

UNIVERSIDAD AUTONOMA DE NUEVO LEON  
FACULTAD DE INGENIERIA MECANICA  
Y ELECTRICA  
DIVISION DE ESTUDIOS DE POSTGRADO



CODIFICACION DE VOZ EN SUBBANDAS

POR

MARIO SANCHEZ MONROY

EN OPCION AL GRADO DE MAESTRO EN CIENCIAS  
DE LA INGENIERIA CON ESPECIALIDAD  
EN TELECOMUNICACIONES

CD. UNIVERSITARIA

ABRIL DEL 2001

TM  
TK7882  
S65  
S2  
2001  
c.1

COODIFICIOM DE  
VOZ EN SUBBANDAS

MS. M. 1



1080117244

UNIVERSIDAD AUTONOMA DE NUEVO LEON

FACULTAD DE INGENIERIA MECANICA  
Y ELECTRICA

DIVISION DE ESTUDIOS DE POSTGRADO



CODIFICACION DE VOZ EN GUEBANDAS

POR

MARIO SANCHEZ MONROY

EN OPCION AL GRADO DE MAESTRO EN CIENCIAS  
DE LA INGENIERIA CON ESPECIALIDAD  
EN TELECOMUNICACIONES

CD. UNIVERSITARIA

ABRIL DEL 2001





UNIVERSIDAD AUTONOMA DE NUEVO LEON

FACULTAD DE INGENIERIA MECANICA

Y ELECTRICA

DIVISION DE ESTUDIOS DE POSTGRADO



CERTIFICACION DE VOZ EN SUBBARRIOS

POR

MARIO SANCHEZ MONROY

EN OPCION AL GRADO DE MAESTRO EN CIENCIAS

DE LA INGENIERIA CON ESPECIALIDAD

EN TELECOMUNICACIONES

CD. UNIVERSITARIA

ABRIL DE 1971

TM

TK7882

.565

52

2001

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN  
FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA  
DIVISIÓN DE ESTUDIOS DE POSTGRADO

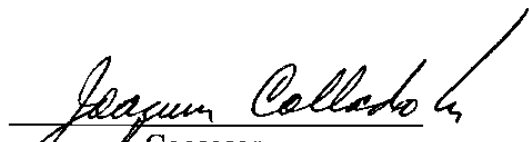
Los miembros del comité de tesis recomendamos que la tesis CODIFICACIÓN DE VOZ EN SUBBANDAS realizada por el Ing. Mario Sánchez Monroy, matrícula 775836, sea aceptada para su defensa como opción al grado de Maestro en Ciencias de la Ingeniería con especialidad en Telecomunicaciones.

El comité de Tesis:



Asesor

Dr. José Antonio de la O Serna



Coasesor

Dr. Joaquín Collado Moctezuma



Coasesor

Dr. Jesús de León Morales



Vo. Bo.

M.C. Roberto Villarreal Garza  
División de Estudios de Postgrado

San Nicolás de los Garza, N.L., Abril de 2001



*Dedicado a mis padres,*

*Carmen Monroy de Sánchez*

*y*

*Mario Sánchez Castro*

*A mis hermanas:*

*Matilde*

*Beatriz*

*Patricia*

*Carolina*

*María Esmeralda*

*Y a todos mis familiares y amigos.*

## **AGRADECIMIENTOS**

**A Dios, por todos sus beneficios.**

**Al Dr. José Antonio de la O Serna, por su paciencia, asesoría en la elaboración de la tesis y su ayuda desinteresada.**

**Al Dr. Joaquín Collado Moctezuma y al Dr. Jesús de León Morales, por sus comentarios en la revisión de la tesis; excelentes profesores durante mis estudios.**

**A los profesores del Programa Doctoral en Ingeniería Eléctrica: Dr. Salvador Acha Daza, Dr. Oscar Moreira Tamayo, Dr. Juan Manuel Ramírez Arredondo y al M.C. José Luis Ortega Soria.**

**Al Centro Universitario RODA, al Club Lince y al Lic. Andrés Javier Gómez Aguirre.**

**Al Ing. Adrián Rodríguez Lara, por su valiosa cooperación con bibliografía.**

**A todos mis compañeros de estudios durante la maestría.**

**A la Facultad de Ingeniería Mecánica y Eléctrica, a la Universidad Autónoma de Nuevo León, y al Programa de Apoyo a la Investigación en Científica y Tecnológica (PAICYT).**

## **RESUMEN**

### **CODIFICACIÓN DE VOZ EN SUBBANDAS**

Publicación No. \_\_\_\_\_

Mario Sánchez Monroy, M.C. de la Ingeniería en Telecomunicaciones.  
Universidad Autónoma de Nuevo León, 2001.

Profesor Asesor: Dr. José Antonio de la O Serna.

En el procesamiento de señales de voz existen tres áreas de desarrollo: la codificación de la señal, el reconocimiento y la producción sintética de voz. Este trabajo se centra en uno de los métodos para codificar la señal o, en otras palabras, representarla eficientemente en un formato digital.

El objetivo del método de codificación en subbandas es dividir la señal de entrada en varios componentes de frecuencia y después codificar cada uno de esos componentes de manera separada. Con esto se busca disminuir lo más posible la tasa de transmisión de la señal sin que se muestre gran deterioro en su calidad. El interés primordial es su aplicación a la telefonía digital, donde debe conservarse la inteligibilidad de la señal permitiendo también la identificación del hablante.

El método aprovecha las características espectrales de la señal de voz, que tiene un mayor contenido de energía en las bandas más bajas, para codificar éstas con una mayor cantidad de bits, asignando menos bits a las bandas de alta frecuencia.

En la tesis se describe la importancia de la codificación en general, se presentan las propiedades de la señal de voz y cómo son utilizadas por la codificación en subbandas. También se exponen las operaciones que se llevan a cabo para el procesamiento de la señal y los factores que influyen en la reconstrucción final. Se explica el diseño de los filtros

espejo en cuadratura para la eliminación del aliasing y la distorsión considerando las operaciones de diezmado y expansión de la señal, necesarias para conservar el número de muestras original. Estas operaciones llevan a la relación con las onduletas, las cuales se describen brevemente y se utilizan para la codificación mediante la aplicación de la transformada discreta de onduleta.

Se describe el problema de cómo elegir los cuantificadores de cada una de las subbandas que es el recurso que se utiliza para la asignación del número de bits por muestra en las mismas. Esto se lleva a cabo tomando en cuenta la contribución de cada banda a la distorsión total, la cual se debe minimizar. Se muestra también la diferencia entre la codificación en subbandas con cuantificación uniforme y logarítmica; la influencia del número de bandas de división y de la frecuencia de muestreo de la señal original, así como el desempeño usando diferentes esquemas de división y distintos tipos de onduleta.

Por último se realiza la evaluación de distintos esquemas de codificación para poder proponer uno que ofrezca el mejor desempeño posible de acuerdo con el objetivo del método.

Resumen.	vi
Índice de figuras.	x
Índice de tablas.	xii

**CAPITULO I  
INTRODUCCIÓN**

1.1 LAS TÉCNICAS DE CODIFICACIÓN	1
1.1.1. Importancia de la codificación.	1
1.1.2. Diferentes tipos de codificadores.	3
1.1.3. Medidas de calidad y desempeño de los codificadores de voz.	6
1.2 LA SEÑAL DE VOZ	10
1.2.1. El espectro de la voz.	10

**CAPITULO II  
PROCESAMIENTO DE LA SEÑAL**

2.1 DESCOMPOSICIÓN DEL ESPECTRO DE VOZ	14
2.1.1. Tipos de división.	14
2.1.2. Descomposición en dos canales.	17
2.2 OPERACIONES BÁSICAS	18
2.2.1. Diezmado de la señal.	18
2.2.2. Interpolación.	22
2.3 FILTRADO	24
2.3.1. Características del filtrado en la codificación de subbandas.	24
2.3.2. Banco de filtros espejo en cuadratura.	25
2.3.3. Distorsión de amplitud y de fase.	28
2.4 ONDULETAS	30
2.4.1. El concepto de onduleta.	30
2.4.2. La transformada discreta de onduleta.	32

**CAPÍTULO III  
CODIFICACIÓN**

3.1. CUANTIFICACIÓN	39
3.1.1. Características de la cuantificación.	39
3.1.2. Ruido de cuantificación y tasa de bits.	41

## ÍNDICE (Continuación)

3.2. LA CODIFICACIÓN PCM	42
3.2.1. Características del sistema PCM.	42
3.2.2. Cuantificación logarítmica	43
3.3. CODIFICACIÓN DE LAS SUBBANDAS	44
3.3.1. Objetivo de la codificación de subbandas.	44
3.3.2. Tasa de transmisión y distorsión.	45
3.3.3. Asignación de la tasa de bits por muestra.	48
3.3.4. Tasa óptima de bits por muestra para bandas octavas.	51

### CAPÍTULO IV DESEMPEÑO DE LA CODIFICACIÓN EN SUBBANDAS

4.1. PROPÓSITO	55
4.2. CODIFICACIÓN CON CUANTIFICACIÓN UNIFORME	56
4.2.1. Ventaja sobre la codificación en banda completa.	56
4.2.2. Influencia del número de subbandas.	58
4.3. CODIFICACIÓN CON CUANTIFICACIÓN LOGARÍTMICA	60
4.3.1. Comportamiento de la cuantificación logarítmica.	60
4.3.2. Reducción de la frecuencia de muestreo.	61
4.3.3. Diferentes esquemas de codificación.	62
4.3.4. Esquema de codificación propuesto.	65
4.3.5. Contribución de las subbandas al desempeño.	65
4.3.6. Tipos de onduleta.	66
4.4. CONSIDERACIONES SOBRE EL PROCESAMIENTO DE LA SEÑAL	68
4.4.1. Normalización de los filtros.	68
4.4.2. Orden de las operaciones de procesamiento.	69
4.5. OBSERVACIONES	70

### CAPÍTULO V CONCLUSIONES

5.1. SÍNTESIS	71
5.2. APORTACIONES	72
5.3. TRABAJOS FUTUROS	73
REFERENCIAS	74

## ÍNDICE DE FIGURAS

1.1.	Comparación de la calidad perceptiva de los codificadores de forma de onda, de fuente e híbridos.	9
1.2.	Espectro típico de una señal de voz muestreada a 8 kHz.	11
1.3.	Señal de voz vocalizada (a) y su espectro (b).	12
1.4.	Señal de voz no vocalizada (a) y su espectro (b).	12
2.1.	Estructura de árbol para descomposición de bandas uniformes.	15
2.2.	Estructura de árbol para descomposición de bandas por octavas.	16
2.3.	Descomposición multibandas.	16
2.4.	Codificación en dos subbandas.	17
2.5.	Diezmador de plegamiento M.	19
2.6.	Efecto del diezmado para M=2.	19
2.7.	Efecto del filtrado y el diezmado en el dominio de la frecuencia para M=2.	21
2.8.	Expansor de despliegue L.	22
2.9.	Efecto de la expansión para L=2.	22
2.10.	Efecto de la expansión y el filtrado en el dominio de la frecuencia para L=2.	23
2.11.	Respuesta característica ideal de los filtros digitales.	25
2.12.	Esquema de la respuesta típica de los filtros espejo.	26
2.13.	Análisis y reconstrucción de $x(n)$ .	27
2.14.	Comparación de las respuestas en frecuencia de un filtro FIR de 20 coeficientes y otro obtenido de la onduleta DB10.	31
2.15.	Descomposición de una señal de voz en cuatro bandas usando la DWT.	36
3.1.	Cuantificación uniforme de elevación centrada (a) y de escalón centrado (b)	40
3.2.	Distorsión y tasa de bits.	47
3.3.	Cuantificadores en un sistema de M bandas uniformes.	48
4.1.	Comparación del desempeño en 2, 3 y 4 subbandas octavas para la palabra 'estudia'.	57
4.2.	Comparación del desempeño en 2, 3 y 4 subbandas octavas para la palabra 'virtud'.	58
4.3.	Desempeño para la codificación de la palabra 'virtud' de 2 a 8 subbandas, a 16, 24 y 32 kbits/.	59
4.4.	Comparación entre la cuantificación logarítmica y la uniforme para una codificación en cuatro subbandas octavas.	60

## ÍNDICE DE FIGURAS (Cont.nuación)

4.5.	Desempeño con frecuencias de muestreo de 6.4 y 8 kHz, a tasas de transmisión de 16 y 32 kbits/s.	61
4.6.	Esquema para codificación a 16 y 32 kbits/s en cinco bandas con tres etapas de filtrado.	62
4.7.	Relación señal a ruido para la codificación en subbandas octavas a 16 kbits/s de <i>Frase 2</i> usando tres familias de onduletas.	67
4.8.	Relación señal a ruido para la codificación en subbandas octavas a 16 kbits/s de <i>Frase 3</i> usando tres familias de onduletas.	67
4.9.	Coefficientes de la onduleta DB10.	68



## ÍNDICE DE TABLAS

1.1.	Distintos tipos de codificadores de voz.	5
1.2.	Medidas de calidad MOS.	8
1.3.	Análisis comparativo de codificadores usados comúnmente.	10
3.1.	Asignación de la tasa de bits para una descomposición en cuatro bandas uniformes y comparación con la codificación en banda completa usando cuantificación uniforme.	51
3.2.	Asignación de la tasa de bits para una descomposición en tres bandas octavas y comparación con la codificación en banda completa usando cuantificación uniforme.	53
4.1.	Desempeño a 16 kbits/s en 5 bandas para distintas distribuciones de $b_k$ .	63
4.2.	Desempeño a 32 kbits/s en 5 bandas para distintas distribuciones de $b_k$ .	64
4.3.	Desempeño de la codificación en subbandas octavas a 16 y 32 kbits/s ( $f_m = 8$ kHz).	64
4.4.	Desempeño de la codificación en subbandas octavas a 16 y 32 kbits/s ( $f_m = 6.4$ kHz).	65
4.5.	Desempeño total y contribución de las subbandas.	66

## INTRODUCCIÓN

### 1.1 LAS TÉCNICAS DE CODIFICACIÓN

#### *1.1.1 Importancia de la codificación*

Cuando se quiere transmitir cualquier tipo de señal a través de un medio específico, siempre es necesario optimizar las características de la señal con el objetivo de poder manejar una mayor cantidad de información. Usualmente esto se realiza mediante alguna técnica de compresión que ofrezca el mínimo de errores en la recuperación de la información.

Según el teorema general de la transmisión de información [11] siempre es posible transmitir información a una velocidad menor que la capacidad del canal con un error relativamente pequeño, aun en presencia de ruido.

El concepto de capacidad de canal, así como también la medida de la información y la codificación se estudian en la teoría de la información.

Los medios de transmisión de señales de información que se usan cotidianamente son: enlaces de ondas electromagnéticas, líneas de transmisión alámbricas y fibras ópticas. Cada uno de ellos tiene propiedades bien definidas en lo general y lo específico. Aunque existe una enorme ventaja en las transmisiones por fibra óptica, no obstante, en lo que al presente trabajo se refiere, la transmisión alámbrica sigue teniendo una importancia fundamental, pues en telefonía se sigue haciendo amplio uso de las redes convencionales. No olvidemos tampoco el crecimiento de los sistemas inalámbricos, sobre todo en telefonía celular, los cuales requieren, además de la codificación, sistemas especiales para poder proteger la privacidad del servicio.

Hay dos limitaciones fundamentales en los medios mencionados: el ancho de banda y el ruido. Es aquí donde las fibras ópticas ofrecen enormes ventajas. Pero la máxima limitación la impone el canal de menor capacidad y, en nuestro caso, siempre están presentes los otros dos medios. La meta es transmitir más cantidad de información en menos tiempo, por lo que se necesita tomar en cuenta el ancho de banda y el ruido. Las comunicaciones digitales permiten lograr esto eficientemente, además constantemente se desarrollan nuevos métodos de codificación.

Por lo que se refiere a la limitación del ancho de banda el teorema de Nyquist establece que la capacidad o velocidad máxima  $R$  de datos a que puede transmitirse la información por un canal de ancho de banda  $B$ , utilizando  $V$  niveles discretos es [11]

$$R = 2 \cdot B \cdot \log_2 V \quad \text{bits/seg} \quad (1.1.1)$$

Así, en un canal de 3 kHz sin ruido no se podrían transmitir señales representadas en dos niveles lógicos a una velocidad superior a los 6 kbits/seg.

Si consideramos un canal con ruido, el teorema de Shannon establece que

$$R = B \cdot \log_2(1 + S/N) \quad \text{bits/seg} \quad (1.1.2)$$

donde S/N representa la relación señal a ruido medida en decibeles. Así, en un canal de 3 kHz con una relación señal a ruido de 30 dB no se podría transmitir a más de 30 kbits/seg; el máximo número de niveles discretos que se podrían usar par codificar la señal y enviarla por ese canal sería igual a 32, de acuerdo con la ecuación (1.1.1).

Para comprender un poco más la importancia de la codificación, consideremos el siguiente ejemplo. Supongamos que deseamos transmitir una hora de música muestreada a 44.1 kHz y codificada a 16 bits (calidad de CD) por medio de un módem que sólo puede transmitir a 14.4 kbits/seg. La cantidad total de información a transmitir sería:

$$(44,100 \text{ muestras/seg}) \times (16 \text{ bits/muestra}) \times (3,600 \text{ seg}) = 2.73 \text{ Gbits}$$

transmitiendo esta información a 14.4 kbits/seg tendríamos

$$(2.73 \text{ Gbits}) / (14,400 \text{ bits/seg}) = 176,400 \text{ seg} = 49 \text{ horas}$$

De lo anterior resulta que necesitaríamos un poco más de dos días para poder transmitir nuestra información; esto utilizando sólo un canal, pues si quisiéramos transmitir en calidad estéreo necesitaríamos el doble de tiempo.

La codificación de información ofrece algunas otras ventajas. La señal codificada es fácil de reconocer para el detector y puede ser regenerada eficientemente; puede también ser encriptada para aplicaciones que requieran privacidad. Además, la codificación provee un formato uniforme para diferentes tipos de señales.

Existen muchos métodos para codificar eficientemente la información y enseguida los resumiremos brevemente.

### ***1.1.2 Diferentes tipos de codificadores***

Un ejemplo actual sobre el tema de la codificación de subbandas lo constituyen los estándares MPEG (*Motion Picture Expert Group*). Cualquier persona que haga uso de

Internet sabe que puede obtener de la red archivos de audio con su música favorita los cuales pueden ser escuchados en su propia computadora o ser grabados en un CD para coleccionarlos. Al referirnos a tales archivos los designamos por su extensión: MP3. Es ampliamente usado y es sólo una de las formas de codificación MPEG, pues se refiere al algoritmo de compresión MPEG-1 Layer 3.

El algoritmo de compresión MPEG es un estándar ISO (*International Organization for Standardization*) desarrollado para sistemas que tienen que ver con la sincronización audio-video; por lo que la compresión de audio es sólo una pequeña parte de dicho proyecto.

Básicamente un codificador MPEG comprime todo aquello que es audible tomando en cuenta las características de percepción del oído a la señal concreta que se escuchará. El que una persona escuche un sonido o no, depende de la frecuencia del sonido y de que su amplitud esté por encima o por debajo del umbral de audición de la persona a esa frecuencia. Por ejemplo un instrumento de orquesta tocando *fortissimo* hará, en mayor o menor medida, que los sonidos de algunos de los demás instrumentos se vuelvan inaudibles para nuestro oído. Este fenómeno se conoce con el nombre de enmascaramiento (*masking*) y las técnicas que lo aprovechan se denominan técnicas de codificación perceptiva (*perceptual coding*) [4,12].

Los codificadores comúnmente usados para la compresión de las señales de voz se pueden dividir en tres clases: de forma de onda (*waveform coders*), de fuente (*source coders*) e híbridos (*hybrid coders*).

Los codificadores de forma de onda intentan producir una señal que sea lo más parecida posible a la original y los algoritmos que utiliza son generalmente de baja complejidad. Los codificadores de subbandas son de este tipo.

Los codificadores de fuente (mejor conocidos como vocoders) trabajan usando un modelo de la forma en que fue generada la señal y asocian a la señal codificada un parámetro de dicho modelo que es transmitido al decodificador.

Los codificadores híbridos combinan los dos anteriores para aprovechar sus características.

La tabla 1.1 muestra los codificadores más conocidos ordenados de acuerdo con la clasificación mencionada. La descripción de estos codificadores queda fuera de los propósitos de este trabajo y solamente se hará referencia a sus niveles de desempeño como punto de comparación con los codificadores de subbandas. Sólo el PCM será tratado con brevedad posteriormente por utilizarse para la cuantificación de la información de cada subbanda.

Tabla 1.1. Distintos tipos de codificadores de voz

CODIFICADORES DE FORMA DE ONDA	PCM ( <i>Pulse Code Modulation</i> ) DPCM ( <i>Differential PCM</i> ) ADPCM ( <i>Adaptive DPCM</i> ) ATC ( <i>Adaptive Transform Coding</i> )	
CODIFICADORES DE FUENTE	VOCODERS ( <i>Voice Coders</i> )	
CODIFICADORES HÍBRIDOS	LPC ( <i>Linear Prediction Code</i> )	MPE ( <i>Multi Pulse Excited</i> ) RPE ( <i>Regular Pulse Excited</i> ) CELP ( <i>Code Excited Linear Predictive</i> ) MBE ( <i>Multiband Excitation</i> ) PWI ( <i>Prototype Waveform Interpolation</i> )

Además de los codificadores mencionados en la tabla 1.1, existen también los siguientes, de tecnología más avanzada: AAC (*Advanced Audio Coding*), ASPEC, VOCPACK y AWT (*Adaptive Wavelet Technology*); el tema de las onduletas (wavelets) se tratará posteriormente.

### ***1.1.3. Medidas de calidad y desempeño de los codificadores de voz.***

La medida de la calidad de la voz es una tarea importante, pero al mismo tiempo difícil. La SNR es una de las medidas objetivas más comunes para evaluar el desempeño de un algoritmo de compresión. Esta se obtiene como:

$$\text{SNR} = 10 \log_{10} \left\{ \frac{\sum_{n=0}^M s^2(n)}{\sum_{n=0}^M (s(n) - \hat{s}(n))^2} \right\} \quad (1.1.3)$$

donde  $s(n)$  es la señal de habla original,  $\hat{s}(n)$  es la señal de habla codificada y  $M$  es el número de muestras de la señal.

Normalmente, un algoritmo de codificación de voz es evaluado basándose en la tasa de bits, la calidad de reconstrucción de la voz, la razón de compresión, la complejidad del algoritmo, el retraso introducido y la robustez del algoritmo ante errores de canal e interferencia acústica. La calidad de la voz está clasificada dentro de cuatro categorías generales:

- a) Calidad de transmisión (*broadcast quality*)
- b) Calidad de red o calidad telefónica (*network quality, toll quality*)

c) Calidad de comunicación (*communication quality*)

d) Calidad sintética.

La voz de alta calidad de transmisión está referida a la transmisión de voz en banda ancha, que generalmente se realiza a tasas por encima de los 64 kbits/seg.

La calidad telefónica se refiere a la que es comparable con una señal análoga de voz típica (200-3200 Hz) y que puede ser implementada a tasas superiores a los 16 kbits/seg; aunque hay codificadores que proveen esta calidad incluso hasta los 11 kbits/seg.

La calidad de comunicación es aquella aceptada comúnmente por los usuarios de los teléfonos celulares. Implica algo de degradación de la voz, que a pesar de no ser natural es altamente inteligible y adecuada para telecomunicaciones. Las tasas de transmisión para este criterio pueden ir de los 12 kbits/s hasta los 4 kbits/seg

La calidad sintética es la de la voz generada por computadora. En ella, la voz pierde sus cualidades naturales y, como consecuencia, no se puede reconocer al interlocutor. Se tiene esta calidad por debajo de los 4 kbits/s.

Aunque se hace referencia a la tasa de bits para cada uno de las cuatro categorías anteriores, los niveles de calidad que representan son mediciones subjetivas que determinan como es percibido el sonido por los oyentes y sus impresiones acerca de él.

Una medida ampliamente usada para medir la calidad de la voz codificada es la prueba MOS (*Mean Opinion Score*) o media de las calificaciones de opinión. Usualmente involucra de 12 a 24 oyentes –las pruebas formales de la CCITT (*Consultative Committee on International Telegraphy and Telephony*) involucran de 32 a 64 oyentes– los cuales son instruidos para registrar las medidas de balances fonéticos de acuerdo con una escala de calidad de 5 niveles, como se muestra en la tabla 1.2.



Una calidad de voz excelente implica que es imposible distinguir la voz codificada de la original y sin ruido alguno perceptible. Por otra parte, una mala calidad implica la presencia de ruido extremadamente molesto y artificial en la voz codificada.

Tabla 1.2. Medidas de calidad MOS.

Escala de MOS	Calidad de la Voz
1	Mala
2	Suficiente
3	Regular
4	Buena
5	Excelente

En la prueba MOS los oyentes están habituados y familiarizados con las condiciones de audición y el rango de la calidad de voz en el que se encuentran. La puntuación es obtenida promediando las calificaciones numéricas sobre varios cientos de registros de voz. Los rangos MOS se relacionan con la calidad del mensaje de voz como sigue: un MOS de 4.5 a 5 significa calidad en la transmisión; uno entre 4 y 4.5, calidad telefónica; entre 3.5 y 4, calidad de comunicación; y entre 2.5 y 3.5, calidad sintética.

Las puntuaciones de MOS pueden diferir significativamente de una prueba a otra, por lo que no son medidas absolutas de comparación entre diferentes codificadores.

En la figura 1.1 se muestra una comparación de la calidad de la voz tomando en cuenta el tipo de codificador general y la tasa de bits que se puede obtener con cada uno. Recuérdese que se trata de medidas de calidad subjetivas y que dependen de la experiencia y las impresiones de un gran número de oyentes.

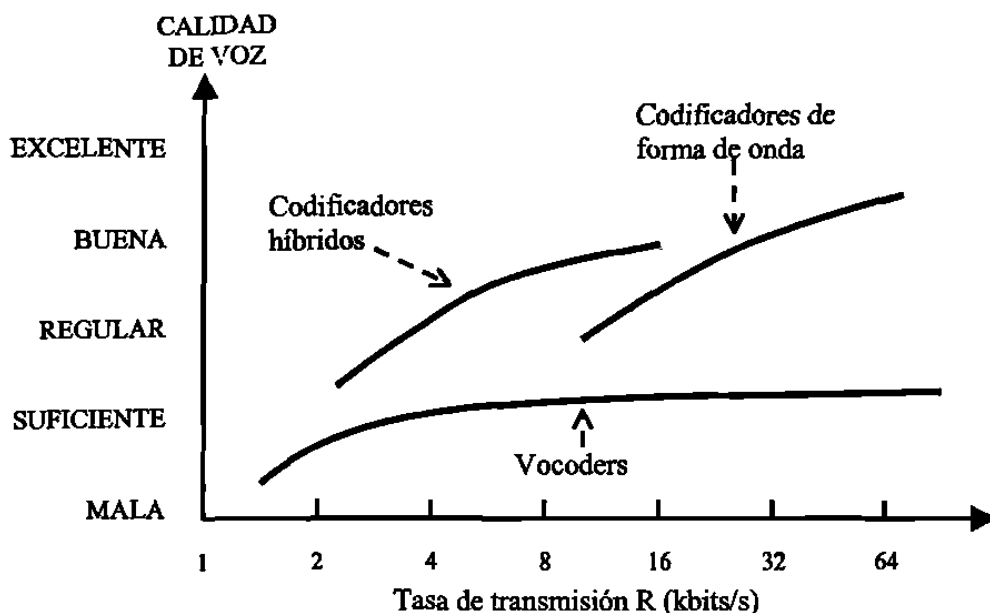


Figura 1.1. Comparación de la calidad perceptiva de los codificadores de forma de onda, de fuente e híbridos.

Las evaluaciones formales subjetivas, como las descritas anteriormente, pueden ser largas y caras; por este motivo los esfuerzos más recientes en calidad de voz se están dirigiendo a *procedimientos de pruebas de evaluación automáticas y hacia medidas objetivas* que son capaces de predecir la calidad subjetiva de la voz.

Las técnicas más comúnmente usadas en la codificación se resumen en la tabla 3. Se puede observar la relación existente entre la tasa de transmisión, la razón de compresión y la calidad de la voz obtenida. La tasa de transmisión de referencia para obtener la razón de compresión es el de la calidad de CD (44.1 kHz, 16 bits/muestra, un solo canal, sólo para el caso de la voz). El último elemento de la tabla se refiere a una técnica que usa onduletas. Más adelante se analizará el empleo de éstas en los codificadores de subbandas. Se debe tomar en cuenta que algunos de estos codificadores permiten la transmisión a tasas de transmisión variables usando frecuencias de muestreo variables. En la tabla 1.3 se asume un rango que va de los 8 a los 16 bits/muestra y una frecuencia de muestreo que oscila entre los 8 y los 44.1 kHz.

Tabla 1.3. Análisis comparativo de codificadores usados comúnmente.

Codificador	Tasa de transmisión	Razón de compresión	Calidad de la voz
PCM	64 kbits/s	1:11	Buena
ADPCM G721	32 kbits/s	1:22	Buena
CELP G728	16 kbits/s	1:44	Regular
GSM	13 kbits/s	1:54	Suficiente
CELP DoD	4.8 kbits/s	1:147	Mala
MPEG III	128 kbits/s	1:6	Excelente
RAPID TRANSIT	128 kbits/s	1:6	Excelente

## 1.2 LA SEÑAL DE VOZ

### 1.2.1 El espectro de la voz

El oído humano es capaz de escuchar señales con frecuencias en el rango de los 20Hz a los 20 kHz. No obstante, la mayoría de la gente no puede escuchar la porción superior de este rango y en algunas personas la frecuencia de corte superior decrece con la edad. En el caso de la voz, su espectro contiene frecuencias que, para efectos prácticos, no se extienden más allá de los 5 kHz. En la figura 1.2 se muestra la magnitud de la transformada de Fourier típica para una señal de voz muestreada a 8 kHz. La magnitud del pico del espectro depende de la fisiología de la persona que habla, así como del lenguaje en que lo hace. Si se considera un silbido humano, la forma de onda es una senoide pura y su transformada de Fourier será un impulso ubicado en la frecuencia del silbido. Si la persona tararea a la misma frecuencia, la transformada de Fourier contendrá la frecuencia fundamental más sus armónicas en múltiplos de dicha frecuencia.

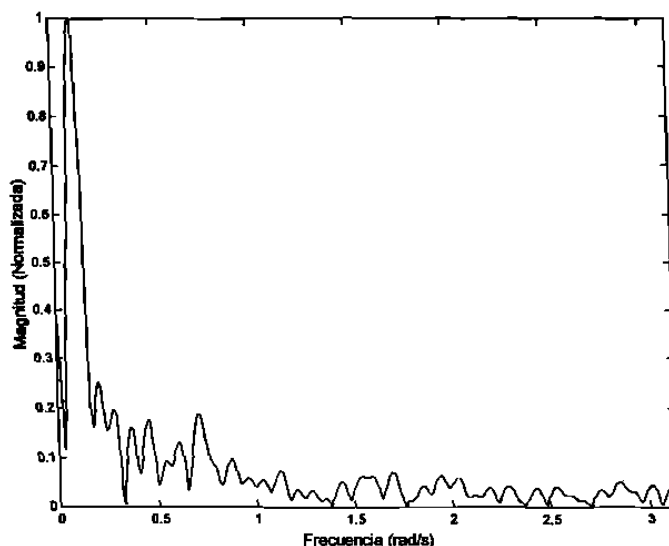


Figura 1.2. Espectro típico de una señal de voz muestreada a 8 kHz.

En telefonía, la experimentación muestra que la porción de la señal de voz entre los 200 Hz y los 3200 Hz es suficiente para la inteligibilidad del habla y el reconocimiento del hablante. Sin embargo dentro de este rango no se tiene una gran fidelidad para señales audibles diferentes a la voz. Para música, por ejemplo, se requieren rangos por encima de los 15 kHz. Los sistemas típicos de entretenimiento casero operan a frecuencias superiores a los 20 kHz. Para comunicaciones digitales, la señal de voz está limitada en banda a 4 kHz (o 3.2 kHz) y muestreada generalmente a 8 kHz (estándar).

Las señales de voz son funciones continuas de tiempo limitado; no son estacionarias, y en el mejor de los casos pueden ser consideradas como quasi-estacionarias en segmentos cortos que van típicamente de 5 ms a 20 ms. Por lo tanto, las propiedades espectrales y estadísticas de la voz son definidas en segmentos cortos.

La voz puede ser clasificada como: vocalizada (/a/, /i/, etc.), no vocalizada (/sh/) y mixta. La figura 1.3 muestra una señal de voz vocalizada y su espectro. Del mismo modo, la figura 1.4 muestra las características de las señales no vocalizadas.

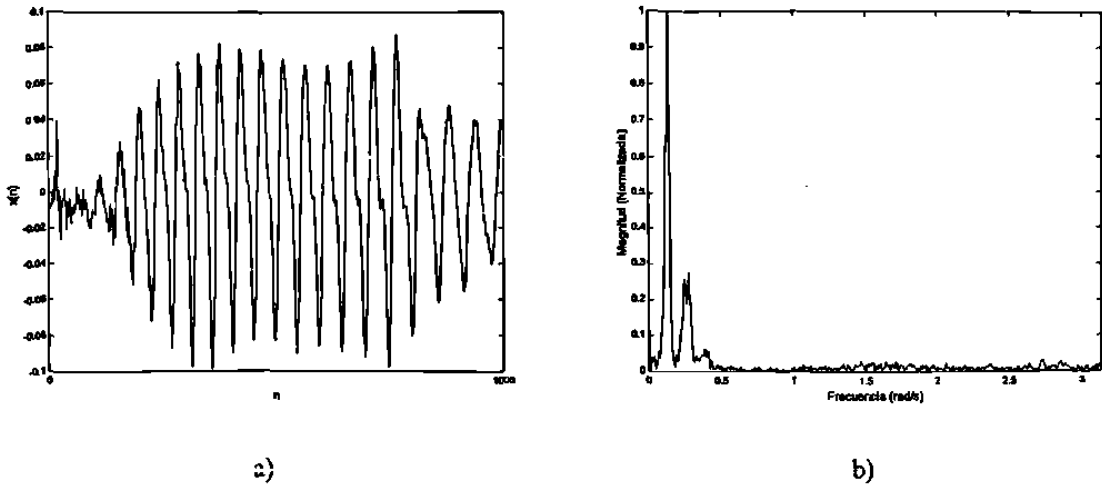


Figura 1.3. Señal de voz vocalizada (a) y su espectro (b).

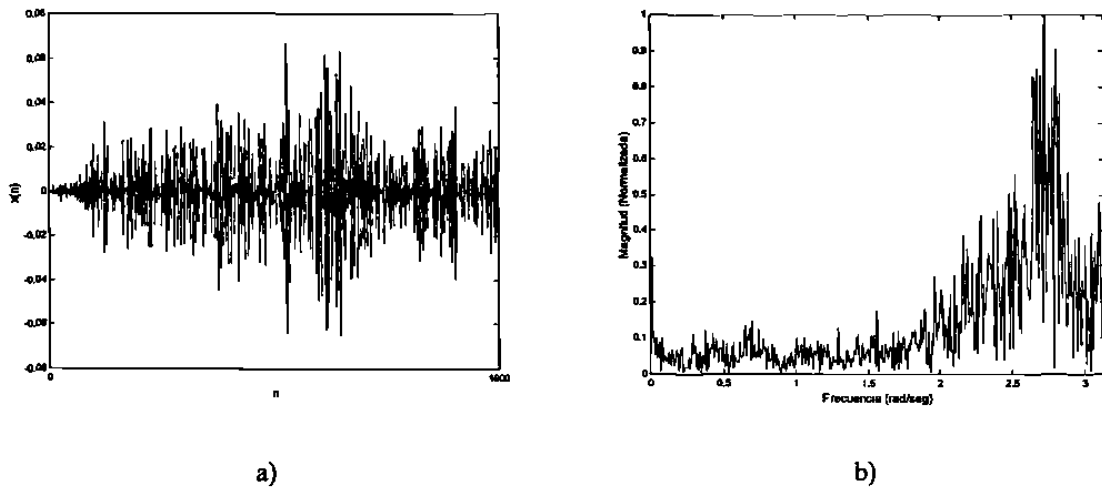


Figura 1.4. Señal de voz no vocalizada (a) y su espectro (b).

La envolvente espectral está caracterizada por un conjunto de picos, los cuales reciben el nombre de formantes. En promedio, hay de tres a cinco picos formantes por debajo de los 5 kHz. Las amplitudes y localizaciones de los tres primeros ocurren usualmente debajo de los 3 kHz y son sumamente importantes en la síntesis de voz y su

percepción. Los formantes más altos influyen en el ancho de banda y las representaciones de voz no vocalizada.

La localización de los formantes es la característica más importante en el procesamiento de señales de voz. El ancho de banda, la amplitud o la pendiente espectral son características secundarias. El primer formante es más importante en la calidad de la percepción, y las altas frecuencias son más importantes para la inteligibilidad.

Normalmente el mayor contenido de energía de una señal de voz se encuentra localizado en las frecuencias más bajas. Esta característica es aprovechada en la codificación de subbandas en la que las bandas de frecuencias más bajas, usualmente, ocupan más bits que las bandas de frecuencias más altas.

# PROCESAMIENTO DE LA SEÑAL

## 2.1 DESCOMPOSICIÓN DEL ESPECTRO DE VOZ.

### 2.1.1 Tipos de división

Originalmente, el método de codificación en sub-bandas fue desarrollado para codificar o comprimir señales de audio [6], aunque actualmente se usa también en imágenes. El método busca descomponer la señal en anchos de banda más pequeños y representarlos, sin redundancia, de tal manera que sea posible reconstruir la señal original sin error. Las bandas de frecuencia más baja contienen la mayor parte de la energía espectral de habla vocalizada. Además, el ruido de cuantificación es más evidente al oído en esas frecuencias. Consecuentemente, se deben usar más bits para las bandas más bajas, y unos pocos para las bandas más altas.

Existen tres esquemas básicos para la división de las subbandas: estructura de árbol de bandas uniformes, estructura de árbol de bandas en octavas y estructura multibanda. Las figuras 2.1, 2.2 y 2.3 muestran, respectivamente, cada uno de estos tipos de descomposición para la etapa de análisis.

La estructura de árbol de bandas uniformes (figura 2.1) utiliza filtros de espejo en cuadratura (*Quadrature Mirror Filter: QMF*) [6,15] los cuales se explicarán en la sección 2.3.2. La señal se divide primero en una banda alta y otra baja, tomando después solo la mitad de las muestras de las señales resultantes, es decir, se realiza un decimado de la señal (esto es lo que se expresa con  $\downarrow 2$ ); cada una de las bandas resultantes se filtra nuevamente con otro par de QMF's repitiendo el mismo procedimiento hasta obtener el número de bandas deseado. Este tipo de división es el utilizado para la codificación MPEG (usa cuatro etapas de filtrado para dar un total de 32 subbandas).

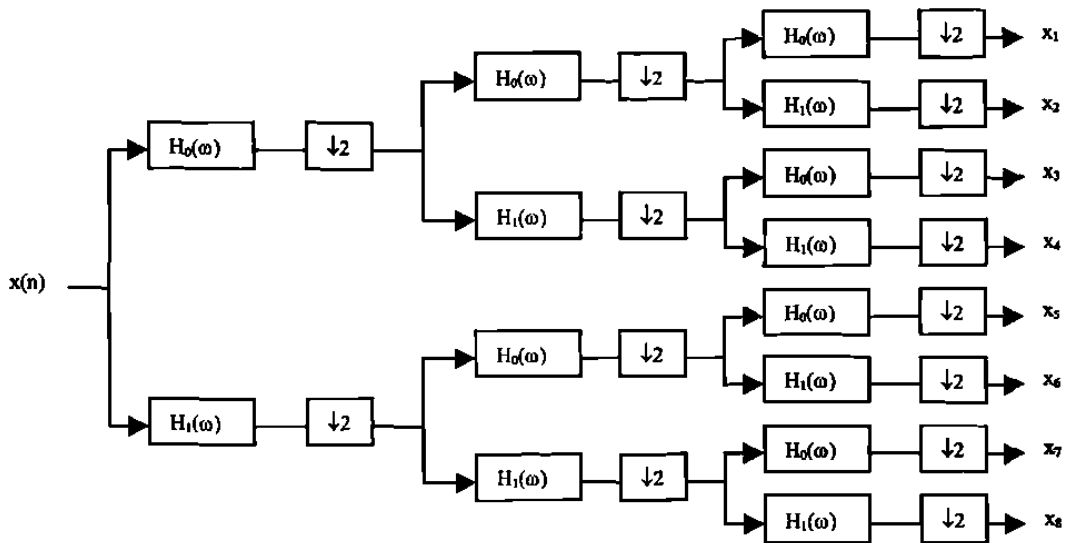


Figura 2.1. Estructura de árbol para descomposición de bandas uniformes.

La estructura de división en octavas<sup>1</sup> (figura 2.2) es similar a la de bandas uniformes. Se inicia filtrando la señal con un par de QMF's del mismo ancho de banda y después se realizan filtrados sucesivos con otros pares de QMF's únicamente para las bandas de frecuencia más baja. El resultado final es que la frecuencia de corte superior de la primera subbanda es la mitad de la segunda; la segunda, la mitad de la tercera; y así sucesivamente.

<sup>1</sup> De el estudio de este tipo de división proviene el desarrollo de las onduletas, como puede verse en [8].



Este tipo de estructura es usado exclusivamente en bancos de filtros de onduletas y son usados extensivamente como una herramienta muy poderosa en codificación de imágenes. La división en bandas uniformes también puede hacerse a partir de filtros con coeficientes de onduleta.

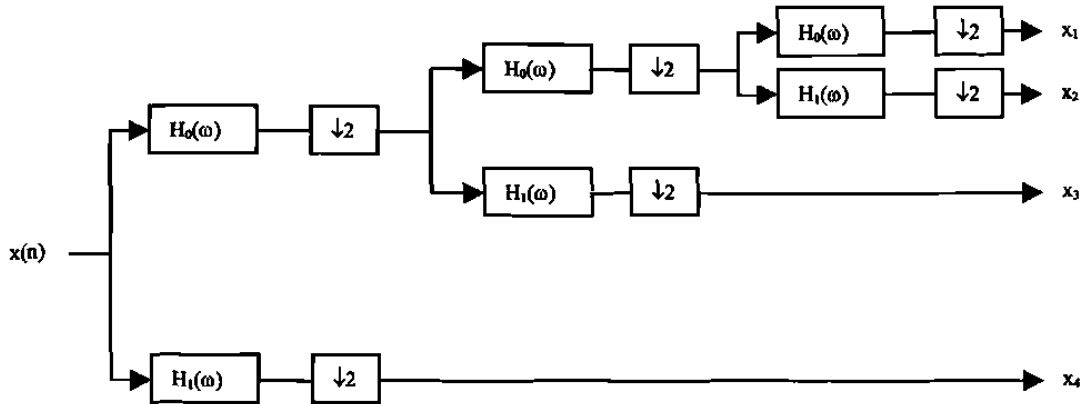


Figura 2.2. Estructura de árbol para descomposición de bandas por octavas.

La descomposición multibanda (figura 2.3) es una división directa, por medio de filtros pasa banda, en un número definido de bandas que pueden tener diferente longitud (con un factor de diezmado  $\downarrow M$  diferente). Su diseño es complejo y usualmente se realiza con bancos de filtros coseno modulado [15, 16]. Uno de los primeros codificadores de subbandas, presentado por Crochiere [18], utiliza una descomposición de este tipo.

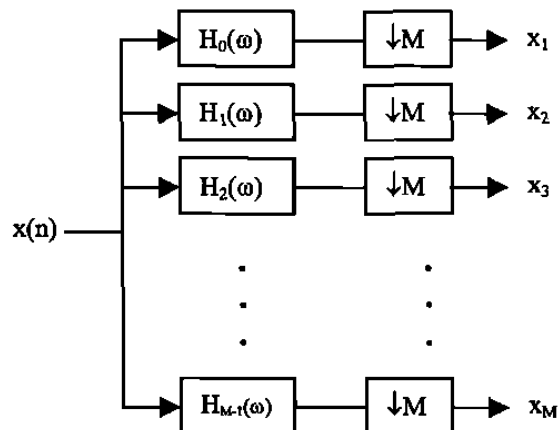


Figura 2.3. Descomposición multibandas.

El proceso básico por el que pasa la señal de voz para la codificación en subbandas en la etapa de análisis es el siguiente:

- Muestreo (paso del tiempo continuo al discreto)
- Descomposición en subbandas por medio del banco de filtros.
- Diezmado de cada subbanda (reducción del número de muestras)<sup>1</sup>
- Digitalización (cuantificación y representación binaria).

### 2.1.2 Descomposición en dos canales

En telefonía, la voz está generalmente muestreada a 8 kHz, aunque se puede muestrear a frecuencias menores si consideramos que se obtiene buena calidad para una señal de voz comprendida entre los 200 y los 3200 Hz. La figura 2.4 es un diagrama simplificado del proceso de codificación; ilustra el sistema de análisis y síntesis para una descomposición en dos canales.

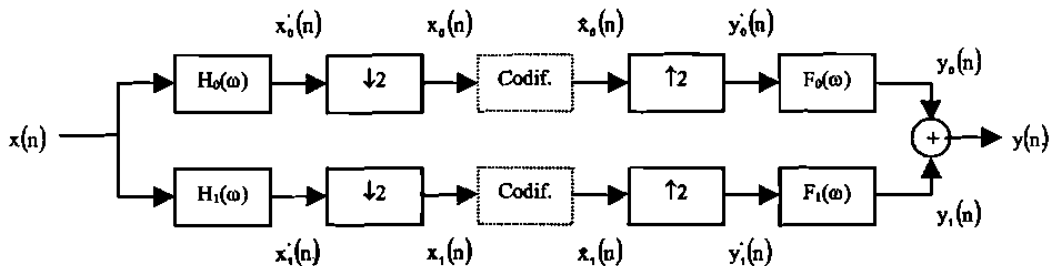


Figura 2.4. Codificación en dos subbandas.

En la etapa de análisis  $H_0(\omega)$  representa el filtro pasa bajas y  $H_1(\omega)$  el pasa altas; la señal muestreada  $x(n)$  pasa a través de estos filtros. En el paso correspondiente al diezmado

<sup>1</sup> La notación que se usará es  $(\downarrow M)$ . Significa tomar solamente cada  $M$ -ésima muestra de la señal. Otra notación que se utiliza es  $M:1$ ; significa reducir de  $M$  muestras a 1 o tomar una de cada  $M$  muestras.

se toma sólo cada segunda muestra de la señal. Después, se realiza la codificación de la señal. En la etapa de síntesis se realiza la interpolación a través de un sobremuestreo o expansión (representado con  $\uparrow 2$ ) insertando ceros en el lugar de las muestras que fueron eliminadas. Después se pasa al banco de filtros de análisis representados como  $F_0(\omega)$  y  $F_1(\omega)$  para el pasa bajas y el pasa altas, respectivamente; se combinan las señales resultantes y se obtiene  $y(n)$ .

El diezmado (o submuestreo) es necesaria para enviar, al final, el mismo número de muestras que la señal  $x(n)$ . El factor de diezmado depende del número de bandas que se utilicen. En este caso, como son dos bandas, se tomo sólo cada segunda muestra a la salida de cada filtro. En el caso de  $M$  filtros, se tomaría sólo una de cada  $M$  muestras, esto únicamente para el caso de una división de bandas uniformes. Para la división en octavas el diezmado es como se ilustra en la figura 2.2, de tal forma que resulta un diezmado mayor para las bandas de frecuencia más baja. Para la descomposición multibanda, el factor de diezmado de cada banda debe ser igual a la frecuencia de muestreo de la señal original entre la frecuencia de muestreo de cada banda.

## 2.2 OPERACIONES BASICAS

### 2.2.1 Diezmado de la señal.

La operación de extraer cada  $M$ -ésima muestra de una señal se conoce comúnmente como diezmado. La secuencia original  $x(n)$  es reemplazada por una nueva secuencia  $y_d(n)$ , la cual es simplemente cada  $M$ -ésimo valor de  $x(n)$ .

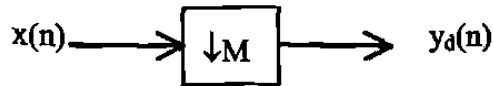


Figura 2.5. Diezmador de plegamiento M.

La figura 2.5 muestra el diezmador de plegamiento M, que opera tomando una secuencia de entrada  $x(n)$  y produce la siguiente secuencia de salida

$$y_d(n) = x(Mn) \quad (2.2.1)$$

donde M es un entero. El diezmador sólo retiene las muestras de  $x(n)$  que ocurren en múltiplos de M. La operación también es conocida como submuestreo.

La figura 2.6 ilustra la operación de diezmado para  $M=2$ , que es la que nos interesa para el análisis de dos subbandas y la aplicación de los QMF's.

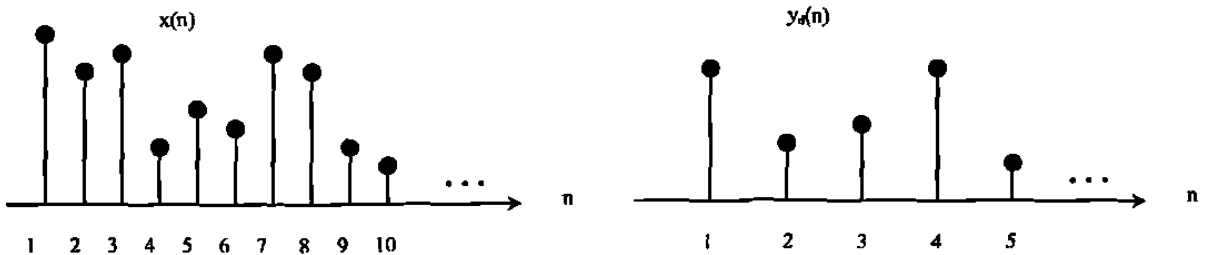


Figura 2.6. Efecto del diezmado para  $M=2$ .

El efecto del diezmado sobre  $x(n)$  equivale a una compresión de la señal en el dominio del tiempo discreto y, por lo tanto, su expansión en el dominio de la frecuencia.

Considerando la transformada de la señal diezmada  $y_d(n)$ , tenemos que:

$$Y_d(\omega) = \sum_{-\infty}^{\infty} y_d(n) e^{-j\omega n} = \sum_{-\infty}^{\infty} x(Mn) e^{-j\omega n} \quad (2.2.2)$$

si cambiamos  $x(Mn)$  por  $x_1(m)$ , se puede escribir:

$$Y_d(\omega) = \sum_{m=-\infty}^{\infty} x_1(m) e^{-j\omega m/M} \quad (2.2.3)$$

tomando en cuenta que  $x_1(m)$  es una secuencia que es igual a cero cuando  $m$  no es un múltiplo entero de  $M$ , de tal forma que:

$$x_1(m) = \begin{cases} x(m), & \text{si } m \text{ es múltiplo de } M \\ 0, & \text{de otra manera.} \end{cases} \quad m = 0, 1, 2, 3, \dots \quad (2.2.4)$$

$x_1(m)$  también puede ser escrita como el producto de una función de submuestreo  $C_M$  y la señal original:

$$x_1(m) = C_M(m)x(m) \quad (2.2.5)$$

$$C_M(m) = \begin{cases} 1, & \text{si } m \text{ es múltiplo de } M \\ 0, & \text{de otra manera.} \end{cases} \quad (2.2.6)$$

Una función que cumple con la condición establecida por la ecuación (2.2.6) para el caso concreto de  $M=2$  es:

$$C_2(m) = \frac{1}{2} (1 + \cos m\pi) \quad (2.2.7)$$

y que convenientemente manipulada se puede escribir como:

$$C_2(m) = \frac{1}{2} \sum_{k=0}^1 e^{-j\pi km} \quad (2.2.8)$$

A partir de la ecuación (2.2.8) se puede generalizar para  $C_M(n)$ , para obtener:

$$C_M(m) = \frac{1}{M} \sum_{k=0}^{M-1} e^{-\frac{j2\pi km}{M}} \quad (2.2.9)$$

Sustituyendo la ecuación (2.2.9) en la ecuación (2.2.3) se obtiene:

$$Y_d(\omega) = \frac{1}{M} \sum_{k=0}^{M-1} X\left(\frac{\omega - 2\pi k}{M}\right) \quad (2.2.10)$$

De acuerdo con la ecuación (2.2.10), la relación entre el espectro de la señal diezmada y el espectro de la señal original cuando  $M=2$  está dada por:

$$Y_d(\omega) = \frac{1}{2} \left[ X\left(\frac{\omega}{2}\right) + X\left(\frac{\omega}{2} - \pi\right) \right] \quad (2.2.11)$$

El diezmado de  $x(n)$  también produce interferencia debida al traslape del espectro de la señal (aliasing) la cual no permite recuperar la señal original. Para evitar esto, la señal se filtra antes del diezmado de tal forma que  $|\omega| < \pi/M$ . La figura 2.7 muestra el resultado de este procedimiento para  $M=2$ .

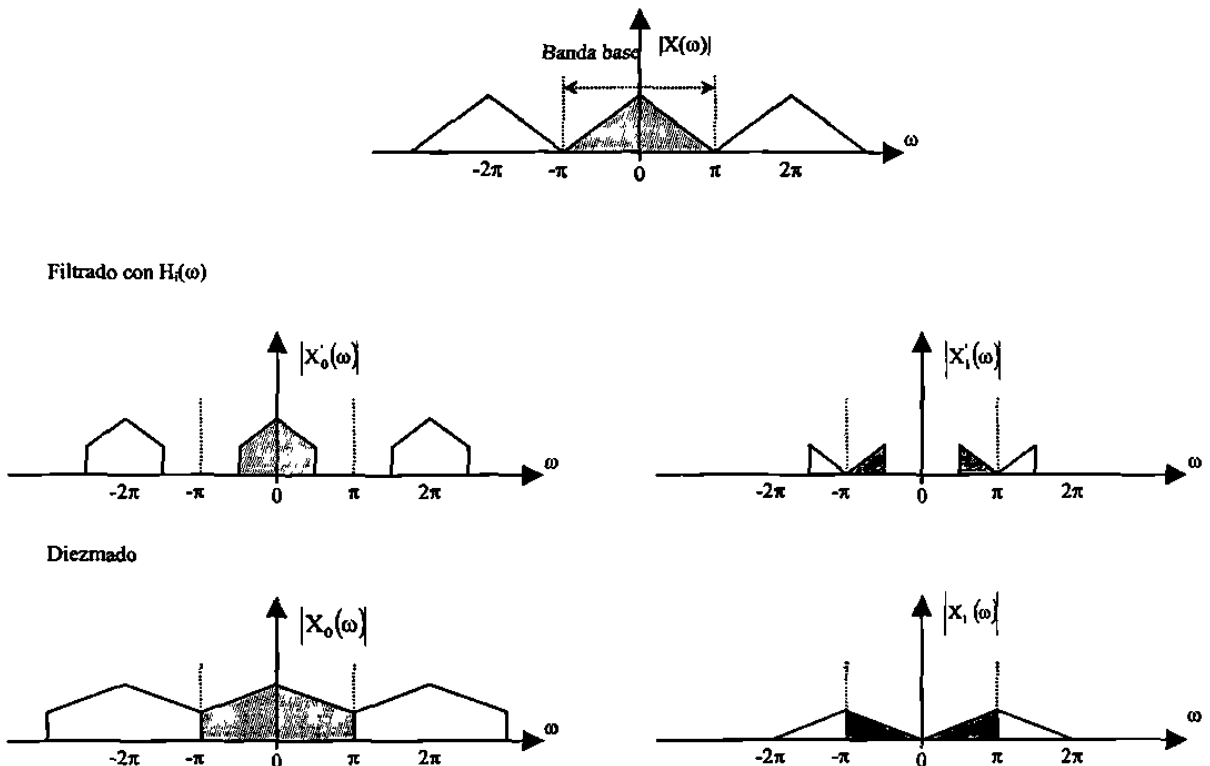


Figura 2.7. Efecto del filtrado y el diezmado en el dominio de la frecuencia para  $M=2$ .

### 2.2.2 Interpolación

El término interpolación se aplica más propiamente al proceso final de reconstrucción de la señal y que se realiza por medio del filtrado. En este apartado se refiere únicamente al proceso inverso del diezmado, el cual es básicamente una expansión del tiempo de la señal que se está procesando. El nombre más adecuado para este proceso es el de expansión.

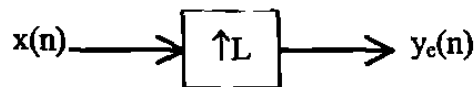


Figura 2.8. Expansor de despliegue L.

El proceso de expansión se lleva a cabo insertando o intercalando ceros en el lugar de las muestras que fueron removidas con el diezmado. La figura 2.8 muestra el expansor de despliegue L, el cual toma una entrada  $x(n)$  y produce la siguiente secuencia de salida

$$y_e(n) = \begin{cases} x(n/L) & \text{si } n \text{ es un entero múltiplo de } L \\ 0 & \text{de otra manera} \end{cases} \quad (2.2.12)$$

donde L es un entero. La operación también es conocida como sobremuestreo. La figura 2.9 muestra la expansión en el dominio del tiempo para  $L=2$ .

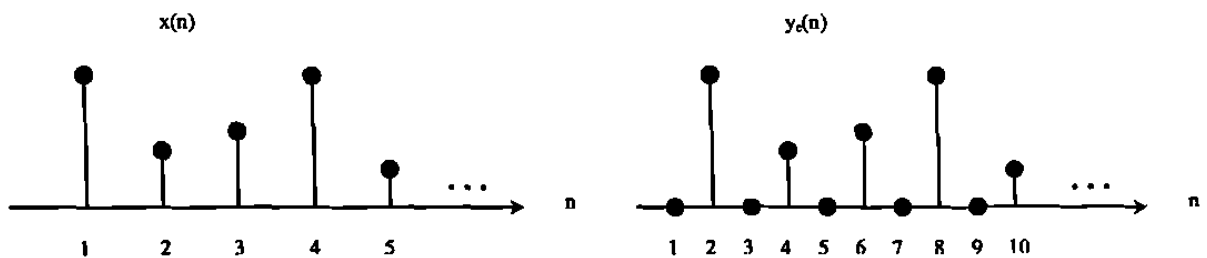


Figura 2.9. Efecto de la expansión para  $L=2$ .

La expansión de la señal en el dominio del tiempo produce la compresión de su espectro en el dominio de la frecuencia. creándose también copias o imágenes del mismo.

La figura 2.10 muestra la operación de expansión y filtrado en el dominio de la frecuencia para  $L=2$ .

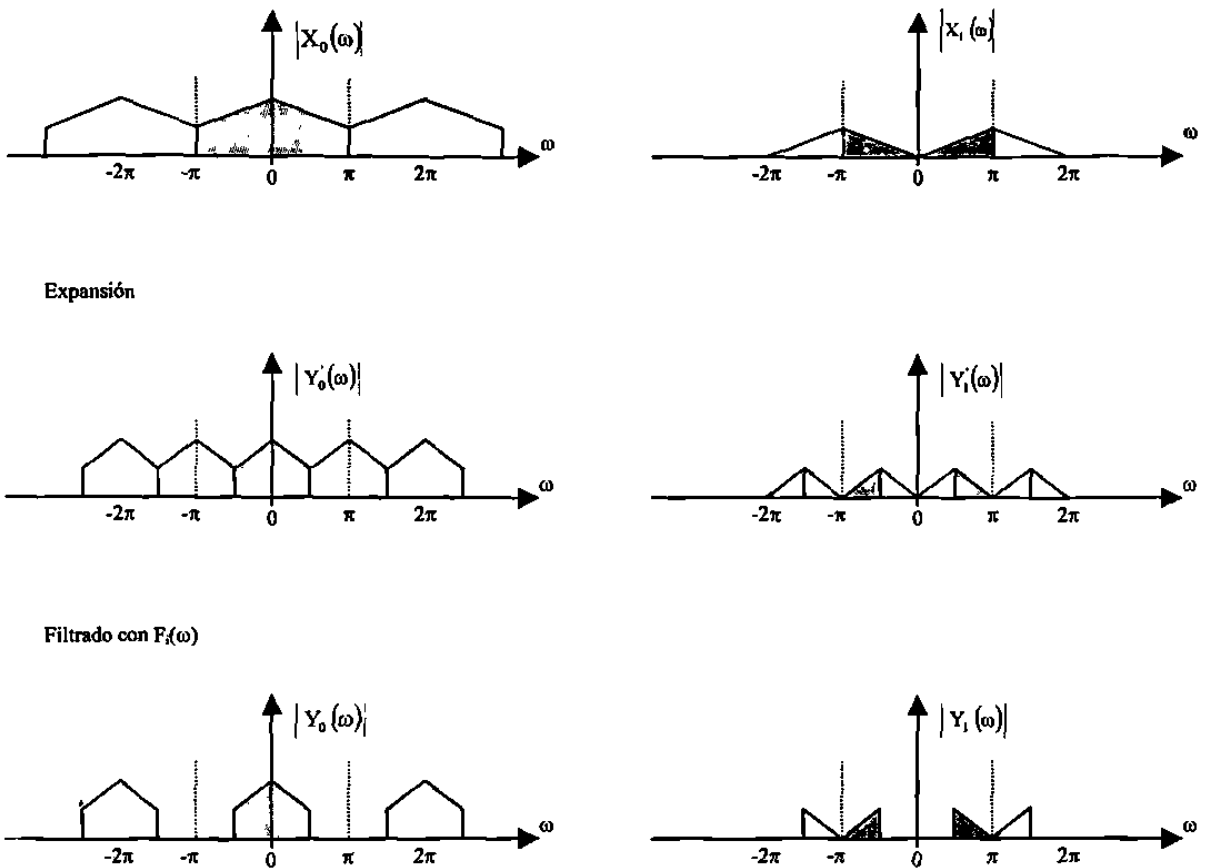


Figura 2.10. Efecto de la expansión y el filtrado en el dominio de la frecuencia para  $L=2$ .

La transformada de la señal expandida  $y_e(n)$  es:

$$Y_e(\omega) = \sum_{n=-\infty}^{\infty} y_e(n) e^{-j\omega n} \quad (2.2.13)$$



y tomando en cuenta la ecuación (2.2.12) se tiene que:

$$Y_e(\omega) = \sum_{k=-\infty}^{\infty} y_e(kL)e^{-j\omega kL} = \sum_{k=-\infty}^{\infty} x(k)e^{-j\omega kL} = X(\omega L) \quad (2.2.14)$$

## 2.3 FILTRADO.

### 2.3.1 Características del filtrado en la codificación de subbandas

Una consideración importante en el desarrollo de la codificación en subbandas es el diseño del banco de filtros. En la sección anterior se ha mostrado el proceso básico de la señal en este tipo de codificación considerando filtros ideales en el proceso de división y reconstrucción del espectro de la señal. Sin embargo el proceso de diezmado produce aliasing; la expansión produce réplicas, y además, las respuestas de los filtros siempre presentarán traslapes de frecuencia.

Tomando en cuenta la figura 2.4 y el proceso mostrado en las figuras 2.7 y 2.10, se esperaría que la señal reconstruida  $y(n)$  fuera igual que la señal original  $x(n)$ , pero no es así. La señal  $y(n)$  es, realmente, una aproximación de  $x(n)$ . Lo que hace que sean diferentes son: el efecto del aliasing, la distorsión de amplitud, la distorsión de fase y también la cuantificación de las subbandas cuando la señal ha sido codificada .

Para el caso de dos canales, el aliasing puede ser completamente eliminado con una buena elección de los filtros de síntesis. En el caso de los bancos de  $M$  filtros con bandas no uniformes (descomposición multibanda), las condiciones para una reconstrucción perfecta y la cancelación del aliasing son más complejas [15].

Los filtros que se usan para la codificación de subbandas son filtros digitales de respuesta a impulso finita (FIR) con coeficientes reales. Cabe recordar que las gráficas de las figuras 2.7 y 2.10 representan respuestas en frecuencia digital y, como puede

observarse, los espectros son periódicos con periodo igual a  $2\pi$  rad/s. Además, si la señal fue muestreada a la frecuencia de Nyquist, su transformada ocupará todo el intervalo  $[0, \pi]$ , lo que corresponderá a toda la banda de la señal analógica original ( ejemplo: 4 kHz para el estándar de voz ).



Figura 2.11. Respuesta característica ideal de los filtros digitales.

Con lo que respecta a las respuestas de los filtros, son pasa bajas los que tienen máxima ganancia en las vecindades de 0 rad/seg y atenúan las frecuencias cercanas a  $\pi$  rad/s; asimismo, los pasa altas tienen máxima ganancia en frecuencias alrededor de  $\pi$  rad/s y atenúan las frecuencias cercanas a 0. Esto se ilustra en la figura 2.11.

### 2.3.2 Banco de filtros espejo en cuadratura.

Un banco de filtros digital es un conjunto de filtros digitales con una entrada común o una salida común. Ambos casos están mostrados en el sistema de la figura 2.4. En el lado izquierdo del sistema está el banco de análisis formado por los filtros de análisis  $H_0(\omega)$  y  $H_1(\omega)$ ; del lado derecho está el banco de síntesis formado por los filtros de síntesis  $F_0(\omega)$  y  $F_1(\omega)$ . En general para una división de  $M$  subbandas, los filtros de análisis se representan con  $H_k(\omega)$  y los de síntesis con  $F_k(\omega)$ ,  $k=0, 1, 2, \dots, M-1$ .

En un banco de filtros de análisis uniforme de  $M$  subbandas, todos los filtros pueden ser obtenidos a partir de  $H_0(\omega)$ , como lo expresa la ecuación (2.3.1). Cada filtro  $H_k(\omega)$  será una versión desplazada de  $H_0(\omega)$ . La cantidad de traslape entre las respuestas de estos filtros depende del grado del filtro  $H_0(\omega)$  que se use.

$$H_k(\omega) = H_0\left(\omega - \frac{2\pi k}{M}\right) \quad k = 0, 1, \dots, (M-1) \quad (2.3.1)$$

Para el caso de la división en octavas únicamente se utiliza una división en dos subbandas aplicada repetidamente sobre la banda baja. Si se consideran  $j$  etapas de filtrado, se obtiene al final una división de  $j+1$  subbandas.

En el caso de  $M=2$ , los filtros son de media banda y si observáramos las gráficas de sus respuestas veríamos que  $H_1(\omega)$  sería como el reflejo de  $H_0(\omega)$ , o el filtro “espejo” de  $H_0(\omega)$ , es decir sus respuestas en frecuencia son simétricas alrededor de  $\pi/2$ . De acuerdo con la ecuación (2.3.1) la relación entre  $H_0(\omega)$  y  $H_1(\omega)$  es:

$$H_1(\omega) = H_0(\omega - \pi) \quad (2.3.2)$$

La figura 2.12 muestra el esquema de la respuesta típica de los filtros espejo.

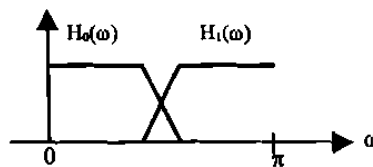


Figura 2.12. Esquema de la respuesta típica de los filtros espejo.

En términos de la respuesta a impulso, a partir de los coeficientes de un filtro de media banda pasa bajas  $h(k)$ ,  $k=0,1,2,3,\dots,N$ , se puede hacer una elección de los coeficientes de los dos filtros de análisis  $h_0(k)$  y  $h_1(k)$ :

$$h_0(k) = h(k) \quad (2.3.3)$$

$$h_1(k) = (-1)^k h(k) \quad (2.3.4)$$

Esta elección es básicamente una alternación de los signos de  $h(k)$  para obtener el filtro pasa altas  $h_1(k)$ , es una de las primeras implementaciones que se usaron en la

construcción del banco de filtros [6]. Posteriormente, para asegurar la ortogonalidad entre  $h_0(k)$  y  $h_1(k)$  se optó por invertir el orden de los coeficientes de  $h(k)$  en la ecuación (2.3.4) y después alternar los signos, con la condición de que el grado  $N$  del filtro sea impar; como se expresa en la siguiente ecuación.

$$h_1(k) = (-1)^k h(N-k) \quad (2.3.5)$$

Después de la elección de los filtros pasa bajas y pasa altas para el banco de filtros de dos canales, se debe considerar también en su diseño el efecto del aliasing. En un banco de filtros de dos canales, el aliasing se puede evitar eligiendo convenientemente los filtros de síntesis.

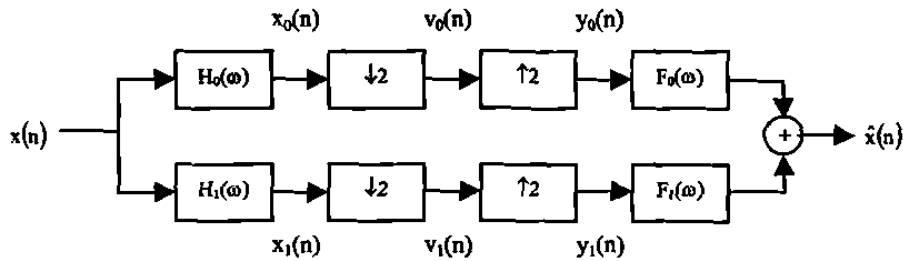


Figura 2.13. Análisis y reconstrucción de  $x(n)$ .

Siguiendo el esquema mostrado en la figura 2.13, se tiene que:

$$X_k(\omega) = H_k(\omega)X(\omega), \quad k=0,1. \quad (2.3.6)$$

Las transformadas de las señales diezmadas  $v_k(n)$ , de acuerdo con la ecuación (2.2.11), son:

$$V_k(\omega) = \frac{1}{2} \left[ X_k\left(\frac{\omega}{2}\right) + X_k\left(\frac{\omega}{2} - \pi\right) \right] \quad (2.3.7)$$

El segundo término de la ecuación anterior representa el aliasing. La transformada  $Y_k(\omega)$ , tomando en cuenta la ecuación (2.2.14), es:

$$Y_k(\omega) = V_k(2\omega) = \frac{1}{2}[X_k(\omega) + X_k(\omega - \pi)] = \frac{1}{2}[H_k(\omega)X(\omega) + H_k(\omega - \pi)X(\omega - \pi)] \quad (2.3.8)$$

La señal reconstruida es:

$$\hat{X}(\omega) = F_0(\omega)Y_0(\omega) + F_1(\omega)Y_1(\omega) \quad (2.3.9)$$

Sustituyendo  $Y_k(\omega)$  y reacomodando, se tiene:

$$\hat{X}(\omega) = \frac{1}{2}[H_0(\omega)F_0(\omega) + H_1(\omega)F_1(\omega)]X(\omega) + \frac{1}{2}[H_0(\omega - \pi)F_0(\omega) + H_1(\omega - \pi)F_1(\omega)]X(\omega - \pi) \quad (2.3.10)$$

El segundo término de la ecuación anterior representa la interferencia debida al aliasing, que puede ser disminuida o cancelada con la elección que se haga de  $F_0(\omega)$  y  $F_1(\omega)$ . Para cancelar el aliasing, se escoge:

$$F_0(\omega) = H_1(\omega - \pi) \quad (2.3.11)$$

$$F_1(\omega) = -H_0(\omega - \pi) \quad (2.3.12)$$

que, en términos de la respuesta a impulso son:

$$f_0(n) = (-1)^n h_1(n) \quad (2.3.13)$$

$$f_1(n) = -(-1)^n h_0(n) \quad (2.3.14)$$

Las ecuaciones (2.3.13) y (2.3.14) son las condiciones con las que se debe modelar el banco de filtros para evitar el aliasing, pero no aseguran la reconstrucción perfecta de la señal, que todavía puede ser afectada por la distorsión de amplitud y fase.

### 2.3.3 Distorsión de amplitud y fase.

Aplicando las condiciones establecidas en las ecuaciones (2.3.11) y (2.3.12), la ecuación (2.3.10) se convierte en:

$$\hat{X}(\omega) = \frac{1}{2} [H_0^2(\omega) - H_1^2(\omega)] X(\omega) \quad (2.3.15)$$

si se toma:

$$G(\omega) = \frac{1}{2} [H_0^2(\omega) - H_1^2(\omega)] \quad (2.3.16)$$

entonces:

$$\hat{X}(\omega) = G(\omega) X(\omega) \quad (2.3.17)$$

de esta manera  $G(\omega)$  representa la respuesta en frecuencia del banco de filtros, el cual, para que se pueda obtener una reconstrucción perfecta de  $X(\omega)$ , debe estar libre del aliasing y también de la distorsión de amplitud de fase, lo que equivale a la condición  $G(\omega) = ce^{-j\omega n_0}$ ; o sea,  $G(\omega)$  debe ser de fase lineal. Así, en ausencia de ruido de cuantificación:

$$\hat{X}(\omega) = ce^{-j\omega n_0} X(\omega) \quad \text{ó} \quad x(n) = c \cdot \hat{x}(n - n_0) \quad (2.3.18)$$

Si  $H(\omega)$  es de fase lineal,  $G(\omega)$  lo será también. Esta restricción requiere que la respuesta a impulso de  $h(k)$  sea simétrica, es decir  $h(k) = h(N - k)$ . Desarrollando  $G(\omega)$  de acuerdo con estas restricciones se puede obtener:

$$G(\omega) = \frac{1}{2} e^{-j\omega N} \left[ |H_0(\omega)|^2 - (-1)^N |H_0(\omega - \pi)|^2 \right] \quad (2.3.19)$$

Recordemos que  $N$  representa el orden del filtro y, por lo tanto, el número de coeficientes es  $N-1$ . Si  $N$  es par,  $G(\omega)$  será igual a cero en  $\omega = \pi/2$  y habrá una distorsión de amplitud muy grande en esa frecuencia, por lo que se debe elegir  $N$  impar; en consecuencia:

$$G(\omega) = \frac{1}{2} e^{-j\omega N} \left[ |H_0(\omega)|^2 + |H_1(\omega)|^2 \right] \quad (2.3.20)$$

Si la magnitud de  $G(\omega)$  es constante, se asegura la reconstrucción perfecta de  $X(\omega)$ ; por esto, es deseable que los filtros satisfagan la siguiente condición:

$$|H_0(\omega)|^2 + |H_1(\omega)|^2 = 1 \quad (2.3.21)$$

Las ecuaciones (2.3.13) y (2.3.14) junto con la (2.3.21) aseguran la reconstrucción perfecta de la señal para un banco de filtros de dos canales y también son válidas para una *descomposición en bandas* uniformes de acuerdo con el esquema de la figura 2.1. El análisis completo para una descomposición multibanda como en la figura 2.3, se puede encontrar en [12] y [15]. Para una descomposición en octavas (figura 2.2) también son aplicables las condiciones de reconstrucción perfecta obtenidas para el banco de filtros de dos canales siempre y cuando se introduzca el suficiente retraso  $\sigma$  desplazamiento en las bandas que sufren menos etapas de filtrado; pero en general si se consideran los filtros sin retraso, se puede asumir la reconstrucción perfecta.

## 2.4 ONDULETAS

### 2.4.1 *El concepto de onduleta*

En procesamiento digital de señales, uno de los temas fundamentales es el filtrado, en el cual se consideran generalmente los problemas de retraso, distorsión y recuperación de la señal. En el caso de los filtros de respuesta a impulso finita, se trata de obtener la mejor aproximación del filtro con el menor número de coeficientes y la menor distorsión posible de la señal. El número de coeficientes influye directamente en el retraso de la señal y en el número de operaciones a realizar en el proceso. Por lo que respecta a los filtros FIR, mientras menor sea el número de coeficientes las características de filtrado serán de *mucho menor calidad*, manifestándose en la distorsión de la señal. El problema quizá no sea mucho el tiempo o la velocidad de procesamiento al aumentar el grado del filtro, sino más bien el costo de tener que implementar más unidades de memoria o de retraso o un mayor número de operaciones. Los filtros con base en coeficientes de onduleta pueden salvar eficientemente ese problema, pues con un menor número de coeficientes pueden

obtener la mejor calidad de filtrado. En la figura 2.14 se muestra la diferencia de las respuestas en frecuencia entre un filtro de Fourier de veinte coeficientes diseñado por medio de ventanas y uno de coeficientes de onduleta.

Las onduletas son en realidad bases de funciones obtenidas a partir de una sola función  $\varphi(t)$  llamada función de escalamiento. El procedimiento para su obtención se puede encontrar en [7, 8, 9, 20], pero aquí interesa más su aplicación, pues han demostrado su efectividad en el procesamiento de señales no estacionarias como las señales de voz. Además son una excelente alternativa para la representación de señales en tiempo y frecuencia simultáneamente [12]. También se emplean exitosamente en el análisis de multirresolución, sobre todo en procesamiento de imágenes[1,2].

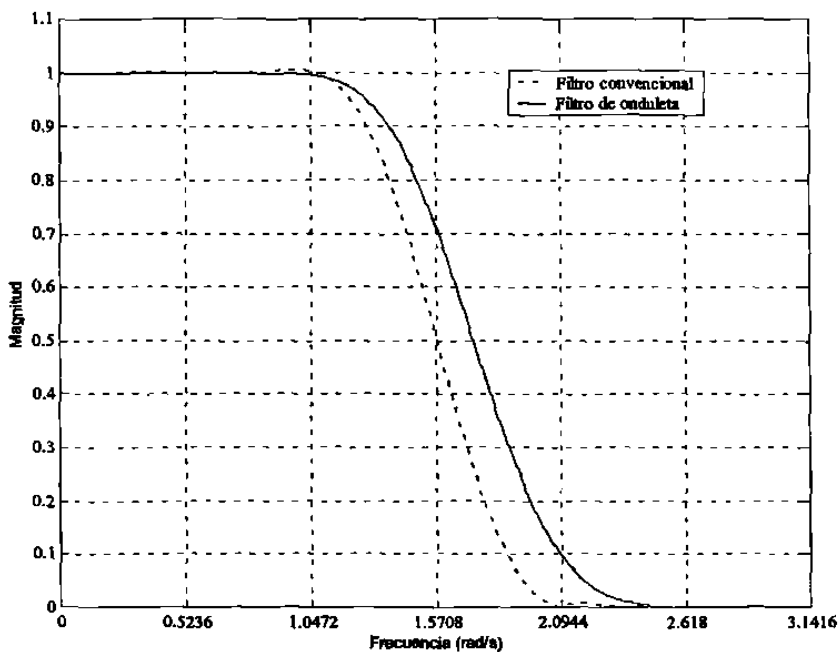


Figura 2.14. Comparación de las respuestas en frecuencia de un filtro FIR de 20 coeficientes y otro obtenido de los coeficientes de la onduleta DB10.<sup>1</sup>

<sup>1</sup> De la familia de onduletas de Daubechies. La información se puede encontrar en [10]



Como se mencionó, las onduletas son bases de funciones en tiempo continuo  $\psi_{j,k}(t)$ , lo que significa que cualquier función  $x(t)$  puede ser representada por:

$$x(t) = \sum_{j,k} b_{j,k} \psi_{j,k}(t) \quad (2.4.1)$$

Donde  $b_{j,k}$  representa los coeficientes de la expansión en onduletas.

Una onduleta típica es comprimida o adelgazada  $j$  veces y desplazada  $k$  unidades de tiempo. Su fórmula es:

$$\psi_{j,k}(t) = \psi(2^j t - k) \quad (2.4.2)$$

Las ecuaciones de escalamiento y de onduleta, respectivamente:

$$\varphi(t) = \sum \sqrt{2}c(k) \varphi(2t - k) \quad (2.4.3)$$

$$\psi(t) = \sum \sqrt{2}d(k) \varphi(2t - k) \quad (2.4.4)$$

son la conexión crucial entre onduletas y filtros. Históricamente, su desarrollo fue separado pero ahora pueden verse juntos. Un filtro pasa bajas  $c(k)$  determina la función de escalamiento  $\varphi(t)$ , mediante la solución de (2.4.3). Después los coeficientes del pasa altas  $d(k)$  producen las onduletas según la ecuación (2.4.4).

#### 2.4.2. La transformada discreta de onduleta

Se puede probar [12] que cualquier función *continua* puede ser representada por la siguiente expansión, definida en términos de una escala dada y sus onduletas derivadas.

$$x(t) = \sum_{k=-\infty}^{\infty} c_{j_0}(k) \varphi_{j_0,k}(t) + \sum_{j=j_0}^{\infty} \sum_{k=-\infty}^{\infty} d_j(k) \psi_{j,k}(t) \quad (2.4.5)$$

En esta expansión, la primera suma da una función que es una aproximación burda de la función  $x(t)$  a una escala  $j$ . Para cada incremento de  $j$  en la segunda suma es agregada

una resolución más fina que incrementa los detalles. En la ecuación (2.4.5)  $c_{j_0}(k)$  y  $d_j(k)$  representan los coeficientes de la expansión en onduletas.

A partir de la elección de  $j_0$  se obtiene la escala más burda, cuyo espacio es completado por  $\phi_{j_0,k}(t)$ . El resto de la función es expandida por las onduletas produciendo los detalles de alta resolución de la función. El conjunto de coeficientes  $c_{j_0}(k)$  y  $d_j(k)$  en la expansión de onduleta representada por la ecuación (2.4.5) es llamado transformada discreta de onduleta [5, 12] (*DWT, Discrete Wavelet Transform*) de la función  $x(t)$ . Estos coeficientes de onduleta, bajo ciertas condiciones, pueden describir completamente a la función original, y de una forma similar a los coeficientes de las series de Fourier, pueden ser usados para análisis, descripción, aproximación y filtrado. Si la función de escalamiento está bien diseñada, entonces, a una escala alta, las muestras de la señal están muy cerca de los coeficientes de escalamiento.

Para trabajar directamente con los coeficientes de onduleta, se presenta la relación entre los coeficientes de expansión a una escala dada en términos de aquellos a una escala más alta. Esta relación es especialmente práctica si tomamos en cuenta que la señal original es usualmente desconocida y solamente está disponible una versión muestreada de la misma a una resolución dada. Como se mencionó antes, para un buen diseño de las funciones de escalamiento o de onduleta, las muestras de la señal discretas pueden ser lo más aproximados posible a los coeficientes de escalamiento. Los coeficientes de escalamiento y de onduleta a una escala  $j$  están relacionados con los coeficientes de escalamiento de la escala  $j+1$  por medio de las siguientes dos relaciones.

$$c_j(k) = \sum_m h_0(m - 2k) c_{j+1}(m) \quad (2.4.6)$$

$$d_j(k) = \sum_m h_1(m - 2k) c_{j+1}(m) \quad (2.4.7)$$

En las ecuaciones (2.4.6) y (2.4.7)  $c_{j+1}(k)$  representa la señal que se desea descomponer en una resolución más baja  $c_j(k)$  y sus detalles  $d_j(k)$ . Respectivamente,  $c_j(k)$  y  $d_j(k)$ , son las bandas de baja y alta frecuencia de  $c_{j+1}(k)$ .

Las ecuaciones (2.4.6) y (2.4.7) establecen que los coeficientes de escalamiento a la escala más alta, junto con los filtros de escalamiento y onduleta  $h_0(k)$  y  $h_1(k)$ , respectivamente, pueden ser usados para calcular los coeficientes de la DWT a escalas más bajas. En la práctica se asume que los valores de una señal discreta en su resolución original corresponden a los coeficientes de escalamiento. Para un sistema dado de onduletas, con filtros de onduleta conocidos  $h_0(k)$  y  $h_1(k)$ , es posible usar (2.4.6) y (2.4.7) reiteradamente para calcular los coeficientes de la DWT a todas las escalas más bajas deseadas.

Mediante el procesamiento de señales multitasas [13,15,16], es posible calcular las dos sumatorias  $c_j(k)$  y  $d_j(k)$  usando dos filtros FIR de análisis diseñados de acuerdo con las ecuaciones (2.3.13) y (2.3.14). Las salidas de esos filtros son calculadas sólo para los índices pares y los filtros son usados con esos índices negativos. Esas diferencias pueden ser incorporadas dentro de la operación de filtrado diezmado la salida del filtro e invirtiendo el orden de los coeficientes del filtro. Los cálculos se siguen realizando hasta que se obtengan  $c_{j_0}(k)$  y  $d_{j_0}(k)$ . El conjunto de esos coeficientes forman la DWT de la señal original  $x(k)$ ; el número de coeficientes será igual al número de puntos en la señal discreta original debido al diezmado de dos a uno que se realiza en cada etapa del proceso

Debido a que el par de filtros usado en el cálculo de la DWT son filtros pasa bajas y pasa altas complementarios el resultado final nos dará secuencias que provienen de diferentes bandas de frecuencia de la señal original. La DWT divide el ancho de banda de la señal en una manera logarítmica. Los filtros son conocidos como filtros de análisis en la literatura referente a bancos de filtros y también en la de onduletas.

Con referencia a la ecuación (2.4.5), en la cual la señal original continua es escrita en su forma de expansión de onduletas, pueden ser derivadas relaciones matemáticas para los filtros de síntesis. Los coeficientes de más alta resolución están relacionados con los coeficientes de resolución más baja y los coeficientes de onduleta por medio de la siguiente ecuación.

$$c_{j+1}(k) = \sum_m c_j(m)h_0(k-2m) + \sum_m d_j(k)h_1(k-2m) \quad (2.4.8)$$

Esta ecuación indica cómo secuencias de DWT a una resolución  $j$  pueden ser usadas de forma reiterada para reconstruir los coeficientes de escalamiento a la resolución más alta disponible.

Recordemos que el banco de filtros de análisis calcula eficientemente la DWT usando bancos de diezmado de filtros digitales. Similarmente, el banco de síntesis calcula eficientemente la DWT inversa reconstruyendo la señal original discreta usando filtros digitales de interpolación.

En resumen, la DWT para una función discreta  $x(k)$  es obtenida a través de las siguientes relaciones, considerando  $c_j(k) = x(k)$  y  $j = J, J-1, \dots, j_0+1$ .

$$c_{j-1}(k) = \sum_m h_0(m-2k)c_j(m) \quad (2.4.9)$$

$$d_{j-1}(k) = \sum_m h_1(m-2k)c_j(m) \quad (2.4.10)$$

De donde se obtiene el conjunto de los coeficientes de la transformada:

$$\{d_{J-1}(k), d_{J-2}(k), \dots, d_{j_0+1}(k), d_{j_0}(k), c_{j_0}(k)\}$$

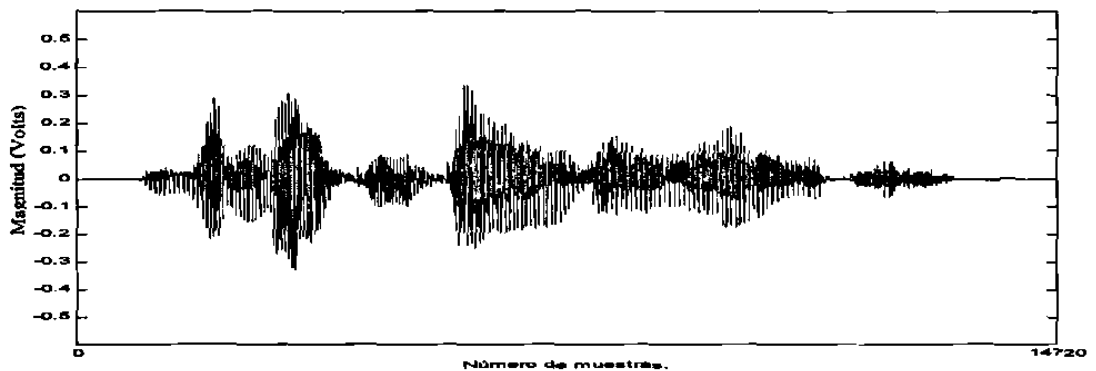
Si  $x(k) = c_j(k)$  y  $j = j_0, j_0+1, \dots, J-1$ ; la IDWT se obtiene a partir de la siguiente ecuación.

$$c_{j+1}(k) = \sum_m c_j(m)f_0(k-2m) + \sum_m d_j(k)f_1(k-2m) \quad (2.4.11)$$

Las señales de voz requieren que su procesamiento sea en tiempo corto (20 ms, o 160 muestras para una señal muestreada a 8 kHz). Aquí es donde el análisis con filtros de onduleta tiene ventaja sobre el análisis con la transformada de Fourier en tiempo corto sobre la recuperación completa de la señal, pues elimina la necesidad de la ventana deslizante y la distorsión de la señal casi no existe.

En la figura 2.15 se ilustra la aplicación de la DWT para la descomposición de una señal de voz muestreada a 8 kHz en cuatro subbandas: 0–500, 500–1000, 1000–2000 y 2000–4000 Hz. Para la obtención de la DWT se usaron los coeficientes de onduleta DB5.

Señal original



Coefficientes  $c_0$

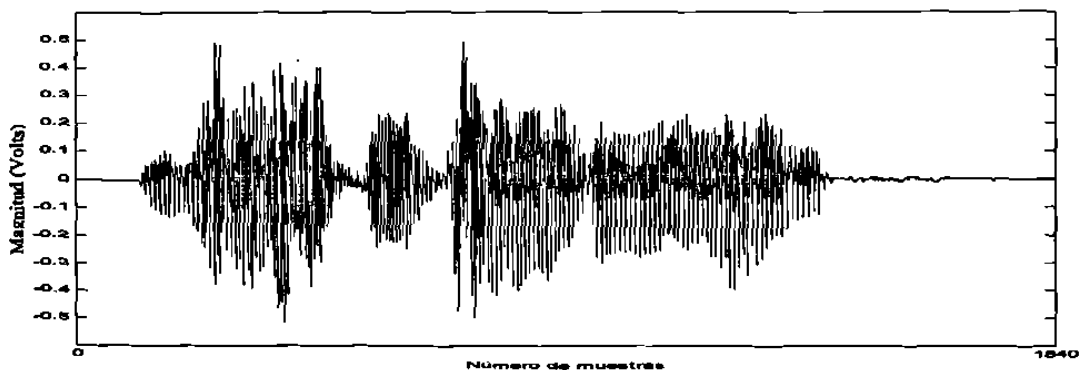


Figura 2.15. Descomposición de una señal de voz en cuatro bandas usando la DWT.

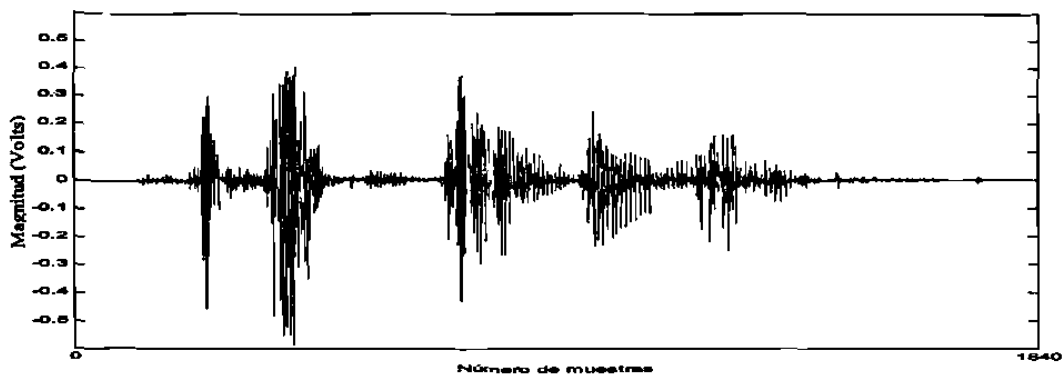
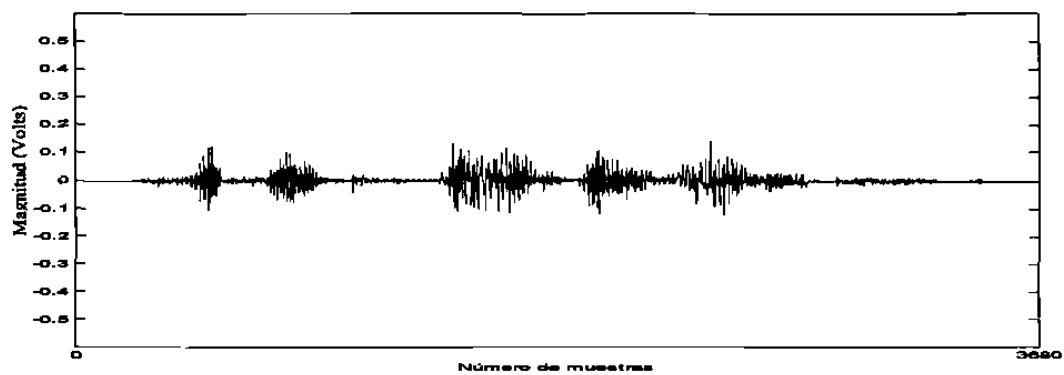
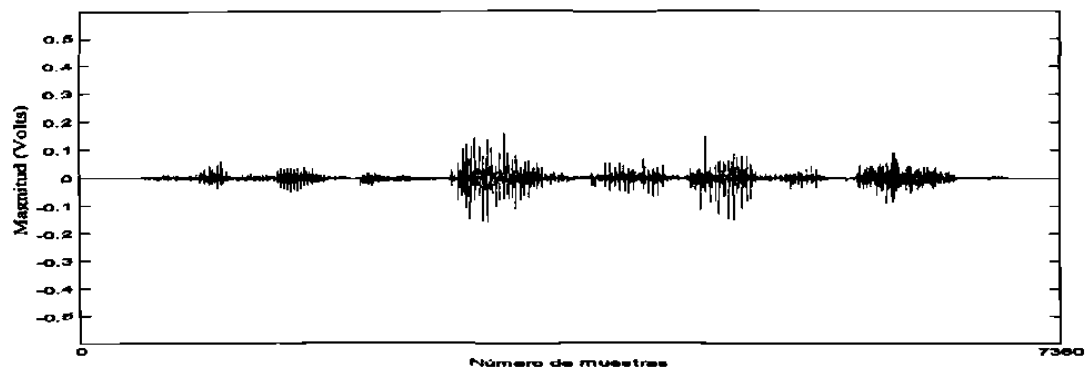
**Coefficientes d0****Coefficientes d1****Coefficientes d2**

Figura 2.15. (Continuación).

Si se observa por segmentos la descomposición de la señal mostrada en la figura 2.15, veremos que algunos de ellos sólo tienen significancia para alguna banda en particular debido a que su magnitud es casi nula en las demás; de esta forma se puede enfocar el trabajo de la codificación a las bandas que sean más necesarias asignándoles un mayor número de bits por muestra, como se verá en el capítulo siguiente.

# CODIFICACIÓN

## 3.1. CUANTIFICACIÓN

### *3.1.1 Características de la cuantificación*

En un enlace de comunicaciones digitales, la cuantificación de la amplitud de la señal determina la mayor parte de la distorsión total así como también la tasa de bits necesaria para la transmisión de la señal hacia el receptor. El procedimiento consiste en transformar la amplitud  $x(n)$  de una señal dada en otra amplitud  $y(n)$  tomada de un conjunto finito de amplitudes posibles. Es un proceso no lineal y no invertible. La función que realiza dicha transformación es, por definición, una función escalonada y se expresa por medio de la relación mostrada en la ecuación (3.1.1).

$$y = Q(x) \tag{3.1.1}$$

La cuantificación puede ser uniforme o no uniforme [10, 11, 14], y, dependiendo de que se asigne o no el cero a un nivel, también puede ser de elevación centrada (ver figura 3.1) o de escalón centrado (figura 3.1) [14]. La cuantificación uniforme es un requisito de



procesamiento de la señal en un sistema digital; sin embargo también se usan cuantificadores no lineales para aplicaciones de transmisión y almacenamiento de señales tales como las de voz.

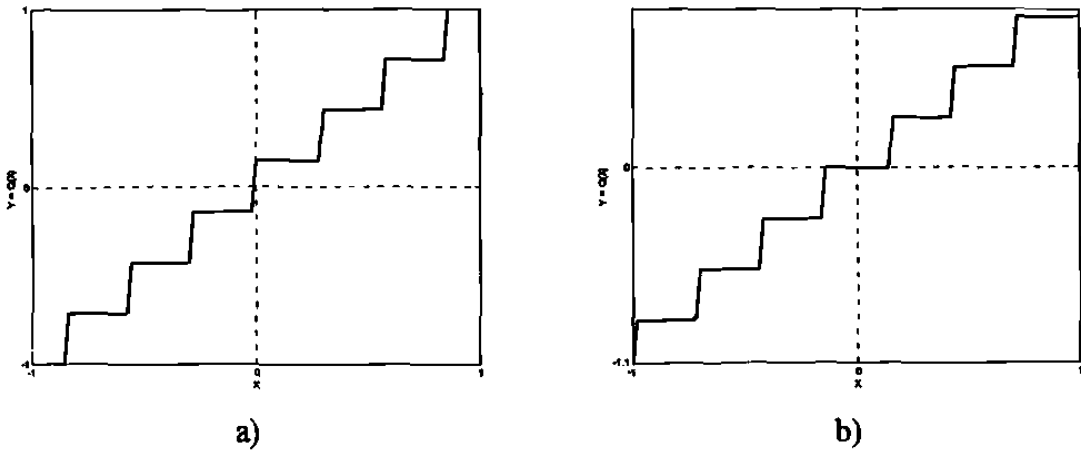


Figura 3.1. Cuantificación uniforme de elevación centrada (a), y de escalón centrado (b).

En la mayoría de las operaciones de procesamiento de señales, la función  $Q$  es independiente de  $n$  (cuantificación sin memoria [14]) y generalmente se usan cuantificadores uniformes definidos por:

$$x_{k+1} - x_k = \Delta \quad (3.1.2)$$

Donde  $\Delta$  es el tamaño del escalón de cuantificación,  $k = 1, 2, \dots, L$  es el nivel de cuantificación y  $L$  representa el número de niveles.

El error que existe entre la entrada  $x$  y la salida y es llamado error de cuantificación, y se define como la diferencia entre  $x$  e  $y$ .

$$q = x - y = x - Q(x) \quad (3.1.3)$$

En la cuantificación uniforme el error de cuantificación  $q$  está siempre en el rango  $-\Delta/2$  a  $\Delta/2$ :

$$-\frac{\Delta}{2} < q \leq \frac{\Delta}{2} \quad (3.1.4)$$

### 3.1.2 Ruido de cuantificación y tasa de bits.

Para determinar los efectos de la cuantificación se adopta una aproximación estadística; se supone que el error de cuantificación es aleatorio y se modela como un ruido que se añade a la señal original [14]. Si la señal original está dentro del rango del cuantificador, el error de cuantificación está limitado en magnitud de acuerdo con la ecuación (3.1.4).

La cantidad más importante para la comparación de los desempeños de los cuantificadores es la varianza  $\sigma_q^2$  del error de cuantificación (potencia del ruido):

$$\sigma_q^2 = \sum_{k=l}^L \int_{x_k}^{x_{k+1}} (x - y_k)^2 p_x(x) dx \quad (3.1.5)$$

donde  $p_x(x)$  es la función de densidad de probabilidad de  $x$ .

Es usual relacionar la varianza  $\sigma_q^2$  del error con la varianza  $\sigma_x^2$  de la señal de entrada como lo expresa la ecuación (3.1.6). Esta relación es conocida como factor de desempeño del cuantificador y representa la varianza del error de cuantificación por unidad de varianza de la señal original.

$$\epsilon_q^2 = \frac{\sigma_q^2}{\sigma_x^2} \quad (3.1.6)$$

El recíproco de  $\epsilon_q^2$  es la relación señal a ruido de cuantificación SNR (en decibeles):

$$\text{SNR} = 10 \log_{10} \left( \frac{\sigma_x^2}{\sigma_q^2} \right) = 10 \log_{10} \left( \frac{1}{\epsilon_q^2} \right) \quad (3.1.7)$$

En algunos casos es de utilidad representar el desempeño del cuantificador como en la ecuación (3.1.8). La constante  $C$  es generalmente del orden de 1 a 10.

$$\sigma_q^2 = c \cdot 2^{-2b} \sigma_x^2 \quad (3.1.8)$$

Con la notación de la ecuación anterior se puede obtener una expresión general para calcular el número  $b$  de bits por muestra necesarios para una SNR dada:

$$b = \frac{1}{2} \log_2 \left( \frac{\sigma_x^2}{\sigma_q^2} \right) + \frac{1}{2} \log_2 c \quad (3.1.9)$$

Para un valor dado de  $\sigma_q^2$ , el número mínimo necesario de bits por muestra está dado por la teoría de distorsión [14], y es de la forma:

$$\min\{b\} = \frac{1}{2} \log_2 \left( \frac{\sigma_x^2}{\sigma_q^2} \right) - \alpha_b \quad (3.1.10)$$

Donde  $\alpha_b = 0$ , para una fuente de ruido gaussiano equiprobable, como es el caso para el desarrollo de las ecuaciones de este capítulo.

## 3.2. LA CODIFICACIÓN PCM

### 3.2.1 Características del sistema PCM

La codificación por modulación de pulsos PCM es el sistema de digitalización de mayor consumo de bits. Fue también el primero en desarrollarse como método de codificación, y aún hoy es ampliamente usado, sobre todo en telefonía. Es el sistema de codificación que ha sido más estudiado y también el más simple conceptualmente. También es ampliamente aceptado como un estándar de comparación contra otros desarrollos para la

digitalización de forma de onda. Por otra parte, casi la mayoría de los codificadores involucran pasos de codificación o decodificación PCM.

En pocas palabras, un codificador PCM no es otra cosa que un muestreo de forma de onda seguido por una cuantificación de amplitud. La frecuencia de muestreo estándar para un canal de voz en PCM es de 8 kHz y generalmente la señal original se limita a una banda de 3200 Hz, para evitar los efectos del aliasing. Si el número de bits por muestra que se utilizan para la codificación es  $b$ , la tasa de transmisión o *bit rate*  $R$  de un sistema PCM, o de cualquier otro, esta dada por medio de la ecuación (1.1.1) o, de forma más simple, por medio de la (3.2.1), en donde  $f_m$  representa la frecuencia de muestreo.

$$R = f_m b \quad (3.2.1)$$

La relación señal a ruido SNR de un sistema PCM para aplicaciones de voz se relaciona directamente con el número de bits por muestra por medio de la ecuación (3.2.2), siempre y cuando la señal haya sido cuantificada en su rango dinámico completo:

$$\text{SNR}_{\text{PCM}}|_{\text{dB}} = 6.02b \quad (3.2.2)$$

### 3.2.2 Cuantificación logarítmica

Las señales de voz tienen la característica de que las amplitudes pequeñas de la señal ocurren más frecuentemente que las grandes. Un cuantificador uniforme, en la codificación PCM, asignará el mismo espacio entre niveles sucesivos a través del rango dinámico completo de la señal. Un mejor desarrollo en este sistema, emplea una *cuantificación no uniforme* [10, 11, 14]; ésta se realiza haciendo pasar la señal por un dispositivo no lineal y después por un cuantificador uniforme. Este tipo de cuantificación en el estándar PCM es un cuantificador logarítmico cuyas características se encuentran descritas por medio de la ecuación (3.2.3), con la entrada  $x$  y la salida  $y$  normalizadas;  $\text{sgn}(x)$  es el signo de  $x$ .

$$y = \text{sgn}(x) \frac{\log(1 + \mu|x|)}{\log(1 + \mu)} \quad (3.2.3)$$

El parámetro  $\mu$  es seleccionado para obtener la característica de compresión deseada; en el sistema norteamericano  $\mu = 255$  y su característica se conoce como la ley- $\mu$ . En Europa existe un estándar similar conocido como la ley-A . La elección de  $\mu = 255$  da como resultado una reducción de cerca de 24 dB en la potencia del ruido de cuantificación con referencia a la cuantificación uniforme. En la reconstrucción de la señal a partir de los valores cuantificados es usada la relación logarítmica inversa de la ecuación (3.2.3) [14].

A una frecuencia de muestreo de 8 kHz con una representación de 8 bits por muestra y utilizando cuantificación logarítmica, la tasa de transmisión resultante en la codificación de voz es de 64 kbits/s (estándar). Usando cuantificación uniforme se requiere una representación de 12 bits por muestra, aproximadamente, para obtener el mismo nivel de fidelidad, lo que significa una tasa de transmisión igual a 96 kbits/s.

Los sistemas comerciales de PCM usan la tasa de transmisión de 64 kbits/s, el cual es muy importante en la transmisión digital ya que representa una unidad estándar para la tecnología digital. Por ejemplo, los sistemas digitales de audio son diseñados para operar a tasas que son múltiplos o submúltiplos de 64 kbits/s. En la red digital de servicios integrados ISDN (*Integrated Services Digital Networks*) pueden ser asignadas tasas de canal múltiplos o submúltiplos de 64 kbits/s no sólo para telefonía, sino también para la transmisión digital de otras señales tales como datos, gráficos o vídeo.

### 3.3 CODIFICACIÓN DE LAS SUB-BANDAS

#### 3.3.1 *Objetivo de la codificación de subbandas*

Como se mencionó anteriormente el propósito de la codificación en subbandas es dividir la señal de entrada en varios componentes de frecuencia y después codificar cada

uno de esos componentes de manera separada. Esta división elimina la redundancia de la señal y proporciona al canal un conjunto de entradas sin correlación. A su vez, tiene la ventaja de que el número de bits usados para codificar cada banda de frecuencia puede ser diferente y también variable; de esta manera la precisión de la codificación puede ser ubicada donde sea más necesaria. La banda total de la señal de voz típicamente es dividida en dos o más subbandas por medio del banco de filtros, que es la parte más compleja del codificador.

Al obtener la división en subbandas, las frecuencias de muestreo de cada una de ellas son reducidas de acuerdo con el número de divisiones. Por ejemplo para una división uniforme en cuatro bandas de una señal muestreada a 8 kHz, las frecuencias de muestreo resultantes para cada banda serán de 2 kHz; en la división en octavas, las frecuencias de muestreo resultantes para cada una de las cuatro subbandas serían 1, 1, 2 y 4 kHz, de la banda más baja a la más alta.

Por medio de una apropiada asignación de bits en las diferentes bandas, el número de niveles de cuantificación y la varianza del error de reconstrucción pueden ser controlados de forma separada para cada una de las bandas; además el error total de reconstrucción puede también ser controlado como una función de la frecuencia. En las bandas de frecuencia más bajas, donde el tono y la estructura formante deben ser conservados con precisión, puede ser usado un número de bits por muestra más grande, mientras que en las bandas de frecuencia superiores, donde ocurren sonidos no vocalizados similares al ruido, pueden usarse menos bits.

### ***3.3.2 Tasa de transmisión y distorsión***

En un codificador de  $M$  subbandas, cada componente  $x_k(n)$  resulta con una nueva tasa de muestreo  $f_k$  y es codificada usando  $b_k$  bits por muestra en cada subbanda. La tasa de transmisión  $R$  es la suma de las tasas de transmisión  $R_k$  necesarias para codificar cada banda individualmente, por lo tanto:

$$R = \sum_{k=1}^M R_k = \sum_{k=1}^M f_k b_k \quad \text{bits / s} \quad (3.3.1)$$

En el caso especial de bandas uniformes  $f_k = f_m/M$ ; si  $f_m$  representa la frecuencia de muestreo de la señal original. Así, la ecuación (3.3.1) se puede simplificar, para bandas uniformes, como:

$$R = \frac{f_m}{M} \sum_{k=1}^M b_k \quad (3.3.2)$$

La ecuación (3.3.2) se puede reducir a la (3.2.1), si tomamos  $b$  como el número de bits por muestra promedio que se usarían para codificar la señal en banda completa:

$$b = \frac{1}{M} \sum_{k=1}^M b_k \quad (3.3.3)$$

La distorsión  $D$  introducida por la cuantificación se mide directamente mediante la varianza del ruido de cuantificación [12] mostrada en la ecuación (3.1.5):

$$D = \sigma_q^2 \quad (3.3.4)$$

Del mismo modo que para las tasas de transmisión de las subbandas, la distorsión total es la suma de las distorsiones de las subbandas como lo expresa la ecuación (3.3.5). El problema de cómo elegir los cuantificadores de cada una de las subbandas es el recurso que se utiliza para el problema de asignación del número de bits por muestra en las mismas. Se trata de minimizar una función que describa la calidad de la aproximación bajo una restricción dada, o sea, el número de bits que pueden ser usados para codificar la señal y, por lo tanto, la tasa de transmisión. Así, el problema de la asignación de bits consiste en minimizar:

$$D = \sum_{k=1}^M D_k \quad (3.3.5)$$

de manera que se satisfaga la ecuación (3.3.6). Este problema también puede plantearse en dirección inversa, es decir, minimizar la tasa de bits dada una distorsión máxima.

$$\sum_{k=1}^M R_k \leq R \quad (3.3.6)$$

Para obtener una solución óptima del problema se debe considerar tanto la tasa de bits como la distorsión introducida por los cuantificadores en cada banda. La figura 3.2 muestra esquemáticamente una curva de la relación entre la tasa bits y la distorsión; la solución óptima a la asignación de bits estaría representada por medio de los puntos en los cuales la pendiente es constante para las curvas tasa-distorsión de cada una de las sub-bandas [12, 14].

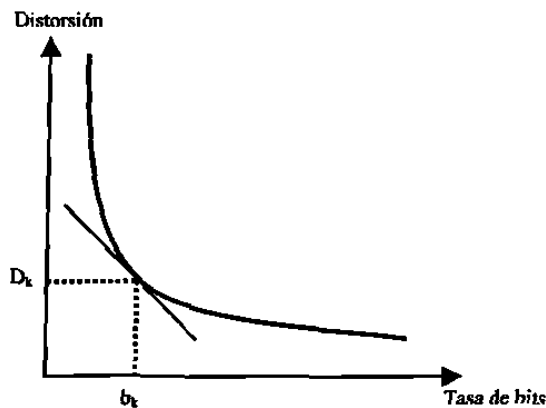


Figura 3.2. Distorsión y tasa de bits.

En resumen, el problema de asignación de bits consiste en lo siguiente: para cada canal  $k$ , se debe elegir de un conjunto de cuantificadores  $\{Q_k\}$ . Eligiendo un cuantificador dado  $Q_k$ , se producirá una distorsión  $D_k$  y una tasa de bits  $b_k$  para el canal  $k$ ; se debe encontrar la combinación de cuantificadores en los distintos canales que produzcan la mínima distorsión total.

En la figura 3.2 se muestra un esquema de codificación de bandas uniformes que incluye los cuantificadores. La distorsión de cada banda a la salida de los cuantificadores



está dada por la varianza del error de cuantificación; la distorsión que hay que minimizar es la suma de las distorsiones en la salida de los filtros de síntesis.

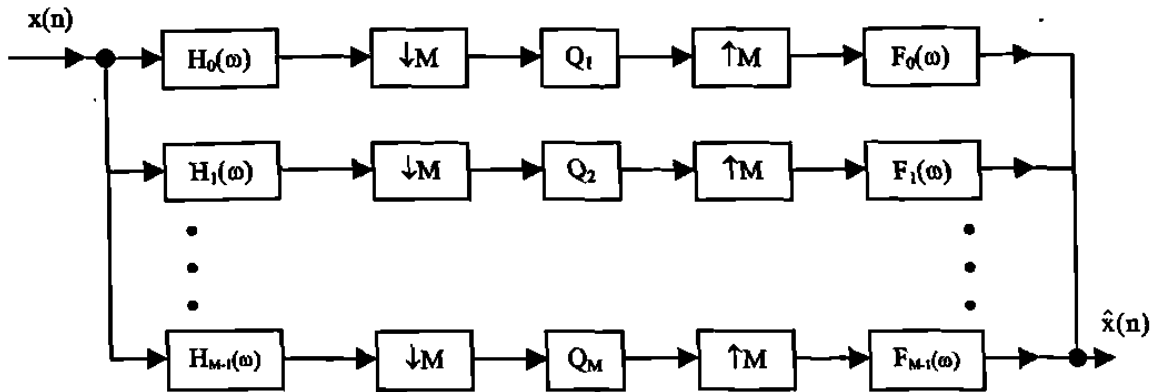


Figura 3.3. Cuantificadores en un sistema de M bandas uniformes

### 3.3.3 Asignación de la tasa de bits por muestra

De acuerdo con la ecuación (3.1.8), para un número dado de bits  $b$ , la varianza del error de cuantificación es proporcional a la varianza de la señal y al factor  $2^{-2b}$ . La distorsión en cada banda después de la cuantificación en la figura 3.2 está dada por la varianza del error de cuantificación. Se puede probar [15] que después de haber pasado por los filtros de síntesis, la distorsión de cada banda  $D_k$  es igual a la varianza del error de cada banda,  $\sigma_{qk}^2$ , entre el número de bandas del sistemas si se considera una división en bandas uniformes; por lo que, de acuerdo con la ecuación (3.3.5), la distorsión total de la reconstrucción es:

$$D = \frac{1}{M} \sum_{k=1}^M \sigma_{qk}^2 \quad (3.3.7)$$

La ecuación (3.3.3) representa la media aritmética de la tasa de bits por muestra de todas las bandas tomando en cuenta la contribución de cada banda. De la misma manera, la ecuación (3.3.7) representaría también la media aritmética de las varianzas de del error de

cuantificación de todas las subbandas. Tomando en cuenta que la media aritmética siempre es mayor o igual que la media geométrica, se puede establecer:

$$D \geq \left( \prod_{k=1}^M \sigma_{qk}^2 \right)^{\frac{1}{M}} \quad (3.3.8)$$

El lado derecho de la desigualdad (3.3.8) es la media geométrica de las varianzas del error de las subbandas. De acuerdo con la ecuación (3.1.8) tendríamos:

$$D \geq c \left( \prod_{k=1}^M 2^{-2b_k} \sigma_{xk}^2 \right)^{\frac{1}{M}} = c \left( 2^{-2 \sum \frac{b_k}{M}} \right) \left( \prod_{k=1}^M \sigma_{xk}^2 \right)^{\frac{1}{M}} \quad (3.3.9)$$

En la ecuación (3.3.9),  $\sigma_{xk}^2$  representa la varianza de la señal en cada subbanda. Se asume que la constante  $c$  es de la misma magnitud para todos los cuantificadores; esto es válido si las entradas de los cuantificadores tienen la misma distribución estadística. Usando la ecuación (3.3.3) se obtiene:

$$D \geq c \cdot 2^{-2b} \left( \prod_{k=1}^M \sigma_{xk}^2 \right)^{\frac{1}{M}} \quad (3.3.10)$$

La distorsión mínima que puede ser obtenida, como se ve en la desigualdad (3.3.10), dependerá de la señal de entrada  $x(n)$  y del número promedio de bits  $b$ . Esto sólo puede conseguirse si la desigualdad (3.3.8) planteada inicialmente se convierte en una igualdad, lo cual se lograría si  $\sigma_{qk}^2$  fuera la misma para todas las bandas, o sea:

$$\sigma_{qk}^2 = \sigma_q^2 = D \quad k = 1, 2, \dots, M \quad (3.3.11)$$

Entonces, a partir de la ecuación (3.1.8) se puede obtener que:

$$2^{2b_k} = c \frac{\sigma_{xk}^2}{\sigma_q^2} \quad (3.3.12)$$

Y, tomando la igualdad de (3.3.10), obtenemos:

$$2^{2b_k} = \frac{\sigma_{xk}^2}{2^{-2b} \left( \prod_{i=1}^M \sigma_{xi}^2 \right)^{\frac{1}{M}}} \quad (3.3.13)$$

Finalmente, despejando  $b_k$  se puede obtener:

$$b_k = b + \frac{1}{2} \cdot \log_2 \frac{\sigma_{xk}^2}{\left( \prod_{i=1}^M \sigma_{xi}^2 \right)^{\frac{1}{M}}} \quad (3.3.14)$$

En la tabla 3.1 se muestra la asignación de bits para una descomposición uniforme en cuatro subbandas de acuerdo con el número de bits por muestra promedio; también se dan las tasas de transmisión;  $SNR_S$  es la relación señal a ruido de la codificación en subbandas y  $SNR_C$  es la relación señal a ruido de la codificación en banda completa. La cuantificación es uniforme. Se puede observar la ventaja contra la codificación en banda completa con una ganancia de la codificación en subbandas de cerca de 8 dB, aproximadamente.

**Tabla 3.1.** Asignación de la tasa de bits para una descomposición en cuatro bandas uniformes y comparación con la codificación en banda completa.

<b>b</b>	<b><math>b_k</math></b>	<b><math>R_k</math> (kbits/s)</b>	<b>R (kbits/s)</b>	<b><math>SNR_S</math> (dB)</b>	<b><math>SNR_C</math> (dB)</b>
8	{12, 7, 6, 7}	{24, 14, 12, 14}	64	53.9	45.4
4	{8, 3, 2, 3}	{16, 6, 4, 6}	32	27.0	20.6
3	{7, 2, 1, 2}	{14, 4, 2, 4}	24	22.8	14.1

### 3.3.4 Tasa óptima de bits por muestra para bandas octavas

De la misma manera en que se obtuvo la tasa de bits óptima para una división en bandas uniformes, se puede obtener también la ecuación para la asignación de bits para una descomposición en bandas no uniformes. Combinando las relaciones mostradas en (3.2.1) y (3.3.1), la expresión que representa la tasa de bits por muestra promedio para una división en bandas no uniformes y, por lo tanto, también para bandas octavas es:

$$b = \frac{1}{f_m} \sum_{k=1}^M f_k b_k = \sum_{k=1}^M \frac{1}{m_k} b_k \quad (3.3.15)$$

En la ecuación (3.3.15) la tasa de bits por muestra promedio depende de las frecuencias de muestreo de cada subbanda o, lo que es lo mismo, del factor de diezmado  $m_k$  empleado en la obtención de cada subbanda, que es igual a  $f_m/f_k$ . Para el caso de bandas uniformes, también se puede decir que la distorsión expresada en la ecuación (3.3.7) depende del factor de diezmación de cada subbanda  $M$ ; entonces, para el caso de bandas no uniformes podemos expresar la distorsión como:

$$D = \frac{1}{f_m} \sum_{k=1}^M f_k \sigma_{qk}^2 \quad (3.3.16)$$

La ecuación (3.3.16) representa la media aritmética de los errores de las subbandas obtenida tomando el cuenta el número de muestras de cada subbanda; así, la desigualdad establecida en (3.3.8), puede replantearse nuevamente modificando la media geométrica para que dependa también de los mismos parámetros:

$$D \geq \left( \prod_{k=1}^M (\sigma_{qk}^2)^{f_k} \right)^{\frac{1}{f_m}} \quad (3.3.17)$$

Siguiendo el mismo procedimiento usado para obtener la ecuación (3.3.14), la ecuación para la asignación óptima de la tasa de bits en la división en subbandas no uniformes es:

$$b_k = b + \frac{1}{2} \cdot \log_2 \frac{\sigma_{xk}^2}{\left( \prod_{i=1}^M (\sigma_{xi}^2)^{f_i} \right)^{\frac{1}{f_m}}} \quad (3.3.18)$$

Para bandas octavas, la relación  $f_m/f_k$  puede ser simplificada y representarla directamente por medio de potencias de 2.

La tabla 3.2 muestra la asignación de bits para una división en tres subbandas octavas: 0-1000, 1000-2000, 2000-4000 Hz, que equivale a dos etapas de filtrado de señal al igual que la división en cuatro bandas uniformes del inciso anterior.

Tabla 3.2. Asignación de la tasa de bits para una descomposición en tres bandas octavas y comparación con la codificación en banda completa.

<b>b</b>	<b>b<sub>k</sub></b>	<b>R<sub>k</sub> (kbits/s)</b>	<b>R (kbits/s)</b>	<b>SNR<sub>s</sub> (dB)</b>	<b>SNR<sub>c</sub> (dB)</b>
8	{11, 7, 7}	{22, 14, 28}	64	54.1	45.4
4	{7, 3, 3}	{14, 6, 12}	32	33.2	20.6
3	{6, 2, 2}	{12, 4, 8}	24	23.4	14.1

Se puede observar una ligera ventaja sobre la codificación en bandas uniformes y también se ve claramente como se le asigna mayor número de bits a la banda más baja.

Obviamente, de acuerdo con lo mostrado en la tabla 3.1 y en la 3.2, la distorsión producida por un sistema de codificación en subbandas es menor que la distorsión de un sistema PCM (con cuantificación uniforme). Eso ya nos da la seguridad de poder reducir la tasa de bits de la señal de voz esperando conservar la calidad del mensaje. La ganancia  $G_{CSB}$ , en términos de la distorsión, de la codificación en subbandas sobre la codificación PCM es la distorsión de la codificación PCM (ecuación 3.3.4) entre la distorsión de la codificación en subbandas (ecuación 3.3.17); tomando en cuenta las relaciones (3.1.8) y (3.3.16) la ganancia puede expresarse como:

$$G_{CSB} = \frac{\sigma_q^2}{\left( \prod_{k=1}^M (\sigma_{qk}^2)^{f_k} \right)^{\frac{1}{f_m}}} = \frac{\sigma_x^2}{\left( \prod_{k=1}^M (\sigma_{xk}^2)^{f_k} \right)^{\frac{1}{f_m}}} = \frac{\frac{1}{f_m} \sum_{k=1}^M f_k \sigma_{xk}^2}{\left( \prod_{k=1}^M (\sigma_{xk}^2)^{f_k} \right)^{\frac{1}{f_m}}} \quad (3.3.19)$$

Como se ve en la ecuación (3.3.19) la ganancia de la codificación en subbandas  $G_{CSB}$  puede interpretarse como la razón de la media aritmética a la media geométrica de las varianzas de las subbandas, y ya que la media aritmética siempre es mayor o igual que la media geométrica, entonces la ganancia siempre será mayor o igual a 1. Esta ganancia indica cuántas veces es mejor la codificación en subbandas sobre la codificación en banda completa en términos de la distorsión.

En el capítulo siguiente se analizarán los efectos del número de subbandas y de la cuantificación no uniforme en la codificación en subbandas.

# DESEMPEÑO DE LA CODIFICACION EN SUBBANDAS

## 4.1 PROPÓSITO

En este capítulo se presentan la evaluación del desempeño de los codificadores en subbandas desarrollados de acuerdo con lo expuesto en los tres capítulos anteriores. Para medir el desempeño se utiliza la relación señal a ruido, descrita ya con anterioridad, en diversos esquemas propuestos para la codificación. La relación señal a ruido describe el grado de aproximación de la señal codificada a la señal original y es la medida más adecuada para los codificadores de forma de onda, como es el caso, y se usa para efectos de comparación entre los diversos esquemas. Sin embargo sigue siendo importante el grado de inteligibilidad logrado después de la recuperación de la señal, o sea la percepción de la señal codificada, esto es, la calidad subjetiva. Se mostrará la diferencia entre la codificación en subbandas con cuantificación uniforme y logarítmica; la influencia del número de bandas de división y de la frecuencia de muestreo de la señal original, así como el desempeño usando diferentes esquemas de división y distintos tipos de onduleta. La pruebas de asignación de bits, codificación y medidas de desempeño se realizaron usando las librerías de procesamiento de señales, comunicaciones y onduletas de MATLAB.



Los archivos de voz utilizados para las pruebas fueron grabados con un software de edición de audio para ser leídos y procesados posteriormente en MATLAB. Las señales originales de voz, son señales con formato digital, cuantificadas inicialmente con una resolución de 16 bits ( 65,535 niveles ). Se escogieron palabras con diferentes espectros de frecuencia para obtener los primeros resultados y después se aplicaron éstos a la codificación de frases completas.

## 4.2 CODIFICACIÓN CON CUANTIFICACIÓN UNIFORME.

### 4.2.1. *Ventaja sobre la codificación en banda completa*

En el capítulo anterior se mencionó la ventaja de la codificación en subbandas, contra la codificación en banda completa. El mejor desempeño de la cuantificación en subbandas se deriva del hecho de la minimización de la distorsión total tomando en cuenta la tasa de bits óptima para cada banda en lugar de asignar una tasa de bits única para la señal. La asignación de bits apropiada, sin importar el tipo de división que se realice, se da por medio de la ecuación (3.3.18); pero se debe hacer notar que  $b_k$  debe tomar valores enteros, por lo que se hace necesario redondear los valores obtenidos. Esto puede influir en el desempeño de la codificación aumentando o disminuyendo la relación señal a ruido dependiendo de la asignación de bits que se escoja.

La figura 4.1 muestra una comparación de la codificación en dos (0-2 y 2-4 kHz), tres (0-1, 1-2, 2-4 kHz) y cuatro (0-0.5, 0.5-1, 1-2, 2-4 kHz) subbandas octavas y la codificación en banda completa; los resultados corresponden a la palabra '*estudia*' muestreada a 8 kHz. Para la división se usaron los coeficientes de onduleta DB5.

En la figura 4.1 el eje X representa la tasa de transmisión R en múltiplos de la frecuencia de muestreo de acuerdo con la relación (3.2.1), lo que significa que el número de bits por muestra en banda completa b toma valores enteros desde 2 hasta 8 bits para la

codificación. El eje Y representa la relación señal a ruido después de la reconstrucción de la señal medida de acuerdo con la ecuación (3.1.7); la codificación de cada banda por separado, la reconstrucción y las variables involucradas fueron explicadas en la sección 3.3.4. El incremento del número de subbandas se obtiene aplicando reiteradamente la DWT sobre la banda de frecuencia más baja.

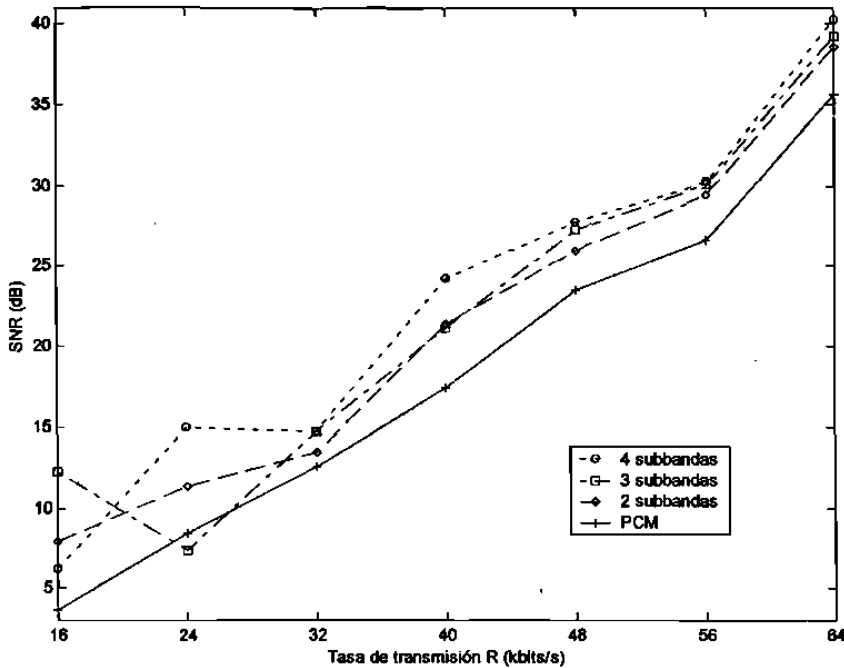


Figura 4.1. Comparación del desempeño en 2, 3 y 4 subbandas octavas para la palabra 'estudia'.

En la figura 4.1 se puede observar cual podría ser el número de subbandas adecuado, para una tasa de transmisión definida en el caso de la palabra 'estudia'; por ejemplo, a 16 kbts/s la mejor elección es dividir en tres subbandas, mientras que para 64 kbts/s se puede escoger la división en cuatro subbandas.

En la sección 3.3.4 se mostró que la ganancia  $G_{CSB}$  de la codificación en subbandas debe ser mayor o igual a uno. En la figura 4.1 se observa, sin embargo, que a 24 kbts/s la

codificación a 3 subbandas tiene una relación señal a ruido menor que la codificación en banda completa; esto es consecuencia del redondeo efectuado para obtener  $b_k$ . Esto mismo se puede observar en la figura 4.2 para 40 kbits/s; los resultados corresponden a la palabra 'virtud', muestreada a 8 kHz y dividida en dos tres y cuatro subbandas octavas. Se puede observar también que a 32 kbits/s, lo mejor es una división en cuatro subbandas; a 16 kbits/s se puede dividir en 3 o 4 subbandas.

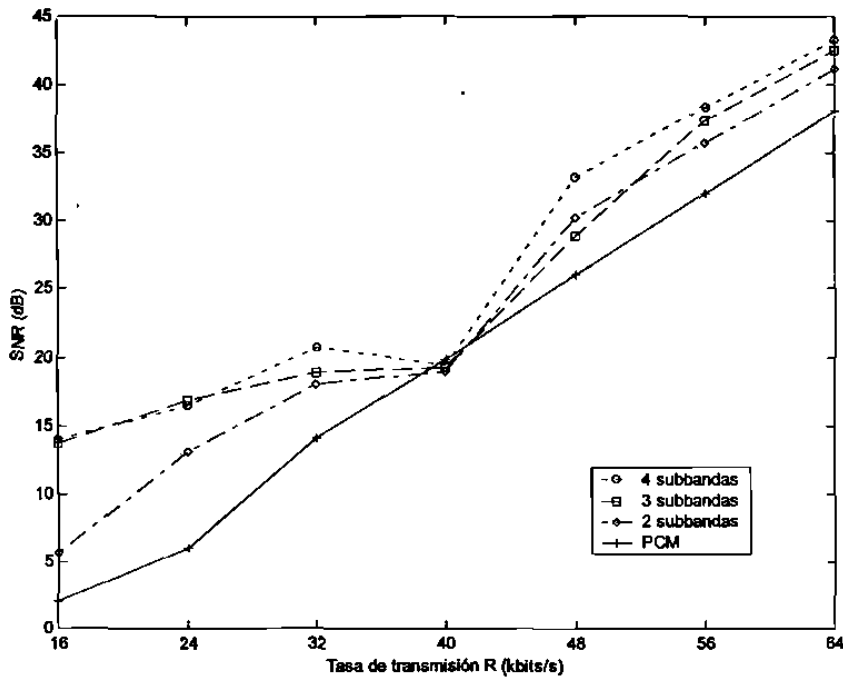


Figura 4.2. Comparación del desempeño en 2, 3 y 4 subbandas octavas para la palabra 'virtud'.

#### 4.2.2 Influencia del número de subbandas

El incremento del número de subbandas, idealmente, tendería a mejorar la relación señal a ruido hasta llegar a cierto límite después de unas seis o siete etapas de filtrado, en el caso de la división con onduletas. En la práctica, la relación señal a ruido no siempre va en aumento pero sí tiende a ser constante conforme aumentan las subbandas. En la figura 4.3

se muestra el desempeño de 2 a 8 subbandas octavas para la codificación de la palabra 'virtud' a 16, 24 y 32 kbits/s (codificación a 2, 3 y 4 bits), la frecuencia de muestro es de 8 kHz. El eje X representa el número de bandas octavas. Recuérdese que el incremento del número de bandas se obtiene de la aplicación sucesiva de la DWT sobre la banda de frecuencia más baja. El eje Y es la relación señal a ruido de la señal reconstruida, medida de acuerdo con la ecuación (3.1.7). Nos interesan sobre todo los resultados a 16 y 32 kbits/s (submúltiplos de 64 kbits/s), en cuyo caso los mejores desempeños generalmente se obtienen en 3 y 4 subbandas octavas.

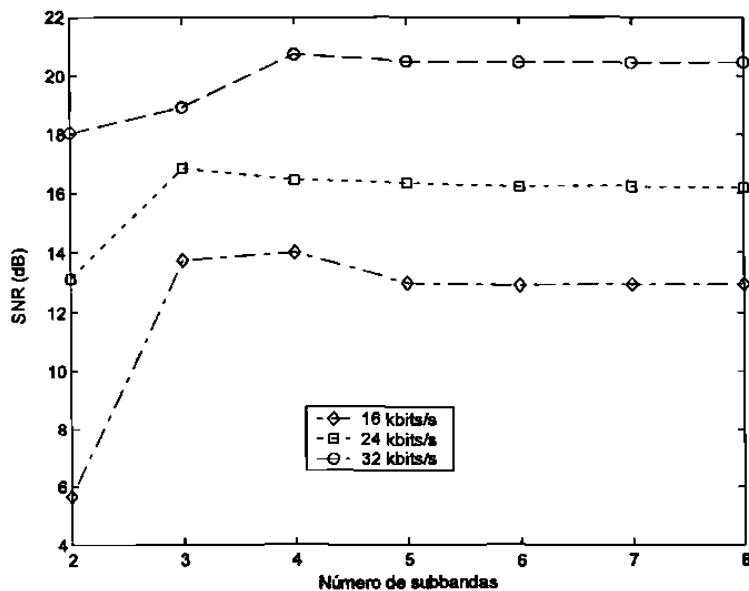


Figura 4.3. Desempeño para la codificación de la palabra 'virtud' de 2 a 8 subbandas, a 16, 24 y 32 kbits/s.

La figura 4.3 es un ejemplo de lo que sucede generalmente con las señales de voz codificadas a menos de 40 kbits/s, pero arriba de esa tasa de transmisión puede no suceder de la misma manera y, en la mayoría de los casos, se tienen generalmente desempeños más constantes y es más fácil identificar el número de subbandas concreto que daría la mejor relación señal a ruido o el número de bandas que queda excluido para ser usado.

## 4.3 CODIFICACION CON CUANTIFICACIÓN LOGARÍTMICA

### 4.3.1 Comportamiento de la cuantificación logarítmica

En la sección 3.2.2 se describió brevemente el objetivo de la cuantificación logarítmica para la codificación PCM como una forma para reducir la tasa de transmisión necesaria para codificar la señal. En el caso de la codificación en subbandas el objetivo es mantener la mejor calidad posible de la señal, por lo que se analiza el efecto de la cuantificación logarítmica al ser aplicada a las subbandas por separado. Para esta aplicación se escogió la constante  $\mu$  en cada subbanda para que sea igual a  $2^{bk} - 1$ . La figura 4.4 muestra la comparación entre el uso de la cuantificación uniforme y la logarítmica para una señal dividida en cuatro subbandas octavas codificada desde 16 a 64 kbits/s. La señal original corresponde a la palabra 'libertad' muestreada a 8 kHz.

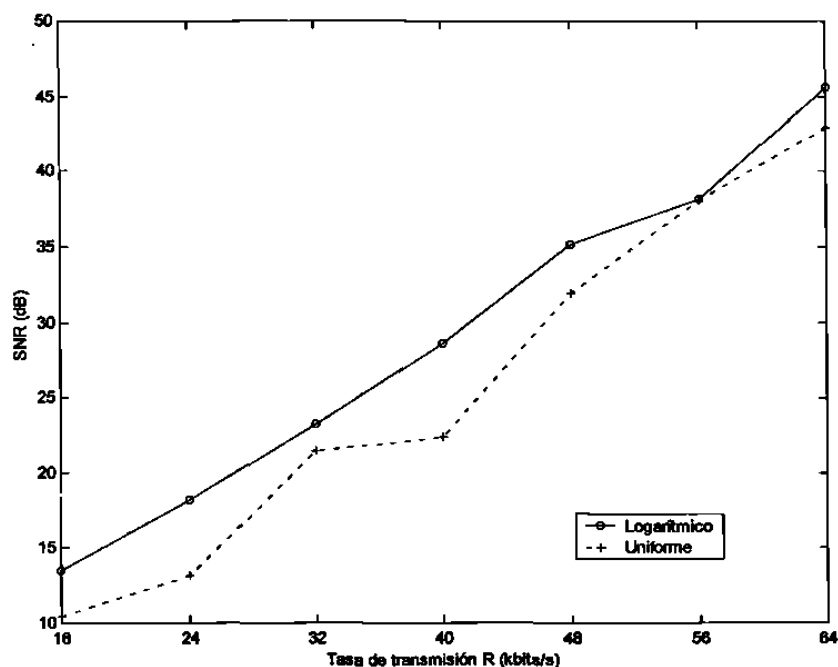


Figura 4.4. Comparación entre la cuantificación logarítmica y la uniforme para una codificación en cuatro subbandas octavas.

Aunque la diferencia es poca entre la cuantificación uniforme y la logarítmica, también puede ser aprovechada para obtener un mejor desempeño de la codificación. Generalmente, la ganancia, en los casos analizados va de 1 dB para las tasas de transmisión más bajas hasta 5 o 6 dB para las tasas más altas. En ocasiones puede no haber cambio en la relación señal a ruido para la tasa de 16 kbits/s.

### 4.3.2 Reducción de la frecuencia de muestreo

Una variable más que se puede manejar para obtener un mejor resultado es la frecuencia de muestreo. Todos los resultados anteriores fueron obtenidos usando una frecuencia de muestreo de 8 kHz. En la sección 1.2.1 se mencionó que es suficiente, en telefonía, para la inteligibilidad y la identificación de la voz un rango de frecuencias entre los 200 y los 3200 Hz; es decir, todavía podemos llevar la frecuencia de muestreo hasta los 6400 Hz. La figura 4.5 muestra la comparación de la codificación de la palabra 'virtud' a 16 y 32 kbits/s, con  $f_m = 6.4$  y  $f_m = 8$  kHz. La cuantificación es logarítmica.

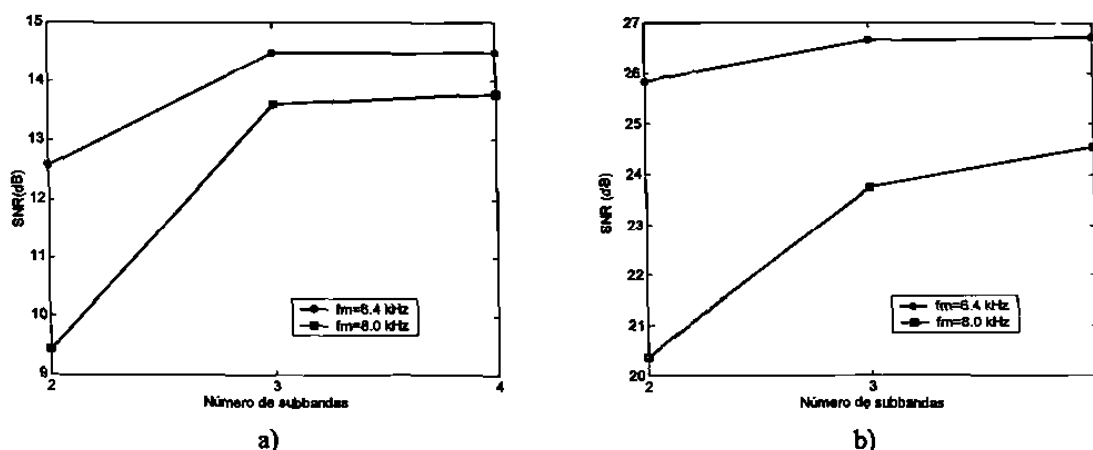


Figura 4.5. Desempeño con frecuencias de muestreo de 6.4 y 8 kHz a tasas de transmisión de a) 16 kbits/s, y b) 32 kbits/s

Se puede observar en la figura 4.5 que se incrementa la relación señal a ruido para la codificación a 16 y 32 kbits/s al utilizar la frecuencia de muestreo de 6.4 kHz. La calidad de la señal de voz codificada es buena para el caso de 32 kbits/s; y para 16 kbits por segundo,

aunque es inteligible se puede notar, al escucharla, una gran diferencia con la señal original. En los casos analizados, la mejor calidad para 32 kbits/s se obtiene para una división en cuatro subbandas octavas. Para 16 kbits/s, hay muy poca diferencia si se divide la señal en tres o cuatro subbandas. Generalmente, la diferencia de las relaciones señal a ruido, a 16 kbits/s, entre la codificación en banda completa y la codificación en subbandas representa un porcentaje que va del 60 al 90% de la SNR de la codificación en subbandas; a 32 kbits/s la diferencia representa de un 25 a un 35%.

#### 4.3.3. Diferentes esquemas de codificación.

Hasta aquí, se ha visto ya como influyen el tipo de cuantificación, el número de subbandas y la frecuencia de muestreo en el desempeño de la codificación en subbandas. Se ha puesto énfasis en la descomposición en octavas pero no se excluyen otros esquemas de división, pues la asignación de la tasa de bits puede usarse independientemente del esquema de separación que se utilice. En el caso de la aplicación de la DWT o el uso de filtros espejo no se está limitado a emplear solamente división en octavas o en bandas uniformes, sino que además se pueden hacer combinaciones. Por ejemplo, la figura 4.6 muestra el esquema propuesto por Crochiere [14] para codificar a 16 y 32 kbits/s en cinco bandas a partir de una señal muestreada a 8 kHz. Los rangos de las subbandas son: 0-500, 500-1000, 1000-2000, 2000-3000 y 3000-4000 Hz.

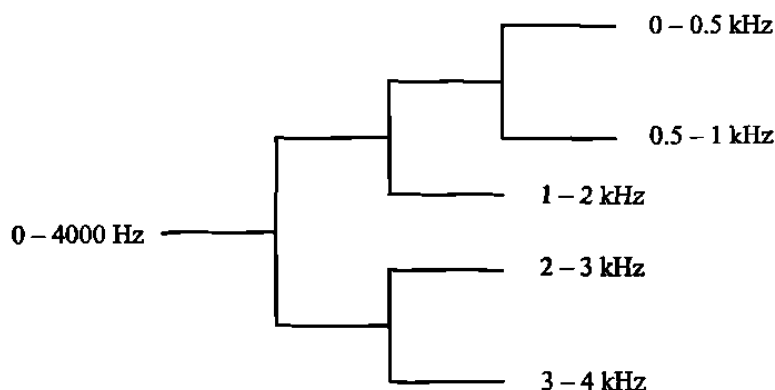


Figura 4.6 Esquema para codificación a 16 y 32 kbits/s en 5 bandas.

La asignación del número de bits por muestra propuesto por Crochiere para cada banda en la codificación a 16 kbits por segundo es  $\{4, 4, 2, 2, 0\}$ ; es decir, en realidad sólo se codificarían cuatro bandas; para la codificación a 32 kbits/s la asignación de bits sería  $\{5, 5, 4, 4, 3\}$ . Usando la ecuación (3.3.18) para asignar  $b_k$  a la codificación de diferentes señales de voz, se pueden obtener otras distribuciones del número de bits por muestra en cada banda, las cuales mejoran el desempeño de la cuantificación, por ejemplo:  $\{5, 5, 2, 0, 1\}$ ,  $\{6, 4, 2, 0, 1\}$ , y  $\{7, 3, 2, 0, 1\}$  en la codificación a 16 kbits/s;  $\{7, 7, 4, 2, 3\}$ ,  $\{8, 6, 4, 2, 3\}$  y  $\{9, 7, 4, 1, 3\}$  en la codificación a 32 kbits/s. El mejor desempeño de la codificación se obtiene con  $b_k = \{6, 4, 2, 0, 1\}$  para 16 kbits/s, y  $b_k = \{9, 7, 4, 1, 3\}$  para 32 kbits/s. La tabla 4.1 muestra la comparación de la relación señal a ruido (en dB) con tres frases distintas para las distribuciones de  $b_k$  correspondientes a 16 kbits/s usando el esquema de división de la figura 4.6. La codificación fue hecha con cuantificación logarítmica, y es así para todos los resultados posteriores.

Tabla 4.1. Desempeño a 16 kbits/s en 5 bandas para distintas distribuciones de  $b_k$  con  $f_m = 8$  kHz.

$b_k$		$\{4, 4, 2, 2, 0\}$	$\{5, 5, 2, 0, 1\}$	$\{6, 4, 2, 0, 1\}$	$\{7, 3, 2, 0, 1\}$
SNR (dB)	Frase 1	15.16	17.09	18.72	16.47
	Frase 2	13.84	15.13	15.29	14.29
	Frase 3	14.48	16.11	16.63	14.70

La tabla 4.2 muestra la comparación de la señal a ruido (en dB) con tres frases distintas para las distribuciones de  $b_k$  correspondientes a una tasa de transmisión de 32 kbits/s, de acuerdo con el esquema de la figura 4.6.



Tabla 4.2. Desempeño a 32 kbits/s en 5 bandas para distintas distribuciones de  $b_k$  con  $f_m = 8$  kHz.

$b_k$		{5, 5, 4, 4, 3}	{7, 7, 4, 2, 3}	{8, 6, 4, 2, 3}	{9, 7, 4, 1, 3}
SNR (dB)	Frase 1	19.38	27.00	27.78	29.07
	Frase 2	21.57	25.65	25.89	25.27
	Frase 3	19.64	25.39	25.51	25.85

La codificación de las mismas frases de los ejemplos anteriores, con las mejores distribuciones de  $b_k$  usando descomposición en octavas se muestran en la tabla 4.3, tanto para 16 como 32 kbits/s. Se da la relación señal a ruido (en dB) para cada frase.

Tabla 4.3. Desempeño de la codificación en subbandas octavas a 16 y 32 kbits/s, de señales de voz muestreadas a 8 kHz.

R		16 kbits/s	32 kbits/s		
$b_k$		{5, 1, 1}	{8, 4, 2}	{7, 5, 4, 3}	{9, 7, 4, 2}
SNR (dB)	Frase 1	17.01	27.65	25.18	27.51
	Frase 2	12.20	22.02	24.29	22.07
	Frase 3	16.30	23.35	24.34	23.20

Se puede notar la ventaja que tiene la codificación a 5 bandas según el esquema de la figura 4.6 sobre la codificación en bandas octavas para señales de voz muestreadas a 8 kHz, sobre todo a 32 kbits/s donde la diferencia es notable.

#### 4.3.4. Esquema de codificación propuesto.

Para señales de voz muestreadas a 6.4 kHz codificadas a 16 y 32 kbits/s, se obtienen mejores resultados con la descomposición en octavas. En la tabla 4.4 se puede ver la asignación óptima del número de bits por muestra y la relación señal a ruido resultante (en dB) para las mismas tres frases que en los ejemplos anteriores. Este esquema de codificación en subbandas octavas es el que se propone, con sus respectivas asignaciones de bits por subbanda en 16 y 32 kbits/s.

Tabla 4.4. Desempeño de la codificación en subbandas octavas a 16 y 32 kbits/s, de señales de voz muestreadas a 6.4 kHz

		R	16 kbits/s		32 kbits/s
		$b_k$	{5, 3, 1}	{5, 5, 3, 1}	{7, 7, 5, 4}
SNR (dB)	Frase 1		19.34	18.46	30.70
	Frase 2		15.13	15.00	27.84
	Frase 3		17.64	17.89	28.92

Trabajar con una frecuencia de muestreo de 6.4 kHz significa que la asignación de bits promedio  $b$  en banda completa para el caso de 16 kbits/s es de 2.5 bits por muestra.

#### 4.3.5. Contribución de las subbandas al desempeño.

En la sección 3.3 se explicó la manera en que se elige la asignación óptima de bits tomando en cuenta las varianzas del ruido de cuantificación de las subbandas. Idealmente, las varianzas del ruido de las subbandas deben ser iguales, de manera que contribuyan de la misma manera a la distorsión total buscando obtener el mejor desempeño; en la práctica esto se ve afectado debido al redondeo para la codificación de cada banda con un número entero de bits por muestra. Por otra parte, es obvio que la relación señal a ruido será mayor en aquellas bandas codificadas con un mayor número de bits por muestra; pero la relación

señal a ruido total calculada tomando en cuenta las contribuciones de distorsión de las subbandas, debe ser igual que la relación señal a ruido medida a partir de la señal reconstruida, de manera que se cumplan las condiciones establecidas en las secciones 3.3.3 y 3.3.4.

La tabla 4.5 muestra los datos de las distorsiones y el desempeño de las subbandas para una de las frases codificadas a 32 kbits/s según el esquema propuesto en la sección 4.3.4. Las subbandas están representadas como X0, X1, X2, X3. El total se refiere a la distorsión (D) y la relación señal a ruido (SNR) calculadas tomando en cuenta la contribución de cada banda; los datos representados por Xr son los que se obtienen directamente de la señal reconstruida y la original.

Tabla 4.5. Desempeño total y contribución de las subbandas

	X0	X1	X2	X3	Total	Xr
D (dB)	-38.3	-43.5	-40.4	-44.5	-41.9	-41.4
SNR (dB)	32.0	32.9	20.0	15.4	28.3	27.8

#### 4.3.6. Tipos de onduleta

Si se efectúa la descomposición en subbandas y se realiza la reconstrucción sin realizar la cuantificación, no importa el tipo de onduleta que se use, pues no hay diferencias significantes entre una y otra. Con respecto a su efecto en el desempeño después de la codificación, se analizaron las onduletas de Coifman, de Daubechies y las onduletas simétricas [9] contenidas en la librería de onduletas de MATLAB. Todos los resultados obtenidos hasta aquí fueron obtenidos usando los coeficientes del filtro de la onduleta DB5, de la familia de onduletas de Daubechies DBn (El número de coeficientes es  $2n$ , y el grado del filtro  $2n - 1$ ). En las dos figuras siguientes se muestra el desempeño para las

onduletas DB1 a DB10, las onduletas de Coifman  $COIF1$  a  $COIF5$  ( $COIFn$ , coeficientes =  $6n$ , grado =  $6n-1$ ) y las simétricas  $SYM2$  a  $SYM8$  ( $SYMn$ , coeficientes =  $2n$ , grado =  $2n-1$ ); los resultados son para la codificación a 16 kbits/s de la Frase 2 (figura 4.7) y la Frase 3 (figura 4.8) analizadas anteriormente. La frecuencia de muestreo de las señales corresponde a 6.4 kHz.

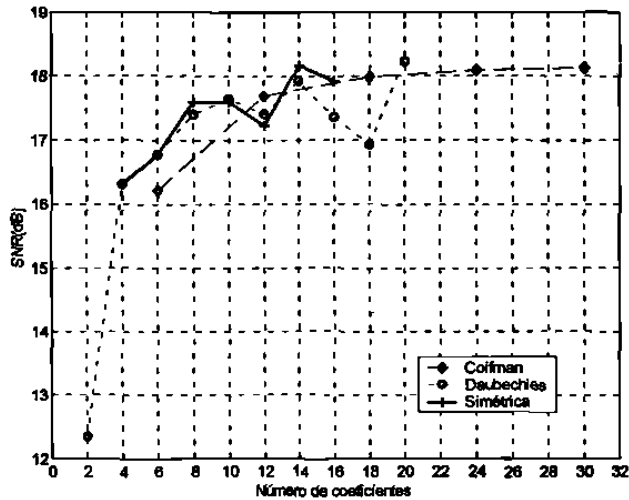


Figura 4.7. Relación señal a ruido para la codificación en subbandas octavas de Frase 2 usando tres familias de onduletas.

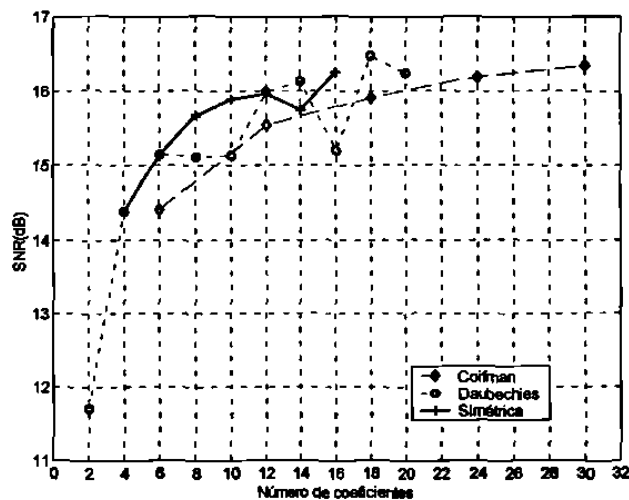


Figura 4.8. Relación señal a ruido para la codificación en subbandas octavas de Frase 3 usando tres familias de onduletas.

La familia de onduletas de Coifman muestra un comportamiento más regular que las de Daubechies y las simétricas. Sin embargo, con la onduleta DB10 (20 coeficientes), se puede obtener una relación señal a ruido mayor que con la onduleta COIF5 (30 coeficientes), la que presenta casi el mismo desempeño que la onduleta DB7. La distribución de los coeficientes de la onduleta DB10 se muestra en la figura 4.8.

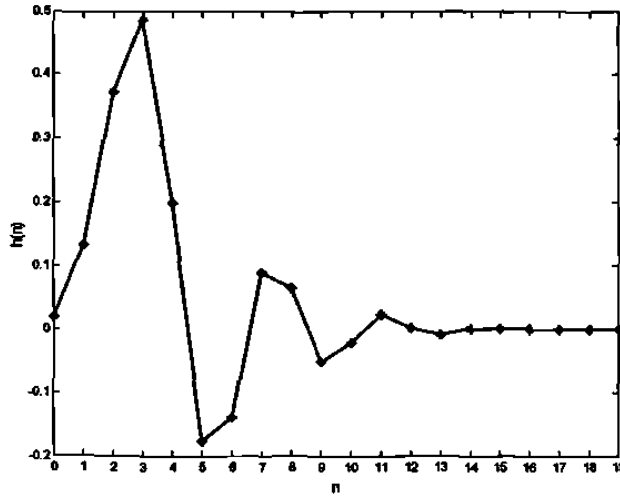


Figura 4.9. Coeficientes de la onduleta DB10.

Usando la onduleta DB10 para la codificación a 16 kbits/s en tres subbandas octavas con  $f_m = 6.4$  kHz y  $b_k = \{5, 3, 1\}$  la relación señal a ruido mejora ligeramente. Los desempeños en las frases 1, 2, y 3 serían: 19.61, 16.23 y 18.23 dB, respectivamente.

## 4.4 CONSIDERACIONES SOBRE EL PROCESAMIENTO DE LA SEÑAL

### 4.4.1 Normalización de los filtros

En el capítulo dos se establecieron las condiciones que deben cumplir los filtros de análisis y síntesis para poder obtener una reconstrucción perfecta de la señal sin aliasing y sin distorsión de amplitud. En la ecuación (2.3.21) se estableció la condición deseable que deben tener los filtros para evitar la distorsión pero, para ser consecuentes con la ecuación

(2.3.20), la ecuación (2.3.21) debería estar igualada a 2. Por tal motivo, generalmente se suelen multiplicar por 2 los filtros de la etapa de síntesis y, así, se cumple con la condición de no distorsión. Al trabajar directamente con los coeficientes de onduleta no hace falta realizar la normalización de los filtros, pues la solución implícita al usuario es multiplicar por  $\sqrt{2}$  los coeficientes tanto de los filtros de análisis como los de síntesis, solución que también se puede poner en práctica al trabajar con los bancos de filtros espejo. Para obtener los resultados de este capítulo se trabajó directamente con los coeficientes de onduleta de la librería de onduletas de MATLAB.

#### ***4.4.2. Orden de las operaciones de procesamiento.***

En la figura 2.4 se mostró el esquema básico de la codificación en dos subbandas, sobre el que se desarrolló todo el análisis del capítulo 2. En la etapa de síntesis aparece primero el filtrado y después la diezmación. En la aplicación para la codificación en subbandas resulta ineficiente representar la secuencia muestreada directamente en su forma original, es de ahí de donde se origina la necesidad del diezmado. La ineficiencia se debe a que si primero se filtra la señal en dos subbandas y después es diezmada, la mitad de las operaciones realizadas serían desperdiciadas, pues al diezmar se destruyen la mitad de las muestras de la señal que fueron calculados con el filtrado; del mismo modo, si se realiza una división en M bandas se conservaría información redundante de la señal a la vez que se incrementaría el número de muestras a transmitir en el mismo factor M. Para evitar esto, generalmente se invierte el orden del filtrado y el diezmado en el esquema de la figura 2.4, y del filtrado con la expansión, aplicando lo que se conoce como las identidades de Noble, dando lugar a la implementación polifásica. Ésta se basa en la separación de la señal en sus muestras pares y nones para después ser filtradas; entonces el número de operaciones queda reducido pudiendo hacer los cálculos a una mayor velocidad. Para mayor referencia sobre estas operaciones se puede consultar [3,8,15].

#### 4.5. OBSERVACIONES.

Aunque el desempeño obtenido para los casos observados representan un tipo de medición objetiva, se puede decir que las cualidades subjetivas de las señales de voz con los mejores desempeños de codificación son bastante buenas para los esquemas propuestos, sobre todo a 32 kbits/s. Como se mencionó en el primer capítulo, la obtención de mediciones subjetivas requiere de oyentes entrenados, lo cual queda fuera del alcance del trabajo. Se obtienen buenos resultados para las señales de voz muestreadas a 8 kHz y también a 6.4 kHz; que es la frecuencia de muestreo del esquema de codificación propuesto al final. El uso de la DWT o los bancos de filtros espejo en cuadratura no se restringe a la división en octavas, como ya se ha mostrado, aunque sea ésta la que se propone. Se mostró además el por qué de la elección de las onduletas de Daubechies para el procesamiento, aunque en general se puede usar cualquier otra familia ya que la diferencia es muy poca. El esquema de codificación propuesto es para subbandas octavas a partir de una señal de voz muestreada a 6.4 kHz. La asignación del número de bits de la banda más baja a la más alta para el caso de 16 kbits/s (tres subbandas) sería {5, 3, 1}, y para 32 kbits/s (cuatro subbandas) sería {7, 7, 5, 4}.

## **CONCLUSIONES**

### **5.1 SINTESIS**

Siempre es posible transmitir información a una velocidad menor que la capacidad del canal con un error relativamente pequeño, aun en presencia de ruido. Esto se logra mediante la codificación de la información, la cual provee un formato uniforme para diferentes tipos de señales. Las señales codificadas pueden ser reconocidas, procesadas y regeneradas eficientemente, además de que pueden ser encriptadas para aplicaciones que requieren de privacidad.

Uno de los métodos para codificar eficientemente señales de voz es la codificación *en subbandas*. Éste se clasifica dentro de los codificadores de forma de onda, que intentan producir una señal lo más parecida posible a la original. El interés primordial se encuentra en su aplicación a la telefonía digital.

El espectro de la voz contiene frecuencias que no van más allá de los 5 kHz. En telefonía es suficiente la porción entre los 200 y los 3200 Hz; a esto se refiere la calidad telefónica cuyos requisitos son la inteligibilidad de la voz y el reconocimiento del hablante. Generalmente la señal de voz es muestreada a 8 kHz, pero las características del método presentado permiten reducir la frecuencia de muestreo hasta los 6.4 kHz.

El mayor contenido de energía de una señal de voz se encuentra localizado en las frecuencias más bajas. Esta característica es aprovechada en la codificación de subbandas cuyo propósito es dividir la señal de entrada en varios componentes de frecuencia y después



codificar cada uno de esos componentes de manera separada. La división es efectuada mediante el empleo de filtros espejo en cuadratura junto con el diezmado y la interpolación; su diseño ayuda a evitar los efectos de aliasing y distorsión permitiendo una reconstrucción perfecta de la señal. Estas operaciones se pueden realizar directa y eficientemente mediante el empleo de la transformada discreta de onduleta que es de gran utilidad para el procesamiento de señales no estacionarias como son las señales de voz.

El método tiene la ventaja de que el número de bits usados para codificar cada banda de frecuencia puede ser diferente y también variable; de esta manera la precisión de la codificación puede ser ubicada donde sea más necesaria. Debido a la distribución del espectro de voz se asignan más bits a las bandas de baja frecuencia y menos bits a las bandas de frecuencias más altas.

Por medio de una apropiada asignación de bits en las diferentes bandas, el número de niveles de cuantificación y la varianza del error de reconstrucción pueden ser controlados de forma separada para cada una de las bandas; además el error total de reconstrucción puede también ser controlado como una función de la frecuencia como es expresado en la ecuación (3.3.18).

## 5.2 APORTACIONES.

Con el objeto de reducir la tasa de transmisión de las señales de voz sin demasiado deterioro de su calidad; se analizó la codificación en subbandas tomando en cuenta la influencia del número de subbandas, el tipo de cuantificación, la frecuencia de muestreo y el tipo de onduleta utilizado para el análisis y la reconstrucción.

En las redes de transmisión digital modernas pueden ser asignadas tasas de canal múltiples o submúltiplos de 64 kbits/s para diversos tipos de señales.

La tasa de transmisión de voz en el estándar de telefonía es de 64 kbits/s. Con la codificación de voz en subbandas es posible reducir la tasa de transmisión a 32 y 16 kbits/s

obteniendo buenos resultados. A 16 kbits/s se pueden obtener ganancias de hasta 10 dB sobre la codificación en banda completa.

En el esquema propuesto para la codificación de voz en subbandas la señal es muestreada a 6.4 kHz y dividida en subbandas octavas; cada banda es codificada usando cuantificación logarítmica. El análisis y reconstrucción de la señal se puede realizar usando la familia de onduletas de Daubechies eligiendo diez o más coeficientes de onduleta.

Para la codificación a 32 kbits/s se realiza una división en cuatro subbandas octavas: 0-400, 400-800, 800-1600 y 1600-3200 Hz. La asignación de bits en cada banda es, respectivamente, {7, 7, 5, 4}.

En la codificación a 16 kbits/s se divide la señal en tres subbandas octavas: 0-800, 800-1600 y 1600-3200 Hz. La asignación de la tasa de bits es {5, 3, 1}.

### 5.3 TRABAJOS FUTUROS

El análisis para la codificación de las subbandas fue realizado usando el esquema de codificación más simple, la codificación PCM. Otros estudios podrían incluir técnicas diferentes en la codificación de las subbandas, como por ejemplo la codificación delta, la adaptiva o la predicción lineal.

El esquema de codificación presentado se realiza utilizando únicamente las propiedades de la señal, pero el esquema se puede ampliar hacia técnicas que involucren los criterios perceptivos del oído y que aprovechen fenómenos tales como el enmascaramiento de señales auditivas mencionados en el primer capítulo.

## REFERENCIAS

- [1] A. Cohen and R.D. Ryan, *Wavelets and Multiscale Signal Processing*, Chapman and Hall (1995).
- [2] A. N. Akansu and R.A. Haddad, *Multiresolution Signal Decomposition*, Academic Press (1992).
- [3] A.N. Akansu and M. J. T. Smith, eds., *Subband and Wavelet transforms*, Kluwer (1995).
- [4] Brian C.J. Moore, *An Introduction to the Psychology of Hearing*. Academic Press 1989.
- [5] C. Sidney Burrus, Ramesh A. Gopinath, Haitao Guo; *Wavelets and Wavelet Transforms*. Prentice Hall, 1998.
- [6] D. Esteban and C. Galand, *Application of quadrature mirror filters to split band voice coding systems*, International Conference on Acoustics, Speech and Signal Processing. Washington, DC. May 1977.
- [7] G. Kaiser, *A Friendly Guide to Wavelets*. Birkhäuser, 1994.
- [8] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley Cambridge Press (1996).
- [9] I. Daubechies, *Ten Lectures on Wavelets*, SIAM (1992).
- [10] J.R. Deller Jr., J.G. Proakis and J.H.L. Hansen, *Discrete Time Processing of Speech Signals*. Macmillan, 1993.
- [11] Jhon G. Proakis, *Digital Communications*. McGraw Hill, 1995.
- [12] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice Hall (1995).
- [13] N. J. Fliege, *Multirate Digital Signal Processing*, John Wiley (1994).
- [14] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall (1984).
- [15] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*. Prentice Hall, 1993.
- [16] R. Crochiere and L. Rabiner, *Multirate Digital Signal Processing*, Prentice Hall, 1983.

- [17] R. Crochiere et al, "Real-time Speech Coding", IEEE Trans. COM-30(4), Apr. 1982.
- [18] R. Crochiere, S. Webber, and J. Flanagan, "Digital Coding of Speech in Sub-bands", The Bell Tech. J., Vol 55(8), Oct. 1976.
- [19] S. Mallat, Wavelet Signal Processing, Academic Press (1996).
- [20] Y. Meyer, Wavelets: Algorithms and Applications, SIAM (1993).

