 8)}(u)$; con lo cual se encuentra localizada en $t = 8$, ya que $w(u - 8)$ denota la ventana cuadrada, desplazada 8 unidades a la derecha.

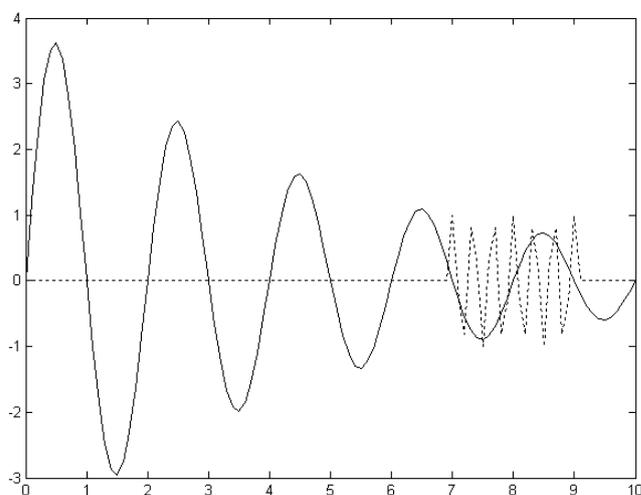
Figura 17. Ventana cuadrada para $\omega = \pi$ y $t = 8$.



En general, la función de análisis será localizada al respectivo “tiempo de análisis” t . Entonces la transformación provee no solo información global de la aparición de ciertas frecuencias circulares, sino que además el tiempo de su aparición.

Existe una desventaja en este procedimiento, y es que; si uno se encuentra interesado en pequeños detalles de la señal alrededor de cierto tiempo t , la correspondiente frecuencia de análisis debe incrementarse. Como se ve en la figura 18 para $\omega = 6\pi$.

Figura 18. Ventana cuadrada para $\omega = 6\pi$ y $t = 8$.



Se observa que el ancho de la ventana es constante y no adaptiva. Si uno se interesase en un detalle muy pequeño de la señal (altas frecuencias) en un intervalo muy pequeño, en este caso de $t = 8$, señales que no serían de nuestro interés también serían analizadas. Por lo que acercamientos en mayor detalle no son posibles.

2.1.3 La transformada wavelet

La transformada wavelet tiene esta propiedad de acercamiento. En comparación con la transformada de Fourier, la transformada wavelet no busca frecuencias circulares sino más bien tamaños de detalles a a cierto tiempo t . En lugar de tamaños de detalles se hablará de “factores de escala”, ambas notaciones serán usadas de forma equivalente.

Como se mencionó anteriormente, frecuencias altas corresponden a pequeños detalles y viceversa, por lo que cuando se compara las transformadas wavelet y Fourier se tiene que tomar en cuenta que las frecuencias y el tamaño de los detalles son inversamente proporcionales uno del otro.

Por lo que debe de existir una constante β que:

$$\alpha = \frac{\beta}{\omega} \quad (2-12)$$

Ahora consideremos una función de análisis (real o compleja) g , oscilando alrededor del eje u , que decrece de forma rápida cuando $u \rightarrow \pm\infty$ y que además $\int_{-\infty}^{\infty} g(u) du = 0$. Este tipo de función es lo que se denomina *wavelet*. En la relación de la ecuación (2-12) la constante β depende de g .

Partiendo de g se puede considerar la siguiente familia de funciones: $g_{(a,t)}(u) = \frac{1}{\sqrt{a}} g\left(\frac{u-t}{a}\right)$. Los miembros de esta familia son generados a partir de g al desplazar la función t unidades seguido de un encogimiento ($a < 1$) o dilatación ($a > 1$) en el ancho de la función. Por lo que la transformada *wavelet* ahora puede ser escrita como:

$$L_g f(a, t) = \int_{-\infty}^{+\infty} f(u) \overline{g_{(a,t)}(u)} du \quad (2-13)$$

Para la transformación *wavelet* tipo *Haar*

$$g(u) = \begin{cases} 1 & 0 < u < \frac{1}{2} \\ -1 & \frac{1}{2} < u < 1 \\ 0 & \text{otro lugar} \end{cases} \quad (2-14)$$

Figura 19. Transformada *wavelet*: señal y función de análisis para $a=1/2$.

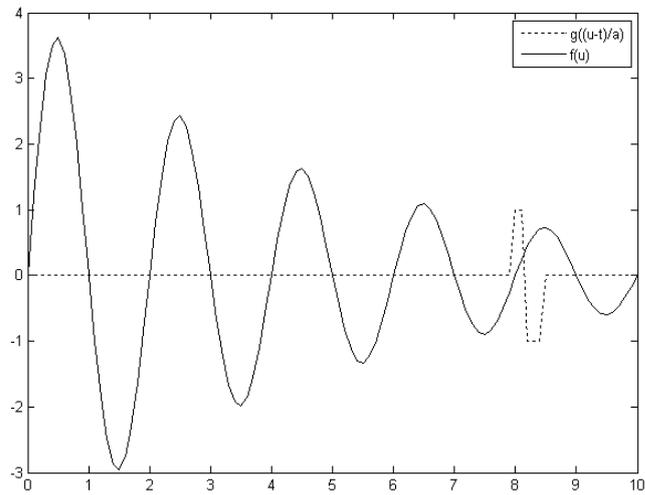
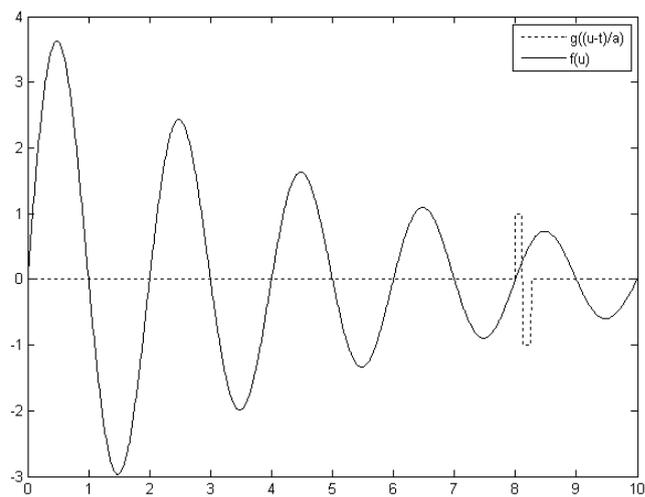


Figura 20. Transformada *wavelet*: señal y función de análisis para $a=1/4$.



Para la figura 19 se utilizó un valor de $a = \frac{1}{2}$ y $t = 8$; mientras que para la figura 20 se utilizó un valor de $a = \frac{1}{4}$ y $t = 8$. Si se comparan con las figuras 17 y 18 se nota que la transformada *wavelet* proporciona el aumento que carecía la STFT. Cuando se busca un detalle cada vez más y más pequeño (frecuencias cada vez más altas) con la transformada *wavelet*, la correspondiente función de análisis oscila más rápido y es comprimida. Hasta ahora se ha presentado de forma básica la transformación *wavelet* esta se profundizará más en el capítulo 5.

2.2 Transformaciones discretas

En esta sección se discutirán las transformadas discretas más importantes. Se comenzará con la transformada Z, la cual es una herramienta fundamental para describir la relación entrada/salida de los sistemas invariantes con el tiempo, LTI. Se discuten además variaciones de la expansión de series de Fourier, llamadas la transformada discreta en el tiempo de Fourier, DFT (*Discrete Fourier Transform*), y la transformada rápida de Fourier, FFT (*Fast Fourier Transform*). La presentación de la transformada discreta *wavelet* se deja para el capítulo 5, donde se presentará un estudio más completo acerca de la misma.

2.2.1 La transformada Z

La transformada Z convierte una señal de tiempo discreto, la cual es una secuencia de números reales o complejos, en una representación compleja en el dominio de la frecuencia. La transformada Z de una señal discreta en el tiempo $x[n]$ es definida como:

$$X(z) = \sum_{-\infty}^{\infty} x[n]z^{-n} \quad (2-15)$$

Nótese que el índice de tiempo n es discreto, mientras que z es un parámetro continuo. Más aún, z es complejo, incluso si $x[n]$ es real.

En general, la convergencia de la suma de (2-15) depende de la secuencia $x[n]$ y el valor de z . Para la mayoría de secuencias solo se tiene una convergencia en ciertas regiones del plano z , llamado región de convergencia, ROC (*Region of Convergence*). La ROC puede ser determinada al encontrar los valores de r para los cuales:

$$\sum_{-\infty}^{\infty} |x[n]r^{-n}| < \infty \quad (2-16)$$

Por lo que $|X(z)|$ es finita si $x[n]r^{-n}$ es absolutamente sumable.

La transformada Z, como se ha definido en la ecuación (2-15), se denomina comúnmente transformada Z bilateral, en contraste con la transformada Z unilateral, que se define de la siguiente manera:

$$X(z) = \sum_0^{\infty} x[n]z^{-n} \quad (2-17)$$

Puede verse claramente que la transformada Z unilateral y bilateral coinciden si $x[n] = 0$ para $n < 0$.

La transformada Z inversa está dada por

$$x[n] = \frac{1}{2\pi j} \oint_C X(z)z^{n-1}dz \quad (2-18)$$

La integración tiene que ser llevada en contra de la agujas del reloj en una trayectoria cerrada C en el plano complejo, la cual encierra el origen y se encuentra en la región de convergencia de $X(z)$.

Algunas veces la transformada Z inversa se puede encontrar por inspección, si la expresión de la transformada Z se reconoce o se puede encontrar en alguna tabla, pero puede ser posible obtener una expresión alternativa de $X(z)$ como suma de términos más simples, cada uno de los cuales puede ser encontrado en alguna tabla. Este es el caso de las funciones racionales, ya que es posible realizar una descomposición en fracciones parciales simples e identificar fácilmente las secuencias correspondientes a cada término.

Suponiendo que $X(z)$ se expresa como un cociente de polinomios en z^{-1} , como es

$$X(z) = \frac{\sum_{k=0}^M b_k z^{-k}}{\sum_{k=0}^N a_k z^{-k}} \quad (2-19)$$

La cual es una transformación frecuentemente usada en el estudio de sistemas lineales e invariantes con el tiempo. Una expresión equivalente es

$$X(z) = \frac{z^N \sum_{k=0}^M b_k z^{M-k}}{z^M \sum_{k=0}^N a_k z^{N-k}} \quad (2-20)$$

La ecuación (2-20) indica que para estas funciones, habrá M ceros y N polos en posiciones del plano complejo distintas de cero. Además, habrá $M - N$ polos en $z = 0$ si $M > N$, o $N - M$ ceros en $z = 0$ si $N > M$. Es decir, las transformadas Z de la forma de la ecuación (2-19) tienen el mismo número de polos y ceros en el plano z finito, y no hay polos ni ceros en $z = \infty$. Para realizar la descomposición en fracciones simples de $X(z)$, es más conveniente notar que $X(z)$ se puede expresar de la forma

$$X(z) = \frac{b_0 \prod_{k=1}^M (1 - c_k z^{-1})}{a_0 \prod_{k=1}^N (1 - d_k z^{-1})} \quad (2-21)$$

Siendo c_k los ceros distintos de cero de $X(z)$ y d_k los polos distintos de cero de $X(z)$. Si $M < N$, y los polos son todos de primer orden, entonces $X(z)$ se puede expresar así:

$$X(z) = \sum_{k=1}^N \frac{A_k}{1 - d_k z^{-1}} \quad (2-22)$$

En la tabla II se presentan algunas propiedades de la transformada Z, la demostración de estas propiedades queda fuera del objetivo de este texto por lo que aquí no se presentan.

Tabla II. Propiedades importantes de la transformada Z.

Propiedad	Relación
Linealidad	$\alpha x[n] + \beta y[n] \leftrightarrow \alpha X(z) + \beta Y(z)$
Convolución	$x[n] * y[n] \leftrightarrow X(z)Y(z)$
Desplazamiento	$x[n - n_0] \leftrightarrow z^{-n_0} X(z)$
Escalado/Modulación	$a^n x[n] \leftrightarrow X\left(\frac{z}{a}\right)$ $e^{j\omega n} x[n] \leftrightarrow X(e^{-j\omega} z)$
Inversión del Tiempo	$x[-n] \leftrightarrow X\left(\frac{1}{z}\right)$
Derivadas	$nx[n] \leftrightarrow -z \frac{dX(z)}{dz}$
Conjugación	$x^*[n] \leftrightarrow X^*(Z^*)$
Paraconjugación	$x^*[-n] \leftrightarrow [X(Z)]^* _{ z =1}$
Multiplicación por coseno	$\cos[\omega n] x[n] \leftrightarrow \frac{1}{2} [X(e^{j\omega} z) + X(e^{-j\omega} z)]$
Multiplicación por seno	$\sin[\omega n] x[n] \leftrightarrow \frac{j}{2} [X(e^{j\omega} z) - X(e^{-j\omega} z)]$

2.2.2 La transformada discreta en el tiempo de Fourier

La transformada discreta en el tiempo de Fourier de una secuencia $x[n]$ está definida como

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n} \quad (2-23)$$

Debido a la periodicidad de 2π del exponente complejo, $X(e^{j\omega})$ es periódica: $X(e^{j\omega}) = X(e^{j(\omega+2\pi)})$. Si $x[n]$ es obtenida por un muestreo regular de una señal continua en el tiempo $x_{ct}(t)$ de tal manera que $x_{ct}(t) = x[nT]$, donde T es el período de muestreo, ω puede ser entendida como la frecuencia normalizada $\omega = 2\pi fT$.

Las propiedades de la transformada discreta en el tiempo de Fourier son derivadas fácilmente de las propiedades de la transformada Z escogiendo $z = e^{j\omega}$. En la tabla III se muestran algunas propiedades de la transformada de Fourier en tiempo discreto.

Tabla III. Propiedades importantes de la transformada de Fourier en tiempo discreto.

Propiedad	Relación
Desplazamiento	$x[n - n_0] \leftrightarrow e^{-j\omega n_0} X(e^{j\omega})$
Convolución	$x[n] * y[n] \leftrightarrow X(e^{j\omega})Y(e^{j\omega})$
Multiplicación en el dominio del tiempo	$x[n]y[n] \leftrightarrow \frac{1}{2\pi} X(e^{j\omega}) * Y(e^{j\omega})$ $x[n]y[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j(\omega-\tau)}) * Y(e^{j\tau}) d\tau$

La transformada inversa puede ser encontrada si la secuencia $x[n]$ es absolutamente sumable, por lo que puede ser obtenida a partir de $X(e^{j\omega})$ por medio de

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega n} d\omega \quad (2-24)$$

Cabe notar que la ecuación (2-24) es la inversa de la transformada Z, evaluada en el círculo unitario.

2.2.2.1 Teorema de Parseval

Como en el caso de señales continuas en el tiempo, la energía de la señal puede ser calculada en el dominio de la frecuencia y el tiempo. Si una señal $x[n]$ es absolutamente sumable cuadráticamente ($(x \in \ell_1(-\infty, \infty)) \cap (x \in \ell_2(-\infty, \infty))$), entonces

$$E_x = \sum_{-\infty}^{\infty} |x[n]|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(e^{j\omega})|^2 d\omega \quad (2-25)$$

Esta expresión puede ser entendida como una expansión en series de $X(e^{j\omega})$ para un espectro periódico de 2π , donde los valores de $x[n]$ son los coeficientes de la serie expandida.

2.2.2.2 Relación entre la transformada Z y la transformada en tiempo discreto de Fourier

Se puede notar que existe una relación estrecha entre la transformada Z y la transformada de Fourier en tiempo discreto. En particular, si se substituye la variable compleja z por la variable compleja $e^{j\omega}$, la transformada Z se reduce a la transformada de Fourier.

Este es uno de los motivos para la utilización de la notación $X(e^{j\omega})$ en la transformada de Fourier. Cuando ésta existe, es simplemente $X(z)$ con $z = e^{j\omega}$. Esto corresponde a restringir la variable z para que tenga módulo unidad. Es decir, para $|z| = 1$ la transformada Z equivale a la transformada de Fourier. Se puede expresar la variable compleja en forma polar

$$z = re^{j\omega} \quad (2-26)$$

Si z se representa de esta forma en la ecuación (2-15) se convierte en

$$X(re^{j\omega}) = \sum_{-\infty}^{\infty} (x[n]r^{-n})e^{-j\omega n} \quad (2-27)$$

La ecuación (2-27) se puede interpretar como la transformada de Fourier del producto de la secuencia original $x[n]$ por la secuencia exponencial r^{-n} . Cuando $r = 1$ se reduce a la transformada de Fourier de $x[n]$.

Como la transformada Z es una función de variable compleja, es conveniente representarla en el plano complejo z . En el plano z , el contorno correspondiente a $|z| = 1$ es la circunferencia de unidad. La transformada Z evaluada en la circunferencia unidad es la transformada en tiempo discreto de Fourier.

La representación de la transformada Fourier en una serie de potencias no converge para todas las secuencias, no siempre se produce un valor finito para la suma infinita. De igual manera, la transformada Z no converge para todos los valores de z . En el caso de la transformada de Fourier, si la secuencia es absolutamente sumable, esta transformada converge a una función continua de ω .

Para la transformada Z, la multiplicación por la secuencia exponencial real r^{-n} , hace que la transformada Z converja aunque la transformada de Fourier no la haga. En resumen, si la región de convergencia de la transformada Z incluye la circunferencia unidad, la transformada de Fourier de la secuencia también converge.

2.2.3 La transformada discreta de Fourier

Para el caso de secuencias de duración finita, es posible desarrollar una representación de Fourier alternativa denominada transformada discreta de Fourier (DFT, *Discrete Fourier Transform*). La DFT es una secuencia, no una función de variable continua, y corresponde a muestras equi-espaciadas en frecuencia de la transformada de Fourier de la señal. Junto con su importancia teórica, tiene un papel crucial en la realización de una gran variedad de algoritmos de tratamiento digital de señales. El motivo es que existen algoritmos eficientes para el cálculo de la DFT.

Suponiendo una secuencia $\tilde{x}[n]$ ¹ que es periódica de período N, de forma que $\tilde{x}[n] = \tilde{x}[n + rN]$ para cualquier valor entero de n y r . Como en el caso de señales de tiempo continuo, esa secuencia se puede representar mediante un desarrollo en serie de Fourier en forma de una suma de de secuencias exponenciales complejas armónicamente relacionadas. Es decir, exponenciales complejas con frecuencias que son múltiplos enteros de la frecuencia fundamental ($2\pi/N$) asociada a la secuencia periódica $\tilde{x}[n]$. El desarrollo de la serie de Fourier viene dado por:

$$\tilde{x}[n] = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{X}[k] e^{j(2\pi/N)kn} \quad (2-28)$$

¹ Se utilizará la tilde($\tilde{}$) para indicar las secuencias periódicas siempre que sea importante distinguir entre secuencias periódicas y no periódicas.

Debido a la periodicidad de las exponenciales armónicamente relacionadas, sólo se necesita un conjunto de N exponenciales complejas para expresar la serie de Fourier.

Los coeficientes de desarrollo de Fourier $\tilde{X}[k]$ en la ecuación (2-28) se pueden encontrar a partir de:

$$\tilde{X}[k] = \sum_{n=0}^{N-1} \tilde{x}[n] e^{-j(2\pi/N)kn} \quad (2-29)$$

Las ecuaciones (2-28) y (2-29) forman una pareja de análisis-síntesis que se denomina representación mediante el desarrollo en serie de Fourier discreto (DSF, *Discrete Fourier Series*) de una secuencia periódica. Muchas veces se usa la notación:

$$W_N = e^{-j(2\pi/N)} \quad (2-30)$$

Con esta notación, la pareja de análisis-síntesis se expresa de la siguiente manera:

$$\text{Ecuación de análisis:} \quad \tilde{X}[k] = \sum_{n=0}^{N-1} \tilde{x}[n] W_N^{kn} \quad (2-31)$$

$$\text{Ecuación de síntesis:} \quad \tilde{x}[n] = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{X}[k] W_N^{-kn} \quad (2-32)$$

Estos resultados se pueden expandir para secuencias de duración finita, donde nos interesa el intervalo entre cero y $(N - 1)$. Fuera de este intervalo, $X[k]$ y $x[n]$ son cero. Esto porque siempre podemos asociar a cada secuencia finita de longitud N una secuencia periódica de manera que

$$\tilde{x}[n] = \sum_{r=-\infty}^{\infty} x[n - rN] \quad (2-33)$$

La ecuación (2-33) puede ser expresada como

$$\begin{aligned}\tilde{x}[n] &= x[(n \text{ módulo } N)] \\ &= x[((n))_N]\end{aligned}\tag{2-34}$$

La secuencia de longitud finita puede ser recuperada a partir de $\tilde{x}[n]$ mediante

$$x[n] = \begin{cases} \tilde{x}[n], & 0 \leq n \leq N - 1 \\ 0, & \text{en el resto} \end{cases}\tag{2-35}$$

Por lo que para secuencias de duración finita las ecuaciones (2-31) y (2-32) estarán definidas en el intervalo $0 \leq n \leq N - 1$ y se convierten en las ecuaciones (2-36) y (2-37), a estas ecuaciones se les denominará ecuaciones de análisis y síntesis de la transformada discreta de Fourier, DFT.

$$\text{Ecuación de análisis:} \quad X[k] = \sum_{n=0}^{N-1} x[n] W_N^{kn}\tag{2-36}$$

$$\text{Ecuación de síntesis:} \quad x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] W_N^{-kn}\tag{2-37}$$

Al definir la DFT, simplemente se reconoce que se está interesado sólo en los valores de $x[n]$ en el intervalo $0 \leq n \leq N - 1$ ya que $x[n]$ es cero fuera de este intervalo, y que solo estamos interesados en los valores de $X[k]$ en el intervalo $0 \leq k \leq N - 1$, ya que son los únicos valores necesarios en la ecuación (2-32).

La tabla IV muestra algunas propiedades importantes de la transformada discreta de Fourier para secuencias de longitud finita. Se debe tomar en cuenta el supuesto de longitud finita y la periodicidad implícita de la representación de secuencias de longitud finita mediante la DFT.

Tabla IV. Propiedades importantes de la transformada discreta de Fourier.

Propiedad	Relación
Linealidad	$\alpha x[n] + \beta y[n] \leftrightarrow \alpha X[k] + \beta Y[k]$
Desplazamiento circular	$X_\mu[m] = W_N^{-\mu m} X[m]$ $W_N^{kn} x[n] \leftrightarrow \sum_{n=0}^{N-1} x[n] W_N^{n(m+k)}$ $W_N^{kn} x[n] \leftrightarrow X[((m+k))_N]$
Dualidad	$x[n] \leftrightarrow X[k]$ $X[n] \leftrightarrow Nx \left[((-k))_N \right], \quad 0 \leq k \leq N-1$
Multiplicación y convolución circular	$Y[m] = X_1[m]X_2[m]$ $y[n] = \sum_{p=0}^{N-1} x_1[p]x_2[((n-p))_N]$
Conjugación compleja	$x^*[n] \leftrightarrow X^*[N-n]$ $x^*[N-n] \leftrightarrow X^*[n]$

2.2.4 La transformada rápida de Fourier

Al conjunto de algoritmos para calcular la transformada discreta de Fourier se le denomina transformada rápida de Fourier (FFT, *fast Fourier transform*). El propósito de estos algoritmos es el cálculo de una DFT de N puntos que corresponda a N muestras de la transformada de Fourier en las N frecuencias equiespaciadas $\omega_k = 2\pi k/N$. Para el cálculo de la DFT en el intervalo de frecuencias $0 \leq \omega \leq 2\pi$ es posible que existan algoritmos más eficientes y flexibles, pero son menos eficientes que los algoritmos FFT para todos los valores de la DFT.

La complejidad y eficiencia de un determinado algoritmo o su realización, depende tanto de la tecnología disponible como de la aplicación. Utilizaremos el número de multiplicaciones y sumas aritméticas como medida de la complejidad computacional. Esta medida es simple de aplicar, y el número de multiplicaciones y sumas está directamente relacionado con la velocidad de cómputo cuando los algoritmos se realizan en computadores digitales de propósito general o en microprocesadores de propósito específico. Pero también, existen otras medidas que pueden ser más apropiadas en algunas ocasiones. Por ejemplo, en aplicaciones VLSI son importantes el área del chip y los requisitos de potencia. Estas consideraciones pueden no estar relacionadas directamente con el número de operaciones aritméticas.

Para una señal de entrada compleja $x[n]$ de longitud N , la implementación de una matriz W requiere N^2 multiplicaciones complejas. La idea detrás de la FFT es factorizar W en el producto de matrices con menor complejidad que de forma conjunta su costo de implementación sea menor que la DFT de forma directa.

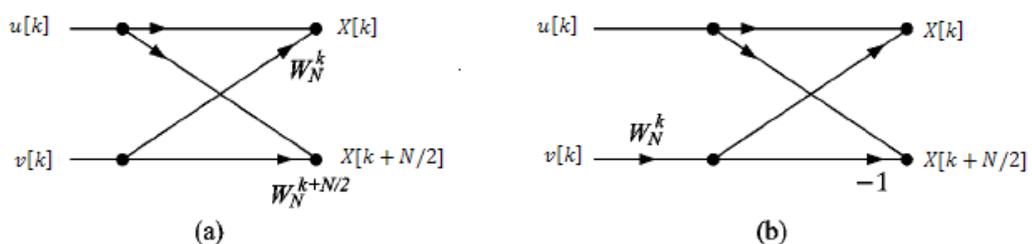
En este apartado se pretende describir de forma cualitativa algunos algoritmos de la FFT cuya longitud es una potencia de dos. En particular, se discutirán los algoritmos *radix-2*, *radix-4* y el *split-radix*.

2.2.4.1 Radix-2 decimación en el tiempo FFT

Consideremos una DFT de N puntos, donde N es una potencia de dos. El primer paso para una implementación rápida es el descomponer la señal en el tiempo $x[n]$ en sus componentes pares e impares. Ahora se ha descompuesto una DFT de N puntos en dos DFT de $N/2$ puntos y algunas operaciones extra para combinar las dos salidas de las DFT. Con esto, se ha reducido la complejidad de N^2 a una complejidad de $\frac{N^2}{2} + N$. Como N se ha considerado como una potencia de dos, el mismo principio de descomposición puede ser usado para DFT más pequeñas y la complejidad puede ser reducida. Esta descomposición se puede llevar a cabo hasta que se alcanza una DFT de dos puntos.

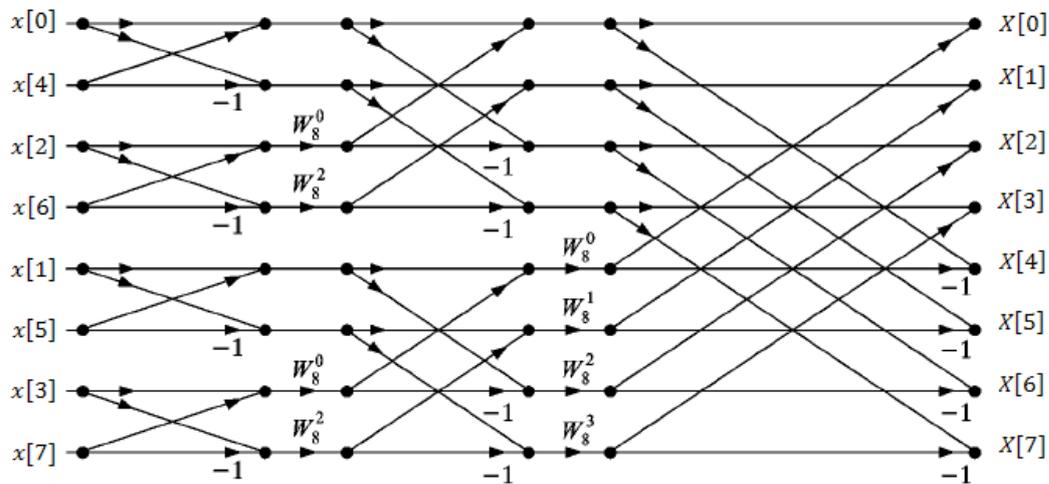
Resulta que todas las etapas de la FFT se encuentran compuestas de los llamados grafos de mariposa como se muestran en la figura 21. En este caso, ambos grafos son equivalentes pero el de la figura 21(b) nos ahorra una multiplicación compleja.

Figura 21. Diagramas de flujo equivalentes.



El diagrama de flujo completo para una FFT de 8 puntos se muestra en la figura 22. Se puede ver que, los valores de salida aparecen en su orden natural, pero los valores de entrada aparecen en un orden permutado. Esto es para todos los N . El orden de los valores de entrada es conocido como el orden del bit invertido.

Figura 22. Diagrama de flujo para una FFT de 8 puntos.



Como la operación mariposa en cada etapa de la FFT es independiente una de otra, los cálculos computacionales de la FFT pueden ser realizados “en lugar”. Esto es, el par de salidas de una mariposa es escrito sobre la entrada. Luego que esto se ha hecho para todas las mariposas de una etapa de procesamiento, uno puede proceder a la siguiente etapa. Por lo que, solo una memoria $N + 1$ de tamaño es necesaria para el computo de una FFT de N puntos.

La complejidad computacional de la FFT es de la siguiente manera. Cada etapa de la FFT requiere $N/2$ multiplicaciones complejas y N sumas. El número de etapas es $\log_2 N$. Esto lleva a un total de $\frac{1}{2}N \log_2 N$ de multiplicaciones complejas y a $N \log_2 N$ de sumas.

Sin embargo, como las DFT de 2 puntos no requieren multiplicaciones, y además que la DFT de 4 puntos involucra multiplicaciones con $1, -1, j$ y $-j$ solamente, el número actual de multiplicaciones complejas es incluso menor que $\frac{1}{2}N \log_2 N$.

2.2.4.2 Radix-4 FFT

La decimación en frecuencia *radix-4* es derivada a partir de la separación de la DFT en cuatro sub secuencias. Por lo que, se ha remplazado la computación de una DFT de N puntos por cuatro de $N/4$ puntos. Una de estas cuatro DFT no requiere computación alguna, y las otras requieren una multiplicación compleja por punto. Comparado con la FFT *radix-2* esto significa $3 \times (N/4)$ multiplicaciones en lugar de $N/2$ multiplicaciones para los factores de raíz unitaria compleja (*twiddle factors*). Sin embargo, el algoritmo *radix-4* requiere solamente $N/4$ puntos de la DFT, y requiere solo la mitad de etapas que un FFT *radix-2*. Por lo tanto, el número total de multiplicaciones es menor para el caso del *radix-4*.

2.2.4.3 Split-radix FFT

La FFT *split-radix*, es una mezcla de los algoritmos *radix-2* y *radix-4*, requiere un menor número de operaciones que todos los algoritmos de FFT conocidos. Es fácil de programar en una computadora. La aproximación *radix-2* es usada para el cómputo de las frecuencias numeradas pares, y la aproximación *radix-4* es usada para el cómputo de las dos sub secuencias de longitud $(N/4)$ de las frecuencias numeradas impares. La aplicación del algoritmo *split-radix* requiere menos multiplicaciones que un algoritmo FFT *radix-2* o *radix-4*, porque menos multiplicaciones complejas ocurren.

3. ESTUDIO DE LA PSICOACÚSTICA

3.1 Sistema auditivo humano

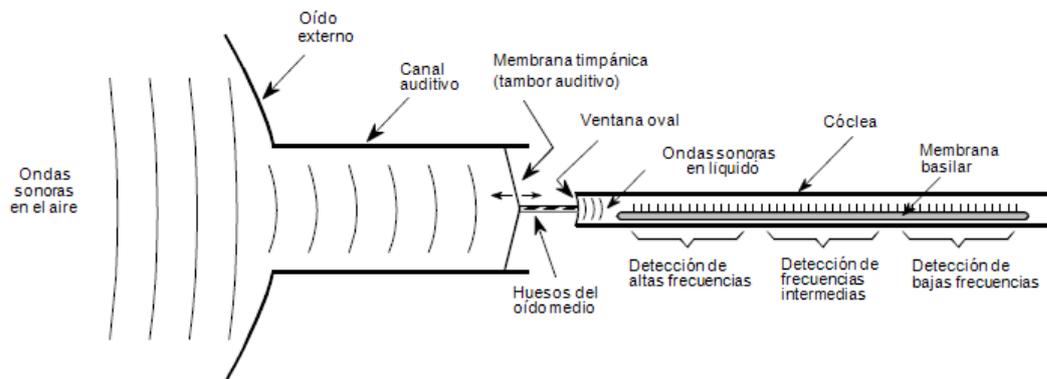
El oído humano es un órgano extremadamente complejo. Para hacer las cosas aún más difíciles, la información de dos oídos es combinada en una red de neuronas muy compleja, el cerebro humano. En este apartado se pretende dar una descripción básica del fenómeno de la audición humana; aunque aún existen algunos efectos sutiles y fenómenos pocamente entendidos en la audición humana.

La figura 23 muestra las estructuras y procesos principales que componen la audición humana. El oído externo está compuesto por dos partes, la solapa de piel y cartílago adherida al lado de la cabeza, y el canal auditivo, un tubo de unos 0.5 cm de diámetro que se extiende unos 3 cm dentro de la cabeza. Estas estructuras direccionan sonidos ambientales al los órganos sensitivos del oído medio e interno localizados de forma segura dentro de los huesos de la cabeza. Estirada al final del canal auditivo existe una delgada capa de tejido llamada la membrana timpánica o tímpano.

Las ondas sonoras golpean el tímpano y lo hacen vibrar. El oído medio es un conjunto de pequeños huesos que transfieren estas vibraciones a la cóclea (oído interno) donde son convertidas en pulsos neurales. La cóclea es un tubo lleno de líquido de aproximadamente 2 mm de diámetro y 3 cm de longitud. Aunque en la figura 23 se muestra como una estructura larga, la cóclea se encuentra curvada y se parece más a una concha de caracol.

Cuando una onda sonora trata de pasar de aire a líquido, solo una pequeña fracción del sonido es transmitido a través de la interface, mientras que la energía restante es reflejada. Esto porque el aire posee una impedancia mecánica baja (baja presión acústica y alta velocidad de partículas resultado de baja densidad y alta compresibilidad), mientras que el líquido posee una alta impedancia mecánica. Es decir, se requiere un mayor esfuerzo mover la mano en agua que el que se necesita para moverla en el aire. Esta diferencia en impedancia mecánica resulta en que la mayor parte del sonido sea reflejada de la interface aire-líquido.

Figura 23. Diagrama funcional del oído humano.



El oído medio es una red de ajuste de impedancia que aumenta la fracción de energía del sonido del líquido entrante en oído interno. La mayor parte de la conversión de impedancia resulta de la diferencia de área entre el tímpano (recibiendo sonido del aire) y la ventana oval (transmitiendo sonido en el líquido). El tímpano posee un área de unos 60 mm^2 , mientras la ventana oval tiene un área aproximadamente de unos 4 mm^2 . Dado que la presión es igual a la fuerza dividida por el área, esta diferencia de área aumenta la presión de la onda sonora 15 veces.

Contenida dentro de la cóclea se encuentra la membrana bacilar, la estructura que soporta alrededor de 12,000 células sensoriales que forman el nervio coclear. La membrana bacilar es rígida cerca de la ventana oval, y se vuelve más flexible hacia su extremo opuesto, lo que permite que funcione como un analizador de espectro. Cuando es expuesta a señales de altas frecuencias, la membrana bacilar resuena donde es rígida, resultando en una excitación de las células nerviosas cerca de la ventana oval. De igual manera, sonidos de baja frecuencia excitan las células nerviosas al extremo de la membrana bacilar. Esto hace que fibras específicas en el nervio coclear respondan a frecuencias específicas. Esta organización es llamada el principio de lugar, y es preservada en toda la vía auditiva en el cerebro.

Otro esquema de codificación de información también usado por la audición humana es llamado el principio de *Volley*. Aquí las células nerviosas transmiten información al generar pulsos eléctricos de corta duración llamados potenciales de acción. Una célula nerviosa en la membrana bacilar puede codificar información de audio al producir un potencial de acción en respuesta a cada ciclo de la vibración. Por ejemplo, una onda sonora de 200 Hz puede ser representada por una neurona produciendo 200 potenciales de acción por segundo. Sin embargo, esto solo funciona para frecuencias por debajo a los 500 Hz, siendo la razón máxima a la que las neuronas pueden producir potenciales de acción. El oído humano supera este problema al permitir que varias células nerviosas tomen turnos para realizar esta única tarea. Por ejemplo, un tono de 3000 Hz puede ser representada por diez células nerviosas actuando alternadamente 300 veces por segundo. Esto extiende el principio de *Volley* hasta 4 kHz, por encima de la cual el principio de lugar es usado exclusivamente.

Es común expresar la intensidad del sonido en una escala logarítmica, llamada decibel SPL (por sus siglas en inglés *Sound Power Level*). A partir de la intensidad del sonido se puede obtener una expresión para el SPL como se muestra en (3-1)

$$L_{SPL} = 10 \log_{10} \left(\frac{I}{I_0} \right) \quad (3-1)$$

En esta escala, 0 dB SPL es una onda sonora con potencia de 10^{-12} watts/m², I_0 , siendo el sonido más débil detectable por el oído humano. Una conversación normal se encuentra en los 60 dB SPL, mientras que un daño doloroso ocurre en el oído en unos 140 dB SPL. La tabla V muestra la relación entre la intensidad del sonido percibida y el volumen

Otra forma de expresar el valor de SPL es a partir de los niveles de presión de la señal, la cual está definida como:

$$L_{SPL} = 20 \log_{10} \left(\frac{P}{P_0} \right) \quad (3-2)$$

Donde el valor de presión más débil detectable, P_0 , es de 2×10^{-5} N/m².

La diferencia entre los sonidos más fuertes y finos que los humanos pueden escuchar están entre 120 dB, un rango de un millón en amplitud. Quienes escuchan pueden detectar un cambio de volumen cuando la señal es alterada en 1 dB (un cambio del 12% en amplitud). En otras palabras, existen 120 niveles de volumen que pueden ser percibidos desde el más fino suspiro al más fuerte trueno. La sensibilidad del oído es impresionante: cuando se escuchan sonidos muy débiles, el tímpano vibra menos que el diámetro de una sola molécula.

Tabla V. Unidades de la intensidad del sonido.

	Watts/cm²	Decibeles SPL	Ejemplo de Sonido
	10 ⁻²	140 dB	Dolor
	10 ⁻³	130 dB	
	10 ⁻⁴	120 dB	Desconfort
	10 ⁻⁵	110 dB	Conciertos de rock
	10 ⁻⁶	100 dB	
	10 ⁻⁷	90 dB	Límite OSHA para el ruido industrial
	10 ⁻⁸	80 dB	
	10 ⁻⁹	70 dB	
	10 ⁻¹⁰	60 dB	Conversación normal
	10 ⁻¹¹	50 dB	
	10 ⁻¹²	40 dB	Más debil audible a 100 Hz
	10 ⁻¹³	30 dB	
	10 ⁻¹⁴	20 dB	Más debil audible a 10 kHz
	10 ⁻¹⁵	10 dB	
	10 ⁻¹⁶	0 dB	Más debil audible a 3 kHz
	10 ⁻¹⁷	-10 dB	
	10 ⁻¹⁸	-20 dB	

↑
 Fuerte
 ↓
 Suave
 ↓

La percepción de volumen y la potencia del sonido se relaciona de forma aproximada en un exponente de 1/3. Por ejemplo, si se aumenta la potencia del sonido en un factor de diez, quienes escuchen reportaran que el volumen se ha aumentado en un factor de dos. ($10^{1/3} \approx 2$). Este es un problema cuando se desean eliminar sonidos ambientales indeseables, como por ejemplo, el estéreo en un apartamento contiguo. Suponiendo que se cubra el 99% de las paredes con algún material a prueba de sonidos, fallando solamente en 1% en la superficie debido a paredes, esquinas, ventanas, etc. Aunque la potencia del sonido se ha disminuido a un 1% de su valor inicial, el volumen percibido solo ha disminuido en un 20% o ($0.01^{1/3} \approx 0.2$).

El rango de audición humana se considera que está entre 20 Hz y 20 kHz, pero es más sensible a sonidos que estén entre 1 kHz y 4 kHz.

Por ejemplo, un escucha puede detectar un sonido tan bajo de 0 dB SPL a una frecuencia de 3 kHz, pero se requieren 40 dB SPL para una frecuencia de 100 Hz (un incremento en amplitud de 100). Quien escucha puede decir que dos tonos difieren entre sí si sus frecuencias difieren más de un 0.3% a 3 kHz. Esto aumenta a un 3% a frecuencias de 100 Hz. En forma de comparación, teclas contiguas en un piano difieren en un 6% en frecuencia.

La ventaja principal de tener dos oídos es la habilidad de identificar la dirección del sonido. Los humanos podemos detectar la diferencia entre dos fuentes de sonido que se encuentran situadas tan cerca como tres grados entre ellas, como la altura de una persona a 10 metros de distancia. Esta información direccional es obtenida de dos maneras distintas. Primero, frecuencias por encima de 1 kHz son fuertemente sombreadas por la cabeza. Es decir, el oído más cercano al sonido recibe una señal más fuerte que el oído en la parte opuesta de la cabeza.

La segunda clave de la direccionalidad es que el oído en el lado más lejano de la cabeza escucha el sonido ligeramente más tarde que el oído más cercano, debido a la distancia más grande de la fuente. Basándonos en el tamaño típico de una cabeza (unos 22 cm) y la velocidad del sonido (de unos 340 metros por segundo), una discriminación angular de tres grados requiere una precisión de unos 30 microsegundos. Por lo que este tiempo requiere del principio de *Volley*, esta clave para la direccionalidad es usada predominantemente para sonidos menores que 1 kHz.

Ambas de estas fuentes de información direccional son ayudadas en gran parte por la habilidad de poder girar la cabeza y observar el cambio en las señales. Una sensación interesante ocurre cuando un escucha es presentado con exactamente los mismos sonidos en los dos oídos, como el de escuchar un sonido monoaural a través de audífonos. El cerebro llega a la conclusión de que el sonido proviene del centro de la cabeza del escucha.

Mientras que el ser humano puede detectar la dirección del sonido, lo hace de forma pobre al identificar la distancia de la fuente de sonido. Esto es porque existen pocas claves disponibles en una onda sonora que provean esta información. La audición humana débilmente percibe que sonidos de alta frecuencia se encuentran cerca, mientras que sonidos de baja frecuencia se encuentran lejos. Esto es porque las ondas sonoras disipan sus altas frecuencias conforme se propagan en largas distancias.

El contenido de eco es otro punto débil en la distancia, dando la percepción de estar en una habitación. Por ejemplo, sonidos en grandes auditorios contendrán ecos con intervalos de unos 100 milisegundos, mientras que 10 milisegundos es típico para pequeñas oficinas. Algunas especies han evolucionado para resolver este intervalo de rangos al usar un sonar activo.

Por ejemplo, los murciélagos y delfines producen chasquidos y chirridos que se reflejan por objetos cercanos. Al medir el intervalo entre la transmisión y el eco, estos animales pueden localizar objetos con hasta 1 cm de resolución. Algunos experimentos han mostrado que algunos humanos, particularmente ciegos, pueden usar la localización por eco activo en una pequeña escala.

3.2 Timbre

La percepción de un sonido continuo, como la nota de un instrumento musical, a menudo es dividido en tres partes: volumen, tono y timbre. El volumen es una medida de la intensidad de la onda sonora. El tono es la frecuencia de la componente fundamental en el sonido, esto es, la frecuencia con la cual la forma de onda se repite así misma. Mientras existen efectos sutiles en estas dos percepciones, son de forma directa asociadas con cantidades físicas.

El timbre es más complicado, siendo determinado por el contenido armónico de la señal. La figura 24 ilustra dos formas de onda, cada una formada al sumar una onda sinusoidal de 1kHz con una amplitud de la unidad, a una onda sinusoidal de 3 kHz con una amplitud de un medio. La diferencia entre las dos formas de onda es que la mostrada en (b) tiene la frecuencia más alta invertida antes de la suma. Dicho de otra manera, la tercera armónica (3 kHz) posee un cambio de fase de 180 grados comparado con la primera armónica (1 kHz). A pesar de las dos distintas formas de onda en el dominio del tiempo, estas dos señales suenan idénticamente. Esto es porque la audición se encuentra basada en la amplitud de las frecuencias, y es bastante insensible a su fase.

La forma de la onda en el dominio del tiempo se encuentra indirectamente relacionada con la audición, y usualmente no es considerada en los sistemas de audio.

La insensibilidad del oído a la fase puede ser entendida al examinar como el sonido se propaga a través del ambiente. Supongamos que se escucha a una persona que se encuentra hablando en un cuarto pequeño. Gran parte del sonido que llega hasta el oyente es reflejado en las paredes, techo y el piso. Ya que la propagación del sonido depende de la frecuencia (como la atenuación, reflexión y resonancia), distintas frecuencias llegarán a los oídos a través de distintos caminos. Esto significa que la fase relativa de cada frecuencia cambiará conforme se mueva en el cuarto. Ya que el oído hace caso omiso de estas variaciones de fase, se percibe que la voz es invariable cuando se cambia de posición. Desde un punto de vista físico, la fase de una señal de audio se vuelve aleatoria cuando se propaga a través del ambiente complejo. Es decir, el oído es insensible a la fase porque no contiene información útil.

Figura 24. Detección de fase del oído humano.

Figura 24 (a). Suma de armónicas en fase.

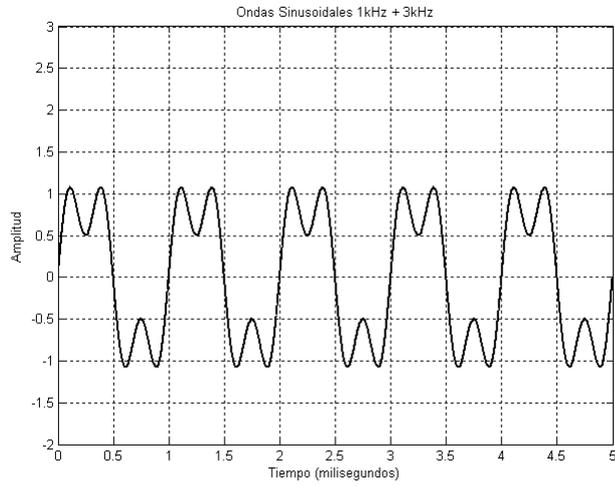
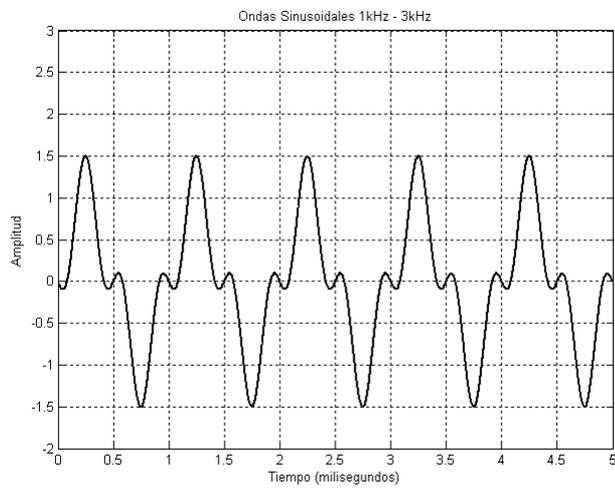


Figura 24 (b). Suma de armónicas con fase de 180 grados.



No se podría decir que el oído es completamente sordo a la fase. Esto es porque un cambio de fase puede recomodar la secuencia de tiempo de una señal de audio y hacer que posea una mayor duración. Aunque el oído puede distinguir entre dos sonidos en base a su duración, esto es algo que no sucede comúnmente en un ambiente normal.

Supongamos que se le pide a un violinista que toque una nota, sea en este caso A por encima de C medio. Cuando la forma de onda es mostrada en un osciloscopio, se parece mucho a la forma de diente de sierra de la figura 25 (a). Este es un resultado de la resina pegajosa aplicada a las fibras del arco del violinista. Conforme el arco es movido a través de la cuerda, la onda es formada y la cuerda se pega al arco, hasta que es tirada atrás, donde la onda se corta. Este ciclo se repite una y otra vez resultando en una forma de onda de diente de sierra.

Figura 25. Forma de onda generada por un violín.

Figura 25 (a). Forma de onda en el dominio del tiempo.

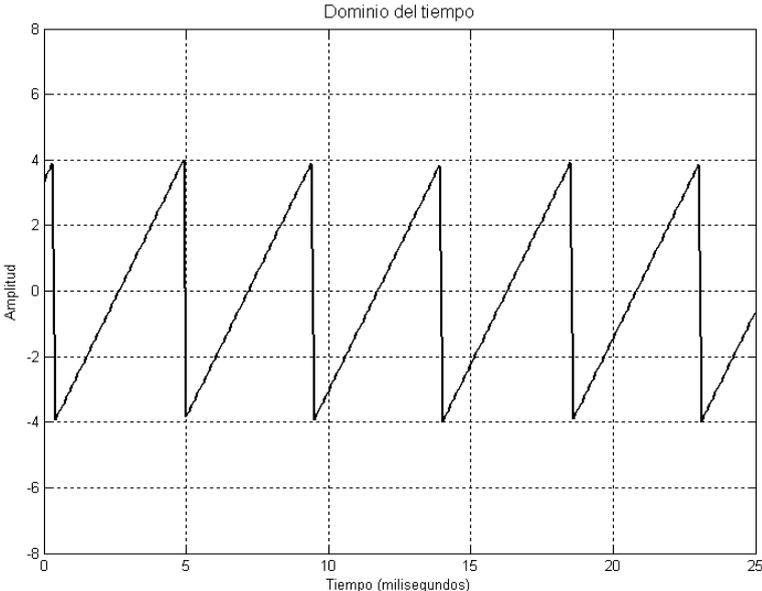
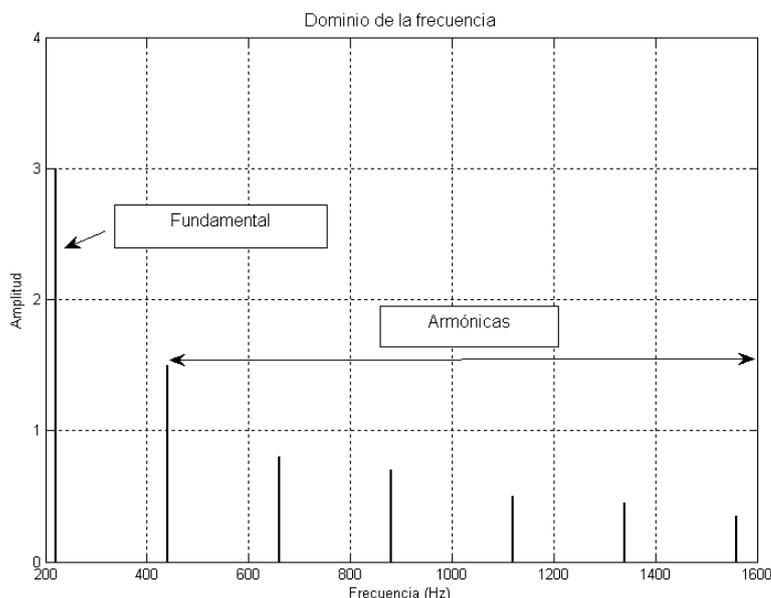


Figura 25 (b). Forma de onda en el dominio de la frecuencia.



La figura 25(b) muestra como este sonido es percibido por un oído, una frecuencia de 220 Hz, más armónicas de 440, 660, 880 Hz, etc. Si esta nota fuese tocada en otro instrumento, la forma de onda se vería distinta; sin embargo, el oído aún escucharía una frecuencia de 220 Hz más las armónicas. Ya que los dos instrumentos producen la misma frecuencia fundamental para esta nota, suenan similar, y se dice que tienen un tono idéntico. Ya que la amplitud relativa de las armónicas es diferente, ellas no suenan idénticamente, y se dice que tienen distinto timbre.

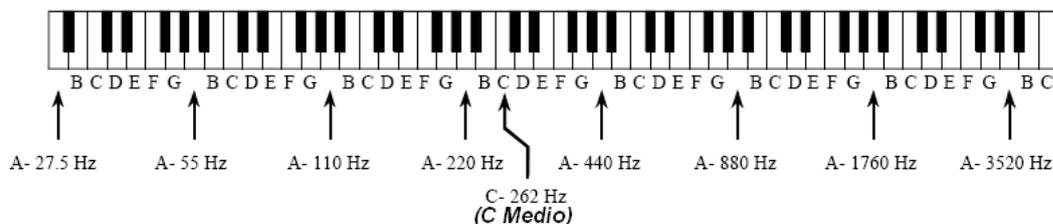
Comúnmente se dice que el timbre está determinado por la forma de onda. Esto es cierto, pero ligeramente engañoso. La percepción del timbre surge del oído al detectar armónicas. Mientras el contenido armónico está determinado por la forma de la onda, la insensibilidad del oído respecto a la fase hace esta relación en una sola vía. Es decir, una forma de onda particular posee un único timbre, mientras un timbre particular tiene un número infinito de formas de onda.

El oído se encuentra bastante acostumbrado a escuchar frecuencias fundamentales más armónicas. Si un escucha se le presenta una combinación de ondas sinusoidales de 1 kHz y 3 kHz, reportará que este sonido es natural y acogedor. Si se usaran ondas sinusoidales de 1 kHz y 3.1 kHz sonará de forma cuestionable.

Esta es la base fundamental de la escala musical, como muestra el teclado del piano en la figura 26. Presionando la tecla que se encuentra más a la izquierda produce una frecuencia fundamental de 27.5 Hz, más armónicas de 55, 110, 220, 440, 880 Hz, etc. (existen armónicas entre estas frecuencias, pero no son importantes para esta discusión). Estas armónicas corresponden a la frecuencia fundamental producida por otras teclas en el teclado.

Específicamente, cada siete teclas blancas es un armónico de la tecla izquierda lejana. Esto es, la octava tecla de la izquierda tiene una frecuencia fundamental de 55 Hz, la 15ava tecla tiene una frecuencia fundamental de 110 Hz, etc. Siendo armónicas una de la otra, estas teclas suenan de forma similar cuando son tocadas, y son armoniosas cuando son tocadas al unísono. Por esta razón, todas son llamadas la nota A. De esta misma manera, la tecla blanca inmediatamente a la derecha de cada A es llamada B, y son armónicas una de otra. Este patrón se repite para las siete notas: A, B, C, D, E, F y G.

Figura 26. El teclado de un piano.



El término octava significa un factor de dos en frecuencia. En un piano, una octava comprende ocho teclas blancas. La frecuencia del piano se duplica cada siete teclas blancas, y todo el teclado se extiende un poco más de siete octavas. El rango de audición humana generalmente se encuentra limitada de 20 Hz a 20 kHz, correspondiendo a aproximadamente $\frac{1}{2}$ octava a la izquierda, y dos octavas a la derecha del teclado de un piano. Ya que las octavas se encuentran basadas en la duplicación de la frecuencia cada cierto número fijo de teclas, son una representación logarítmica de la frecuencia. Esto es importante porque la información de audio generalmente se encuentra distribuida de esta manera.

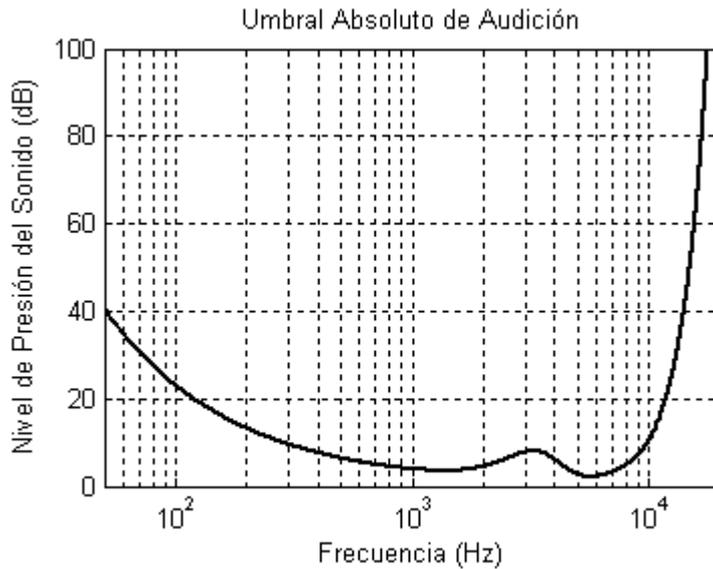
Por ejemplo, como tanta información existe en la octava de 50 Hz a 100 Hz, como en la octava que comprende las frecuencias de 10 kHz a 20 kHz. Aunque el piano solo cubre cerca de un 20% de las frecuencias que el humano puede escuchar (4 kHz de 20 kHz), puede producir más del 70% de información auditiva que los humanos pueden percibir (7 de 10 octavas). De igual manera, la frecuencia más alta que un humano puede detectar cae de 20 kHz a 10 kHz en el transcurso de la vida adulta. Sin embargo, esto es una pérdida de solamente el 10% de la habilidad para escuchar (una octava de diez). Esta distribución logarítmica de información afecta de forma directa el período de muestreo de la señal de audio.

3.3 Umbral absoluto y umbral de enmascaramiento

El umbral absoluto de audición representa la intensidad mínima requerida para que una persona pueda detectar un tono puro en la ausencia de cualquier otro sonido. El umbral absoluto puede ser descrito en el dominio de la frecuencia por la curva dada en (3-3), mostrada en la figura 27.

$$T = 3.64 \left(\frac{f}{1000} \right)^{-0.8} - 6.5e^{-0.6 \left(\frac{f}{1000} - 3.3 \right)^2} + 10^{-3} \left(\frac{f}{1000} \right)^4 \quad (3-3)$$

Figura 27. Umbral absoluto de audición.



El proceso de enmascaramiento ocurre en el oído cuando un sonido es inaudible o menos audible por la presencia de otro sonido. El umbral de enmascaramiento o la mínima distorsión notable, JND por sus siglas en inglés, es el umbral de detección de una señal (prueba o enmascarada), en la presencia de otra señal (máscara). El umbral de enmascaramiento puede ser concebido como una modificación en el umbral absoluto donde la presencia de la máscara aumenta el umbral de detección. De forma equivalente, el umbral absoluto puede ser concebido como un caso especial del umbral de enmascaramiento donde no existe ninguna máscara presente. Naturalmente, esto lleva a la idea que los umbrales de enmascaramiento son dependientes de la señal y que sus formas están determinadas por la presencia (o ausencia) de varias componentes espectrales.

3.4 Fenómenos psicofísicos

Mucho del conocimiento del sistema auditivo humano y el hecho de desarrollar modelos que se asemejen a este, como se verá en el capítulo 8, provienen de experimentos psicofísicos más que del entendimiento fisiológico del sistema auditivo. Las mediciones psicofísicas son generalmente obtenidas a partir de extensos exámenes a personas que son representativos de la población en general. A continuación se presentan algunos resultados de esta experimentación.

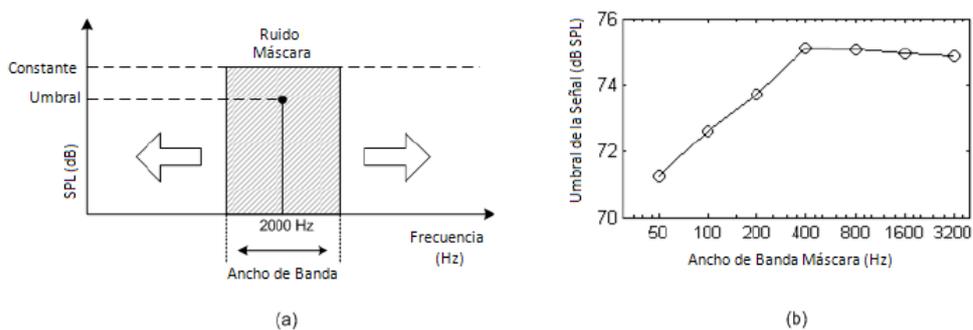
3.4.1 Bandas críticas y filtros auditivos

Como se ha dicho anteriormente, el oído realiza un tipo de análisis espectral de la señal entrante donde es descompuesta en distintas componentes de frecuencia. Como resultado, la selectividad en frecuencia del sistema auditivo periférico ha sido modelado como un banco de filtros pasa banda conocidos como filtros auditivos. Los filtros auditivos se encuentran en la membrana bacilar y poseen una resolución en frecuencia caracterizada por la escala de banda crítica. La escala de banda crítica es de aproximadamente 100 Hz para frecuencias por debajo de los 500Hz y 20% de la frecuencia central para frecuencias por encima de los 500Hz. Las bases fisiológicas de la resolución en bandas críticas de los filtros auditivos no son del todo comprendidas, pero parece ser el resultado del principio de lugar que ocurre sobre la membrana bacilar. Algunos resultados a partir de mediciones psicofísicas han mostrado que las bandas críticas corresponden a una distancia constante a lo largo de la membrana bacilar, donde una banda crítica es igual a 0.9 mm según *Moore*, 1.5 mm de acuerdo *Schroeder* y 1.3 mm de acuerdo a *Zwicker* y *Scharf*. A continuación se describen tres experimentos que han servido para caracterizar los filtros auditivos y las bandas críticas.

3.4.1.1 Experimento de Fletcher: ensanchamiento de banda

En este experimento la detección del umbral fue hecho a una señal sinusoidal (la prueba) y fue realizado bajo la presencia de ruido en la banda de paso (la máscara). Bajo estas condiciones el valor JND fue encontrado. Este proceso se repitió varias veces variando el ancho de banda de la máscara, los resultados obtenidos mostraron un aumento en la energía total de la máscara. A partir de lo obtenido, se podría asumir que grandes anchos de banda implicarían altos niveles de potencia para el enmascarado, y por lo tanto niveles más altos de JND. Este fue el caso, hasta cierto punto, ya que para cierto valor de JND este permanecería constante independientemente del ancho de banda de la máscara, ver figura 28. Fletcher propuso que el sistema auditivo funcionaba como un banco de filtros con distintas bandas de paso, y cuando se detectaba una señal se hacía uso de uno de estos de filtros, con lo que el ancho de banda en que se trabajaba era el de la banda crítica del filtro.

Figura 28. Experimento de Fletcher.



El experimento de Fletcher también llevó a modelo de potencia espectral, PSM. El PSM es un conjunto suposiciones acerca del proceso de enmascarado que se pueden resumir en las siguientes:

1. El sistema auditivo contiene un arreglo de filtros lineales pasa banda traslapados.
2. Cuando se detecta una señal en un ambiente con ruido, el escucha se asume que usa solamente un filtro centrado en una frecuencia próxima a la de la señal.
3. Solamente las componentes del ruido que pasan por este filtro tienen un efecto de enmascaramiento de la señal.
4. El umbral de enmascaramiento de la señal es determinado por la cantidad de ruido que pasa por el filtro auditivo. El umbral se asume a un valor constante de la razón señal a ruido, K , a la salida del filtro, ignorando fluctuaciones de corta duración de la máscara y/o fases relativas de la señal.

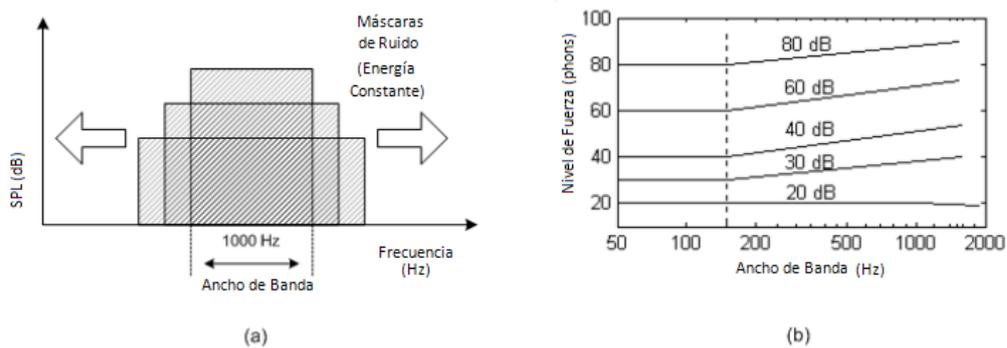
En la realidad, ninguna de estas suposiciones del modelo de potencia son estrictamente correctas. Por ejemplo, los filtros auditivos no son realmente lineales, el oído no siempre usa un solo banco para detectar una señal, el ruido que ocurre fuera de una banda de paso puede contribuir en la detección de la señal, y las fluctuaciones dependientes del tiempo en la máscara no siempre pueden ser ignoradas. Sin embargo, el concepto del banco de filtros auditivos es útil y el modelo del espectro de potencia permite una forma conveniente de calcular el umbral de enmascaramiento en el dominio de la frecuencia, como por ejemplo el uso de la función de densidad espectral en el dominio de Fourier.

3.4.1.2 Experimento de Zwicker: ruido recortado

Zwicker también realizó experimentos acerca de la escala de banda crítica. Uno de sus experimentos consistió en juntar una señal tonal con dos distintas fuentes de ruido recortado como se muestra en la figura 29(a).

El experimento involucraba el encontrar el umbral de detección de la prueba mientras se movían los ruidos recortados, y en el proceso determinar la distancia entre los dos ruidos recortados que causaría que los niveles de umbral cambiaran. Se encontró que lo niveles del umbral se mantendrían iguales siempre y cuando la separación se mantuviera sobre cierto ancho de banda crítico. Pero cuando se separaban más de eso, el valor del umbral comenzaba a decaer rápidamente como se ve en la figura 29(b).

Figura 29. Experimento de Zwicker.



Usando el modelo de potencia espectral, el cuarto enunciado dice que un valor fijo de la energía de enmascarado resultará en un valor fijo del nivel de la máscara. Esto es lo que sucede siempre y cuando el ruido recortado se mantenga en una banda crítica. Pero cuando la energía de la máscara (que llega al filtro) decrece, el nivel de enmascarado también decrece de acuerdo al radio K. Esto es lo que sucedió cuando el ruido fue separado más de la banda crítica. Con este procedimiento, Zwicker realizó mediciones del ancho de banda crítico de un amplio rango de frecuencias y llegó a una expresión para la escala de banda crítica dada por

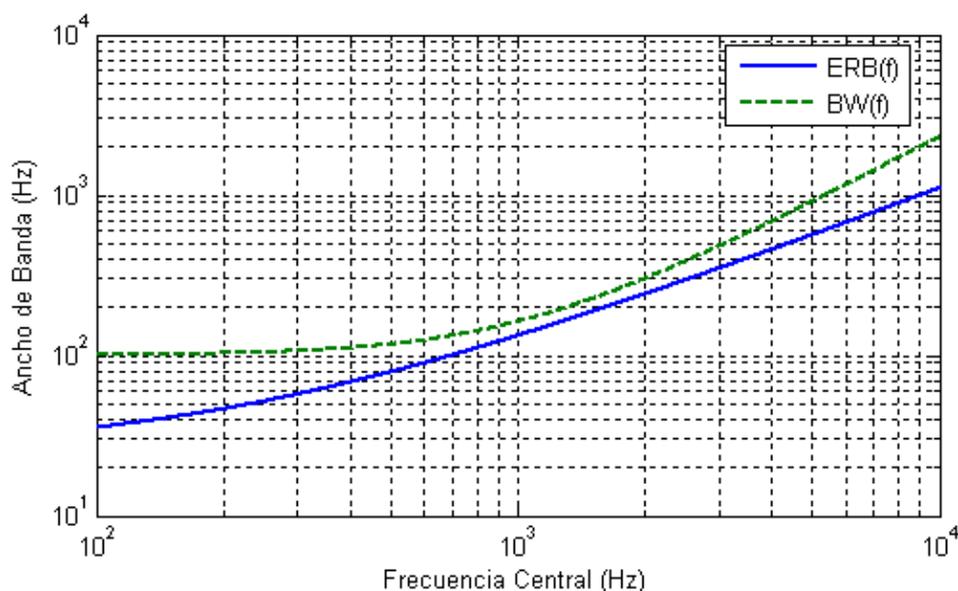
$$BW_c(f) = 25 + 75 \left[1 + 1.4 \left(\frac{f}{1000} \right)^2 \right]^{0.69} \quad (3-4)$$

A partir de las suposiciones de una K constante y que el ruido fuera simétrico alrededor del tono enmascarado, se puede manipular la separación del ruido recortado, δf , y encontrar los cambios correspondientes en P_s con la finalidad de derivar la forma de $W(f)$ a partir de (3-5). De algunas formas estimadas de filtros, la banda crítica fue medida para distintas frecuencias y los resultados fueron ajustados por la siguiente curva

$$ERB(f) = 24.7 \left[4.37 \left(\frac{f}{1000} \right) + 1 \right] \quad (3-6)$$

Donde ERB representa el ancho de banda rectangular equivalente. En la figura 31 se muestra la gráfica para $BW_c(f)$ y $ERB(f)$. Se puede notar la pequeña diferencia existente cuando $BW_c(f)$ permanece casi plano por debajo de los 500 Hz, mientras que $ERB(f)$ sigue decreciendo por debajo de los 500 Hz.

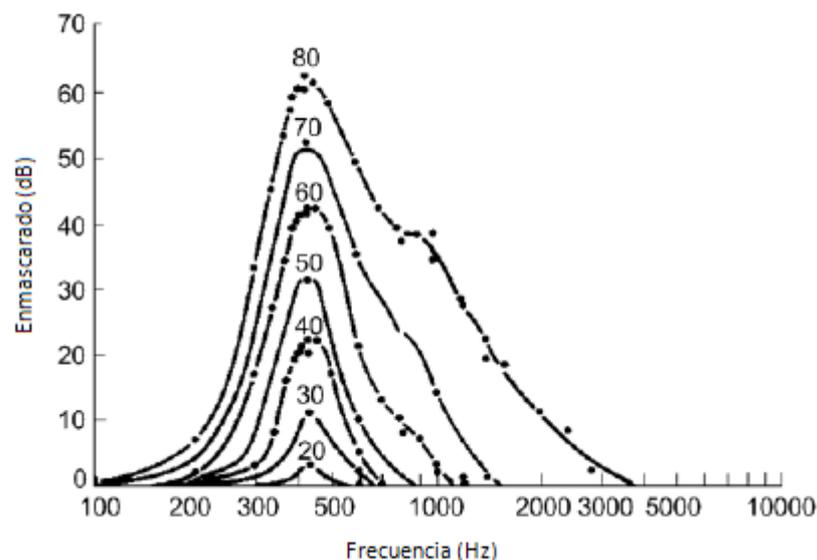
Figura 31. Bandas críticas de Zwicker y Moor.



3.4.2 Patrones de enmascaramiento y patrones de excitación

Hasta ahora el enmascarado descrito representa un enmascarado que la señal de prueba experimenta a una frecuencia fija, mientras que la frecuencia de la máscara es variada. Si se deja fija la frecuencia de la máscara y se varía la frecuencia de la señal de prueba, se obtiene lo que se conoce como patrón de enmascaramiento o audiogramas enmascarados. De hecho, esto fue lo que muchos experimentos tempranos hicieron para caracterizar la dispersión del enmascarado. En la figura 32 se muestra un ejemplo de los patrones de enmascaramiento para una máscara de ruido de banda estrecha y una señal tonal de prueba.

Figura 32. Patrón de enmascaramiento de ruido de banda estrecha centrado en 410 Hz.



Los patrones de enmascarado representan tres niveles de intensidad de la máscara centrada en 410 Hz. Otras curvas como estas han sido medidas para una variedad de máscaras a diferentes frecuencias y con distintas propiedades tonales.

Nótese que los patrones de enmascaramiento difieren de los filtros auditivos en que un patrón de enmascaramiento describe como una señal como máscara aumenta el nivel de umbral de enmascaramiento, mientras que los filtros auditivos representan la resolución con la cual el oído experimenta este enmascaramiento.

Se ha encontrado que los patrones de enmascaramiento exhiben varias no linealidades y propiedades dependientes de la señal que reflejan los procesos activos y complejos del oído. Específicamente los patrones de enmascaramiento no son simétricos en la frecuencia, son no lineales respecto a la intensidad de la máscara y la frecuencia central, dependientes de las cualidades tonales de la señal, y tampoco son aditivas. Como resultado, el enmascaramiento que resulta de una señal típica de entrada que contiene ambas componentes, tonal y parecidas al ruido, a varias frecuencias y con varias intensidades no es simplemente igual a la suma de las componentes individuales de enmascaramiento de la señal.

La determinación de una adecuada representación del enmascaramiento que resulte para las señales de audio más simples es entonces una tarea muy compleja y difícil. En un modelo de enmascaramiento simple, sin embargo, varias suposiciones necesitan ser hechas para que su cálculo pueda ser simplificado. Una de estas suposiciones es el de aproximar la forma del patrón de enmascaramiento con una sola función prototipo llamada la función de dispersión dada como

$$SF_{dB}(x) = 15.81 + 7.5(x + 0.474) - 17.5\sqrt{1 + (x + 0.474)^2} \quad (3-7)$$

Donde x es la distancia en la escala de *Bark*. La unidad *Bark* corresponde al ancho de una banda crítica. Nótese que la función de dispersión no toma en cuenta la tonalidad, la no linealidad, o la frecuencia central de la máscara.

Se ha sugerido que los patrones de enmascaramiento derivados de los experimentos descritos arriba reflejan una actividad más básica y sutil que el proceso de escuchar, siendo los denominados patrones de excitación. Un patrón de excitación es esencialmente una curva que representa el patrón de actividad neural en la membrana basilar cuando hay una señal de entrada. Es usado como un indicador crudo para lo que el oído siente opuesto a lo que el oído recibe. Se ha sugerido que el patrón de enmascaramiento y el patrón de excitación son aproximadamente paralelos en sus formas y se encuentran separados por una pequeña distancia. Más aún, los patrones de excitación han sido encontrados útiles en predecir el enmascaramiento en señales más complejas cuando múltiples máscaras y componentes enmascaradas se encuentran presentes. B. Moore establece una conexión con los filtros auditivos y los patrones de excitación, donde desarrolla un procedimiento para derivar el patrón de excitación a partir del filtro de audición.

3.4.3 Asimetría del enmascaramiento

En la exposición anterior se aludieron algunos elementos importantes respecto de la máscara y el enmascarado como son las características tonales o similares al ruido que poseen las mismas. Las diferencias de enmascaramiento de señales tonales, por ejemplo una señal sinusoidal, y las similares al ruido, por ejemplo ruido en la banda de paso, se conoce como la asimetría del enmascaramiento. Varios investigadores han estudiado la naturaleza asimétrica del enmascaramiento y han presentado una variedad de resultados útiles, algunos de los cuales se describen a continuación.

Como se está trabajando con dos tipos de señales con dos tipos distintos de roles, es conveniente distinguir cuatro tipos de escenarios de enmascaramiento, los cuales son: ruido enmascarando tono, tono enmascarando tono, tono enmascarando ruido, y ruido enmascarando ruido.

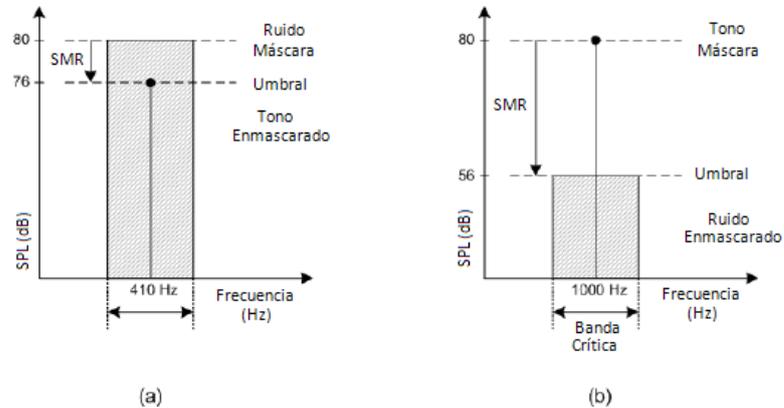
Ruido Enmascarando Tono (RET):

La figura 33(a) muestra un experimento de este tipo donde el ruido de banda estrecha centrado en 410 Hz representa a la máscara y el tono a 410 Hz representa al enmascarado. Se ha encontrado que el valor JND de la prueba se encuentra directamente relacionado con el nivel de la máscara y la forma tiende a ser similar la función de distribución de (3-7). Para la máscara y el enmascarado de la figura 33(a) el valor mínimo de la relación señal a máscara, SMR, en el punto donde se da el máximo enmascaramiento, fue de 4 dB para una máscara de 80 dB, y de 3 dB para una máscara de 60 dB.

Tono Enmascarando Tono (TET):

En el caso donde la máscara como el enmascarado son señales tonales, se ha encontrado que el enmascarado es mayor para frecuencias de prueba ligeramente mayores y ligeramente menores que la frecuencia de la máscara. Para una señal de máscara de 400 Hz, el valor mínimo de SMR fue de 19 dB para una máscara de 80 dB, de 15 dB para una máscara de 60 dB, y de 14 dB para una máscara de 40dB. Cuando la frecuencia de máscara y la prueba fueron cercanas la una de la otra, el enmascarado fue interrumpido por un fenómeno llamado *beating* donde la máscara y el enmascarado interactúan para crear una sensación fluctuante de volumen. Cuando las señales se encontraban muy lejanas entre sí se dieron otros efectos de no linealidad, hacen de los análisis más complicados de llevar.

Figura 33. Asimetría del enmascarado.



Tono Enmascarando Ruido (TER)

En este tipo de experimentos una señal tonal es la máscara y otra señal tonal es el enmascarado. En la figura 33(b) se muestra un tono enmascarador centrado a 1000 Hz y ruido de banda estrecha también centrado a 1000 Hz. De nuevo, la curva JND se ha encontrado que es dependiente del nivel y la frecuencia de la máscara y puede ser aproximada por medio de la función de dispersión que se presenta en (3-7). Pero comparado al escenario de RET, los niveles de enmascaramiento se han encontrado que son menores debido a las propiedades inferiores de enmascaramiento de las señales tonales. El valor mínimo de SMR para las señales de la figura 33(b) fue de 21 dB para una máscara de 60 dB, 24 dB para una máscara de 80 dB, y de 28 dB para una máscara de 90 dB.

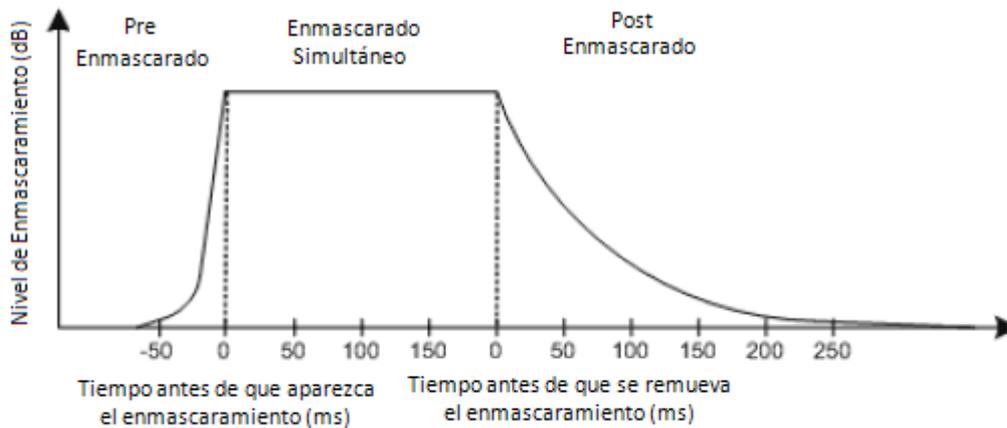
Ruido Enmascarando Ruido (RER):

Medir el enmascarado de una señal de ruido con otra señal de ruido es difícil de medir debido a las relaciones complejas de fase que se desarrollan entre la máscara y el enmascarado. Estudios realizados acerca de este enmascaramiento encontraron valores mínimos de SMR que estaban en los 26 dB.

3.4.4 Enmascaramiento temporal

Los conceptos de enmascaramiento descritos hasta ahora han sido presentados únicamente en el dominio de la frecuencia, también conocido como enmascaramiento simultáneo. Por otra parte, el enmascaramiento temporal o enmascaramiento no simultáneo es el enmascaramiento que ocurre en el dominio del tiempo antes de que comience el enmascarado y luego que este se remueve como se muestra en la figura 34.

Figura 34. Enmascaramiento no simultáneo.



Este enmascaramiento consiste del pre enmascarado o enmascarado hacia atrás, y del post enmascarado o enmascarado hacia adelante. Lo que esto significa es que el oído comienza a experimentar un enmascaramiento un poco antes de que la señal de enmascaramiento aparezca y sigue experimentándolo un tiempo después de que la señal desaparece.

Aunque existen una variedad de estudios acerca del pre enmascaramiento, no se ha sugerido una adecuada teoría de que lo causa o cómo puede ser predicho, más aún parece haber una discrepancia acerca de varios resultados. Por ejemplo, los valores medidos se encuentran entre 2 y 20 ms, y se sugiere que estos dependen en gran medida del entrenamiento que tenga el escucha.

4. PROCESADO DIGITAL DE AUDIO

El procesamiento de señales es cada vez más común en cualquier tipo de aplicación y cuando se trabaja con señales de audio esto puede llegar a mejorar muchos factores de la misma; como lo son la calidad de sonido, reducción de ruido, ecualización y otras muchas más. Además de esto el procesamiento digital ha permitido un desarrollo mucho más complejo de aplicaciones, muchas en tiempo real, que permiten una compresión y transmisión de las mismas sobre canales con un ancho de banda limitado. En este capítulo se pretende dar una descripción de los elementos que implican el procesamiento de una señal de audio. Se comienza por describir como una señal continua es convertida a una señal discreta y viceversa, y lo que esto implica. Posteriormente se presentan los filtros digitales y sus distintas características, también se presenta el análisis multi-resolución como una herramienta alternativa y de creciente aplicación. En particular, el análisis multi-resolución será de especial utilidad cuando se trabaje con la transformada wavelet.

4.1 ACD y DAC

La mayor parte de las señales con las que nos encontramos son señales continuas y como se explicó en el apartado anterior, el sistema auditivo humano representa estas señales en pulsos eléctricos que finalmente van a nuestro cerebro. La Conversión Analógica a Digital (*ADC*, por sus siglas en inglés) y la Conversión Digital a Analógica (*DAC*) son procesos que permiten a las computadoras el interactuar con estas señales.

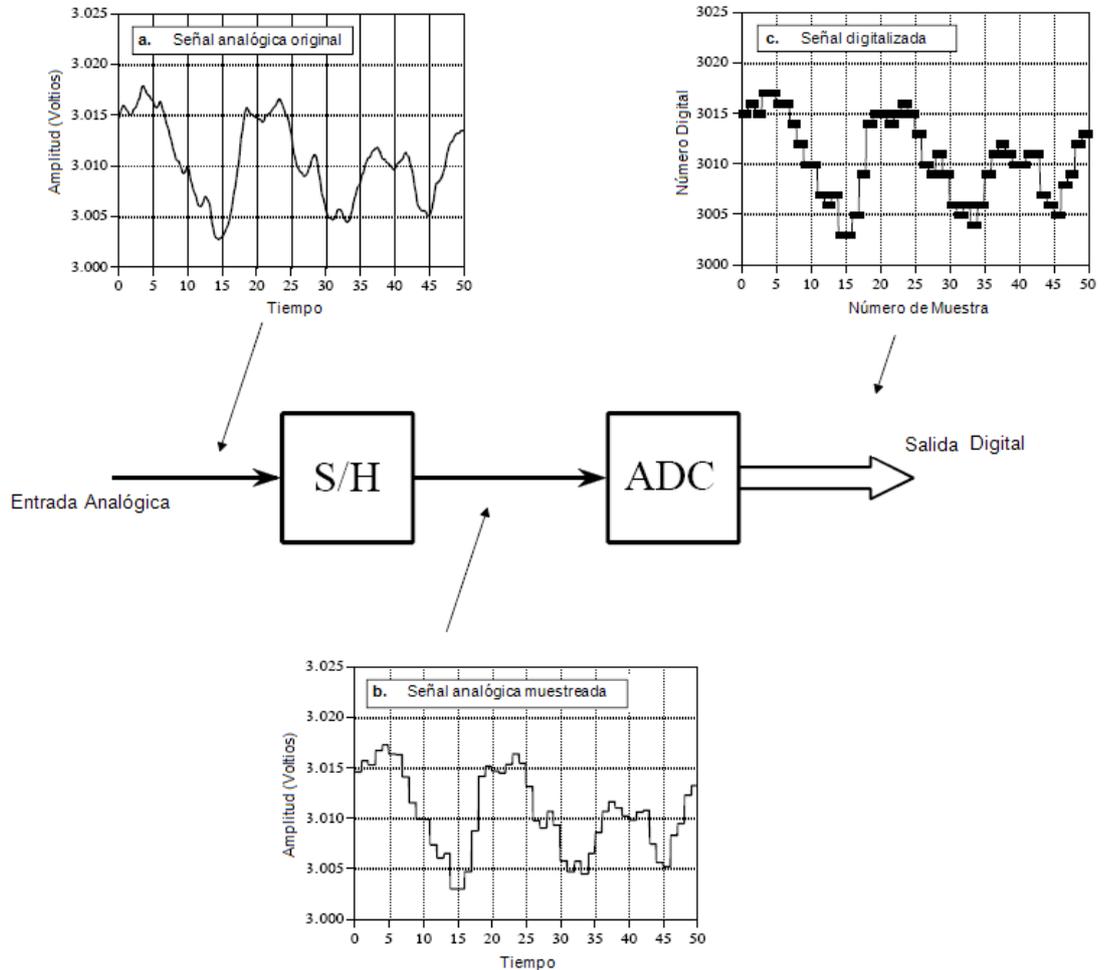
La información digital es diferente de su contraparte continua en dos aspectos importantes: su muestreo y su cuantización. Estos dos conceptos restringen cuanta información puede contener una señal digital. En este apartado se pretende mostrar el manejo de información: entender que información se debe retener, y que información se puede dar el lujo de perder. A cambio, esto dicta la selección de la frecuencia de muestreo, el número de bits, y el tipo de filtrado.

4.1.1 Cuantización

En la figura 35 se muestra la forma de onda de una conversión analógica a digital. En la figura 35 (a) se muestra la señal analógica a ser digitalizada, siendo una señal de voltaje que varía con el tiempo. Se asumirá que el voltaje variará de 0 a 4.095 volts, correspondiendo a los valores entre 0 y 4095 que produciría un digitalizador de 12 bits. Puede verse que el diagrama de bloques se encuentra partido en dos secciones, el muestreo y retención (*S/H, Sample and Hold*), y el convertidor analógico a digital (ADC).

El circuito de muestreo y retención se requiere para mantener el voltaje que entra al ADC constante mientras la conversión se lleva a cabo. Sin embargo, no es esta la razón por la que se muestra aquí: el dividir la digitalización en dos etapas es un modelo teórico importante que conlleva el entendimiento de la digitalización. El hecho de que se vea como un elemento electrónico es un bono adicional. Como se muestra por la diferencia entre (a) y (b), la salida en el circuito de muestreo y retención es permitida únicamente a intervalos periódicos, en el cual la salida es idéntica a valor instantáneo de la señal de entrada. Variaciones que ocurran en la señal de entrada fuera de los tiempos de muestreo son completamente ignorados. Esto es, el muestreo convierte la variable independiente (el tiempo en este caso) de continua a discreta.

Figura 35. Formas de onda de la digitalización de una señal.

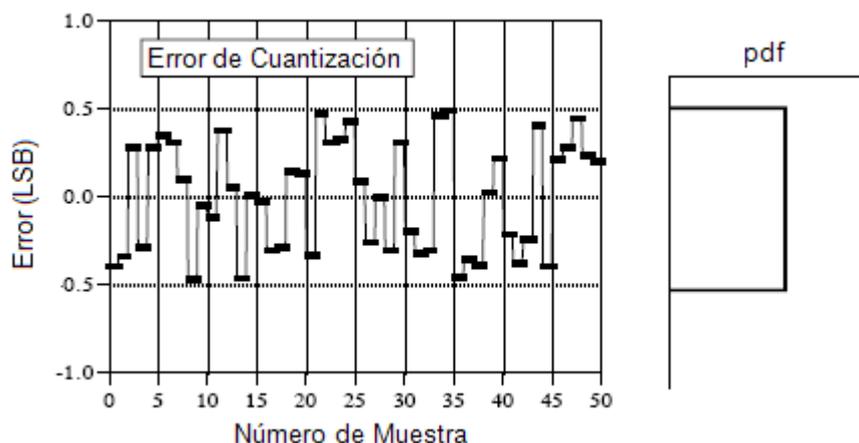


Como se muestra por las diferencias entre (b) y (c), el ADC produce un valor entero que se encuentra entre 0 y 4095 para cada una de las regiones planas en (b). Esto introduce a un error, ya que cada voltaje puede obtener un valor entre 0 y 4.095 V. Por ejemplo, tanto 2.56000 V y 2.56001 V serán convertidos en el valor digital de 2560. En otras palabras, la cuantización convierte la variable dependiente de continua a discreta.

Se ha evitado comparar (a) y (c) porque esto englobaría el muestreo y la cuantización. Es importante que se analicen de forma separada porque degradan la señal en distinta manera, así como también son controlados por distintos parámetros electrónicos. Existen además casos donde uno es usado sin el otro. Por ejemplo, el muestreo sin cuantización es usado en filtros capacitivos con switcheo.

Primero se considera los efectos de la cuantización. Cualquier muestra en una señal digitalizada puede tener un máximo de error de $\pm\frac{1}{2}$ LSB, *Least Significant Bit*, que es la distancia entre dos niveles de cuantización adyacentes. En la figura 36 se puede apreciar el error del ejemplo anterior, que se obtiene al substrae la señal muestreada y retenida con la señal digitalizada. Un factor importante de análisis es que el error de cuantización es muy parecido al ruido aleatorio.

Figura 36. Error de cuantización.



Esto permite que se desarrolle un importante modelo del error de cuantización. En muchos casos, la cuantización resulta en nada más que la adición de una cantidad de ruido aleatorio a la señal. El ruido aditivo es uniformemente distribuido entre $\pm\frac{1}{2}$ LSB, posee una media de cero, y una desviación estándar de $1/\sqrt{12}$ LSB (~ 0.29 LSB). Por ejemplo, al pasar una señal analógica a través de un digitalizador de 8 bits proporciona un ruido RMS de: $0.29/256$, o aproximadamente $1/900$ del valor máximo de escala. Una conversión de 12 bits agrega un ruido de: $0.29/4096 \approx 1/14000$, mientras que una conversión de 16 bits agrega: $0.29/65536 \approx 1/227000$. Ya que el error de cuantización es ruido aleatorio, el número de bits determina la precisión de los datos.

Este modelo proporciona bastante ayuda, ya que el ruido aleatorio generado por la cuantización solamente se agregará al cualquier tipo de ruido presente en la señal analógica. Por ejemplo, si se toma una señal con un máximo de amplitud de 1.0 V, y ruido aleatorio de 1.0 mV RMS. Digitalizar esta señal a 8 bits resulta convertir 1.0 V al número digital 255, y 1.0 mV en 0.255 LSB. El ruido aleatorio se combina al sumar sus varianzas. Esto es, las señales son sumadas en cuadratura, por lo que el ruido total de la señal digitalizada es: $\sqrt{0.255^2 + 0.29^2} = 0.386$ LSB. Siendo un aumento de casi el 50% del ruido originalmente presente en la señal analógica.

Si se utilizaran 12 bits, se tendría un ruido aleatorio de 1.0 mV equivalente a 4.095 LSB. Combinando el ruido aleatorio y el de cuantización, $\sqrt{4.095^2 + 0.29^2} = 4.105$ LSB prácticamente no habrá un incremento en el ruido y no se perderá nada por la cuantización. Cuando se debe tomar la decisión de cuántos bits son necesarios en el sistema, se deben hacer dos preguntas: primero, se debe de determinar cuánto ruido se encuentra presente en la señal análoga; y segundo, cuánto ruido puede ser tolerado en la señal digital.

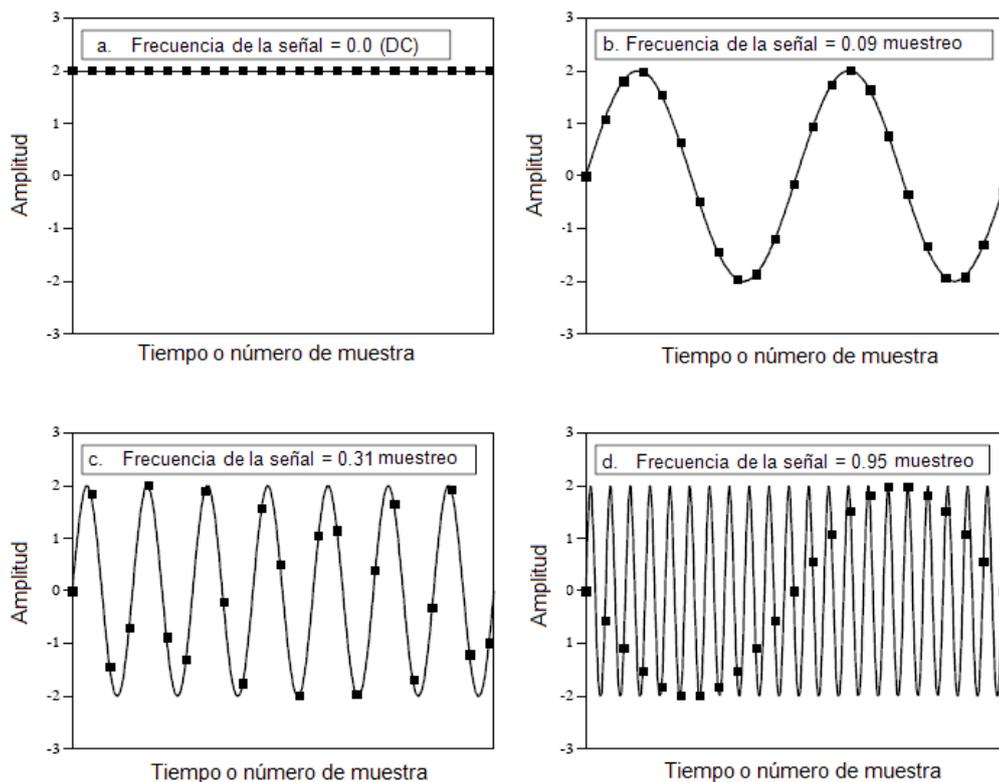
Este modelo no es válido cuando el error de cuantización no puede ser tomado como aleatorio. Esto ocurre comúnmente cuando la señal analógica se mantiene en el mismo valor por muchas muestras consecutivas. La salida se mantiene retenida o “pegada” en el mismo valor digital por muchas muestras continuas, aunque la señal pueda estar cambiando entre $\pm\frac{1}{2}$ LSB. En lugar de existir un ruido aleatorio aditivo, el error de cuantización se parece más a un efecto de determinación de umbral o a una distorsión extraña.

Una técnica común para mejorar la digitalización de estas señales que varían lentamente es conocida como difuminado. Esta consiste en agregar una cantidad pequeña de ruido a la señal analógica, con el fin de que el ruido agregado permita variaciones aleatorias entre valores adyacentes. Para clarificar esto un poco se puede imaginar el caso en que una señal análoga es un voltaje constante de 2.0001 V, y donde las variaciones de los niveles 2000 y 2001 son de un décimo de la señal analógica. Sin la técnica de difuminado, el tomar 10000 muestras de la señal producirán 10000 números idénticos, todas con el valor de 2000. A continuación, se hace el mismo experimento con la adición de una pequeña cantidad de ruido. Las 10000 muestras ahora variarán entre dos niveles, con aproximadamente 90% de ellas tomando un valor de 2000, y un 10% tomando un valor de 2001. Tomando el valor promedio de todas las muestras resulta en un valor cercano de 2000.1. Aunque la medición de un solo valor posee la limitación de $\pm\frac{1}{2}$ LSB, la estadística de un número grande de muestras puede ser mucho mejor. En resumen, el agregar ruido provee mucha más información.

4.1.2 Teorema de muestreo

El principio de un muestreo adecuado es simple. Suponiendo que se muestre la señal de alguna manera, si la señal analógica puede ser reconstruida de forma exacta por medio de las muestras, se habrá hecho un muestreo adecuado. Aunque parezca que la información muestreada es insuficiente o confusa, se establece que la información importante ha sido adquirida si se puede revertir el proceso.

Figura 37. Muestreo de señales.



En la figura 37 se pueden apreciar distintas sinusoidales antes y después de su digitalización. La línea continua representa la señal analógica entrando en el ADC, mientras que las marcas cuadradas son las señales digitales que abandonan el ADC. En(a), la señal analógica es un valor constante de DC, una onda cosenoidal de frecuencia cero. Ya que la señal analógica es una serie de líneas rectas entre cada una de las muestras, toda la información necesaria para reconstruir la señal analógica se encuentra contenida en todos los datos digitales. Por lo que es muestreada de forma adecuada.

La onda senoidal mostrada en (b) posee una frecuencia de 0.09 la frecuencia de muestreo. Esto podría representar, por ejemplo, una onda senoidal de 90 ciclos por segundo muestreada a 1000 muestras por segundo. Dicho de otra manera, se toman 11.1 muestras sobre un ciclo completo de la sinusoidal. Esta situación es más complicada que el caso anterior, porque la señal analógica no puede ser reconstruida solamente al trazar líneas continuas entre los puntos donde hay datos. Entonces, ¿estos datos representan adecuadamente la señal analógica? La respuesta es sí, esto porque no habrá otra sinusoidal, o combinación de sinusoidales, que produzca este patrón de muestras. Estas muestras corresponden a una sola señal analógica, y entonces la señal analógica puede ser reconstruida de forma exacta. De igual manera, se encuentra correctamente muestreada.

En (c), la situación es un poco más compleja, ya que se aumenta la frecuencia de la señal sinusoidal a 0.31 la frecuencia de muestreo. Esto resulta en solamente 3.2 muestras por ciclo de la señal sinusoidal. Aquí las muestras son tan escasas que parecen no seguir la forma general de la señal analógica. De igual manera que en el caso anterior estas muestras representan a la señal analógica. Las muestras son una representación única de la señal analógica. Por lo que se encuentra muestreada adecuadamente.

En (d), la frecuencia analógica es llevada a 0.95 el período de muestreo, con 1.05 muestras por ciclo de la sinusoidal. Ahora, estas muestras no representan de forma adecuada los datos. Estas muestras representan una señal sinusoidal diferente que la contenida en la señal analógica. En particular, la señal original de frecuencia 0.95 se falsifica a sí misma con una señal sinusoidal de frecuencia 0.05 en la señal digital. Este fenómeno de sinusoidales cambiando frecuencia durante el muestreo se conoce como *aliasing*.

Como los datos digitales no se encuentran relacionados de forma única, una reconstrucción ambigua es posible. No existe nada que sugiera que la señal analógica original era de 0.95 y no de 0.05. Este es un ejemplo de un muestreo inadecuado.

Esto conlleva a punto fundamental en el DSP, el teorema de muestreo. Frecuentemente se denomina el teorema de muestreo de Shannon, o el teorema de muestreo de Nyquist. El teorema de muestreo indica que una señal continua puede ser muestreada de forma correcta, solo si no contiene componentes espectrales de frecuencia por encima de un medio de la frecuencia de muestreo. Por ejemplo, una tasa de muestreo de 2000 muestras por segundo requiere que la señal analógica este compuesta de frecuencias por debajo de 1000 ciclos por segundo. Si frecuencias por encima de este límite se encuentran presentes se encuentran presentes en la señal, estas serán encubiertas con frecuencias entre 0 y 1000 ciclos por segundo, combinando con cualquiera información legítima que se encuentre ahí.

En la figura 38 se muestra como las frecuencias son cambiadas durante el *aliasing*. El punto clave a entender es que la señal digital no puede contener frecuencias por encima de un medio de la frecuencia de Nyquist.

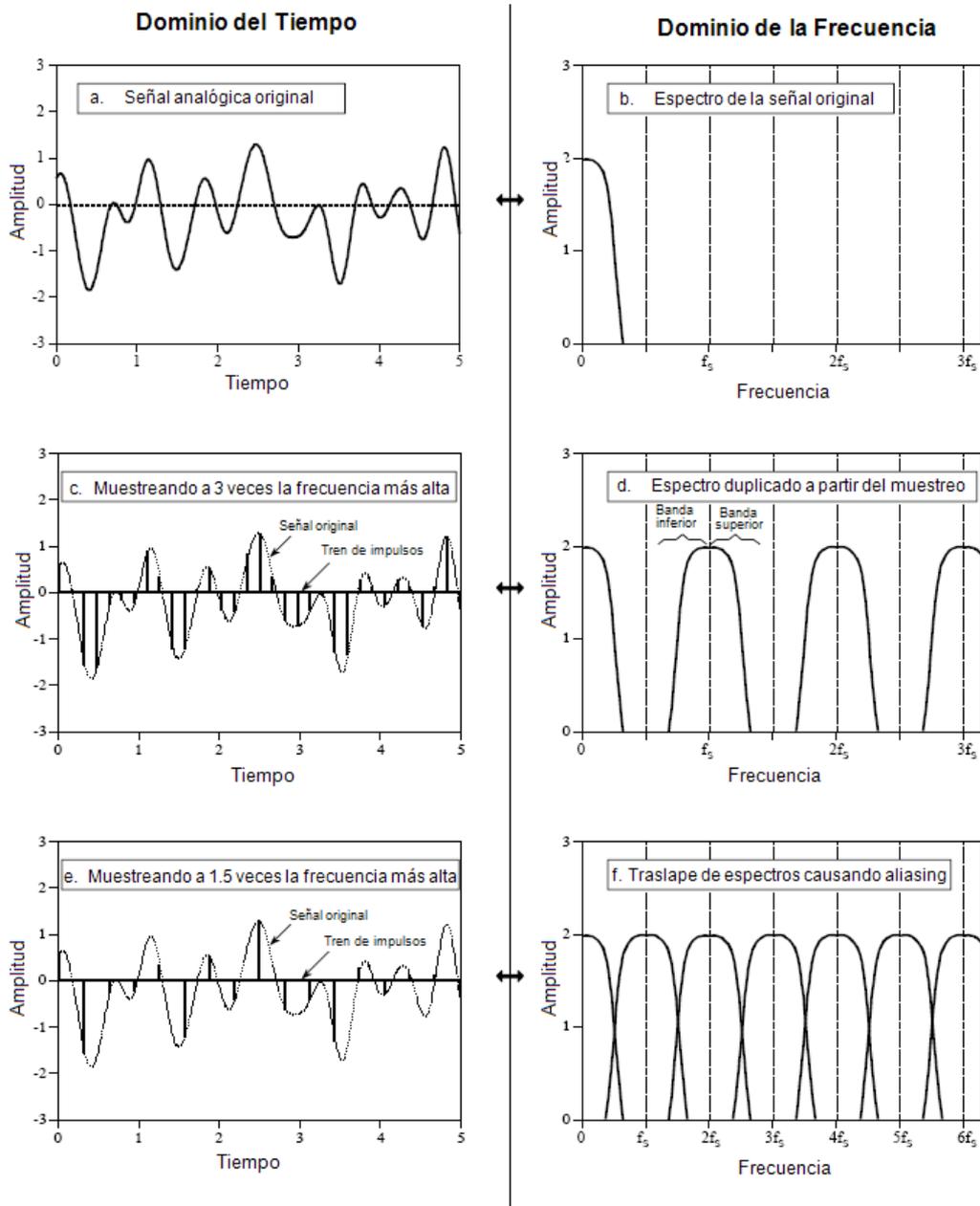
El efecto de *aliasing* es un problema doble: la información puede ser perdida en las bajas y altas frecuencias. Suponiendo que se tiene una señal digital que posee una frecuencia de 0.2 la frecuencia de muestreo. Si esta señal hubiese sido obtenida por un muestreo adecuado, la señal analógica original debió de tener una frecuencia de 0.2. Si existió *aliasing* durante el muestreo, la frecuencia de la señal digital pudo haberse originado por un número infinito de frecuencias en la señal analógica: 0.2, 0.8, 1.2, 1.8, 2.2, etc.

Así como el efecto de *aliasing* puede cambiar la frecuencia durante el muestreo, también puede cambiar la fase. Por ejemplo, si se observa la figura 37 (d), se puede observar que la señal se encuentra invertida de la señal original. Solamente dos cambios de fase son posibles: 0° (no hay cambio de fase) y 180° (inversión). El cambio de fase cero ocurre para frecuencias analógicas de 0 a 0.5, 1.0 a 1.5, 2.0 a 2.5, etc. Una inversión de fase ocurre para frecuencias de 0.5 a 1.0, 1.5 a 2.0, 2.5 a 3.0, y así sucesivamente.

Ahora se pretenderá explicar qué es lo que pasa con la información cuando una señal continua es convertida en discreta. El problema en esto son dos cosas completamente distintas; una es una forma de onda continua mientras que la otra es un arreglo de números. De ahí que se introduce el concepto del tren de impulsos.

En la figura 39 se muestra un ejemplo de una señal analógica. En (c) se ve la señal muestreada usando un tren de impulsos. El tren de impulsos es una señal continua que consiste en una serie de pequeños pulsos que se ajustan a la señal original en los instantes de muestreo. Cada impulso es infinitesimalmente estrecho. Fuera de los tiempos de muestre el valor de la forma de onda es cero. Se debe recordar que el tren de impulsos es un concepto teórico, no una forma de onda que puede existir en un circuito electrónico. Debido a que la señal analógica y el tren de impulsos son formas de onda continuas, se puede hacer una comparación entre estas dos.

Figura 39. El teorema de muestreo en el dominio del tiempo y la frecuencia.



Entender la relación entre el tren de impulsos y la señal discreta es sencillo; en términos del contenido de información, estas son idénticas. Si una es conocida, es trivial el calcular la otra. Se puede considerar que se ha alcanzado nuestro fin último si se entienden las consecuencias de cambiar la forma de onda en la figura 39 (a) en la forma de onda de la figura 39 (c).

Muestrear la señal en (a) por medio de un tren de impulsos produce la señal en (c), y su espectro de frecuencia mostrado en (d). Este espectro es una duplicación del espectro de la señal original. Cada múltiplo de la frecuencia de muestreo, f_s , $2f_s$, $3f_s$, $4f_s$, etc., ha recibido una copia y un reflejo de izquierda a derecha del espectro de la señal original. La copia se llama banda lateral superior, mientras que la otra copia se denomina banda lateral inferior.

El muestreo ha generado nuevas frecuencias, y aunque ha sucedido esto, se puede decir que se ha muestreado de forma correcta. Esto porque la señal en (c) puede ser transformada de nuevo en la señal en (a) al eliminar todas las frecuencias por encima de $\frac{1}{2}f_s$. Esto es, un filtro analógico paso bajo convertirá el tren de impulsos, (b), de nuevo en la señal analógica original, (a).

En el dominio del tiempo, el muestreo es logrado al multiplicar la señal original con el tren de impulsos con amplitud unitaria. El espectro de frecuencia de este tren de impulsos unitario es también un tren de impulsos unitarios, con pulsos que ocurren a múltiplos de la frecuencia de muestreo. Cuando dos señales son multiplicadas en el dominio del tiempo, su espectro de frecuencia es convolucionado. Esto resulta en una duplicación del espectro de frecuencia en la ubicación de cada pulso del espectro del tren de impulsos. Al ver la señal original como una composición de frecuencias positivas y negativas lleva al caso de las bandas laterales superior e inferior, respectivamente.

En la figura 39 (e) se muestra un ejemplo de un muestreo inadecuado, resultado de una tasa de muestreo muy baja. La señal analógica aún posee frecuencias mayores a 3.3 kHz, pero la tasa de muestreo ha sido bajada a 5 kHz. Se puede ver que f_s , $2f_s$, $3f_s$, $4f_s$, etc., poseen un espaciamiento menor en (f) que en (d). El espectro de frecuencia, (f), muestra el problema: la porción duplicada del espectro ha invadido la banda entre cero y un medio de la frecuencia de muestreo.

Aunque (f) muestra que estas frecuencias han mantenido su identidad, en la realidad ellas son sumadas y forman una sola distribución. Como no existe una forma de separar el traslape de frecuencias, se pierde información, y la señal original no puede ser reconstruida. Este traslape ocurre cuando la señal analógica contiene frecuencias mayores que un medio de la frecuencia de muestreo, con lo que, se ha probado el teorema de muestreo.

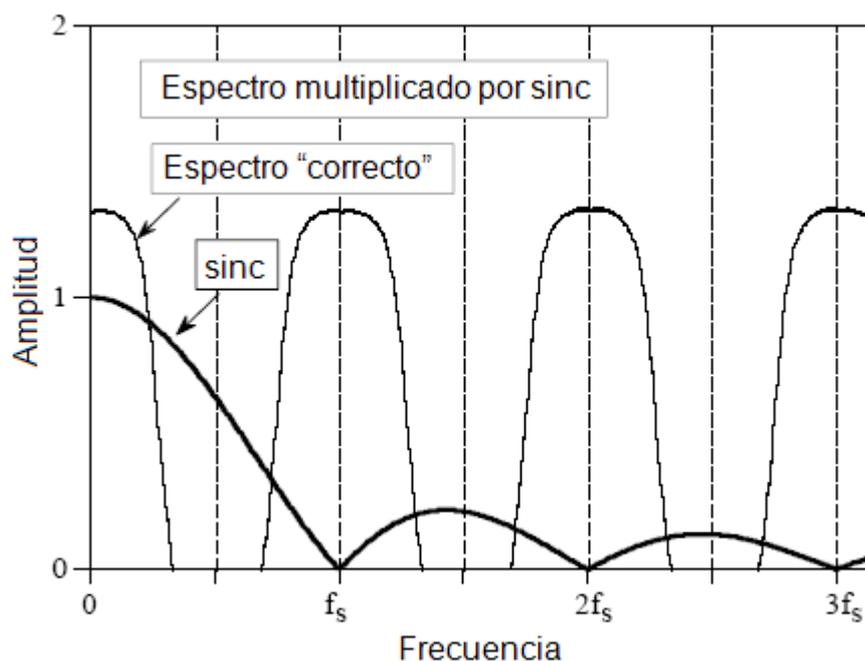
4.1.3 Conversión digital a analógico

En teoría, la manera más simple para convertir una señal digital a analógica consiste en sacar las muestras almacenadas en memoria y convertirlas en un tren de impulsos. Luego la señal analógica original podría ser perfectamente reconstruida al pasar el tren de impulsos por un filtro paso bajo, con su frecuencia de corte igual a un medio de la frecuencia de muestreo. En otras palabras, la señal original y el tren de impulsos poseen un espectro de frecuencia idéntico por debajo de la frecuencia de Nyquist. A frecuencias altas, el tren de impulsos contiene una duplicación de su información, mientras que la señal original no contiene nada (asumiendo que no ocurre *aliasing*).

Aunque este método es puramente matemático, es difícil generar estos pulsos estrechos en la electrónica. Casi todos los DACs retienen el último valor hasta que otra muestra es recibida. Esta técnica se denomina retención de orden cero, el equivalente de muestreo y retención usado en los ADCs. El procedimiento de retención de orden cero produce una señal con forma de escalera.

Este procedimiento puede ser entendido como la convolución de un tren de impulsos con un pulso rectangular, con un ancho igual al período de muestreo. Esto resulta en la multiplicación de la transformada de Fourier del pulso rectangular, en este caso la función *sinc*, con la transformada de Fourier del tren de impulsos, figura 40.

Figura 40. Espectro de frecuencia de filtro rectangular y señal reconstruida.



El filtro analógico usado para convertir la señal de retención de orden cero en la señal analógica deseada necesita hacer dos cosas: remover todas las frecuencias por encima de un medio de la frecuencia de muestreo, e incrementar la amplitud de las frecuencias por el recíproco del sistema de retención de orden cero, $1/\text{sinc}(x)$.

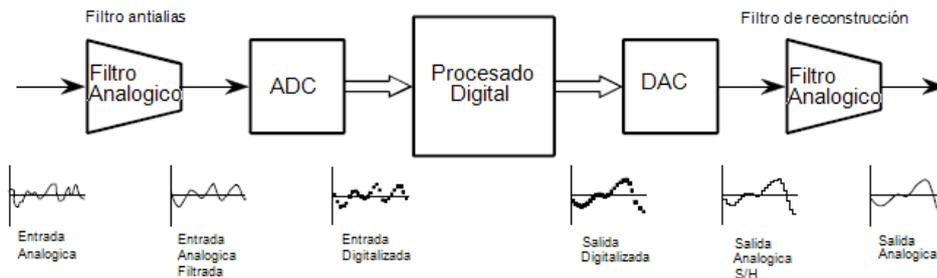
Este aumento de $1/\text{sinc}(x)$ puede ser manejado de cuatro maneras: (1) ignorarlo y aceptar las consecuencias, (2) diseñar un filtro analógico que incluya la respuesta $1/\text{sinc}(x)$, (3) usar una técnica de múltiples tasas de muestreo o (4) hacer la corrección en software antes del DAC.

4.1.4 Filtros analógicos

Un diagrama de un sistema de DSP se muestra en la figura 41, esto es, como indica el teorema de muestreo. Antes de encontrar la conversión analógico a digital la señal de entrada es procesada con un filtro pasa bajo que elimina las frecuencias por encima de la frecuencia de Nyquist. Esto se hace para prevenir lo que se conoce como *aliasing* durante el muestreo, y se conoce como un filtro *antialias*. En el otro extremo, la señal digitalizada es pasada por un convertidor digital a analógico y luego por otro filtro paso bajo ajustado a la frecuencia de Nyquist. Este filtro es llamado filtro de reconstrucción, y podría incluir el procedimiento de retención de orden cero.

Desafortunadamente, existe un problema muy grande con este modelo simple: las limitaciones de los filtros electrónicos pueden ser tan malas como los problemas que intentan resolver.

Figura 41. Diagrama de bloques de un sistema de DSP.



Tres tipos de filtros analógicos son usados comúnmente: Chebyshev, Butterworth y Bessel. Cada uno de ellos es diseñado para optimizar un parámetro de rendimiento distinto. La complejidad de cada filtro puede ser ajustada al seleccionar el número de polos y ceros. Entre más polos existan en el filtro mayor será la complejidad electrónica del mismo, así como también tendrá un mejor rendimiento. Cada uno de estos nombres describe lo que el filtro realiza, no un arreglo particular de resistencias y capacitores. Por lo que para los fines del procesamiento digital de señales son más importantes sus características que la forma en que son construidos.

Cada filtro optimiza un parámetro en particular en expensas de cualquier otro. El filtro Chebyshev optimiza el rizado consiguiendo una caída de la respuesta en frecuencia más pronunciada en frecuencias bajas; el filtro Butterworth optimiza una banda de paso plana; y el filtro de Bessel optimiza la respuesta al impulso por su linealidad.

La selección del filtro *antialias* depende en gran medida en un aspecto: cómo es presentada la información en la señal que se pretende procesar. Comúnmente se utilizan dos métodos por los cuales se codifica la señal analógica, codificación en el dominio del tiempo y la codificación en el dominio de la frecuencia. La diferencia entre una y otra es crítica en DSP.

En la codificación en el dominio de la frecuencia, la información está contenida en ondas sinusoidales que combinadas forman la señal. La digitalización de este tipo de señales involucra un filtro *antialias* que posea un corte agudo, como un filtro Chebyshev, Elíptico o Butterwoth; y aunque posean una mala respuesta al impulso, la información codificada no se ve afectada por este tipo de distorsión.

Por otra parte, la codificación en el dominio del tiempo usa la forma de la señal para codificar la información. El teorema del muestreo es de ayuda para entender como las señales codificadas en el dominio del tiempo deben de ser digitalizadas. Por ejemplo, si se deseara digitalizar la forma de una señal codificada en el dominio del tiempo, se debe tomar en cuenta que de existir alguna componente de alta frecuencia presente en la señal (debido a ruido o a la misma forma de onda) al usar un filtro *antialiasing*, o al no usarlo, la señal reconstruida será distinta.

La razón por la cual la señal original no puede ser reconstruida por las muestras obtenidas radica en que la señal original contiene componentes de frecuencia mayores que un medio de la frecuencia de muestreo. Como estas frecuencias no pueden existir en la señal digitalizada, la señal reconstruida no puede contenerlas de igual manera.

El filtro de Chebyshev ataca este problema al remover de forma agresiva las componentes de alta frecuencia. Esto resulta en una señal analógica que puede ser muestreada y reconstruida de forma perfecta. Sin embargo, la señal analógica reconstruida es idéntica a la señal filtrada, no a la señal original. Aunque nada se pierde en el muestreo, la forma de onda ha sido distorsionada de forma severa por el filtro *antialias*.

El filtro de Bessel está diseñado para este problema. Su salida se aproxima mucho a la señal original, dando un gentil redondeo a los bordes. Al ajustar la frecuencia de corte del filtro, la suavidad de los bordes puede ser compensada por la eliminación de las componentes de alta frecuencia de la señal. Usando más polos en el filtro permite un mejor balance entre estos dos parámetros. Un lineamiento común es el de ajustar la frecuencia de corte a un cuarto de la frecuencia de muestreo. Esto resulta en aproximadamente dos muestras sobre la porción de aumento de cada borde. Nótese que tanto el filtro de Bessel y Chebyshev han removido el ruido de alta frecuencia presente en la señal.

La última opción es el no usar un filtro *antialias*. Esto proporciona una gran ventaja ya que el valor de cada muestra es idéntico al valor de la señal analógica. La desventaja es que puede surgir *aliasing* y puede darse en dos formas distintas. Primero, interferencias y ruido de altas frecuencias se convertirán en muestras sin importancia. Segundo, al no existir información entre muestras mucha información de que ocurre o qué es lo que hay en esta señal no es conocida. Se puede decir que no hay un filtro adecuado para cada caso, solamente con un buen entendimiento de lo que se planea hacer luego de que los datos son adquiridos se puede estimar que filtro es de mayor ayuda.

4.2 Filtros digitales

Los filtros digitales son usados para dos propósitos generales: separación de las señales que han sido combinadas, y la restauración de señales que han sido distorsionadas de alguna manera. Los filtros analógicos pueden ser usados para estas mismas tareas; sin embargo, los filtros digitales pueden alcanzar resultados superiores.

La separación de señales es necesitada cuando una señal se ha contaminado con interferencia, ruido u otras señales. La restauración de señales es usada cuando esta se ha distorsionado de alguna manera. Por ejemplo, una grabación de audio hecha con un equipo de grabación pobre puede ser filtrada para representar el sonido como ocurrió realmente.

Los filtros digitales pueden alcanzar un rendimiento miles de veces mejor que un filtro analógico. Esto hace una dramática diferencia con los problemas de filtrado analizados.

Con los filtros analógicos, el énfasis se sitúa en manejar las limitaciones de la electrónica, como la precisión y la estabilidad de las resistencias y los capacitores. En comparación, los filtros digitales son tan buenos que el rendimiento del filtro es comúnmente ignorado. El análisis se sitúa en las limitaciones de las señales, y temas más teóricos acerca de su procesamiento.

Cada filtro lineal tiene una respuesta al impulso, una respuesta al escalón y una respuesta en frecuencia. Cada una de estas respuestas contiene la información completa acerca del filtro, pero de forma distinta. Si alguna de las tres es especificada, las otras dos son fijas y pueden ser calculadas de forma directa. Estas tres representaciones son importantes, porque ellas describen como el filtro reaccionará bajo distintas circunstancias.

La forma más directa de implementar un filtro digital es el de convolucionar la señal de entrada con la respuesta al impulso del filtro. Todos los filtros lineales posibles pueden ser hechos de esta manera. Cuando la respuesta al impulso es usada en esta forma, los diseñadores de filtros le denominan núcleo del filtro.

En este caso una serie de coeficientes del filtro son aplicados a las muestras, al combinar muestras anteriores con los coeficientes. Una muestra de la salida es sumada con anteriores muestras multiplicadas por los coeficientes del filtro. La muestra de entrada es multiplicada con el primer coeficiente del filtro, la anterior muestra de entrada con el segundo coeficiente del filtro, y así se prosigue. La suma de todos los productos es la muestra resultante filtrada. Como una fórmula, la convolución tiene la siguiente forma:

$$y[n] = \sum_{k=0}^{N-1} x[n-k]h[k] \quad (4-1)$$

La convolución puede ser escrita usando una notación matricial, donde las señales son escritas como vectores. Puede ser escrita como:

$$[y_n \quad y_{n+1} \quad y_{n+2} \quad \Lambda] = \begin{bmatrix} h_1 & h_0 & & \\ & h_1 & h_0 & \\ & & h_1 & h_0 \\ & & & 0 \end{bmatrix} \begin{bmatrix} x_{n-1} \\ x_n \\ x_{n+1} \\ M \end{bmatrix} \quad (4-2)$$

Los espacios vacíos en la matriz representan ceros. Solo un filtro de longitud 2 es mostrado. La matriz de convolución es una matriz circulante a la derecha, específicamente una matriz *Toeplitz*, donde cada fila se encuentra desplazada a la derecha respecto a su antecesora.

Para recuperar los coeficientes del filtro, existen una gran variedad de métodos, que resultan en filtros con diferentes propiedades. Para filtros paso bajo y paso alto, una pendiente en su banda de transición y planicie en su banda de paso y su banda de rechazo son restricciones y requerimientos importantes en su diseño.

Existe otra forma de realizar filtros digitales, llamada recursión. Cuando un filtro es implementado por convolución, cada muestra en la salida es calculado al ponderar las muestras en la entrada, y sumaras. Los filtros recursivos son una extensión de esto, usando valores previamente calculados de la salida, así como también de puntos de la entrada. En lugar de usar un núcleo de filtro, los filtros recursivos son definidos por un conjunto de coeficientes recursivos. La respuesta al impulso de un filtro recursivo se encuentra compuesta por sinusoidales que decaen exponencialmente en amplitud. En principio, esto hace sus respuestas al impulso infinitamente largas. Sin embargo, la amplitud eventualmente cae por debajo del valor mínimo para ser consideradas como ruido, y las muestras restantes pueden ser ignoradas.

Los filtros recursivos son llamados filtros IIR por sus siglas en inglés *Infinite Impulse Response*. En comparación, los filtros llevados a cabo por convolución son llamados filtros FIR (*Finite Impulse Response*). Los filtros llevados a cabo por convolución poseen un mejor rendimiento que los filtros que usan recursión, pero se ejecutan de manera más lenta.

La respuesta al escalón es la salida cuando la entrada es un escalón. Como el escalón es la integral del impulso, la respuesta al impulso es la integral de la respuesta al impulso. Esto permite dos formas de obtener la respuesta al escalón: (1) insertar una forma de onda escalón en el filtro y esperar su respuesta, o (2) integrar discretamente la respuesta al impulso. La respuesta en frecuencia puede ser encontrada al tomar la DFT (usando algún algoritmo FFT) de la respuesta al impulso.

4.2.1 Filtros FIR

4.2.1.1 Filtro de desplazamiento de promedio

Es el más común en DSP, ya que es el más fácil de entender y usar. A pesar de su simplicidad, es un filtro óptimo para una tarea común: reducir el ruido aleatorio mientras mantiene una respuesta aguda. Esto hace que sea el filtro adecuado para señales codificadas en el dominio del tiempo. Sin embargo, este filtro es el peor filtro para señales codificadas en el dominio de la frecuencia, con poca habilidad para separar una banda de frecuencias de otra. Entre la familia de estos filtros se encuentran el Gaussiano, Blackman, y el de paso múltiple de desplazamiento de promedio. Estos poseen un rendimiento un poco mejor en el dominio de la frecuencia, con el costo extra de un incremento en el tiempo de computación.

4.2.1.2 Filtro de ventana *sinc*

Estos filtros son usados para separar una banda de frecuencias de otra. Son bastante estables, producen pocas sorpresas, y pueden ser llevados a niveles de rendimiento increíbles. Estas características excepcionales en el dominio de la frecuencia son obtenidas a expensas de su pobre rendimiento en el dominio del tiempo, incluyendo un rizo excesivo y sobresaltos en la respuesta al escalón. Cuando es llevado por convolución estándar, los filtros de ventana *sinc* son fáciles de programar, pero lentos de ejecutar. Si se aplica la FFT se puede aumentar de forma dramática la velocidad de computación de estos filtros.

4.2.1.3 Filtro modificados

Muchos filtros tienen alguna de las cuatro respuestas en frecuencia: paso bajo, paso alto, paso banda o rechaza banda. Estos filtros poseen una respuesta arbitraria en frecuencia, ajustada a las necesidades de una aplicación particular. El DSP excede por mucho esta área en comparación con los filtros analógicos. Dos usos importantes se pueden dar con este tipo de filtros: deconvolución, una manera de restaurar señales que han sufrido una convolución indeseada, y filtrado óptimo, el problema de separar señales con un espectro de frecuencia que se traslapa.

4.2.1.4 Convolución FFT

Este tipo de filtro usa la técnica de suma de traslape, *overlap-add*, en el cual se parte una señal en pequeños segmentos para un procesamiento más sencillo.

Al utilizar este método junto con la transformada rápida de Fourier, se permite que las señales se convolucionen al multiplicar su espectro de frecuencia. Para núcleos de filtros mayores que 64 puntos, la convolución FFT es más rápida que la convolución estándar, mientras que produce los mismos resultados.

4.2.2 Filtros IIR

4.2.2.1 Filtros de un solo polo

La utilidad de un filtro de este tipo es la habilidad de poder usarlos para muchas aplicaciones como por ejemplo: supresión de componente de DC, supresión de ruido de altas frecuencias, moldeado de formas onda, etc. Son fáciles de programar, fáciles de ejecutar y producen pocas sorpresas.

Estos filtros poseen poca habilidad para separar una banda de frecuencias de otra, es decir que funcionan bien en el dominio del tiempo pero pobremente en el dominio de la frecuencia. Esta respuesta en frecuencia puede ser mejorada si se añaden varias etapas. Esto puede ser realizado de dos maneras. Primero, la señal puede ser pasada por el filtro varias veces. Segundo, la transformada z puede ser usada para encontrar los coeficientes de recursión que combinan la cascada en una sola etapa. Cualquiera de las dos maneras es usada comúnmente.

4.2.2.2 Filtros de banda estrecha

Una necesidad común en la electrónica y en el DSP es el de aislar una banda de frecuencias muy estrecha de un ancho de banda muy grande. Por ejemplo, si se deseara aislar los tonos de señalización de una red de telefonía. Existen dos tipos de respuesta en frecuencia: la banda de paso y el rechazo de banda (también se le conoce como *notch filter*).

4.2.2.3 Filtros de respuesta de fase

Existen tres tipos de respuesta de fase que un filtro puede tener: fase cero, fase lineal y fase no lineal. El filtro con fase cero o nula se caracteriza con una respuesta al impulso que es simétrica entrono a la muestra cero. La forma en sí no es de importancia, sino que las muestras negativas son una imagen reflejada de las muestras positivas. Cuando se toma la transformada de Fourier de esta forma de onda simétrica, la fase es totalmente cero. La desventaja de un filtro de fase cero es que requiere el uso de índices negativos, que pueden ser inconvenientes en su uso. El filtro de fase lineal es una manera de lidiar con esto. La respuesta al impulso es igual que la de uno de fase cero, excepto que ha sido desplazada para usar únicamente muestras positivas.

La respuesta al impulso aún es simétrica, pero ha sido desplazada de cero. Este desplazamiento resulta en una línea recta, que da el nombre de fase lineal. La inclinación de esta línea es directamente proporcional a la cantidad que se haya desplazado. En muchos casos el filtro de fase lineal es equivalente al filtro de fase cero.

Una respuesta al impulso que no es simétrica de izquierda a derecha posee una respuesta en fase que no es una línea recta, se dice que este filtro posee una fase no lineal.

El interés de filtros con fase cero y fase lineal es que poseen bordes de izquierda y derecha que se ven iguales, mientras que los filtros de fase no lineal poseen bordes de izquierda y derecha que se ven distintos. En muchas aplicaciones no se tolera que existan diferencias entre los bordes de izquierda y derecha ya que esta diferencia puede ser malinterpretada como una característica propia de la señal que es analizada.

4.2.2.4 Filtros Chebyshev

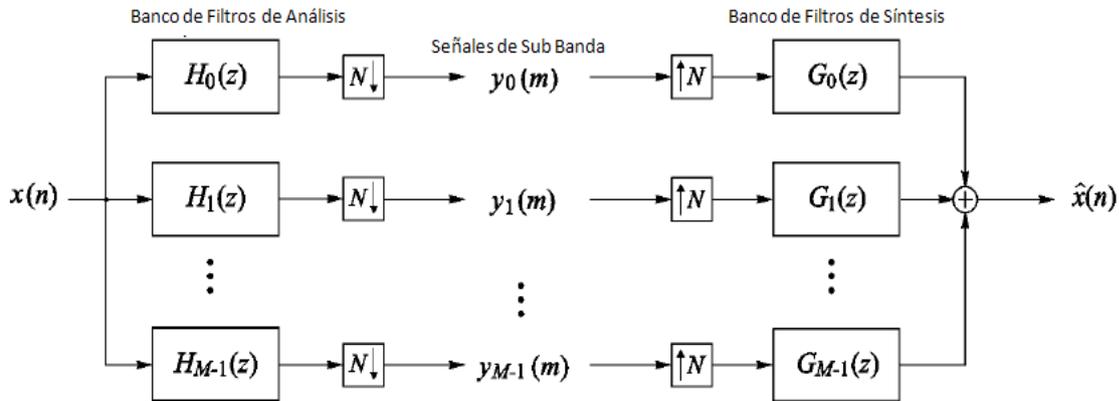
Los filtros Chebyshev son usados para separar una banda de frecuencias de otra. Aunque no pueden alcanzar el rendimiento de un filtro de ventana *sinc*, ellos pueden ser más que adecuados para muchas aplicaciones. El principal atributo de un filtro Chebyshev es su velocidad, típicamente poseen una velocidad de un orden mucho mayor que el de una ventana *sinc*. Esto es porque pueden ser generados por medio de recursiones en lugar de convoluciones. El diseño de estos filtros se basa en la aplicación de la transformada Z.

4.3 Multi-resolución

Los bancos de filtros son arreglos de filtros pasa bajo, pasa banda y pasa alto usados para la descomposición y composición espectral de señales. Estos bancos juegan un papel importante en muchas aplicaciones modernas de procesamiento digital de audio. La razón de su popularidad es el hecho de que permiten de forma sencilla la extracción de componentes espectrales de una señal con una eficiente implementación. Ya que muchos bancos de filtros involucran varias tasas de muestreo, se les conoce también como sistemas multi-resolución. Para dar un ejemplo, la figura 42 muestra un banco de filtros de M -canales.

La señal de entrada es descompuesta en M señales de sub banda aplicando M filtros de análisis con diferentes bandas de paso. Con lo que cada señal de sub banda posee información de la señal de entrada en una banda particular de frecuencias. Los bloques con flechas apuntando hacia abajo en la figura 42 indican un sub muestreo (decimación) por un factor N , y los bloques con flechas apuntando hacia arriba indican un sobre muestreo (interpolación) con un factor N . Sub muestrear por N significa que solamente la N –ésima muestra es tomada. Esta operación sirve para reducir o eliminar las redundancias en las M señales de sub banda. Sobre muestreo por N significa la inserción de $N - 1$ ceros consecutivos entre las muestras. Esto permite recuperar la tasa de muestreo original. El sobre muestreo es seguido por filtros que reemplazan los ceros con valores significativos. En el caso de $M = N$ se habla de sub muestreo crítico porque este es el factor máximo de sub muestreo para la que se puede dar una perfecta reconstrucción. La reconstrucción perfecta significa que la señal original es una copia de la señal original con no otra distorsión que una variación en el tiempo y en la escala de amplitud.

Figura 42. Banco de filtros de M-canales.



Desde un punto de vista matemático, un banco de filtros realiza una expansión en series, donde las señales de sub banda son los coeficientes, y las variaciones en el tiempo $g_k(n - iN), i \in Z$, de las respuestas al impulso $g_k(n)$ del filtro de síntesis forman la base. La principal diferencia a la transformación en bloques es que la longitud de la respuesta al impulso de los filtros es usualmente más grande que N así que las secuencias se traslapan entre sí.

4.3.1 Operaciones básicas de la multi-resolución

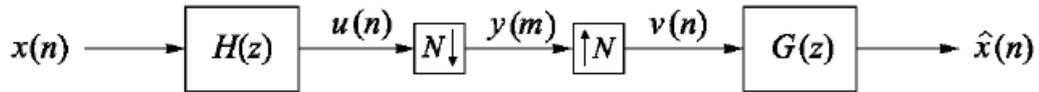
4.3.1.1 Decimación e interpolación

Las decimación y la interpolación ocurren en todo tipo de sistemas multi-resolución, de ahí que es importante comprender la interpretación espectral de los mismos. Para esto, se considera la configuración de la figura 43. La secuencia $v[n]$ resulta de insertar ceros en $y[m]$. Debido a las distintas tasas de muestreo se obtienen las siguientes relaciones entre $Y(z)$ y $V(z)$:

$$Y(z^N) = V(z) \tag{4-3}$$

Luego del sub muestreo y el sobre muestreo por N los valores de $v[nN]$ y $u[nN]$ siguen siendo iguales, mientras que las otras muestras de $v[n]$ son cero.

Figura 43. Componentes típicos de un banco de filtros.



Usando la siguiente correspondencia

$$\frac{1}{N} \sum_{k=0}^{N-1} e^{\frac{2\pi jkn}{N}} = \begin{cases} 1, & \frac{n}{N} \in Z \\ 0, & \text{cualquier otro lugar} \end{cases} \quad (4-4)$$

La relación entre $v[n]$ y $u[n]$ puede ser escrita como

$$v[n] = u[n] \frac{1}{N} \sum_{k=0}^{N-1} W_N^{-kn} \quad (4-5)$$

Donde:

$$W_N = e^{\frac{2\pi j}{N}} \quad (4-6)$$

La transformada Z está dada como

$$\begin{aligned} V(Z) &= \sum_{n=-\infty}^{\infty} v[n]z^{-n} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} \sum_{n=-\infty}^{\infty} u[n](W_N^k z)^{-n} \end{aligned}$$

$$V(Z) = \frac{1}{N} \sum_{k=0}^{N-1} U(W_N^k z) \quad (4-7)$$

La relación entre $Y(z)$ y $U(z)$ se encuentra a partir de las expresiones (4-7) y (4-3):

$$Y(z) = \frac{1}{N} \sum_{k=0}^{N-1} U\left(W_N^k z^{\frac{1}{N}}\right) \quad (4-8)$$

Con (4-8) y observando que $U(z) = H(z)X(z)$ se puede establecer la siguiente relación entre $Y(z)$ y $X(z)$:

$$Y(z) = \frac{1}{N} \sum_{k=0}^{N-1} H\left(W_N^k z^{\frac{1}{N}}\right) X\left(W_N^k z^{\frac{1}{N}}\right) \quad (4-9)$$

A partir de (4-9) y (4-3) y observando que $\hat{X}(z) = G(z)Y(z^N)$ se puede concluir

$$\hat{X}(z) = \frac{1}{N} \sum_{k=0}^{N-1} G(z)H(W_N^k z)X(W_N^k z) \quad (4-10)$$

El espectro de las señales descritas anteriormente se muestra en la figura 44 para el caso de una señal de entrada $x[n]$ de banda estrecha, que no conlleva a efectos de aliasing. Esto significa que los productos $G(z)H(W_N^k z)X(W_N^k z)$ en (4-10) son cero para $i \neq 0$. En el caso general cuando se da aliasing los espectros se traslapan. Esto se muestra en la figura 45, donde las áreas sombreadas indican las componentes de aliasing debido al sub muestreo.

Es claro que $x[n]$ solo puede ser recuperado de $y[m]$ si no existe aliasing. Sin embargo, el caso donde existe aliasing es la forma normal de operación de los bancos de filtros multi-resolución. La razón por la cual estos filtros permiten la reconstrucción perfecta está en el hecho de que pueden ser diseñados de tal manera que las componentes que poseen aliasing de las distintas ramas paralelas se compensan a la salida.

Figura 44. Espectro para las señales decimadas e interpoladas sin efectos de aliasing.

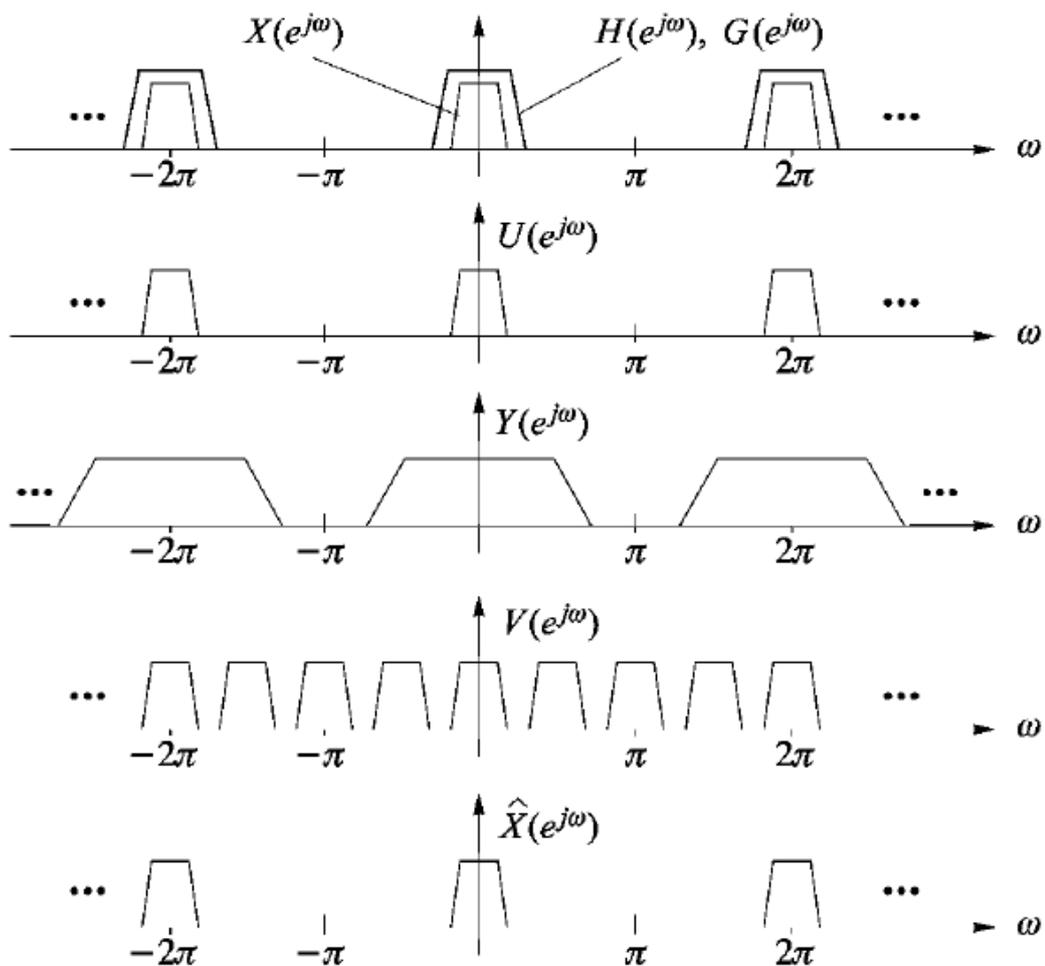
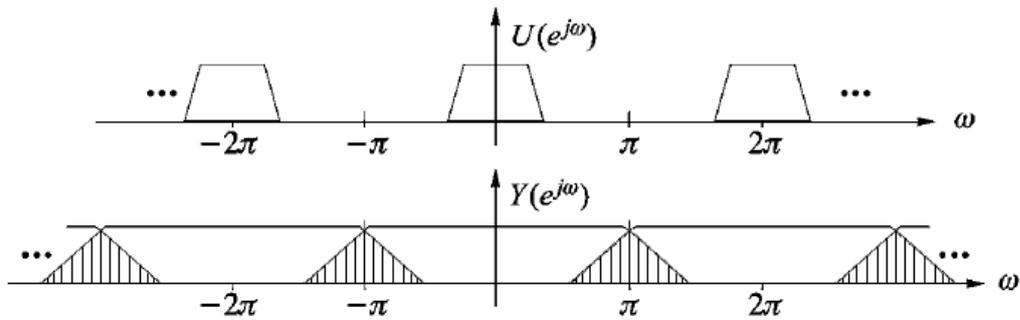


Figura 45. Espectro para las señales decimadas e interpoladas bajo los efectos de aliasing.



4.3.1.2 Descomposición polifásica

Una descomposición polifásica es aquella en la que la secuencia $x[n]$ es descompuesta en sub secuencias $x_i[m]$, como las mostradas en la figura 38. Al intercalar todos los $x_i[m]$ de nuevo conlleva a la señal original $x[n]$. Varios tipos de descomposiciones polifásicas son conocidas, las cuales son descritas a continuación.

Tipo-1

Una descomposición polifásica tipo-1 de la secuencia $x[n]$ en M componentes está dada por

$$X(z) = \sum_{\ell=0}^{M-1} z^{-\ell} X_{\ell}(z^M) \quad (4-11)$$

Donde:

$$X_{\ell}(z) \leftrightarrow x_{\ell}[n] = x[nM + \ell] \quad (4-12)$$

La figura 38 muestra este tipo de descomposición.

Tipo-2

La descomposición polifásica tipo-2 está dada por

$$X(z) = \sum_{\ell=0}^{M-1} z^{-(M-1-\ell)} X'_{\ell}(z^M) \quad (4-13)$$

Donde:

$$X'_{\ell}(z) \leftrightarrow x'_{\ell}[n] = x_{\ell}[nM + M - 1 - \ell] \quad (4-14)$$

La única diferencia entre la descomposición tipo-1 y la del tipo-2 está en el indexado:

$$X_{\ell}(z) = X'_{M-1-\ell}(Z) \quad (4-15)$$

Tipo-3

La descomposición polifásica tipo-3 se define como

$$X(z) = \sum_{\ell=0}^{M-1} z^{\ell} \bar{X}_{\ell}(z^M) \quad (4-16)$$

Donde:

$$\bar{X}_{\ell}(z) \leftrightarrow \bar{x}_{\ell}[n] = x_{\ell}[nM - \ell] \quad (4-17)$$

La relación con las componentes descomposición polifásica tipo-1 es

$$X_0(z) = \bar{X}_0(Z)$$

$$X_{\ell}(z) = z^{-1} \bar{X}_{M-\ell}(Z), \quad \ell = 1, \dots, M-1 \quad (4-18)$$

La descomposición polifásica es usada frecuentemente tanto para señales como para filtros.

4.3.2 Aplicación de la multi-resolución.

Existe una tendencia muy grande en la electrónica de remplazar circuitos analógicos con algoritmos digitales. La conversión de datos es un ejemplo de esto. Por ejemplo, si se considera el diseño de una grabadora de voz, un sistema que digitalizará la voz, la almacenará de forma digital, y luego la reconstruirá para su posterior reproducción. Para recrear un habla inteligible, el sistema debe capturar frecuencias entre 100 y 3000 Hz. Sin embargo, la señal analógica producida por el micrófono también posee frecuencias mayores, por ejemplo de 40 kHz.

La forma clásica de trabajar este problema es el de pasar esta señal analógica por un filtro paso bajo Chebyshev de 8 polos a una frecuencia de corte de 3 kHz, y luego muestrearla a 8 kHz. En el otro extremo, el DAC reconstruye la señal analógica a 8 kHz con un sistema de retención de orden cero. Otro filtro Chebyshev con una frecuencia de 3 kHz es usado para producir la señal de voz final.

Existen muchos beneficios en muestrear más rápido que este análisis directo. Por ejemplo, si se rediseñara este grabador usando un muestreo de 64 kHz. El filtro *antialias* tiene una tarea más sencilla: pasar todas las frecuencias por debajo de 3 kHz, mientras rechaza todas las frecuencias por encima de 32 kHz. Una simplificación similar ocurre en el filtro de reconstrucción. Lo que permite que a un muestreo mayor el filtro de ocho polos sea remplazado por un sistema simple de capacitor y resistencia. El problema es que ahora el sistema se encuentra inundado con datos por la alta tasa de muestreo.

Utilizando la multi-resolución se podrían usar más de una sola tasa de muestreo. El ejemplo de la grabadora de voz funcionaría de la siguiente manera. Primero, al pasar la señal de voz a través de un filtro paso bajo RC simple y muestrear a una frecuencia de 64 kHz. Los datos resultantes contienen las señales deseadas de 100 a 3000 Hz, pero también poseen una banda indeseable entre 3 kHz y 32 kHz. Segundo, se procede a remover estas frecuencias indeseables por software, al usar un filtro paso bajo digital a 3 kHz. Tercero, volviendo a muestrear la señal digital de 64 kHz a 8 kHz al simplemente descartar siete de cada ocho muestras, un procedimiento llamado decimación. Los datos digitales resultantes son equivalentes a los producidos por un filtrado analógico y un muestreo de 8 kHz.

La técnica multi-resolución también puede ser usada en la salida de nuestro ejemplo. Los datos a 8 kHz son retirados de la memoria y convertidos a una tasa de muestreo de 64 kHz, usando la interpolación. Lo que involucraría introducir siete muestras, con un valor de cero, entre cada una de las muestras obtenidas desde la memoria. La señal resultante es un tren de impulsos digitales, conteniendo la banda de frecuencias deseada de entre 100 y 3000 Hz, además de una duplicación espectral entre 3 kHz y 32 kHz. Posteriormente todo por encima de 3 kHz es removido con un filtro digital paso bajo. Luego de la conversión a una señal digital por medio de un DAC, una red RC simple es todo lo necesario para producir la señal de voz fina.

La conversión multi-resolución es valiosa por dos razones: (1) reemplaza componentes análogos con software, una ventaja económica clara en producciones en masa, y (2) se pueden alcanzar altos niveles de rendimiento en aplicaciones críticas. Por ejemplo, un sistema de audio de discos compactos usa esta técnica para alcanzar la mejor calidad de sonido. Este incremento de rendimiento es un resultado de reemplazar los componentes analógicos (1% de precisión), con algoritmos digitales (0.0001% de precisión de error por redondeo). Por lo que los filtros digitales presentan un rendimiento mucho mayor en distintas aéreas.

4.4 Desventajas del tratamiento clásico

Las transformaciones en el dominio de la frecuencia son ya una herramienta estándar en el procesado de señales: muchas investigaciones se han hecho, y muchas mejoras se han alcanzado. Algoritmos rápidos con bajo tiempo de computación han sido desarrollados, así como también su aplicación en tiempo real hacen que sea una de las bases en el estudio de señales. Muchas aplicaciones musicales hoy en día usan la transformada de Fourier para el procesado de alta definición en aplicaciones de software y hardware. Por ejemplo, los archivos mp3 usan una derivación de la DFT para una compresión de alta calidad.

A pesar de su idoneidad para calcular el espectro de frecuencia, ciertos inconvenientes para información de audio, en especial en tiempo real, han sido identificados:

1. El análisis produce bandas de frecuencia igualmente espaciadas. Esto no corresponde a la precepción humana de las frecuencias. Para obtener una resolución aceptable en bajas frecuencias, la resolución en altas frecuencias se encuentra “sobre detallada”. Por lo que, la transformada de Fourier es ineficiente para estos propósitos.
2. Cuando los datos son inventanados para reducir errores, es incluso más ineficiente, cuando técnicas como la suma de traslape son usadas.
3. Para aplicaciones en tiempo real donde la STFT es usada, un compromiso entre la resolución del tiempo y la frecuencia debe de ser aceptado debido al principio de incertidumbre.

Muchas alternativas en el análisis espectral han sido desarrolladas para superar las limitaciones de la transformada de Fourier. Sin embargo, todas tienen sus respectivas limitaciones, y ningún método general supera todas las limitaciones sin introducir nuevas limitaciones. Estas muchas veces proveen mejores soluciones en algunos campos específicos. Especialmente, muchas de ellas no proveen una resíntesis, por lo que no son adecuadas para el procesamiento de audio. Algunas de estas alternativas son el análisis espectral por auto regresión, codificación lineal predictiva, funciones de *Walsh*, etc.

En conclusión, se puede decir que la STFT ha sido aplicada de forma exitosa en sistemas de procesamiento musical en tiempo real. Sus profundas investigaciones proveen una solución adecuada para la solución de este tipo de problemas. Sin embargo, sus problemas inherentes la hacen ver como un compromiso más que una solución adecuada para este campo.

Muchas investigaciones se han hecho apuntando a una nueva representación de tiempo/frecuencia, que pueda resolver esta incerteza en el tiempo/frecuencia. La transformada *wavelet* proporciona una nueva investigación en este campo.

5. WAVELETS

5.1 Introducción

La historia de las *wavelets* se origina en 1910, con Alfred Haar, un matemático alemán, que desarrolló la llamada función *Haar* y la matriz *Haar*. Siendo ésta un tipo especial de matriz: de 2 operaciones (traslación – compresión y dilatación –desplazamiento) sobre un vector madre, la matriz es construida, todos los vectores son perpendiculares unos de otros, debido a este vector madre. Con este esquema fue posible crear matrices ortogonales de cualquier tamaño, todos estos vectores basados en un único vector. Esta es la primera construcción conocida de una *wavelet*, mientras que el término aún no había sido establecido en ese tiempo.

Posteriormente, se fueron desarrollando investigaciones para superar y entender las limitaciones de la transformada de Fourier. Un campo de interés fue el de partir un fenómeno complejo en varias piezas más sencillas. En los años 30, estas fueron las técnicas de *Littlewood-Paley*, desarrolladas más ampliamente en los años 50 y 60 llevando a la aplicación de la teoría de *Calderon-Zygmund*. En los años 70 la descomposición atómica, como la teoría de descomposición en el espacio de *Hardy* fue ampliamente usada. G. *Weiss* y R. *Coifman* aportaron mucha investigación en estas descomposiciones atómicas.

En 1980, A. *Grossman* y J. *Morlet* definieron de forma general las *wavelets* en el contexto de la física cuántica. Poco tiempo después, J. *Strömberg* descubrió la primera *wavelet* ortogonal.

Luego en los 80s, Y. Meyer y otros grupos independientes realizaron cálculos discretos de las técnicas de *Littlewood-Paley*, luego de entender que estas podrían ser un sustituto de las técnicas de Fourier. Es ahí donde Grossmann y Morlet sugieren el nombre de “*wavelets*” en lugar de la “Teoría de *Littlewood-Paley*”.

5.2 Funciones *wavelet*

Las *wavelets* son funciones en el tiempo continuo que poseen propiedades especiales; usualmente la letra ψ es usada para la función *wavelet*. La función necesita desaparecer desde $-\infty$ hasta ∞ . Esto lleva a decir que las *wavelets* son ondas localizadas.

Otro requerimiento es que su integral sea cero. De ahí, que necesite por lo menos de un cambio de signo, haciendo que su forma se parezca a una pequeña onda –el nombre de *wavelet* está basado en esta propiedad, siendo la traducción del francés *ondelette* al inglés *wavelet*.

Las funciones *wavelet* son los puntos de partida para el análisis, siendo comparables con las ondas sinusoidales de la FT. Existe un desarrollo matemático elevado cuando se desean *wavelets* con algunas propiedades especialmente buenas. Las *wavelets* que aquí se presentan fueron diseñadas con estas propiedades, por lo que su construcción no ha sido para nada trivial.

5.2.1 Wavelet Haar

La *wavelet* Haar es un tipo especial de *wavelet*. Posee únicamente 2 coeficientes del filtro. La función de la *wavelet* es una onda cuadrada; por lo que señales suaves no pueden ser bien aproximadas. Es la única *wavelet* que a la vez es simétrica y ortogonal. Debido a su velocidad de computación, es perfecta para el procesamiento en tiempo real. Sin embargo, la calidad no es suficiente; cualquier modificación en los coeficientes *wavelet* resulta en un gran nivel de *aliasing*.

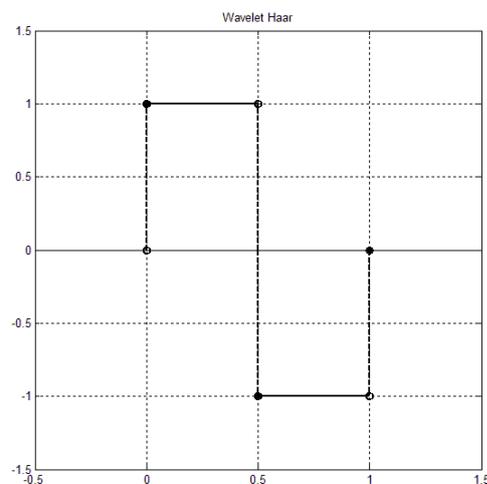
La función de análisis, *wavelet* Haar, se puede escribir como:

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \text{otras regiones} \end{cases} \quad (5-1)$$

mientras que, la función de escala es:

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & \text{otras regiones} \end{cases} \quad (5-2)$$

Figura 46. Función *wavelet* Haar.



5.2.2 Wavelet Daubechies

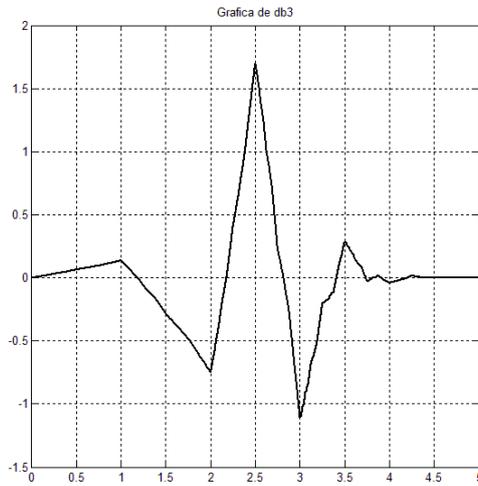
A finales de los años 80's las *wavelets* ortogonales y de soporte compacto creadas por Ingrid *Daubechies* ganaron mucha atención. Estas fueron unas de las primeras que permitieron el análisis discreto de *wavelets* de forma práctica.

Ella las construyo al diseñar filtros ortogonales que cumplieran con ser máximamente planas a una frecuencia de cero y a un medio de la frecuencia de muestreo. Así que, la restricción para el diseño fue el de tener el máximo número de momentos de desvanecimiento a partir de un tamaño de soporte dado. Para un número de momentos de desvanecimiento p , el filtro tiene una cantidad de $2p$ coeficientes. La restricción de un soporte mínimo conlleva a una resolución temporal máxima. Los filtros y *wavelets* resultantes son llamados *Daubechies* p o solamente Dp . Para el caso especial de $p = 1$, la *wavelet* resultante es el tipo Haar.

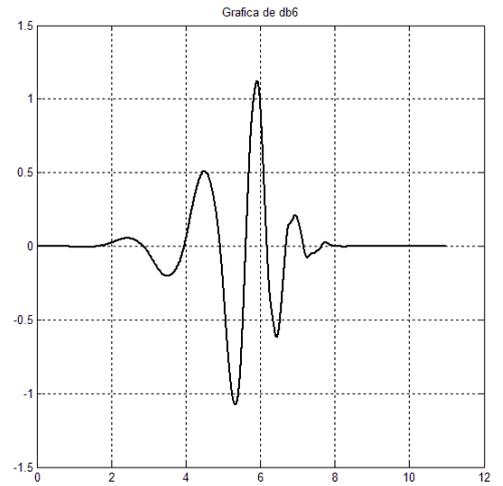
La mayoría de las *wavelets* *Daubechies* no son simétricas –por el contrario, muchas son bastante asimétricas. Para valores pequeños $p > 1$, estas no son suaves pero aun son continuas. Al aumentar p , la función *wavelet* se vuelve más suave. Por ejemplo, la *wavelet* $D2$ tiene singularidades en los puntos $p/2^n$ (con p y n enteros) donde es diferenciable por la izquierda pero no por la derecha. Debido a su planicie, el filtro no separa las bandas de frecuencia muy bien. La pendiente de respuesta en frecuencia del filtro es proporcional a la raíz cuadrada de $2p$.

Las *wavelets* *Daubechies* no están definidas en términos de las funciones de escala y *wavelet*; de hecho, no son posibles de escribir en una forma cerrada. Las graficas de la Figura 47 son generadas usando el algoritmo de cascada, una técnica numérica que consiste en obtener la transformada inversa de $[1\ 0\ 0\ 0\ 0\ \dots]$ un número apropiado de veces.

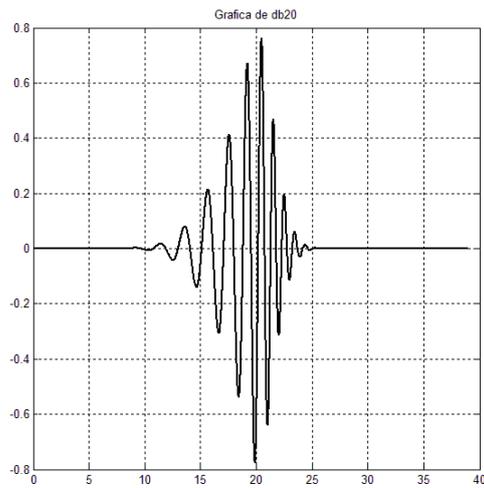
Figura 47. Familia de funciones wavelet Daubechies.



(a).

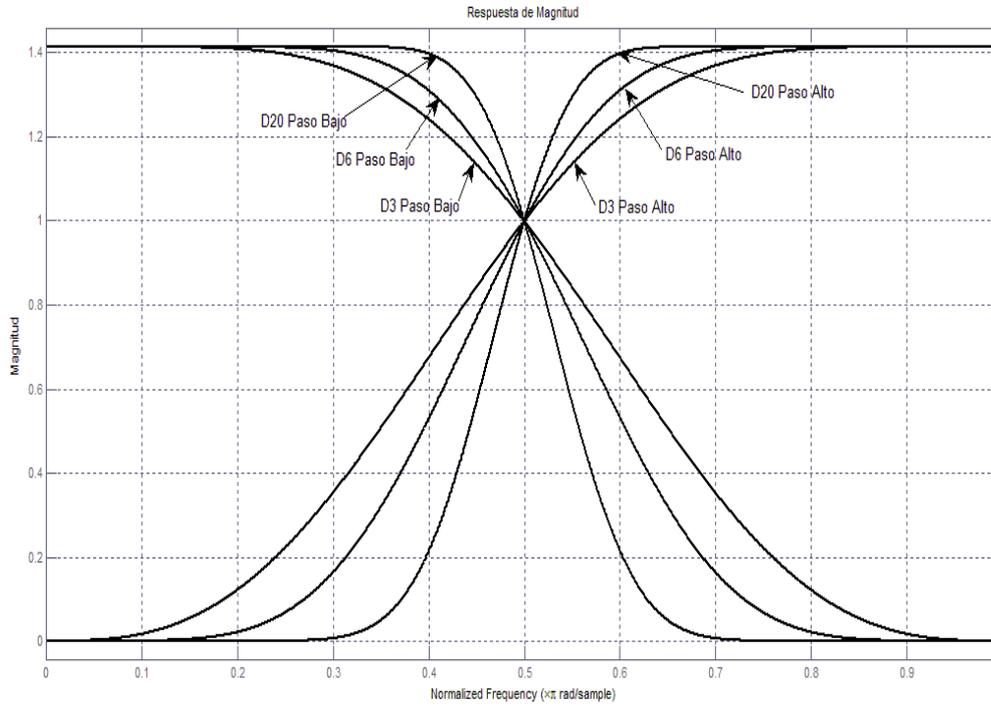


(b).



(c).

Figura 48. Respuesta en frecuencia de funciones *wavelet Daubechies*.



La Figura 47 muestra la gráfica de las funciones (a) *wavelet D3*, (b) *wavelet D6* y (c), *wavelet D20*. En la Figura 48 se muestra la respuesta en frecuencia de los respectivos filtros. Se puede notar que entre más elevado es el valor de p , más empinada es la curva de transición.

5.2.3 Otras *wavelet* ortogonales

Daubechies construyó otra serie de *wavelets* ortogonales denominadas *symmlets*, cuyas propiedades son tan buenas como las de la familia *Daubechies* (soporte compacto, p momentos de desvanecimiento) pero fueron diseñadas con el requerimiento de optimizar la simetría y la fase lineal. Aún, cuando es imposible para las *wavelets* ortogonales, no son perfectamente simétricas.

Otra familia de *wavelets*, también construida por *Daubechies*, es la denominada *Coiflets*. Ella las construyo por pedido de Ronald Coifman, quien necesitaba *wavelets* similares a las de la familia *Daubechies*, pero con una restricción adicional en la función de escala: no solo la función *wavelet*, sino también la función de escala debía tener p momentos de desvanecimiento. Esto presenta la ventaja de que los coeficientes de aproximación pueden ser aproximados por las muestras de la señal por sí mismos. Sin embargo, el soporte, y en consecuencia la longitud de los filtros, es mayor (la longitud del filtro es $6p$ en lugar de $2p$), así que esta propiedad adicional conlleva un costo en eficiencia.

Una familia especial de *wavelets* es el de las *wavelets Meyer*. En este caso la *wavelet* y la función de escala son construidas en el dominio de la frecuencia con una función auxiliar. Su soporte es infinito, pero aún así las funciones tienen un decaimiento rápido. Son infinitamente diferenciables; de ahí que sean simétricas y ortogonales, pero no poseen ningún momento de desvanecimiento. Los filtros FIR no pueden ser construidos, así que la implementación del banco de filtros no es posible.

5.2.4 Wavelets crudas

Algunas *wavelets* que carecen de muchas propiedades interesantes se denominan *wavelets* “crudas”; la *wavelet Morlet* y el sombrero mexicano son un caso de este tipo de *wavelets*. Ambas poseen una expresión explícita para ψ , pero la función de escala no puede ser construida. No poseen soporte compacto, ni tampoco momentos de desvanecimiento, y no son ortogonales. Debido a estas limitaciones, no se pueden hacer cálculos para construir algún filtro, por lo que solo se puede realizar la transformada *wavelet* continua. Su utilidad radica en la posibilidad de ser usadas para demostraciones matemáticas, ya que la función *wavelet* existe como fórmula.

5.2.5 Wavelets biortogonales

Existe una buena cantidad de *wavelets* biortogonales estudiadas. La mayor ventaja de las *wavelets* biortogonales es la posibilidad de crear transformadas simétricas: tanto la función *wavelet* como la función de escala son simétricas. Esto requiere una longitud impar en ambos filtros de análisis. La función de escalado y *wavelet* son distintas para el análisis y resíntesis, así para una transformación de un banco de filtros, dos filtros de análisis y dos filtros de resíntesis distintos son necesarios. Una práctica común para transformadas biortogonales es el indicar la función de escala y *wavelet* de análisis con $\tilde{\phi}$ y $\tilde{\psi}$, respectivamente.

Es aparente que los filtros pueden tener propiedades distintas para el análisis y resíntesis. Consecuentemente, propiedades útiles para el análisis son diseñadas en el filtro de análisis (como los momentos de desvanecimiento) mientras que los filtros de resíntesis pueden ser diseñados respecto a otras propiedades útiles para la reconstrucción (como la regularidad).

Battle y *Lemarié* introdujeron *wavelets* biortogonales basados en *splines* polinomiales. Para *splines* de grado m , la función *wavelet* resultante tiene $m + 1$ momentos de desvanecimiento. A diferencia de las *wavelets* Daubechies, no poseen soporte compacto; filtros finitos solo pueden ser aproximados al cortar los bordes. Sin embargo, la función *wavelet* posee una caída exponencial, así que truncar los filtros de forma razonable no introduce mucho error. Las funciones *wavelet* generadas a partir de *splines* polinomiales pueden ser especificadas explícitamente en el dominio de la frecuencia, y como son *splines* polinomiales, son $m + 1$ veces continuamente diferenciables, resultando en *wavelets* suaves. Para un valor impar de m , estas *wavelets* son simétricas.

Un esquema de ortogonalización permitiría hacer la familia de filtros Battle y Lemarié ortogonal. En resumen, las wavelets tipo spline proporcionan una regularidad máxima con simetría y soporte mínimo.

Otras wavelets biortogonales son las *Binlets*, basadas también en splines. Ellas son simétricas, poseen soporte compacto y los coeficientes son binarios: todos los coeficientes del filtro son enteros divididos por la misma potencia de 2. Esto permite una implementación eficiente en computadoras.

5.3 Transformada *wavelet* continua

La transformada *wavelet* $W_x(b, a)$ de una señal $x(t)$ continua en el tiempo está definida como:

$$W_x(b, a) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (5-3)$$

Por lo que, la transformada *wavelet* es calculada como el producto de $x(t)$ y la versión trasladada y escalada de la función $\psi(t)$, denominada función *wavelet*.

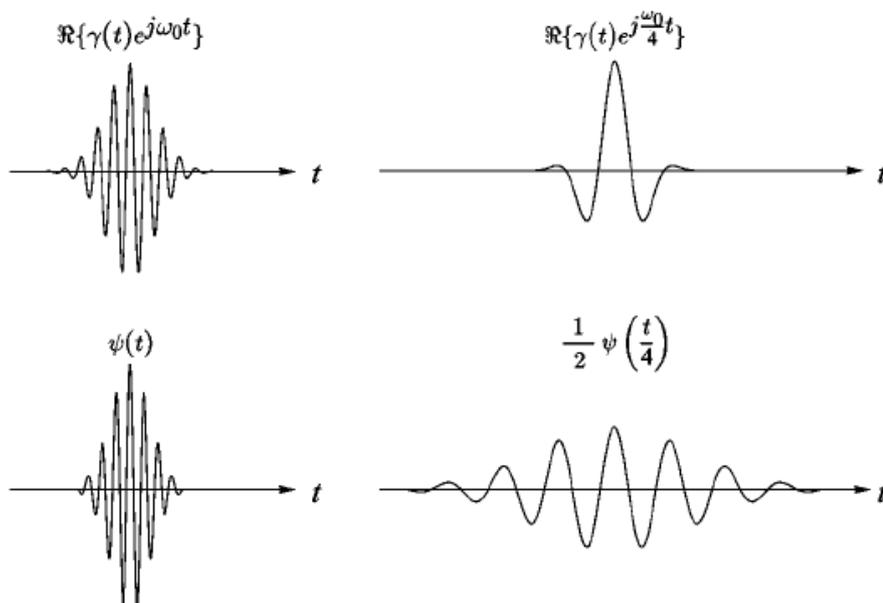
Si se considera a $\psi(t)$ como una respuesta al impulso en la banda de paso, el análisis *wavelet* puede ser entendido como un análisis en esta banda. Al variar el parámetro de escala a la frecuencia central y el ancho de banda de la banda de paso son influenciadas. Las variaciones de b simplemente significan una traslación en el tiempo; así que, para un valor fijo de a la transformada en la ecuación (5-3) puede ser vista como la convolución de $x(t)$ con la función *wavelet* revertida en el tiempo y escalada:

$$W_x(t, a) = |a|^{-\frac{1}{2}} x(t) * \psi_a(t), \quad \psi_a(t) = \psi^*\left(\frac{-t}{a}\right) \quad (5-4)$$

El factor de $|a|^{-\frac{1}{2}}$ es introducido para asegurar que todas las funciones escaladas $|a|^{-\frac{1}{2}} \psi^*(t/a)$ con $a \in \mathbb{R}$ tengan la misma energía.

Ya que la función de análisis $\psi(t)$ está escalada y no modulada como el núcleo de la STFT, un análisis *wavelet* es usualmente llamado análisis en tiempo-escala en lugar de un análisis en tiempo-frecuencia. Sin embargo, ambas son relacionadas de forma natural una con la otra por medio de la interpretación de la banda de paso.

Figura 49. Comparación de los núcleos de análisis de la STFT y la WT.



La figura 49 muestra ejemplos de los núcleos de la STFT y de la transformada *wavelet*. Como se puede ver, una variación del retraso en el tiempo b y/o el parámetro de escala a no tienen efecto alguno en la forma del núcleo de la transformada *wavelet*. Sin embargo, la resolución en el tiempo y la frecuencia de la transformada *wavelet* dependen de a .

La transformada *wavelet* se puede interpretar como un análisis de octavas con constante Q; ya que, para un análisis de altas frecuencias (un a pequeño) se tiene una buena localización en el tiempo y una pobre resolución en la frecuencia. Por otro lado, para bajas frecuencias de análisis, se tiene buena resolución en frecuencia pero pobre en el tiempo.

Cuando se utiliza la transformada *wavelet* se debe de asegurar que la señal pueda ser perfectamente reconstruida a partir de esta representación. De otra manera la representación sería completa o parcialmente inservible. Para asegurar una reconstrucción perfecta de la transformada *wavelet* se debe cumplir la siguiente condición:

$$C_{\psi} = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty \quad (5-5)$$

Donde $\Psi(\omega)$ denota la transformada de Fourier de la *wavelet*. Esto se conoce como la condición de admisibilidad de la función *wavelet* $\psi(t)$.

Obviamente, para satisfacer la condición en (5-5) la *wavelet* debe cumplir con

$$\psi(0) = \int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (5-6)$$

Más aún, $|\Psi(\omega)|$ debe decrecer rápido para $|\omega| \rightarrow 0$ y para $|\omega| \rightarrow \infty$, esto es, $\psi(t)$ debe de ser una respuesta al impulso de una banda de paso.

La transformada *wavelet* puede ser obtenida por medio de la interpretación de un análisis en la banda de paso y la transformada inversa de Fourier. En base a la relación de Parseval se puede aplicar la transformada inversa de Fourier al espectro enventanado de $X(\omega)\Psi^*(a\omega)$, lo que conduce a:

$$W_x(b, a) = |a|^{\frac{1}{2}} \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) \Psi^*(a\omega) e^{j\omega b} d\omega \quad (5-7)$$

La resolución en el tiempo-frecuencia puede ser descrito por medio de la ventana en el tiempo-frecuencia asociada a la *wavelet*. El centro (t_0, ω_0) y los radios Δ_t y Δ_ω pueden ser calculados por medio del análisis de variable aleatoria. El valor de t_0 y Δ_t de la ventana pueden ser definidos de forma análoga al valor medio y la desviación estándar de la variable aleatoria respectivamente:

$$t_0 = \int_{-\infty}^{\infty} t \cdot \frac{|\psi(t)|^2}{\|\psi\|^2} dt \quad (5-8)$$

$$\Delta_t = \left(\int_{-\infty}^{\infty} (t - t_0)^2 \cdot \frac{|\psi(t)|^2}{\|\psi\|^2} dt \right)^{\frac{1}{2}}. \quad (5-9)$$

De igual manera los valores para ω_0 y Δ_ω serían:

$$\omega_0 = \int_{-\infty}^{\infty} \omega \cdot \frac{|\Psi(\omega)|^2}{\|\Psi\|^2} d\omega \quad (5-10)$$

$$\Delta_\omega = \left(\int_{-\infty}^{\infty} (\omega - \omega_0)^2 \cdot \frac{|\Psi(\omega)|^2}{\|\Psi\|^2} d\omega \right)^{\frac{1}{2}}. \quad (5-11)$$

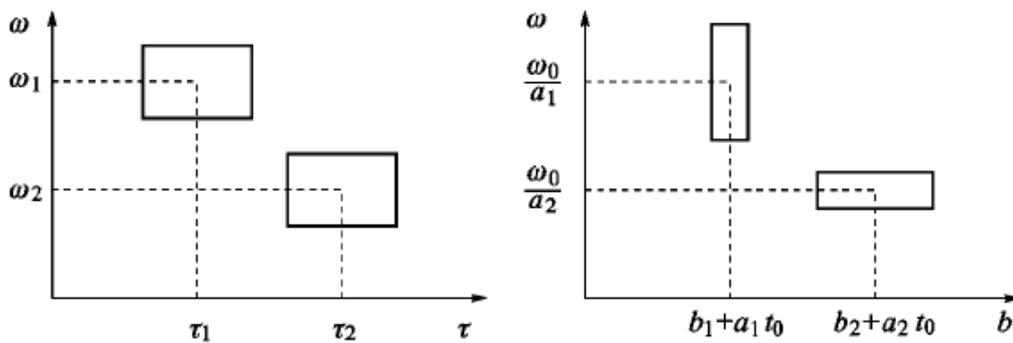
Para el centro y el radio de la función escalada $\psi\left(\frac{t}{a}\right) \leftrightarrow |a|\Psi(a\omega)$ se tiene $\left\{a \cdot t_0, \frac{1}{a}\omega_0\right\}$ y $\left\{a \cdot \Delta_t, \frac{1}{a}\Delta_\omega\right\}$, respectivamente. Esto significa que la transformada *wavelet* $W_x(b, a)$ provee información de la señal $x(t)$ y de su espectro $X(\omega)$ en una ventana de tiempo-frecuencia.

$$[b + a \cdot t_0 - a \cdot \Delta_t, b + a \cdot t_0 + a \cdot \Delta_t] \times \left[\frac{\omega_0}{a} - \frac{\Delta_\omega}{a}, \frac{\omega_0}{a} + \frac{\Delta_\omega}{a} \right] \quad (5-12)$$

El área $4\Delta_t\Delta_\omega$ es independiente del parámetro a y b y está determinado solamente por la *wavelet* $\psi(t)$ usada.

La ventana de tiempo se comprime cuando a se hace pequeño, y se expande cuando a se hace grande. Por otro lado, la ventana de frecuencia se amplía cuando a se hace pequeño, y se comprime cuando a se hace grande. Un análisis de ventana pequeño lleva a una buena resolución de tiempo pero a una mala resolución de frecuencia. La figura 50 muestra las distintas resoluciones de la transformada rápida de Fourier, izquierda, y la transformada *wavelet*, derecha.

Figura 50. Comparación de la resolución de la STFT y la WT.



La ecuación (5-3) muestra que si la señal es escalada $\left(x(t) \rightarrow x\left(\frac{t}{c}\right)\right)$, la representación *wavelet* $W_x(b, a)$ también estará escalada; excepto esto, $W_x(b, a)$ no sufrirá ninguna otra modificación. Por esta razón se habla de una transformada invariante afín. Más aún, la transformada *wavelet* es invariante a la traslación, es decir, un desplazamiento de la señal $(x(t) \rightarrow x(t - t_0))$ conlleva a un desplazamiento de t_0 de la representación *wavelet* $W_x(b, a)$, pero aparte de esto no sufre ninguna otra modificación.

5.3.1 Wavelets para el análisis en escala de tiempo

En el análisis de señales en la escala de tiempo nuestra meta es inferir ciertas propiedades de la señal a partir de la transformada *wavelet* de una manera conveniente.

Wavelets analíticas son especialmente adecuadas para este propósito. Como una señal analítica, estas contienen únicamente frecuencias positivas. Para la transformada de Fourier de la *wavelet* analítica $\psi_{b,a}(t)$ se tiene lo siguiente:

$$\Psi_{b,a}(\omega) = 0 \text{ para } \omega \leq 0 \quad (5-13)$$

Considerando ahora la señal real $x(t) = \cos(\omega_0 t)$. Su espectro es

$$X(\omega) = \pi[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)] \leftrightarrow x(t) = \cos(\omega_0 t) \quad (5-14)$$

Substituyendo $X(\omega)$ con (5-14) en (5-7) lleva a

$$\begin{aligned} W_x(b, a) &= \frac{1}{2} |a|^{\frac{1}{2}} \int_{-\infty}^{\infty} (\delta(\omega - \omega_0) + \delta(\omega + \omega_0)) \Psi^*(a\omega) e^{j\omega b} d\omega \\ &= \frac{1}{2} |a|^{\frac{1}{2}} [\Psi^*(a\omega_0) e^{j\omega_0 b} + \Psi^*(-a\omega_0) e^{-j\omega_0 b}] \end{aligned} \quad (5-15)$$

De ahí, para una *wavelet* analítica

$$W_x(b, a) = \frac{1}{2} |a|^{\frac{1}{2}} \Psi^*(a\omega_0) e^{j\omega_0 b} \quad (5-16)$$

Como únicamente el argumento del exponencial complejo en la ecuación (5-16) depende de b , la frecuencia de $x(t)$ puede ser inferida de la fase de $W_x(b, a)$. Por esto, cualquier línea horizontal en el plano de tiempo-frecuencia puede ser considerado. La magnitud de $W_x(b, a)$ es independiente de b , así que la amplitud de $x(t)$ puede ser vista como independiente del tiempo. Esto significa que la magnitud de $W_x(b, a)$ muestra directamente la distribución de energía de la señal en el tiempo-frecuencia.

Un escalograma es el cuadrado de la magnitud de la transformada *wavelet*:

$$|W_x(b, a)|^2 = \left| |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \right|^2 \quad (5-17)$$

Los escalogramas, igual que los espectogramas, pueden ser representados como imágenes en donde la intensidad es expresada en distintos tonos de grises. La figura 51 muestra escalogramas para $x(t) = \delta(t)$. En la figura 51(a) se muestra un escalograma para una *wavelet* real, mientras que, en la figura 51(b) se muestra un escalograma para una *wavelet* analítica. Se puede ver que las *wavelets* analíticas deberían ser escogidas de tal manera que se visualice la distribución de energía de la señal en relación al tiempo y la frecuencia (y escalado, respectivamente).

5.3.2 Reconstrucción integral y semi-discreta

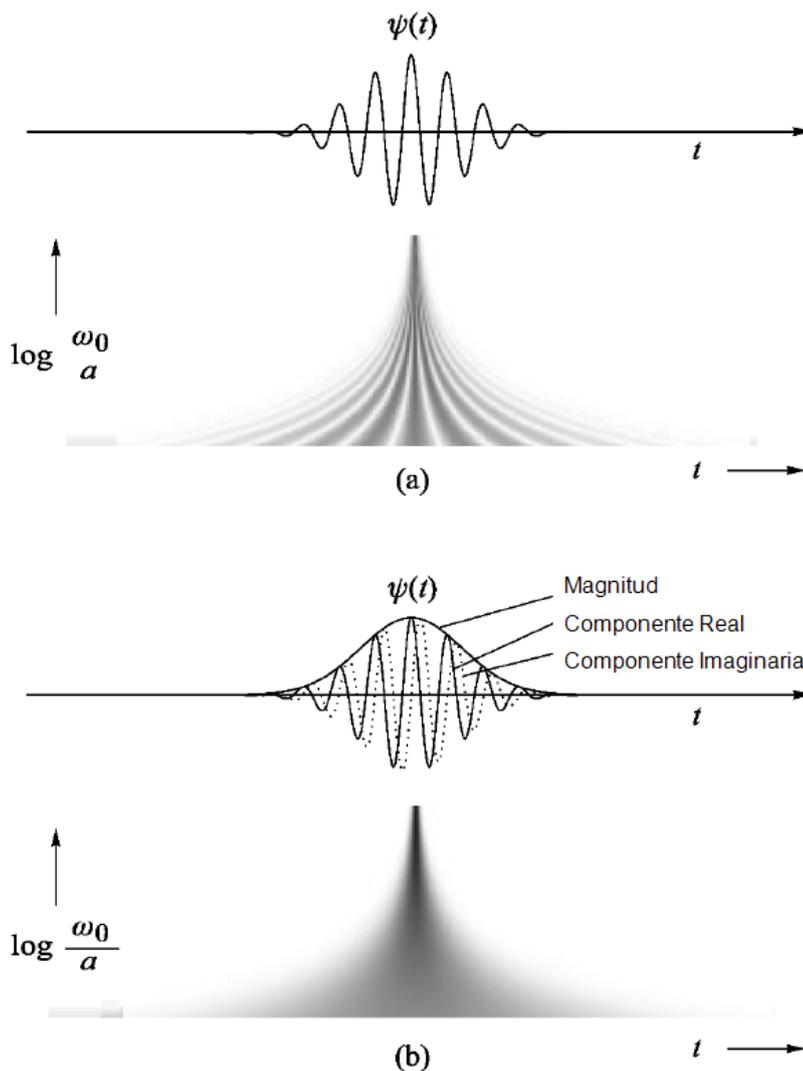
Se considerarán dos variantes de la transformada *wavelet*, y éstas difieren en la forma en que son reconstruidas. Específicamente, se verá la reconstrucción integral desde todo el plano de tiempo-frecuencia y luego a una reconstrucción semi-discreta.

La transformada inversa de la CWT esta dada como:

$$\langle x, y \rangle = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_x(b, a) W_y^*(b, a) \frac{da db}{a^2} \quad (5-18)$$

Calcular los coeficientes *wavelet* en cada escala posible es una cantidad muy grande de trabajo, y genera una cantidad asombrosa de datos. De ahí que, si se escogen escalas y posiciones basados en potencias de dos, nuestro análisis puede ser igual de exacto y eficiente. A este tipo de análisis se le conoce como *wavelets* diádicas semi-discretas.

Figura 51. Escalogramas para wavelet real y analítica.



Se habla de *wavelets* diádicas semi-discretas si cada señal $x(t) \in L^2(R)$ puede ser reconstruida a partir de valores semi-discretos $W_x(b, a_m)$, dónde $a_m, m \in Z$ se encuentran arreglados de forma diádica:

$$a_m = 2^m \quad (5-19)$$

La transformada *wavelet* es calculada a través de las líneas $W_x(b, 2^m)$:

$$W_x(b, 2^m) = 2^{-\frac{m}{2}} \int_{-\infty}^{\infty} x(t) \psi^*(2^{-m}(t - b)) dt \quad (5-20)$$

5.4 Series wavelet

Anteriormente se mencionó que nuestro análisis se puede reducir a escalas de 2 sin perder exactitud y eficiencia, es más, la naturaleza booleana de los computadores hace que esta aproximación sea aparentemente transparente en sus microprocesadores. A continuación se presentan algunas consideraciones a partir de esta aproximación.

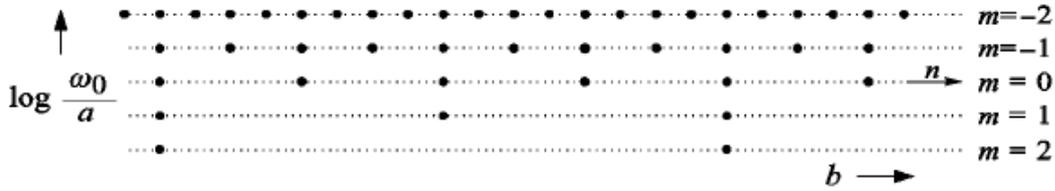
5.4.1 Muestreo diádico

La reconstrucción a partir de valores discretos puede ser obtenida si se consideran los siguientes puntos:

$$a_m = 2^m, b_{mn} = a_m nT = 2^m nT \quad (5-21)$$

Esto conlleva a los valores $W_x(b_{mn}, a_m) = W_x(2^m nT, 2^m)$. En la figura 52 se puede observar el esquema de muestreo.

Figura 52. Muestreo diádico de la transformada *wavelet*.



Los valores de $\{W_x(2^m nT, 2^m), m, n \in Z\}$ forman la representación de $x(t)$ respecto a la *wavelet* $\psi(t)$ y el esquema de muestreo.

No se podría asumir que cualquier conjunto de $\psi_{mn}(t), m, n \in Z$ permiten la reconstrucción de todas las señales $x(t) \in L^2(R)$. De ahí que deba existir un conjunto dual $\check{\psi}_{mn}(t), m, n \in Z$, y ambos conjuntos deben atravesar $L^2(R)$.

El conjunto dual no necesariamente debe ser construido a partir de *wavelets*. Sin embargo, nos interesa el caso donde $\check{\psi}_{mn}(t)$ se encuentra derivado como:

$$\check{\psi}_{mn}(t) = 2^{-\frac{m}{2}} \cdot \check{\psi}(2^{-m}t - nT), \quad m, n \in Z \quad (5-22)$$

Para una *wavelet* $\psi(t)$ dada, la posibilidad de reconstrucción perfecta es dependiente del intervalo de muestreo T . Si se escoge un valor muy pequeño de T , sobre muestreo, los valores $W_x(2^m nT, 2^m), m, n \in Z$ son bastante redundantes, y la reconstrucción es muy sencilla. Por lo que las funciones $\check{\psi}_{mn}(t), m, n \in Z$ son linealmente dependientes, y un número infinito de conjuntos duales $\check{\psi}_{mn}(t)$ existen. La cuestión en sí sobre la existencia de un conjunto dual $\check{\psi}_{mn}(t)$ puede ser obtenida si se examinan dos marcos límites A y B . Se puede demostrar que la existencia de un conjunto dual se puede garantizar si se cumple con la condición de estabilidad de los marcos límites. Además, si se utilizara un marco ajustados, $A = B$, la reconstrucción perfecta con $\check{\psi}_{mn}(t) = \psi_{mn}(t)$ es posible.

Si se escoge un valor grande de T tal que las muestras $W_x(2^m nT, 2^m)$, $m, n \in \mathbb{Z}$ no contengan redundancia alguna (muestreo crítico), las funciones $\psi_{mn}(t), m, n \in \mathbb{Z}$ son linealmente independientes. Si además cumple con la condición de estabilidad de los marcos límites con $0 < A \leq B < \infty$ se cumple con las condiciones para las denominadas *wavelets* biortogonales.

Como caso especial, se tienen las *wavelets* ortonormales, siendo recíprocas sobre sí mismas.

Para el caso ortonormal, las funciones $\psi_{mn}(t), m, n \in \mathbb{Z}$ pueden ser usadas para el análisis y síntesis. Los casos ortonormales siempre tienen los mismos límites de marco (marcos ajustados), siendo una relación especial de la forma de Parseval.

5.4.2 Descomposición por octavas

Un análisis de octavas es muchas veces insuficiente. Se preferiría descomponer cada octava en M sub bandas, de tal manera que se mejoraría la resolución de frecuencia por un factor de M .

Se considera el caso donde el período de muestreo usado para todas las sub bandas M de una octava es el mismo. Esto corresponde a una jerarquización de un análisis de M *wavelets* diádicas con las *wavelets* escaladas

$$\psi^{(q)}(t) = 2^{\frac{q}{2M}} \cdot \psi\left(\frac{q}{2^M}t\right), \quad q = 0, 1, \dots, M - 1 \quad (5-23)$$

Se puede generalizar un poco más este concepto al expresar:

$$a_m = a_0^m, \quad b_{mn} = a_m nT, \quad m, n \in \mathbb{Z} \quad (5-24)$$

Con un valor arbitrario $a_0 > 1$. Se puede dar un un análisis de M *wavelets* diádicas con las *wavelets*

$$\psi^{(q)}(t) = a_0^{\frac{q}{2M}} \cdot \psi\left(a_0^{\frac{q}{M}}t\right), \quad q = 0, 1, \dots, M - 1 \quad (5-25)$$

5.5 Transformada wavelet discreta

La transformada discreta *wavelet*, DWT, es calculada de forma análoga a la CWT. Aquí se presenta la DWT diádica, la cual está escalada en potencias de 2, resultando en las siguientes transformadas discretas:

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k) \quad (5-26)$$

$$d_{j,k} = \int_{-\infty}^{\infty} f(t) \psi_{j,k}(t) dt \quad (5-27)$$

Los coeficientes $d_{j,k}$ son los coeficientes *wavelet*, siéndo analogos a los obtenidos en la forma continua de la transformada *wavelet*.

La transformada discreta inversa se encuentra a partir de la suma de las *wavelets* trasladadas y dilatadas, que se encuentran ponderadas por los coeficientes $d_{j,k}$.

$$f(t) = \sum_{j,k} d_{j,k} \psi_{j,k}(t) \quad (5-28)$$

La base fundamental de estas transformadas esta en la condición de biortogonabilidad, en donde, cada espacio representa una función linealmente independiente, que al ser sumadas son cubiertas todas la bandas de frecuencias.

Es decir, como cada subespacio cubre cierta banda de frecuencias, al aplicar la intersección a este grupo de subespacios, se obtiene toda la banda de frecuencias de análisis.

El escalar de forma diádica hace que la señal escalada sea un elemento que pertenece a un subespacio mayor que el de la señal normal. Si se genera una secuencia de funciones, siéndo estas proyecciones de la señal original sobre los distintos subespacios, la convergencia de la misma será la señal original. A partir de estas dos condiciones se puede entender cómo se genera el espacio vectorial que contiene las señales en las distintas sub bandas.

Si se asume que los subespacios V_m se encuentran abarcados por las versiones escaladas y desplazadas en el tiempo de la función $\phi(t)$:

$$V_m = \text{span}\{\phi(2^{-m}t - n), n \in Z\} \quad (5-29)$$

Por lo que, las señales de sub banda $x_m(t) \in V_m$ son expresadas como:

$$x_m(t) = \sum_{n=-\infty}^{\infty} c_m(n)\phi_{mn}(t) \quad (5-29)$$

Con:

$$\phi_{mn}(t) = 2^{-\frac{m}{2}}\phi(2^{-m}t - n) \quad (5-30)$$

Donde la función $\phi(t)$ es conocida como la función de escala.

5.6 Análisis de bancos de filtros con Q constante

El problema de un espaciamiento igualitario en las bandas de frecuencia de la FT ha llevado a una variedad de análisis de transformadas de bancos de filtros de constante Q. Éstas han sido usadas en las investigaciones de audio desde finales de los años 70's. Ejemplos de estas transformaciones son la transformada de auditorio y la transformada de frecuencia delimitada por Q. De igual manera la transformada *wavelet* puede ser clasificada como una técnica con Q constante.

Q puede ser visto como el cociente del ancho de banda y la frecuencia central del filtro. Así que aumentado la frecuencia, el ancho de banda se vuelve más grande en este análisis. Las bandas de análisis son delgadas para bajas frecuencias y anchas para altas frecuencias. La FT, sin embargo, puede ser clasificada como una transformada con ancho de banda constante.

La longitud de la ventana de análisis es de igual manera proporcional a la frecuencia que se analiza: ventanas grandes son usadas para analizar bajas frecuencias, ventanas pequeñas para altas frecuencias. De esta manera, el principio de incertidumbre no es evitado, pero es usado efectivamente. El análisis de Q constante intercambia tiempo contra la resolución de frecuencia "dentro" de la transformada: incerteza temporal pero alta resolución en bajas octavas (bandas de análisis estrechas) y alta resolución temporal con baja resolución de frecuencia en altas octavas.

Como los transientes tienden a ser cortos y poseer componentes de alta frecuencia, este esquema de constante Q permite una buena localización de los eventos.

El oído posee una respuesta en frecuencia similar a una de constante Q, especialmente por encima de 500Hz; el sistema auditivo realiza un tipo de análisis de banco de filtros con anchos de banda dependientes de la frecuencia. Estas bandas son llamadas bandas críticas.

El análisis de filtros de constante Q puede ser realizado al aplicar varios filtros paso bajo (y opcionalmente filtros paso alto) sucesivamente a la señal, o al aplicar varios filtros pasa banda a la misma señal. Otras aproximaciones existen, por ejemplo basados en algoritmos de FFT. Mientras que bancos de filtros de Q constante son típicamente menos eficientes en su cálculo, no necesitan tanto cálculo: por ejemplo, para analizar 4 octavas con una resolución de media nota (12 medias notas por octava), un análisis de constante Q necesitaría 48 bandas (cada una cubriendo la banda de frecuencia de cada media nota), mientras que el análisis de Fourier necesita aproximadamente 200 bandas.

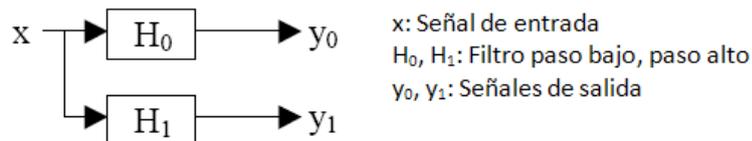
5.7 Bancos de filtros

Un banco de filtros es un conjunto de filtros, que separan las componentes de frecuencia de la señal en diferentes señales, cada una con un subgrupo de frecuencias. La combinación de las bandas de paso de los filtros cubre todo el rango de frecuencias, así que los filtros son complementarios. Un banco de filtros simple consiste de un filtro paso bajo y un filtro paso alto, cada uno con una frecuencia de corte a un medio del ancho banda.

Aplicando este banco de filtros a la señal resultan dos nuevas señales, una con la mitad de las frecuencias bajas y la otra con la mitad de las frecuencias altas. La FT puede ser considerada un banco de filtros especial: divide la señal en muchas señales sinusoidales.

Usualmente filtros paso bajo y paso alto son usados. Un diagrama de bloques de este banco de filtros se vería así:

Figura 53. Banco de filtros simple.

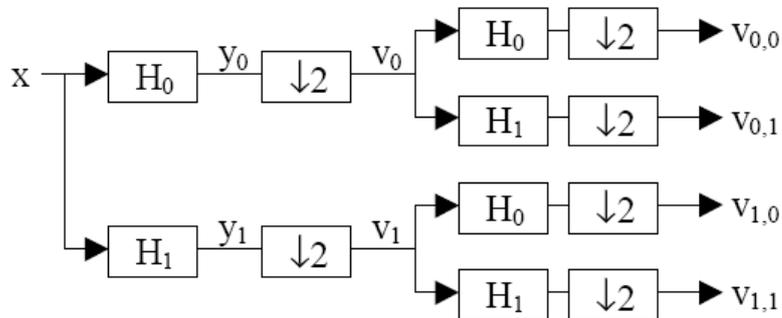


Para construir un banco de filtros con más de 2 bandas de frecuencia, y_0 puede ser filtrado nuevamente por 2 filtros, un filtro pasa bajo y uno paso alto que dividirán las bandas de nuevo en 2 bandas.

Sin embargo, es posible separar aún más las bandas de frecuencia usando los mismos filtros paso bajo y paso alto. El ancho de banda de y_0 y y_1 es de un medio del ancho de banda de la señal original. Siguiendo el teorema del muestreo, estas pueden ser representadas exactamente con la mitad del número de muestras. Esto es hecho de forma exacta con decimadores. Ellos reducen la señal a la mitad de sus muestras, al tomar cada 2^{da} muestra. Esto se denomina reducción de muestreo o *downsampling*, su operador es usualmente indicado por $\downarrow 2$. El decimar resulta en una señal con la mitad del número de muestras, pero representan el mismo intervalo de tiempo que la señal original. Por lo que, la razón de muestreo es dividida también. La salida decimada puede ser filtrada de nuevo usando los mismos filtros para de nuevo dividir en contenidos de frecuencia bajos y altos.

Un banco de filtros con 4 bandas de salida puede ser construido usando este diagrama de bloques:

Figura 54. Banco de filtros de 2 canales y 4 salidas.



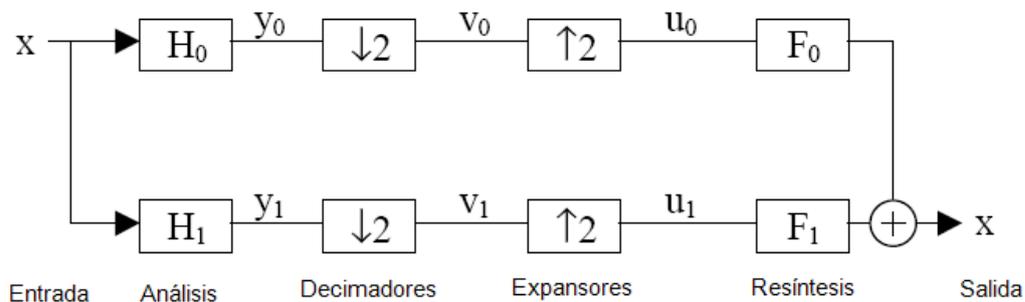
Por lo tanto, este banco de filtros trabaja en etapas sucesivas. El número de filtros por etapa es llamado canal. Es posible el crear un banco de filtros de 3 canales con bandas de paso bajo, paso alto y pasa banda (para un rango de frecuencias de un filtro paso bajo y paso alto). En general, se refiere a un banco de filtros de M -canales. Cuando las bandas de frecuencia son distribuidas equitativamente, se pueden usar decimadores para reducir el muestreo, indicado por $\downarrow M$. Un M -decimador toma cada M –ésima muestra y descarta el resto.

5.7.1 Reconstrucción perfecta

Bajo ciertas condiciones, un banco de filtros es reversible, con lo que la señal de entrada puede ser obtenida de las bandas. La reconstrucción es bastante útil, el banco de filtros se convierte en un par transformada/inversa directo. Al final, depende de los filtros si una reconstrucción perfecta es posible. Para la reconstrucción, sobre muestreo o *upsampling* debe de ser realizado para eliminar el efecto de la decimación.

Esto es hecho al insertar un cero luego de cada muestra. Adicionalmente, 2 filtros de resíntesis F_0 y F_1 son necesarios para suavizar los ceros, revirtiendo los filtros de análisis paso bajo y paso alto. Las muestras resultantes son obtenidas al añadir las salidas de los filtros de resíntesis. La figura 55 muestra un banco de filtros de análisis de 2 canales, seguido de un banco de filtros de resíntesis:

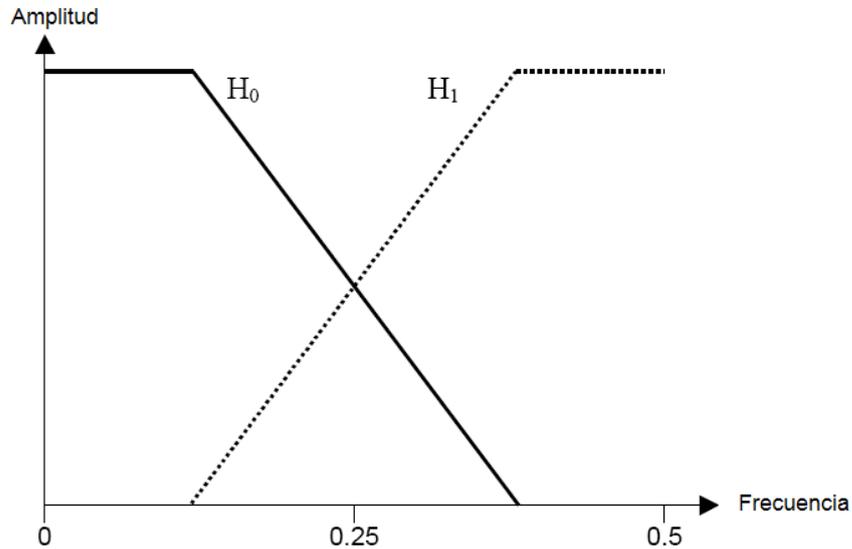
Figura 55. Banco de filtros de análisis y resíntesis.



Existen muchos aspectos para obtener una reconstrucción apropiada para un banco de filtros. A continuación se presenta una selección.

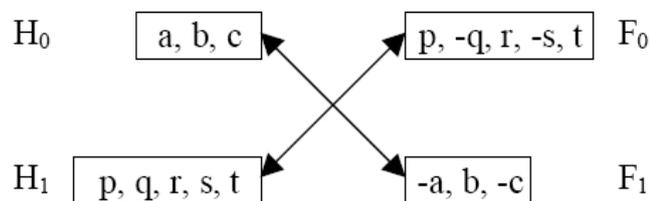
Como los filtros digitales no tienen un corte ideal, la respuesta en frecuencia de los filtros paso bajo y paso alto se traslapan: el paso bajo deja pasar componentes de frecuencia de la banda alta de paso, de igual manera, el filtro paso alto deja pasar frecuencias bajas (ver figura 56). Este aspecto, causa *aliasing* cuando se da una reducción del muestreo. La solución para una reconstrucción perfecta es el de diseñar los filtros de reconstrucción F_0 y F_1 de tal manera que puedan cancelar el efecto de *aliasing* de los filtros de análisis.

Figura 56. Traslape de frecuencias de filtros paso alto y paso bajo.



Como F_0 y F_1 dependen fuertemente en los filtros de análisis, es conveniente el calcularlos de forma directa a partir de H_0 y H_1 . Existe una manera sencilla de calcular F_0 y F_1 usando el patrón de signos alternantes. F_0 es derivado a partir de H_1 al cambiar el signo de cada segundo coeficiente del filtro partiendo del segundo coeficiente. F_1 es calculado de forma análoga, pero el cambio de signo comienza con el primer coeficiente del filtro.

Figura 57. Patrón alternante de signos.

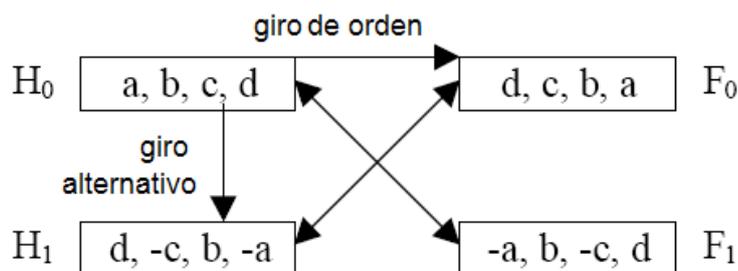


La figura 57 muestra un ejemplo de cómo construir F_0 y F_1 : a, b, c y p, q, r, s, t son los coeficientes de H_0 y H_1 , respectivamente. Obviamente, obtener los filtros de análisis a partir de los filtros de resíntesis funciona igualmente bien.

Los filtros de análisis dependen uno del otro. Como se dijo anteriormente, la respuesta en frecuencia debe ser complementaria. Para obtener una reconstrucción perfecta en conjunción con los filtros de reconstrucción, se deben cumplir más condiciones. El poder alcanzar las condiciones perfectas de reconstrucción y explicarlas sobrepasa la finalidad de esta tesis por lo que no se hará énfasis a la parte matemática tras este proceso.

Sin embargo, existe un método para derivar H_1 a partir de H_0 cuando ambos filtros deben tener la misma longitud. Es denominado giro alternativo o *alternating flip* y se muestra en la figura 58.

Figura 58. Patrón de giro alternativo.



Los coeficientes de H_1 están invertidos y cambiados de signo con H_0 . Aún así, las condiciones para la perfecta reconstrucción y la cancelación de *aliasing* son impuestas en los filtros. Entonces, los filtros llevan a un banco de filtros ortogonal. Se dice que es ortogonal, ya que la matriz de convolución es ortogonal, es decir la transpuesta de esta matriz es la matriz inversa.

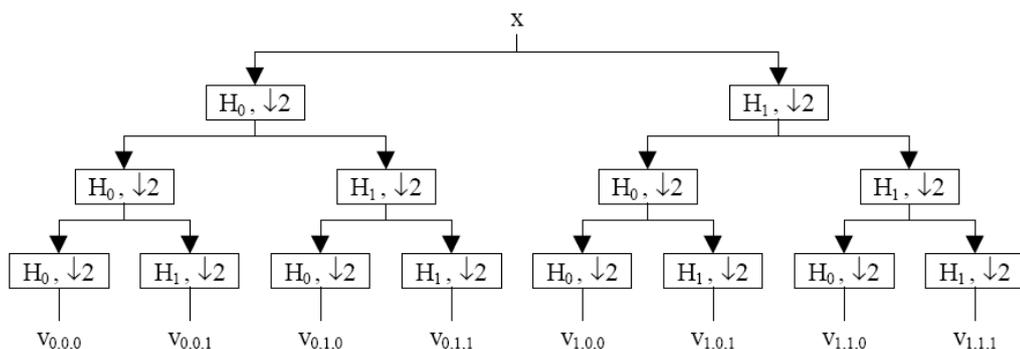
Existe otra forma más sencilla de obtener H_1 : y es a partir de H_0 usando el patrón de signos alternativos. Estos filtros son llamados bancos de filtros espejo en cuadratura o *quadrature mirror filters*, QMF.

La ortogonalidad no es requerida para una perfecta reconstrucción. El requisito mínimo es biortogonalidad de la matriz de transformaciones y su contraparte inversa. Por lo que, ortogonalidad es un caso especial donde un banco de filtros es biortogonal a sí mismo. Bancos de filtros biortogonales no necesariamente tienen la misma longitud para H_0 y H_1 .

5.7.2 Descomposición en paquetes

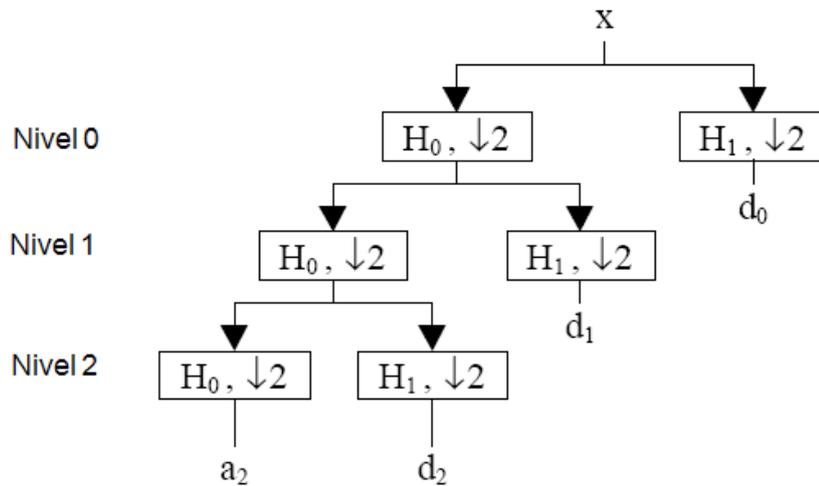
Los bancos de filtros *wavelet* son perfectos bancos de filtros de reconstrucción. Ellos aparecen en arboles de filtros: en el primer nivel, el espectro de frecuencia es dividido en la mitad baja y alta. Luego de reducir el muestreo, estas pueden ser divididas de nuevo, a un nivel específico. El modelo de paquetes de arboles *wavelet*, *wavelet packet tree*, se muestra en la figura 54: cada salida del filtro de reducción de muestreo es dividida de nuevo en 2 señales. La figura 59 muestra una descomposición de 3 niveles.

Figura 59. Paquetes de arboles *wavelet*.



El segundo método es el árbol *wavelet*. Aquí, la salida del filtro paso alto ya no es separada. Es llamado algoritmo piramidal o algoritmo de *Mallat*.

Figura 60. Árbol *wavelet*.



El árbol *wavelet* tiene algunas propiedades importantes. Las señales producidas en su salida son denominadas detalles d y aproximaciones a , son referidos como los coeficientes *wavelet*. Las aproximaciones son las salidas decimadas del filtro paso bajo; de igual manera los detalles vienen del filtrado paso alto. En cada nivel, las aproximaciones son separadas aún más, y solamente la aproximación del último nivel es retenida. El número de niveles determina el conjunto de coeficientes de detalle resultantes. Los detalles del nivel 0, d_0 , poseen la mitad del número de coeficientes como el número de muestras de la señal original, debido a la decimación. Consecuentemente, d_1 tiene un cuarto del número de coeficientes, y así sucesivamente.

Cuando se ve el significado de los coeficientes, parecerá obvio que d_0 contiene la mitad de las frecuencias superiores de la señal original, d_1 contiene el rango de frecuencias desde un cuarto a un medio de las frecuencias, etc.

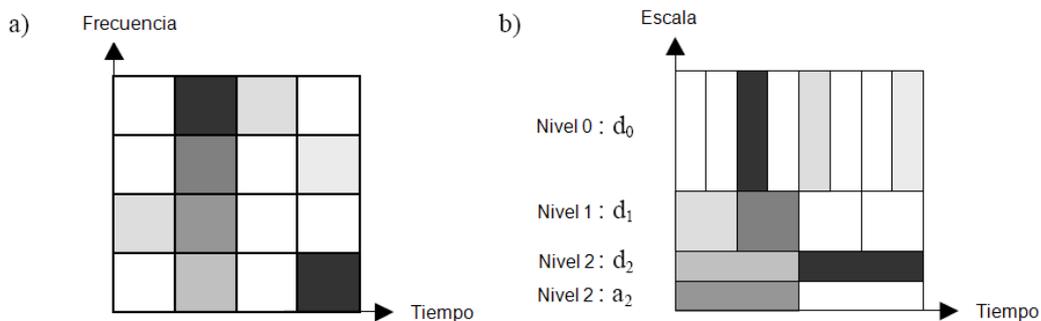
Como estas son octavas, el árbol wavelet divide de forma efectiva la señal en octavas. La aproximación del último nivel contiene las restantes frecuencias bajas. Por ejemplo, si se tiene una frecuencia de muestreo de 40kHz, una descomposición de 3 niveles (como el de la figura 53) da origen a d_0 con frecuencias desde 10kHz a 20kHz, d_1 desde 5kHz a 10kHz, d_2 desde 2.5kHz a 5kHz, d_3 desde 1.25kHz a 2.5kHz y aproximaciones con un rango de 0Hz a 1.25kHz.

Considerando la resolución en el tiempo de los coeficientes se puede confirmar que es una transformada de constante Q. Cada detalle posee un medio de la resolución de tiempo que el anterior nivel de detalle. Los coeficientes d_0 poseen un resolución alta en el tiempo, la mitad de la señal original. Bajando más en el árbol, las localizaciones en el tiempo bajan, pero, por otra parte, la banda de frecuencia se hace menor, resultando en una mejor resolución en frecuencia. Esta es la transformada *wavelet* discreta, DWT.

En la literatura, la dimensión de los niveles no es llamada "frecuencia"; por el contrario, la dimensión es llamada escala. La salida de la WT es entonces, un dominio tiempo-escala. Los coeficientes *wavelet* pueden ser representados en una red de tiempo-escala. La figura 61 muestra una comparación de la representación de la STFT y la WT.

En este diagrama cada caja representa un coeficiente de la salida de la transformada. Debido a la distinta resolución en el tiempo y en la frecuencia, las cajas de cada coeficiente no tienen el mismo alto y ancho para la transformada *wavelet*. El color de las cajas representa la magnitud del coeficiente.

Figura 61. Comparación STFT y WT.



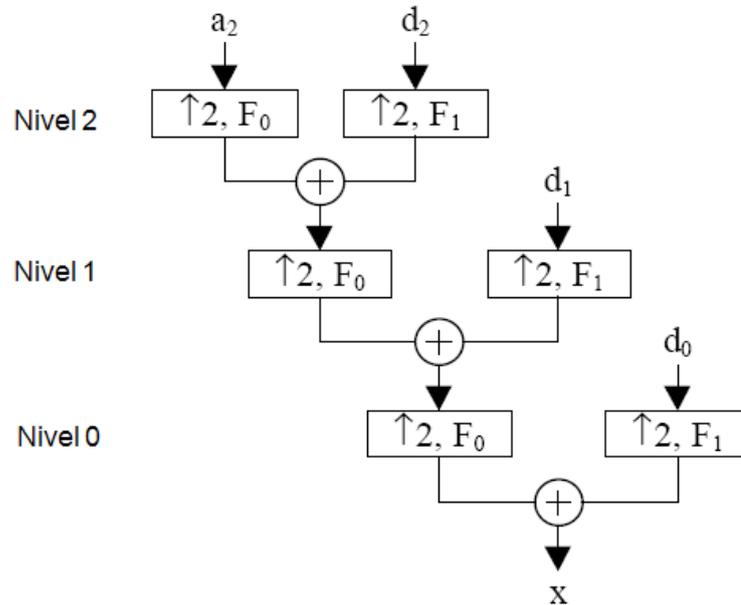
Puede verse que la señal posee un pequeño transiente en la tercera caja de los coeficientes de nivel 0. Usando la STFT, este transiente no puede ser localizado muy bien en el tiempo. Por otra parte, el sonido que lleva el 2^{do} coeficiente del nivel 2, puede analizarse considerablemente mejor por su contenido de frecuencia que la STFT. Sin embargo, la WT no puede localizarlo precisamente en el tiempo.

De forma general, el número de niveles no está limitado, excepto por la longitud de la señal. Como cada nivel trabaja en la mitad del número de muestras, se llegará un momento en el que no habrá ninguna muestra para realizar una descomposición.

5.7.3 Resíntesis wavelet

Como los filtros de un banco de filtros *wavelet*, WFB, necesitan ser adecuados para una perfecta reconstrucción se deben cumplir las condiciones anteriormente especificadas. Los coeficientes de detalle y aproximación son sobre muestreados y filtrados por un filtro de resíntesis. La suma de estos da origen a los coeficientes del próximo nivel. Este proceso es repetido hasta que se alcanza el nivel 0.

Figura 62. Resíntesis *wavelet* de 3 niveles.



Para que exista una resíntesis completa, no todos los coeficientes son necesarios. Por lo que, los paquetes de arboles *wavelet* normalmente no siempre se dividen en cada filtro paso alto y paso bajo, sino en algunos seleccionados.

5.8 Relación entre los bancos de filtros y la transformada *wavelet*

Normalmente en la práctica las funciones *wavelet* no son usadas para el cálculo de la DWT. Incluso, no son el punto de partida para desarrollar una *wavelet*. Más aún, son derivadas a partir de los filtros paso bajo y paso alto del correspondiente banco de filtros de reconstrucción.

A partir de la función de escalado se puede obtener la ecuación de dilatación:

$$\phi(t) = \sqrt{2} \sum_{k=0}^N h_0(k) \phi(2t - k) \quad (5-31)$$

Siendo h_0 los coeficientes del filtro paso bajo y N el número de coeficientes del filtro.

Se puede observar que la ecuación de dilatación es recursiva sobre sí misma, por lo que no siempre existe una solución para $\phi(t)$. La función de escala es una función de tiempo continuo, pero es poco probable que sea continua; incluso puede que no sea suave e incluya discontinuidades.

Finalmente, la función *wavelet* ψ puede ser calculada a partir de la función de escala con la ecuación *wavelet*:

$$\psi(t) = \sqrt{2} \sum_{k=0}^N h_1(k) \phi(2t - k) \quad (5-32)$$

Siendo h_1 los coeficientes del filtro paso alto y N el número de coeficientes del filtro.

Se puede ver que si se conoce la función de escala, la *wavelet* madre puede ser determinada de forma directa, sin recursión alguna.

El esquema de cálculo a partir del banco de filtros posee la misma salida que la DWT, los coeficientes de detalle son exactamente los coeficientes $d_{j,k}$ calculados por la DWT. El índice de escala j se convierte en el nivel, mientras que el índice de traslación k corresponde al plano del tiempo en el banco de filtros. De ahí que, es posible calcular la DWT completamente sin las funciones *wavelet*, como es hecho con la estructura de árbol del banco de filtros.

Cada nivel del banco de filtros corresponde a una escala de la *wavelet*. Consecuentemente, cuando los filtros son conocidos, la DWT puede ser calculada exclusivamente con el banco de filtros. Como el banco de filtros usa la decimación para el escalado, se refiere a una DWT decimada.

La *wavelet* depende del filtro paso alto; por lo que es lógico que cree los coeficientes de detalle para una descomposición usando la CWT o la DWT. Por su parte, la función de escala corresponde al filtro paso bajo, así que al aplicarlo, los coeficientes de aproximación restantes pueden ser obtenidos o resíntetizados.

En la figura 63, 64 y 65 se puede ver un resumen de lo anteriormente expuesto. En la figura 63. (a) y 63 (b) se puede observar la función de escala y la función *wavelet* respectivamente.

En la figura 64 se muestran los coeficientes de los filtros de análisis y resíntesis, y en la figura 65 la respuesta en frecuencia de los filtros de análisis.

Figura 63. Wavelet tipo Daubechies 2.

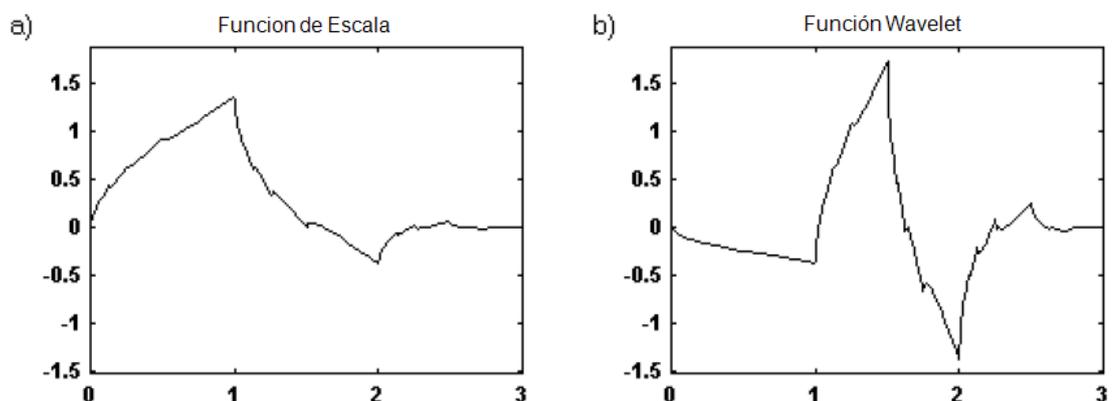


Figura 64. Coeficientes de filtro paso bajo y paso alto.

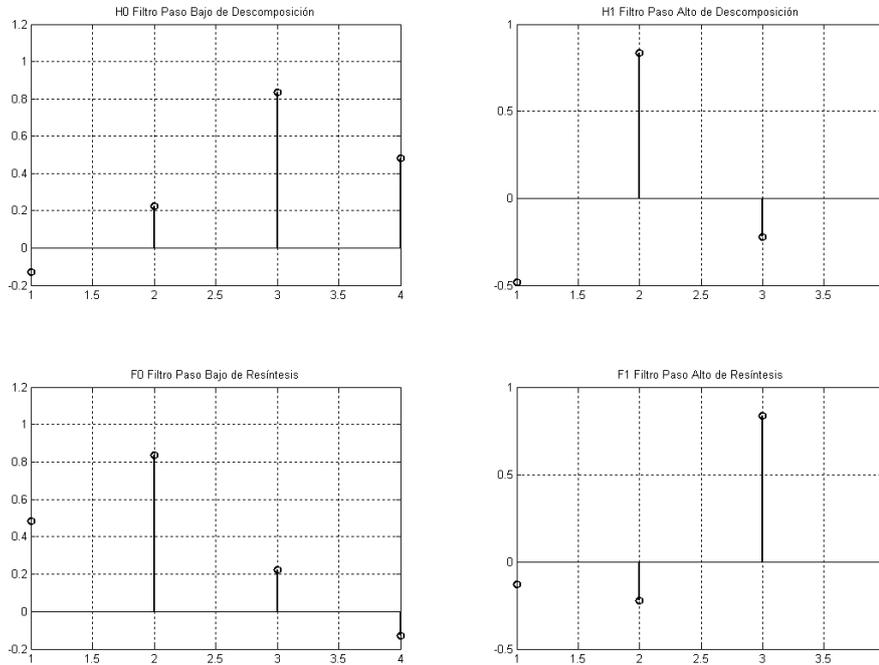
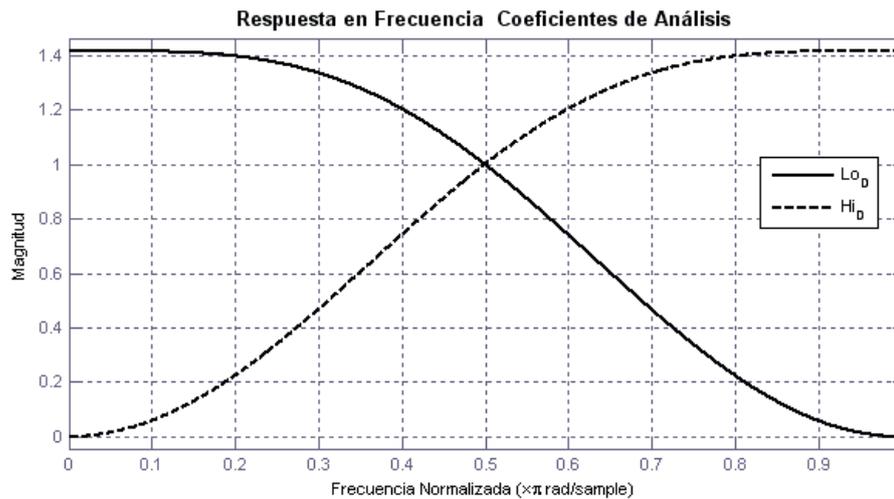


Figura 65. Respuesta en frecuencia de los filtros de análisis.



5.9 Propiedades de la transformada *wavelet*

La transformada *wavelet* es lineal. Esto significa que la transformada de la suma de dos señales es igual que la suma de sus transformadas. Además, la transformada conserva energía, es decir, la energía de la señal es igual a la energía de los coeficientes. La propiedad local de granularidad permite la localización de eventos de la señal original.

Las *wavelets* aproximan a la señal. Entonces, la forma de la *wavelet* determina la precisión. Un aspecto importante para la precisión es el número de momentos de desaparición, *vanishing moments*, de la función *wavelet*. Una *wavelet* con N momentos se puede aproximar a un polinomio de orden $N - 1$. Por lo que, entre mayor sea la cantidad de estos momentos, mayor “concentrada” la forma en que los coeficientes *wavelet* describen la señal.

La clasificación de las transformadas *wavelet* se pueden dar en:

1. Dependiendo de la existencia de la función de escala. Esto es cierto cuando el filtro paso bajo es conocido y la ecuación de dilatación tiene una solución.
2. Dependiendo si el filtro existe (y si el esquema de cálculo del banco de filtros existe). Algunas *wavelets* son especificadas como una función para la función *wavelet*, y no todas ellas pueden ser expresadas como filtros.
3. Ortogonabilidad. Un banco de filtros puede ser ortogonal o biortogonal.

4. Simetría. Funciones *wavelet* y de escalado simétricas llevan a sus respectivos filtros simétricos. *Wavelets* ortogonales compactas no pueden ser simétricas en ambos filtros, paso bajo y paso alto, como lo demuestra la construcción del patrón alternante de signos. Sin embargo, *wavelets* ortogonales pueden ser anti simétricas. Los filtros simétricos y anti simétricos son de fase lineal.
5. Soporte compacto. Funciones escaladas compactas soportadas y funciones *wavelet* conllevan a filtros de duración finita, FIR. *Wavelets* no compactas deberían de tener una caída rápida, por lo que los filtros FIR pueden ser bien aproximados.
6. Suavidad. Para muchas aplicaciones de procesamiento de señales, la suavidad es importante, al variar los coeficientes *wavelet* la señal resultante será una señal suave. La suavidad está relacionada con la regularidad de la función *wavelet*.

5.10 Aplicaciones de las *wavelet*

Las *wavelets* son usadas de forma satisfactoria en el procesado de señales. Hoy en día se encuentran bien establecidos algoritmos *wavelet* en el procesado digital de imágenes. La habilidad de la WT de extraer las características fundamentales (más importantes para el ojo) resultan en una alta compresión sin perder mucha calidad. La calidad de compresión ha mostrado ser superopr que la compresión usual JPEG, la cual se basa en la FT. El FBI adoptó la compresión *wavelet* para su archivo de imágenes de huellas dígitalas. De igual manera, para la compresión de video es usada la transformada *wavelet*, ejemplos de estas compresiones son el codec Dirac, Pixlet y Blink que se enfocan en la alta definición de los videos.

La reducción de ruido funciona muy bien; ya que los coeficientes bajos tienden a contener ruido en un espectro amplio de forma no correlacionada. Si se establecen los coeficientes por debajo de cierto umbral, el ruido puede ser eliminado de la imagen. Este método funciona muy bien para señales de audio.

Otros campos del procesamiento de señales, donde la WT es eficiente, incluye la detección de singularidades o quiebres, determinar la evolución a largo plazo de la señal, y el reconocimiento de patrones.

Para el procesamiento de audio, se han realizado una variedad de experimentos. De igual manera, la compresión de sonido ha sido desarrollada de forma exitosa con buenos resultados.

Más aún, las *wavelets* pueden ser usadas en el álgebra lineal, por ejemplo, en la resolución de sistemas lineales por medio de la transformada *wavelet*; demostrando su aplicación en la multiplicación rápida de matrices.

5.10.1 La transformada *wavelet* en el procesamiento de señales de audio

Como se ha dicho anteriormente la WT provee de ciertas características especialmente útiles en el procesamiento de señales de audio. Su descomposición de múltiple resolución permite una buena localización temporal de altas frecuencias; mientras que, ofrece una buena resolución en frecuencia para bajas frecuencias.

Un evento de alta frecuencia puede ser analizado por muchas *wavelets* “rápidas”, pequeñas y de alta frecuencia. Notas bajas pueden ser analizadas con *wavelets* “lentas”, largas y de baja frecuencia.

Los transientes no estacionarios pueden ser localizados y analizados muy bien. Generalmente, esto falla con el análisis de Fourier.

La descomposición logarítmica de la banda de frecuencias imita mucho la percepción humana de las frecuencias. La WT ofrece una banda de frecuencias logarítmicamente igual (octavas), mientras que la FT posee una baja resolución logarítmica para las bajas frecuencias. La WT adopta todas las ventajas de una transformada de constante Q.

Para el procesado en tiempo real, la WT no necesita una ventana especial para ser aplicada, ya que descompone el tiempo por sí misma. Así que presenta la ventaja de localidad.

Muchos aspectos dependen de la *wavelet* de análisis. Se necesita investigación, para determinar que *wavelet* es adecuada para cada aplicación específica. Adicionalmente, para el computo de la descomposición *wavelet*, se debe de decidir cuántas escales o niveles deben de ser calculados. La WT necesita una parametrización más considerable que la FT.

La WT puede ser calculada de forma eficiente con el algoritmo del banco de filtros piramidal. Aunque su complejidad hace que su computo consuma más tiempo que el cálculo de la FFT. Sin embargo, es lo suficientemente rápida para el análisis en tiempo real y la resíntesis de datos de audio.

La WT crea bandas de frecuencia de octavas amplias. Así que un análisis fino no puede ser posible. Usar más bandas por nivel o escala podría ser una solución, aunque no ha sido investigado mucho al respecto de la eficiencia computacional en lo que respecta su cálculo. Para muchas aplicaciones, el ancho de banda por octavas es lo suficientemente fino, para algoritmos de detección de tono, una resolución en frecuencias es necesaria.

En comparación con la FT, se puede decir que la WT provee las propiedades, que son bien adaptadas para analizar y procesar señales reales de audio.

6. WAVELET HAAR

6.1 Introducción

En el capítulo anterior se presentaron las bases de la teoría wavelet y ahora se presentan estos conceptos por medio de la wavelet Haar. La wavelet Haar es posiblemente la wavelet más sencilla de estudiar, y aunque no le resta complejidad, permite una descripción matemática muy útil. Usar la wavelet Haar permitirá entender el trabajo de las wavelets, para luego implementar wavelets más complejas; mostrando así su aplicación a funciones de tiempo continuo, como también al análisis multi-resolución.

Se comenzará con el desarrollo de la teoría wavelet en tiempo continuo para la wavelet Haar, descomponiendo una función continua $f(t) \in L^2(R)$, se calcularán sus coeficientes y luego se reconstruirán en distintos grados de resolución. Posteriormente se mostrará la relación entre una función wavelet y la función de escala, y el espacio en donde esta función se genera. Por último se presenta el análisis multi-resolución y la transformada wavelet, con lo cual, se estará en la capacidad de analizar señales por medio de wavelets.

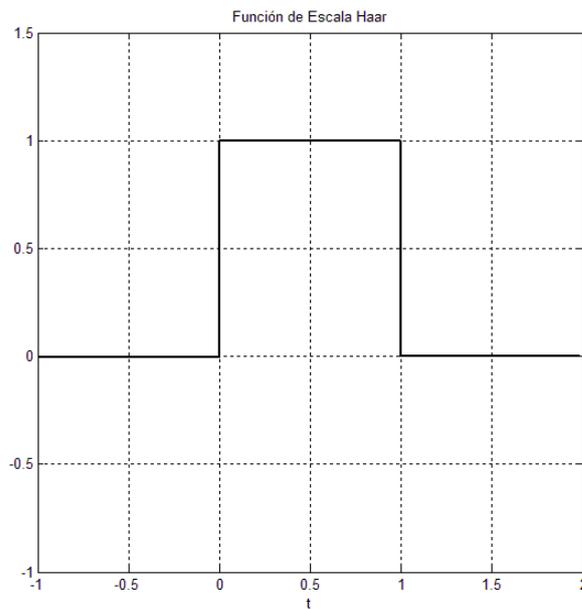
6.2 Función Haar de escala

Sea la función $\phi(t)$ perteneciente a $L^2(R)$, definida de la siguiente manera:

$$\phi(t) = \begin{cases} 1 & \text{si } 0 \leq t \leq 1 \\ 0 & \text{el resto.} \end{cases} \quad (6-1)$$

Esta función se denomina función de escalamiento, cuya gráfica se muestra en la figura 66.

Figura 66. Función de escala Haar.



Se definen entonces un conjunto de funciones de escalamiento en términos de traslaciones enteras de la función básica de escalamiento $\phi(t)$:

$$\phi_k(t) = \phi(t - k) = \begin{cases} 1 & \text{si } t_1 = k \leq t \leq k + 1 = t_2 \\ 0 & \text{el resto.} \end{cases} \quad (6-2)$$

$$k \in Z \wedge \phi \in L^2(R)$$

El sub espacio de $L^2(R)$ generado por esta función es definido como

$$v_0 = \overline{\text{Span}_k\{\phi_k(t)\}} \quad (6-3)$$

para todos los enteros k desde $-\infty$ a ∞ . El superrayado denota clausura, esto significa que

$$f(t) = \sum_k c_k \phi_k(t) \quad k \in Z \quad (6-4)$$

para cualquier $f(t) \in V_0 \in L^2(\mathbb{R})$.

Es decir, cualquier función $f(t)$ que esté en V_0 puede ser representada por una combinación lineal del conjunto de funciones $\phi_k(t)$ con sus respectivos coeficientes c_k . Para un rápido cálculo de estos coeficientes es necesario que $\phi_k(t)$ sea ortonormal.

Ejemplo 6.1

Representar la función $f(t) = t^2$ con $t \in [-3,3]$, mediante la función de escalamiento $\phi_k(t)$ con $t \in [-3,2]$

Solución

Representaremos la función continua $f(t) = t^2$, como una combinación lineal de la forma:

$$f(t) = c_{-3}\phi(t+3) + c_{-2}\phi(t+2) + c_{-1}\phi(t+1) + c_{-0}\phi(t) + c_1\phi(t-1) + c_2\phi(t-2) \quad (6-5)$$

Para calcular el coeficiente c_{-3} se hará la multiplicación vectorial de $\phi(t+3)$ por (6-5)

$$t^2 = [c_{-3}\phi(t+3) + c_{-2}\phi(t+2) + c_{-1}\phi(t+1) + c_{-0}\phi(t) + c_1\phi(t-1) + c_2\phi(t-2)] \odot \phi(t+3) \quad (6-6)$$

$$\langle t^2, \phi(t+3) \rangle = \langle c_{-3}\phi(t+3), \phi(t+3) \rangle + \cdots + \langle c_2\phi(t-2), \phi(t+3) \rangle \quad (6-7)$$

Como se mencionó anteriormente, $\phi_k(t)$ debe ser ortonormal, por lo que es ortogonal y todos los productos de la derecha de (6-7) serán cero, excepto $\langle c_{-3}\phi(t+3), \phi(t+3) \rangle$

$$\langle t^2, \phi(t+3) \rangle = \langle c_{-3}\phi(t+3), \phi(t+3) \rangle \quad (6-8)$$

$$\langle t^2, \phi(t+3) \rangle = \int_{-3}^{-3+1} c_{-3}\phi(t+3)\phi(t+3)dt \quad (6-9)$$

donde los límites de integración fueron calculados a partir de (6-2).

Resolviendo la integral de (6-9) se obtiene:

$$\langle t^2, \phi(t+3) \rangle = c_{-3} \quad (6-10)$$

El hecho de que en el lado derecho solo resulte c_{-3} , significa que el conjunto de funciones $\phi_k(t)$ son normales sobre sí mismas. Con lo cual se cumple lo que anteriormente se había mencionado acerca de que el conjunto de funciones $\phi_k(t)$.

Ahora, si se trabaja sobre el lado izquierdo de (6-10) se debe resolver:

$$c_{-3} = \int_{-3}^3 t^2\phi(t+3) dt \quad (6-11)$$

$$c_{-3} = \int_{-3}^{-2} t^2\phi(t+3) dt + \int_{-2}^3 t^2\phi(t+3) dt$$

Como $\phi(t + 3)$ se encuentra definida de $[-3, -2]$

$$c_{-3} = \int_{-3}^{-2} t^2 dt + \int_{-2}^3 t^2 \times 0 dt \quad (6-12)$$

$$c_{-3} = 6.\overline{33}$$

Ahora se procederá a calcular c_{-2}

$$t^2 = [c_{-3}\phi(t + 3) + c_{-2}\phi(t + 2) + c_{-1}\phi(t + 1) + c_{-0}\phi(t) + c_1\phi(t - 1) + c_2\phi(t - 2)] \odot \phi(t + 2) \quad (6-13)$$

$$\langle t^2, \phi(t + 2) \rangle = \langle c_{-3}\phi(t + 3), \phi(t + 2) \rangle + \dots + \langle c_2\phi(t - 2), \phi(t + 2) \rangle \quad (6-14)$$

De igual manera, como se determinó para c_{-3} , $\phi_k(t)$ es ortogonal, por lo que todos los productos punto de la parte derecha de (6-14) serán cero, excepto $\langle c_{-2}\phi(t + 2), \phi(t + 2) \rangle$.

$$\langle t^2, \phi(t + 2) \rangle = \langle c_{-2}\phi(t + 2), \phi(t + 2) \rangle \quad (6-15)$$

Por la normalidad de las funciones, el lado derecho de (6-15) será solamente c_{-2} , por lo que se puede reducir a:

$$\langle t^2, \phi(t + 2) \rangle = c_{-2} \quad (6-16)$$

$$c_{-2} = \int_{-3}^3 t^2 \phi(t + 2) dt \quad (6-17)$$

La función $\phi(t + 2)$ se encuentra definida solamente para $[-2, -1]$, por lo que la ecuación (6-17) se puede expresar como:

$$c_{-2} = \int_{-3}^{-2} t^2 \phi(t + 2) dt + \int_{-2}^{-1} t^2 \phi(t + 2) dt + \int_{-1}^3 t^2 \phi(t + 2) dt \quad (6-18)$$

$$c_{-2} = \int_{-3}^{-2} t^2 \times 0 dt + \int_{-2}^{-1} t^2 dt + \int_{-1}^3 t^2 \times 0 dt$$

$$c_{-2} = 2.\overline{33}$$

Del cálculo de estos dos coeficientes se puede observar que cualquier coeficiente c_k puede ser calculado como:

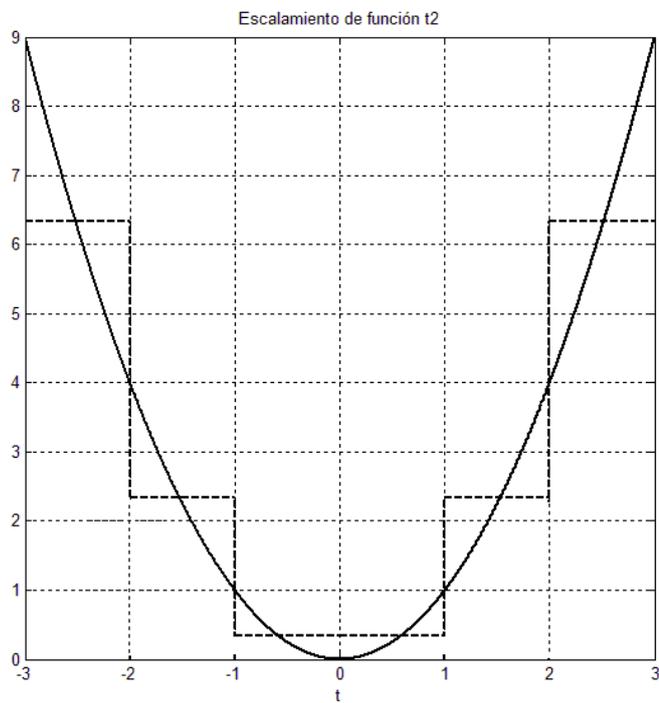
$$c_k = \langle f(t), \phi_k(t) \rangle = \int_{t_1}^{t_2} f(t) dt \quad (6-19)$$

Por lo que siempre que $f(t) \in L^2(R)$ y $\phi_k(t)$ sea ortonormal. Se podrán calcular los coeficientes restantes usando (6-19). La tabla VI muestra el cálculo de los seis coeficientes necesarios. Con estos seis coeficientes se puede reconstruir $f(t) = t^2$ mediante la ecuación (6-5), que proporciona la gráfica mostrada en la figura 67.

Tabla VI. Coeficientes de reconstrucción para la función $f(t) = t^2$ usando la función de escalamiento en v_0 .

Coeficiente	Valor Obtenido
c_{-3}	6.33
c_{-2}	2.33
c_{-1}	0.33
c_0	0.33
c_1	2.33
c_2	6.33

Figura 67. Escalado de función $f(t) = t^2$.



Es evidente que la representación no es lo suficientemente adecuada solamente con estos coeficientes por lo que, para una mejor representación de cualquier señal perteneciente a $L^2(\mathbb{R})$ por medio de una función de escalamiento, debemos ser capaces de crear una nueva familia de funciones, que además de tener las capacidades de trasladarse, puedan también ser escaladas para lograr así una mejor resolución.

Definiendo un conjunto de funciones de escalamiento

$$\phi_{j,k}(t) = \phi(2^j t - k) = \begin{cases} 1 & \text{si } t_1 = \frac{k}{2^j} \leq t \leq \frac{k+1}{2^j} = t_2 \\ 0 & \text{el resto.} \end{cases} \quad (6-20)$$

con $\phi_{j,k}(t) \in V_j$, que se define como

$$V_j = \overline{\text{Span}_k\{\phi_k(2^j t)\}} = \overline{\text{Span}_k\{\phi_{j,k}(t)\}}$$

donde $\phi_{j,k}(t)$ es la función básica de escalamiento trasladada y escalada, y V_j es un sub espacio de $L^2(\mathbb{R})$ generado por (6-20).

Ejemplo 6.2

Representar la función $f(t) = t^2$ con $t \in [-3,3]$, mediante la familia de funciones $\phi_{j,k}(t)$.

Solución

Para representar esta función trabajaremos en el espacio v_1 , o lo que es lo mismo, $j = 1$. Por lo que la función $f(t)$ se puede expresar como la siguiente combinación de funciones

$$t^2 = \sum_{k=-6}^5 c_k \phi(2t - k) \quad (6-21)$$

y trabajando de manera análoga al ejemplo 1, usando las ecuaciones (6-6) a (6-9), se llega al siguiente resultado para el coeficiente c_{-6}

$$\langle t^2, \phi(2t + 6) \rangle = \langle c_{-6} \phi(2t + 6), \phi(2t + 6) \rangle$$

$$\langle c_{-6} \phi(2t + 6), \phi(2t + 6) \rangle = \int_{-3}^{-2.5} c_{-6} dt$$

donde los límites se encuentran a partir de (6-20)

$$\langle c_{-6} \phi(2t + 6), \phi(2t + 6) \rangle = \frac{c_{-6}}{2}$$

$$\frac{c_{-6}}{2} = \langle t^2, \phi(2t + 6) \rangle$$

$$c_{-6} = \int_{-3}^{-2.5} t^2 dt = 7.58\bar{3}$$

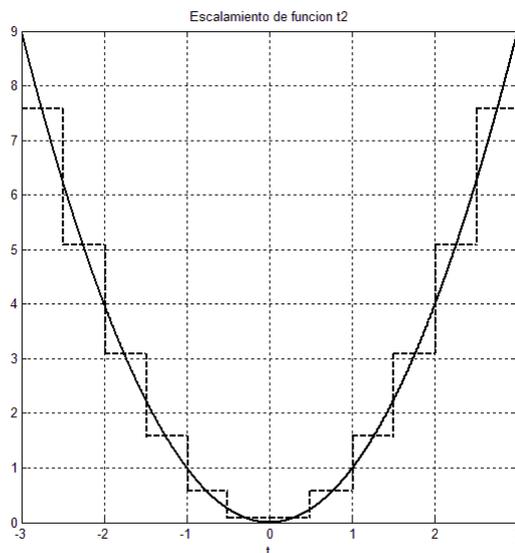
En la tabla VII se muestra el resto de coeficientes, así mismo, en la figura 68 se muestra la reconstrucción de $f(t)$.

En esta figura se presenta una clara mejoría en la resolución de la representación de la función cuadrática, debido que $\phi_{j,k}(t)$ en V_1 está definida en un intervalo de tiempo más pequeño que V_0 .

Tabla VII. Coeficientes de reconstrucción para la función $f(t) = t^2$ usando la función de escalamiento en v_1 .

Coeficiente	Valor Obtenido
c_{-6}	7.58
c_{-5}	5.08
c_{-4}	3.08
c_{-3}	1.58
c_{-2}	0.58
c_{-1}	0.08
c_0	0.08
c_1	0.58
c_2	1.58
c_3	3.08
c_4	5.08
c_5	7.58

Figura 68. Escalado de función $f(t) = t^2$ usando la función de escalamiento en v_1 .



6.2.1 Relación de ortogonalidad y normalidad

En el ejemplo anterior se puede observar que para V_1 , los coeficientes quedaban divididos por 2, lo que indica que las funciones de escalamiento en este espacio no son ortonormales. Se puede ver que esto no ocurre solamente en V_1 , sino que en todo V_j .

$$\langle \phi_{j,k}(t), \phi_{j,k}(t) \rangle = \int_{t_1}^{t_2} \phi_{j,k}(t) \phi_{j,k}(t) dt \quad (6-22)$$

$$\text{con } t_1 = \frac{k}{2^j} \text{ y } t_2 = \frac{k+1}{2^j}.$$

La ecuación (6-22) se reduce entonces a:

$$\langle \phi_{j,k}(t), \phi_{j,k}(t) \rangle = \frac{1}{2^j} \quad (6-23)$$

Por lo tanto debe existir una constante r_j que normalice este producto, y nos permita hacer ortonormales al conjunto de funciones $\phi_{j,k}(t)$. De tal manera que se cumpla

$$\langle r_j \phi_{j,k}(t), r_j \phi_{j,k}(t) \rangle = 1 \quad (6-24)$$

$$\langle r_j \phi_{j,k}(t), r_j \phi_{j,k}(t) \rangle = \int_{t_1}^{t_2} r_j \phi_{j,k}(t) r_j \phi_{j,k}(t) dt = \frac{r_j^2}{2^j}$$

$$\frac{r_j^2}{2^j} = 1$$

$$r_j = 2^{\frac{j}{2}}$$

Esta es la constante de normalización que permite que el conjunto de funciones $\phi_{j,k}(t)$ sea ortonormal.

La nueva familia de funciones $\phi_{j,k}(t)$ ² queda definida como

$$\phi_{j,k}(t) = 2^{\frac{j}{2}} \phi(2^j t - k) = \begin{cases} 2^{\frac{j}{2}} & \text{si } t_1 = \frac{k}{2^j} \leq t \leq \frac{1+k}{2^j} = t_2 \\ 0 & \text{el resto.} \end{cases} \quad (6-25)$$

Una propiedad importante de la función de escalamiento $\phi_{j,k}(t)$, es que puede ser representada por una combinación lineal de ella misma, trasladada y escalada. Por medio de la ecuación de dilatación se puede encontrar $\phi(t)$ que es

$$\phi(t) = \sqrt{2} \sum_{k=0}^1 h_0(k) \phi(2t - k) \quad (6-26)$$

$$\phi(t) = \sqrt{2} h_0(0) \phi(2t) + \sqrt{2} h_0(1) \phi(2t - 1)$$

Se escogen los valores que hacen que se cumpla el intervalo de tiempo en que está definida $\phi(t)$, esto es, $\phi_{1,0}(t)$ está definida en $t \in [0,0.5)$ y $\phi_{1,1}(t)$ está definida en $t \in [0.5,1)$.

Los valores de $h_0(0)$ y $h_0(1)$, siendo los coeficientes del filtro paso bajo, pueden ser obtenidos de la siguiente manera:

$$h_0(0) = \sqrt{2} \int \phi(t) \phi(2t) dt$$

$$h_0(0) = \sqrt{2} \left\{ \int_0^{1/2} 1 \times 1 dt + \int_{1/2}^1 1 \times 0 dt \right\}$$

$$h_0(0) = \frac{1}{\sqrt{2}}$$

² A partir de este momento, cuando se refiera a $\phi_{j,k}(t)$ se estará hablando de la ecuación (6-25)

$$h_0(1) = \sqrt{2} \int \phi(t)\phi(2t - 1)dt$$

$$h_0(1) = \sqrt{2} \left\{ \int_0^{1/2} 1 \times 0 dt + \int_{1/2}^1 1 \times 1 dt \right\}$$

$$h_0(1) = \frac{1}{\sqrt{2}}$$

De esta manera la función de escala puede ser representada por sí misma trasladada y escalada según

$$\phi(t) = \phi(2t) + \phi(2t - 1) \quad (6-27)$$

6.3 Función wavelet Haar

En la sección anterior se observó que es posible obtener una mejor aproximación de la señal cuando las funciones de escalamiento ocupan el espacio V_1 que utilizando la función de escalamiento que ocupa el espacio V_0 . Sin embargo, las características de una señal pueden ser mejor descritas, no incrementando el tamaño del espacio de las funciones de escalamiento, sino definiendo un espacio W_j de funciones levemente diferentes a las funciones de escalamiento, que representen la diferencia que existe entre un espacio V_j y un espacio V_{j+1} tal que:

$$V_{j+1} = V_j \oplus W_j \quad (6-28)$$

por lo que se puede decir que el espacio W_0 corresponde al complemento del espacio V_0 en el espacio V_1 .

La función que expande el espacio W_0 se conoce como función wavelet, y se define de la forma

$$\psi(t) = \begin{cases} 1 & 0 \leq t \leq \frac{1}{2} \\ -1 & \frac{1}{2} \leq t \leq 1 \\ 0 & \text{otras regiones} \end{cases} \quad (6-29)$$

para la wavelet Haar, ver figura 69, que al igual que la función de escalamiento que expande el espacio V_0 , puede ser representada en el intervalo $[0,1]$ como una combinación lineal de las funciones de escalamiento que expanden el espacio V_1 de la siguiente manera

$$\psi(t) = h_1(0)\sqrt{2}\phi(2t) + h_1(1)\sqrt{2}\phi(2t - 1) \quad (6-30)$$

por lo que al realizar el producto interno de (6-30) por $\sqrt{2}\phi(2t)$ y luego por $\sqrt{2}\phi(2t - 1)$, se obtienen los siguientes coeficientes

$$h_1(0) = \sqrt{2} \int_0^{\frac{1}{2}} \psi(t) dt \quad (6-31)$$

y

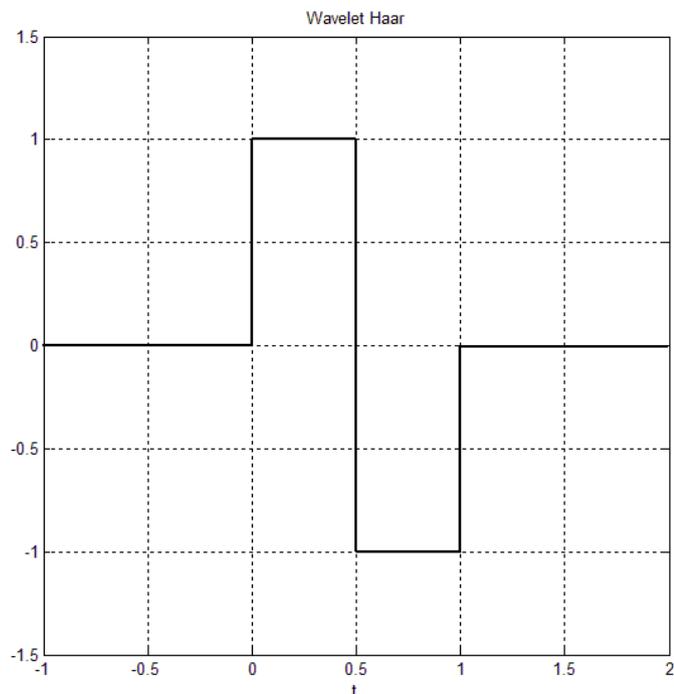
$$h_1(1) = \sqrt{2} \int_{\frac{1}{2}}^1 \psi(t) dt \quad (6-32)$$

respectivamente, y luego de integrar los coeficientes se obtiene

$$h_1(0) = -h_1(1) = \frac{1}{\sqrt{2}}$$

estos coeficientes permiten que se mantenga la normalidad de $\psi(t)$.

Figura 69. Función wavelet Haar.



La ortogonalidad de W_0 se establece del hecho de que los espacios V_0 y V_1 son ortogonales, y por lo tanto cualquier espacio con $j = 0, \pm 1, \pm 2, \dots$ también lo es, entonces el espacio W_0 al ser el complemento de V_0 en V_1 es ortogonal. Al igual que con la función escalamiento, es posible obtener una representación de la diferencia que existe entre aproximar una señal con un nivel de resolución j y aproximar la misma señal con un nivel de resolución $j + 1$, mediante el producto interno de esta señal con un set de funciones que expandan el espacio W_j , donde j será elegido de acuerdo al grado de aproximación que se desee.

La ortogonalidad de W_0 , y por consecuencia $\psi(t)$, puede ser demostrada de la siguiente manera. Definiendo a la función wavelet desplazada en el tiempo por una constante k como

$$\psi_k(t) = \begin{cases} 1 & k \leq t \leq k + \frac{1}{2} \\ -1 & k + \frac{1}{2} \leq t \leq k + 1 \\ 0 & \text{otras regiones} \end{cases} \quad (6-33)$$

y definiendo de la misma manera otra función $\psi_m(t)$ con $m > k$, entonces realizando el producto punto entre ellas de la forma

$$\langle \psi_k(t), \psi_m(t) \rangle = \int_k^{m+1} \psi_k(t) \psi_m(t) dt \quad (6-34)$$

se obtiene que

$$\langle \psi_k(t), \psi_m(t) \rangle = \int_k^{k+1} \psi_k(t) \times 0 dt + \int_m^{m+1} 0 \times \psi_m(t) dt \quad m, k \in Z \quad (6-35)$$

con lo cual se comprueba la ortogonalidad del espacio W_0 y la función wavelet.

Ejemplo 6.3

Representar la función $f(t) = t^2$ sobre el intervalo $[-3,3]$, mediante la función wavelet $\psi_k(t)$ con $t \in [-3,2]$

Solución

La función continua $f(t) = t^2$ se representará como una combinación lineal de la forma:

$$t^2 = d_{-3}\psi(t+3) + d_{-2}\psi(t+2) + d_{-1}\psi(t+1) + d_{-0}\psi(t) + d_1\psi(t-1) + d_2\psi(t-2) \quad (6-36)$$

en el espacio W_0 .

Para determinar el coeficiente d_{-3} se procede a realizar la multiplicación vectorial de (6-36) por $\psi(t + 3)$, obteniendo

$$\langle t^2, \psi(t + 3) \rangle = \langle d_{-3}\psi(t + 3), \psi(t + 3) \rangle + \dots + \langle d_2\psi(t - 2), \psi(t + 3) \rangle \quad (6-37)$$

Todos los productos de la derecha de (6-37) a excepción de $\langle d_{-3}\psi(t + 3), \psi(t + 3) \rangle$ son cero, con lo que se obtiene

$$\langle t^2, \psi(t + 3) \rangle = \langle d_{-3}\psi(t + 3), \psi(t + 3) \rangle \quad (6-38)$$

$$\int_{-3}^{-3+1} t^2 \psi(t + 3) dt = \int_{-3}^{-3+1} d_{-3} \psi(t + 3) \psi(t + 3) dt \quad (6-39)$$

Para obtener la solución de (6-39) se debe tener en cuenta los casos cuando $\psi(t + 3) = 1$

$$\int_{-3}^{-3+\frac{1}{2}} t^2 dt = \frac{d_{-3}^+}{2} \quad (6-40)$$

y cuando $\psi(t + 3) = -1$

$$\int_{-3+\frac{1}{2}}^{-3+1} -t^2 dt = \frac{d_{-3}^-}{2} \quad (6-41)$$

La solución para $\psi(t + 3) = 0$ resulta ser trivial. La resolución de las integrales (6-40) y (6-41) entrega dos coeficientes, los cuales son

$$d_{-3}^+ = \frac{91}{12} \text{ y } d_{-3}^- = -\frac{61}{12}$$

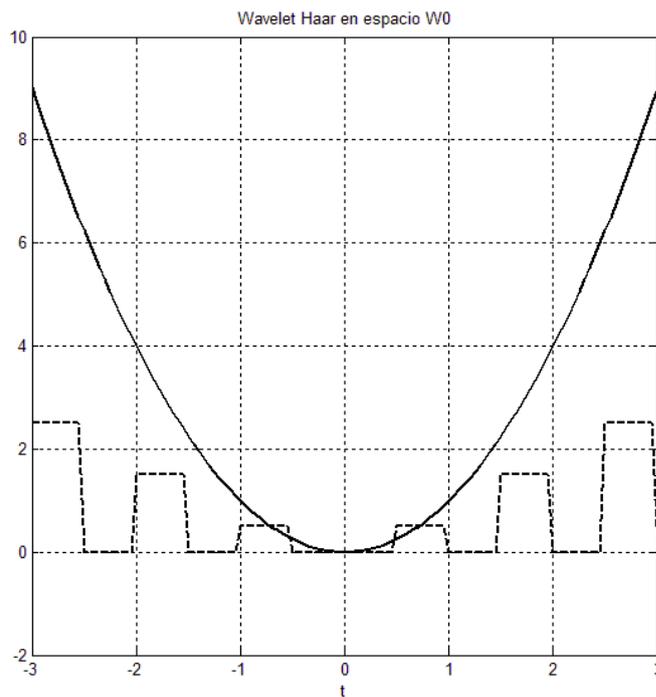
en donde el coeficiente d_{-3} resulta ser la suma de $d_{-3}^+ + d_{-3}^-$, con lo que se obtiene que $d_{-3} = 2.5$.

En la tabla VIII se muestra el resto de coeficientes obtenidos de forma análoga y en la figura 70 la gráfica de estos coeficientes.

Tabla VIII. Coeficientes para la función $f(t) = t^2$ usando la función wavelet Haar en W_0 .

Coeficiente	Valor Obtenido
d_{-3}	2.5
d_{-2}	1.5
d_{-1}	0.5
d_0	-0.5
d_1	-1.5
d_2	-2.5

Figura 70. Aproximación de la función $f(t) = t^2$ mediante la función wavelet en el espacio W_0 .



Ejemplo 6.4

Representar la función $f(t) = t^2$ con $t \in [-3,3]$, mediante la familia de funciones wavelet $\psi_{j,k}(t)$ con un valor de $j = 1$.

Solución

El representar $f(t) = t^2$ para $j = 1$, significa que la aproximación que se obtendrá de la función se encuentra en el sub espacio V_1 , esta representación de t^2 es la combinación lineal de $\psi_{1,k}(t)$ por lo que resulta

$$t^2 = d_{-6}\psi(t+6) + d_{-5}\psi(t+5) + \dots + d_5\psi(t-5) \quad (6-42)$$

Para obtener cada coeficiente se procede de forma análoga a como se hizo en (6-37), (6-38) y (6-39). El coeficiente d_{-6} es

$$\int_{-3}^{-3+\frac{1}{2}} t^2 \psi(t+6) dt = \int_{-3}^{3+\frac{1}{2}} d_{-6} \psi(t+6) \psi(t+6) dt \quad (6-43)$$

cuando el valor de $\psi(t+6) = 1$ se tiene

$$\int_{-3}^{-3+\frac{1}{4}} t^2 dt = \frac{d_{-6}^+}{4} \quad (6-44)$$

y cuando $\psi(t+6) = -1$ se tiene que es

$$\int_{-3+\frac{1}{4}}^{-3+\frac{1}{2}} t^2 dt = \frac{d_{-6}^-}{4} \quad (6-45)$$

Al resolver (6-44) y (6-45) se obtienen los respectivos valores de los coeficientes d_{-6}^+ y d_{-6}^-

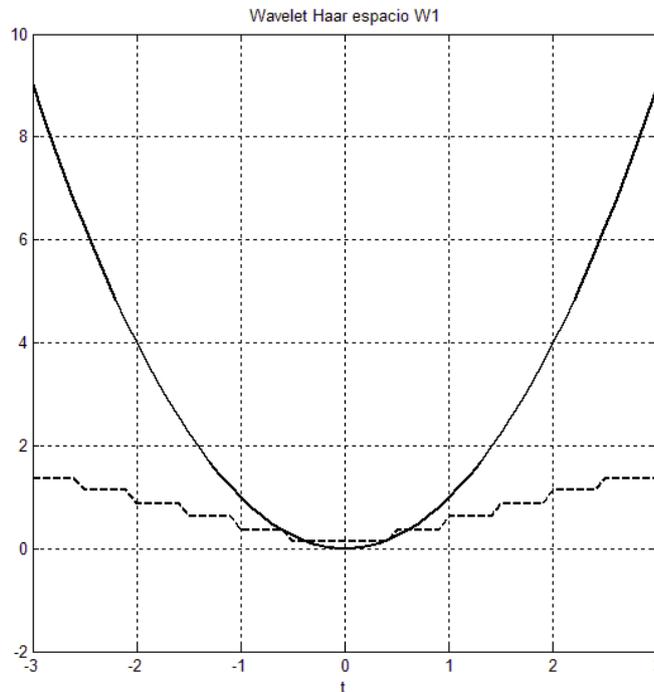
$$d_{-6}^+ = 8.27 \text{ y } d_{-6}^- = -6.9$$

La tabla IX muestra los 12 coeficientes obtenidos, y en la figura 71 se presenta la gráfica de la aproximación a partir de estos 12 coeficientes.

Tabla IX. Coeficientes para la función $f(t) = t^2$ usando la función wavelet Haar en W_1 .

Coeficiente	Valor Obtenido
d_{-6}	1.37
d_{-5}	1.13
d_{-4}	0.87
d_{-3}	0.63
d_{-2}	0.37
d_{-1}	0.13
d_0	-0.13
d_1	-0.37
d_2	-0.63
d_3	-0.87
d_4	-1.13
d_5	-1.37

Figura 71. Aproximación de la función $f(t) = t^2$ mediante la función wavelet en el espacio W_1 .



6.3.1 Relación de ortogonalidad y normalidad

De ambos ejemplos se puede ver que los coeficientes obtenidos quedaban divididos por 2 y 4 respectivamente, es claro entonces, que las wavelets para estos espacios no son ortogonales. En el apartado 6.2.1 se determinó una constante r_j que permita la ortogonalidad de la función de escalamiento, de igual manera, existe una constante r_j que permite que la función wavelet sea ortogonal.

Se esperarían entonces que se cumpliera con

$$\langle r_j \psi_{j,k}(t), r_j \psi_{j,k}(t) \rangle = 1 \quad (6-46)$$

pero a partir de (6-40), (6-41), (6-44) y (6-45) se observa que se obtiene

$$\langle \psi_0(t), \psi_0(t) \rangle = \pm \frac{1}{2} \quad (6-47)$$

$$\langle \psi_0(t), \psi_0(t) \rangle_{\pm} = \frac{1}{4}$$

por lo que en este caso $r_0 = \sqrt{2}$ y $r_1 = 2$ para que estos productos sean 1. Es de notar que r_1 se puede expresar a partir de r_0 , como $r_1 = \sqrt{2} \times \sqrt{2}$ y en forma más general, r_j a partir de r_0 puede ser escrito como

$$r_j = 2^{\frac{j}{2}} \quad (6-48)$$

A partir de (6-48) ya se puede expresar un conjunto de funciones ortonormales mediante escalamiento y traslación que sea de la forma

$$\psi_{j,k}(t) = 2^{\frac{j}{2}} \psi_{j,k}(2^j t - k) = \begin{cases} 2^{\frac{j}{2}} & \frac{k}{2^j} \leq t \leq \frac{k}{2^j} + \frac{1}{2^{j+1}} \\ -2^{\frac{j}{2}} & \frac{k}{2^j} + \frac{1}{2^{j+1}} \leq t \leq \frac{k+1}{2^j} \\ 0 & \text{otras regiones} \end{cases} \quad (6-49)$$

con $\psi_{j,k}(t) \in W_j$ que se define como

$$W_j = \overline{\text{Span}_k\{\psi_k(2^j t)\}} = \overline{\text{Span}_k\{\psi_{j,k}(t)\}} \quad (6-50)$$

Donde $\psi_{j,k}(t)$ es la función wavelet base trasladada y escalada, W_j es un sub espacio de $L^2(R)$ generado por (6-49).

6.4 Análisis multi-resolución

En el análisis multi-resolución la idea es aproximar una función $f(t)$ en distintos niveles de resolución $\{f_1(t), f_2(t), f_3(t), \dots\}$, con lo que se obtiene una descomposición multi-escala de la forma:

$$f(t) = f_0(t) + \sum_{j \geq 0} g_j(t) \quad (6-51)$$

donde cada función $g_j(t) = f_{j+1}(t) - f_j(t)$ representa el error que se obtiene al aproximar $f_{j+1}(t)$ mediante $f_j(t)$, es decir, la variación entre dos niveles sucesivos de resolución.

En el capítulo 5 se presentó la idea de la transformada wavelet discreta, DWT, y que por medio de los coeficientes de la función de escala y los de la función wavelet se puede reconstruir la señal transformada. Antes de presentar una expresión alternativa para (6-51), donde abarca el uso de la DWT, se pretende dar una explicación del espacio en donde se trabaja el análisis multi-resolución.

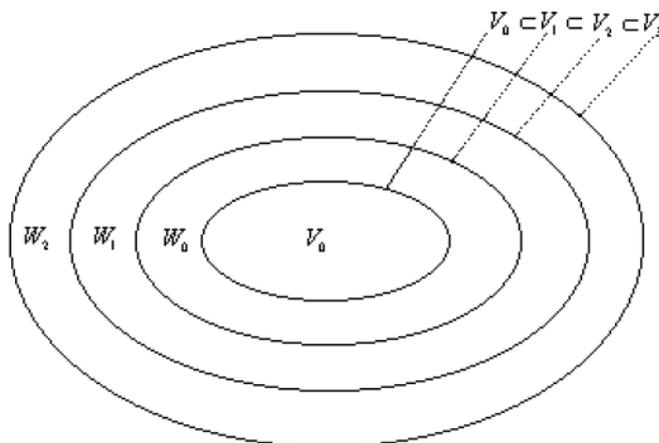
En un análisis multi-resolución se requiere que exista un anidamiento de los espacios generados por las funciones de escala. De tal manera que el espacio que contiene a las señales de más alta resolución también contenga a las de más baja resolución. Con lo cual, los elementos de un espacio son simplemente versiones escaladas de los elementos del siguiente espacio. Una forma de ver esto fue cuando se obtuvo (6-27), en donde la función de escala se representa por una versión de ella misma escalada más otra versión de ella misma escalada y trasladada.

Ahora bien, al aumentar j en la función de escala se mejorará la resolución de la representación de la función $f(t)$. Sin embargo, se puede obtener una mejor representación de la señal, no mediante el aumento de j , si no que al definir un nuevo espacio W_j o espacio wavelet. Este espacio es el complemento ortogonal de V_j en V_{j+1} , es decir que

$$V_j \oplus W_j = V_{j+1}$$

En la figura 72 se muestra la relación existente entre los espacios de escalamiento y los espacios wavelet. En este caso $V_3 = V_0 \oplus W_0 \oplus W_1 \oplus W_2$, lo que hace expandible esto a todos lo espacios siguientes como $L^2 = V_0 \oplus W_0 \oplus W_1 \oplus W_2 \oplus \dots$ donde V_0 es el espacio inicial, expandido para la función de escala $\phi(t - k)$, y los espacios W_j darán la información más detallada de la señal a medida que j crece.

Figura 72. Espacios escalados y wavelet



La escala que se use para expandir el espacio inicial será una cuestión a considerar y su elección depende en gran medida del análisis que se realice y de la señal en sí. Por lo que se podría pensar en representar un espacio inicial con una resolución tan alta o tan baja como sea necesaria.

La función que expande el espacio W_j es la wavelet madre $\psi_{j,k}(t)$. Como $W_0 \subset V_1$, la función wavelet $\psi(t)$ puede ser representada por una suma de funciones escala, escaladas y trasladadas, de la forma de (5-33). De la misma manera esto se cumple para $\phi(t)$, ya que $V_0 \subset V_1$, por lo que la función de escala puede representarse como (5-32).

Los coeficientes $h_1(n)$ se pueden encontrar por recursión de la siguiente manera

$$h_1(n) = (-1)^n h_0(1 - n) \quad (6-52a)$$

y si se tuviera una función finita de largo N , se tiene

$$h_1(n) = (-1)^n h_0(N - 1 - n) \quad (6-52b)$$

Con esto una función pertenecerá al espacio W_j si puede ser representada por la función prototipo de una wavelet madre de la forma de (6-49). Se puede ver que la variable j dirá en que espacio wavelet está trabajando la wavelet madre, siendo análogo para la función de escala.

6.5 Transformada wavelet discreta (DWT)

Como se mencionó una mejor representación de una señal se obtiene no mediante el aumento del espacio V_j , si no que el al combinar con la función de escala una función wavelet, con todas las características anteriormente enunciadas.

A partir de (5-30), que es una función de sub banda en el espacio V_{j_0} , y de (5-28), que se encuentra en el espacio wavelet W_j , se puede encontrar una expresión alternativa a (6-51). El espacio generado es de la forma $L^2 = V_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus \dots \oplus W_{j_n}$, donde j_0 es la escala del espacio inicial y j_n es la escala máxima del espacio wavelet usado.

Se pueden establecer expresiones para $f_0(t)$ y $g_j(t)$ para una señal de longitud 2^N si se usa (5-30) y (5-28) respectivamente, con lo cual la DWT es

$$f(t) = \sum_{k=0}^{2^{j_0}-1} c_{j_0,k} \phi_{j_0,k}(t) + \sum_{j=j_0}^{N-1} \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(t) \quad (6-53)$$

Donde $c_{j_0,k}$ son los coeficientes de escala, $d_{j,k}$ son los coeficientes wavelet y j_0 entrega el espacio inicial V_{j_0} , que será el espacio de menor resolución, y dependiendo de este j_0 es que el resto de los índices seguirán corriendo. A partir de la longitud de la señal, 2^N , se limitará el nivel de descomposición de la misma, ya que no tiene sentido representar una señal que se encuentra en un espacio V_{j_n} en el mismo espacio.

Ejemplo 6.5

Aplicar la DWT para el sistema Haar a la función $f(t) = \sin(t)$, con una frecuencia de muestreo de $\frac{2^3}{2\pi} Hz$.

Solución

Al muestrear la función $f(t) = \sin(t)$ con $t = nt_0 = n \left(\frac{2\pi}{2^3}\right)$ se obtiene el siguiente vector $f(t) = \left[0, \frac{1}{\sqrt{2}}, 1, \frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}, -1, -\frac{1}{\sqrt{2}}\right]$, para $0 \leq t \leq 2\pi$.

Este intervalo se trasladará a $[0,1]$ con el fin de trabajar en el mismo dominio de la función escalamiento y la wavelet Haar, siendo estas las funciones que utilizaremos para descomponer y luego reconstruir la señal³ $f(t)$.

La construcción de $f(t)$ a partir de (6-53) es

$$f(t) = \sum_{k=0}^{2^0-1} c_{0,k} \phi_{0,k}(t) + \sum_{j=0}^{3-1} \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(t) \quad (6-54)$$

$$= c_{0,0} \phi(t) + \sum_{j=0}^{2} \sum_{k=0}^{2^j-1} d_{j,k} 2^{\frac{j}{2}} \psi(2^j t - k)$$

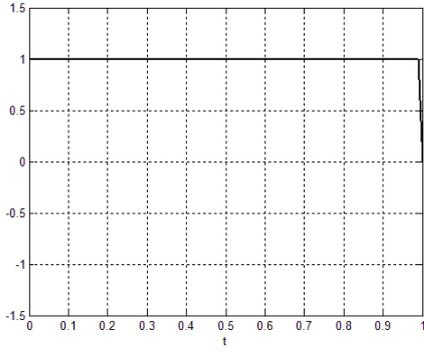
$$\sin(t) = c_{0,0} \phi(t) + d_{0,0} \psi(t) + d_{1,0} \sqrt{2} \psi(2t) + d_{1,1} \sqrt{2} \psi(2t - 1) + d_{2,0} 2 \psi(4t) + d_{2,1} 2 \psi(4t - 1) + d_{2,2} 2 \psi(4t - 2) + d_{2,3} 2 \psi(4t - 3) \quad (6-55)$$

Tabla X. Intervalos de tiempo para $f_{j,k}(t)$.

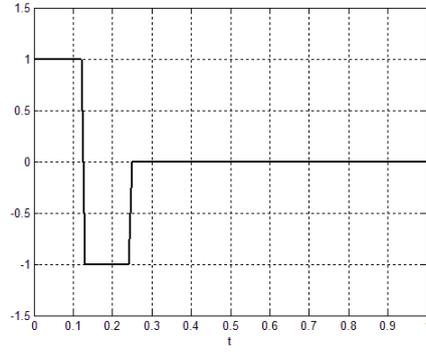
Función $f_{j,k}(t)$	Intervalo donde $f_{j,k}(t) = 1$	Intervalo donde $f_{j,k}(t) = -1$
$\phi(t)$	0.000 - 1.000	No existe
$\psi(t)$	0.000 - 0.500	0.500 - 1.000
$\sqrt{2} \psi(2t)$	0.000 - 0.250	0.250 - 0.500
$\sqrt{2} \psi(2t - 1)$	0.500 - 0.750	0.750 - 1.000
$2 \psi(4t)$	0.250 - 0.125	0.125 - 0.250
$2 \psi(4t - 1)$	0.250 - 0.375	0.375 - 0.500
$2 \psi(4t - 2)$	0.500 - 0.625	0.625 - 0.750
$2 \psi(4t - 3)$	0.750 - 0.875	0.875 - 1.000

³ Nótese que se utiliza el concepto de señal a la hora de realizar la transformación y el concepto función cuando se describe matemáticamente la señal.

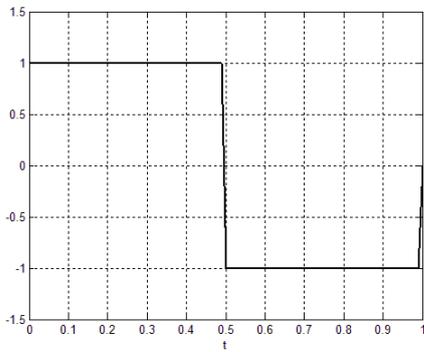
Figura 73. Función de escalado y wavelet para espacio W_2



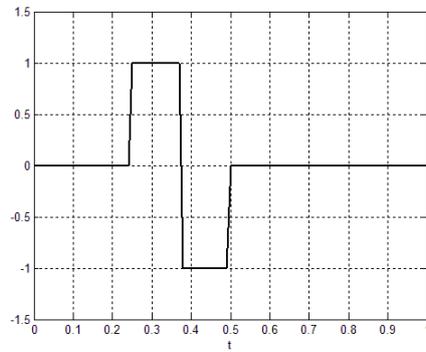
a.



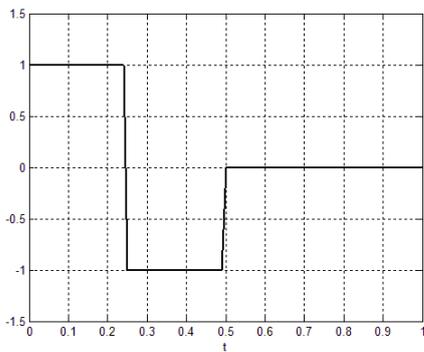
e.



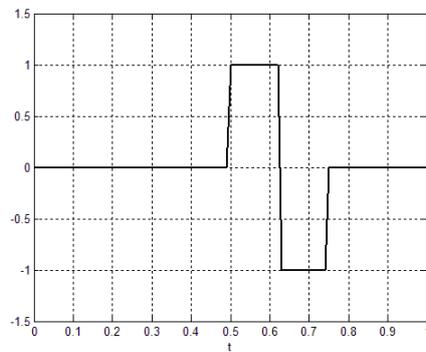
b.



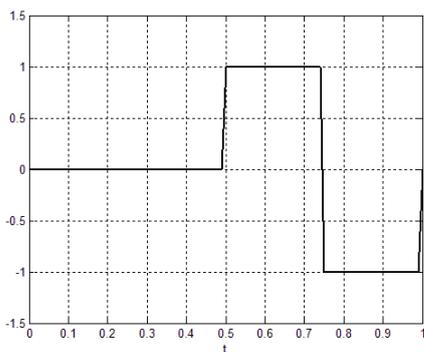
f.



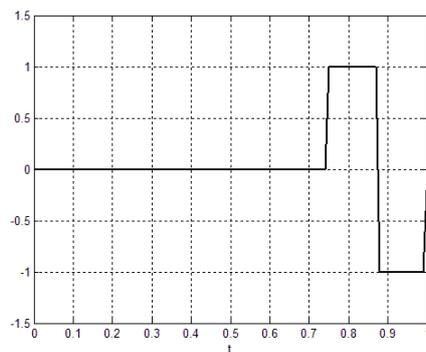
c.



g.



d.



h.

En la tabla X se muestran los intervalos de tiempo donde se encuentran definida la función de escalado y las funciones wavelet obtenidas para (6-55). En la figura 73 se muestra la grafica de cada una de estas funciones.

Existe una manera más conveniente de representar (6-55) a partir de los tiempos donde cada función de escala y wavelet se encuentran definidas. Esto se logra a través de una representación matricial de la siguiente manera

$$\begin{bmatrix} 0 \\ 1 \\ \frac{1}{\sqrt{2}} \\ 1 \\ 1 \\ \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \\ -1 \\ 1 \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = 2^{\frac{j_0}{2}} \begin{bmatrix} 1 & 1 & \sqrt{2} & 0 & 2 & 0 & 0 & 0 \\ 1 & 1 & \sqrt{2} & 0 & -2 & 0 & 0 & 0 \\ 1 & 1 & -\sqrt{2} & 0 & 0 & 2 & 0 & 0 \\ 1 & 1 & -\sqrt{2} & 0 & 0 & -2 & 0 & 0 \\ 1 & -1 & 0 & \sqrt{2} & 0 & 0 & 2 & 0 \\ 1 & -1 & 0 & \sqrt{2} & 0 & 0 & -2 & 0 \\ 1 & -1 & 0 & -\sqrt{2} & 0 & 0 & 0 & 2 \\ 1 & -1 & 0 & -\sqrt{2} & 0 & 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} c_{0,0} \\ d_{0,0} \\ d_{1,0} \\ d_{1,1} \\ d_{2,0} \\ d_{2,1} \\ d_{2,2} \\ d_{2,3} \end{bmatrix} \quad (6-56)$$

donde cada columna de la matriz de 8×8 representa los coeficientes de la función de escalado o la función wavelet. Por ejemplo, la primera columna, representa $\phi(t)$ ya que todos los valores son 1 en cualquier intervalo de tiempo entre 0 y 1. La segunda columna representa a $\psi(t)$, ya que la mitad del tiempo los coeficientes serán 1 y la otra mitad serán -1, etc.

El cálculo de los coeficientes es encontrado por

$$\begin{bmatrix} c_{0,0} \\ d_{0,0} \\ d_{1,0} \\ d_{1,1} \\ d_{2,0} \\ d_{2,1} \\ d_{2,2} \\ d_{2,3} \end{bmatrix} = 2^{-\frac{j_0}{2}} \mathbf{A}^{-1} \begin{bmatrix} 0 \\ 1 \\ \frac{1}{\sqrt{2}} \\ 1 \\ \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \\ -1 \\ \frac{1}{\sqrt{2}} \\ -1 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \quad (6-57)$$

donde A es la matriz de 8×8 , y en este caso $j_0 = 0$. El valor de los coeficientes es

$$\begin{bmatrix} c_{0,0} \\ d_{0,0} \\ d_{1,0} \\ d_{1,1} \\ d_{2,0} \\ d_{2,1} \\ d_{2,2} \\ d_{2,3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0.6036 \\ -0.1768 \\ 0.1768 \\ -0.1768 \\ 0.0732 \\ 0.1768 \\ -0.0732 \end{bmatrix} \quad (6-58)$$

En el capítulo siguiente se verá la importancia que tiene la descomposición de estos coeficientes, en particular, en la limpieza y compresión de señales.

Ahora que se conoce el procedimiento para determinar los coeficientes, se puede descomponer la señal para distintos espacios iniciales V_{j_0} .

Para $j_0 = 1$ se tiene

$$\begin{bmatrix} c_{1,0} \\ c_{1,1} \\ d_{1,0} \\ d_{1,1} \\ d_{2,0} \\ d_{2,1} \\ d_{2,2} \\ d_{2,3} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0.4268 \\ -0.4268 \\ -0.1768 \\ 0.1768 \\ -0.1768 \\ 0.0732 \\ 0.1768 \\ -0.0732 \end{bmatrix} \quad (6-59)$$

Para $j_0 = 2$ se tiene

$$\begin{bmatrix} c_{2,0} \\ c_{2,1} \\ c_{2,2} \\ c_{2,3} \\ d_{2,0} \\ d_{2,1} \\ d_{2,2} \\ d_{2,3} \end{bmatrix} = \frac{1}{\sqrt{2}\sqrt{2}} \begin{bmatrix} 0.1768 \\ 0.4268 \\ -0.1768 \\ -0.4268 \\ -0.1768 \\ 0.0732 \\ 0.1768 \\ -0.0732 \end{bmatrix} \quad (6-60)$$

Para $j_0 = 3$ se tiene

$$\begin{bmatrix} c_{3,0} \\ c_{3,1} \\ c_{3,2} \\ c_{3,3} \\ c_{3,4} \\ c_{3,5} \\ c_{3,6} \\ c_{3,7} \end{bmatrix} = \frac{1}{\sqrt{2}\sqrt{2}\sqrt{2}} \begin{bmatrix} 0 \\ 2 \\ \sqrt{8} \\ 2 \\ 0 \\ -2 \\ \sqrt{8} \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \quad (6-61)$$

Se puede observar que al acercarse el espacio inicial V_{j_0} , el espacio V_{j_m} , en el cual reside la señal original, se disminuirán los coeficientes wavelet, hasta llegar a representar la señal original solo con coeficientes de escala. Se nota que la última descomposición entrega una reconstrucción perfecta de la señal original, debido a que fue representada sólo con funciones de escala pertenecientes al espacio al cual pertenece la señal original.

El gran tamaño de los cálculos matemáticos hace necesaria la búsqueda de una manera más eficiente para el uso de la DWT. La DWT sería impráctica si su ejecución fuese tan complicada como para ser implementada en un microprocesador, de ahí que los bancos de filtros sean de gran ayuda para el desarrollo de la transformada rápida wavelet, FWT.

6.6 Transformada rápida wavelet (FWT) y banco de filtros

Una de las principales razones por las cuales la transformada wavelet es una potente herramienta matemática para el análisis de señales, es que permite el diseño y aplicación de rápidos algoritmos para el cálculo computacional.

6.6.1 Análisis o descomposición de señales

Se desea trabajar con un modelo de señal f_n , que parta de una señal análoga f , correspondiente a un set de valores discretos cuya longitud será de $N = 2^n$. Este modelo se pretende aproximar como una combinación lineal de funciones de escala de la forma:

$$f(t) = f_n(t) = \sum c_{n,k} 2^{\frac{n}{2}} \phi(2^n t - k) \quad (6-62)$$

El principal objetivo de la descomposición de una señal mediante la DWT se basa en que; a partir de $f_n \in V_n$ la señal original se puede expresar solamente con funciones de escala, muy apropiado para un análisis multi-resolución. Es posible expresar f_n como

$$f_n = f_{n-1} + g_{n-1} \quad (6-63)$$

dado que:

$$V_{n-1} \subset V_n \quad \wedge \quad W_{n-1} \subset V_n \quad (6-64)$$

$$V_n = V_{n-1} \oplus W_{n-1} \quad (6-65)$$

y

$$f_{n-1} \in V_{n-1} \quad \wedge \quad g_{n-1} \in W_{n-1} \quad (6-66)$$

El desarrollo de bancos de filtros y el diseño de algoritmos rápidos no se relaciona en forma directa con las funciones wavelet y escala, sino más bien con los coeficientes relacionados a estas funciones. Por lo que es importante encontrar los coeficientes $c_{n-1,k}$ y $d_{n-1,k}$ en términos de $c_{n,k}$.

Se sabe que tanto $\phi(t)$ y $\psi(t)$ generan bases ortogonales en L^2 de tal manera que el cálculo de $c_{n-1,k}$ y $d_{n-1,k}$ se realiza a través del producto interno de la señal con la función de escala y wavelet respectivamente

$$c_{n-1,k} = \langle f(t), \phi_{n-1,k}(t) \rangle = 2^{\frac{n-1}{2}} \int_{-\infty}^{\infty} f(t) \phi(2^{n-1}t - k) dt \quad (6-67)$$

$$d_{n-1,k} = \langle f(t), \psi_{n-1,k}(t) \rangle = 2^{\frac{n-1}{2}} \int_{-\infty}^{\infty} f(t) \psi(2^{n-1}t - k) dt \quad (6-68)$$

A partir de (5-31) y (5-32) se pueden tener expresiones alternativas para $\phi(2^{n-1}t - k)$ y $\psi(2^{n-1}t - k)$ la cuales son

$$\phi(2^{n-1}t - k) = \sqrt{2} \sum_p h_0(p) \phi(2^n t - 2k - p) \quad (6-69)$$

$$\psi(2^{n-1}t - k) = \sqrt{2} \sum_p h_1(p) \psi(2^n t - 2k - p) \quad (6-70)$$

Reemplazando en las integrales (6-67) y (6-68) y haciendo cambio de variable $m = 2k + p$ resulta

$$c_{n-1,k} = \int_{-\infty}^{\infty} f(t) \sum_p h_0(m - 2k) 2^{\frac{n}{2}} \phi(2^n t - m) dt \quad (6-71)$$

$$d_{n-1,k} = \int_{-\infty}^{\infty} f(t) \sum_p h_1(m - 2k) 2^{\frac{n}{2}} \psi(2^n t - m) dt \quad (6-72)$$

e intercambiando el orden de la sumatoria y la integral se tiene

$$c_{n-1,k} = \sum_p h_0(m - 2k) \int_{-\infty}^{\infty} f(t) 2^{\frac{n}{2}} \phi(2^n t - m) dt \quad (6-73)$$

$$d_{n-1,k} = \sum_p h_1(m - 2k) \int_{-\infty}^{\infty} f(t) 2^{\frac{n}{2}} \psi(2^n t - m) dt \quad (6-74)$$

En las ecuaciones (6-73) y (6-74) se observa que el término dentro de la integral es el mismo para ambas ecuaciones, aún más, resulta que es $c_{n,m}$. Con esto se ha podido establecer una representación de los coeficientes de escala y wavelets a un nivel de resolución más bajo en términos de los coeficientes escala en un nivel de resolución más alto

$$c_{n-1,k} = \sum_p h_0(m - 2k) c_{n,m} \quad (6-75)$$

$$d_{n-1,k} = \sum_p h_1(m - 2k) c_{n,m} \quad (6-76)$$

Las ecuaciones (6-75) y (6-76) corresponden a una convolución discreta, donde la secuencia de entrada dada por $c_{n,k}$ es convolucionada con h_0 para obtener una representación “suave”, dada por los coeficientes de escala $c_{n-1,k}$, y con h_1 para obtener los detalles de la señal, dados por los coeficientes wavelet $d_{n-1,k}$. Es claro ahora de que h_0 funciona como un filtro paso bajo y h_1 como uno paso alto.

Ahora (6-62) se puede escribir como

$$f(t) = \sum_k c_{n-1,k} 2^{\frac{n-1}{2}} \phi(2^{n-1}t - k) + \sum_k d_{n-1,k} 2^{\frac{n-1}{2}} \psi(2^{n-1}t - k) \quad (6-77)$$

Al aplicar esta operación sobre una señal digital real, los datos de salida obtenidos serán del doble a los de entrada, por lo que se aplica un subsampling que realiza un diezmado de la señal original.

6.6.2 Síntesis o reconstrucción de señales

Al proceso de reconstrucción a partir de los datos obtenidos en el análisis se le denomina síntesis, y consiste en la aplicación de la inversa de la transformada wavelet discreta, IDWT. Lo que se desea hacer es poder presentar los coeficientes de escala en un nivel de resolución más alto mediante una combinación de los coeficientes de escala y wavelets de un nivel de resolución más bajo. Esto se puede obtener si se substituye $\phi(2^{n-1}t - m)$ y $\psi(2^{n-1}t - m)$ con expresiones equivalentes derivadas de (5-31) y (5-32).

La nueva expresión para (6-77) es

$$f(t) = \sum_m c_{n-1,m} \sum_p h_0(p) 2^{\frac{n}{2}} \phi(2^n t - 2m - p) + \sum_m d_{n-1,m} \sum_p h_1(p) 2^{\frac{n}{2}} \phi(2^n t - 2m - p) \quad (6-78)$$

al multiplicar ambos lados de (6-78) por $\phi(2^n t - k)$ e integrar respecto al tiempo

$$\begin{aligned} \int_{-\infty}^{\infty} f(t) \phi(2^n t - k) dt \\ = \sum_m c_{n-1,m} \sum_p h_0(p) 2^{\frac{n}{2}} \int_{-\infty}^{\infty} \phi(2^n t - 2m - p) \phi(2^n t - k) dt \\ + \sum_m d_{n-1,m} \sum_p h_1(p) 2^{\frac{n}{2}} \int_{-\infty}^{\infty} \psi(2^n t - 2m - p) \phi(2^n t - k) dt \end{aligned} \quad (6-78)$$

El lado izquierdo de (6-78) se observa que es $c_{n,k}$, además los productos de la derecha, al ser $\phi_{n,k}$ ortonormal, son cero excepto en

$$\begin{aligned} c_{n,k} = \sum_m c_{n-1,m} \sum_p h_0(p) \delta(k - (2m + p)) \\ + \sum_m d_{n-1,m} \sum_p h_1(p) \delta(k - (2m + p)) \end{aligned} \quad (6-79)$$

Cambiando de variable por $q = 2m + p$

$$c_{n,k} = \sum_m c_{n-1,m} \sum_p h_0(p) \delta(k - q) + \sum_m d_{n-1,m} \sum_p h_1(p) \delta(k - q)$$

(6-80)

y como $\delta(k - q)$ está definida por

$$\delta(k - q) = \begin{cases} 1 & \text{si } k = q \\ 0 & \text{si } k \neq q \end{cases} \quad (6-81)$$

se tiene finalmente

$$c_{n,k} = \sum_m c_{n-1,m} h_0(2m - k) + \sum_m d_{n-1,m} h_1(2m - k) \quad (6-82)$$

Así pues, como en el análisis se hace un filtrado y un subsampleo, en la síntesis se realiza un subsampleo y luego un filtrado. En el subsampleo se insertan ceros entre muestras con el fin de aumentar al doble la longitud de las componentes de entrada de tal manera que la señal de salida luego del filtrado tenga la misma longitud que la señal original.

Ejemplo 6.6

Aplicar una descomposición y reconstrucción a la señal discreta $X = \{4, 1, 2, 3, 1, 6, 2, 5\}$ utilizando la wavelet Haar.

Solución

En el desarrollo de este capítulo se han encontrado que los coeficientes h_0 y h_1 son:

$$h_0(0) = h_0(1) = \frac{1}{\sqrt{2}}$$

$$h_1(0) = \frac{1}{\sqrt{2}} \quad h_1(1) = -\frac{1}{\sqrt{2}}$$

Como la aproximación está relacionada con un promedio y el detalle con diferencias, a partir de (6-75) y (6-76) se puede tener la primera descomposición

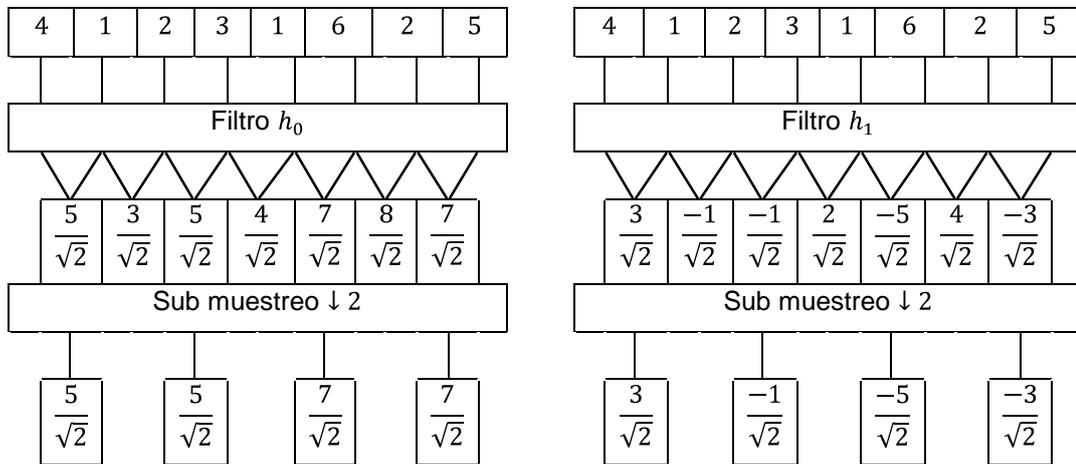
$$c_{J-1,k} = \frac{1}{\sqrt{2}}(c_{J,2k} + c_{J,2k+1}) \quad (6-83)$$

$$d_{J-1,k} = \frac{1}{\sqrt{2}}(c_{J,2k} - c_{J,2k+1}) \quad (6-84)$$

donde los c_j son los coeficientes correspondientes a la señal original. J es el nivel más alto de resolución con el que se puede trabajar y $N = 2^J$ es la longitud de la señal original. En este caso se tiene que $N = 8 = 2^3$.

En la figura 74 se muestra el algoritmo de descomposición, ahí se puede ver como se realiza la convolución discreta entre los filtros h_0 y h_1 con la señal original y posteriormente el subsampleo. Con esto se obtienen dos conjuntos de coeficientes, uno que posee la aproximación de la señal original y el otro que posee los detalles, ambos son de longitud 4. Ahora se puede mantener el detalle y se procede a aplicar el algoritmo a los coeficientes de aproximación dando como resultado una nueva aproximación más general y un nuevo detalle, de longitud 2. Se puede llevar a cabo una vez más este proceso para obtener un detalle y una aproximación de longitud 1.

Figura 74. Algoritmo de análisis, DWT



La reconstrucción de la señal a partir de los coeficientes de escala y wavelet que pertenecen al primer nivel de descomposición se pueden obtener si se suman y restan (6-83) y (6-84), obteniendo las siguientes expresiones

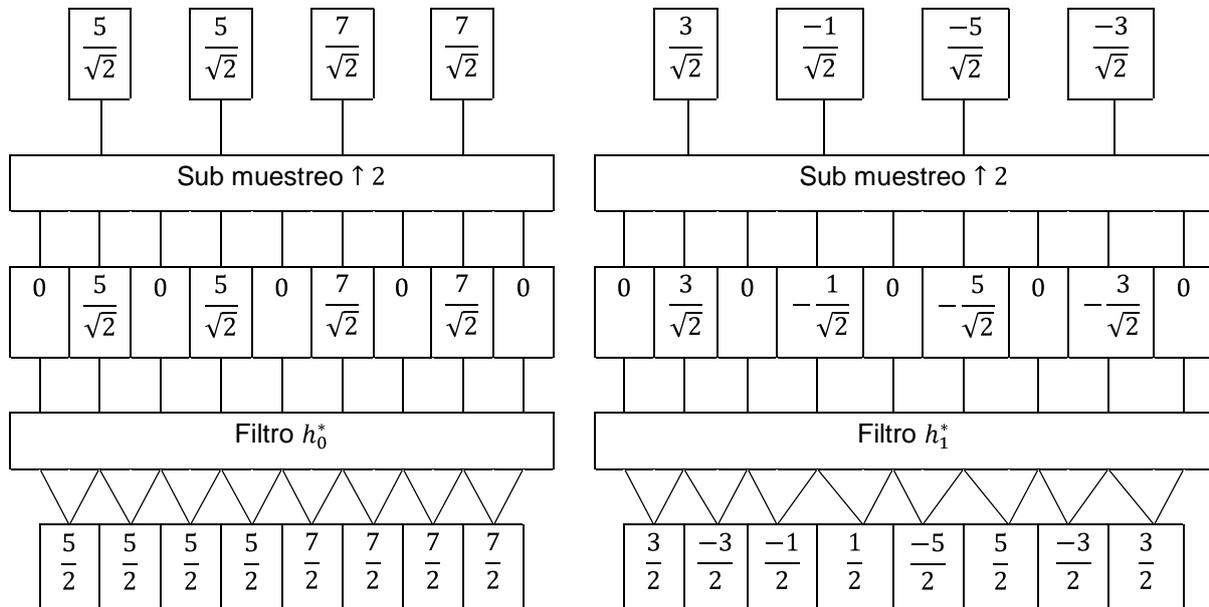
$$c_{J,2k} = \frac{1}{\sqrt{2}}(c_{J-1,k} + d_{J-1,k}) \quad (6-85)$$

$$c_{J,2k-1} = \frac{1}{\sqrt{2}}(c_{J-1,k} - d_{J-1,k}) \quad (6-86)$$

Estas dos expresiones permiten la reconstrucción perfecta de la señal.

En la figura 75 se muestra el algoritmo de síntesis, se puede ver el proceso de sub muestreo y posteriormente la convolución para reconstruir la señal. La suma de los coeficientes de reconstrucción y escala wavelet se muestra en la figura 76.

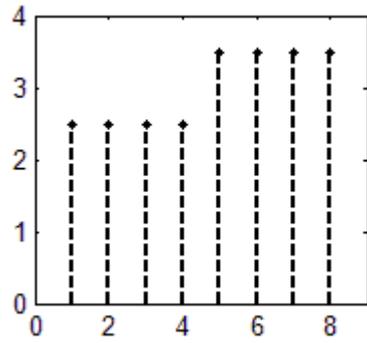
Figura 75. Algoritmo de síntesis, IDWT



En la figura 74 y 75 se observa que la descomposición y reconstrucción se ha realizado sólo una vez. Sin embargo, el objetivo principal de un análisis wavelet es el obtener el vector de la DWT, ya que contiene información que útil a la hora de realizar un procesamiento de señales; como la compresión y reducción de ruido. En las figuras 77 y 78 se muestra el proceso de descomposición completo al que se hace referencia.

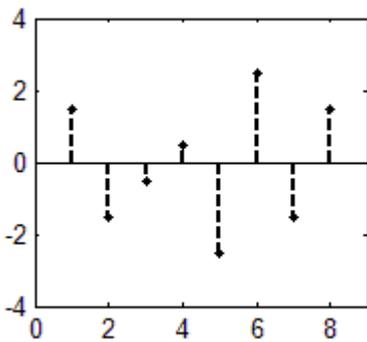
Figura 76. Reconstrucción a partir de aproximación y detalles

Reconstrucción de la aproximación

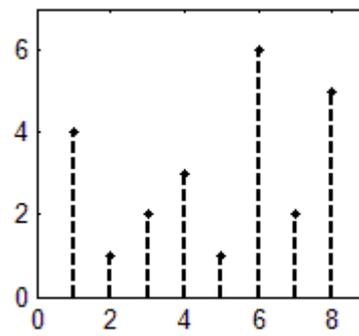


+

=



Reconstrucción del detalle



Señal reconstruida

Figura 77. Análisis multi-resolución para determinar vector DWT

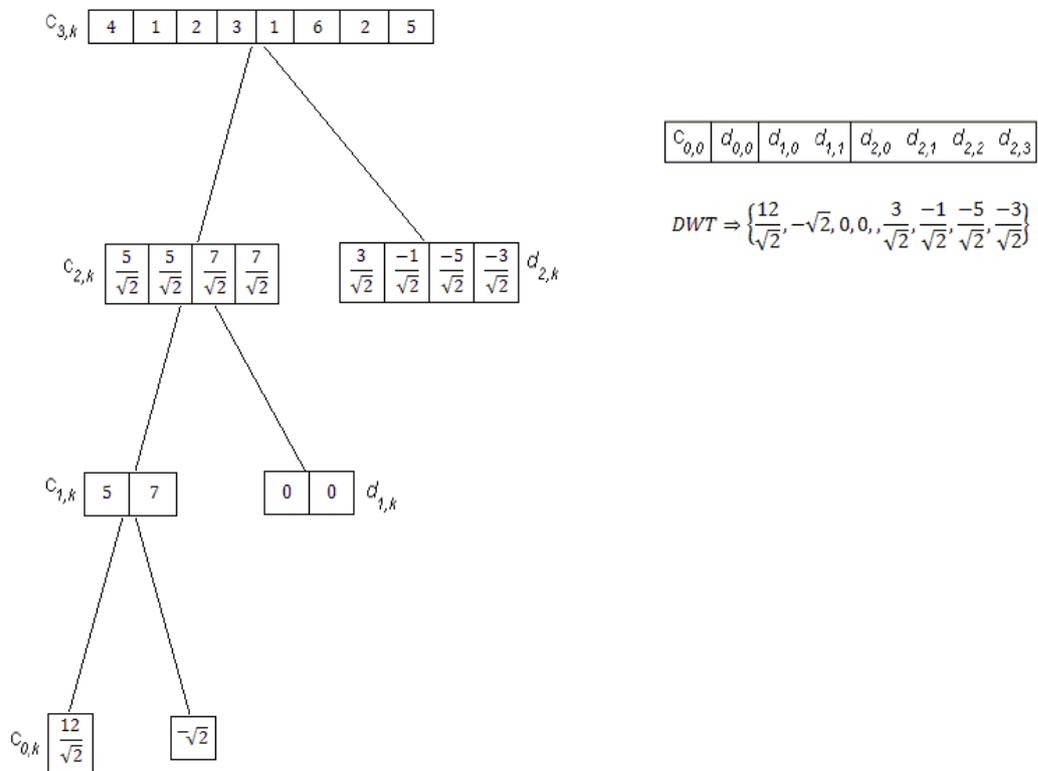
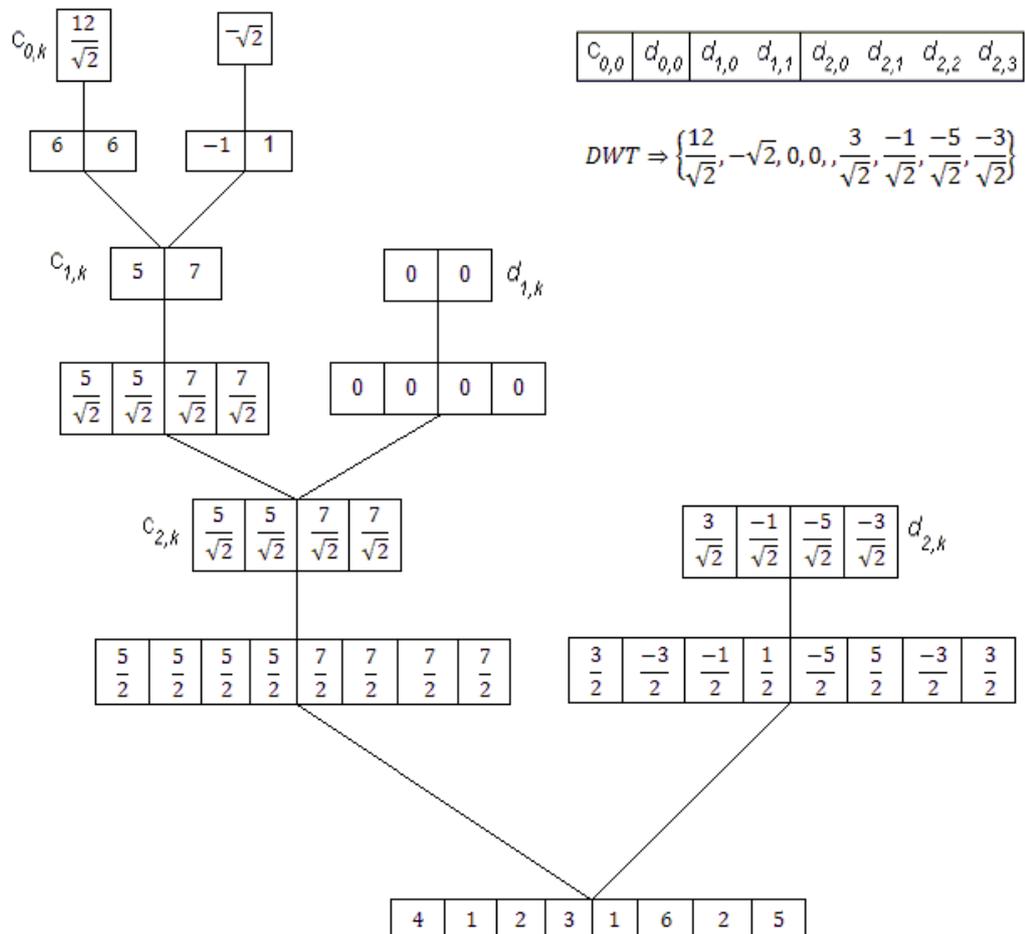


Figura 78. Reconstrucción de señal a partir del vector DWT



7. WAVELETS EN EL PROCESADO DE SEÑALES DE AUDIO

7.1 Introducción

Dos de los grandes éxitos de las wavelets son la eliminación de ruido y la compresión de señales, comúnmente considerados problemas de alta complejidad. La corrección de ruido basado en el análisis wavelet conlleva a algoritmos simples y potentes que son más fáciles de ajustar que los métodos tradicionales basados en el análisis funcional.

La compresión de señales por medio de wavelets constituye un campo que parece ser muy competitivo por razones muy similares a las que hacen que el análisis de reducción de ruido por medio de bandas wavelet funcione. De hecho, con las señales que estamos interesados, en muchos casos la representación wavelet se puede realizar usando pocos coeficientes.

La eficiencia del proceso de eliminación de ruido y la compresión de señales depende en gran medida de la wavelet utilizada, por lo que antes de describir estos procesos se deben considerar los requerimientos que una señal de audio implica.

7.2 Requerimientos y consideraciones

La calidad de descomposición *wavelet* depende especialmente de la habilidad de aproximar la señal con las *wavelets*. Cuando la *wavelet* aplicada no se asemeja a la forma de la señal analizada, los coeficientes *wavelet* no extraerán las características importantes de la señal, resultando en muchos coeficientes distintos de cero que aproximen la señal. De ahí que, a un mejor análisis, una menor cantidad de coeficientes *wavelet*.

Las señales musicales son casi siempre un tipo de onda suave, significativamente más suave que las imágenes. Las imágenes pueden contener bordes afilados, líneas finas y un alto contraste. Los filtros cortos que corresponden a *wavelets* rugosas, como *Daubechies 2*, han probado aproximar bastante bien imágenes. Sin embargo, las señales de audio, llevan al requerimiento de una *wavelet* lo suficientemente suave, es decir, con una alta regularidad.

Otro factor importante es el tamaño de la banda de transición del filtro paso bajo y el filtro paso alto. Bandas de transición muy grandes causan un traslape de las bandas de paso bajo y las bandas de paso alto. Lo que produce que las bandas de salida del banco de filtros no se encuentren bien separadas, y los efectos de *aliasing* se refuercen cuando los coeficientes son cambiados. Especialmente en aplicaciones donde los coeficientes *wavelet* se encuentran relacionados directamente con la frecuencia, como el cambio de tono, una amplia separación de respuesta en frecuencia de paso alto y paso bajo también es importante.

Una respuesta de fase lineal es crucial para filtros de audio de alta calidad. Cuando los filtros no poseen por lo menos una fase lineal aproximada, ciertas frecuencias son retrasadas en el dominio *wavelet*. La transformada inversa deshace esta distorsión de fase.

Sin embargo, cuando los coeficientes *wavelet* son cambiados, pueden ocurrir modificaciones indeseables a las frecuencias que se encuentran “fuera de fase”. Una respuesta de fase lineal se puede alcanzar si se usan filtros simétricos.

Por último, distintas *wavelets* poseen distintas localizaciones temporales. *Wavelets* con soporte compacto pueden localizar el tiempo de un evento mejor que otras. Así que, para un análisis temporal, una *wavelet* corta es necesaria, entre más rápido decaiga mucho mejor. Esto entra en conflicto con la habilidad de separar en bandas de frecuencia y la suavidad requerida, para este caso filtros grandes proveen mejores resultados.

Para el caso del procesado en tiempo real, las transformaciones *wavelet* llevan al requerimiento de un algoritmo lo suficientemente rápido como para que el procesador sea capaz de calcular la transformada *wavelet*, así como su inversa, mucho más rápido que la porción de audio que es reproducida. Por ejemplo, si se trabajan con porciones de audio de 20ms de duración, cualquier procesado de esta porción no debe de tardar más de 20ms –de otra manera el flujo de estas porciones estaría entrecortada. Entre más rápido haya sido completado el procesamiento, mejor será, y en el tiempo restante el procesador puede ser usado para un procesado adicional de la señal de audio, tareas del sistema operativo, etc. Adicionalmente, se necesita un espacio libre, para que el sistema en tiempo real opere estable en cualquier momento, como cuando ocurren altos picos de uso del procesador. Este espacio libre necesita ser especialmente grande para sistemas operativos con preferencias a la multitarea, ya que el sistema puede interrumpir el procesado de la porción de audio para realizar otra tarea.

Como la longitud de los filtros afecta de forma directa el tiempo de cómputo para el análisis y resíntesis, de ahí que se prefieran filtros cortos. Sin embargo, y en general; más momentos de desvanecimiento y bandas de transición conllevan a filtros más grandes. Siendo esto preferido para filtros de audio, se puede realizar un compromiso razonable entre la longitud del filtro.

Las computadoras normalmente pueden procesar números enteros más rápido que números de punto flotante. Por lo que se podría pensar en usar transformadas *wavelet* basadas en números enteros. Sin embargo, valores enteros de muestras no son muy adecuados para un procesado de audio de alta calidad, ya que pueden resultar en errores de redondeo o señales no tan suaves como las esperadas.

Los sistemas de procesado de audio de alta calidad se pueden asumir que trabajan con señales en un formato de punto flotante, así que el uso de una transformada de valores enteros sería de poco beneficio, mientras que reduciría la calidad total. Todas las PCs de hoy en día se encuentran equipadas con un procesador de punto flotante bastante rápido, así que el impacto en el desempeño no es muy importante.

En conclusión, la elección de la *wavelet* “*adecuada*” depende en gran medida del análisis que se pretenda hacer. Sin embargo, la necesidad de una respuesta de fase lineal limita el uso de cierto tipo de *wavelets*. Por otra parte, la resolución temporal no es una preocupación primordial, por el contrario es de mayor interés una pendiente abrupta en los filtros. De ahí que las *wavelets* biortogonales⁴ y las *Symlets* serán las elegidas por cumplir con estas características, destacando las biortogonales por su simetría.

⁴ El uso de *wavelets* biortogonales implica el doble de cálculo para los filtros de descomposición y reconstrucción por lo que se sacrifica el tiempo de ejecución por la eficiencia.

7.3 Reducción de ruido de una señal de audio

El objetivo de la reducción de ruido es el de reconstruir una señal útil a partir de observaciones dañadas por ruido aditivo. La reconstrucción de la señal se puede obtener al excluir el ruido $w(n)$ de una señal que posee la forma

$$y(n) = x(n) + w(n) \quad (7-1)$$

En este caso $w(n)$ podría ser un proceso Gaussiano de ruido blanco, AWGN, el cual es estadísticamente independiente de $x(n)$. La reducción de ruido se alcanzaría al aplicar una operación no lineal a la representación wavelet de $y(n)$.

En el contexto de la STFT este proceso se da por medio de la sustracción de espectros. De hecho, la reducción de ruido por medio de wavelets se encuentra relacionada de forma estrecha con la sustracción de espectros.

La principal diferencia entre estos dos métodos está en el hecho de que las wavelets usadas para la reducción de ruido son de valor real mientras que para la STFT son complejas.

El proceso de eliminación de ruido es el siguiente. Primero, la señal $y(n)$ se descompone usando un banco de filtros de bandas por octavas, con lo que se realiza la DWT. Luego, los coeficientes son manipulados de tal manera que se remueven las componentes de ruido. Posteriormente se reconstruye la señal.

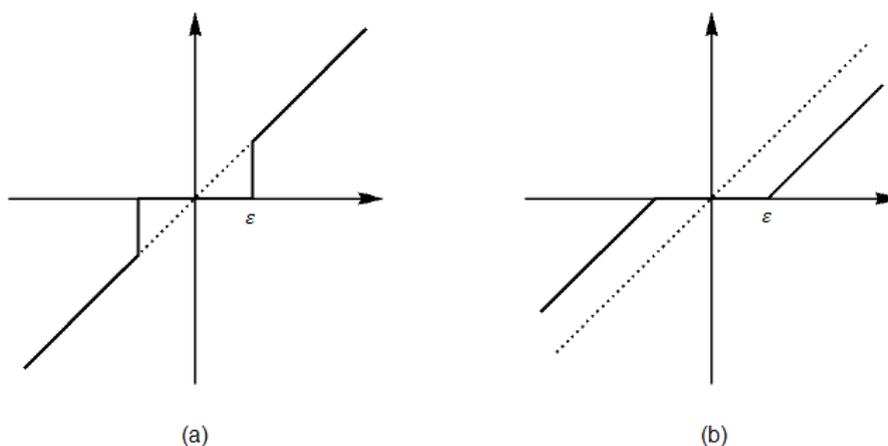
Existen dos enfoques propuestos: el umbral duro y el umbral suave. Las siguientes no linealidades describen los dos tipos de umbrales:

$$\hat{y}(n) = \begin{cases} y(n), & y(n) > \varepsilon \\ y(n), & y(n) < -\varepsilon \\ 0, & |y(n)| \leq \varepsilon \end{cases} \quad (\text{Duro}) \quad (7-2)$$

$$\hat{y}(n) = \begin{cases} y(n) - \varepsilon, & y(n) > \varepsilon \\ y(n) + \varepsilon, & y(n) < -\varepsilon \\ 0, & |y(n)| \leq \varepsilon \end{cases} \quad (\text{Suave}) \quad (7-3)$$

La figura 79 muestra el concepto del umbral duro en (a), y el umbral suave en (b), claramente se observa que el umbral duro crea discontinuidades en $\hat{y}(n)$

Figura 79. Umbral duro y suave.



Básicamente, la idea del umbral es que $x(n)$ puede ser representado con menos coeficientes wavelet, mientras que el ruido posee características de un amplio ancho de banda y se distribuye en todos los coeficientes. Por ejemplo, esto aun sigue siendo verdadero si $x(n)$ es una señal paso bajo, mientras que $w(n)$ es ruido blanco.

Con el procedimiento del umbral se establecen los coeficientes pequeños que representan a $w(n)$ a cero, mientras que los coeficientes grandes debidos a $x(n)$ son ligeramente modificados. Por lo que al escoger un valor umbral ε adecuado, la señal $\hat{y}(n)$ reconstruida por los coeficientes wavelet manipulados contendrá mucho menos ruido que el que posee $y(n)$.

En la práctica, el problema es el de escoger ε , porque usualmente no se conoce *a priori* la cantidad de ruido que posee la señal. Si ε es muy pequeño, el ruido no será removido efectivamente. Si es muy grande, la señal será distorsionada.

7.3.1 Umbral universal

Varios estudios han determinado que la elección del umbral depende del nivel de ruido σ de los datos de entrada. El nivel de ruido se propone que sea la desviación media absoluta de los coeficientes wavelet obtenidos en el primer nivel de descomposición dividida por 0.6745.

$$\sigma = \frac{\sum_k \frac{|d_{J-1,k} - \mu_{J-1,k}|}{n}}{0.6745} \quad (7-4)$$

Donde $\mu_{J-1,k}$ representa el valor medio de los coeficientes wavelet en el primer nivel de descomposición.

A partir de este valor del nivel de ruido *Donoho* y *Johnstone* propusieron el umbral universal⁵, T , definido como

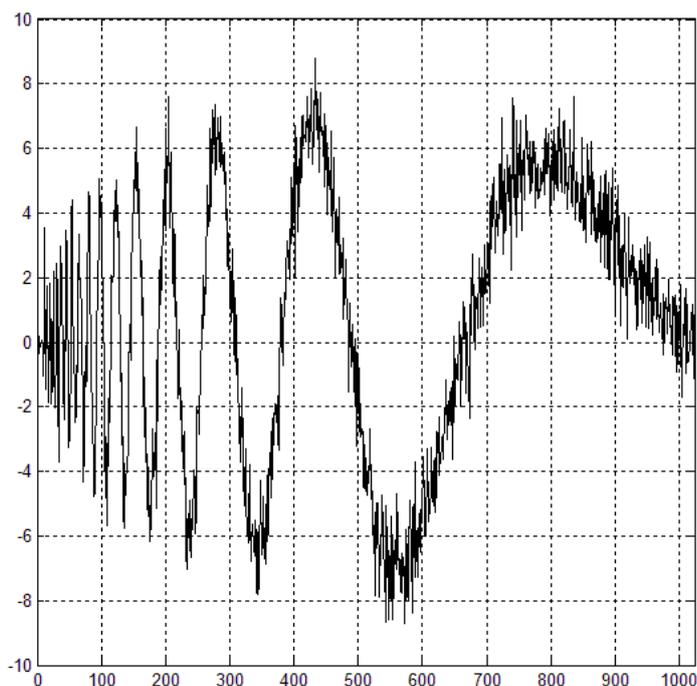
$$T = \sigma\sqrt{2\log(N)} \quad (7-5)$$

⁵ Suponiendo ruido blanco ε con distribución de probabilidad Gaussiana de varianza σ^2 , esta no se encuentra limitada, pero se sabe que $\lim_{n \rightarrow \infty} P\{\max_{1 \leq i \leq n} |\varepsilon_i| > \sigma\sqrt{2\log(N)}\} = 0$, con lo que se puede considerar que $|\varepsilon_i| < \sigma\sqrt{2\log(N)}$.

7.3.2 Ejemplo de reducción de ruido estacionario

En este ejemplo se realizará un análisis wavelet a una señal de efecto Doppler afectada por ruido como la mostrada en la figura 81. Se utilizará la wavelet biortogonal de cuarto orden tanto en la descomposición como en la reconstrucción, *bior4.4*. Se escoge esta wavelet por su soporte compacto, simetría y la posibilidad de reconstrucción perfecta.

Figura 80. Señal producida por efecto Doppler con ruido aditivo.



La estrategia con la que se abordará el problema del ruido consiste en la descomposición de la señal hasta que se alcance una aproximación donde el ruido esté ausente o se atenúe lo suficiente (a_4 o a_5). Posteriormente se complementará esta aproximación con los detalles atribuibles a la señal deseada y se descartarán los que se consideren ruido.

El proceso de reducción de ruido se realizó con el *wavelet* toolbox de MATLAB, a continuación se muestra el procedimiento realizado

1. Cálculo del nivel de ruido

$$\sigma = 1.108$$

2. Cálculo del valor umbral

$$T = 4.124$$

3. Descomposición de la señal hasta el quinto nivel, figura 81.
4. Aplicación del umbral suave (para evitar discontinuidades) sobre los coeficientes de detalle, figura 82.
5. Reconstrucción de la señal a partir de los nuevos coeficientes wavelet, figura 83.
6. El error cuadrático medio

$$E = 4.14\%$$

7. Porcentaje de energía retenida (potencia señal original/potencia señal reconstruida)

$$PER = 94.15\%$$

Figura 81. Análisis y detalle de señal Doppler.

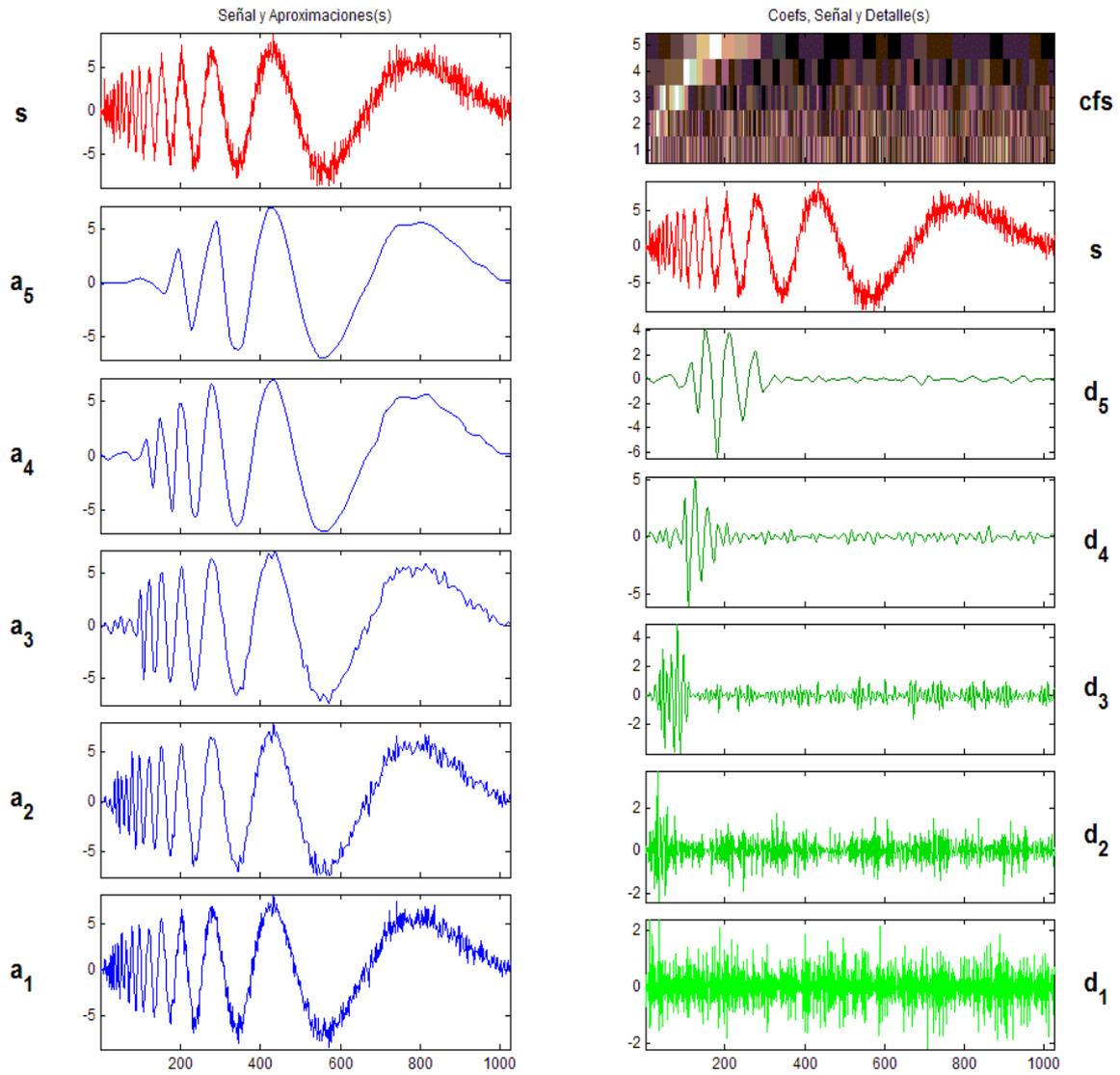


Figura 82. Aplicación del umbral suave sobre los coeficientes de detalle.

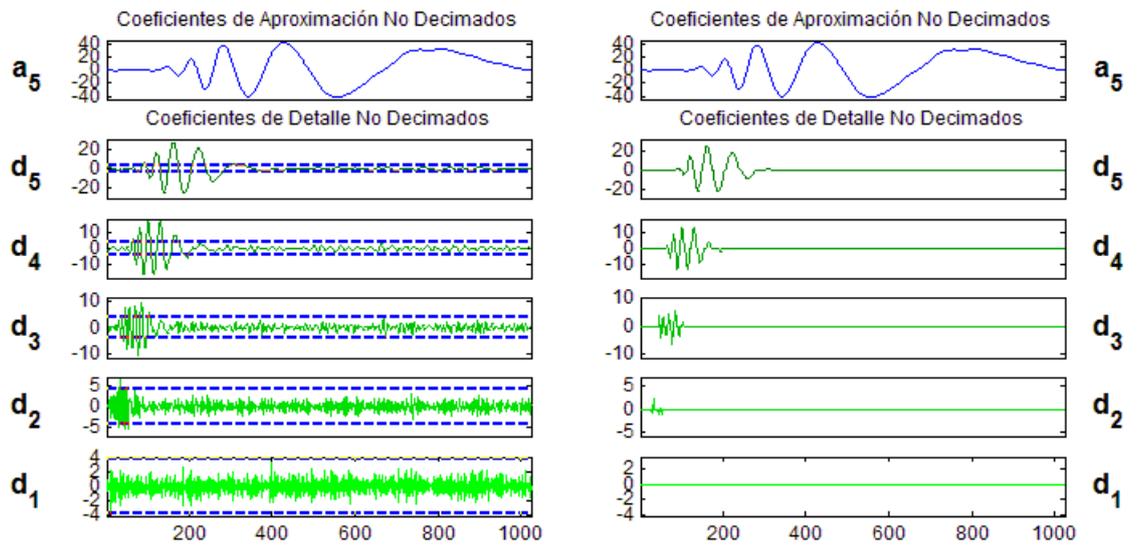
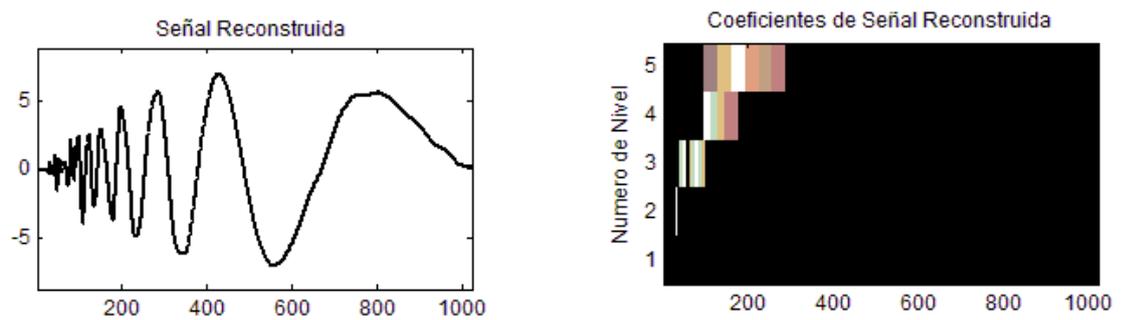


Figura 83. Reconstrucción de la señal a Doppler.



Se puede observar que la reducción de ruido se realizó únicamente al limitar los coeficientes de detalle mientras que los coeficientes de análisis se dejaron intactos. Esto conlleva a dividir la señal con ruido en dos partes, la primera contiene los coeficientes de análisis de la forma

$$\hat{A}_d = \sum_k \hat{\beta}_{5,k} \phi_{5,k} \quad (7-6)$$

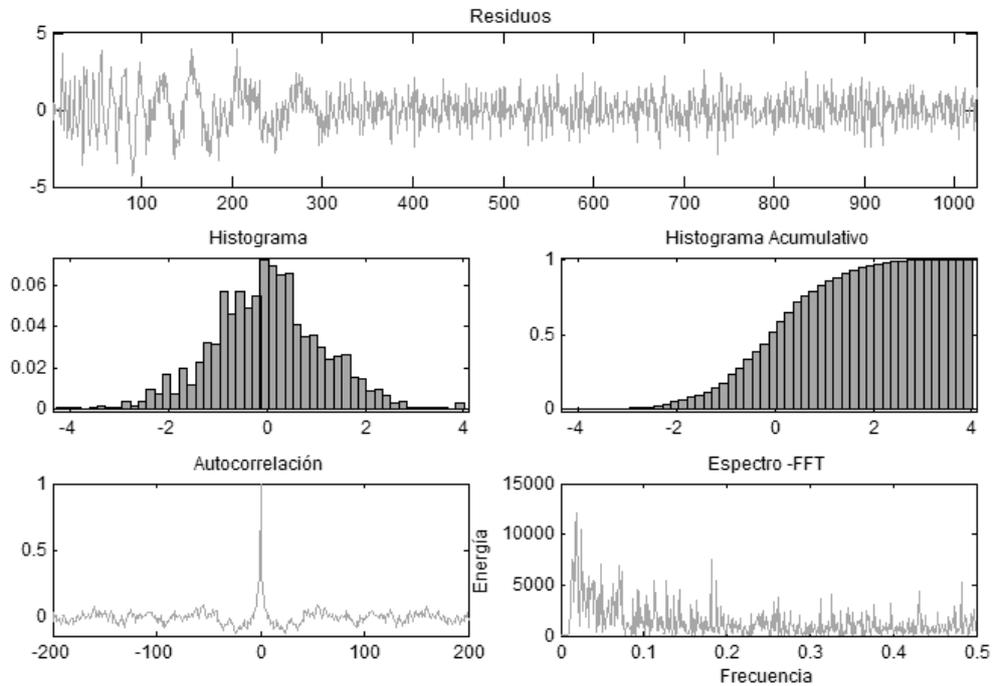
Los estimadores $\hat{\beta}_{5,k}$ son esencialmente los coeficientes de análisis de la señal original y \hat{A}_d es la aproximación a_5 . La segunda parte de la señal se compone de los coeficientes de análisis y tiene la forma

$$\hat{D}_d = \sum_{1 \leq j \leq 5} \sum_k \hat{\alpha}_{j,k} \psi_{j,k} \quad (7-7)$$

Donde los estimadores $\hat{\alpha}_{j,k}$ son los coeficientes de detalle que se les aplicó el umbral suave.

Si se observa la figura 84 se pueden hacer algunas conclusiones acerca de los residuos que en el análisis anterior se consideraron como ruido. En el histograma se observa como la distribución de los residuos se aproxima muy bien a una distribución de probabilidad Gaussiana; y al no existir una autocorrelación, se comprueba que los datos eliminados pueden ser considerados únicamente como ruido. Además si se observa que el contenido espectral de los residuos está constituido por componentes de distintas frecuencias se confirma nuestra hipótesis de que lo eliminado es solamente ruido.

Figura 84. Residuos de la señal a Doppler.



7.3.3 Ejemplo de reducción de ruido no estacionario

El modelo empleado en 7.3.2 es muy sencillo y simplista ya que se han hecho algunas suposiciones acerca del ruido que no se pueden pasar por alto. Se pueden tomar dos casos particularmente:

- El ruido presenta múltiples cambios en la varianza;
- La estructura del ruido se desconoce

7.3.3.1 Ruido con cambios en la varianza

La idea detrás de este proceso consiste en definir umbrales dependientes del tiempo para poder manejar el proceso no estacionario del ruido. De manera más precisa, el modelo asume (igual que antes) que las observaciones son iguales a la señal de interés superpuesta con ruido, como se observa en (7-1), pero ahora, la varianza del ruido puede cambiar con el tiempo.

Existen distintos valores de varianza en diferentes intervalos de tiempo, siendo estos valores desconocidos así como también los intervalos de tiempo donde ocurren.

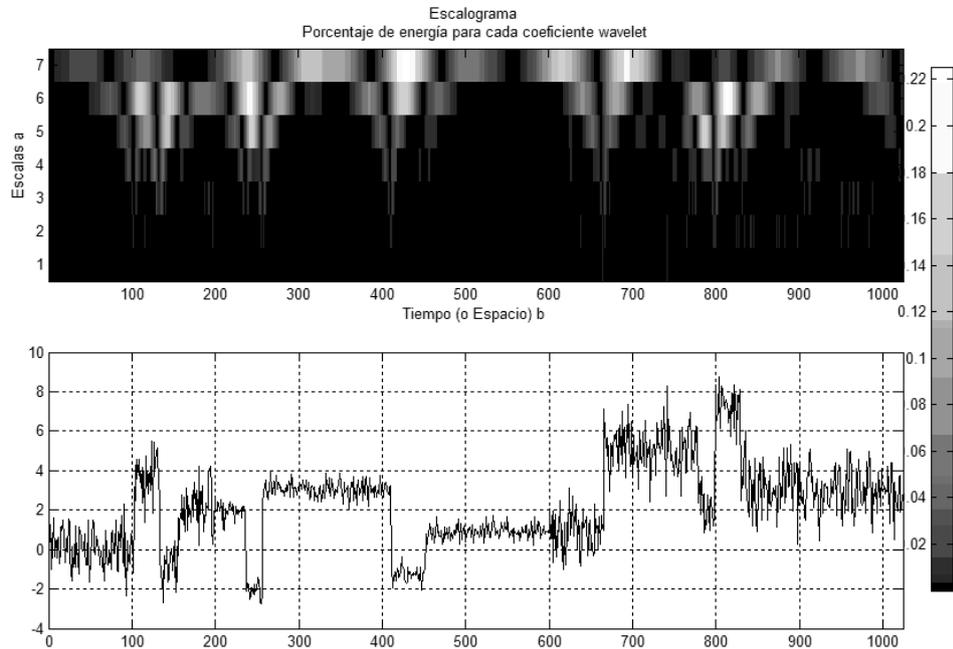
Surge entonces el primer problema, y es el de estimar cuáles son los puntos donde se da un cambio en la varianza, o es lo mismo, los intervalos de tiempo donde se dan estos cambios. El algoritmo aquí descrito se basa en el propuesto por Marc Lavielle acerca de la detección puntos de cambio utilizando programación dinámica.

El algoritmo consiste en descomponer la señal de interés por medio de la transformación wavelet en un primer nivel. Con esto se obtiene un detalle de primer nivel, que carece de la señal de interés, sin embargo captura las características principales desde un punto de vista de detección de variaciones. Siempre y cuando la señal de interés posea una representación wavelet dispersa. Usando este detalle se substituye un porcentaje de los valores mayores por la media, por ejemplo un 4%, con lo cual se pretende eliminar los valores grandes. La finalidad de este proceso es el de obtener una distribución más uniforme en los distintos intervalos de tiempo y a partir de esta nueva representación se pueden estimar los puntos donde suceden cambios en la varianza.

Al igual que en el ejemplo anterior la complejidad consiste en establecer los valores umbrales y a esto se suma el de establecer los intervalos de tiempo donde existen cambios en la varianza⁶. A continuación se describe el proceso de reducción de ruido realizado con el *wavelet* toolbox de MATLAB utilizando la wavelet *symlet* de nivel 4, *sym4*, a una señal como la que se muestra en la figura 85.

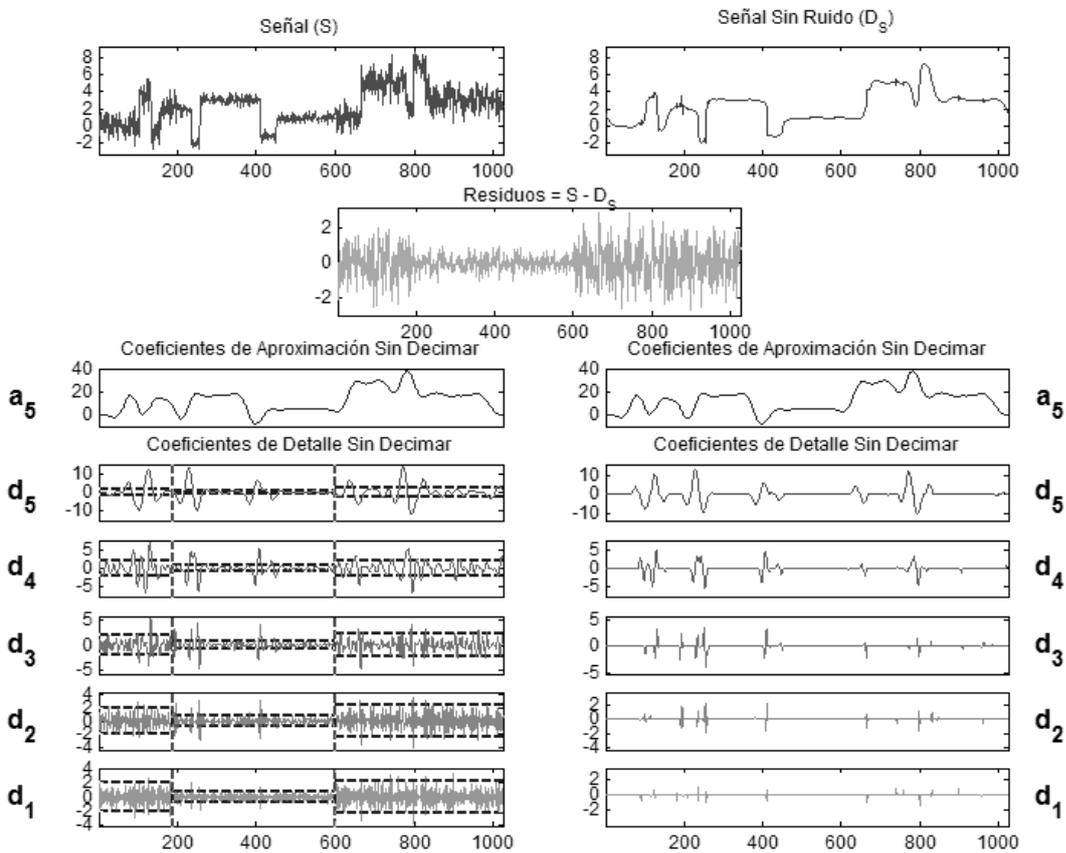
⁶ Consultar apéndice 1 para más información acerca de cómo se obtuvieron estos resultados.

Figura 85. Señal con ruido de múltiples varianzas.



El primer paso, como se mencionaba, es el de establecer los puntos donde la varianza cambia; una vez establecidos se puede proceder a determinar la varianza en estos intervalos y con esto, los valores umbrales correspondientes. Escogidos estos valores: ¿cuál método de aplicación del umbral es conveniente? En el ejemplo anterior la selección del umbral suave llevó a una señal sin discontinuidades, mientras que la selección del umbral fuerte conlleva a una señal con más discontinuidades que puede conservar más sus bordes y forma. La figura 86 muestra el proceso de descomposición de la señal y la aplicación de los distintos niveles de umbral para cada intervalo sobre los detalles. En la figura 87 se muestra la diferencia de aplicar el umbral duro y el umbral suave.

Figura 86. Reducción de ruido usando múltiples umbrales.



En este caso la elección del umbral duro presenta un PER del 93.11% mientras que el umbral suave presenta un PER del 87.75%. Por otra parte, el error con el umbral duro es de 33.80% y el del umbral suave es de 17.82%. Una diferencia de 5.36% del PER produce casi el doble de error al aproximar la señal reconstruida.

A partir de estos resultados se puede observar que la elección del umbral suave implica una mayor pérdida en nivel de energía de la señal pero un nivel de error menor en su aproximación. Por el contrario un umbral duro proporciona un nivel de energía mayor pero conlleva a un error mayor en la señal sintetizada. La figura 88 muestra la comparación de residuos. Arriba se muestran los residuos obtenidos aplicando solamente un valor umbral y abajo se muestran los residuos obtenidos al aplicar varios umbrales.

Figura 87. Aplicación del umbral suave y duro.

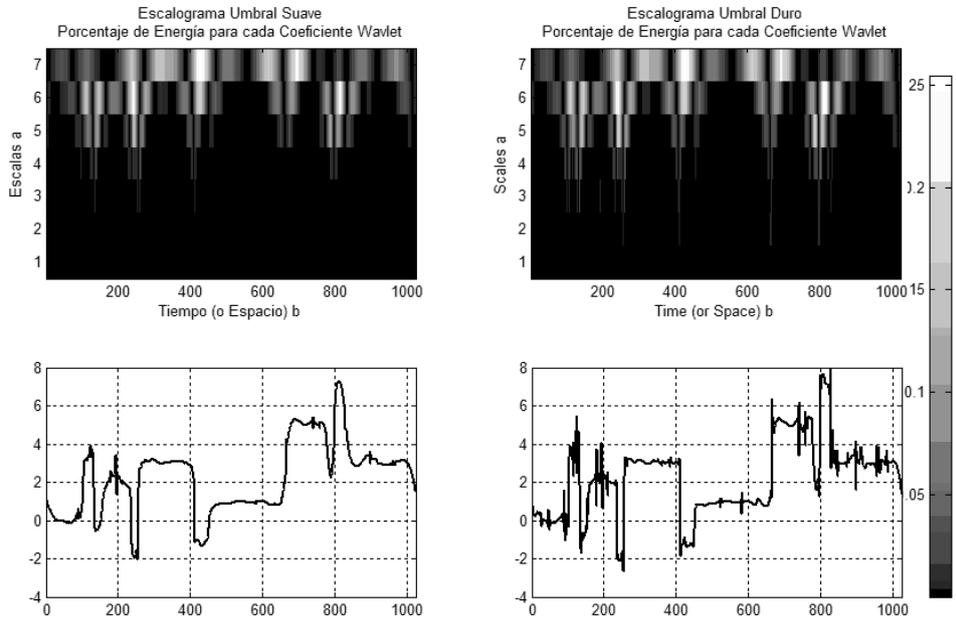
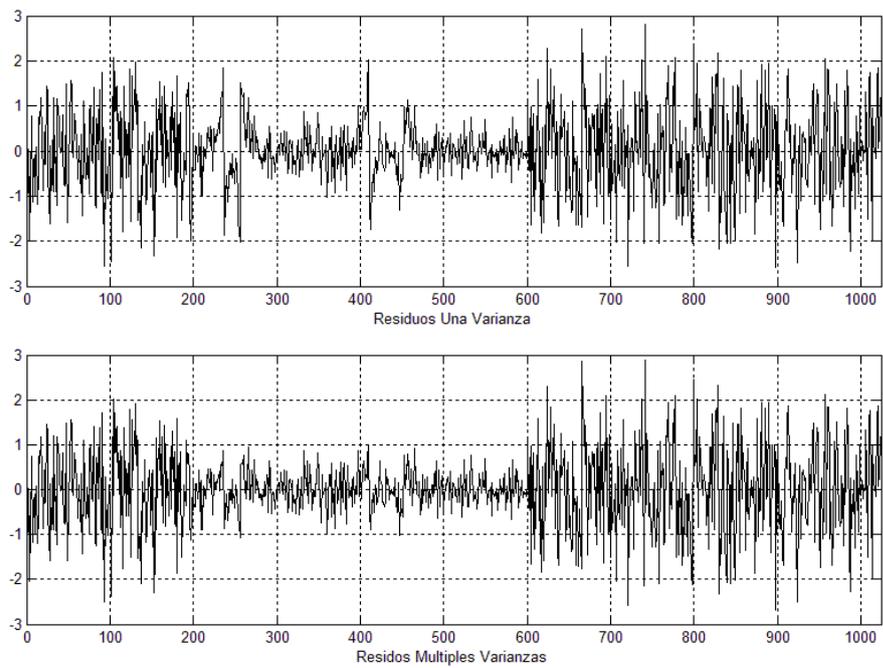


Figura 88. Comparación de residuos.



Se puede observar que en el intervalo de 200 a 600 se suprimió algo más que ruido. Abajo, con la división en distintos intervalos, se observa una región más uniforme en este mismo intervalo.

7.3.3.2 Ruido con estructura desconocida

En el análisis de señales reales con ruido, en particular las de audio, el conocimiento de la estructura del ruido es totalmente desconocida. Se pueden considerar dos casos: que el ruido sea desconocido pero que tenga una varianza homogénea por nivel de escala o que; el ruido sea desconocido y que posea un varianza dependiente del tiempo y la escala.

En el caso de que el ruido posea una varianza homogénea por nivel se debe de establecer un valor de umbral sobre cada nivel de detalle. Este umbral se establece al encontrar la varianza σ_j de cada detalle, siendo esto una extensión de lo descrito 7.3.2.

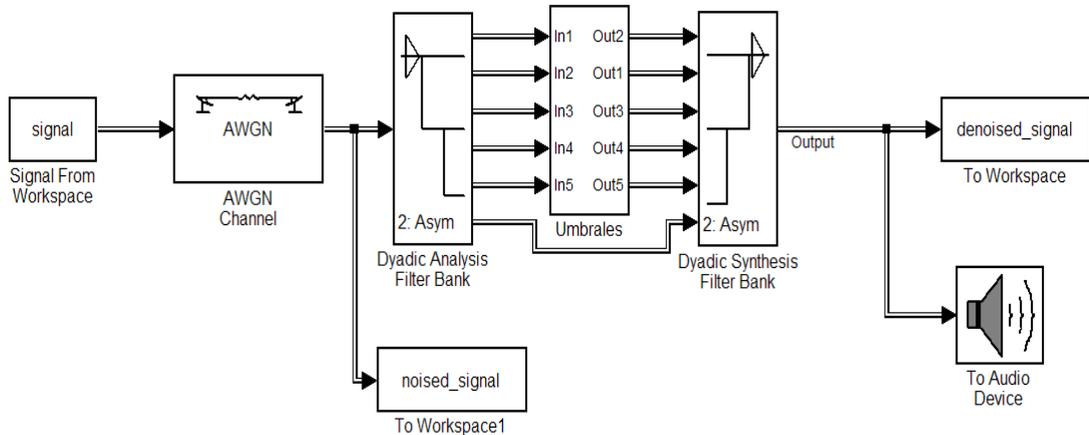
Si el ruido que posee la señal se supone variante con el tiempo y la escala se debe trabajar sobre cada nivel de detalle. En este caso el procedimiento es una extensión de lo explicado en 7.3.3.1, pero ahora este proceso se debe de realizar en cada nivel.

7.3.4 Aplicación: reducción de distintos niveles de ruido a una señal de audio

En los apartados anteriores se mostró la manera en que se puede eliminar el ruido, estacionario o no, de una señal. A partir de esto, resulta la cuestión de qué tan viable podría ser implementar este tipo de procesamiento a señales reales, específicamente a señales de audio.

La figura 89 presenta una simulación realizada en la versión 7.1 de *Simulink* en conjunto con el *Signal Processing Blockset*.

Figura 89. Diagrama de reducción de ruido usando Simulink.



La simulación consiste en el análisis de una señal de audio en formato WAV que se subdivide en pequeñas tramas de 512 muestras cada una, la cual se especifica en el bloque SIGNAL. Esta señal es pasada por un canal que le introduce ruido Gaussiano blanco, AWGN, cuya relación S/N se especifica en este bloque. Luego esta señal con ruido es descompuesta por medio de un análisis diádico de banco de filtros de 5 niveles utilizando la wavelet Symlet de cuarto orden. Al realizar esta descomposición se obtienen 5 niveles de detalle y una aproximación. El umbral universal suave fue el utilizado y se especifica con el bloque UMBRALES.

Se utilizaron tres distintos criterios para establecer el umbral que se le aplicaría a cada detalle, estos fueron⁷:

1. A partir del nivel de ruido $\hat{\sigma}_1$ del primer nivel de detalle (UPD).
2. A partir del nivel de ruido $\hat{\sigma}_j$ sobre cada nivel de detalle. (UCD)
3. Discriminación a partir de lo escuchado, es decir ensayo y error.

⁷ Ver apéndice 2 para los distintos algoritmos.

Posteriormente, con estos nuevos detalles y la aproximación se procede a reconstruir la señal usando un banco de filtros diádico con las mismas características usadas en el análisis. La simulación se puede realizar en tiempo real, por lo que además es muy conveniente para aplicaciones que podrían necesitar un algoritmo con este requerimiento.

Tabla XI. Energía y potencia al aplicar distintos criterios de umbral.

Relación S/N	Umbral Sobre Primer Detalle		Umbral Sobre Todos los Detalles		Umbral Escuchado	
	Energía Retenida	Potencia Umbral	Energía Retenida	Potencia Umbral	Energía Retenida	Potencia Umbral
90	0.96	-35.32	0.97	-37.52	0.99	-40.80
40	0.95	-35.10	0.97	-37.44	0.99	-40.74
34	0.95	-34.58	0.97	-37.22	0.99	-40.57
32	0.94	-34.25	0.97	-37.05	0.98	-38.84
30	0.94	-33.81	0.97	-36.81	0.97	-36.72
28	0.93	-33.24	0.97	-36.45	0.96	-35.28
26	0.92	-32.54	0.96	-35.95	0.93	-33.46
24	0.90	-31.71	0.96	-35.29	0.91	-32.13
22	0.88	-30.75	0.95	-34.47	0.87	-30.51
20	0.85	-29.70	0.93	-33.50	0.82	-29.00
18	0.81	-28.58	0.92	-32.44	0.77	-27.70
15	0.73	-26.90	0.88	-30.85	0.68	-26.06
12	0.64	-25.35	0.84	-29.48	0.58	-24.48
10	0.58	-24.47	0.82	-28.76	0.50	-23.43
8	0.52	-23.73	0.79	-28.20	0.43	-22.53
5	0.46	-22.90	0.77	-27.65	0.34	-21.51
0	0.39	-22.15	0.75	-27.20	0.25	-20.37

Figura 90. Relación S/N vs. potencia umbral.

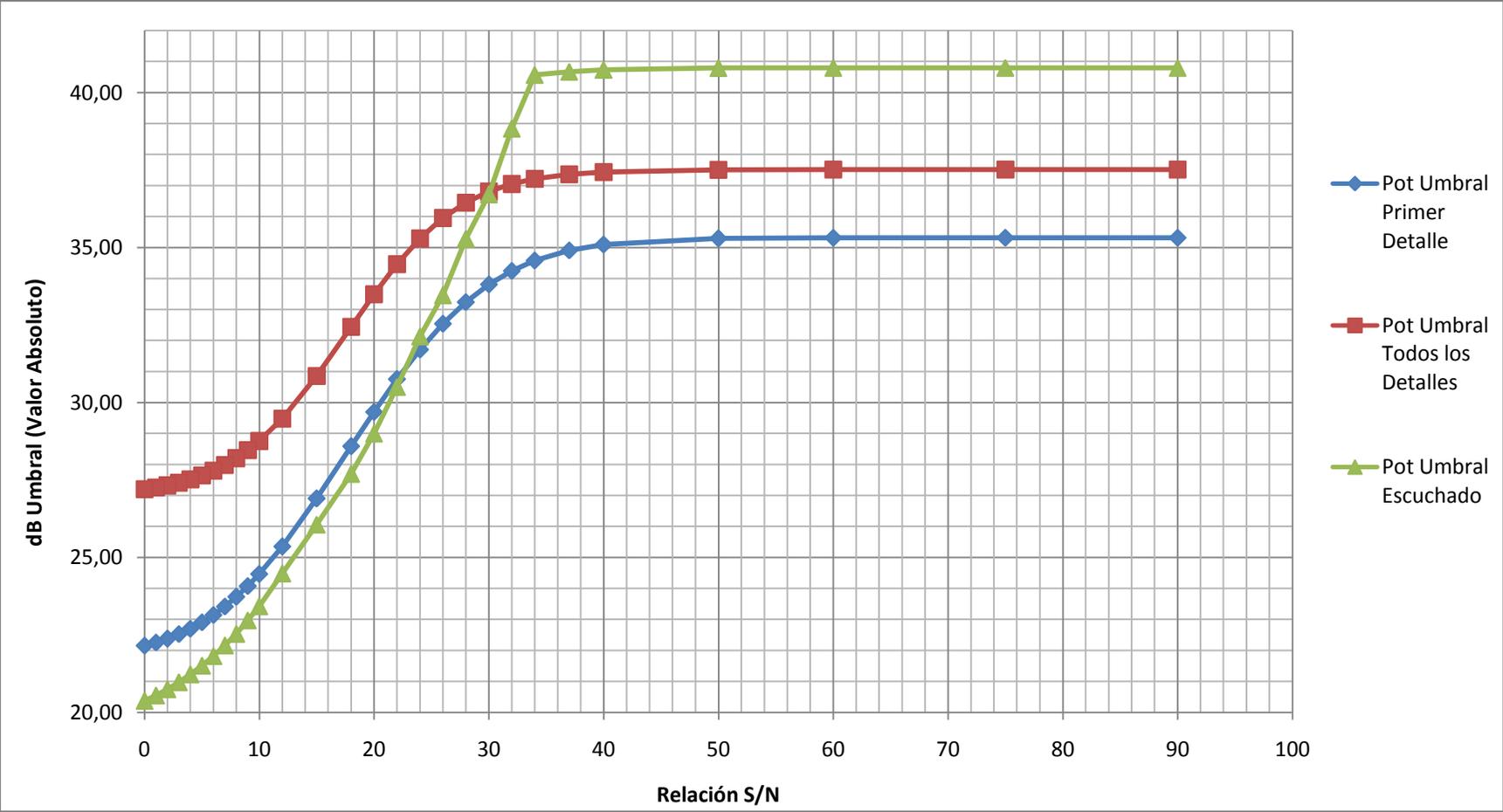


Figura 91. Relación S/N vs. energía retenida.

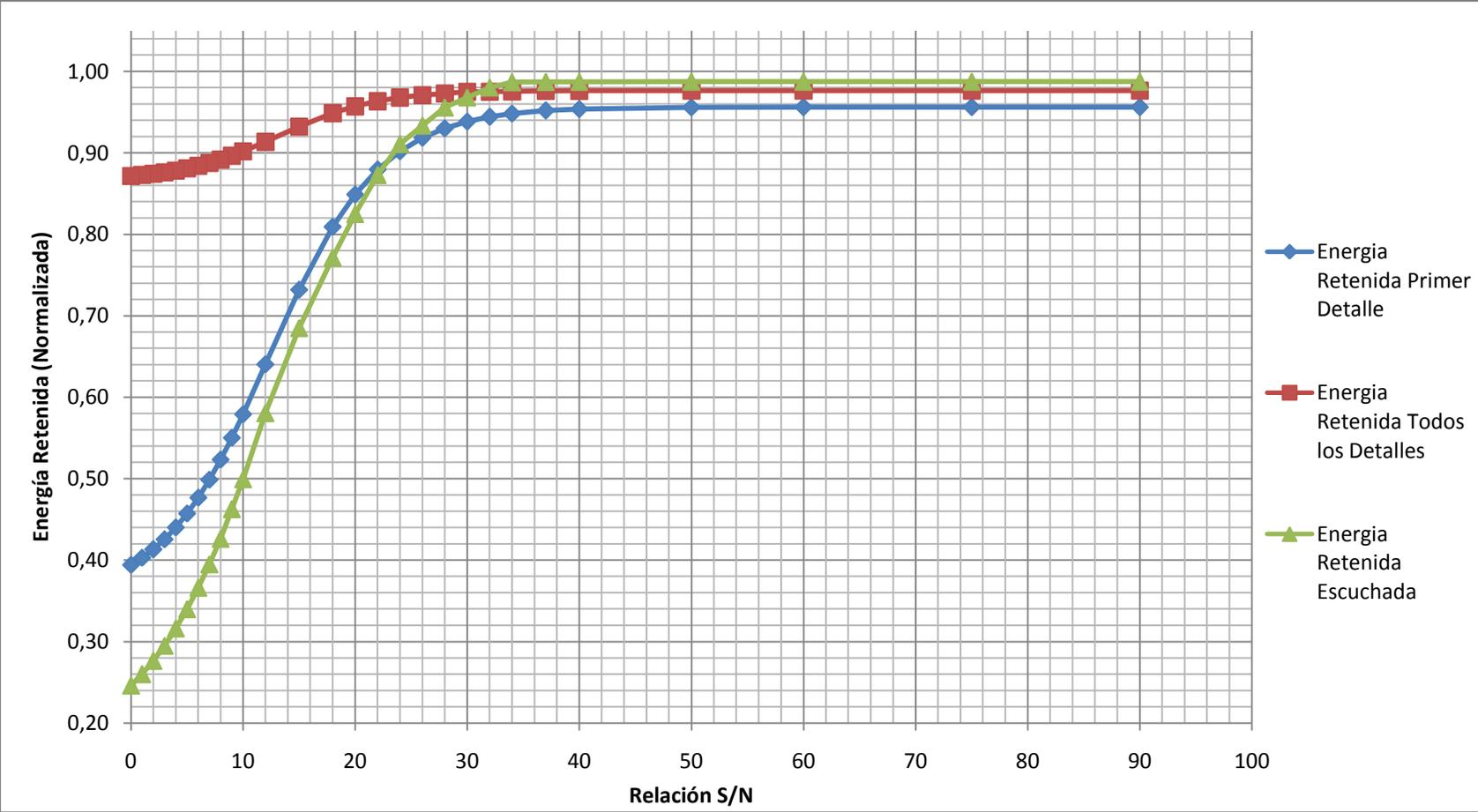
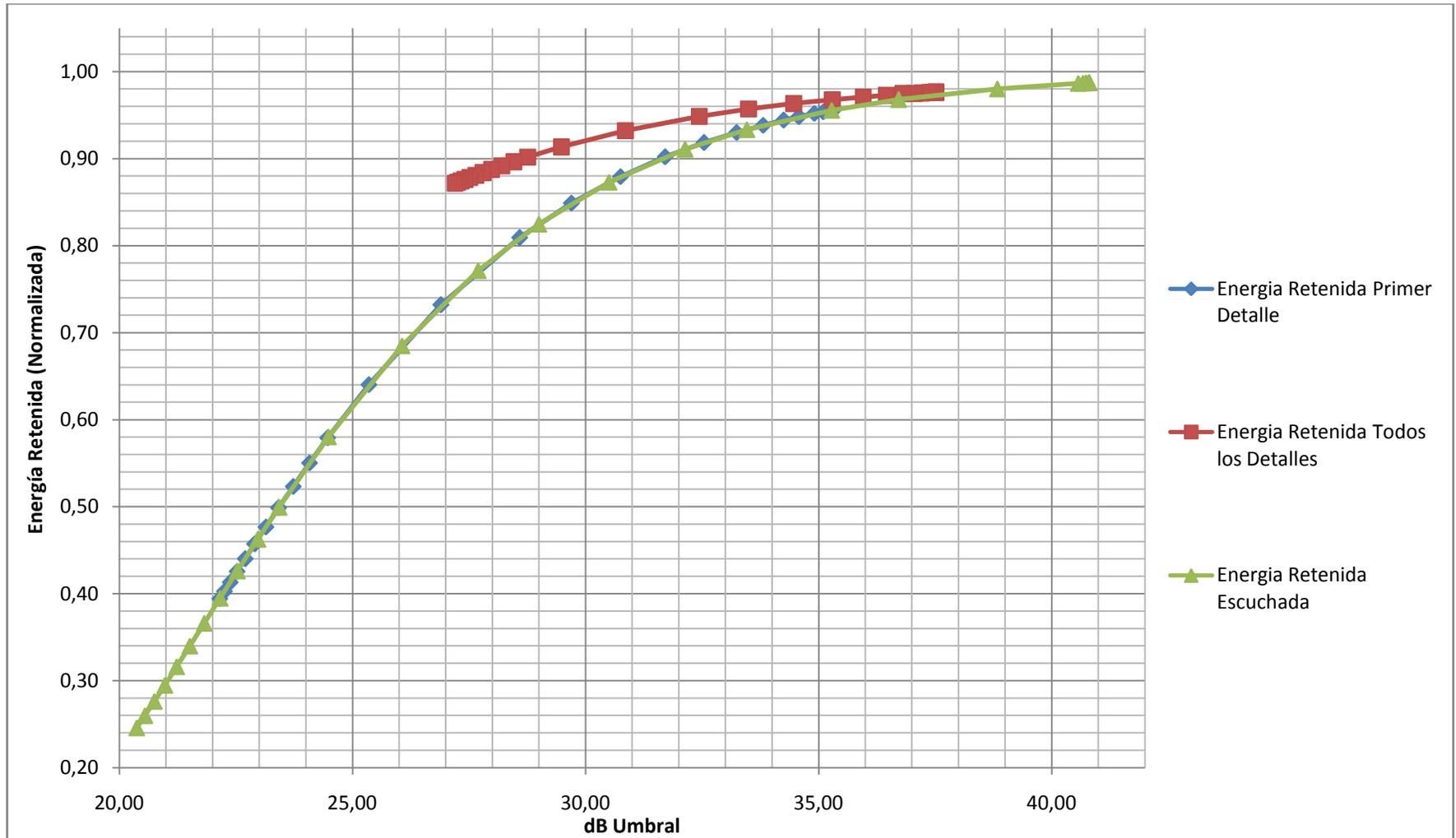


Figura 92. Nivel de potencia del umbral vs. energía retenida.



La tabla XI muestra los datos obtenidos en cada una de las distintas simulaciones y a partir de estos se obtienen las graficas de las figuras 90,91 y 92 que permitirá hacer un análisis cualitativo de las mismas.

En la figura 90 se muestra la grafica de la relación S/N vs. la potencia de cada umbral. Se observa que la potencia usada para discriminar por el UCD es mayor que la usada por el UPD. Es de especial interés la región que se encuentra entre los 0dB y los 40dB de la relación señal ruido ya que presenta el comportamiento dinámico de cada umbral. Se puede notar que el UPD funciona mucho más próximo al que el oído discriminaría, teniendo un comportamiento casi igual a este sobre los 23dB. A partir de los 28dB se observa como el UCD se aproxima mucho mejor a la respuesta del oído, teniendo un valor muy próximo a este sobre los 30dB. Después de los 40dB se ve como ambos umbrales buscan estabilizarse a un valor máximo de potencia. En esta región la eficiencia del UCD es mucho mejor frente al UPD.

La figura 91 muestra la relación S/N vs. la energía retenida de la señal. Es interesante observar como la energía retenida por el UPD es superior a la que se obtiene a partir de un umbral establecido por el escucha hasta los 23dB, que es el punto cuando el nivel de potencia discriminada por el UPD es igual a la del oído. Lo mismo sucede a los 30dB para UCD. Posteriormente los valores de la energía retenida es menor a la obtenida al discriminar por el oyente estabilizándose a partir de los 40dB. Es de notar como el objetivo del UCD es el de mantener la energía retenida, aunque esto implique más ruido en la señal resultante.

En la figura 92 se muestra el nivel de potencia del umbral vs. la energía retenida. A partir de esta gráfica y lo anteriormente mencionado se pueden establecer las siguientes conclusiones acerca de los tres métodos de umbral. En primer lugar se observa como UPD trabaja más parecido al comportamiento del oído, y justifica la elección de un umbral de este tipo cuando se desea un comportamiento como el mismo.

En segundo lugar se ve que un UCD proporciona una cantidad de energía retenida mayor al trabajar sobre niveles de potencia del umbral más altos. Sin embargo a partir de los resultados obtenidos y a la percepción del oyente, que es lo que interesa al final de todo, se puede establecer que un UPD funciona lo suficientemente bien como para ser adoptado a la hora de realizar una limpieza de ruido. Proporcionando además la ventaja de que necesita menos cálculos de procesamiento que el UCD al requerir el cálculo de un solo umbral.

7.4 Compresión de señales de audio

La compresión de señales tanto en el contexto de señales de audio, imágenes y otras más connota al proceso de reducir y crear una representación de esta señal que use una cantidad menor de bits que la señal original. El objetivo principal es el de reducir los requerimientos de almacenamiento o los tiempos de transmisión cuando esta información es comunicada sobre un canal. Idealmente se desea que esta compresión sea reversible. Esto es, dados ciertos datos comprimidos, se desea poder reconstruir la señal original. Cuando esto es posible, se dice que el proceso de compresión es sin pérdidas; de otra manera, se tiene una compresión con pérdidas.

Supongamos ahora el caso de una señal discreta $s(n)$ que se comprime y luego se reconstruye llevando a la señal $\hat{s}(n)$. El error en la señal se encuentra definido como

$$e(n) \equiv s(n) - \hat{s}(n) \quad (7-8)$$

Con esto la señal reconstruida puede ser vista de forma alternativa como una versión contaminada por ruido aditivo de la señal original. Por lo que al reescribir (7-8) como

$$s(n) = \hat{s}(n) + e(n) \quad (7-9)$$

Lo que resulta en algo que ya anteriormente se ha trabajado, la ecuación (7-1). En el caso de que $e(n)$ sea igual a cero se tiene una compresión sin pérdidas.

La noción detrás del proceso de compresión se encuentra basada en que las regularidades de la señal pueden ser bien aproximadas usando los siguientes elementos: un número pequeño de coeficientes de aproximación (de acuerdo a un nivel elegido adecuadamente) y algunos de los coeficientes de detalle.

El algoritmo de compresión es muy similar al proceso de reducción de ruido, y consiste igualmente de tres etapas:

1. Descomposición de la señal por medio de la DWT.
2. Selección y aplicación del umbral a los coeficientes, en este caso se aplica el umbral duro.
3. Reconstrucción de la señal a partir de la IDWT.

7.4.1 Métodos de compresión

7.4.1.1 Umbral global

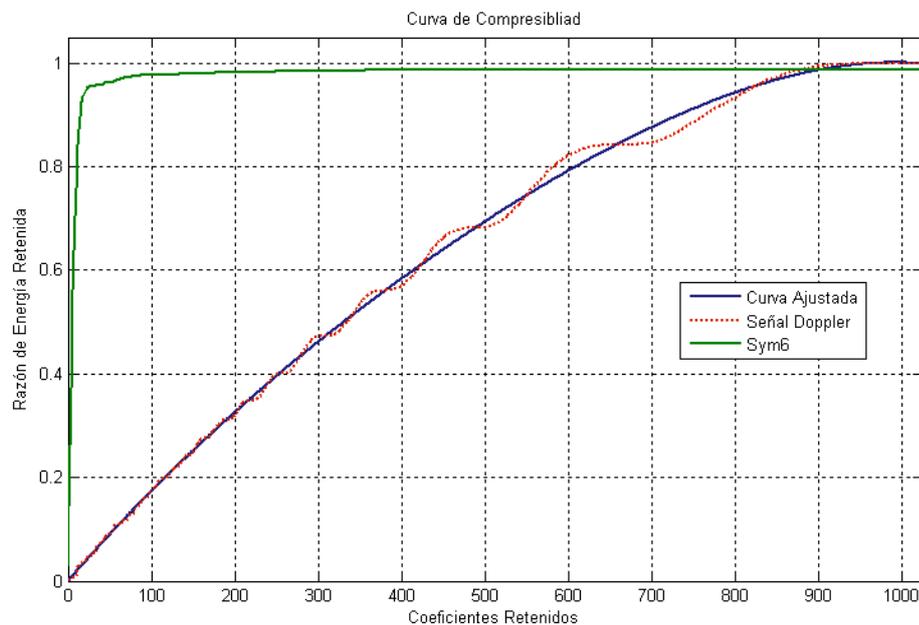
La manera más sencilla de trabajar la compresión por medio de wavelets es el de tomar la función de energía finita f y considerar una base ortonormal L^2 wavelet, a partir de la cual se obtendría una expansión de f sobre este espacio.

Al obtener una expresión de f en este nuevo espacio, el objetivo principal del umbral global sería el de mantener los M coeficientes de valor absoluto más grandes, para obtener una aproximación f_M de la señal. En general, esta descomposición conlleva cierta pérdida. El umbral global se calcula al ajustar el porcentaje de coeficientes a ser truncados.

En la sección 7.3.2 se trabajó con una señal de efecto Doppler con ruido, partiendo de esta señal se puede construir una curva de compresibilidad para la misma. Si en el eje x se presenta el número de coeficientes preservados y en el eje y la razón de energía retenida luego de la compresión, se obtiene la figura 93.

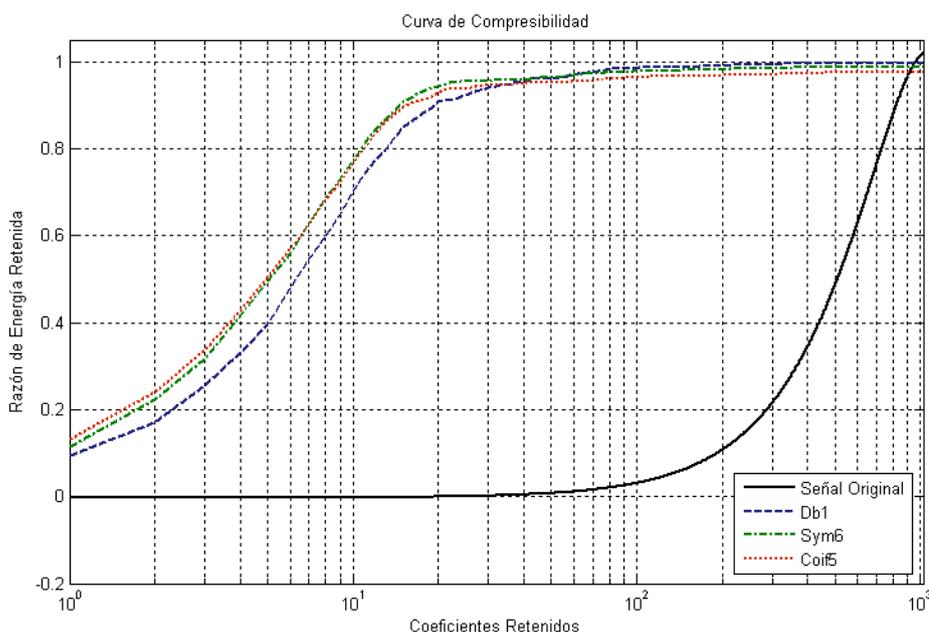
Se puede observar que una buena cantidad de la energía de la señal se concentra en pocos coeficientes de la representación wavelet en comparación de la representación usual de la señal. La señal ha sido descompuesta hasta un octavo nivel usando la wavelet *sym6*. El coeficiente más grande de la descomposición, el cual es el coeficiente de aproximación del octavo nivel, contiene 11.4% de la energía. Mientras que, el coeficiente más grande de la señal contiene únicamente un 0.23% de la energía.

Figura 93. Curva de compresibilidad de la señal Doppler.



Con otro tipo de wavelets la curva de compresibilidad esencialmente posee el mismo aspecto. Por ejemplo, en la figura 94 se presentan las curvas de compresibilidad para las wavelets *db1*, *sym6* y *coif5* en una escala logarítmica para los coeficientes retenidos. Se puede observar que mientras más momentos de desvanecimiento posea la wavelet usada, mejor será la capacidad de compresión del espacio wavelet. Sin embargo, el costo de esto es una menor capacidad de retener toda la energía de la señal si se usaran todos los coeficientes. Se puede mostrar que estos valores de energía son tolerables si ven las diferencia obtenidas para *db1-sym6* que es de %0.88 y para *db1-coif5* que es de 2.17%.

Figura 94. Curva de compresibilidad para distintas wavelets.



7.4.1.2 Umbral por nivel

Esta estrategia se basa en la propuesta por *Birge* y *Massart*; la cual, consiste en dejar todos los coeficientes de aproximación en el nivel de descomposición J , mientras que el número de coeficientes de detalle mantenidos del nivel j al nivel J está dado por:

$$n_i = \frac{M}{(J + 2 - j)^\alpha} \quad (7-10)$$

Siendo α un parámetro de compresión, el cual usualmente es igual a 1.5. La elección de este método se basa en la rápida convergencia de f que puede llegar a alcanzarse. El valor de M denota la escasez de distribución de los coeficientes wavelet. Si L denota la longitud de los coeficientes de aproximación más gruesos, nivel más alto, el valor de M toma los valores de la tabla XII, a partir de la señal utilizada.

En resumen, este método consiste en mantener una aproximación tosca y seleccionar unos cuantos coeficientes finos para los niveles bajos de detalle, que contengan las fluctuaciones de alta frecuencia de la señal.

Tabla XII. Valores de M.

Escasez	M
Alta	L
Media	$1.5L$
Baja	$2L$

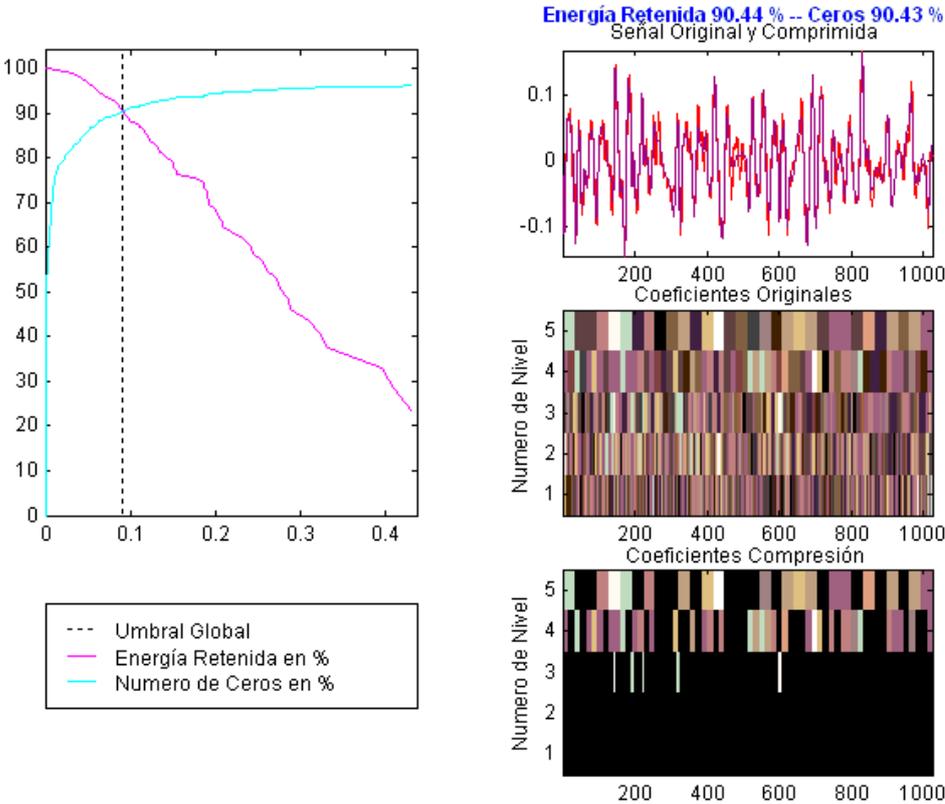
7.4.2 Comparación de los umbrales de compresión

En los apartados anteriores se presentaron los dos tipos de umbral aplicados para la compresión de señales. A continuación se presenta la aplicación de estos umbrales a una señal de audio de 1024 muestras, el cual podría ser el punto de partida para un esquema similar al usado en la sección 7.3.4.

La figura 95 muestra los resultados obtenidos de la compresión de la señal de audio mencionada. En este caso se utilizó la wavelet *sym6* para aplicar el umbral global.

La grafica del lado izquierdo de la figura 95 presenta dos curvas: el porcentaje de energía retenida frente al valor del umbral global seleccionado; y el porcentaje de ceros frente al valor del umbral global. Se observa que conforme el valor umbral aumenta la energía retenida va disminuyendo y la cantidad de coeficientes que se vuelven cero van aumentando. El valor umbral fue seleccionado en el punto donde el porcentaje de energía retenida y el porcentaje de ceros fuera la misma con la finalidad de que la energía sacrificada fuera compensada por el porcentaje de ceros introducido. Los resultados muestran que el porcentaje de energía retenida fue del 90.44% mientras que el porcentaje de ceros fue del 90.43%. Las tres graficas del lado derecho de la figura 95 presentan los resultados antes y después de la compresión realizada.

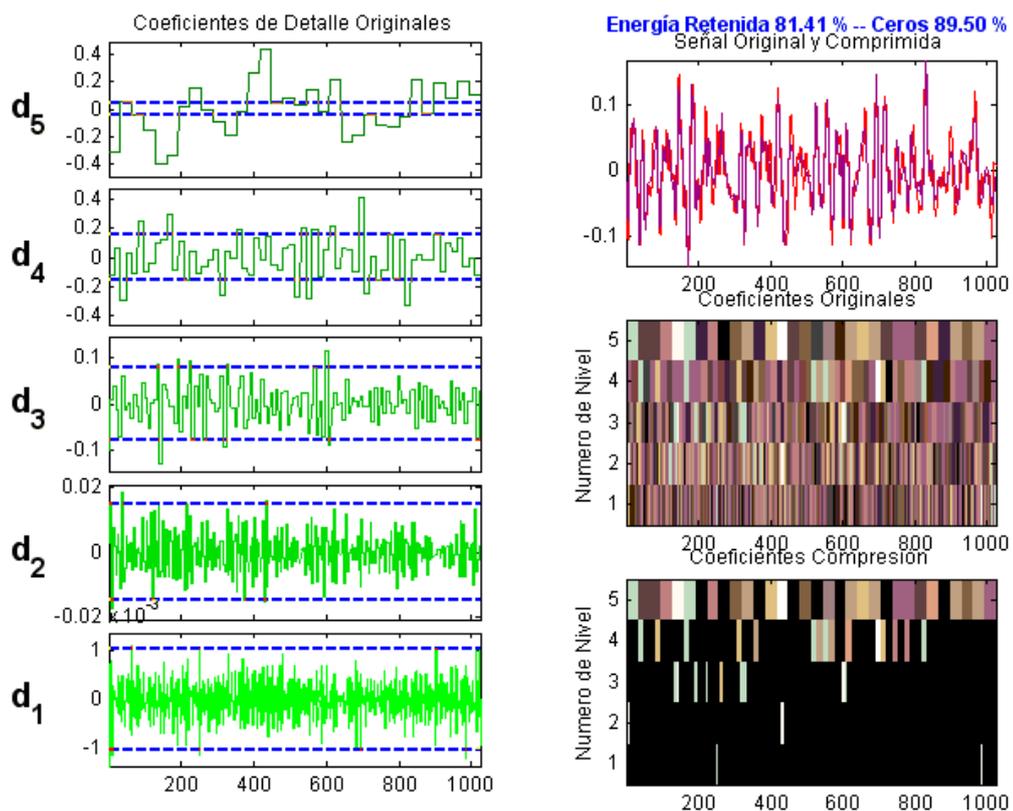
Figura 95. Compresión aplicando el umbral global.



En la figura 96 se presenta la compresión de la misma señal de audio usando el criterio del umbral por nivel y la wavelet *sym6*. Las graficas al lado izquierdo de la figura muestran los distintos niveles de umbral obtenidos a partir de (7-10), donde la escasez de coeficientes fue asumida como baja.

Los resultados muestran que con la aplicación de este umbral el porcentaje de energía retenida fue del 81.41% y el porcentaje de ceros fue del 89.50%. Las graficas del lado derecho muestran los resultados antes y después de la compresión.

Figura 96. Compresión aplicando el umbral por nivel.



Los resultados obtenidos para los dos umbrales muestran que el umbral global retiene un mayor porcentaje de energía de la señal frente al umbral por nivel con casi el mismo porcentaje de ceros para ambos casos.

Sin embargo desde el punto de vista de la calidad de audio de la señal se puede observar a partir de los escalogramas resultantes de la figura 95 y la figura 96 que el umbral por nivel retiene componentes de otros niveles que el umbral global descartó totalmente.

Ambos ejemplos mostraron que la compresión conlleva a una pérdida de la señal original. En el caso de que la señal comprimida contenga algún tipo de ruido, al comprimirla, también se eliminaría gran parte del ruido al preservar los coeficientes más grandes. La diferencia entre la compresión de señales y la eliminación de ruido radica en la forma en que se escoge el nivel de umbral. En el caso de la compresión, el umbral viene dado por limitantes externos como: el ancho de banda, memoria disponible, razones de compresión. Por el contrario, en el caso de la reducción de ruido el umbral se selecciona de forma automática al estimar el nivel de ruido usando la información dada por el modelo para reducirlo.

8. WAVELETS EN LA CODIFICACIÓN PERCEPTUAL DE AUDIO

8.1 Introducción

Los codificadores perceptuales son codificadores con pérdidas que introducen error de codificación en la señal codificada mientras tratan de minimizar la percepción de sus efectos. El error de codificación es el resultado del pre filtrado, aliasing, y componentes de cuantización como son el (re)muestreo y la (re)cuantización. De estas tres, el error de cuantización es la que típicamente se desea minimizar. Codificadores tradicionales han confiado en mediciones de distorsión como el error cuadrático medio para minimizar el error de codificación. El resultado ha sido una optimización de la relación señal a ruido, SNR, pero no necesariamente una mejora en la calidad de cómo el escucha lo percibe. Sin embargo, en la codificación perceptual el ruido de codificación es modelado de tal manera que la calidad percibida de la señal de audio sea optimizada de acuerdo a algunos criterios subjetivos.

El criterio de distorsión perceptual en un codificador perceptual usualmente es calculado por una etapa psicoacústica que trata de modelar el comportamiento del sistema auditivo humano. El criterio perceptual, al que comúnmente se refiere como la mínima distorsión notable, JND, o el umbral de enmascaramiento, provee del umbral debajo del cual el ruido de codificación se mantiene imperceptible.

El umbral de enmascaramiento usualmente se presenta por una curva en el dominio de la frecuencia que cubre el rango de audición humana, y se representa de forma más general por un contorno en el tiempo-frecuencia cuando se toma en cuenta el enmascaramiento temporal. Un codificador perceptual esencialmente trata de controlar el ruido de cuantización a manera de que su forma permanezca por debajo del umbral de enmascaramiento en el dominio de la frecuencia.

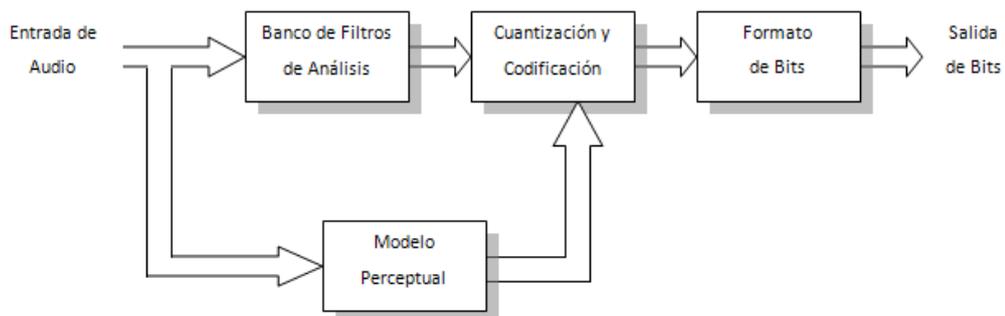
Este capítulo pretende dar una breve descripción de los requerimientos básicos para una codificación perceptual de audio. Posteriormente se hace énfasis en los requerimientos del WFB, resolución temporal y de frecuencia, a partir de los cuales se plantea los problemas que implican su uso. Finalmente se presenta una solución a estos problemas y lo que implica en el diseño del codificador cuando este es implementado.

8.2 Estructura de un codificador perceptual de audio genérico

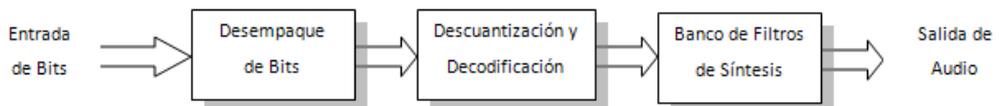
La figura 97 muestra el diagrama de la estructura de un codificador y un decodificador perceptual de audio genérico. El codificador, que se muestra en la figura 97(a), trabaja en la entrada con la señal de audio y proporciona los bits de salida codificados. El decodificador, mostrado en la figura 97(b), trabaja con los bits de entrada codificados y reconstruye la señal original a la salida. Las tres etapas en el decodificador, como resultado, son las operaciones inversas de las tres etapas en el codificador. Más específicamente, el banco de filtros de análisis, la cuantización y codificación, y la etapa de formato de bits del codificador corresponden al banco de filtros de síntesis, la descuantización y decodificación, y al desempaque de bits del decodificador, respectivamente.

La etapa extra en el codificador es el modelo psicoacústico, que no es necesario en el decodificador ya que la información es implícitamente codificada como información adjunta. Por lo que los codificadores perceptuales son asimétricos ya que el codificador necesita de un mayor poder computacional que el decodificador. Esto lo hace deseable en ciertas aplicaciones donde un servidor codifica la señal para muchos clientes.

Figura 97. Codificador y decodificador perceptual de audio genérico.



(a) Codificador



(b) Decodificador

El codificador funciona de la siguiente manera. La señal de entrada típicamente es segmentada en bloques o *frames* contiguos para que cada bloque pueda ser procesado individualmente. Como se vio en 7.3.4 algunas de las ventajas de la segmentación fueron: la reducción de la carga computacional, la localización de la señal en el tiempo y la operación en tiempo real.

El poder localizar la señal en el tiempo significa el poder aplicar un enmascaramiento en el dominio de la frecuencia. Dentro del codificador, el bloque de entrada primero entra al banco de filtros y a la etapa psicoacústica. El banco de filtros transforma la señal en una representación en el dominio de la frecuencia o en una representación conjunta de tiempo-frecuencia.

El modelo psicoacústico primero aplica una transformación de alta resolución en frecuencia y luego aplica reglas psicoacústicas para calcular el dominio en frecuencia del umbral de enmascaramiento. La salida de ambos, el banco de filtros y el modelo psicoacústico, es pasada por la etapa de cuantización y codificación donde ocurre realmente la reducción de bits. La etapa de codificación y cuantización decide como son asignados los bits entre los coeficientes del banco de filtros y un cuantizador es usado para (re)cuantizar los coeficientes del banco de filtros. Muchas veces, una etapa adicional de codificación sin pérdida es aplicada para remover también la redundancia estadística. Los coeficientes cuantizados, junto con ciertos bits de identificación, son finalmente formateados en los paquetes de salida. El decodificador, en la etapa de recepción, simplemente aplica las operaciones inversas.

8.2.1 Consideraciones del banco de filtros

La elección del banco de filtros óptimo ha sido un tema de mucha investigación y discusión en el desarrollo de codificadores perceptuales. Sin embargo existen ciertas características de diseño que se pueden tomar en cuenta.

A continuación se presentan estas características.

1. Resolución tiempo-frecuencia

Los bancos de filtros deben considerar la resolución tiempo-frecuencia del sistema auditivo humano, la resolución que provea de la mejor ganancia de codificación, y la resolución que provea de un control adecuado de los eventos en el tiempo. Es claro que un banco de filtros fijo no puede cumplir con todos los requerimientos antes descritos y de ahí que se necesite de un esquema adaptivo.

2. Separación de canales

También conocida como la localización en frecuencia, se refiere a que tan bien un canal del banco de filtros es separado de otros canales. En la práctica el traslape entre las distintas bandas siempre ocurre en los bancos de filtros debido al carácter no ideal de operación de los filtros.

La separación entre canales es importante en el contexto de codificación perceptual debido a que se asume que los canales adyacentes son independientes y no se traslapan cuando se aplican los resultados perceptuales. Esto es muy importante para las componentes tonales de la señal, que requieren una alta resolución en frecuencia pero también una alta separación en frecuencia. La magnitud de cuan separados se encuentran los canales ha demostrado tener un impacto directo en el desempeño del codificador perceptual.

3. Manejo de fronteras y bloqueo de artefactos

La segmentación o enventanado de la señal de entrada en pequeños bloques conlleva al bloqueo de artefactos en las fronteras de los mismos. Esto puede ser percibido como distorsiones en la señal reconstruida, particularmente para porciones de la señal que son estacionarias. Para evitar el bloqueo de artefactos, la codificación del ruido debe poseer cierta correlación en las fronteras de los bloques. La minimización de distorsiones en la frontera típicamente se hace al aplicar una ventana de suma y traslape, o al usar una transformada *lapped*.

4. Reconstrucción perfecta

La reconstrucción perfecta requiere que la señal reconstruida sea idéntica a la señal de entrada, con un posible retraso, en la ausencia de cualquier error de codificación. Aunque la reconstrucción perfecta no es un requerimiento estricto, ya que hay codificadores que no utilizan bancos de filtros con esta característica, generalmente simplifica el diseño del sistema de codificación.

5. Decimación máxima

Los bancos de filtros máximamente decimados, o críticamente muestreados, proveen del mismo número de coeficientes en el dominio de la transformada aplicada que el número de coeficientes en el dominio del tiempo de la señal original.

Es decir, el número de muestras entrantes por segundo en el tiempo es igual al número de muestras por segundo en el dominio de la frecuencia. Como el objetivo principal es el de disminuir la tasa de datos mientras se mantiene la calidad de audio, los sistemas críticamente muestreados son deseados.

6. Complejidad computacional

Aunque ahora ha venido siendo de menos importancia, la complejidad computacional y el retraso en la codificación aún pueden ser requerimientos importantes para algunas aplicaciones. Por lo que los bancos de filtros de análisis y síntesis deben proveer de algoritmos rápidos y eficientes a la hora de ser aplicados.

8.2.2 Modelo psicoacústico

El objetivo del modelo psicoacústico para un codificador de audio es el de determinar el umbral de enmascaramiento generado por una señal localizada en el tiempo, donde la máscara puede ser asumida como ruido (blanco) de cuantización. Idealmente, esto requeriría encontrar la curva de enmascaramiento de la señal de entrada tomando en cuenta el enmascaramiento temporal y en frecuencia, con lo cual la curva resultante se podría representar en el plano tiempo-frecuencia. Aún más, el modelo psicoacústico necesitaría manejar todos los tipos de señales que contuvieran cualquier combinación de componentes tonales y de ruido a distintas frecuencias e intensidades.

Sin embargo, en la realidad, cada modelo psicoacústico realiza ciertas suposiciones y simplificaciones que hacen del diseño práctico, mientras se introducen ciertas inexactitudes. Como resultado, siempre existe un costo entre exactitud y complejidad existente en cada modelo dado. La mayoría de modelos psicoacústicos pueden ser ampliamente divididos en tres categorías, que son: modelos fisiológicos, modelos en base a la excitación de patrones, y modelos en base a patrones de enmascaramiento.

8.2.2.1 Modelo fisiológico

Los modelos fisiológicos tratan de simular los mecanismos internos del sistema auditivo basándose en el conocimiento existente y el entendimiento del oído. Como se mencionó en el capítulo 3, aún existe mucho del sistema auditivo que no es entendido, particularmente la base fisiológica que se observa en los experimentos psicofísicos, con lo que el diseño se hace muy complicado. Sin embargo los modelos fisiológicos representan la forma ideal de aproximar este problema y eventualmente puede llevar a soluciones que provean de soluciones adecuadas.

Un modelo fisiológico propuesto es el de modelar el oído por una serie de etapas que asemejaran el oído externo, medio e interno. De manera más específica, las etapas de procesamiento fueron: 1) un filtro acústico para el oído externo y medio, 2) un banco de filtros de 103 bandas para la membrana basilar y coclear, 3) una celda de cabellos internos rectificadores/filtro paso bajo, 4) una etapa de procesamiento neural y 5) una eta de estimación de la curva de enmascaramiento. Existen otros modelos aún más complejos donde los patrones de excitación fueron obtenidos al usar un modelado extensivo del oído interno.

8.2.2.2 Modelo en base a la excitación de patrones

Los modelos en base a la excitación de patrones trabajan primero al calcular el patrón de excitación que es generado por la señal enmascarada, por ejemplo la señal de entrada; y el patrón generado por la señal enmascarada y la máscara, como por ejemplo la señal reconstruida. Se dice entonces que existe un enmascaramiento completo si la diferencia entre las dos curvas difiere no más de cierto umbral sobre todas las frecuencias. El umbral es tomado entre 0.1 y 1 dB. El funcionamiento de este modelo se basa en que todas las señales, máscara y enmascarada, generan cierto patrón de actividades en las neuronas de acuerdo a un conjunto complejo de reglas y la máscara, como el ruido de cuantización, puede ser detectado solamente en la presencia de una señal, en este caso la señal de entrada, y si esta contribución adicional en la actividad neural excede cierta cantidad que el oído es capaz de detectar.

En general, los patrones de excitación no conllevan a un umbral de enmascaramiento ya que solamente son capaces de indicar si una señal que es ruido puede ser detectada o no. Debido a esto, el modelo basado en la excitación de patrones ha sido sugerido más como una herramienta para evaluar modelos psicoacústicos que para ser usado como un modelo que calcule el umbral de enmascaramiento.

8.2.2.3 Modelo en base a patrones de enmascaramiento

Los modelos basados en los patrones de enmascaramiento son considerados los más sencillos de todos, esto debido a que aplican un conjunto de reglas simples para obtener el umbral de enmascaramiento. Los modelos basados en los patrones de enmascaramiento son los de uso más común en los codificadores de audio actuales. Un modelo basado en los patrones de enmascaramiento contiene las siguientes etapas:

1. La densidad espectral de potencia, PSD, de la señal de entrada primero es calculada y luego es dividida de acuerdo a la escala de bandas críticas.
2. La tonalidad de cada componente de enmascaramiento en la señal es identificada.
3. Las funciones de dispersión son aplicadas para cada componente de enmascaramiento.
4. El umbral de enmascaramiento promedio es encontrado al combinar los umbrales de enmascaramiento individuales mientras se toma en cuenta la tonalidad de cada componente de enmascaramiento.
5. El umbral absoluto de audición también es tomado en cuenta, con lo que un enmascaramiento que se encuentre por debajo del umbral absoluto en cualquier punto es elevado al nivel de este umbral.

Se puede notar que cada paso de este modelo se basa en alguna de las reglas descritas en el capítulo 3. El primer paso se basa en el modelo del espectro de potencia de la sección 3.4.1.1. El segundo paso aplica la asimetría del enmascaramiento de la sección 3.4.3. El tercer paso utiliza la función de dispersión descrita en la sección 3.4.2 para cada componente de enmascaramiento.

En el cuarto paso se asume que las componentes de enmascaramiento trabajan de forma lineal y pueden simplemente ser sumadas. El quinto paso aplica el umbral absoluto de audición descrito en la sección 3.3.

Sin embargo, este modelo posee algunas suposiciones que son importantes de resaltar. Primero, el enmascaramiento temporal no es tomado en cuenta y solo se aplica al dominio de la frecuencia. Segundo, la función de dispersión usada se obtiene de un modelo simple (una máscara- un enmascarado) y se aplica a un modelo más complejo (múltiples máscaras- múltiples enmascarados); además de ignorar la dependencia en la frecuencia de los patrones de enmascaramiento.

Tercero, el modelo de bandas críticas se basa en un modelo discreto mientras que el oído posee un modelo de bandas críticas que es continuo, por lo que los umbrales de enmascaramiento no poseen límites de bandas críticas como los propuestos.

A pesar de estos inconvenientes, los modelos basados en patrones de enmascaramiento han mostrado un buen desempeño con una complejidad relativa en una variedad de codificadores de audio que han alcanzado altas tasas de compresión, ejemplos de estos son el MPEG-1 Layer III, MPEG-AAC y el PAC propietario de Lucent.

8.2.3 Etapa de cuantización y codificación

La meta de la etapa de cuantización y codificación esencialmente es alcanzar una representación de los datos lo más compacta posible mientras se introduce la mínima distorsión perceptual posible. La etapa de cuantización y codificación es la parte inteligente del algoritmo de codificación que determina la estrategia de codificación total.

La etapa de cuantización y codificación usualmente es diseñada en tres sub etapas. Primero, una estructura de control determina como los bits son distribuidos entre los coeficientes espectrales usando los resultados del modelo psicoacústico. Segundo, una etapa de cuantización (o re cuantización) mapea los coeficientes del banco de filtros en una representación de baja resolución de acuerdo a la distribución de bits determinado por la estructura de control. Tercero, los datos cuantizados a veces son codificados por una etapa adicional de codificación sin pérdidas con la finalidad de remover redundancias estadísticas. En general, estas tres etapas trabajan en conjunto de una forma interactiva y pueden ser diseñadas para dar un esquema sencillo como un bloque de compansión, o un esquema más complejo como el de análisis-por-síntesis con codificación sin ruido.

La etapa de cuantización y codificación puede ser diseñada para incorporar cualquier número de técnicas de codificación y herramientas que han aparecido en varios codificadores de audio. En general, la estrategia de codificación en un codificador perceptual involucra mucho su flexibilidad en términos de su complejidad y desempeño, y libertad en términos de su implementación actual. Es aquí cuando implementaciones específicas contribuyen significativamente en el desempeño general. Como ejemplo de esto se tiene la codificación de audio usando el estándar MPEG-1 Layer III que provee de todos los lineamientos para garantizar una inter operabilidad entre distintas implementaciones, pero la suficiente libertad para que distintas implementaciones puedan agregar sus propias mejoras. Esta es la razón del porque existen distintas implementaciones del algoritmo MP3 como el estándar ISO, LAME, y Fraunhofer y sus distintos desempeños.

8.2.4 Codificación sin ruido

La codificación sin ruido típicamente es un paso adicional incluido en la etapa de codificación con la finalidad de obtener una reducción en la tasa de bits. Algunas técnicas conocidas en la codificación de audio incluyen el método de agrupamiento usado en el estándar MPEG-1 Layer II y la codificación Huffman usada en el estándar MPEG-1 Layer III y AAC. Además se ha sugerido el uso de codificaciones como la codificación por duración de longitud (*run-length coding*), codificación aritmética por porciones de bit (*bit-sliced arithmetic coding*) y el algoritmo LZW.

8.3 Resolución tiempo-frecuencia

Se ha observado que la estrategia de codificación de audio podría involucrar dos criterios. El primero, basado en un esquema de compresión estadística (sección 7.4) buscaría maximizar o minimizar la distribución de la señal en el dominio wavelet de acuerdo a un criterio estadístico. Bajo este criterio se encuentran algunos tipos de codificación que adaptan la estructura del árbol de descomposición y el filtro base con lo que optimizan la concentración de energía en el dominio wavelet. El segundo, involucra la codificación perceptual, donde el criterio perceptual es calculado por un modelo psicoacústico, de tal manera que la distorsión de codificación pueda ser modelada para minimizar la distorsión percibida. La aplicación de este criterio perceptual a un WFB significa que la elección de la estructura del árbol de descomposición y el filtro base, y la manera en que cada conjunto de coeficientes de cada banda son codificados, primero necesita considerar los requerimientos perceptuales de la señal de audio y las características del sistema auditivo humano.

A diferencia de la codificación de imágenes, donde se piensa que la descomposición wavelet como un análisis multi-resolución, la codificación perceptual trata al WFB como un banco de filtros que provee una representación en el dominio del tiempo-frecuencia de la señal de entrada.

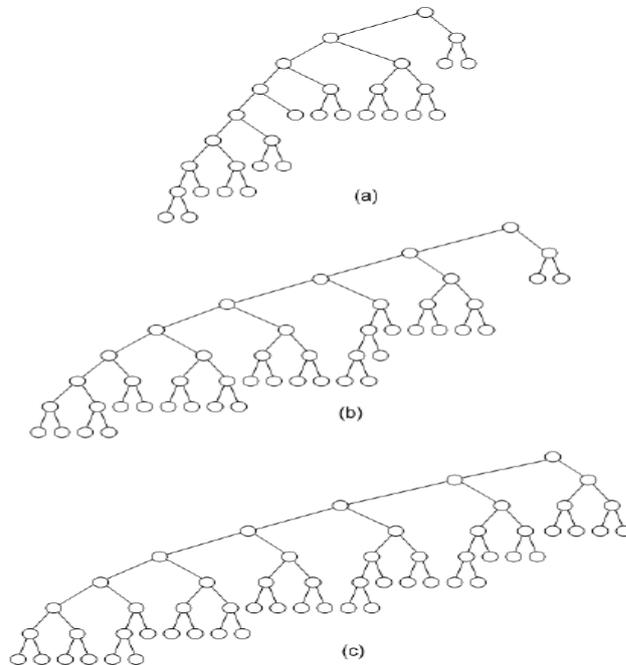
8.3.1 Estructura del árbol wavelet

Como se ha ido observando en el desarrollo de esta tesis, la estructura del WFB determina la resolución en el tiempo-frecuencia de la señal descompuesta y de acuerdo a las propiedades de la señal (estacionarias, transientes o de tonos periódicos) cierta resolución tiempo-frecuencia es preferida. Como los codificadores perceptuales son diseñados para eliminar las irrelevancias perceptuales (y luego las redundancias estadísticas), la resolución tiempo-frecuencia necesita ser derivada del modelo psicoacústico.

Resulta que en muchos casos los modelos psicoacústicos proveen de información en el dominio de la frecuencia con una resolución de banda crítica, y el diseño de muchos codificadores de audio wavelets se han basado en estructuras de árbol con resultados similares a los de bandas críticas. La figura 98 muestra la estructura de tres WFB, en (a) el de 16 bandas, en (b) el de 24 bandas y en (c) el de 29 bandas.

Los codificadores wavelet basados en estructuras de árbol adaptivas han mostrado ser más flexibles cuando se está tratando con señales de audio variantes en el tiempo. Destacan de estos codificadores algoritmos como el propuesto por Wickerhauser que usa un criterio basado en la entropía de la señal, y el propuesto por Srinivasan y Jamieson que se basa en un criterio perceptual. Existen además otros esquemas en desarrollo que tratan de utilizar tanto un criterio estadístico como perceptual.

Figura 98. Estructura de tres WFB.



8.3.2 Base del filtro wavelet

La elección del filtro base para el WFB ha recibido menos atención que la estructura de árbol y existe una falta de consenso acerca del mismo.

Se observó de la sección 7.4.1.1 que un mayor número de momentos de desvanecimiento proporcionaba una mejor capacidad de compresión. Este aumento en el desempeño se podría justificar del hecho de que filtros más grandes generalmente proporcionan una mejor selectividad en frecuencia y de ahí que proporcionen una mejor ganancia en la codificación. De igual manera Sinha y Tewfik encontraron que filtros con un número máximo de momentos de desvanecimiento proveían de resultados “casi optimos”.

Sin embargo, otro estudio hecho por Kudumakis y Sandler mostró que cuatro distintas familias de filtros: Daubechies-A (fase mínima), Daubechies-B (simetría máxima), Johnston, y Smith-Barnwell, proporcionaban desempeños similares cuando se comparaba la SNR y se hacían pruebas de audio. En otro experimento realizado por Philipe, Moreau y Mainard se concluyó que la ganancia de codificación era el criterio más relevante para seleccionar el filtro, y más aún, la selectividad en frecuencia y la regularidad (respecto al número de momentos de desvanecimiento) era de menos importancia.

En muchos casos la elección del filtro base se da a partir de la capacidad del filtro de aproximar a la señal, sección 7.2, y de ahí que se escojan filtros como el biortogonal o el symlet. Sin embargo, existen codificadores wavelet que utilizan la wavelet Daubechies aunque se ha mostrado que su naturaleza es más adecuada para el análisis de imágenes.

8.3.3 Conclusión

Se puede notar que mientras existe un consenso acerca de la estructura de árbol, no existe un resultado definitivo acerca del filtro base a usar.

Sin embargo, se puede decir que mientras la estructura de árbol determina la resolución tiempo-frecuencia general, el filtro base determina las propiedades de localización de las distintas bandas wavelet. Esta propiedad de localización es muy importante en la codificación de audio y será expuesta más adelante en este capítulo, en particular para la elección del filtro base.

8.4 Representación de señales de audio usando WFB

Cuando se utiliza un WFB para representar una señal de audio el objetivo del WFB es el de presentar una representación de la señal de entrada de acuerdo a las características de la misma. En el caso de la codificación perceptual sucede que la representación debe de ajustarse a la resolución del umbral de enmascaramiento. En el caso ideal, esta resolución sería determinada por el modelo psicoacústico y variaría por cada bloque de audio, pero el usar una resolución fija de tiempo-frecuencia significa que:

1. Una descomposición de bandas críticas es lo suficientemente eficiente.
2. Si los artefactos temporales son muy importantes, deben de ser identificados por otros métodos ya que el modelo psicoacústico es ineficiente para estos casos.

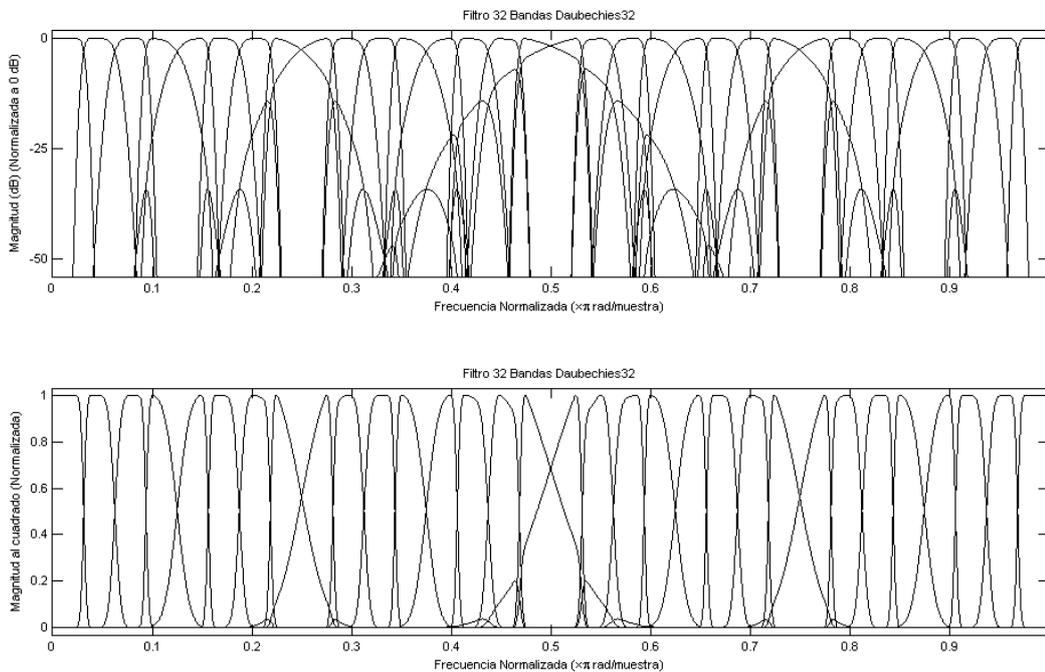
A pesar de que ambos requerimientos son muy importantes, el objetivo de este capítulo es el de describir los requerimientos del WFB como una posible solución para el banco de filtros, de ahí que se dará más interés a la descomposición de bandas críticas.

8.4.1 Representación de las sub bandas en el dominio de la frecuencia

La separación de las bandas en un banco de filtros, en particular una descomposición de árbol, depende en gran medida del filtro usado como base. Utilizar un filtro que posea una banda de transición muy grande, por ejemplo Daubechies, afectará al descender en los distintos niveles de descomposición llegando al caso en que exista un traslape de bandas y se produzcan errores por aliasing.

La figura 99 muestra las sub bandas resultantes de la aplicación de un banco de filtros de 32 bandas a partir de una estructura de árbol simétrico, la grafica superior se encuentra en escala logarítmica y la inferior en escala lineal. La razón del por qué ocurre esto está bastante ligado en la forma en que los bancos de filtros trabajan y cómo hacen para obtener cada sub banda.

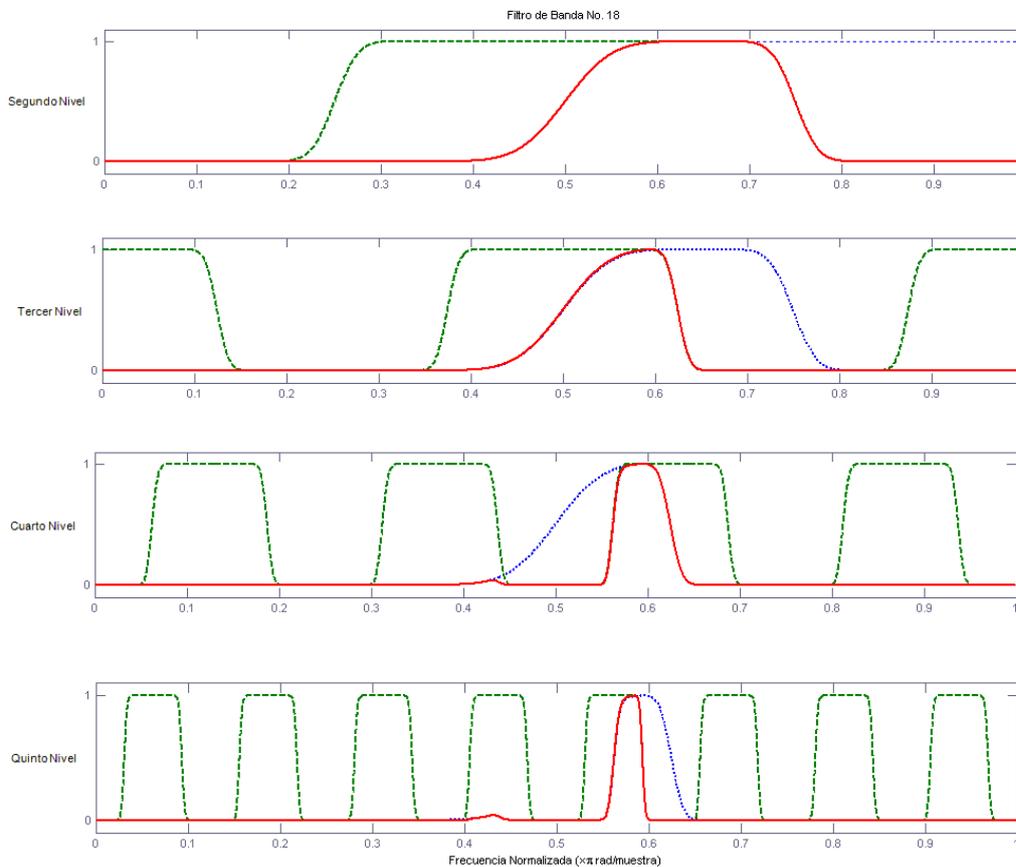
Figura 99. Filtro de 32 bandas Daubechies.



Si se parte de los dos filtros base, paso bajo y paso alto, y se observa que en cada descomposición estos son trasladados y reducidos en frecuencia se puede pensar entonces que los demás filtros son de alguna manera una versión trasladada y combinada de estos filtros en los distintos niveles de descomposición; es decir, la convolución de varios filtros. En la figura 100 se muestra la manera en que se obtuvo el filtro para la sub banda no.18.

En este caso, se obtuvo al realizar la convolución de los filtros $H_1(z)$, paso alto del primer nivel, y $H_1(z^2)$, paso alto del segundo nivel. Posteriormente este nuevo filtro fue convolucionado con $H_0(z^4)$, paso bajo del tercer nivel. Este proceso se repitió entonces con $H_1(z^8)$ y finalmente con $H_1(z^{16})$ como se muestra en la figura 100.

Figura 100. Filtro para la sub banda no.18.



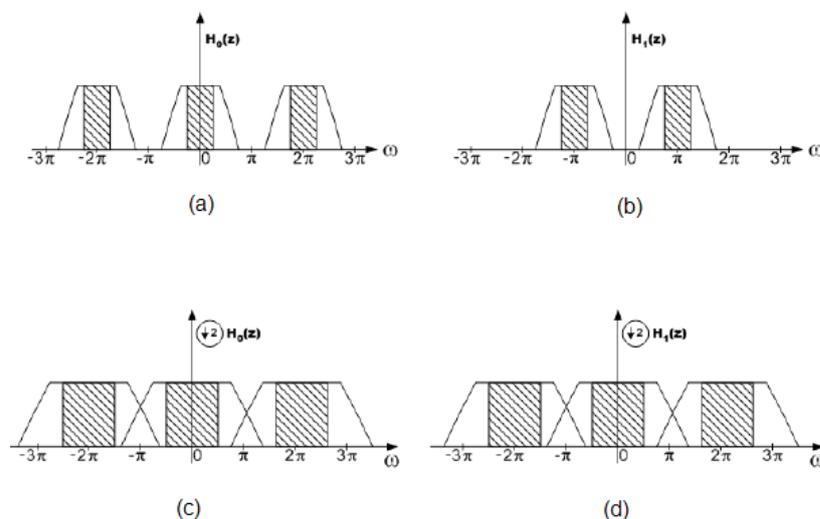
A partir de esto se puede entender que por la forma en que los filtros se convolucionan se da el traslape de bandas y el error por aliasing, esta es la mayor desventaja al usar un banco de filtros iterativo como el WFB.

Esto desde el punto de vista de la codificación perceptual de audio no es adecuado ya que no hay una adecuada separación de los canales ni una buena resolución en frecuencia, por lo que debe corregido.

8.4.2 Orden de las sub bandas

Cuando se obtiene una descomposición por medio de una estructura de árbol el orden que este presenta en sus ramas inferiores no es el de una secuencia lógica; sino más bien, posee cierto orden que depende del filtro usado. Esto se puede explicar de la siguiente manera, suponiendo que se tiene una señal $x[n]$ y es pasada por un banco de filtros en cuadratura, QMF, el cual posee dos filtros $H_0(z)$ y $H_1(z)$. Se obtienen dos distintas señales $x_0[n]$ y $x_1[n]$ cuyo espectro de frecuencia se muestra en la figura 101(a) y 101(b) respectivamente.

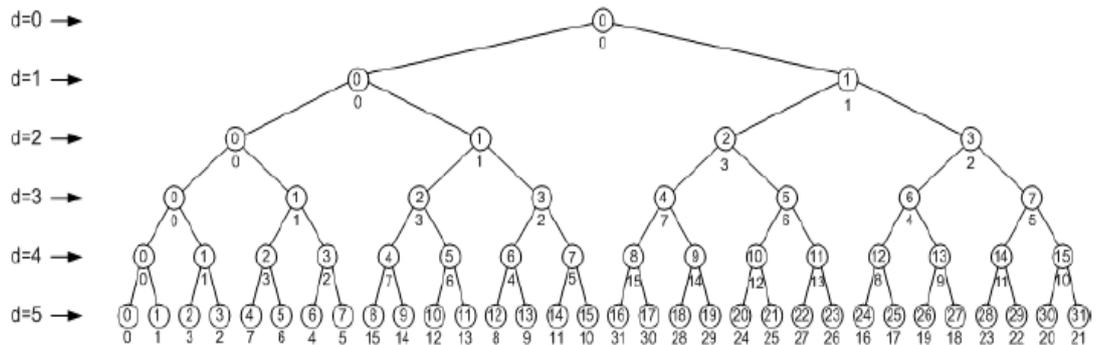
Figura 101. Efectos de filtrado y sub muestreo.



El filtrado seguido del sub muestreo puede ser interpretado en el dominio de la frecuencia como una multiplicación seguida por una expansión. La expansión es hecha respecto a las frecuencias centrales $2\pi m$ para $m \in \mathbb{Z}$. En la figura 101(a) la región sombreada representa las bajas frecuencias luego de aplicar el filtro $H_0(z)$ y en la figura 101(b) la región sombreada representa las altas frecuencias obtenidas al aplicar el filtro $H_1(z)$. Luego, en la figura 101(c) se muestra el resultado de realizar el sub muestreo a la señal $x_0[n]$. En este caso la región de “baja frecuencia” antes de sub muestrear se mantiene como de “baja frecuencia” luego de sub muestrear; y la región de “alta frecuencia” antes de sub muestrear sigue siendo de “alta frecuencia” luego de sub muestrear. En la figura 101(d) se observa el resultado de sub muestrear $x_1[n]$. Para esta señal, las componentes de “baja frecuencia” antes de sub muestrear se convierten en componentes de “alta frecuencia”; y las componentes de “alta frecuencia” antes de sub muestrear se convierten en componentes de “baja frecuencia” luego de sub muestrear. Esto significa que un filtro paso alto seguido de un sub muestreo invertirá el orden de frecuencias del espectro de la señal resultante.

Para una descomposición con estructura de árbol, una señal puede ser iterada muchas veces antes de alcanzar la rama final, y dependiendo del camino que tome, puede haber cualquier número de etapas paso bajo y paso alto. Ocurrendo una inversión en el espectro de frecuencia en las etapas paso alto. Como se mencionaba anteriormente, las ramas de una estructura de árbol no necesariamente siguen una secuencia lógica de izquierda a derecha, ni tampoco ocurre que la rama paso bajo proporciona las bajas frecuencias y la rama paso alto proporciona las frecuencias paso alto. En la figura 102 se muestra una estructura de árbol de 32 bandas, donde encerrado en un círculo se muestra el número de nodo o la secuencia natural. El otro número que aparece es el de la secuencia relacionada si cada banda estuviera seguida una tras de otra.

Figura 102. Secuencia natural y secuencia relacionada a cada banda en un WFB de 32 bandas.



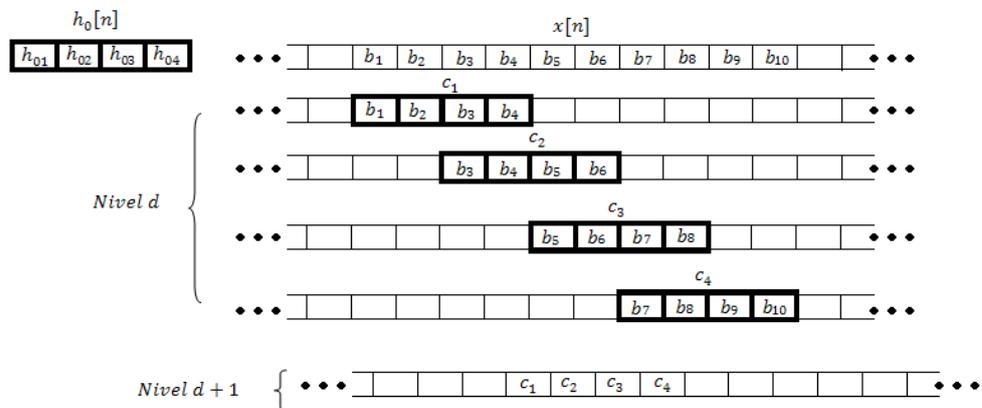
Esto es muy importante cuando se diseñan filtros para la codificación de audio, ya que se esperaría que las bandas reflejen un orden lógico en el dominio de la frecuencia. Lo anteriormente expuesto no ha sido muy estudiado ya que usualmente se suele trabajar solamente en la forma asimétrica del WFB.

8.4.3 Localización en el dominio del tiempo

El hecho de que los filtros se obtengan por un proceso de convolución también juega un papel importante en la localización en el dominio del tiempo. De hecho, la aplicación de filtros grandes resulta en un tiempo para realizar la convolución mucho más elevado que cuando se aplican filtros pequeños. En la figura 103 se muestra la aplicación de un filtro de longitud 4 (cuadros de línea gruesa), sobre una señal $x[n]$ (recuadros lado derecho). Suponiendo en este caso que se tiene una señal de entrada $x[n]$ y en la descomposición de nivel 0 se obtuvieron los coeficientes b_k . Al pasar al siguiente nivel el filtro $h_0[n]$ realiza una descomposición por medio de la convolución de los coeficientes b_k y los coeficientes del filtro h_{0k} .

El resultado son cuatro coeficientes para el primer nivel de descomposición. Resulta interesante notar que se necesitan de 10 coeficientes del nivel 0 para obtener los 4 coeficientes del primer nivel. Un solo coeficiente del segundo nivel necesitaría de los 4 coeficientes del primer nivel, o 10 del nivel 0 para poder ser obtenido.

Figura 103. Coeficientes necesarios al convolucionar con un filtro.



A partir de lo expuesto anteriormente se puede determinar el número de coeficientes necesarios en el nivel de descomposición 0 a partir de la siguiente expresión:

$$T_{soporte} = 2^d + (2^d - 1)(L - 2) \quad (8-1)$$

Donde d es el nivel de descomposición y L la longitud del filtro.

Al pasar de un nivel a otro existe un sub muestreo de 2 muestras reduciendo la resolución en el tiempo por un factor de 2; con lo que, idealmente se esperaría que este tiempo fuese de $T_{ideal} = 2^d$. Esto solamente sucede para $L = 2$.

Tabla XIII. Soporte en el tiempo para distintos filtros.

Nivel	Ideal	Longitud del Filtro					
		2	4	8	16	32	64
1	2	2	4	8	16	32	64
2	4	4	10	22	46	94	190
3	8	8	22	50	106	218	442
4	16	16	46	106	226	466	946
5	32	32	94	218	466	962	1954
6	64	64	190	442	946	1954	3970
7	128	128	382	890	1906	3938	8002
8	256	256	766	1786	3826	7906	16066

La tabla XIII muestra el cálculo del tiempo de soporte (muestras necesarias) para distintos filtros de longitud L en el nivel d . Se puede ver como las wavelets poseen un comportamiento no ideal en el dominio del tiempo y como también existe un traslape entre coeficientes wavelet adyacentes. Aunque la cantidad de traslape depende en el nivel y la longitud del filtro, también juegan un papel importante las características en el dominio del tiempo del filtro paso bajo y paso alto. A pesar de esto, los valores de la tabla XIII son una buena opción para entender la relación entre la localización en el tiempo y la localización en frecuencia. Se puede notar que una localización en el tiempo ideal requiere de un filtro de longitud $L = 2$, pero a cambio de esto se tendría una pobre localización en la frecuencia por la longitud del mismo.

8.5 Optimización del WFB

El problema del traslape de bandas durante la descomposición por medio del WFB ha sido analizado por muy pocos autores. A continuación se da una breve descripción de dos métodos para disminuir el traslape entre bandas, siendo este tema interesante para futuras investigaciones.

8.5.1 Minimización del traslape entre bandas por umbral

Una manera de disminuir el error de aliasing producido por bandas situadas fuera de lugar es calculando la cantidad de distorsión introducida en todas las bandas cuando una banda es cuantizada y asegurarse que la suma de todas las distorsiones se mantenga por debajo del umbral de enmascaramiento para todas las bandas. Se puede formular a como

$$\sum_{i=0}^{M-1} \sigma_i^2 |G_i(\omega)|^2 \leq T(\omega) \quad (8-2)$$

Para todas las frecuencias ω , donde σ_i^2 es el ruido de cuantización, $G_i(\omega)$ la respuesta del canal en la banda i y $T(\omega)$ la energía del umbral de enmascaramiento.

Este criterio esencialmente toma en cuenta la cantidad de ruido de codificación que es introducido en una sola banda por todas las bandas (incluyendo la banda en cuestión) y trata de minimizar la asignación de bits para cada banda mientras se mantiene la distorsión total por debajo del umbral de enmascaramiento. El valor de σ_i^2 comúnmente se realiza por un proceso de optimización debido a su complejidad.

Aunque este método toma de forma explícita el error asociado con todas las componentes fuera de banda, la complejidad asociada en el diseño de este método y el alto requerimiento para representar ciertas bandas implica serias desventajas en su realización. Se esperaría que el WFB se ajustara a la codificación de audio y no que la codificación de audio se ajustara al WFB.

8.5.2 Algoritmo modificado de Remez para filtros ortogonales en cuadratura

El algoritmo modificado de Remez propuesto por Rioul y Duhamel provee de una solución ortogonal en cuadratura a partir de algunas limitaciones adicionales en la longitud del filtro (L), la banda de transición (B), y el número de momentos de desvanecimiento (K). El algoritmo también maximiza la atenuación en la banda de rechazo luego de satisfacer los requerimientos dados. El algoritmo modificado de Remez se basa en el diseño de filtros de media banda seguidos de una factorización, donde el filtro paso bajo factorizado es de la forma

$$H_0(z) = (1 + z^{-1})^K Q(z) \quad (8-3)$$

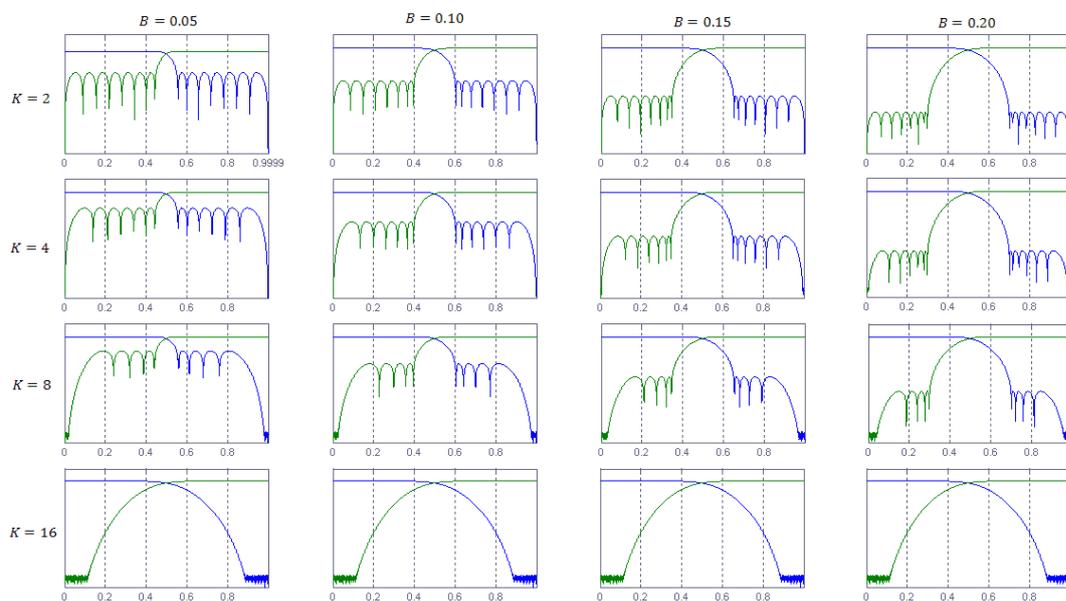
Donde $Q(z)$ es un polinomio que no contiene polos o ceros en $z = -1$.

Para soluciones que involucren wavelets, el filtro $H_0(z)$ también debe de cumplir con las condiciones de admisibilidad (sección 5.3) y ortogonalidad (sección 6.2.1), siendo estas

$$\sum_n h_0[n] = \sqrt{2} \quad \text{y} \quad \sum_n h_0[n]h_0[n + 2k] = \delta[k] \quad (8-4)$$

Si $H_0(z)$ es de longitud L , la condición de ortogonalidad representa $L/2$ ecuaciones (o limitantes), lo que deja un máximo de $L/2$ grados de libertad para K . La condición de admisibilidad también implica la existencia de $1 \leq K \leq L/2$ momentos que son cero.

Figura 104. Filtros obtenidos a partir del algoritmo modificado de Remez.



El algoritmo modificado de Remez primero impone la regularidad K en el filtro $H_0(z)$ y luego utiliza los $L/2 - K$ grados de libertad para satisfacer la limitación de la banda de transición B para posteriormente maximizar la atenuación en la banda de rechazo. Como resultado de esto la regularidad, la banda de transición, y la atenuación en la banda de rechazo son los tres requerimientos que compiten a la hora de emplear el algoritmo. Para una regularidad máxima de $K = L/2$ el algoritmo proporciona la solución Daubechies. La figura 104 muestra la aplicación del algoritmo modificado de Remez⁸ para distintos valores de B y K en una escala logarítmica.

⁸ El apéndice 3 presenta el algoritmo de Remez modificado escrito por Oliver Rioul usado para obtener estos filtros.

Se puede observar que el filtro con la mayor banda de atenuación ($B = 0.20$) y la menor regularidad ($K = 2$) proporciona la mejor atenuación en la banda de rechazo, y el filtro con la menor banda de atenuación ($B = 0.05$) y la mayor regularidad ($K = 16$) proporciona la peor atenuación en la banda de rechazo. Para otros valores conforme se aumenta el valor de B y se disminuye el valor de K se mejora la atenuación en la banda de rechazo. Además, el número de oscilaciones en la banda de rechazo se encuentran relacionadas con los grados de libertad restantes luego de que se impone la regularidad K , siendo de $L/2 - K$, como consecuencia del teorema de alternación usado en el algoritmo.

8.5.3 Filtro de 32 bandas modificado

En la sección 8.4.1 se mostró como el uso de un banco de filtros basado en la wavelet Daubechies proporcionaba un conjunto de sub bandas que se traslapaban unas con otras. Sin embargo, se ha mostrado que se pueden crear wavelets que cumplan con los requerimientos de regularidad, banda de transición y longitud del filtro a partir de un criterio de optimización.

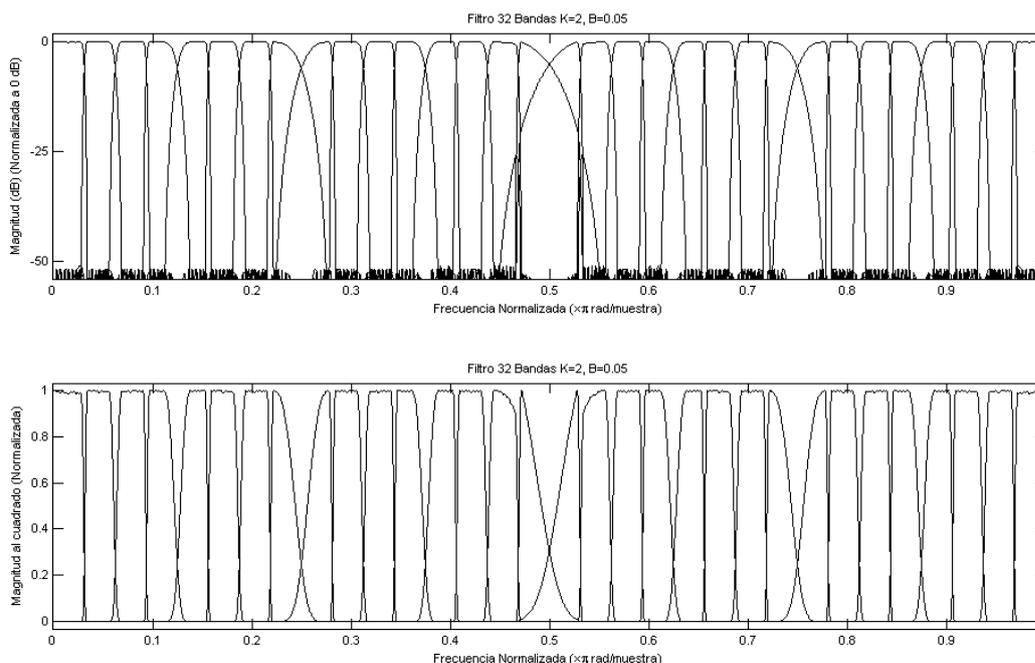
Se ha mostrado además que el uso de filtros con cortes agudos en la banda de transición mejora la eliminación de los lóbulos laterales y el traslape entre bandas adyacentes puede ser eliminado hasta cierto nivel. Sin embargo, la elección de la banda de transición afecta de forma directa los niveles de atenuación en la banda de rechazo, por lo que ambos requerimientos no pueden ser alcanzados a la vez.

En la figura 99 se observa como los lóbulos laterales son una de las mayores causas del error por aliasing razón por la cual se ha decidido por eliminar las mismas con la restricción de una regularidad mínima ($K = 2$).

El valor de B es distinto para cada filtro de sub banda, y podría ser tema de mucha más investigación, sin embargo se limitará a un valor constante ($B = 0.05$) para todas la bandas con la finalidad de eliminar los lóbulos laterales sacrificando la atenuación en la banda de rechazo.

La figura 105 muestra los resultados de la implementación de este nuevo filtro de longitud 32, claramente se observa una mejora frente a los resultados de la figura 99.

Figura 105. Filtro de 32 bandas con $L = 32$, $K = 2$ y $B = 0.05$.



El Moving Pictures Experts Group ha establecido en el códec MPEG-1 que una codificación de alta calidad necesita de 96 dB de atenuación en la banda de rechazo; sin embargo, el resultado obtenido para este filtro fue de 21 dB. Una manera de satisfacer la necesidad de una banda de transición específica y una banda de rechazo con 96 dB es el de usar filtros con longitudes más grandes. Por ejemplo, el uso de un filtro de longitud $L=64$ proporciona una banda de rechazo de 106 dB.

Como muestra la ecuación (8-1) el uso de filtros más grandes implica un tiempo de soporte más grande, para el caso del filtro de longitud 32 se necesitarían de 7906 muestras para un coeficiente a una descomposición de 8 niveles o de 179.3 ms a un muestreo de 44100 kHz. En el caso del filtro de longitud 64 se necesitarían 16066 o 364.3 ms. Estos resultados muestran que la localización en el tiempo es pobre para estos filtros e incluso ineficiente cuando existe un pre enmascaramiento en el tiempo (aproximadamente de 50 ms).

A partir de los resultados obtenidos se puede observar como el diseño de un WFB para una resolución de bandas críticas implica sacrificar ya sea la resolución temporal o la resolución en frecuencia. La resolución en frecuencia es sacrificada con una separación de los canales de sub banda, permitir lóbulos laterales o disminuir la atenuación en la banda de rechazo. Por su parte, la resolución en el tiempo es sacrificada con filtros más grandes que necesitan más tiempo en su descomposición.

Finalmente, se puede notar que cada requerimiento no puede ser maximizado en su totalidad, de acuerdo a las necesidades de la aplicación, el diseñador debe de escoger que es más importante para su buen funcionamiento.

CONCLUSIONES

1. La calidad de una descomposición *wavelet* depende especialmente de la habilidad de aproximar la señal con las *wavelets*; y en particular, para una señal de audio se debe de tomar en cuenta el carácter estacionario y no estacionario que presentan las mismas.
2. Las señales de audio llevan al requerimiento de una *wavelet* con una alta regularidad, es decir un buen soporte en el dominio de la frecuencia.
3. *Wavelets* con soporte compacto pueden localizar el tiempo de un evento mejor que otras, ya que poseen una mejor resolución en el tiempo.
4. Los requerimientos de una alta regularidad y un soporte compacto para una *wavelet* son imposibles de alcanzar al mismo tiempo, por lo que se debe de realizar un compromiso entre la resolución en frecuencia y la resolución en el tiempo a fin de escoger el requerimiento más importante.

5. La utilización de bancos de filtros wavelet permite la descomposición de una señal en una serie de bandas críticas, sin embargo se necesita de un proceso de optimización para alcanzar una separación tolerable de estas bandas. Este proceso de optimización se basa en la longitud, regularidad y separación entre bandas del filtro wavelet a usar.
6. El procesado en tiempo real lleva al requerimiento de un algoritmo lo suficientemente rápido por lo que un soporte compacto es preferible en este caso.
7. La reducción de ruido y la compresión de señales son dos procesos inherentes de la transformada wavelet y su eficacia radica en la simple aplicación de un umbral a los coeficientes de análisis y/o detalle obtenidos.
8. La diferencia entre la compresión de señales y la eliminación de ruido radica en la forma en que se escoge el nivel de umbral. En el caso de la compresión, el umbral viene dado por limitantes externos como: el ancho de banda, memoria disponible, razones de compresión. Por el contrario, en el caso de la reducción de ruido el umbral se selecciona de forma automática al estimar el nivel de ruido usando la información dada por el modelo para reducirlo.

RECOMENDACIONES

1. Se sugiere impartir en la carrera de ingeniería electrónica el curso de análisis multiresolución y transformada wavelet enfocado en el tratamiento digital de señales.
2. Es aconsejable crear en la carrera de ingeniería electrónica el área de procesamiento digital de señales donde se podría incluir el curso de comunicaciones 4 y el ya mencionado curso de multiresolución y transformada wavelet.
3. Fomentar en los estudiantes de las carreras de ingeniería eléctrica y electrónica el uso de paquetes computacionales como MATLAB, así como también el desarrollo de laboratorios en este tipo de paquetes por los catedráticos de la facultad.
4. Enfocar el curso de electrónica aplicada 1 al procesamiento de señales de audio, mientras que el curso de electrónica aplicada 2 podría estar enfocado al procesamiento imágenes.

REFERENCIAS BIBLIOGRÁFICAS

1. Shpak, Dale J. **A generalized Remez method for the design of FIR digital filters.** IEEE Transactions on circuits and systems, Vol. 37, no. 2, febrero 1990.
2. Tzanetakis, George y otros. **Audio Analysis using the Discrete Wavelet Transform.** Princeton University.
3. Zanartu, Matias. **Audio compression using wavelet techniques.** Electrical and Computer Engineering. Purdue University. Estados Unidos.
4. Bosi, Marina y Davidson, Grant. **High quality, low-rate audio transform coding for transmission and multimedia applications.** Dolby Laboratories, Inc. Estados Unidos: 1993.
5. Rams, Martínez y otros. **Implantes cocleares: desarrollo y perspectivas.** Revista mexicana de ingeniería biomédica, Vol. XXVII, no. 1 Junio 2006 pp. 45 - 54.
6. Qiao, Feng y Milam, Rachael. **Moments and vanishing wavelet moments.**

7. Wendt, Christopher y Petropulu, Athina P. **Pitch determination and speech segmentation using the discrete wavelet transform.** Electrical and Computer Engineering Department Drexel University.

8. Vidakovic, Brani y Mueller, Peter. **Wavelets for kids. A tutorial introduction.** Duke University. Estados Unidos. 28 pp.

BIBLIOGRAFÍA

1. Akansu, Ali N. y Haddad, Richard A. **Multiresolution signal decomposition. Transforms, subbands, wavelets.** 2a ed. Estados Unidos: Academic Press, 2001. 499 pp.
2. Bahoura, Mohammed y Rouat, Jean. **Wavelet noise reduction: application to speech enhancement.** Université du Québec à Chicoutimi. Canadá.
3. Bömers, Florian. **Wavelets in real time digital audio processing: analysis and sample implementations.** University of Mannheim, Department of Computer Science IV. Alemania: 2000. 119 pp.
4. Faundez, Pablo y Fuentes, Álvaro. **Procesamiento digital de señales acústicas utilizando wavelets.** Universidad Austral de Chile. 143 pp.
5. Lee, Peter. **Wavelet filter banks in perceptual audio coding.** University of Waterloo. Canadá: 2003. 134 pp.

6. Mallat, Stéphane. **A wavelet tour of signal processing**. 2a ed. Estados Unidos: Academic Press, 1998. 637 pp.
7. Mertins, Alfred. **Signal analysis. Wavelets, filter Banks, time-frequency transforms and applications**. Estados Unidos: John Wiley & Sons, 1999. 317 pp.
8. Misiti, Michel y otros. **Wavelets and their applications**. Estados Unidos: ISTE Ltd, 2007. 330 pp.
9. Patel, Ankit y otros. **Lossless sound compression using the discrete wavelet transform**. Universidad de Harvard. Estados Unidos: 2002.
10. Rao, Raghuveer M. y Ajit S. Bopardikar. **Wavelet transforms introduction to the theory and applications**. Estados Unidos: Addison-Wesley, 1998. 310 pp.
11. Rioul, P. y Duhamel, P. **A Remez exchange algorithm for orthonormal wavelets**. IEEE Transactions on circuits and systems-II: analog and digital signal processing. Vol. 41, no. 8. agosto 1994.
12. Smith, W. Steven. **The scientist and engineer's guide to digital signal processing**. 2a ed. Estados Unidos: California Technical Publishing, 1999. 650 pp.

13. Stark, Hans-Georg Stark. **Wavelets and signal processing**. Holanda: Springer, 2005. 150 pp.
14. Stearns, Samuel D. **Digital signal processing with examples in MATLAB**. Estados Unidos: CRC Press, 2002. 334 pp.
15. Suter, Bruce W. **Multirate and wavelet signal processing**. Vol. 8. Estados Unidos: Academic Press, 1998. 199 pp.
16. Vaidyanathan, P. P. **Multirate systems and filter banks**. Estados Unidos: Prentice Hall, 1993. 910 pp.
17. Walker, James S. **A primer on wavelets and their scientific applications**. 2a ed. Estados Unidos: Chapman & Hall/CRC, 2008. 294 pp.
18. Weeks, Michael. **Digital signal processing using MATLAB and wavelets**. Estados Unidos: Infinity Science Press LLC, 2007. 452 pp.

ANEXO 1

```
% Determinación de varianza para ruido no estacionario
% Este programa estima los puntos donde la varianza del ruido que
contiene
% una señal son distintos.
% Se basa en el algoritmo propuesto por Marc Lavielle para la
determinar
% estos puntos

%Ingresar la señal a analizar
%Default (Comentar las dos líneas si se desea usar otra señal)
load nblocr1;
x = nblocr1;

%Quitar comentarios para utilizar con otras señales

%x = input ('¿Señal a analizar?');
% analisis_varianzas = x;

%Paso 1. Obtener el ruido al suprimir la aproximación.
% Realizar una descomposición de primer nivel usando
% la wavelet sym4.
wname = 'sym4';
lev = 1;
[c,l] = wavedec(x,lev,wname);
% Reconstrucción del detalle en el nivel 1.
det = wrcoef('d',c,l,wname,1);

%Paso 2. Remover casi toda la señal, reemplazando el 2% de los más
grandes
% valores por la media.
x = sort(abs(det));
% Ordena de menor a mayor los elementos

retencion = input...
('Ingrese el nivel de retención de la señal entre 0 y 1');
v2p100 = x(fix(length(x)*retencion));
% Selecciona el porcentaje de retención que tendrá la señal.
% Entre 0 y 1

ind = find(abs(det)>v2p100);
```

```

    % Encuentra la posición de los valores que son mayores que v2p100
    % (El 2% buscado)

    det(ind) = mean(det);
    % Convierte los valores de detalle que son mayores al 98% en el
valor
    % de la media

%Paso 3. Usando la función wvarchg se pueden estimar los puntos donde
    % existen cambios
%Los parametros son:
    % El retraso mínimo entre dos puntos que cambian es d = 10.
    % El máximo número de puntos donde pueden haber variaciones 5.
    [cp_est,kopt,t_est] = wvarchg(det,5)

%Rutina para descomponer señal para analizar varianzas

x = analisis_varianzas;
[c,l] = wavedec(x,lev,wname);
% Reconstruct detail at level 1.
det = wrcoef('d',c,l,wname,1);

t0 = 1;

% Descomposición de la señal en distintos intervalos para encontrar su
% varianza en estos intervalos.

for j = 1:length(cp_est)
    nivel = num2str(j);
    signal = strcat('s',nivel);
    sn = genvarname(signal,'who');

        switch j
            case 1
                for i = t0:cp_est(j);
                    s(i) = det(i);
                end

            otherwise
                for i = cp_est(j-1)+1:cp_est(j)
                    s(i-cp_est(j-1)) = det(i);
                end
        end

        eval([sn ' = s;']);
    t0 = 1+cp_est(j);
end

```

```

for i = cp_est(length(cp_est))+1:length(det);
    nivel = num2str(length(cp_est)+1);
    signal = strcat('s',nivel);
    sn = genvarname(signal, 'who');
    s(i-cp_est(length(cp_est))) = det(i);
end
eval(['sn ' = s;']);

% El valor de sigma representa la varianza obtenida sobre cada
intervalo
% de interés.
sigma1 = (mad(s1,1))/0.6745
sigma2 = (mad(s2,1))/0.6745
sigma3 = (mad(s3,1))/0.6745
sigma4 = (mad(s4,1))/0.6745

N21 = cp_est(2)-cp_est(1);
N32 = cp_est(3)-cp_est(2);
N43 = cp_est(4)-cp_est(3);
N54 = length(x)-cp_est(4);

% El valor umbral se encuentra en base al propuesto por Donoho y
Johnstone
% pero ahora aplicado cada uno sobre cada intervalo.
umbral1 = sigma1*(sqrt(2*log(N21)))
umbral2 = sigma2*(sqrt(2*log(N32)))
umbral3 = sigma3*(sqrt(2*log(N43)))
umbral4 = sigma3*(sqrt(2*log(N54)))

```


ANEXO 2

A continuación se presenta el diagrama de bloques de la simulación realizada en Simulink para la reducción de ruido de una señal. La simulación consta de dos casos: (1) reducción de ruido con múltiples umbrales; y (2) reducción de ruido con un único umbral.

La figura 106 muestra el diagrama de bloques de ambos casos, el cuál es el mismo. Sin embargo, la diferencia radica en el bloque denominado **Umbrales**.

Los bloques que componen la simulación son los siguientes:

Signal: Aquí se especifica la señal así como también la frecuencia de muestreo (44100Hz) y el número de muestras (512).

AWG: En este bloque se da la inserción de ruido blanco Gaussiano especificando la SNR. Por defecto se tiene en 20 dB.

Dyadic Analysis Filter Bank: Este bloque se encarga de especificar el banco de filtros de análisis a usar, se pueden usar wavelets ya definidas o establecer los coeficientes del filtro a usar. Por defecto se tiene una descomposición con la wavelet sym4 de 5to nivel.

Dyadic Synthesis Filter: Funciona de igual manera que el filtro de análisis solamente que estos son los parámetros para la reconstrucción.

Noised_signal: Almacena en una variable de este mismo nombre la señal con ruido antes de ser procesada.

Denoised_signal: Almacena en una variable de este mismo nombre la señal sin ruido luego de ser procesada.

To Audio Device: Establece los parámetros para la adecuada reproducción en la PC.

Umbrales: Consiste de la aplicación del umbral suave. Este puede ser múltiple y se basa en la determinación del nivel de ruido de cada nivel de detalle, o puede ser único en base al primer nivel de detalle. La figura 107 muestra el caso para los múltiples niveles mientras que la figura 111 el caso de un único umbral.

Figura 106. Diagrama de simulación de reducción de ruido por medio de Simulink.

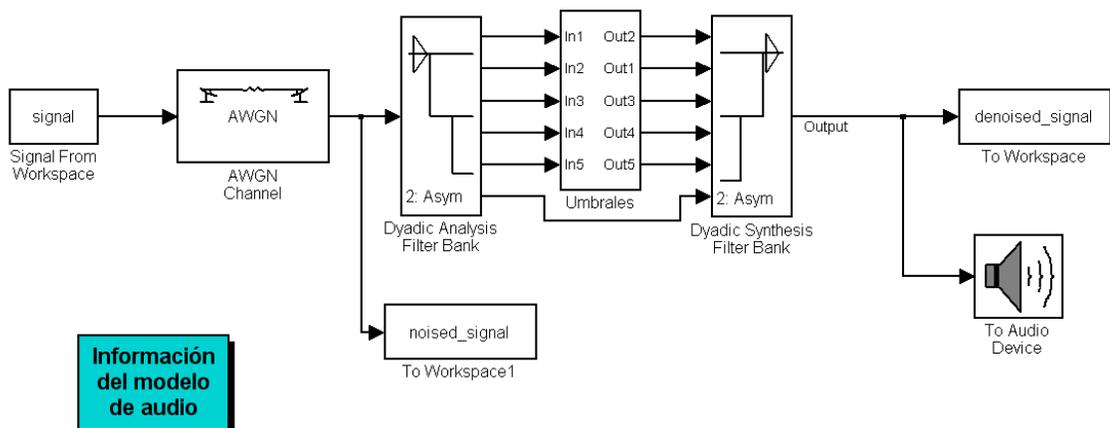


Figura 107. Diagrama de interno del bloque umbral para el caso de múltiples umbrales.

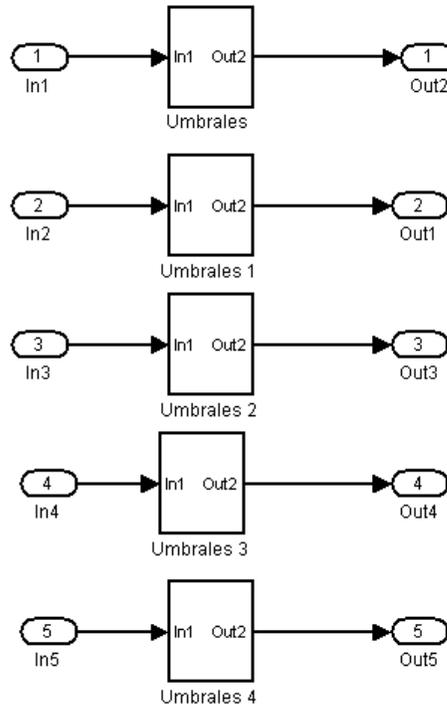


Figura 108. Diagrama de interno del bloque umbrales.

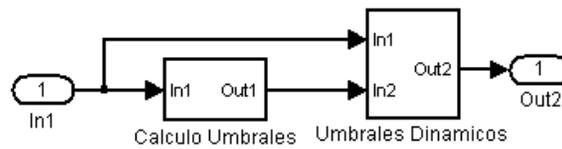


Figura 109. Diagrama de interno del bloque calculo umbrales.

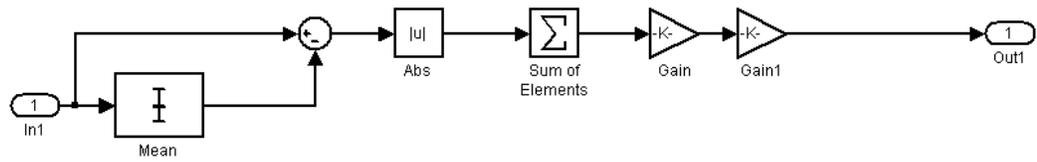


Figura 110. Diagrama de interno del bloque umbrales dinámicos.

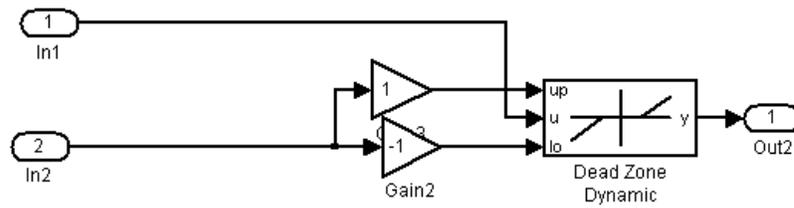
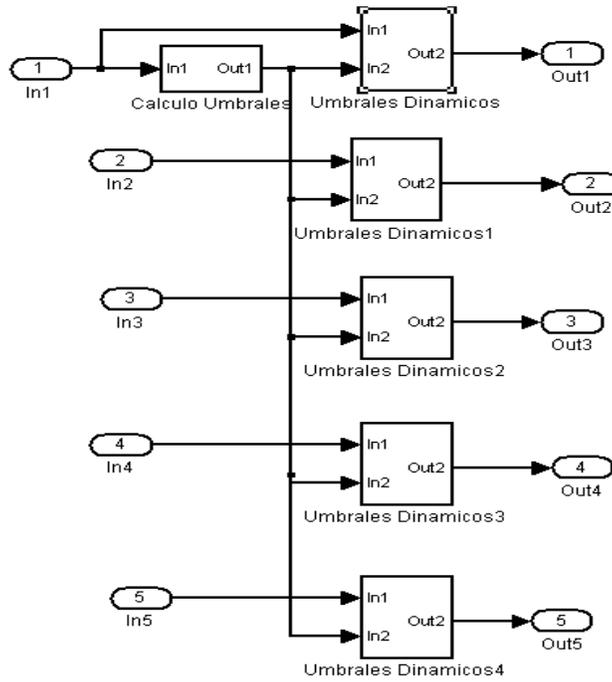


Figura 111. Diagrama de interno del bloque umbral para el caso de un único umbral.



ANEXO 3

```
function [p,r,delta]=remezwav(L,K,B)
%REMEZWAV Obtiene la wavelet optima usando el algoritmo de intercambio
de
% REMEZWAV(L,K,B) proporciona la respuesta al impulso P(z) de máxima
% selectividad en frecuencia. El filtro H(z) de longitud L satisface
los K
% grados de regularidad.
%El ancho de banda normalizado es dado por B. Este argumento es
opcional si
%K==L/2.

%[P,R]=REMEZWAV(L,K,B)también proporciona los ceros de P(z), con lo
cual
%puede ser usado para determinar H(z).

%[P,R,DELTA]=REMEZWAV(L,K,B) también proporciona la máxima desviación
si
%K<L/2.

%Crédito a su autor original Olivier Rioul

computeroots=(nargout>1);

%%%%%%%%%%%%%% Paso 1 %%%%%%%%%%%%%%%
if rem(L,2),error('L must be even');end
if rem(L/2-K,2)
    K=K+1;
end
if K>L/2
    error('K is too large')
end
N=L/2-K;

%%%%%%%%%%%%%% Paso 2 %%%%%%%%%%%%%%%
% Solución Daubechies PK(z)=z^(-(2K-1))+AK(Z^2)
if K==0
    AK=0;
else
    binom=pascal(2*K,1);
    AK=binom(2*K,1:K)./(2*K-1:-2:1);
    AK=[AK AK(K:-1:1)]; AK=AK/sum(AK);
end
%%%%%%%%%%%%%% Paso 2' %%%%%%%%%%%%%%%
```

```

% Factor Daubechies PK(z)=((1+z^(-1))/2)^(2K) QK(z)
if computeroots && K>0
    QK=binom(2*K,1:K); QK=QK.*abs(QK);
    QK=cumsum(QK); QK=QK./abs(binom(2*K-1,1:K));
    QK=[QK QK(K-1:-1:1)]; QK=QK/sum(QK)*2;
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Paso 3 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Solución Daubechies PK(z)
if K==L/2
    p=zeros(1,2*L-1);
    p(1:2:2*L-1)=AK; p(L)=1;
    if computeroots
        r=[roots(QK);-ones(L,1)];
    end
    return
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Paso 4 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Polinomio Daubechies PK(x)=1+x*DK(x^2)
if K==0
    DK=0;
else
    binom=pascal(K,1);
    binom=binom(K,:);
    DK=binom./(1:2:2*K-1);
    DK=fliplr(DK)/sum(DK);
end

wp=(1/2-B)*pi;          % Frecuencia de corte
gridens=16*(N+1);      % Densidad
found=0;                % Prueba booleana para ciclo Remez

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Paso I %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Estimación inicial para yk
a=min(4,K)/10;
yk=linspace(0,1-a,N+1);
yk=(yk.^2).*(3+a-(2+a)*yk);
yk=1-(1-yk)*(1-cos(wp)^2);
ykold=yk;
iter=0;
while 1 % Ciclo Remez
iter=iter+1;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Paso II %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Calculo de delta
Wyk=sqrt(yk).*((1-yk).^K);
Dyk=(1-sqrt(yk).*polyval(DK,yk))./Wyk;
for k=1:N+1
    dy=yk-yk(k); dy(k)=[];
    dy=dy(1:N/2).*dy(N:-1:N/2+1);

```

```

Lk(k)=prod(dy);
end
invW(1:2:N+1)=2 ./Wyk(1:2:N+1);
delta=sum(Dyk./Lk)/sum(invW./Lk);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Paso III %%%%%%%%%
% Calculo de R(y) en una malla más fina
Ryk=Dyk-delta.*invW; Ryk(N+1)=[];
Lk=(yk(1:N)-yk(N+1))./Lk(1:N);
y=linspace(cos(wp)^2,1-K*1e-7,gridens);
yy=ones(N,1)*y-yk(1:N)'.*ones(1,gridens); % yy contiene a y-yk en cada
linea
ind=find(yy==0); % Se evita la división por cero 0
if ~isempty(ind), yy(ind)=1e-30*ones(size(ind)); end
yy=1./yy;
Ry=((Ryk.*Lk)*yy)./(Lk*yy);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Paso IV %%%%%%%%%

% Encontrar el próximo yk
Ey=1-delta-sqrt(y).*(polyval(DK,y)+((1-y).^K).*Ry);
k=find(abs(diff(sign(diff(Ey))))==2)+1;
% N extrema
if length(k)>N % Puede ocurrir si L y K son grandes
k=k(1:N);
end
yk=[yk(1) y(k)]; % N+1 extrema incluyendo wp
if K==0, yk=[yk 1]; end % extrema en y==1 sumado
if all(yk==ykold), break; end % PARAR CICLO REMEZ
ykold=yk;

end % CICLO REMEZ
fprintf('%g iterations in remezwav.\n',iter);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Paso A %%%%%%%%%
% Calculo de la respuesta al impulso
w=(0:2*N-2)*pi/(2*N-1); y=cos(w).^2;
yy=ones(N,1)*y-yk(1:N)'.*ones(1,2*N-1);
ind=find(yy==0);
if ~isempty(ind), yy(ind)=1e-30*ones(size(ind)); end
yy=1./yy;
Ry=((Ryk.*Lk)*yy)./(Lk*yy);
Ry(2:2:2*N-2)=-Ry(2:2:2*N-2);
r=Ry*cos(w'*(2*(0:N-1)+1)); % IDFT (parcial real)
r=r/(2*N-1); r=[r r(N-1:-1:1)];
p1=[r 0]+[0 r];
pp=p1; % guardar p1 para su próximo uso
for k=1:2*K, p1=[p1 0]-[0 p1]; end
if rem(K,2), p1=-p1; end
p1=p1/2^(2*K+1);
p1(N+1:N+2*K)=p1(N+1:N+2*K)+AK; % sumar respuesta Daubechies:
p(1:2:2*L-1)=p1; p(L)=1;

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Paso A' %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Calculo de las raices
if computeroots
Q(1:2:2*length(pp)-1)=pp;
for k=1:2*K, Q=[Q 0]-[0 Q]; end
if rem(K,2), Q=-Q; end
Q=Q/2;
if K>0 % Agregar el factor QK Daubechies
Q(2*N+1:L-1)=Q(2*N+1:L-1)+QK;
else
Q(L)=1;
end
r=[roots(Q); -ones(2*K,1)];
end

```