

TRABAJO ESPECIAL DE GRADO

DESARROLLO DE UNA PLATAFORMA AUTOMATIZADA PARA LA ESTIMACIÓN DE LA CALIDAD DE SISTEMAS DE COMPRESIÓN DE SONIDO

Presentado ante la Ilustre
Universidad Central de Venezuela
por el Br. Barillas R., José R.
para optar al título de
Ingeniero Electricista

Caracas, 2007

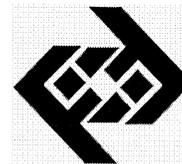
TRABAJO ESPECIAL DE GRADO

DESARROLLO DE UNA PLATAFORMA AUTOMATIZADA PARA LA ESTIMACIÓN DE LA CALIDAD DE SISTEMAS DE COMPRESIÓN DE SONIDO

Tutor Académico: Prof. Luis Fernández

Presentado ante la Ilustre
Universidad Central de Venezuela
por el Br. Barillas R., José R.
para optar al título de
Ingeniero Electricista

Caracas, 2007



CONSTANCIA DE APROBACIÓN

Caracas, 12 de febrero de 2007

Los abajo firmantes, miembros del Jurado designado por el Consejo de Escuela de Ingeniería Eléctrica, para evaluar el Trabajo Especial de Grado presentado por el Bachiller José Barillas, titulado:

“DESARROLLO DE UNA PLATAFORMA AUTOMATIZADA PARA LA ESTIMACIÓN DE LA CALIDAD DE SISTEMAS DE COMPRESIÓN DE SONIDO”

Consideran que el mismo cumple con los requisitos exigidos por el plan de estudios conducente al Título de Ingeniero Electricista en la mención de Comunicaciones, y sin que ello signifique que se hacen solidarios con las ideas expuestas por los autores, lo declaran APROBADO.


Prof. Zeldivar Bruzual
Jurado


Prof. Luis Fernández
Tutor Académico


Prof. Francisco Marchena
Jurado



DEDICATORIA

A mis padres.
A mi esposa.
A mis suegros.

AGRADECIMIENTOS

“...conducir por orden mis pensamientos,
comenzando por los objetos más fáciles de conocer,
para ascender poco a poco, como por grados,
hasta el conocimiento de los más compuestos...”

Del “*Discurso del Método*”

René Descartes

Agradezco a Dios el Don dado todos los días de comenzar a amar de nuevo.

Le agradezco infinitamente a mi esposa Mayra por su constancia, amor y perseverancia. Incansable compañera y amor eterno.

Les agradezco a mis padres su compromiso incondicional y dedicación. Su entrega a mis necesidades espero retribuir las en grande.

A mis suegros y cuñados que siempre me han apoyado y querido incondicionalmente.

Agradezco a mi familia, valioso soporte desde siempre.

A mis compañeros universitarios que me enseñaron valiosas lecciones de vida. A Austin Martínez por su dedicación y entrega, Luis Maracara por su apoyo en los momentos difíciles y a Ricardo Rodríguez por enseñarme que la vida se vive mejor en paz. Con ellos logré apreciar lo importante del trabajo en equipo.

Al Profesor Pedro Pinto por sus útiles y oportunos consejos que me hicieron seguir adelante en los estudios.

A mi tutor Académico, Profesor Luis Fernández, quien ha tenido la paciencia y la comprensión necesarias para apoyarme en este proyecto.

Barillas R., José R.

**DESARROLLO DE UNA PLATAFORMA AUTOMATIZADA PARA
LA ESTIMACIÓN DE LA CALIDAD DE SISTEMAS DE
COMPRESIÓN DE SONIDO**

Tutor Académico: Prof. Luis Fernández. Tesis. Caracas. U.C.V. Facultad de Ingeniería. Escuela de Ingeniería Eléctrica. Ingeniero Electricista. Opción: Comunicaciones. Institución: U.C.V. 2007. 127 h. + anexos.

Palabras claves: Compresión de audio; Modelo psicoacústico de codificación de audio; Evaluación de calidad de audio comprimido; MOS objetivo; Normas ISO/IEC Audio; MATLAB.

Resumen. Como último objetivo se plantea el desarrollo de una plataforma funcional de evaluación de audio comprimido por medio del modelo psicoacústico, utilizando para ello primero la implementación de un modelo de compresión, siguiendo la norma ISO/IEC 11172-3 con codificación basada en la percepción psicoacústica y luego su comparación con codificadores de uso público. Ambas herramientas se desarrollan en entorno MATLAB. De tal forma, se obtiene un indicador objetivo de calidad de audio percibido que no depende de la evaluación subjetiva de algún oyente humano. La plataforma presentada, así como la implementación del compresor, están debidamente comentadas para facilitar futuras actualizaciones requeridas.

ÍNDICE GENERAL

	Pág.
CONSTANCIA DE APROBACIÓN	
DEDICATORIA	
AGRADECIMIENTOS	
RESUMEN	V
LISTA DE ABREVIATURAS	X
LISTA DE FIGURAS	XI
LISTA DE GRÁFICAS	XIII
LISTA DE TABLAS	XIV
INTRODUCCIÓN	1
CAPÍTULO I	
EL MODELO PSICOACÚSTICO	6
1. Métodos y procedimientos en Psicoacústica	6
2. Experimentos realizados.	8
2.1 Sensibilidad del oído humano	8
2.2 Enmascaramiento en frecuencia	9
2.3 Enmascaramiento temporal.	9
3. Percepción	10
3.1 Enmascaramiento simultáneo.	11
3.2 Ruido usado para mediciones psicoacústicas.	15
3.2.1 Enmascaramiento de sonidos senoidales por ruido de banda ancha	15
3.2.2 Enmascaramiento de sonidos senoidales por ruido de banda angosta.	16
3.3 Enmascaramiento de sonidos senoidales por sonidos senoidales.	17
3.4 Enmascaramiento de sonidos senoidales por sonidos complejos.	18
3.5 Enmascaramiento temporal.	19
3.5.1 Pre-Enmascaramiento.	20
3.5.2 Post-Enmascaramiento.	20
4. Umbrales de la audición.	21
4.1 Umbrales absolutos.	21
4.1.1 Umbral de audibilidad.	22
4.1.2 Umbrales de frecuencia.	24
4.2 Umbrales diferenciales.	25
4.2.1 Umbrales diferenciales de intensidad.	25
4.2.2 Umbrales diferenciales de frecuencia.	26
5. Codificación basada en la percepción.	28
5.1 Codificación de audio de alta calidad.	28
5.1.1 Bandas críticas.	29
5.1.2 Enmascaramiento auditivo.	30
5.2 Principios básicos de la codificación de audio perceptiva.	30
5.2.1 Banco de filtros.	31

5.2.2 Modelo Perceptivo o psicoacústico.	31
5.2.3. Cuantificación y codificación	31
5.2.3.1 El codificador.	32
5.2.3.2. Decodificación.	32
5.2.3.3 Formato y empaquetamiento de la trama de bits.	33
CAPÍTULO II	
PROCEDIMIENTOS BÁSICOS GENERALES A LA ENTRADA DEL CODIFICADOR MPEG	35
1. Codificador MPEG	35
1.1 Generalidades	35
1.2 Filtro Pasa Alto de entrada	36
1.3 Filtro de Análisis de sub-banda	36
1.4 El modelo Psicoacústico	39
1.5 Asignación de Bits o Ruido.	39
1.5.1. Método de asignación de bits o ruido.	39
1.6 El <i>formateador</i> de la trama	40
2. Capa I.	40
2.1 Asignación de bits.	41
2.2 Cuantificación y codificación de las muestras de la sub-bandas.	42
2.3 La codificación de la asignación de bits.	43
2.4 Datos Auxiliares.	43
2.5 Formato de los datos.	43
3. Capa II.	45
3.1 Modelo Psicoacústico.	45
3.2 Codificación de los Factores de escala (SCF).	45
3.2.1 Codificación de los SCFSI.	46
3.3 Asignación de bits.	46
3.4 Cuantificación y codificación de las muestras de la sub-bandas.	48
3.5 Asignación de bits.	48
3.6 Datos Auxiliares.	49
4. Capa III.	49
4.1 Generalidades.	49
4.2 Modelo Psicoacústico.	54
4.3 Decisión de conmutación de ventana.	56
4.4 Análisis del Banco de filtro Híbrido.	57
4.5 Transformada discreta del coseno modificada MDCT	58
CAPÍTULO III	
MODELOS PSICOACÚSTICOS DE LA ISO/IEC 11172-3	59
1. Modelo Psicoacústico I.	59
1.1 Transformada rápida de Fourier.	61
1.1.1 Ventana Hann	62
1.1.2 Densidad espectral de potencia $X_{(k)}$.	62
1.2. Determinación del Nivel de Presión Sonora	63

1.3. Consideración del Umbral de silencio (<i>Quiet Threshold</i>) para cada componente de la FFT.	63
1.4. Localización de las componentes tonales y No-tonales.	64
1.4.1. Etiquetado del máximo local.	66
1.4.2. Listado de los componentes tonales y cálculo de SPL.	66
1.4.3. Listado de las componentes no tonales y cálculo de la energía.	66
1.5. Diezmado de las componentes tonales y no tonales.	67
1.6. Cálculo de los umbrales individuales de enmascaramiento.	68
1.6.1 Capa I.	69
1.6.2 Capa II.	69
1.7 Cálculos del Umbral de enmascaramiento global (LTg).	71
1.8. Determinación del umbral de enmascaramiento mínimo.	72
1.9. Cálculo de la relación Señal-Máscara (SMR).	72
2. Modelo Psicoacústico II.	73
2.1 El proceso de cálculo de generación del umbral.	73
2.1.1. La longitud de desplazamiento (iblen).	73
2.1.2 La elección del método depende de la implementación.	74
2.1.3 La frecuencia de muestreo.	74
2.2 Comentarios sobre notaciones.	74
2.3 Pasos en el cálculo del umbral.	75
2.3.1 Reconstrucción de 1024 muestras de la señal de entrada.	75
2.3.2 Cálculo del espectro complejo de la señal de entrada.	76
2.3.3 Cálculo predictivo de r y f.	76
2.3.4 Cálculo de la medida de impredecibilidad C_w .	76
2.3.5 Cálculo de la Energía e Impredecibilidad en las particiones de cálculo de umbral.	77
2.3.6 Convolución de la energía <i>particionada</i> y la impredecibilidad con la función de Spreading.	78
2.3.7 Convertir cb_b a tb_b , el índice de tonalidad.	79
2.3.8 Cálculo de la SNR requerida en cada partición.	79
2.3.9 Cálculo de la relación de potencia.	79
2.3.10 Cálculo del umbral de energía actual, nb_b .	79
2.3.11 Esparcir (spread) el umbral de energía sobre las líneas de FFT,	79
2.3.12 Inclusión de los umbrales absolutos	79
2.3.13 Control de Pre-eco	80
2.3.14 Cálculo de las relaciones señal-máscara, SMR_n .	80
3. Modificaciones al modelo Psicoacústico II para su uso con Capa III	81

CAPÍTULO IV	
MEDICIONES DE CALIDAD	87
1. Puntaje medio de opinión subjetiva (MOS).	88
2. Técnicas de medición en el dominio no perceptivo.	90
3. Técnicas de medición perceptiva.	91
3.1. Modelo de variables de salida (<i>Model Output Variables</i> MOV)	94
CAPÍTULO V	
PLATAFORMA DE EVALUACIÓN DE CALIDAD DE AUDIO	96
1. Introducción.	96
2. Estimación del retraso entre señales de audio.	98
3. Transformación perceptiva.	99
4. Medición de diferencias entre señales.	100
4.1 El bloque de medida en tiempo.	100
4.2 El bloque de medida en frecuencia.	101
5. Resultados del uso de la plataforma de evaluación propuesta.	104
5.1 Codificadores utilizados.	104
5.2 Audio seleccionado.	104
6. Resultados.	106
6.1 Respecto al modelo psico-acústico implementado.	106
6.2 Respecto a la plataforma de evaluación implementada.	107
CONCLUSIONES	110
RECOMENDACIONES	114
REFERENCIAS BIBLIOGRÁFICAS	116
BIBLIOGRAFÍA	119
GLOSARIO	123
ANEXOS	128

LISTA DE ABREVIATURAS

ADPCM	PCM adaptativo
ASA	American Standards Association
ASPEC	Adaptive Spectral Perceptual Entropy Coding
ATRAC	Adaptive TRansform Acoustic Coding
Bslbf	bit string, left bit first. Cadena de bits, el bit izquierdo primero (forma en que se escriben las cadenas de bits en la norma ISO/IEC 11172).
Ch	Canal. Si su valor es 0, indica el canal izquierdo de una señal estéreo o el primero de una señal de 2 canales independientes.
CPU	Central Processing Unit, Unidad Central de Proceso (UCP).
DC	Direct Current. Corriente continua.
DPCM	PCM diferencial
DSP	Digital Signal Processor, Procesador Digital de Señal.
Gr	Gránulo de 3x32 muestras de subbanda en Layer II o 18x32 muestras de subbanda en Layer III.
Main_data	parte de la porción del bitstream de audio que contiene los Scalefactors, datos codificados con Huffman e información auxiliar (ancillary).
Main_data_beg	Es la ubicación en la trama del comienzo de Main_data. Esta ubicación corresponde al final de Main_data de la trama anterior más 1 bit. Esta se calcula a partir de main_data_end de la trama previa.
MPEG	Moving Pictures Expert Group
Part2_length	representa el número de bits de main_data utilizado para los SCF.
Rpchof	coeficientes polinomiales remanentes, el orden más alto primero.
Sblimit	Es el número de la subbanda más baja en la cual no se asignan bits.
SCF	Factor(es) de escala.(ScaleFactor)
SCFSI	Selección de Información de los Factores de Escala (Scale Factor Selection Information).

LISTA DE FIGURAS

FIGURAS	Pág.
1. Sensibilidad del oído humano en función de la frecuencia. Umbrales absolutos.	8
2. Enmascaramiento en frecuencia del tono de 1 kHz.	9
3. Enmascaramiento en frecuencia de múltiples tonos.	10
4. Efectos del pre y post enmascaramiento	10
5. Enmascaramiento simultáneo.	11
6. Umbral de enmascaramiento.	13
7. Enmascaramiento de sonidos senoidales por ruido de banda ancha.	15
8. Sonido senoidal enmascarado por ruidos de banda angosta.	16
9. Sonido senoidal enmascarado con niveles variables.	17
10. Enmascaramiento de sonidos senoidales por sonidos senoidales.	18
11. Enmascaramiento de sonidos senoidales por sonidos complejos.	19
12. Post-enmascaramiento.	20
13. Post-enmascaramiento producido en función de la duración del sonido enmascarador.	21
14. Umbrales de audición.	22
15. Umbral de audibilidad en detalle.	23
16. Diferencia entre niveles de presión en el oído humano	24
17. Umbral de mínima variación perceptible de intensidad.	26
18. Umbrales de mínima diferencia perceptible para diferentes frecuencias.	26
19. Umbral de mínima variación perceptible en función de la frecuencia del sonido senoidal de prueba.	27
20. Umbral de mínima diferencia perceptible en Hz en función de la frecuencia del sonido senoidal de prueba.	27
21. Diagrama de bloque sistema de codificación de audio perceptivo MPEG	30
22. Ejemplo de esquema del codificador y decodificador de audio MPEG-1 para capa I y II (un solo canal)	36
23. Diagrama de bloques del Filtro de Análisis de Sub-banda. Banco de	

Filtros Polifásicos.	38
24. Formato de la trama en la capa I.	41
25. Diagrama de bloques codificador MPEG-1 para Capas I y II.	44
26. Partición de la trama en capa II.	45
27. Formato de la trama en la capa II.	49
28. Ejemplo de esquema del codificador/decodificador MPEG 1 Capa III.	50
29. Bloque de codificación con situación de pre-eco.	51
30. Procesamiento de la señal en el banco de filtros para señal estéreo y Bloques Largos en Capa III.	53
31. Efecto del pre-eco en la señal según la longitud del bloque.	55
32. Diagrama de estados de la lógica para la conmutación de ventanas.	56
33. Ejemplo de cancelación del <i>aliasing</i> por medio del MDCT.	57
34. Estudio del modelo Psicoacústico I, según la norma ISO 11172-3.	60
35. Detalle Modelo Psicoacústico implementado en Capa I y II de MPEG.	61
36. Diagrama de bloques del modelo psicoacústico 2: Cálculo del Umbral.	84
37. Diagrama de bloques del modelo Psicoacústico 2: Cálculo del Umbral (Parte 1).	85
38. Diagrama de bloques del Modelo Psicoacústico 2: Cálculo del Umbral (Parte 2).	86
39. Diagrama de Bloques del Modelo Psicoacústico 2 Capa III. Cálculo del umbral para bloques cortos. (Parte 3).	86
40. Auditorio utilizado por el Centro de Investigación de Comunicaciones de Canadá.	89
41. Ejemplo de MOS de la Capa II y Capa III (64 Kbps por canal).	90
42. Modelo perceptivo de la plataforma de evaluación de calidad propuesta	93
43. Esquema general de la plataforma de evaluación de calidad audio comprimido.	98
44. Esquema de medición en tiempo.	100
45. Esquema de medición en frecuencia	101
46. Estructura de evaluación en tiempo y frecuencia de la plataforma diseñada.	102

LISTA DE GRÁFICAS

GRÁFICAS	Pág.
1. Comparación de MOS obtenido por compresor a 96 kbps	106
2. Comparación de MOS obtenido por compresor a 128 kbps.	106
3. MOS obtenido con codificador 1.	108
4. MOS obtenido con codificador 2.	108
5. MOS obtenido con codificador 3	109

LISTA DE TABLAS

TABLAS	Pág.
1. Tamaño de la trama según la frecuencia	43
2. Clase de cada diferencia	45
3. Datos técnicos de la Transformada Rápida de Fourier (FFT)	62
4. Rango de frecuencias por Capa para la determinación del componente tonal	65
5. Valores del subíndice j.	66
6. Cantidad de muestras según la Capa.	70
7. Diferentes mediciones de calidad de audio.	87
8. Descripción de variables en el Modelo de salida (MOV) del sistema PEAQ básico.	95

INTRODUCCIÓN

A lo largo de los años se han desarrollado sistemas que permiten comprimir el sonido con la intención de ahorrar ancho de banda (ó almacenamiento). Casi todos ellos logran elevadas tasas de compresión utilizando algoritmos denominados “con pérdidas”, en el sentido que se desecha información que debería ser imperceptible al oído. La codificación directa de la forma de onda del sonido, como por ejemplo PCM, preserva toda la información de la señal, pero tiene el inconveniente de requerir altos anchos de banda para su transmisión. Sin embargo, un análisis estadístico de las muestras de sonido muestra una alta redundancia entre muestras sucesivas lo cual se ha utilizado en algunos algoritmos de codificación de voz para la reducción del ancho de banda requerido. Por ejemplo, la codificación de voz utilizando el estándar G.711 de la UIT-T requiere de 64 kbps para su transmisión, en tanto que si se utiliza ADPCM como en el caso de G.723 esto se puede bajar a unos 32 kbps. Otros algoritmos de este tipo como el G.729 pueden llegar a reducir el ancho de banda aún más pero a costa de una reducción de la calidad del sonido reconstruido. Otros algoritmos de codificación de voz se basan en modelaje del tracto vocal y pueden lograr compresiones aún mayores, con un compromiso entre calidad y ancho de banda. Un ejemplo de ello es G.723.1, ampliamente utilizado en Internet y en aplicaciones de Voz sobre IP, el cual puede llegar a velocidades de 5,6 y 6,5 kbps (usando dos algoritmos diferentes de acuerdo a la congestión del canal).

Debido al procesamiento que hace el oído humano de la señal reconstruida, no se puede inferir la calidad de la misma haciendo una simple comparación de las formas de onda a la entrada y a la salida del sistema. Por esa razón es necesario incorporar al usuario humano en el sistema de medición, por lo cual la medida debe incorporar este componente subjetivo. Entonces, la “calidad” del sistema vendrá expresada en algún sistema de unidades apropiadas para describir el grado de satisfacción del usuario ante el servicio ofrecido. Hoy día, el mecanismo ampliamente

aceptado para la descripción de la calidad de estos sistemas es el M.O.S. (Mean Opinion Score) el cual representa una media estadística representada en una escala arbitraria del 1 al 5.

JUSTIFICACIÓN

Cuando se trata de aplicaciones de sonido (a diferencia de la voz), debe recurrirse a otros algoritmos que realicen compresiones aún mayores, lo cual implica el desecho de una mayor cantidad de información “redundante”. La aplicación de voz limita el ancho de banda de la señal de entrada a un máximo de 3,4 kHz por lo que al aplicar el Teorema de Nyquist el flujo de datos del codificador PCM es de 64.000 bits por segundo. En el caso del sonido, se limita el ancho de banda de entrada a unos 22 kHz. Si se representan las muestras con 16 bits y se consideran dos canales (para el estéreo) se requerirían $22.000 \times 2 \times 16 \times 2 = 1.408.000$ bits por segundo, lo cual es adecuado para la primera generación de CD (libro rojo) a 1,4 Mbps en la velocidad 1X. Sin embargo, este ancho de banda es prohibitivo para radiodifusión por lo que se requiere de otros algoritmos que provean niveles de compresión más elevados. La codificación ATRAC se basó en el aprovechamiento de las limitaciones del oído humano para obtener altos índices de compresión utilizando codificación con pérdidas. El modelo psico-acústico utilizado fue posteriormente incorporado en los algoritmos de codificación de la familia de estándares MPEG. De allí que surja la necesidad de evaluar la calidad de audio comprimido sin la dependencia del ser humano, como resultado de mediciones intrínsecas al compresor.

PLANTEAMIENTO DEL PROBLEMA

La idea básica es determinar el umbral de sensibilidad del oído humano a cada frecuencia para poder cuantificar el grado de pérdidas permisible en dicha frecuencia. Esto se apoya en que el funcionamiento del oído humano está basado en

un análisis frecuencial por medio de los llamados “pelos acústicos” que flotan en la endolinfa. Cada una de estas células ciliadas resuena a una frecuencia diferente y su respuesta es enviada independiente al cerebro a través del nervio acústico. El modelo psico-acústico aprovecha el efecto de saturación o enmascaramiento que sufren estas células cuando están presentes en el sonido recibido otras componentes frecuenciales de alto nivel cercanas a su frecuencia de resonancia. Como este efecto de enmascaramiento depende del sonido recibido, el mismo varía constantemente y debe ser calculado en tiempo real para poder establecer los umbrales de audición que a su vez permitan asignar la cuantización a emplear. Dicho de otra manera, la determinación de la máscara permite asignar el número de bits con que se va a codificar una frecuencia en particular. La señal de entrada (que es PCM) se filtra a través de un banco de 32 filtros pasa banda y cada una de estas salidas, después de decimada y escalada, es codificada con un número de bits que depende de la relación SMR (Signal to Mask Ratio). De esta manera cada componente frecuencial es codificada con el mínimo número posible de bits sin que la degradación (ocasionada por el ruido de redondeo) sea perceptible por el oído humano por hallarse por debajo del umbral de audición. Utilizando estos algoritmos ha sido posible la compresión del sonido hasta de 12:1 para el popular formato MP3. Es importante resaltar que el estándar no especifica la implementación de cada uno de los bloques funcionales, sino solamente la estructura de los datos y las reglas de decodificación los mismos, por lo que los desarrolladores están en libertad de inventar constantemente nuevos algoritmos en procura de una mejora continua de la calidad del sonido reproducido y de mayores relaciones de compresión.

Al igual que la voz, la estimación de la calidad del sonido requiere del componente subjetivo y se expresa también en términos del M.O.S. Por la misma razón, no tiene sentido una comparación directa de las formas de onda a la entrada y a la salida del sistema, porque hay que incluir el procesamiento de las señales por el usuario humano.

Sin embargo, debe recalcar que el M.O.S. es una medida estadística y por tanto involucra un gran número de personas y puede ser muy costoso en tiempo y dinero por lo que en algunas situaciones no resulta conveniente su aplicación. En algunas aplicaciones se requiere de una herramienta capaz de producir un parámetro equivalente al M.O.S. pero sin necesidad de recurrir a un oyente humano. Esto es la motivación del presente trabajo especial de grado.

OBJETIVO GENERAL

Desarrollar una herramienta de software capaz de producir un parámetro representativo de la calidad de un sistema de compresión de audio que sea equivalente al M.O.S.

OBJETIVO ESPECÍFICO

Desarrollar algoritmos que emulen las funciones del oído-cerebro humano por medio del software MATLAB para investigar sobre los parámetros más representativos de la calidad de un sistema de compresión de audio.

METODOLOGÍA

- 1) Estudiar los mecanismos sensoriales humanos para poder aprovechar sus limitaciones a la hora de diseñar algoritmos de compresión del sonido.
- 2) Estudiar en detalle los estándares de compresión MPEG referentes al audio. En particular se estudiará el modelo psico-acústico empleado en cada uno de ellos.
- 3) Estudiar las señales de prueba utilizadas para el MOS en los sistemas de audio.
- 4) Implementar un modelo psico-acústico por medio del software MATLAB.

- 5) Con este modelo hacer análisis de las señales de prueba para extraer algunas máscaras representativas.
- 6) Hacer pasar las señales de prueba por otros compresores de audio y observar las señales reconstruidas.
- 7) Por medio del modelo psico-acústico realizar comparaciones entre las señales a la entrada y a la salida del codificador MP3 para investigar cuales parámetros serán los más convenientes para definir la calidad del sistema de codificación.
- 8) Someter la herramienta a prueba con otros codificadores de sonido para probar su consistencia.
- 9) Elaborar el informe final.

LIMITACIONES Y ALCANCES

El trabajo de investigación solo se limita a realizar el diseño de la solución para el codificador de audio psicoacústico y la implementación de un prototipo funcional para la evaluación de audio comprimido.

Debido a que el software se desarrolla bajo un estándar con *copyright* ISO no será posible la publicación de estas normas en los anexos. Todo lo correspondiente a la normas ISO junto con su software y/o datos han debido ser cancelados directamente a la ISO.

El software fue desarrollado siguiendo la norma y no hubiera sido posible llevarlo a términos efectivos sin la consulta adicional bibliográfica sobre autores especializados en audio, por lo tanto se considera un trabajo de autoría propia, el cual será publicado en los anexos con fin evaluativo.

La plataforma de evaluación de audio se desarrolló siguiendo algunas ideas propuestas por algunas recomendaciones ya existentes que poseen alto grado de confiabilidad. Asimismo, su publicación en los anexos tendrá fines meramente evaluativos.

CAPÍTULO I

EL MODELO PSICOACÚSTICO

La Psicoacústica es una rama de la psicofísica que estudia la relación existente entre el estímulo de carácter físico y la respuesta de carácter psicológico que el mismo provoca. Estudia la relación entre las propiedades físicas del sonido y la interpretación que hace de ellas el cerebro. [6]

1. Métodos y procedimientos en Psicoacústica

La Psicoacústica es una disciplina empírica. Los resultados se obtienen estadísticamente a partir de los resultados concretos de los experimentos realizados con cada uno de los sujetos del experimento. Si los resultados son muy dispares, no es posible extraer conclusiones. [6]

El diseño del experimento en sí y las condiciones en las que se realiza son críticas para la obtención de resultados válidos. Todo un conjunto de resultados obtenidos puede ser cuestionado si el diseño del experimento no fue lo suficientemente cuidadoso como para tener en cuenta los diferentes factores que podrían influir sobre los resultados. [6]

A continuación se describen algunos métodos utilizados en la psicoacústica, debiendo señalarse que cada uno de ellos será más apropiado para un tipo de investigación y que los diferentes métodos implicarán también diferentes demoras en el logro de resultados satisfactorios. [6]

(a) Método de ajuste: El sujeto tiene control sobre el estímulo.

(b) Método de seguimiento (tracking): El sujeto tiene control sobre el estímulo, pero sólo sobre la dirección en la cual varía el estímulo. (Seguimiento de Békésy.)

(c) Estimación de magnitud: A los estímulos se les asignan números correspondientes a la magnitud percibida en alguna de las dimensiones posibles.

(d) Procedimiento Sí-No: El sujeto debe decidir si una señal está presente o no. De alguna manera es un procedimiento de elección forzada (elección forzada de un intervalo y dos alternativas), dado que el sujeto sólo puede contestar "Sí" o "No", y no por ejemplo "No Sé".

(e) Elección forzada de dos intervalos: Al sujeto se le presentan dos intervalos y debe decidir si la señal ocurre en el primero o el segundo intervalo.

(f) Procedimientos adaptables: En los procedimientos de elección forzada la sucesión de estímulos es escogida por el investigador. En estos casos, los estímulos presentados dependen de las respuestas que vaya dando el sujeto.

(g) Comparación de pares de estímulos: Un par de estímulos AB tiene diferencias en una dimensión, mientras que el siguiente par CD tiene diferencias en otra dimensión. El sujeto debe decidir si la diferencia percibida en el primer par AB es mayor o menor que la percibida en el segundo par CD.

La realización de experimentos a efecto de obtener valores y escalas que pudieran reflejar las características del funcionamiento del sistema auditivo es sólo una parte del trabajo en psicoacústica. La otra componente importante es el diseño de modelos que ayuden a explicar los resultados del experimento, es decir, el funcionamiento del sistema auditivo. La determinación de modelos de funcionamiento del sistema auditivo tiene una estrecha relación con los estudios del funcionamiento del cerebro.[6]

2. Experimentos realizados.

Los modelos psicoacústicos se componen a partir de las percepciones de un grupo de personas entrenadas para rendir al máximo en este campo. Por medio de una serie de experimentos se puede determinar la sensibilidad del oído humano a una serie de fenómenos, de forma que aparezcan resultados útiles para el tratamiento del sonido, como ya se ha visto. Las tres siguientes características de la audición se acompañan del experimento que sirve para cuantificarlas. [7]

2.1 Sensibilidad del oído humano.

Experimento: Se sitúa a la persona aislada de otros sonidos y se le ofrece un tono de 1 kHz al nivel mínimo de sonido posible. Se eleva el volumen hasta que sea justo perceptible. Variar la frecuencia y representar en la gráfica. A este patrón se le denomina umbral absoluto (*Absolute Threshold*) o umbral de silencio (*Quiet Threshold*).[7]

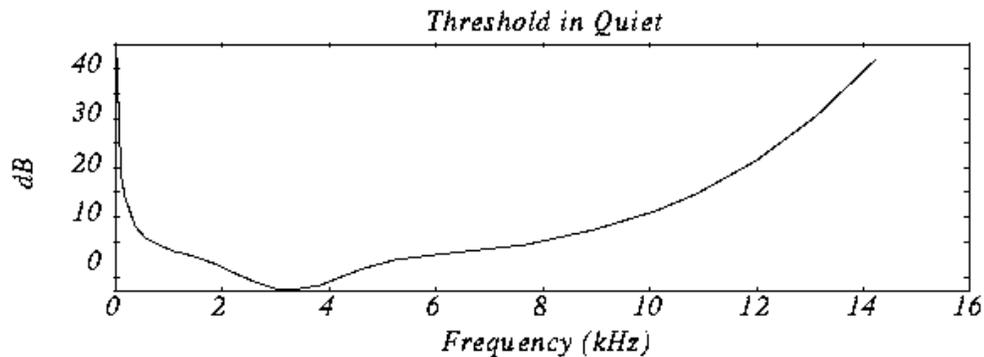


Figura 1. Sensibilidad del oído humano en función de la frecuencia. Umbrales absolutos.
Fuente: Sergio Mangiaterra, Audio-mpeg

Esta curva se ha logrado estimar por medio de la siguiente expresión [18]:

$$Tq(f) = 3,64 \left(\frac{f}{1000} \right)^{-0,8} - 6,5e^{-0,6 \left(\frac{f}{1000} - 3,3 \right)^2} + 10^{-3} \left(\frac{f}{1000} \right)^4 \quad (dB) \quad (f \text{ en kHz.}) \quad \text{Ec. 1.1}$$

2.2 Enmascaramiento en frecuencia.

Experimento: en las mismas condiciones, se ofrece un tono de 1 kHz (enmascarador) a un volumen determinado (60 dB, por ejemplo). Se da un sonido de prueba a 1,1 kHz y se eleva su volumen hasta que sea justo perceptible. Se varía la frecuencia del tono de prueba y trazar la gráfica del límite de audición, ver figura 2. Se repite para todas las frecuencias de tonos de prueba necesarias.[7]

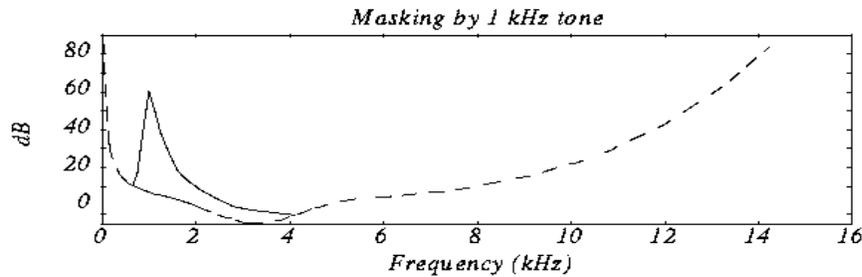


Figura 2. Enmascaramiento en frecuencia del tono de 1 kHz. Fuente: Sergio Mangiaterra, Audio-mpeg

2.3 Enmascaramiento temporal.

Para las siguientes pruebas, se pre-definen los siguientes conceptos. Existen dos tipos básicos de enmascaramiento: el enmascaramiento simultáneo y el no simultáneo. En el simultáneo el sonido de prueba y el enmascarador coinciden temporalmente. En el caso del enmascaramiento no simultáneo, el sonido de prueba puede ser anterior (pre-enmascaramiento) o posterior (post-enmascaramiento) al enmascarador. También puede suceder que el sonido de prueba continúe después de haberse apagado el enmascarador. También en ese caso recibe el nombre de post-enmascaramiento. [6]

Experimento: en las mismas condiciones, se ofrece un tono enmascarador de 1 kHz a 60 dB y un tono de prueba de 1,1 kHz a 40 dB. El tono de prueba no es audible. Parar el tono enmascarador y, con un pequeño retraso, el de prueba. Se

aumenta el retraso hasta que se distinga el tono de prueba. Ver figura 3.[7]

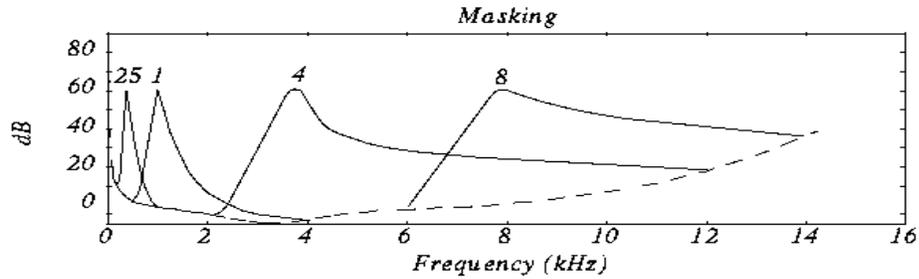


Figura 3. Enmascaramiento en frecuencia de múltiples tonos. Fuente: Sergio Mangiaterra, Audio-mpeg

Si se realiza a la inversa, el enmascaramiento pre-temporal también existe, aunque en un tiempo muy inferior, casi despreciable. Se repite para diversas frecuencias y volúmenes y se representa. Ver figura 4. [7]

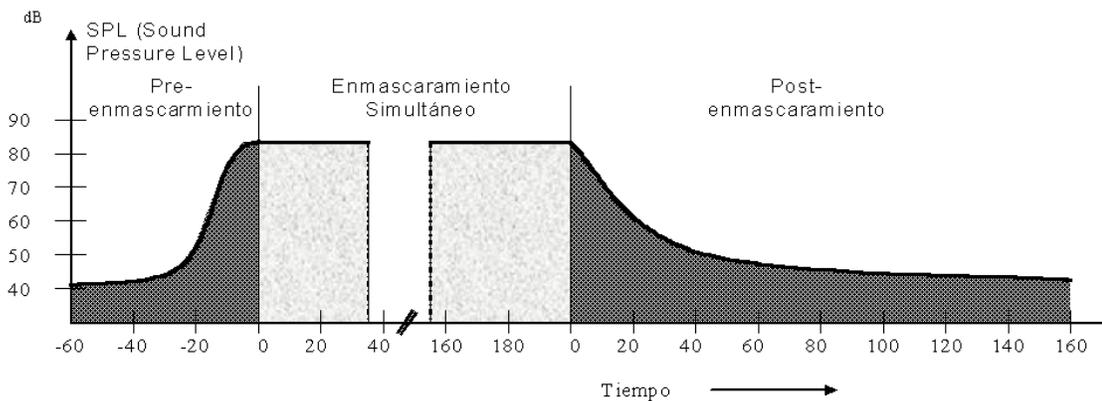


Figura 4. Efectos del pre y post enmascaramiento. (Tiempo en ms.) Fuente: Sergio Mangiaterra, Audio-mpeg

De estos experimentos se obtienen los siguientes parámetros: (a) Percepción y (b) Umbrales de audición, los cuales serán los principales parámetros del oído a modelar matemáticamente.

3. Percepción

La percepción auditiva está basada en el análisis de bandas críticas en la parte más interna del oído donde la transformación de frecuencia-espacio tiene lugar

en la membrana basilar. El espectro de potencia no está representado en una escala lineal de frecuencias sino sobre bandas limitadas de frecuencia denominadas bandas críticas¹. [6]

El sistema de audición puede ser descrito como un banco de filtros pasa banda, constituidos por filtros pasa bandas fuertemente solapados con anchos de bandas del orden de los 100 Hz para frecuencias por debajo de 500 Hz y del orden de 5 kHz para señales de frecuencias superiores. Hasta los 24 kHz son 26 las bandas críticas a ser tenidas en cuenta. [6]

3.1 Enmascaramiento simultáneo.

El enmascaramiento simultáneo es un fenómeno que tiene lugar en el dominio de la frecuencia donde las señales de bajo nivel (por ejemplo podría ser un tono puro) puede volverse inaudible al oído humano si simultáneamente una señal más fuerte (por ejemplo un ruido de banda angosta) está lo suficientemente próximo en frecuencia. El sonido de referencia y el enmascarador coinciden temporalmente. Se puede ver en la figura 5. [6]

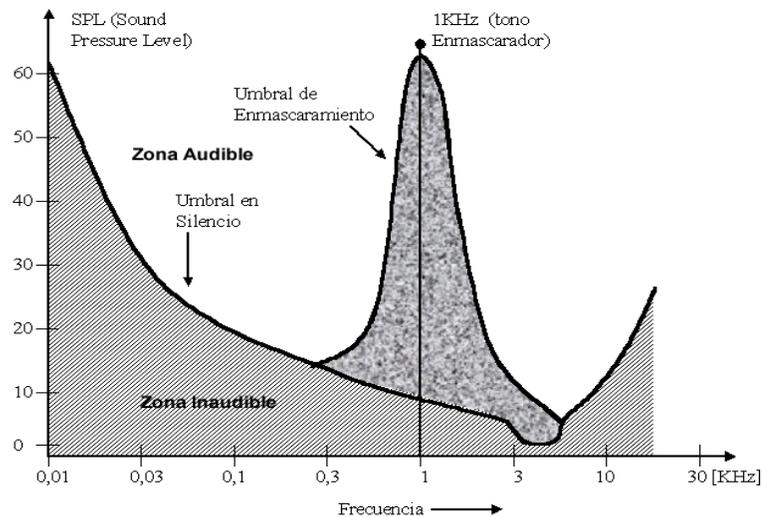


Figura 5. Enmascaramiento simultáneo. Fuente: Sergio Mangiaterra, Audio-mp3

¹ Ver anexo 2. Caracterización del oído humano. (8. Bandas críticas)

Por ejemplo, para un tono de 1 kHz, un umbral de enmascaramiento puede ser medido, de manera que cualquier señal debajo del mismo no será audible. El umbral de enmascaramiento depende del nivel de presión sonora (SPL) y de la frecuencia de la señal enmascarada. [7]

En este caso un tono de 1 kHz con una presión sonora de 60 dB presenta un umbral de enmascaramiento como el que se observa en la figura 5, por lo que cualquier señal cuyo nivel de presión sonora (SPL) se ubique por debajo de este umbral no será percibida por el oído humano. Cabe destacarse que la pendiente de la curva descrita por el umbral de enmascaramiento es más pronunciada hacia las bajas frecuencias (de lo que se deduce que frecuencias cercanas más altas que la enmascaradora pueden ser más fácilmente enmascarables). Por otra parte, la distancia entre el nivel de enmascaramiento y el umbral de enmascaramiento, es menor cuando un ruido enmascara un tono que a la inversa.[7]

Las contribuciones del ruido y de señales de bajo nivel son enmascaradas dentro y fuera de una determinada banda crítica si su SPL está por debajo del umbral de enmascaramiento. Los aportes producidos por el ruido pueden deberse al ruido de codificación, distorsión por *aliasing* y ruidos de transmisión. [7]

Sin la presencia del enmascarador, una señal será inaudible si su SPL está por de bajo del umbral de silencio, el cual depende de la frecuencia y recorre un rango dinámico de 60 dB como se puede observar en la curva inferior de la figura 5. En la figura 6 se pueden ver más detalles del umbral de enmascaramiento.[6]

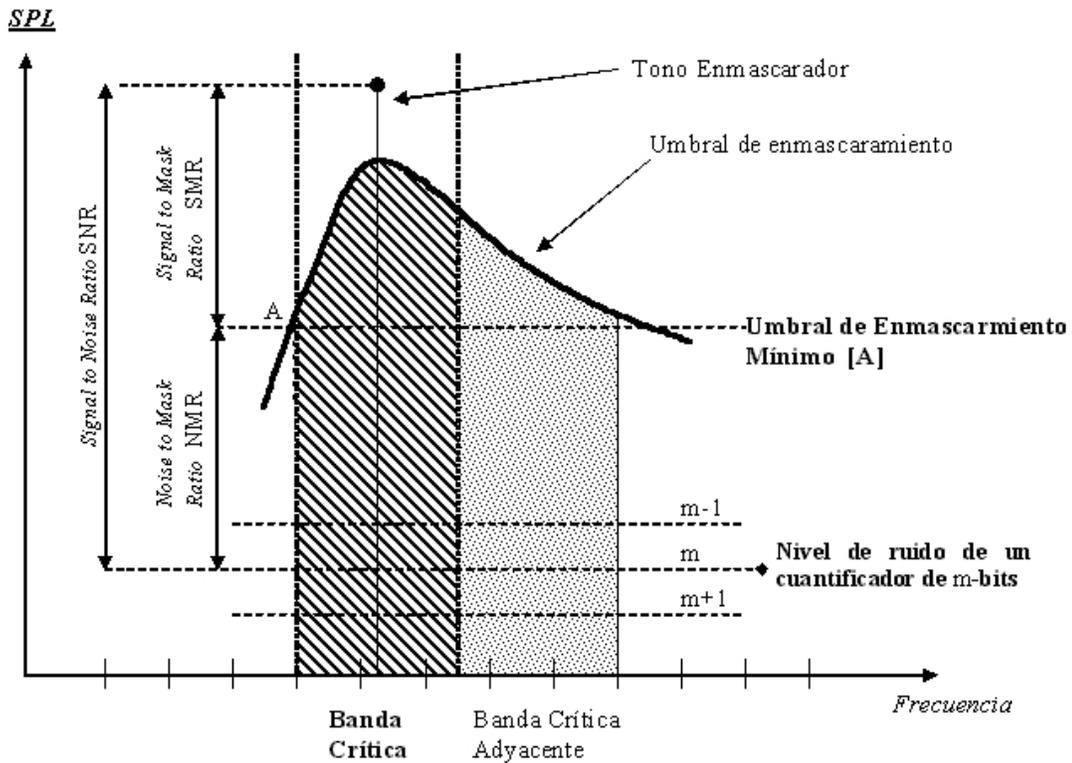


Figura 6. Umbral de enmascaramiento. Fuente: Sergio Mangiaterra, Audio-mpeg

La distancia, o diferencia de potencia, expresada en decibeles, entre el nivel de la señal enmascaradora (un tono en el ejemplo) y el umbral de enmascaramiento mínimo se la denomina Relación Señal Máscara o Relación de señal a enmascaramiento (Signal to Mask Relation SMR). El valor máximo se halla en el borde izquierdo de la Banda Crítica (BC) (punto A). Dentro de la BC el ruido de codificación no será audible en tanto la Relación Señal-Ruido (SNR) sea mayor que la SMR. [7] [2]

Considérese por ahora a la SNR como la relación resultante de una cuantificación de m-bits, luego la distorsión perceptible en una sub-banda dada es entonces medida como la Relación Ruido Máscara o Relación Ruido a Enmascaramiento (Noise to Mask Relation NMR), para un cuantificador de m-bits. Por tanto es posible definir un umbral de enmascaramiento crítico para una banda

crítica, debajo de la cual todas las señales de alto nivel que se encuentran dentro de la banda se hacen inaudibles por efecto de la señal más intensa. [2,7]

La Relación Ruido a enmascaramiento (NMR) se indica mediante:

$$NMR = SNR - SMR \rightarrow SMR = SNR - NMR \quad (\text{dB SPL}) \quad \text{Ec. 2.2}$$

La NMR describe la diferencia entre el ruido de codificación en una determinada sub-banda y el nivel donde una distorsión se vuelve audible; este valor (en dB) debe ser negativo para que el ruido de cuantificación sea inaudible. [2,7].

Hasta ahora se ha descrito el enmascaramiento producido por una sola señal enmascaradora. Si la fuente de señal consta de numerosas señales enmascaradoras el Umbral Global de Enmascaramiento puede ser calculado como el aporte que realizan el resto de las bandas sobre la considerada obteniéndose el nivel de distorsiones perceptibles en función de la frecuencia. [7]

El cálculo del umbral global de enmascaramiento esta basado en la suma (en dB.) de los enmascaramientos individuales de todas las bandas comprendidas en la señal de audio o voz, suficiente para el análisis basado en bandas críticas y es determinado en audio a través de la Transformada Rápida de Fourier (FFT) de orden 512 o 1024. [7]

En un primer paso, los umbrales de enmascaramiento individuales son calculados dependiendo del tipo de enmascarador (tono o ruido) y rango de frecuencia [7]. Este rango de frecuencia depende de cómo el oído percibe todo el ancho de banda, por medio de la membrana basilar y separada en bandas críticas, tal como ya se ha expuesto². Luego, el umbral total de enmascaramiento es calculado por adición de los niveles individuales de enmascaramiento a lo que se le adiciona también el umbral de silencio (esto hace que se asegure que el umbral global de

² En el capítulo III, 1.6.2 , la variable \mathbf{vf} parametriza esta característica dentro del modelo Psicoacústico.

enmascaramiento no se ubique nunca por debajo del de silencio).[7] Los efectos del enmascaramiento, sobrepasando los límites de las bandas críticas, deben ser incluidos en el cálculo. Por último la Relación Señal Mascara Total (TSMR) se determina como la relación entre el máximo de potencia de la señal y el umbral de enmascaramiento Global (o como la diferencia en dB) (ver figura 6) [7]

3.2 Ruido usado para mediciones psicoacústicas.

El ruido blanco tiene una densidad espectral independiente de la frecuencia. Su rango (a los efectos de mediciones en psicoacústica) va de los 20 Hz a los 20 kHz. Se usa también el ruido rosado, el cual tiene una distribución pareja de la energía en cada una de las octavas. El ruido de enmascaramiento uniforme se obtiene modificando el ruido blanco en función de la curva de enmascaramiento (ver figura 7). [6]

3.2.1 Enmascaramiento de sonidos senoidales por ruido de banda ancha

La figura 7 muestra el umbral de audibilidad para un sonido senoidal en función de su frecuencia, enmascarado por ruido blanco (ruido de banda ancha). La línea punteada representa el umbral de audibilidad en silencio. [6]

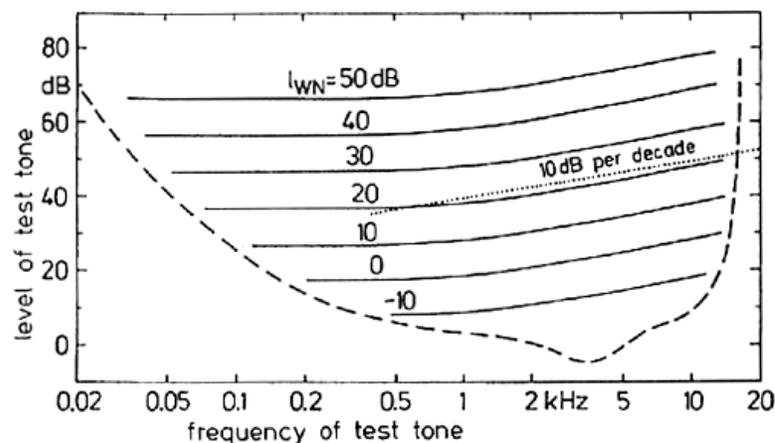


Figura 7. Enmascaramiento de sonidos senoidales por ruido de banda ancha. Fuente: E. Zwicker, H. Fastl: Psychoacoustics. Facts and models) Lwn=Level of white noise

Las curvas de los umbrales indicados para diferentes intensidades del ruido son planas sólo hasta los 500 Hz. Por encima de esa frecuencia las curvas se tuercen, con una pendiente de 10 dB por cada multiplicación por diez de la frecuencia. Resulta interesante observar que la fuerte dependencia de la frecuencia observada en el umbral de audibilidad en silencio, desaparece cuando se enmascara un sonido con ruido de banda ancha. [6] Necesario para algunas mediciones, el ruido de enmascaramiento uniforme se obtiene aplicando un filtro que compense las curvas observadas en la figura anterior, de manera que se obtiene una curva totalmente plana. [6]

3.2.2 Enmascaramiento de sonidos senoidales por ruido de banda angosta

El ruido de banda angosta es un ruido con un ancho de banda igual o menor al ancho de banda crítica. [6]

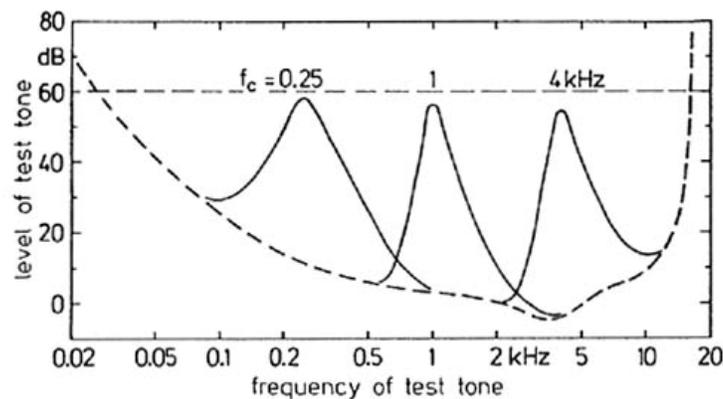


Figura 8. Sonido senoidal enmascarado por ruidos de banda angosta. Fuente: E. Zwicker, H. Fastl: Psychoacoustics. Facts and models)

La figura 8 muestra los umbrales de sonidos senoidales enmascarados por ruidos con ancho de banda igual a la banda crítica, con frecuencias centrales de 250 Hz, 1 kHz y 4 kHz. La intensidad del ruido es siempre 60 dB. Mientras que la forma de la curva alrededor de la frecuencia central es muy similar para 1 y 4 kHz, la misma no lo es en el caso del ruido con frecuencia central en 250 Hz. También puede

observarse que el nivel del umbral disminuye a medida que aumenta la frecuencia: -2 dB para 250 Hz, - 3 dB para 1 kHz y -5 dB para 4 kHz. [6]

La figura 9 muestra el umbral de un sonido senoidal enmascarado por un ruido de banda angosta con ancho igual a la banda crítica con frecuencia central igual a 1 kHz pero con niveles variables. [6]

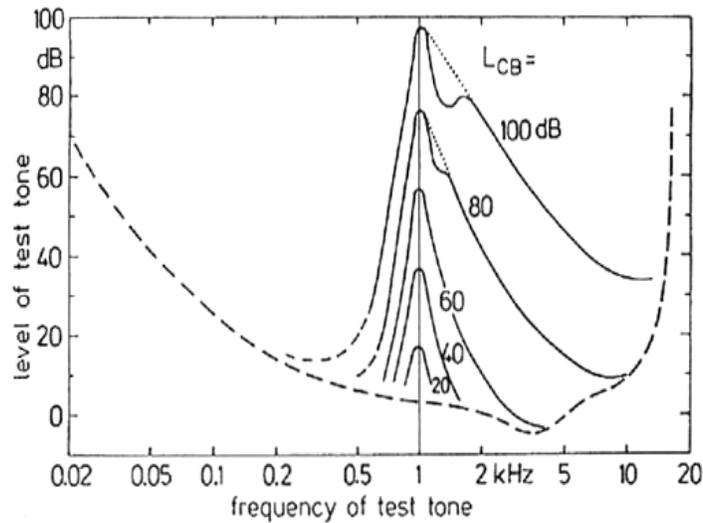


Figura 9. Sonido senoidal enmascarado con niveles variables. Fuente: E. Zwicker, H. Fastl: Psychoacoustics. Facts and models)

Puede observarse una pendiente muy pronunciada para frecuencias por debajo de la frecuencia central del ruido, pero que la pendiente para frecuencias superiores a la frecuencia central del ruido se hace cada vez menos pronunciada a medida que aumenta el nivel del ruido. El umbral está siempre 3 dB por debajo del nivel del ruido. [6]

3.3 Enmascaramiento de sonidos senoidales por sonidos senoidales

La figura 10 muestra el umbral de sonidos senoidales enmascarados por otro sonido senoidal con una frecuencia de 1 kHz, para diferentes niveles del sonido senoidal enmascarador.[6]

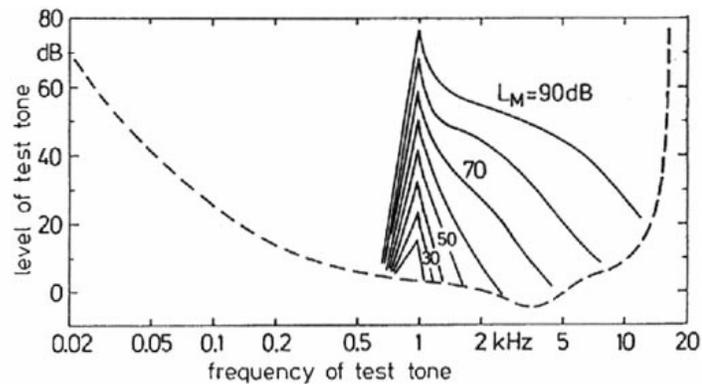


Figura 10. Enmascaramiento de sonidos senoidales por sonidos senoidales. Fuente: E. Zwicker, H. Fastl: Psychoacoustics. Facts and models

Para medir exactamente el umbral se presentan dos tipos de problemas. El primero es que cuando el sonido senoidal de referencia y el sonido senoidal enmascarador tienen frecuencias cercanas el escucha percibe la fluctuación en amplitud (igual a $f_2 - f_1$), lo que dificulta el establecimiento del umbral. [6]

El otro problema es que para algunas frecuencias se hace notorio el sonido de combinación (sonido diferencial), por lo que a un oyente no experimentado le cuesta diferenciar entre la percepción del sonido senoidal de referencia o el diferencial. Esto es particularmente notorio para frecuencias del sonido senoidal de referencia de 1.4 kHz. [6]

3.4 Enmascaramiento de sonidos senoidales por sonidos complejos

La figura 11 muestra el umbral de sonidos senoidales enmascarados por un sonido armónico complejo con una frecuencia fundamental de 200 Hz y otros 9 parciales, todos con la misma amplitud y con fases escogidas aleatoriamente. Se indican umbrales para niveles de 40 y 60 dB del enmascarador. [6]

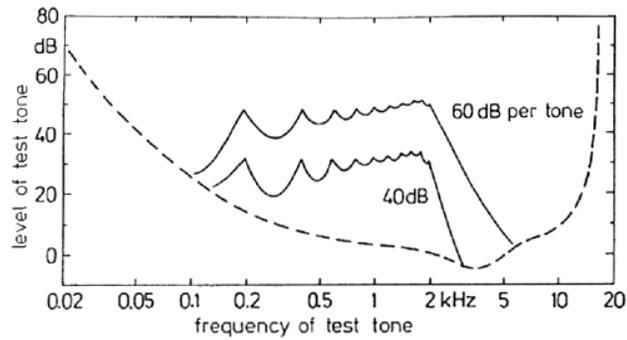


Figura 11. Enmascaramiento de sonidos senoidales por sonidos complejos. Fuente: E. Zwicker, H. Fastl: Psychoacoustics. Facts and models)

Para este caso de sonido complejo la igualdad de fase de sus componentes llevaría a que la envolvente dinámica del sonido comenzara a adquirir la característica de un impulso, lo que podría tener consecuencias adicionales en la determinación de los umbrales. [6]

3.5 Enmascaramiento temporal

Junto con el enmascaramiento simultáneo, otros dos fenómenos juegan un importante rol en la percepción del audio en el ser humano: Pre-enmascaramiento: El sonido de referencia precede al enmascarador. Post-enmascaramiento: El sonido de referencia sucede al enmascarador o el sonido de referencia subsiste después de haberse apagado el enmascarador. En ambos casos de enmascaramiento no simultáneo el sonido de prueba (es decir, el sonido enmascarado) debe ser un sonido de tipo impulso. [7]

Los efectos de enmascaramiento temporal tienen lugar antes y después que una señal comience y termine. Los efectos del pre-enmascaramiento tienen una duración de una décima parte del post-enmascaramiento y según lo determinan nuevos estudios es bastante menor que ese valor. [7]

La duración del post-enmascaramiento está en el orden de los 50 a 200 ms. Ambos efectos serán explotados en los algoritmos de codificación de ISO/MPEG.

3.5.1 Pre-Enmascaramiento

Aún cuando no sea fácilmente imaginable, sonidos que aún no existen pueden enmascarar sonidos ya existentes. No obstante, el fenómeno del pre-enmascaramiento se reduce a lapsos sumamente reducidos, aproximadamente en el orden de los 20 ms posteriores al sonido de prueba. El fenómeno del pre-enmascaramiento dependería de la relación de tiempo entre el sonido de prueba y el enmascarado (los 20 ms mencionados) y no parecería haber influencia del nivel del sonido enmascarador sobre el desplazamiento del umbral. [7]

La explicación podría ser que los sistemas físicos no realizan saltos de tipo abrupto sino que realizan más bien transiciones continuas. De esa manera es posible pensar que el tiempo de ataque de un hecho sonoro (que puede ser del orden de los 20 ms) pueda ser la causa del fenómeno del pre-enmascaramiento. De cualquier manera no se tiene una comprensión cabal del fenómeno del pre-enmascaramiento. Incluso con sujetos entrenados ha sido difícil obtener resultados reproducibles. [7]

3.5.2 Post-Enmascaramiento

El post-enmascaramiento ocurre también para sonidos de prueba de relativamente corta duración (pocas centenas de ms) y está limitado a sonidos que ocurren inmediatamente después del enmascarador (< 200 ms). Ver figura 12.

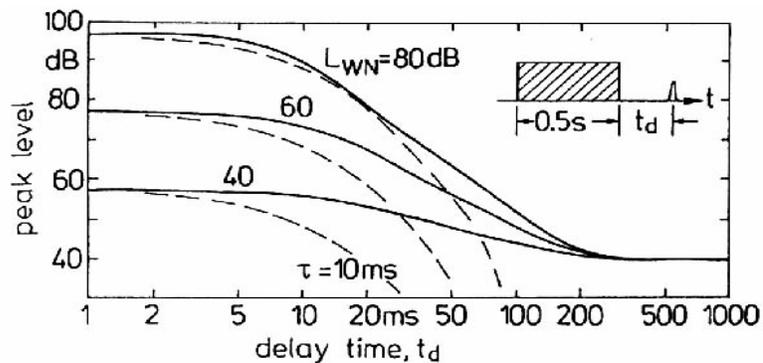


Figura 12. Post-enmascaramiento. Fuente: E. Zwicker, Psychoacoustics. Facts and models.

Lo que se observa en la figura 12 es que durante los primeros 5 ms después de que se apaga el enmascarador prácticamente no se produce ninguna diferencia con respecto al enmascaramiento simultáneo. Luego de esos 5 ms el umbral de enmascaramiento decrece hasta igualarse con el umbral de audibilidad (en silencio) aproximadamente a los 200 ms.

La siguiente figura muestra el post-enmascaramiento producido en función de la duración del sonido enmascarador. Las curvas representan el enmascaramiento para sonidos enmascaradores de 200 ms y de 5 ms de duración.

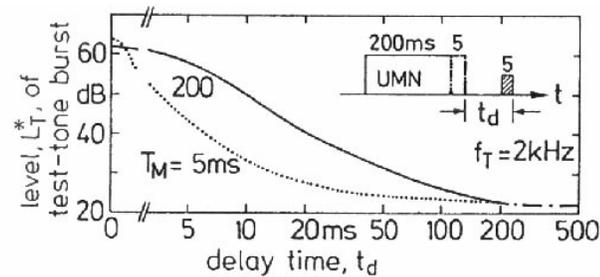


Figura 13. Post-enmascaramiento producido en función de la duración del sonido enmascarador. Fuente: E. Zwicker, H. Fastl: Psychoacoustics. Facts and models

4. Umbrales de la audición

4.1 Umbrales absolutos

Los umbrales absolutos de la audición son aquellos valores de uno de los parámetros del estímulo físico a partir del cual la sensación comienza o deja de producirse. El umbral superior de frecuencias es dependiente de la edad. Con el paso del tiempo se deterioran las células capilares del órgano de Corti, lo que tiene como consecuencia que cada vez se perciba menos las frecuencias agudas.[6] Ver figura 14.

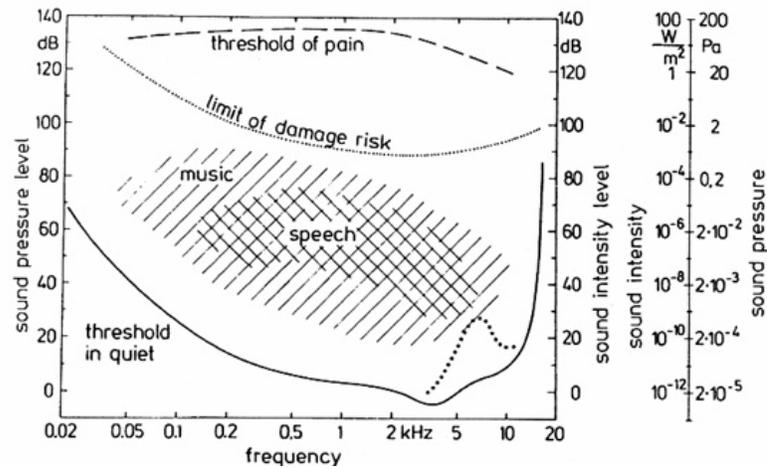


Figura 14. Umbrales de audición. Fuente: E. Zwicker, H. Fastl: Psychoacoustics. Facts and models)

Nótese en la figura que el límite máximo del oído de SPL está alrededor de 100 dB.

4.1.1 Umbral de audibilidad

El umbral de audibilidad está definido por la mínima intensidad o presión necesarias para que un sonido pueda ser percibido.

De la figura 15 se observa que el umbral de audibilidad no depende sólo de la intensidad o presión, sino que también es dependiente de la frecuencia del sonido senoidal de prueba. El sistema auditivo humano tiene un área de mayor sensibilidad entre los 500 y los 3000 Hz., producida principalmente por las curvas de respuesta del sistema auditivo periférico (oído externo, medio e interno). [6]

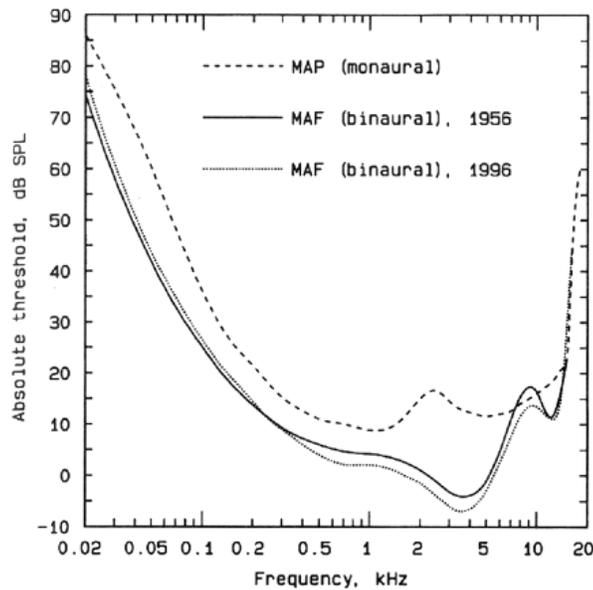


Figura 15. Umbral de audibilidad en detalle. Fuente: B.J.C. Moore: An Introduction to the Psychology of hearing.

Las curvas de la figura 15 muestran dos formas diferentes de medir el umbral de audibilidad, la mínima presión audible (MAP) y el mínimo campo audible (MAF). La mínima presión audible (MAP) se mide colocando pequeños micrófonos dentro del canal auditivo. La información (señal de prueba) es enviada, por lo general, por medio de auriculares. En el caso del mínimo campo audible (MAF) la medición se realiza en ausencia del sujeto, en cámaras anaecoicas, colocando un micrófono en el centro mismo de donde se encontraba la cabeza del sujeto. [6]

Las diferencias fundamentales entre una curva y otra (la zona entre los 1.5 y los 6 kHz) están dadas principalmente por las resonancias producidas en el pabellón y el canal auditivo externo. El oído externo aumenta la presión sonora en el tímpano en unos 15 dB para frecuencias entre 1,5 a 6 kHz. La transmisión del oído medio es más eficiente para frecuencias medias. [6]

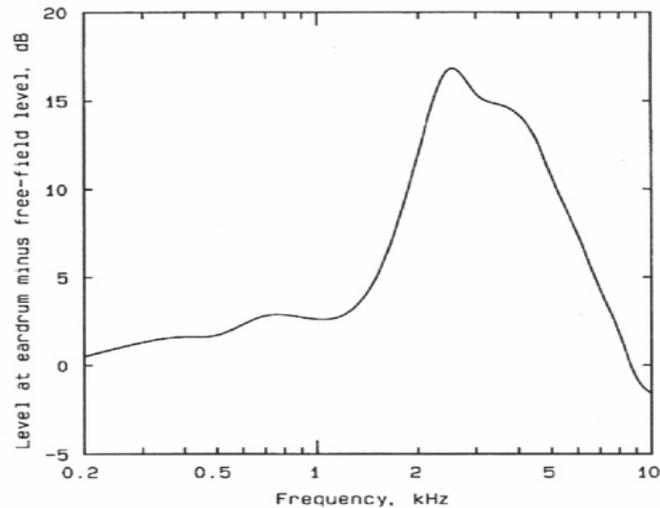


Figura 16. Diferencia entre niveles de presión en el oído humano. Fuente: B.J.C. Moore: An Introduction to the Psychology of hearing.

La figura 16 muestra las diferencias entre el nivel de presión sonora en el tímpano y el nivel de presión sonora en el campo libre, producto de lo expresado anteriormente. Originalmente (curvas calculadas por Fletcher y Munson) el umbral de audibilidad había sido definido como la mínima presión necesaria para percibir un sonido senoidal de 1 kHz. La presión necesaria para ello es de $2 \cdot 10^{-5} \text{ N/m}^2$ (o una intensidad de 10^{-12} W/m^2), valor tomado además como referencia para la determinación de valores absolutos. Es decir, el umbral de audibilidad es de 0 dB para 1 kHz. Sin embargo, cálculos más recientes de las curvas (Robinson y Dadson) mostraron que, si se mantiene el valor de $2 \cdot 10^{-5} \text{ N/m}^2$ como valor de referencia, el umbral de audibilidad es de + 3 dB para 1 kHz. [6]

4.1.2 Umbral de frecuencia

Por lo general se toman los valores 20 Hz y 20.000 Hz como los umbrales de frecuencia de la audición. Es decir, el sistema auditivo humano no percibe señales con frecuencias menores a los 20 Hz o mayores a los 20 kHz. En otra literatura

pueden encontrarse los valores 16 Hz y 16 kHz. [6]

El umbral superior de frecuencias es dependiente de la edad. Con el paso del tiempo se deterioran las células capilares del órgano de Corti, lo que tiene como consecuencia que cada vez se perciban menos las frecuencias agudas. La exposición prolongada a sonidos dañinos puede contribuir a acelerar esta pérdida de percepción de las frecuencias más agudas. [6]

4.2 Umbrales diferenciales

Los umbrales diferenciales de la audición señalan las mínimas variaciones de uno de los parámetros del estímulo físico, necesarias para que se produzca un cambio en la sensación. [6]

Existe una distinción entre umbrales de mínima variación perceptible (MVP) y umbrales de mínima diferencia perceptible (MDP). Los primeros (MVP) se miden variando uno de los parámetros de un sonido (por ejemplo, mediante modulación de amplitud o de frecuencia), mientras que los segundos (MDP) se miden presentando dos señales diferentes al sujeto. [6]

4.2.1 Umbrales diferenciales de intensidad

La figura 17 muestra el umbral de mínima variación perceptible de intensidad, para una frecuencia de modulación de 4 Hz, para un sonido senoidal de 1 kHz y para ruido blanco, en función del nivel de presión sonora del sonido de prueba. [6]

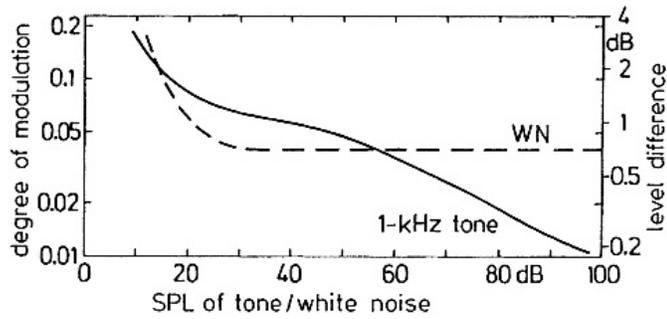


Figura 17. Umbral de mínima variación perceptible de intensidad. Fuente: E. Zwicker, H. Fastl: Psychoacoustics. Facts and models.

La figura 18 muestra los umbrales de mínima diferencia perceptible para diferentes frecuencias (representativas de las distintas frecuencias) en función del nivel de presión sonora del sonido de prueba. [6]

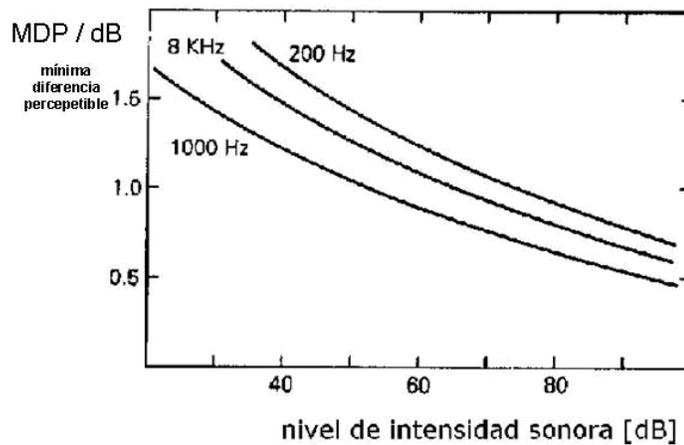


Figura 18. Umbrales de mínima diferencia perceptible para diferentes frecuencias. Fuente: D. Hall: Musical Acoustics.

4.2.2 Umbrales diferenciales de frecuencia

La figura 19 muestra el umbral de mínima variación perceptible en función de la frecuencia del sonido senoidal de prueba. [6]

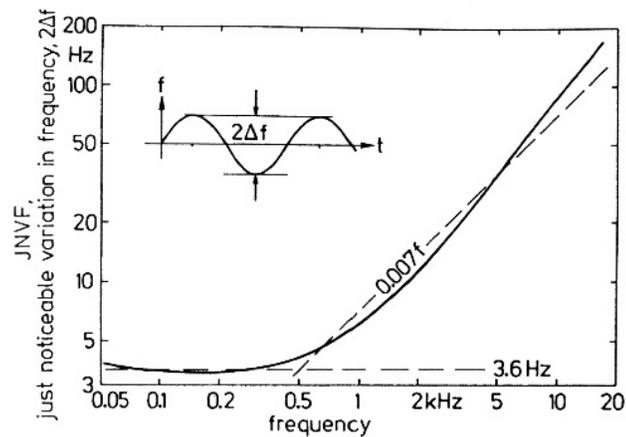


Figura 19. Umbral de mínima variación perceptible en función de la frecuencia del sonido senoidal de prueba. Fuente: E. Zwicker, H. Fastl: Psychoacoustics. Facts and models.

Mientras que la figura muestra una MVP más o menos constante de 3,6 Hz hasta los 500 Hz, a partir de allí el umbral aumenta con una pendiente aproximada de $0.007 \cdot f$, lo que implica que se pueden percibir variaciones del 0,7% de la frecuencia. [6]

La figura 20 muestra el umbral de mínima diferencia perceptible en Hz en función de la frecuencia del sonido senoidal de prueba. [6]

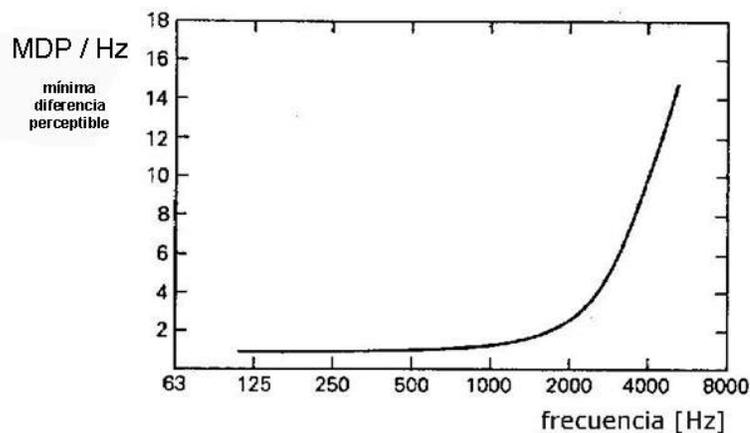


Figura 20. Umbral de mínima diferencia perceptible en Hz en función de la frecuencia del sonido senoidal de prueba. Fuente: D. Hall: Musical Acoustics

La figura muestra que el sistema auditivo es capaz de percibir diferencias de frecuencia de aproximadamente 1 Hz hasta los 500 Hz, mientras que a partir de allí la curva crece con una pendiente de aproximadamente $0.002 \cdot f$, lo que implica percibir diferencias de frecuencia de aproximadamente 0.2%. [6]

Es de notar que los umbrales de mínima variación perceptible y de mínima diferencia perceptible en frecuencia están representados prácticamente por la misma curva, en una relación aproximada de 3:1 (aproximadamente de 3,6 Hz en el de MVP y 1 Hz en el umbral de MDP), y con una diferencia en la pendiente de la curva por encima de los 500 Hz.

Las consecuencias de esto es que por encima de los 5 kHz se pierde la noción de altura, es decir, el sistema auditivo no es capaz de determinar la altura de los sonidos. La mayoría de los sonidos usados comúnmente en música tiene solamente componentes que aparecen en la forma de parciales por encima de los 5 kHz, y no como frecuencias fundamentales. Tómese en cuenta que el DO más agudo en el piano tiene una frecuencia fundamental de 4.224 Hz. La determinación de altura allí se hace difícil.

5. Codificación basada en la percepción

5.1 Codificación de audio de alta calidad

La función básica de los sistemas de codificación perceptiva de audio es comprimir los datos de audio digital en una de tal forma que:

- (a) La compresión sea tan eficiente como sea posible, en este caso el archivo resultante comprimido es tan pequeño como se pueda y
- (b) El audio reconstruido (decodificado) suena exactamente (o tan cerca como sea

posible) al audio original antes de la compresión.

Otros requisitos para las técnicas de compresión de audio incluyen:

(c) La baja complejidad para facilitar el software decodificador o hacer más económico el hardware decodificador y con bajo nivel de consumo de energía y

(d) la flexibilidad suficiente para los diferentes posibles escenarios de aplicación. [9]

El tercer requisito tiene que ver con la capacidad de procesamiento y de transmisión del sistema digital. La técnica que permite lograr estos criterios es la llamada Codificación Perceptiva (*Perceptual Encoding*) y requiere del conocimiento de la Psicoacústica para lograr el objetivo de una compresión eficiente y audición reconocible. La codificación perceptiva es una técnica de compresión con poca pérdida, dado que el archivo decodificado no es una réplica bit a bit de los datos de audio originales. Los codificadores perceptuales para codificación de audio de alta calidad han sido tema de investigación desde la última parte de los años setenta, con una mayor incidencia a partir de 1.986.

El estándar de codificación MPEG es el objetivo de estudio de este proyecto porque alcanza este desempeño notable al explotar dos características psicoacústicas del sistema auditivo humano:

5.1.1 Bandas críticas

Como ya se ha visto en el capítulo destinado a revisar las características auditivas, el oído interno representa el espectro de potencia de señales entrantes sobre una escala no lineal en la forma de bandas de frecuencia limitadas denominadas bandas críticas. La banda de frecuencia audible, que se extiende hasta los 20 kHz, cubre hasta 25 bandas críticas (ver anexo 2. Caracterización del oído humano. Tabla 1), cuyos anchos de banda individuales aumentan con la frecuencia. Como ya se ha determinado, el sistema auditivo puede modelarse como un banco de filtros pasa

banda, compuesto por 25 filtros pasa banda que se traslapan y que tienen anchos de banda menores de 100 Hz para las frecuencias audibles más bajas y hasta 5 kHz para las más altas.

5.1.2 Enmascaramiento auditivo

Este fenómeno ya ha sido explicado dentro de las características auditivas y formas parte del sistema de percepción de oído. El enmascaramiento auditivo es un fenómeno en el dominio de la frecuencia y es más notable cuando ambas señales se encuentran en la misma banda crítica y menos efectiva cuando están en bandas vecinas.

Con estos basamentos en la Psicoacústica del sistema auditivo humano se puede describir la operación del estándar de codificación de audio MPEG.

5.2 Principios básicos de la codificación de audio perceptiva.

Tal como se muestra en el diagrama de bloque, figura 21, el sistema básico de codificación perceptiva, consiste en el siguiente:

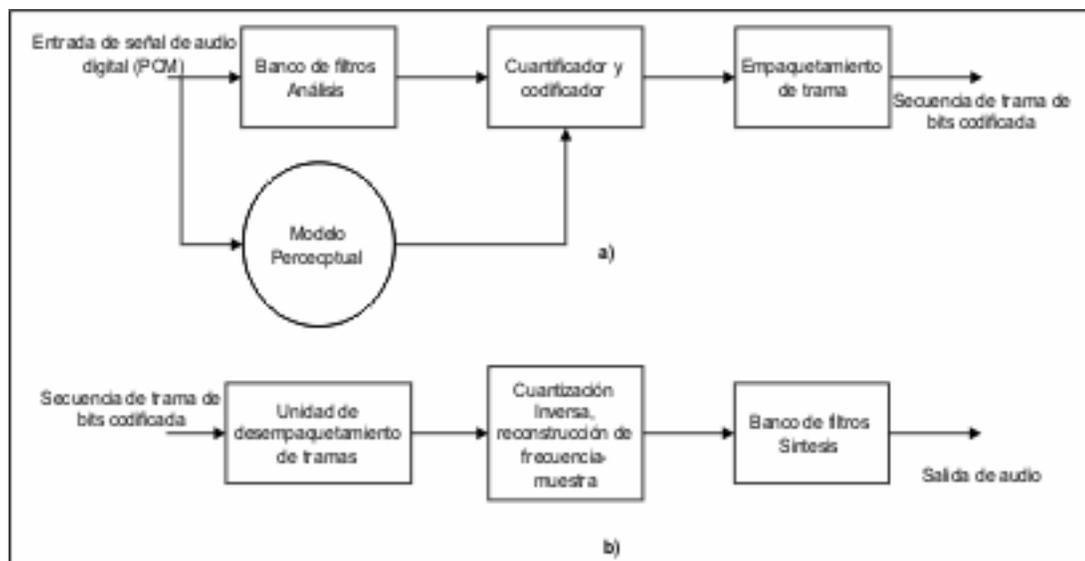


Figura 21. Diagrama de bloque sistema de codificación de audio perceptivo MPEG a)

5.2.1 Banco de filtros: Un banco de filtro es usado para descomponer la señal de entrada en componentes espectrales sub-muestreadas en el dominio de tiempo y frecuencia. Es decir, la señal de audio se descompone en múltiples sub-bandas para la codificación. Este conjunto de filtros se realiza en tres capas denominadas I, II y III, las cuales son de complejidad, retardo y desempeño perceptivo subjetivo crecientes. [9]

El algoritmo en la capa I utiliza un banco de filtros pasa banda que divide la señal de audio en 32 sub-bandas de ancho constante; este banco de filtros se encuentra también en las capas II y III con ciertas variantes, manteniendo un compromiso entre la eficiencia del cómputo y el desempeño perceptual. El algoritmo en la capa II es un simple mejoramiento de la capa I, mejora el desempeño de la compresión codificando los datos en grupos más grandes. El algoritmo de la capa III es mucho más refinado en cuanto a que se diseña para conseguir resoluciones de frecuencia más cercanas a las particiones entre las bandas críticas. Más adelante se explicará con detalle cada una de ellas. Junto con el correspondiente banco de filtros en el decodificador se forma el sistema de análisis y síntesis. [2] [9]

5.2.2 Modelo Perceptivo o psicoacústico: Usando la señal de entrada en el dominio del tiempo y/o la salida del análisis del banco de filtro, dependiendo de la frecuencia y del tiempo, se calcula un estimado del actual umbral de enmascaramiento, usando reglas conocidas por la psicoacústica. A esto se le denomina Modelo Perceptivo del sistema de codificación perceptivo o Modelo Psicoacústico, el cual es el componente clave en el codificador. Su función estriba en analizar el contenido espectral de la señal de audio de entrada y, por ello, en calcular la relación señal a enmascaramiento para cada sub-banda en cada una de las tres capas. [2] [9]

5.2.3. Cuantificación y codificación: La información proveniente del modelo psicoacústico la usa el cuantificador-codificador para decidir cómo distribuir el número de bits disponible para la cuantificación de las señales de sub-banda. Dicha asignación dinámica de bits se realiza de tal modo que se minimice la audición del

ruido de cuantificación. Los componentes espectrales son cuantificados y codificados teniendo en cuenta el ruido, el cual es introducido por la cuantificación, por debajo del umbral de enmascaramiento. Dependiendo del algoritmo, este paso es realizado en diferentes formas, desde un simple bloque de compansión (compresión y expansión) hasta un sistema de análisis por síntesis usando compresión adicional sin ruido. Los codificadores de audio que utilizan el Modelo Perceptivo generalmente calculan la cantidad de ruido por banda crítica que puede ser enmascarado. [1] Esto significa que se oculta el ruido de cuantificación con la máscara Psicoacústica. [2] [9]

5.2.3.1 El codificador.

Procesa la señal digital de audio (PCM) y genera el *bitstream* comprimido para el almacenamiento. El algoritmo del codificador no está estandarizado y pueden ser utilizados varios medios, tales como la estimación del umbral de enmascaramiento auditivo, cuantificación y establecimiento de la escala (escalado) de la señal. De cualquier modo, la salida del codificador debe ser tal que el decodificador que cumpla con las especificaciones de la cláusula 2.4 de la norma MPEG ISO/IEC 11172, producirá una señal de audio compatible para la aplicación propuesta.

Por esta razón el algoritmo del codificador es lo que será objeto de estudio en este proyecto.

5.2.3.2. Decodificación

El decodificador debe procesar el *bitstream* para reconstruir la señal de audio digital. La especificación de este elemento sí esta totalmente definida y debe seguirse en todos sus puntos, según las normas MPEG.

Los datos del *bitstream* son desempaquetados para recuperar las diversas partes de la información. El bloque de reconstrucción recompone la versión cuantificada de la serie de muestras mapeadas. El “mapeador” inverso transforma estas muestras de nuevo a PCM.

5.2.3.3 Formato y empaquetamiento de la trama de bits

Por último, la unidad de empaquetamiento de tramas ensambla las muestras de audio cuantificadas en una secuencia de bits decodificables. Para ensamblar el arreglo de bits generado se utiliza un “formateador” de arreglo de bits, el cual típicamente consiste en los coeficientes espectrales cuantificados y codificados y alguna información paralela como puede ser la información de localización de bits, logrando una corriente de bits decodificable.

De esta forma el modelo Psicoacústico sólo es necesario en el codificador. El decodificador recibe simplemente las operaciones de procesamiento de señales que se realizan en el codificador, convirtiendo la secuencia recibida de bits codificados en una señal de audio en el dominio del tiempo. [2] [9]

En los codificadores basados en la percepción, el proceso de codificación esta controlado por la Relación Señal a Máscara Global (GSMR) versus la curva de Frecuencia. Si la tasa de bits necesaria para un enmascaramiento completo de la distorsión esta disponible, el esquema de codificación será transparente (la señal decodificada será indistinguible de la señal fuente). Cabe señalar que la tasa de bits a la salida del codificador se mantiene constante, pudiendo darse casos donde la tasa disponible excede los requerimientos de la codificación (y por ende la codificación se efectúa en forma óptima). En otros casos la tasa disponible a la salida es insuficiente para las pretensiones del codificador (en cuyo caso la transmisión incluirá ruido de codificación). Si la tasa de bits para un enmascaramiento completo no esta disponible, entonces el umbral de enmascaramiento global sirve como una función de ponderación del error espectral; el espectro de error resultante tiene la forma del umbral de enmascaramiento global.

En los diseños prácticos de codificación basada en la percepción, no se puede sobrepasar los límites establecidos por el enmascaramiento dado que el procesamiento posterior de la señal acústica (por ejemplo filtrados en ecualizadores)

por el usuario final y los múltiples procesos de codificación y decodificación pueden desenmascarar el ruido. Además, dado que el conocimiento acerca del enmascaramiento auditivo es muy limitado, el modelo auditivo usado para en el diseño de un modelo particular de codificador basado en la percepción no es lo suficientemente preciso, por lo que, como un requerimiento adicional, se necesita un margen de seguridad en los diseños prácticos de codificadores.

CAPÍTULO II

PROCEDIMIENTOS BÁSICOS GENERALES A LA ENTRADA DEL CODIFICADOR MPEG

1. Codificador MPEG

1.1 Generalidades

El banco de filtro realiza un mapeo de tiempo a frecuencia. En la norma ISO-11172 se utilizan 2 tipos de bancos de filtros: (a) un banco del tipo polifásico para las capas I y II y (b) un banco del tipo híbrido-polifásico (con MDCT) para la capa III.

Ambos proveen un mapeo específico del dominio tiempo a frecuencia. Los bancos hacen un muestreo crítico (hay tantas muestras en el dominio frecuencia como en el dominio tiempo) y proveen la primera separación de frecuencia en el codificador y los filtros de reconstrucción en decodificador. Las muestras de salida de los bancos de filtros son cuantificadas.

En las capas I y II, se utiliza un banco de filtros con 32 sub-bandas. En cada sub-banda, 12 o 36 muestras son agrupadas para el procesamiento. En la capa III, el banco de filtros tiene una resolución dependiente de la señal, donde puede haber 6x32 muestras frecuenciales (bloque corto) o 18x32 muestras frecuenciales (bloque largo). En el caso de 6x32 muestras frecuenciales, 3 *sets* de cada frecuencia son cuantificados separadamente. [7]. Esto se logra por medio de la MDCT, la cual puede ser seleccionada para generar cada sub-banda, bien sea 6 sub-bandas o bloques cortos (*short window MDCTs*) o 18 sub-bandas llamados bloques largos (*long-window MDCTs*). Los bloques largos proporcionan una mejor resolución temporal. También durante los estados transitorios de la señal, las 2 sub-bandas más bajas en la MDCT pueden ser seleccionadas para que sean ventanas largas, mientras las 30 restantes son seleccionadas a ventanas cortas. Esto ayuda a la adaptación dinámica para la codificación mientras se mantiene una mejor resolución para las bajas frecuencias. [4]

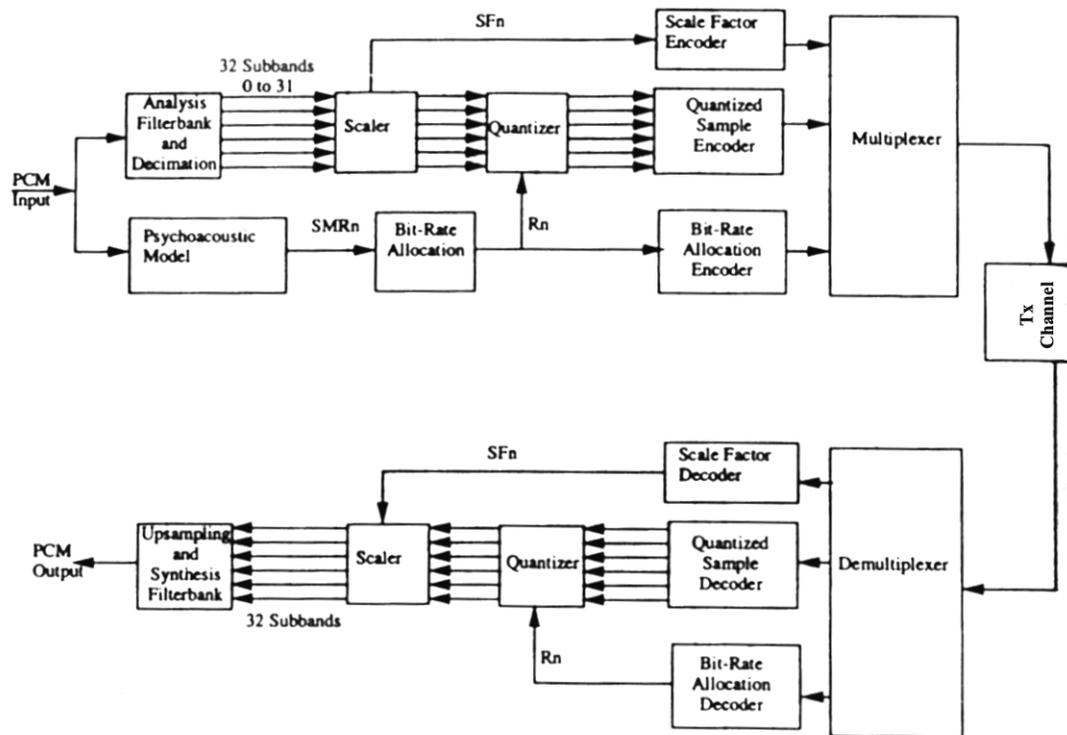


Figura 22. Ejemplo de esquema del codificador y decodificador de audio MPEG-1 para capa I y II (un solo canal). Fuente: [4]

1.2 Filtro Pasa Alto de entrada

El algoritmo de codificación provee una respuesta en frecuencia que arranca desde DC. De cualquier modo, en aplicaciones donde éste no es un requerimiento se recomienda la utilización de un filtro pasa alto que debe incluirse en la entrada del codificador. La frecuencia de corte del filtro debe estar entre 2 y 10 Hz. La utilización de dicho filtro evita un requerimiento innecesario de tasa de bits alta para la sub-banda más baja e incrementa la calidad de audio total.

1.3 Filtro de Análisis de sub-banda

El filtro de análisis de sub-banda se utiliza para partir el ancho de banda de la señal con frecuencia de muestreo $F_s=32$ sub-bandas equiespaciadas con una

frecuencia de muestreo de FS/32. En este proceso se realizan las siguientes operaciones:

- Ingreso de 32 muestras de audio.
- Construcción de un vector X de 512 elementos con las últimas 512 muestras de audio ingresadas. Las 32 muestras de audio ingresadas son desplazadas dentro del vector desde la posición 0 a 31, la más reciente en la posición 0, las últimas 32 se desplazan hacia fuera (se descartan).
- Se enmarca el vector X con el vector C (cuyos coeficientes figuran en la tabla C1 de los anexos) generando el vector z (ver figura 23).
- Se calculan los 64 valores Y_i de acuerdo a la formula dada en el diagrama de flujo.
- Se calculan las 32 muestras de cada sub-banda S_i a través de un matrizado. Los coeficientes para la matriz pueden ser calculados a través de la siguiente fórmula:

$$M_{ik} = \cos\left[\frac{(2i+1)(k-16)\pi}{64}\right] \quad \text{Con } i \text{ desde } 0 \text{ a } 31 \text{ y } k \text{ desde } 0 \text{ a } 63 \quad \text{Ec. 2.1.}$$

En el siguiente diagrama se detallan los pasos donde se transforman las muestras desde el dominio tiempo a las 32 sub-bandas (en el dominio frecuencia) [7].

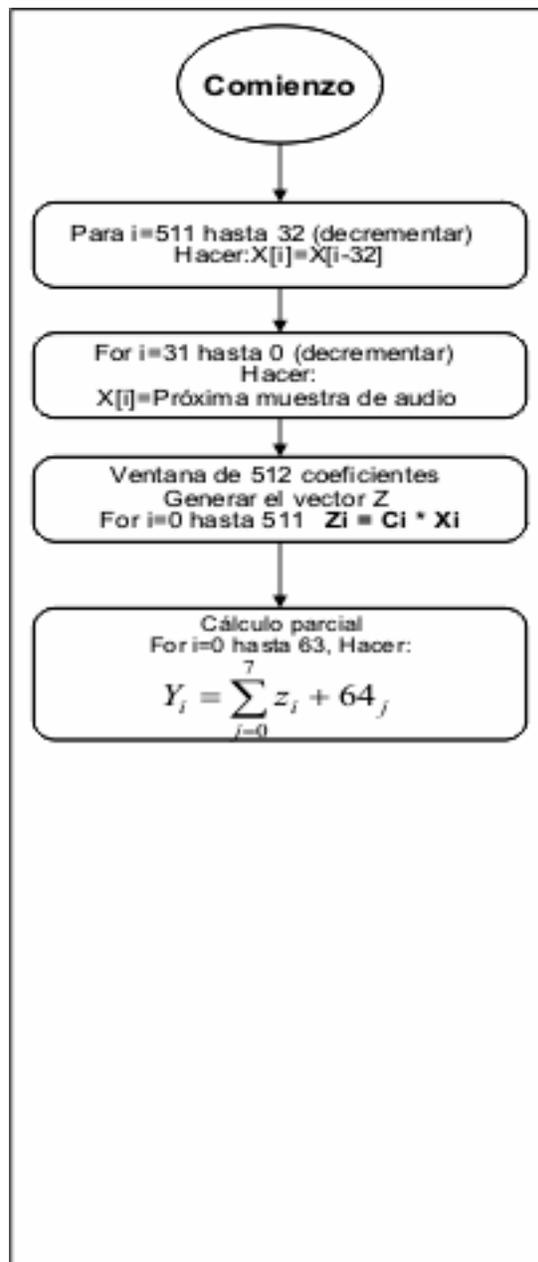


Figura 23. Diagrama de bloques del Filtro de Análisis de Sub-banda. Banco de Filtros Polifásicos. Fuente [12]

1.4 El modelo Psicoacústico

El modelo psicoacústico calcula el nivel de ruido perceptible para cada sub-banda. Este nivel es utilizado en la asignación de bits o ruido para determinar los cuantificadores actuales y niveles de cuantificación. Los modelos psicoacústicos descritos en la norma son dos y, si bien ambos pueden usarse para cualquiera de las capas, el primer modelo se adapta mejor a las capas I y II mientras que el modelo II lo hace mejor para la capa III (con una leve modificación) (ver Capítulo I, Modelo psicoacústico).

En ambos modelos la salida es la Relación Señal Máscara (SMR) para cada banda (en las capas I y II) o grupo de bandas (capa III)

1.5 Asignación de Bits o Ruido

Al *asignador* ingresan tanto las muestras provenientes del banco de filtros y la SMR proveniente del modelo Psicoacústico, y ajusta la asignación de bits (Capas I y II) o la asignación de Ruido (Capa III) de manera de cumplir simultáneamente tanto con el requerimiento de la tasa de bits como también con el requerimiento de enmascaramiento. A bajas tasas estos métodos tratan de dedicar bits de manera que resulte psicoacústicamente tolerable cuando no se puede satisfacer la demanda psicoacústica a la tasa de bits requerida.

1.5.1. Método de asignación de bits o ruido: Existen dos métodos de control de la tasa de bits. En las capas I y II este método es un proceso de asignación de bits (una determinada cantidad de bits se asigna a cada muestra o grupos de muestras) en cada sub-banda.

El método para la capa III es un bucle de asignación de ruido, donde los cuantificadores son variados de un modo organizado y la variable a controlar es el ruido inyectado. En cada caso el resultado es un conjunto de parámetros de cuantificación y las muestras de salida cuantificadas que ingresan al formateador de trama.

1.6 El *formateador* de la trama

El formateador del *bitstream* varía de capa en capa. En las capas I y II se utilizan código PCM fijo para cada muestra de sub-banda, con la excepción que en Capa II las muestras cuantificadas pueden ser agrupadas. En Capa III se utilizan códigos Huffman para representar las muestras de frecuencia cuantificadas. Los códigos Huffman son palabras de longitud variable que permiten una más eficiente representación de las muestras cuantificadas en el *bitstream*, a costo de una mayor complejidad.

El formateador del *bitstream* toma las salidas del banco de filtros cuantificadas junto con la asignación de bits (capas I y II) o la asignación de ruido (capa III) y otra información adyacente (*side info*) requerida y codifica y formatea esa información de una manera eficiente. En el caso de la Capa III, la codificación Huffman se inserta también en este punto.

2. Capa I

El cálculo de los parámetros del modelo psicoacústico puede obtenerse utilizando cualquiera de los modelos detallados en la norma ISO 11172. La extensión (*shift length*) de la FFT es de 384 muestras. Cada modelo provee la SMR para cada sub-banda

El banco de filtro de sub-bandas es como el detallado anteriormente en el punto 1.3. El cálculo de los factores de escala (*Scalefactors: SCF*) se efectúa cada 12 muestras de cada sub-banda. El máximo valor absoluto de estas 12 muestras se determina. Por tratarse de muestras normalizadas estas varían entre -1 y 1 . El mínimo valor de la tabla B1(ver anexos) que resulta mayor que este máximo es usado como **SCF**.

El índice de la tabla B1 está representado por 6 bits, el MSB (*most significant bit*) primero. El **SCF** se transmite sólo si el número de bits asignado a la sub-banda es distinto de cero.

La trama en capa I posee la siguiente estructura:

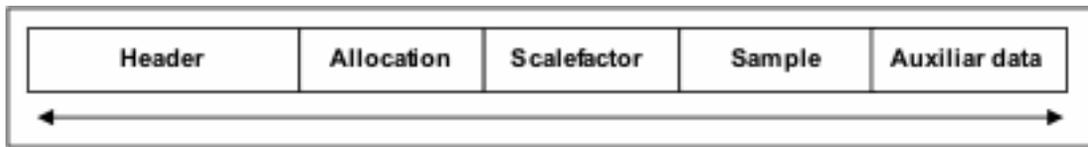


Figura 24. Formato de la trama en la capa I. Fuente [12]

2.1 Asignación de bits

Antes de ajustar a una tasa de bits fija, el número de bits disponibles para codificar las muestras y los SCF deben ser determinados. Este número se determina de la siguiente expresión:

$$\mathbf{Adb} = \mathbf{cb} - (\mathbf{bhdr} + \mathbf{bcr}c + \mathbf{bbal} + \mathbf{banc}) \quad \text{Ec.2.2}$$

donde:

- **Adb**: bits disponibles para la codificación de las muestras
- **Cb**: número total de bits disponibles
- **Bhdr**: cantidad de bits del Header (cabecera) (32 bits)
- **Bcr**c: palabra de control CRC (código de redundancia cíclica) (16 bits)
- **Bbal**: bits utilizados para la asignación
- **Banc**: bits utilizados para el envío de datos auxiliares (no forman parte del audio) en la trama.

El número de bits resultantes es usado para la codificación de las muestras de la sub-bandas y los SCF. El principio utilizado en el procedimiento de asignación es la minimización de la NMR (Relación Ruido Mascara) en la trama, con la restricción que el número de bits utilizado no exceda el disponible para la trama. La cantidad de bits posibles para asignar a una muestra puede encontrarse en el anexo de tablas siendo el rango de 0 a 15 bits (excluyendo la asignación de 1 bit).

El procedimiento de asignación es iterativo, donde en cada paso de iteración se incrementa el número de niveles de las muestras de la sub-banda que mayor beneficio produzca (de acuerdo al modelo psicoacústico).

Primero se calcula **MNR** (relación Máscara-Ruido) para cada sub-banda restando a **SNR** la **SMR**:
$$\mathbf{MNR} = \mathbf{SNR} - \mathbf{SMR} \quad \text{Ec. 2.3}$$

donde la **SNR** puede hallarse en la tabla C2 (ver anexos) y **SMR** es la salida del modelo psicoacústico.

En principio, ningún bit se asigna para las muestras y **SCFs**. El número de bits por muestras (**Bspl**) y el número de bits por SCF (**Bscf**) es inicializado en cero. El procedimiento iterativo comienza y cada bucle de iteración contiene los siguientes pasos:

- Determinación de la mínima **MNR** para todas las sub-bandas. La precisión de la cuantificación de la sub-banda con el mínimo **MNR** se incrementa con el uso del próximo número de bits más alto.
- La nueva **MNR** se calcula.
- **Bspl** (bits por muestra) se actualiza de acuerdo al número de bits adicionales requeridos. Si por primera vez se asigna un número de bits a la sub-banda entonces **Bscf** (bits para SCF) debe incrementarse en 6 bits por lo que los bits disponibles ahora se calcularán:

$$\mathbf{Adb} = \mathbf{cb} - (\mathbf{bhdr} + \mathbf{bcrc} + \mathbf{bbal} + \mathbf{bscf} + \mathbf{bspl} + \mathbf{banc}) \quad \text{Ec. 2.4}$$

El proceso iterativo se repite tantas veces hasta que **Adb** no pueda incrementarse con un nuevo paso de iteración (por el incremento de **Bspl** y **Bscf**).

2.2 Cuantificación y codificación de las muestras de la sub-bandas.

Un cuantificador lineal con una representación de cero simétrico se utiliza para cuantificar las muestras de las sub-bandas. Esta representación evita que pequeños cambios de valor alrededor del cero, cuantificando a diferentes niveles. Cada una de las muestras de las sub-bandas es normalizada dividiendo su valor por el SCF para obtener **X** y cuantificada de acuerdo a los siguientes pasos:

- Cálculo de $\mathbf{AX+B}$

- Tomar los N bits más significativos
- Invertir el MSB

A y B son constantes que se encuentran en la tabla C3 (ver anexos). N representa el número de bits necesarios para codificar el número de pasos (escalones). La inversión del bit más significativo se hace a los efectos de evitar todos '1' en la representación del código, dado que el código con todos '1' se utiliza para la palabra de sincronización.

2.3 La codificación de la asignación de bits.

Se realiza de acuerdo a las tablas impuestas por la norma ISO/IEC 11172.

2.4 Datos Auxiliares

Pueden transmitirse dentro de la trama y su longitud no está acotada, pero éstos reducen la cantidad de bits disponibles para audio, lo que puede resultar en una degradación de la calidad de audio.

La presencia de un patrón de bits dentro de los datos auxiliares que coincida con la palabra de sincronismo puede dañar la sincronización. Esto puede ser más frecuente cuando se utiliza formato libre.

2.5 Formato de los datos.

La información codificada de las sub-bandas se transfiere en tramas. El número de *slots* en una trama varía con la frecuencia de muestreo y la tasa de bits. Cada trama contiene información de 384 muestras de la señal de audio original por lo que la tasa de la trama es de $F_s/384$.

Tabla 1. Tamaño de la trama según la frecuencia. Fuente [13]

F_s (kHz)	Tamaño de la Trama (ms)
48	8
44.1	8,7074
32	12

Una trama puede contener información de uno o dos canales. La longitud de un slot en la Capa I es de 32 bits (4 bytes). El número de *slots* en una trama puede obtenerse de la siguiente manera:

$$\text{Número de slots/trama} = (\text{bitrate}/F_s) * 12 \quad \text{Ec. 2.5}$$

Si el resultado no es un número entero éste se trunca y se requiere de “padding” (relleno). Esto significa que el número de *slots* puede variar entre N y N+1.

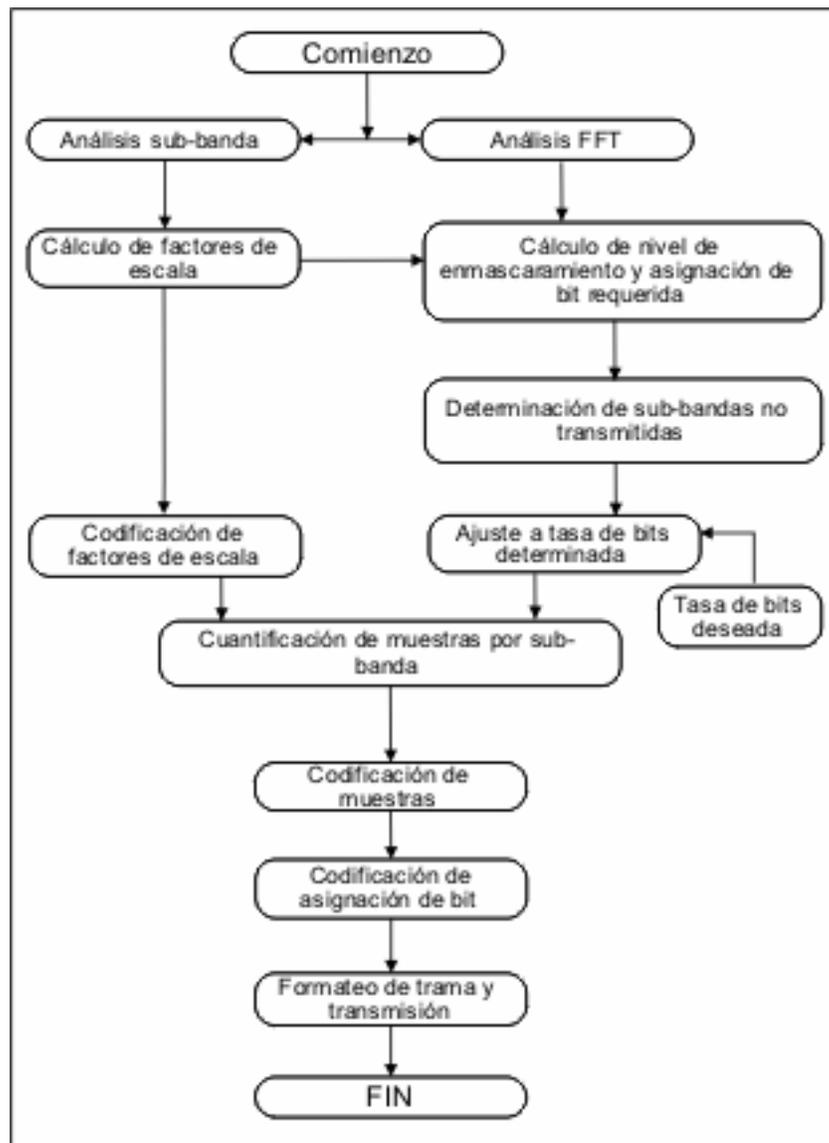


Figura 25. Diagrama de bloques codificador MPEG-1 para Capas I y II. Fuente [12]

3. Capa II

3.1 Modelo Psicoacústico

El modelo psicoacústico puede calcularse con cualquiera de los dos modelos descritos en la norma. Si el modelo 1 se utiliza para calcular los parámetros psicoacústicos, la longitud de ventana de la FFT es de 1152 muestras. Si se utiliza el modelo 2, el cálculo se realiza dos veces con una longitud de ventana (*shift length*) de 576 muestras y el mayor de cada par de SMR se utiliza. Ambos modelos proveen la SMR para cada sub-banda.

El banco de filtro de sub-banda corresponde al descrito anteriormente, en el punto 1.3. El cálculo de los SCF para cada sub-banda se realiza cada 12 muestras. Se determina el máximo valor absoluto de estas 12 muestras y el mínimo valor de la tabla B1 (ver anexos) que sea mayor que este máximo se usa como factor de escala.

3.2 Codificación de los Factores de escala (SCF).

Una trama corresponde a 36 muestras por sub-bandas y por lo tanto contiene tres SCF por sub-banda.

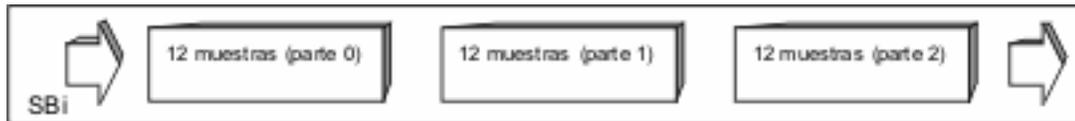


Figura 26. Partición de la trama en capa II. Fuente [12]

Se calculan las diferencias entre los SCF de cada grupo de 12 muestras:

$$Dscf_1 = scf_1 - scf_2 \quad \text{Ec.2.6}$$

$$Dscf_2 = scf_2 - scf_3 \quad \text{Ec.2.7}$$

La clase de cada diferencia se determina como sigue:

Tabla 2. Clase de cada diferencia. Fuente [13]

clase	Dscf
1	Dscf<=-3
2	-3<dscf<0
3	Dscf=0
4	0<dscf<3
5	Dscf>=3

El par de clases de diferencias indican el punto de entrada a la tabla C4 (ver anexos) (Capa II Scalefactors Transmission Patterns). La columna con nombre “*Scalefactor used in the encoder*” da los tres SCF que son actualmente usados. “1”, “2” y “3” significa que el primero, segundo y tercero dentro de una trama, respectivamente. “4” significa el máximo de los tres SCF. Si después de este ajuste de SCF dos o tres son iguales indica que no todos los SCF necesitan ser transmitidos para cierta sub-banda dentro de la trama. Solamente los SCF indicados en la columna de “Transmission Patterns” serán transmitidos. La información que detalla el número y la posición de los SCF en cada sub-banda se denomina “SCF Selection Information”.

3.2.1 Codificación de los SCFSI:

Se codifican utilizando con una palabra de 2 bits que también se encuentra en la tabla C4 (ver anexos). Solamente los SCFSI para las sub-bandas que tienen una asignación de bits distinta de 0 se transmiten.

3.3 Asignación de bits

Antes de ajustarse a una tasa fija de bits, el número de bits disponible (adb) que están disponible para la codificación de las muestras y de los SCF debe ser determinado. Este número puede obtenerse de la siguiente manera:

$$\mathbf{Adb} = \mathbf{cb} - (\mathbf{bhdr} + \mathbf{bcrc} + \mathbf{bbal} + \mathbf{banc}) \quad \text{Ec. 2.8}$$

Donde:

- **Adb**: bits disponibles para la codificación de las muestras
- **Cb**: número total de bits disponibles
- **Bhdr**: cantidad de bits del Header(cabecera) (32 bits)
- **Bcrc**: palabra de control CRC (código de redundancia cíclica) (16 bits)
- **Bbal**: bits utilizados para la asignación
- **Banc**: bits utilizados para el envío de datos auxiliares (no forman parte del audio) en la trama.

El número resultante puede ser utilizado para codificar las muestras de las sub-bandas y los SCF. El principio usado en el procedimiento de asignación es la minimización de la relación ruido a máscara (**NMR**) total sobre la trama con las restricciones que el número de bits utilizados no exceda el número de bits disponible para la esa trama. La tabla B2 (ver anexos) de la norma indica para cada sub-banda el número de posibles escalones que se pueden utilizar para cuantificar las muestras. El número de bits requeridos para representar esas muestras cuantificadas se detalla en la tabla B4 (ver anexos).

La asignación es un procedimiento iterativo donde, en cada paso de iteración el número de niveles de la sub-banda que produzca el mayor beneficio psicoacústico se incrementa.

Primero se calcula la relación NMR para cada sub-banda haciendo:

$$\mathbf{NMR} = \mathbf{SNR} - \mathbf{SMR} \quad (\text{en dB}) \quad \text{Ec.2.9}$$

SNR puede encontrarse en la tabla C5 (ver anexos), mientras que **SMR** es la salida del modelo psicoacústico. Al inicio de la iteración, no se asignan bits para las muestras y **SCF**. El número de bits por muestras (**Bspl**) y el número de bits por SCF (**Bscf**) son inicializados a cero. El procedimiento iterativo comienza y cada bucle de iteración contiene los siguientes pasos:

- Determinación de la mínima **MNR** para todas las sub-bandas.
- La precisión de la cuantificación de la sub-banda con el mínimo **MNR** se incrementa con el uso del próximo número de bits más alto que se indica en la tabla B2 (ver anexos).
- La nueva **MNR** se calcula.

Bspl (bits por muestra) se actualiza de acuerdo al número de bits adicionales requeridos. Si un número de bits distinto de cero se asigna por primera vez a una sub-banda, **bsel** debe ser actualizado y **bscf** es actualizado de acuerdo al número de **SCF** requeridos en la sub-banda, por lo que los bits disponibles ahora se calcularán:

$$\mathbf{Adb} = \mathbf{cb} - (\mathbf{bhdr} + \mathbf{bcrc} + \mathbf{bbal} + \mathbf{bsel} + \mathbf{bscf} + \mathbf{bspl} + \mathbf{banc}) \quad \text{Ec. 2.10}$$

El proceso iterativo se repite tantas veces hasta que **Adb** no pueda incrementarse con un nuevo paso de iteración (por el incremento de **Bsel**, **Bspl** y **Bscf**).

3.4 Cuantificación y codificación de las muestras de la sub-bandas.

Cada una de las 12 muestras por sub-bandas está normalizada, dividiendo su valor por el **SCF** para obtener X y cuantificarlo usando la siguiente fórmula:

- Cálculo de AX+B.
- Tomar los N bits más significativos
- Convertir el MSB

A y B son constantes que se encuentran en la tabla C6 (ver anexos). N representa el número de bits necesarios para codificar el número de pasos (escalones). La inversión del bit más significativo se hace a los efectos de evitar todos '1' en la representación del código, dado que el código con todos '1' se utiliza para la palabra de sincronización.

Dado el número de escalones en que las muestras serán cuantificadas, la tabla B4 (ver anexos), indica si el agrupamiento de muestras tendrá lugar. Si este no será usado, las tres muestras se codifican con palabras de código individuales. Si será usado, se determina que existirá agrupamiento y las tres muestras se codifican en una única palabra de código. Solamente un valor, V_m ($m=3,5,9$), se transmite. La relación entre este valor (V_m) y las tres muestras consecutivas (x,y,z) que son reemplazadas son:

$$V_3 = 9z + 3y + x \quad (v_3 \text{ en } 0...26) \quad \text{Ec. 2.11}$$

$$V_5 = 25z + 5y + x \quad (v_5 \text{ en } 0..124) \quad \text{Ec. 2.12}$$

$$V_9 = 81z + 9y + x \quad (v_9 \text{ en } 0..728) \quad \text{Ec. 2.13}$$

3.5 Asignación de bits.

A los efectos de una codificación más eficiente, solamente se permiten un número limitado de cuantificaciones las cuales pueden ser diferentes para cada sub-banda. Solamente se transmite el índice con la longitud de palabra '**nbal**' en la tabla B2 (ver anexos) del MSB.

3.6 Datos Auxiliares:

Pueden transmitirse dentro de la trama y su longitud no está acotada, pero éstos reducen la cantidad de bits disponibles para audio, lo que puede resultar en una degradación de la calidad de audio. La presencia de una secuencia de bits en los datos auxiliares, igual a la palabra de sincronismo, puede dificultar la sincronización, pero este problema puede ser más frecuente cuando se utiliza el formato libre (Free Format).

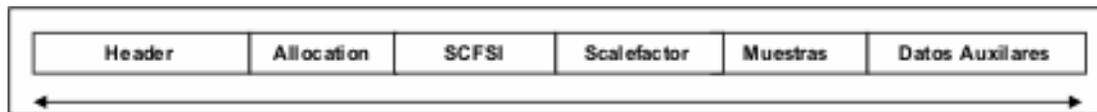


Figura 27. Formato de la trama en la capa II. Fuente [12]

Las diferencias con la capa I son:

- La longitud del slot es de 8 bits (1 byte).
- Un nuevo bloque (SCFSI) se introdujo en la trama.
- La asignación de bits, SCF y muestras han sido sujetas a una codificación más eficiente.

4. Capa III

4.1 Generalidades

La codificación de la capa III introduce muchas características nuevas con respecto a las dos anteriores. El resumen de éstas es:

- Banco de filtros híbrido conmutable (entre dos resoluciones distintas de frecuencia/tiempo).
- Control avanzado de pre-eco (actúa sobre la resolución del banco de filtro híbrido).
- Cuantificación no uniforme.
- Codificación Huffman de las muestras cuantificadas. (Codificación entrópica).

- Una técnica de buffer (denominada *bit reservoir*) que permite un mejor aprovechamiento del bit rate.
- Único de las 3 Capas que soporta codificación con tasa de bits (bit rate) variable.
- Dentro del modo Joint Stereo soporta MS_stereo (Middle Side Stereo) además de Intensity stereo, a diferencia de Capa I y II que solo soportan éste último.

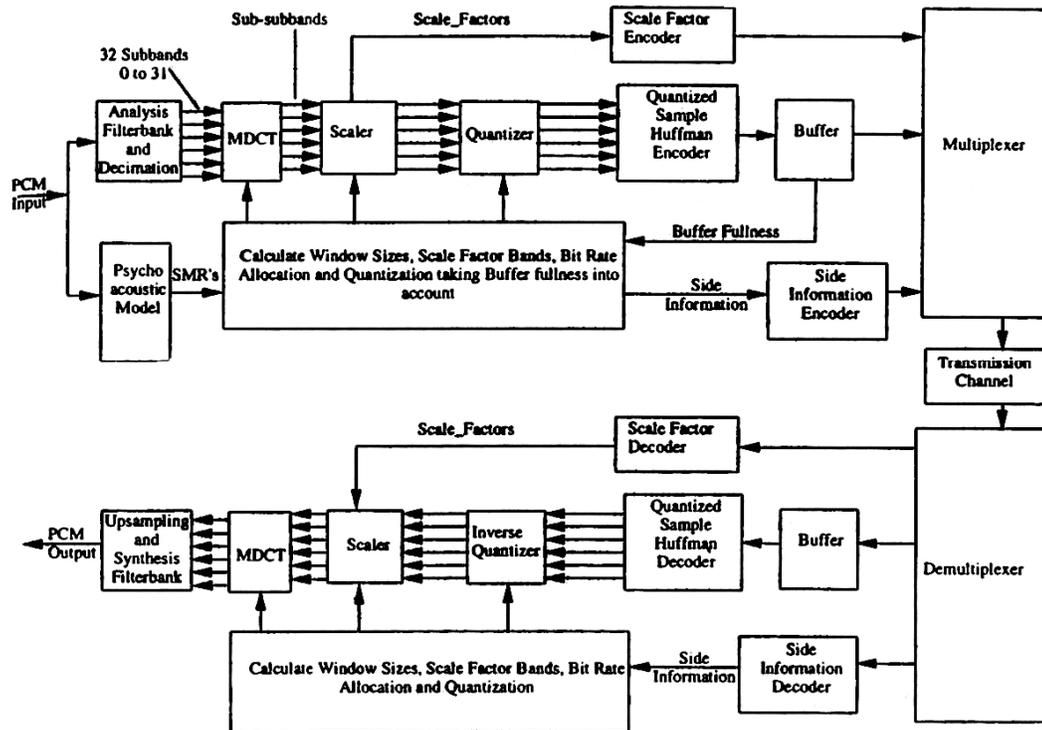


Figura 28. Ejemplo de esquema del codificador/decodificador MPEG 1 Capa III. Fuente: [4]

Las muestras de audio de la entrada se convierten en muestras de sub-banda a través del banco de filtros híbrido que realiza el mapeo de tiempo a frecuencia. La salida de este banco de filtro se aplica a la entrada de MDCT donde a la salida se obtiene una resolución más fina en frecuencia.

De acuerdo a las características psicoacústicas de la señal (la existencia de pre-ecos, o no) se obtendrá a la salida del MDCT una mejor resolución frecuencial (pero pobre resolución temporal) o sea bloques largos (normal); o una mejor

resolución temporal pero más pobre resolución frecuencial (bloques cortos) que permiten enmascarar eficientemente la existencia de pre-ecos.

Una situación de pre-eco se da cuando a un período de poca variación de la señal (bajo requerimiento de bits para codificar) le sigue una variación abrupta de la misma lo que implica una mayor demanda instantánea de bits. Esto se ejemplifica en el siguiente gráfico.

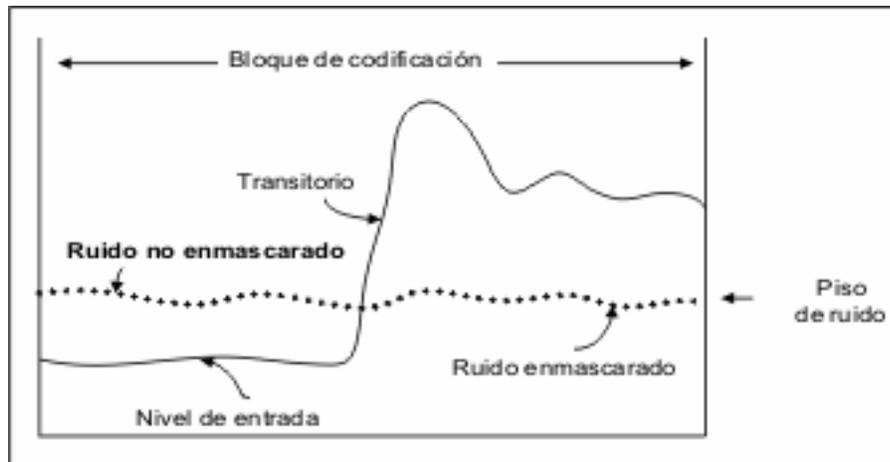


Figura 29. Bloque de codificación con situación de pre-eco. Fuente [7]

A diferencia de capa I o II los *Scalefactors* afectan a un conjunto de líneas espectrales y son gobernados por el modelo psicoacústico (se ajustan más precisamente a las bandas críticas) y existen SCF para bloques cortos y largos.

Los SCF son utilizados para “colorear” el ruido de cuantificación. Si el ruido de cuantificación se colorea con la forma correcta, éste es enmascarado completamente. A diferencia de capa I y II, los SCF de capa III no se refieren al máximo local de la señal cuantificada. En capa III, SCF son utilizados en el decodificador para obtener los factores de división (divisores) para grupos de valores (agrupados en SCF bands). En caso de capa III, estos grupos de valores (SCF bands) se extienden sobre varias líneas espectrales. Estos grupos denominados bandas de factores de escala (SCF bands) son seleccionados de manera de asemejarse a las bandas críticas del oído humano en la forma más cercana posible.

La subdivisión del espectro en SCF bands permanece fija para cada longitud de bloque y frecuencia de muestreo y almacenados en tablas en el codificador y decodificador (tabla B8, anexos). Los SCF son cero (**SCF=0**) para líneas de frecuencia por sobre la línea más alta de la tabla, lo que significa que el factor de multiplicación es 1.

Los SCF son cuantificados logarítmicamente. El escalón de cuantificación se establece con una variable denominada **scalefac_scale**.

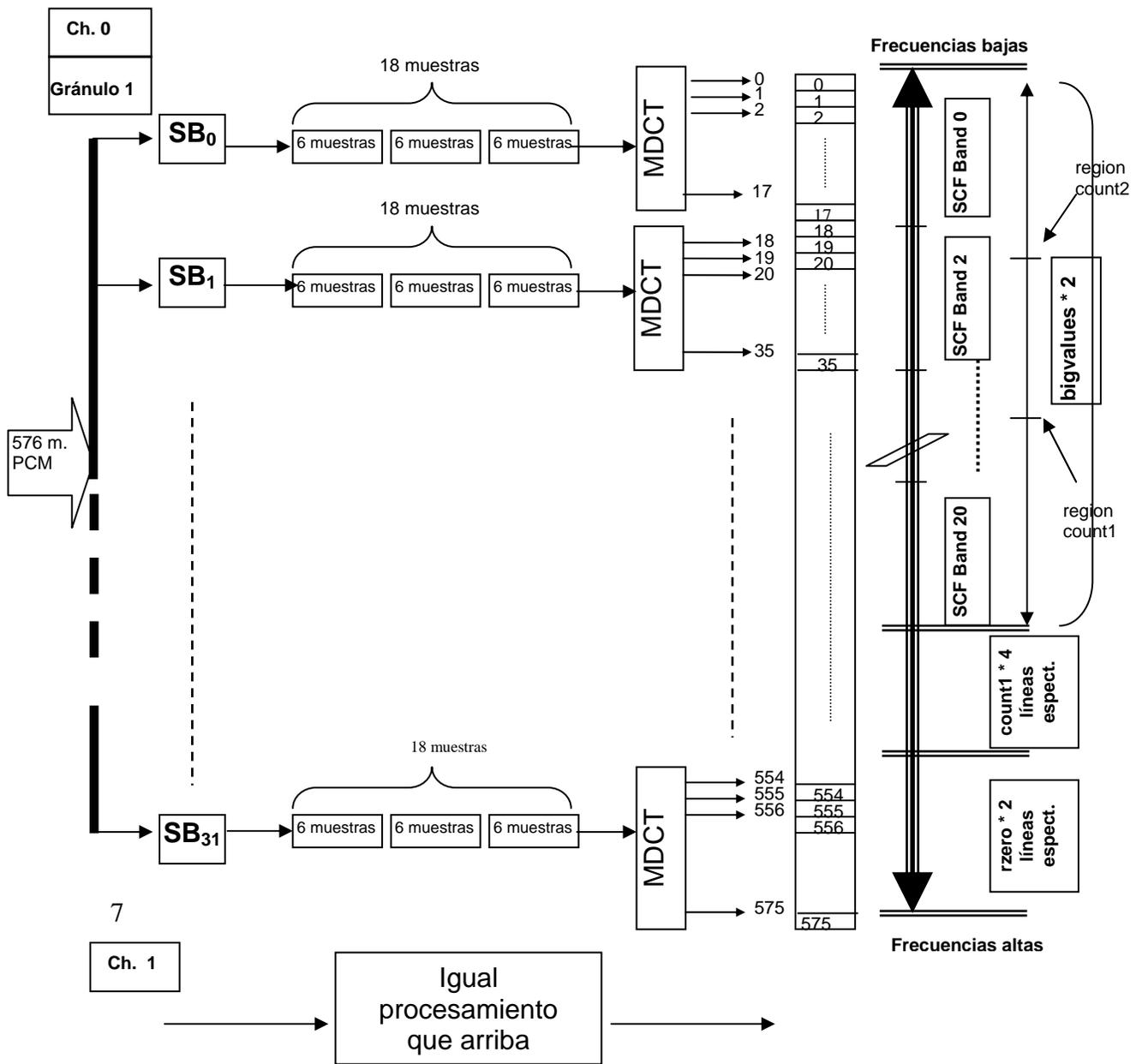


Figura 30. Procesamiento de la señal en el banco de filtros para señal estéreo y Bloques Largos en Capa III. Fuente [12]

4.2 Modelo Psicoacústico

Para el proceso de codificación en Capa III se puede utilizar cualquiera de los modelos psicoacústicos detallados en la norma, aunque en ésta se adopta el modelo 2 para modificarlo de acuerdo a los requerimientos de codificación.

De ahora en adelante siempre se hará referencia al Modelo Psicoacústico 2 dentro del proceso de codificación del Capa III. Las características principales son:

- (a) Este modelo se corre 2 veces por bloque utilizando una longitud de desplazamiento (*shift length*) de 576 muestras.
- (b) Se calcula la relación Señal Máscara (SMR) para cada **SCF Band**.
- (c) Conmutación de ventanas.

Una parte crucial en la codificación en el dominio de la frecuencia de las señales de audio es la aparición de pre-ecos. Considérese un caso dónde a un período de silencio le sigue un sonido de tipo percusivo, (tales como las castañuelas o platillos) dentro del mismo bloque de codificación. Dicho “ataque” producirá errores de cuantificación comparativamente grandes. Al producirse la decodificación de dicho bloque el filtro pasa banda (en el decodificador) esparcirá dicho error de cuantificación sobre todo el bloque, por lo que se volverá audible la distorsión, especialmente a bajo *bitrate* (MPEG-2).

Los pre-ecos pueden ser enmascarados utilizando el efecto temporal de *pre-masking*, si la ventana de tiempo es de corta longitud (del orden de unos pocos milisegundos). No obstante, un mayor porcentaje del *bitrate* total se requiere típicamente para la transmisión de la información lateral (*Side info*) si los bloques son más cortos. La solución pasa entonces por la conmutación de bloques de diferentes longitudes cuando las características de la señal los establezcan (un bloque largo de 576 muestras por canal con mejor resolución frecuencial y más pobre resolución temporal, o 3 bloques de 192 muestras por canal donde la resolución frecuencial es menor pero la temporal es mejor).

En la siguiente figura se observa la señal original de audio (a) y como el efecto del pre-eco es claramente visible cuando la longitud del bloque es 576 (b), mientras que en (c) solo afecta parte del segundo sub-bloque y no al primero por ser el subbloque de longitud 192.

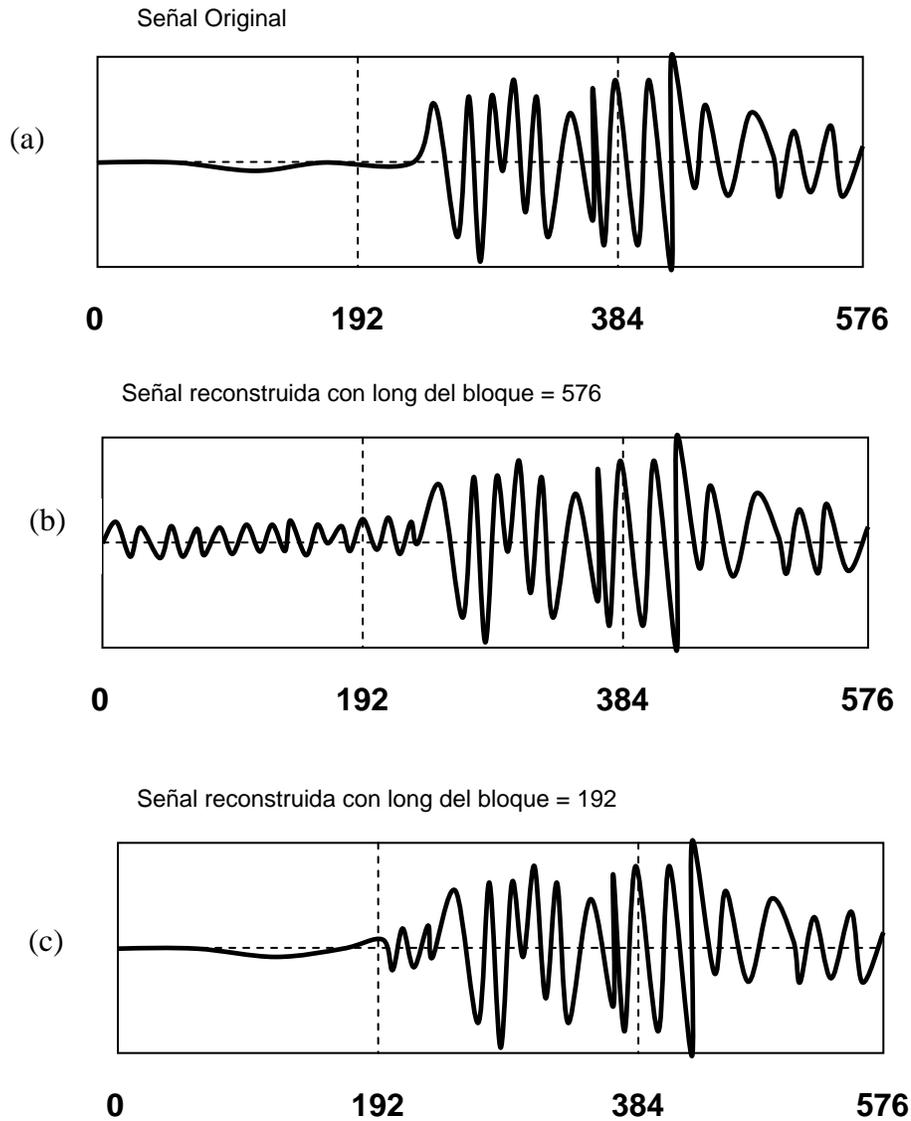


Figura 31. Efecto del pre-eco en la señal según la longitud del bloque. Fuente [7]

4.3 Decisión de conmutación de ventana:

La decisión si el banco de filtros debiese ser conmutado a ventanas cortas se deriva del cálculo del umbral de enmascaramiento por medio de calcular la estimación de la entropía psicoacústica (PE) y conmutando cuando PE excede el valor 1800. Si se alcanza esta condición se inicia la secuencia:

Start (block_type = 1) Short (block_type = 2) Stop (block_type = 3)

En el siguiente diagrama se muestra los posibles estados y cambios para la lógica de conmutación de ventanas.

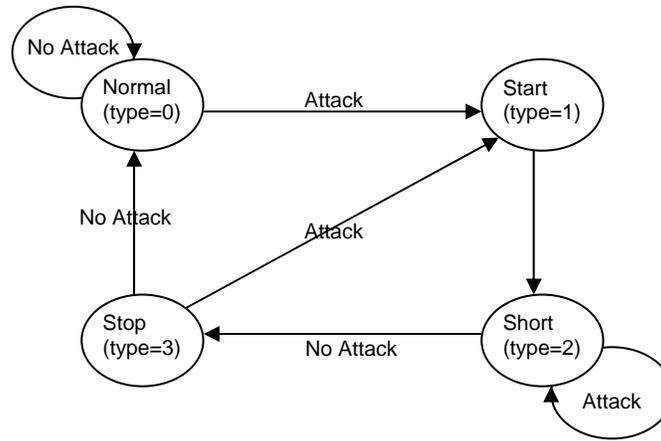


Figura 32. Diagrama de estados de la lógica para la conmutación de ventanas. Fuente [7]

Dos ventanas de MDCT especiales (**start window** y **stop window**) se necesitan en caso de transiciones entre bloques largos y cortos y viceversa para mantener la cancelación del aliasing que provee la MDCT en el dominio tiempo. Esto se observa en la siguiente figura.

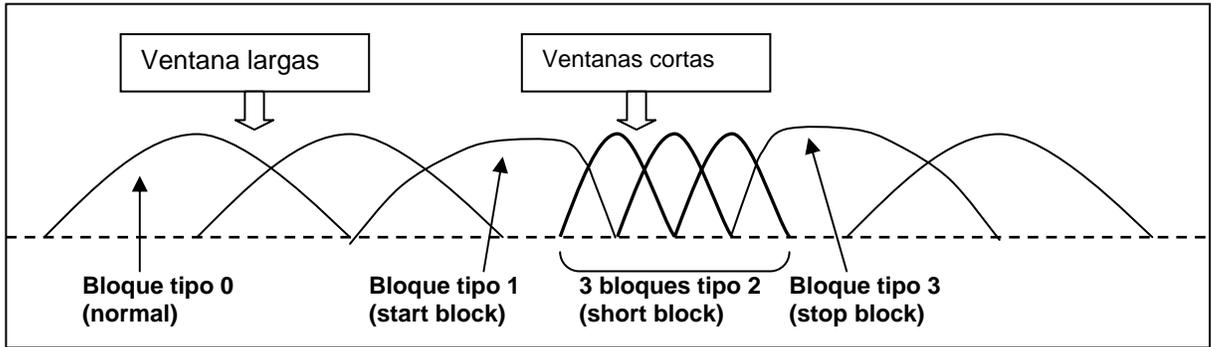


Figura 33. Ejemplo de cancelación del *aliasing* por medio del MDCT. Fuente [7]

4.4 Análisis del Banco de filtro Híbrido:

El análisis de sub-banda del banco de filtros polifásico se ha descrito anteriormente. La salida del banco de filtros polifásico es la entrada a la subdivisión en líneas espectrales usando la MDCT. De acuerdo a la salida del modelo psicoacústico las variables **block_split_flag** y **block_type**, las ventanas y transformadas “normal”, “start”, “short”, y “stop” son usadas. Entonces 18 valores de salida consecutivos de un gránulo y 18 valores consecutivos del gránulo anterior son ensamblados en un bloque de 36 muestras.

Block_type = 0 (enmarcado normal)

$$z_i = x_i' \text{ sen } [\pi/36 * (i + 1/2)] \quad \text{para } i = 0 \text{ a } 35 \quad \text{Ec.2.14}$$

Block_type = 1 (start block)

$$z_i = \begin{cases} x_i' \text{ sen } [\pi/36 * (i + 1/2)] & \text{para } i = 0 \text{ a } 17 \\ x_i & \text{para } i = 18 \text{ a } 23 \\ x_i \text{ sen } [\pi/12 * (i - 18 + 1/2)] & \text{para } i = 24 \text{ a } 29 \\ 0 & \text{para } i = 30 \text{ a } 35 \end{cases} \quad \text{Ec. 2.15}$$

Block_type = 3 (stop block)

$$z_i = \begin{cases} 0 & \text{para } i = 0 \text{ a } 5 \\ x_i \text{ sen } [\pi/12 * (i - 6 + 1/2)] & \text{para } i = 6 \text{ a } 11 \\ x_i & \text{para } i = 12 \text{ a } 17 \\ x_i \text{ sen } [\pi/36 * (i + 1/2)] & \text{para } i = 18 \text{ a } 35 \end{cases} \quad \text{Ec. 2.16}$$

Block_type = 2 (short block)

El bloque de 36 muestras se divide en 3 bloques solapados:

$$Y_i^{(0)} = x'_{i+6} \quad \text{para } i = 0 \text{ a } 11 \quad \text{Ec.2.17}$$

$$Y_i^{(1)} = x'_{i+12} \quad \text{para } i = 0 \text{ a } 11$$

$$Y_i^{(2)} = x'_{i+18} \quad \text{para } i = 0 \text{ a } 11$$

Cada uno de los 3 bloques cortos es “enventanado” separadamente.

$$z_i^{(k)} = y_i^{(k)} \text{ sen } [\pi/12 * (i + 1/2)] \quad \text{para } i = 0 \text{ a } 11 \text{ k} = 0 \text{ a } 2 \quad \text{Ec. 2.18}$$

4.5 Transformada discreta del coseno modificada MDCT.

En la siguiente fórmula: “n” es el número de muestras enmarcadas (para bloques cortos n=12 y para bloques largos n=36). En el caso de un bloque tipo 'corto' cada uno de los 3 bloques cortos es transformado separadamente.

$$x_i = \sum_{k=0}^{n-1} z_k \cos[\pi / 2n \cdot (2k + 1 + n/2)(2i + 1)] \quad \text{para } i = 0 \text{ hasta } n/2 - 1 \quad \text{Ec. 2.19}$$

Donde z_k es la muestra proveniente de la salida del banco de filtros polifásicos después del enventanado.

CAPÍTULO III

MODELOS PSICOACÚSTICOS DE LA ISO/IEC 11172-3

1. Modelo Psicoacústico I

El modelo psicoacústico controla la cuantificación y a su salida se obtiene la SMR. El cálculo del modelo psicoacústico tiene que ser adaptado a la Capa correspondiente. Este ejemplo es válido para Capas 1 y 2. El modelo también puede ser adoptado para la Capa 3. No hay una diferencia principal en aplicación del Modelo Psicoacústico I a la Capa 1 y 2. [12]

Capa 1: se calcula una nueva asignación de bits para cada bloque de 12 muestras por sub-banda o 384 muestras PCM. [12]

Capa 2 y 3: se calcula una nueva asignación de bits para los tres bloques totalizando 36 muestras por sub-bandas correspondiente al $3 * 384$ (1152) muestras PCM. [12]

Se calcula la asignación de bits de las 32 sub-bandas en base de la relación señal máscara de todas las sub-bandas. Por lo tanto, es necesario determinar para cada sub-banda, el máximo nivel de señal y el mínimo umbral de enmascaramiento. [12]

El mínimo de umbral de enmascaramiento se deriva de una transformada rápida de Fourier aplicada a una entrada PCM, seguida por el cálculo del modelo psicoacústico. [12]

La transformada rápida de Fourier en paralelo con el filtro sub-banda compensa la pérdida de selectividad espectral obtenida en bajas frecuencias por el banco de filtros de sub-banda. Esta técnica provee tanto una suficiente resolución en el tiempo para la codificación de la señal de audio (filtro polifásico con una ventana optimizada para mínimos pre-ecos) y una suficiente resolución espectral para el cálculo de sus umbrales enmascaramiento. Pueden ser calculadas las frecuencias y niveles de distorsión por *aliasing*. Esto es necesario para el cálculo de la tasa mínima

en esas sub-bandas que necesiten algunos bits para cancelar las componentes de *aliasing* en el decodificador. La complejidad adicional para calcular la mejor resolución en frecuencia es necesaria solamente en el codificador y no introduce en un retardo adicional o complejidad en el decodificador. [12]

En este capítulo se explica el Modelo Psicoacústico de la norma ISO 11172-3. En la medida en la que se explican los pasos para la obtención de la relación señal a máscara final, se ejemplificarán con gráficas alusivas al procedimiento en cuestión.

El cálculo de la relación señal máscara está basado en los siguientes pasos:

1. Cálculo de la transformada rápida de Fourier para la conversión del dominio tiempo al dominio frecuencia.
2. Determinación del nivel de presión sonora para cada sub-banda.
3. Determinación del umbral de silencio (umbral absoluto).
4. Encontrar las componentes del tipo tonal y no tonal de la señal de audio.
5. Diezmado de los enmascaradores, para obtener solamente aquellos relevantes.
6. Cálculo de los umbrales de enmascaramiento individuales.
7. Determinación del umbral de enmascaramiento global.
8. Determinación del umbral de enmascaramiento mínimo para cada sub-banda.
9. Cálculo de la relación señal máscara para cada sub-banda. [12]

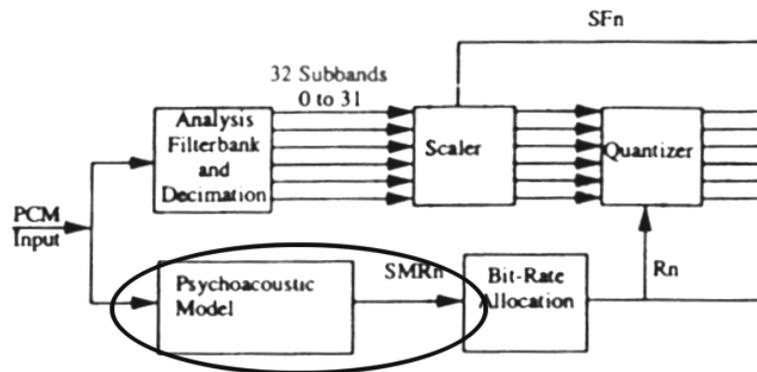


Figura 34. Estudio del modelo Psicoacústico I, según la norma ISO 11172-3. Fuente: [4]

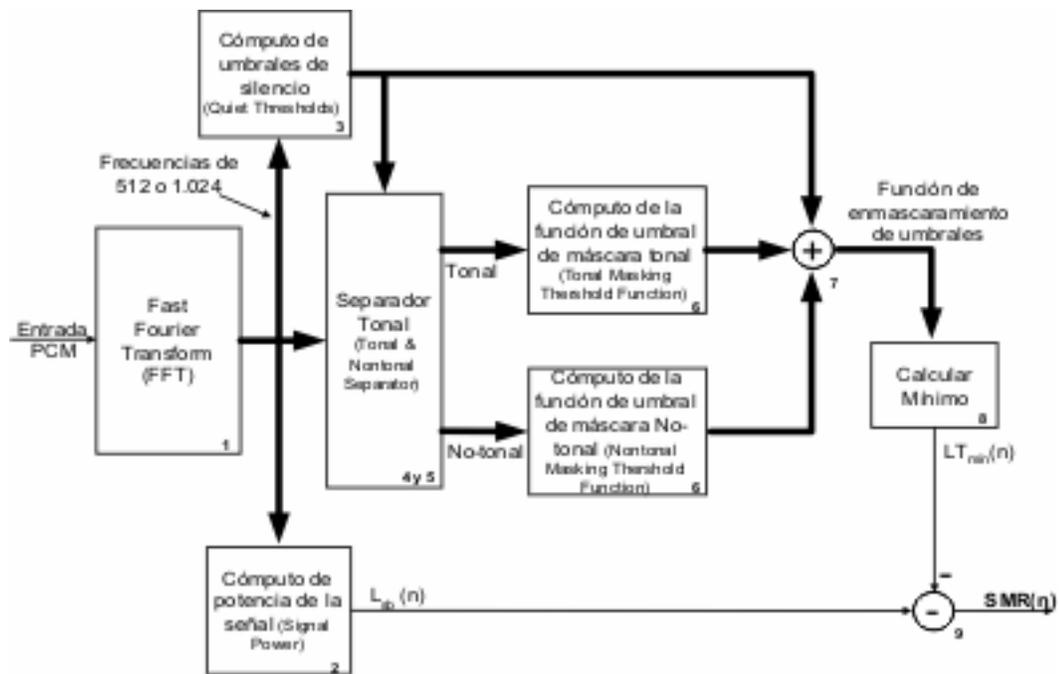


Figura 35. Detalle Modelo Psicoacústico implementado en Capa I y II de MPEG. Fuente: [4]

1.1 Transformada rápida de Fourier.

El umbral de enmascaramiento se obtiene estimando la densidad espectral de potencia que se calcula a través de la transformada rápida de Fourier de orden 512 para la Capa I y de orden 1024 para la Capa II y III. La transformada rápida de Fourier se calcula directamente desde la entrada PCM enmarcada con una ventana del tipo Hann. [12]

Para que haya coincidencia en el tiempo entre la asignación de bits y las muestras de sub-banda correspondientes, la señal PCM entrante que ingresa a la transformada rápida de Fourier, ésta debe retardarse.[12] La resolución espectral resultante es mucho más detallada que las sub-bandas, dando 8 puntos por sub-banda para Capa I y 16 puntos por sub-banda para Capa II. Esta resolución incrementada es necesaria para estimar con exactitud el enmascaramiento, especialmente a bajas frecuencias. [4]

El retardo del filtro de sub-banda es de 256 muestras, lo que corresponde 5,3ms de audio a 48 kHz de tasa de muestreo. Se requiere un desplazamiento de

ventana de 256 muestras para compensar la demora en el banco de filtro de análisis de sub-banda.

La ventana Hann debe coincidir con las muestras por sub-banda de la trama. Para la Capa I esto se adiciona a un desplazamiento de ventana de 64 muestras. Para Capa II se requiere un desplazamiento de ventana de 64 muestras.

Tabla 3. Datos técnicos de la Transformada Rápida de Fourier (FFT).
Fuente:propia

	Capa I	Capa II
Longitud de la transformada	512 muestras	1024 muestras
Tamaño de la ventana (fs=48kHz)	10,67 mseg	21,3 mseg
Tamaño de la ventana (fs=44.1Khz)	11,6 mseg	23,2 mseg
Tamaño de la ventana (fs=32kHz)	16 mseg	32 mseg
Resolución de frecuencia	Frec_muestreo/512	Frec_muestreo/1024

1.1.1 Ventana Hann: se determina con la función $h(i)$. Esta ventana poseerá su coeficiente normalizado por el factor $(8/3)^{1/2}$ para acercar la ventana a la unidad de energía por muestra [21], siguiendo la siguiente fórmula, con N=largo de la FFT:

$$h(i) = (8/3)^{1/2} \cdot 0,5 \cdot \left[1 - \cos\left(\frac{2\pi i}{N}\right) \right] \quad i = 0 \dots N-1 \quad [12] \quad \text{Ec. 3.1}$$

1.1.2 Densidad espectral de potencia $\mathbf{X}_{(k)}$: Donde $\mathbf{s}(\mathbf{l})$ es la señal de muestras de audio de entrada y $\mathbf{x}(\mathbf{l})$ es la señal de entrada normalizada, de acuerdo al largo de la FFT, N (N muestras), y al número de bits por muestra, \mathbf{b} , usando la siguiente relación[12,20]:

$$x(l) = \frac{s(l)}{N(2^{b-1})} \quad \text{Usando una FFT de 512 puntos se obtiene un estimado de } \mathbf{X}_{(k)}: \quad \text{Ec. 3.2}$$

$$X_{(k)} = 96 + 10 \cdot \log_{10} \left| \frac{1}{N} \cdot \sum_{l=0}^{N-1} h_{(l)} \cdot s_{(l)} \cdot e^{[-j \cdot k \cdot l \cdot \frac{2\pi}{N}]} \right|^2 \quad dB \quad k = 0 \dots N/2 \quad [12] \quad \text{Ec. 3.3}$$

$\mathbf{X}_{(k)}$ Es entonces la señal de entrada estudiada por bloque de audio. Se ha hecho una normalización al nivel de referencia de 96 dB SPL, de forma tal que el máximo valor posible corresponda a 96 dB¹. [12] La entrada normalizada, $\mathbf{x}(\mathbf{l})$ es

¹ Ver el Anexo 2. Caracterización el oído

segmentada en cuadros de 12 ms (512 muestras) usando la ventana de Hann sobrepuesta tal que cada cuadro contiene 10,9 ms de datos nuevos.[20]

1.2. Determinación del Nivel de Presión Sonora (*Sound Pressure Level*) (volumen).

El nivel de presión sonora (SPL) L_{sb} en la sub-banda n se calcula como:

$$L_{sb}(n) = MAX[X_{(k)}, 20 \cdot \log_{10}(scf_{max}(n) \cdot 32768) - 10] dB \quad [12] \quad \text{Ec. 3.4}$$

Donde:

- $X_{(k)}$ es el SPL de la línea espectral con un índice k de la FFT con el máximo de amplitud en el rango de frecuencia correspondiente a la sub-banda n . [12]
- **scf(máx)** es en Capa I el SCF y en Capa II el máximo de los 3 SCF de la sub-banda n de la trama. [12]
- El término -10 dB corrige la diferencia entre nivel de pico y RMS, esto es:

$$V_{rms} = \frac{V_p}{\sqrt{2}} \rightarrow 20 \cdot \log_{10}\left(\frac{1}{\sqrt{2}}\right) = -10dB \quad \text{Ec. 3.5}$$

- El nivel de presión sonora (SPL) L_{sb} se calcula para cada sub-banda n. [12]

La ecuación 3.4 significa que se usará el valor más grande entre el máximo valor de la FFT en la sub-banda o un valor normalizado **Scale_factor SCF**. Para tonos puros dentro de cada sub-banda el valor de la FFT será el más grande (máximo local), mientras que para los sonidos parecidos al ruido, el SFC normalizado serán las componentes no-tonales de menor tamaño (mínimos locales).[4]

1.3. Consideración del Umbral de silencio (Quiet Threshold) para cada componente de la FFT.

El umbral en silencio **LTq(k)**, también denominado umbral absoluto (UA) se detalla en las tablas de los anexos N° D.1-a,b,c para Capa I y D.1-d,e,f para Capa II. Estas tablas dependen de la frecuencia de muestreo de la señal PCM de entrada. Los valores están disponibles para cada muestra en el dominio de la frecuencia donde se

calcula el umbral de enmascaramiento.[12] Estos valores son simplemente tomados de las tablas derivadas empíricamente.[4]

Un *offset* dependiente de la tasa total de bits se utiliza para el umbral absoluto. Este *offset* es de -12 dB para tasa de bits ≥ 96 Kbps y 0 dB para tasas de bits < 96 Kbps por canal. [12]

1.4. Localización de las componentes tonales y No-tonales.

La tonalidad de una componente de enmascaramiento tiene influencia en el Umbral de Enmascaramiento (UE). Por este motivo, resulta importante discriminar entre las componentes del tipo Tonal (más parecido a la sinusoides) y No Tonal (más parecido al ruido). Para el cálculo del umbral de enmascaramiento global, es necesario discriminar ambas componentes del espectro de la FFT. [12]

Un tono es básicamente una componente de la FFT que es un máximo local rodeado por más de una o dos componentes significativas dentro de una banda predefinida de frecuencias vecinas². Los anchos de banda predefinidos son más pequeños que una sub-banda para tonos de baja frecuencia y más largos que una sub-banda para tonos de alta frecuencia. Las componentes tonales son identificadas, sus frecuencias marcadas y la potencia de su señal medida. [4]

Cada componente tonal y sus frecuencias vecinas en los anchos de banda predefinidos son entonces removidos del grupo de componentes de la FFT para obtener las componentes No-tonales o parecidas al ruido. Estas son entonces caracterizadas midiendo la potencia de la señal remanente dentro de cada banda de un grupo contiguo de bandas críticas, las cuales son especificadas en las tablas obtenidas empíricamente. Esta operación comienza con la determinación del máximo local, luego se extrae la componente tonal (tipo sinusoides). Los límites de las bandas críticas se dan en las tablas de los anexos N° D.2-a,b y c para Capa I y D.2-d,e y f para Capa II [12]. Se calcula la intensidad de las componentes No-tonales dentro de

² Es una definición de tono en Procesamiento de Señales. No necesariamente es una buena definición tal como el tono es percibido por el oído humano.

un ancho de banda, entonces son asignadas a una frecuencia representativa dentro de cada banda crítica. [4]

El ancho de banda de las bandas críticas varía con la frecuencia. Poseen un ancho de banda de solamente 0,1 kHz para bajas frecuencias hasta un ancho de banda de aproximadamente 4 kHz en altas frecuencias. Estos límites se determinaron en experimentos psicoacústicos³, donde se comprobó que el oído humano tiene mejor resolución en bajas frecuencias que en la región de frecuencias altas. [12]

Para determinar si el máximo local puede ser una componente del tipo tonal se examina un rango de frecuencia (df) alrededor del máximo. Este rango de frecuencia esta dado por la tabla 4 [12]

Tabla 4. Rango de frecuencias por Capa para la determinación del componente tonal.
Fuente:[13]

Sampling rate: 32kHz		
Capa I:	df = 125 Hz	0 kHz < f ≤ 4.0kHz
	df = 187.5 Hz	4.0 kHz < f ≤ 8.0 kHz
	df = 375 Hz	8.0 kHz < f ≤ 15.0kHz
Capa II:	df = 62.5 Hz	0 kHz < f ≤ 3.0 kHz
	df = 93.75 Hz	3.0 kHz < f ≤ 6.0 kHz
	df = 187.5 Hz	6.0 kHz < f ≤ 12.0 kHz
	df = 375 Hz	12.0 kHz < f ≤ 24.0 kHz
Sampling rate: 44.1kHz		
Capa I:	df = 172.266 Hz	0 kHz < f ≤ 5.512kHz
	df = 281.25 Hz	5.512 kHz < f ≤ 11.024 kHz
	df = 562.50 Hz	11.024 kHz < f ≤ 19.982kHz
Capa II:	df = 86.133 Hz	0 kHz < f ≤ 2.756 kHz
	df = 129.199 Hz	2.756 kHz < f ≤ 5.512kHz
	df = 258.398 Hz	5.512 kHz < f ≤ 11.024 kHz
	df = 516.797 Hz	11.024 kHz < f ≤ 19.982kHz
Sampling rate: 48 kHz		
Capa I:	df = 187.5 Hz	0 kHz < f ≤ 6.0 kHz
	df = 281.25 Hz	6.0 kHz < f ≤ 12.0 kHz
	df = 562.50 Hz	12.0 kHz < f ≤ 24.0 kHz
Capa II:	df = 93.750 Hz	0 kHz < f ≤ 3.0 kHz
	df = 140.63 Hz	3.0 kHz < f ≤ 6.0 kHz
	df = 281.25 Hz	6.0 kHz < f ≤ 12.0 kHz
	df = 562.50 Hz	12.0 kHz < f ≤ 24.0 kHz

³ Ver el Capítulo I. Caracterización el oído

Para listar las líneas espectrales $X_{(k)}$ que son componentes tonales y No tonales se deben realizar las siguientes 3 operaciones: [12]

1.4.1. Etiquetado del máximo local.

La línea espectral $X_{(k)}$ se etiqueta como un máximo local si:

$$X_{(k)} > X_{(k-1)} \text{ y } X_{(k)} \geq X_{(k+1)} \quad \text{Ec.3.6}$$

1.4.2. Listado de los componentes tonales y cálculo de SPL

Un máximo local se coloca en la lista de componentes tonales si:

$$X_{(k)} - X_{(k+j)} \geq 7\text{dB} \quad \text{Ec. 3.7}$$

Donde j es elegido de acuerdo a la tabla 5.

Tabla 5. Valores del subíndice j. Fuente:[13]

Capa I:	
J= -2,+2	para $2 < k < 63$
J= -3,-2,+2,+3	para $63 \leq k < 127$
J= -6,..., -2,+2,...,+6	para $127 \leq k \leq 250$
Capa II:	
J= -2,+2	para $2 < k < 63$
J= -3,-2,+2,+3	para $63 \leq k < 127$
J= -6,...,-2,+2,...,+6	para $127 \leq k < 255$
J= -12,...,-2,+2,...,+12	para $255 \leq k < 500$

Si resulta que $X_{(k)}$ es un componente tonal, entonces se listan los siguientes parámetros:

- El índice k de la línea espectral

$$SPLx_{tm(k)} = 10 \cdot \log_{10} \left[10^{\frac{x(k-1)}{10}} + 10^{\frac{x(k)}{10}} + 10^{\frac{x(k+1)}{10}} \right] \text{dB} \quad \text{Ec. 3.8}$$

- La bandera tonal (tonal flag)

Luego, todas las líneas espectrales dentro de las frecuencias analizadas se inicializan a -8 dB. [12]

1.4.3. Listado de las componentes no tonales y cálculo de la energía

Las componentes no tonales (similares a ruido) se calculan a partir de las líneas espectrales remanentes. Para calcular las componentes no tonales a partir de estas líneas espectrales $\mathbf{X}_{(k)}$, se determinan las bandas críticas $z_{(k)}$ usando las tablas D2-a,b y c para Capa I y D2-d,e y f para Capa II (ver anexos). [12]

En Capa I:

- 23 bandas críticas son usadas para las tasas de muestreo de 32 kHz.
- 24 bandas críticas son usadas para las tasas de muestreo de 44.1 kHz.
- 25 bandas críticas son usadas para las tasas de muestreo de 48 kHz

En Capa II:

- 24 bandas críticas son usadas para las tasas de muestreo de 32 kHz.
- 26 bandas críticas son usadas para las tasas de muestreo de 44.1 kHz y 48 kHz.

Dentro de cada banda crítica, la energía de las líneas espectrales remanentes luego que las componentes tonales se establecieron en cero, son sumadas para formar el SPL de la nueva componente No tonal $\mathbf{X}_{nm(k)}$ correspondiente a esa banda crítica.

Los siguientes parámetros son listados: [12]

- El índice K de la línea espectral más cerca de la media geométrica de la banda crítica.
- SPL $X_{nm(k)}$ en dB.
- La bandera no-tonal (Non Tonal Flag)

1.5. Diezmado de las componentes tonales y no tonales.

El diezmado⁴ es un procedimiento que se utiliza para reducir el número de enmascaradores (*maskers*) que son utilizados para el cálculo del umbral de enmascaramiento global [12.] Se detectan sólo los enmascaradotes tonales y no-tonales. Las componentes tonales y no-tonales menores que el umbral de silencio son

⁴ Decantado, filtraje, extracción de entre muchos valores, uno sólo, según criterio acordado.

descartadas. Después, si dos componentes tonales están cercanas en frecuencia las componentes más pequeñas son descartadas [4]:

(a) Las componentes tonales $X_{tm(k)}$ o no-tonales $X_{nm(k)}$ son considerados para el cálculo del umbral de enmascaramiento solamente si:

$$X_{tm(k)} \geq LT_{q(k)} \text{ o } X_{nm(k)} \geq LT_{q(k)} \quad \text{Ec. 3.9}$$

En la expresión, $LT_{q(k)}$ es el umbral absoluto (o umbral en silencio) a la frecuencia del índice k. Estos valores se dan en la tabla de los anexos D.1-a,b, y c para Capa I y D.1-d,e, y f para Capa II. [12]

(b) El diezmo de dos o más componentes tonales dentro de una distancia de 0,5 Bark; se deja la componente con la energía más alta y se remueven las componentes menores de la lista de componentes tonales. Para esta operación, una ventana que se desplaza en la banda crítica se utiliza con ancho de 0,5 Bark. [12]

En adelante, el índice j se utilizará para indicar las componentes tonales y no-tonales relevantes de la lista combinada diezmada. [12]

1.6. Cálculo de los umbrales individuales de enmascaramiento.

De las $N/2$ muestras originales en el dominio de la frecuencia, indexadas en k, solamente un subconjunto de muestras, indexadas en i son consideradas para el cálculo del umbral de enmascaramiento global. Las muestras están en las tablas D.1-a,b y c para Capa I y D.1-d,e y f para Capa II (ver anexos) [12]. Para cada enmascarador relevante de tonalidad se calcula una función de umbral de enmascaramiento que es una función no-lineal derivada empíricamente de un enmascarador de potencia de la señal y de un enmascarador de frecuencia. Igual procedimiento se realiza para cada enmascarador relevante de no-tonalidad. Un valor dado de potencia del enmascarador tonal generalmente provee menos

enmascaramiento que la misma cantidad de potencia de un enmascarador no-tonal. [4]

1.6.1 Capa I:

Para las líneas de frecuencia correspondientes a la región de frecuencia cubierta por las primeras 6 sub-bandas, no se utiliza sub-muestreo. Para las líneas de frecuencia correspondientes a la región de frecuencia cubierta por las siguientes 6 sub-bandas se considera cada segunda línea espectral. Finalmente, en el caso de 44,1 kHz y 48 kHz de tasa de muestreo, en la región de frecuencias correspondientes a las sub-bandas remanentes, son consideradas cada 4 líneas espectrales, hasta 20 kHz. En el caso de 32 kHz de frecuencia de muestreo, en la región de frecuencias correspondientes a las sub-bandas restantes se consideran cada 4 líneas espectrales, hasta 15 kHz. (Ver también las tablas D.1-a,b y c para Capa I, anexos). [12]

1.6.2 Capa II:

Para las líneas de frecuencia correspondientes a la región de frecuencia la cual es cubierta por las primeras 3 sub-bandas, no se utiliza sub-muestreo. Para las líneas de frecuencia correspondientes a la región de frecuencia la cual es cubierta por las siguientes 3 sub-bandas se considera cada segunda línea espectral. Para las líneas de frecuencia correspondientes a la región de frecuencia la cual es cubierta por las siguientes 6 sub-bandas se considera cada 4 líneas espectrales. Finalmente, en el caso de 44,1 kHz y 48 kHz de tasa de muestreo, en la región de frecuencias correspondientes a las sub-bandas remanentes, cada 8 líneas espectrales son consideradas hasta 20 kHz. En el caso de 32 kHz de frecuencia de muestreo, en la región de frecuencias correspondientes a las sub-bandas restantes cada 8 líneas espectrales se consideran hasta 15 kHz. (Ver también las tablas D.1-d,e,f para Capa II, anexos). [12]

El número de muestras, i , en el dominio de las frecuencias sub-muestreadas es diferente dependiendo de las tasas de muestreo y de las Capas. Ver tabla 6.

Tabla 6. Cantidad de muestras según la Capa. Fuente:[3]

Tasa de muestreo (kHz)	Capa I (i=)	Capa II (i=)
32	108	132
44,1	106	130
48	102	126

Para cada componente tonal y no-tonal se le asigna el valor del índice i en el dominio de cada frecuencia sub-muestreada que más cerca se corresponda a la frecuencia de la línea espectral original $X_{(k)}$. Este índice i se da en las tablas D.1-a,b y c para Capa I y D.1-d,e y f para Capa II (ver anexos).

Los umbrales de enmascaramiento individuales para tanto las componentes tonales y No tonales se dan en la siguiente expresión:

$$LT_{tm}[z(j),z(i)] = X_{tm}[z(j)] + av_{tm}[z(j)] + vf [z(j),z(i)] \text{ dB} \quad \text{Ec. 3.10}$$

$$LT_{nm}[z(j),z(i)] = X_{nm}[z(j)] + av_{nm}[z(j)] + vf [z(j),z(i)] \text{ dB} \quad \text{Ec.3.11}$$

En las fórmulas, LT_{tm} y LT_{nm} son los umbrales de enmascaramiento individuales en las bandas críticas (z) (en Bark) de las componentes de enmascaramiento, en la banda crítica del enmascarador (masker) z_m (en Bark). Los valores en dB pueden ser tanto positivos como negativos. El término $X_{tm}[z(j)]$ es el SPL de la componente de enmascaramiento con el número índice j en la banda crítica correspondiente $z(j)$. El término av se denomina el índice de enmascaramiento y vf es la función de enmascaramiento de la componente de enmascaramiento $X_{tm}[z(j)]$. El índice de enmascaramiento av es diferente para los enmascaradores (*maskers*) tonales y No tonales (av_{tm} y av_{nm}). [12]

Para los maskers tonales está dado por:

$$av_{tm} = (-1,525 - 0,275 * z(j) - 4,5) \text{ dB SPL} \quad \text{Ec. 3.12}$$

Para los maskers no-tonales está dado por:

$$av_{nm} = (-1,525 - 0,175 * z(j) - 0,5) \text{ dB SPL} \quad \text{Ec. 3.13}$$

La función de enmascaramiento vf de un enmascarador (*masker*) se caracteriza por pendientes diferentes, inferiores y superiores, las cuales dependen de

la distancia (en Bark) $dz = z(i) - z(j)$ al enmascarador [12]. La variable vf se aproxima al espaciado basilar o patrón de excitación del oído descrito en el Capítulo I (3.2). En la expresión, i es el índice de la línea espectral a la cual se calcula la función de enmascarado y j es la del enmascarador. Las bandas críticas $z(j)$ y $z(i)$ se encuentran en las tablas D.1-a,b y c para Capa I y D.1-d,e y f para Capa II (ver anexos). La función de enmascarado, que es la misma para los enmascaradores tonales y no-tonales está dada por:[12]

$$vf = 17 * (dz+1) - (0,4 * X[z(j)] + 6) \text{ dB para } -3 \leq dz < -1 \text{ Bark} \quad \text{Ec. 3.14}$$

$$vf = (0,4 * X[z(j)] + 6) * dz \text{ dB para } -1 \leq dz < 0 \text{ Bark} \quad \text{Ec. 3.15}$$

$$vf = -17 * dz \text{ dB para } 0 \leq dz < 1 \text{ Bark} \quad \text{Ec. 3.16}$$

$$vf = -(dz-1) * (17 - 0,15 * X[z(j)]) - 17 \text{ dB para } 1 \leq dz < 8 \text{ Bark} \quad \text{Ec. 3.17}$$

En estas expresiones $X[z(j)]$ es el SPL de la j -ésima componente de enmascarado (en dB). Por razones de complejidad de implementación, el enmascaramiento no se considera más allá, si $dz < -3$ Bark o $dz \geq 8$ Bark (LT_{tm} y LT_{nm} se inicializan en -8 dB).

1.7 Cálculos del Umbral de enmascaramiento global (LT_g).

El umbral de enmascaramiento global $LT_g(i)$ en la i -ésima muestra de frecuencia se deriva a partir de las pendientes superiores e inferiores de los umbrales de enmascaramiento individuales de cada uno de los j -ésimos enmascaradores tonales y no tonales y la adición del umbral de silencio o umbral absoluto de audio $LT_q(i)$. Este umbral se da en las tablas D.1-a,b,c para el Capa I y de las tablas D.1-d,e,f para Capa II (ver anexos) [12]. El umbral de enmascaramiento global se obtiene sumando, en el dominio de la potencia⁵, las potencias correspondientes a los umbrales de enmascaramiento individuales, tonales y no-tonales, más el umbral de enmascaramiento en silencio.[4]

⁵ Enmascaradores individuales y umbrales absolutos se especifican en $dB = 10 \log(\text{potencia})$.

Es común que los umbrales de enmascaramiento sean computados por separado para cada canal estéreo. Sin embargo, en algunas aplicaciones es posible utilizar el enmascaramiento de un canal estéreo sobre el otro. [4]

$$LTg(i) = 10 \log_{10} \left(10^{0,1 \cdot LTq(j)} + \sum_{m=1}^M 10^{0,1 \cdot LTm(j,m)} + \sum_{n=1}^N 10^{0,1 \cdot LTnm(j,n)} \right) \text{ dB} \quad \text{Ec. 3.18}$$

El número total de enmascaradores tonales está dado por **M** y el número total de enmascaradores no-tonales por **N**. Para una **i** dada, el rango de **j** puede ser reducido para abarcar esas componentes de enmascaramiento que se hallan entre -8 y $+3$ Bark desde **i**. Fuera de este rango, **LT_{tm}** y **LT_{nm}** se consideran -8 dB. [12]

1.8. Determinación del umbral de enmascaramiento mínimo.

El mínimo nivel de enmascaramiento **LT_{min}(n)** en la sub-banda **n** se determina a partir de la siguiente expresión:

$$LT_{\min}(\mathbf{n}) = \text{MIN} [LT_f(\mathbf{i})] \text{ dB} \quad \text{con } f(i) \text{ en la sub-banda } n. \quad \text{Ec. 3.19}$$

Donde **f(i)** es la frecuencia de la **i**-ésima muestra frecuencial. Las **f(i)** están tabuladas en las tablas de los anexos D.1-a,b,c para el Capa I y de las tablas D.1-d,e,f para Capa II.

Un nivel mínimo de enmascaramiento (**LT_{min}(n)**) se computa para cada sub-banda.[12]

1.9. Cálculo de la relación Señal-Máscara (SMR)

La relación señal máscara se computa para cada sub-banda (**n**) de la siguiente manera: [12]

$$SMR_{sb}(\mathbf{n}) = L_{sb}(\mathbf{n}) - LT_{\min}(\mathbf{n}) \quad \text{en dB.} \quad \text{Ec. 3.20}$$

Luego del cálculo de **SMR(n)** comienza la asignación de bits, cuantificación y codificación para cada **n**-bloque. En el caso estéreo, esto se realiza simultáneamente para ambos canales, lo cual permite una asignación adaptable entre canales. Las

muestras por sub-banda son entonces cuantificados por un cuantificador uniforme, con un número asignado de niveles. Cada Capa posee un criterio para la asignación de bits y la escogencia del número de niveles de cuantificación.[4]

2. Modelo Psicoacústico II

Generalidades

El Modelo Psicoacústico 2 es un modelo independiente que puede ser ajustado y adaptado a cualquiera de las Capas de la Norma ISO-MPEG-Audio. Esta descripción provee la suficiente información para su implementación en las Capas I y II y para su uso en la Capa III, basado en esta información, se incluyen las adaptaciones necesarias descritas al final del presente tema. [12]

2.1 El proceso de cálculo de generación del umbral.

Posee 3 entradas. Estas son:

2.1.1. La longitud de desplazamiento (**iblen**) para el proceso de cálculo del umbral (donde $384 < \mathbf{iblen} < 640$). Este **iblen** debe mantenerse constante sobre cualquier aplicación del proceso de cálculo del umbral. Si, como sucede en Capa III, resulta necesario calcular los umbrales para 2 longitudes de desplazamiento (**iblen**) diferentes, entonces dos procesos serán necesarios, cada uno corriendo con una longitud de desplazamiento fija. En el caso donde **iblen** esté fuera del rango (de 384 a 640) entonces será necesario calcular los umbrales psicoacústicos, tanto con una longitud de ventana como una longitud de desplazamiento diferentes. Hay dos formas de hacer esto:

(a) Usar una longitud de transformada diferente y recalculando los coeficientes de inicio para el modelo o,

(b) Usar la misma longitud de transformadas pero una ventana tipo Hann más corta, apropiada para la data y el problema en cuestión. [12]

2.1.2 La elección del método depende de la implementación. Las muestras **iblen** más nuevas de la señal, con las muestras retrasadas (bien sea en el banco de filtros o en el cálculo psicoacústico) de forma tal que la ventana del cálculo psicoacústico esté centrada en la ventana temporal de la aplicación. [12]

2.1.3 La frecuencia de muestreo. Existe un conjunto de tablas provistas para las frecuencias de muestreo estándares. La frecuencia de muestreo, a igual que el **iblen**, debe necesariamente permanecer constante a lo largo de un proceso de cálculo de umbral. [12]

La salida para el Modelo Psicoacústico 2 es un conjunto de Relaciones Señal Máscara (**SMR_n**), que se adaptan a las Capas como se describe más adelante. [12]

Antes de correr el modelo inicialmente, la matriz utilizada para contener los datos de ventana de la FFT precedente y las matrices que contienen **r** (módulo) y **f** (fase) deben inicializarse para proveer un punto de inicio conocido. [12]

En Capa II, las relaciones de enmascarado psicoacústico deben calcularse 2 veces para cada trama en el codificador. La relación más estricta de cada par de relaciones es utilizada para la asignación de bits a las muestras (**bit allocation**). (se recuerda que en Capa II se puede transmitir la asignación [**allocation**] para un conjunto de 12 o 36 muestras de acuerdo, justamente, a lo arriba expuesto) [12]

2.2 Comentarios sobre notaciones [12]

- **w** Indica que el cálculo se indexa por frecuencia en la línea espectral de la FFT. (Un índice de **w=1** indica DC y un índice de **w=513** se refiere a la línea espectral de la frecuencia de Nyquist.
- **b** Indica que el cálculo esta ordenado (indexado) en el dominio de la partición de cálculo del umbral. En los casos donde el cálculo incluye una convolución o suma en el dominio de la partición de cálculo del umbral, entonces se utiliza **bb** como variable de suma. La numeración de la partición comienza en 1.

- **n** Indica que el cálculo está ordenado (indexado) en el dominio de asignación (allocation) del coder bit (o codebook). Un índice de 1 corresponde a la banda más baja el banco de filtros de sub-banda.

- La función *SPREADING* (esparcir, extender)

Muchos puntos en la descripción del modelo hacen referencia a la función **spreading**. Esta se calcula de la siguiente manera [12]:

$$\mathbf{tmpx} = 1,05 (j-i) \quad \text{Ec. 3.21}$$

Donde **i** es el valor Bark (Ec.2.2) de la señal a ser extendida (spread) y **j** es el valor Bark de la banda a ser extendida dentro de **i**. **tmpx** Es una variable temporal.

$$\mathbf{x} = 8 \cdot \text{MIN} [(\mathbf{tmpx} - 0,5)^2 - 2 \cdot (\mathbf{tmpx} - 0,5) , 0] \quad \text{Ec. 3.22}$$

Donde **x** es una variable temporal y **MIN(a,b)** devuelve el valor más negativo entre a y b.

$$\mathbf{tmpy} = 15,811389 + 7,5 \cdot (\mathbf{tmpx} + 0,474) - 17,5 \cdot [1,0 + (\mathbf{tmpx} + 0,474)^2]^{0,5} \quad \text{Ec. 3.23}$$

Donde **tmpy** es otra variable temporal que representa la función de enmascaramiento sugerida por Schroeder.

if (tmpy < -100) then

$$\mathbf{sprdnf(i,j)} = 0$$

else

$$\mathbf{Sprdnf(i,j)} = 10^{(x+tmpy)/10}$$

end if

Ec. 3.24

2.3 Pasos en el cálculo del umbral

Los siguientes pasos son necesarios par el cálculo de la **SMR_n** utilizada en el codificador.

2.3.1 Reconstrucción de 1024 muestras de la señal de entrada.

La longitud de desplazamiento (**iblen**) de las muestras nuevas está disponible en cada llamada al generador de umbral. El generador de umbral debe

almacenar 1024 muestras **iblen** y concatenarlas (a esas muestras), para reconstruir precisamente 1024 muestras consecutivas de la señal de entrada s_i donde i representa el índice tal que $1 < i < 1024$ para el actual flujo de entrada.[12]

2.3.2 Procedimiento de cálculo del espectro complejo de la señal de entrada.

1. s_i Se enmarca (*windowed*) con una ventana tipo Hann de 1024 puntos. Por ejemplo, con la ecuación (8.1):

$$sw_i = s_i \cdot 0,5 \cdot \left[1 - \cos\left(\frac{2\pi i}{1024}\right) \right] \quad i = 0 \dots N - 1 \quad \text{Ec. 3.25}$$

Nótese que en Capa III una ventana más corta puede usarse en caso de que **window_switching** esté activo, con el apropiado centrado de la ventana. (Ver descripción del codificador capa III, Capítulo II).[12]

2. Se calcula la FFT de sw_i

3. Se calcula la representación polar de la transformada donde se obtiene r_w y f_w que representan las componentes de amplitud y fase de la transformada, respectivamente, de sw_i .[12]

2.3.3 Cálculo predictivo de r y f .

La magnitud r_w y fase f_w , estimadas por un cálculo predictivo, son calculadas a partir los 2 bloques de cálculos de umbral precedentes r y f :

$$r_w = 2r_w(t-1) - r_w(t-2) \quad \text{ver que: } x_{(t)} = x_{(t-1)} + [x_{(t-1)} - x_{(t-2)}] \quad \text{Ec. 3.26}$$

$$f_w = 2f_w(t-1) - f_w(t-2) \quad \text{Ec. 3.27}$$

Donde t representa el bloque actual, $t-1$ apunta al bloque de datos previos y $t-2$ indica el bloque de datos anterior a $t-1$. O sea que el valor predicho es igual al anterior más la variación (con signo) del valor anterior con su predecesor.[12]

2.3.4 Cálculo de la medida de impredecibilidad C_w

La medida de impredecibilidad C_w es:

$$C_w = \frac{\left[(rw \cdot \cos(fw) - \hat{rw} \cdot \cos(\hat{fw}))^2 + [rw \cdot \sin(fw) - \hat{rw} \cdot \sin(\hat{fw})]^2 \right]^{0.5}}{rw + \text{abs}(\hat{rw})} \quad \text{Ec. 3.28}$$

Sacrificando calidad (*performance*), esta medición puede ser calculada sobre la porción más baja de las líneas de frecuencia solamente. Los cálculos se deberían extender desde DC hasta por lo menos 3 kHz o, preferiblemente, 7 kHz. Un límite superior que es inferior a 5.5 kHz puede reducir el *performance* considerablemente con respecto a los tests subjetivos realizados al algoritmo de audio. El valor de C_w por encima de este límite debería ser inicializado en 0,3. Mejores resultados se obtienen calculando C_w hasta los 20 kHz [12]. La impredecibilidad da, entonces, una idea del error de predicción cometido.

2.3.5 Cálculo de la Energía e Impredecibilidad en las particiones de cálculo de umbral.

Las particiones de cálculo del umbral proveen una resolución de aproximadamente una línea de FFT o de 1/3 de Bark (lo que resulte más ancho). A bajas frecuencias, una sola línea de la FFT constituirá una partición de cálculo. En altas frecuencias, muchas líneas se combinarán en una sola partición de cálculo. Un conjunto de valores de partición se dan para cada una de las 3 tasas de muestreo en la tabla de los anexos D.3, correspondiente al cálculo de partición. Estos elementos de las tablas serán usados para el proceso de cálculo de umbral. Existen varios elementos en cada entrada a la tabla:

La energía de cada partición, e_b es:

$$e_b = \sum_{\omega=\omega_{lowb}}^{\omega_{highb}} rw^2 \quad \text{Ec. 3.29}$$

y la impredecibilidad ponderada, c_b es:

$$e_b = \sum_{\omega=\omega_{lowb}}^{\omega_{highb}} rw^2 \cdot C_w \quad \text{Ec. 3.30}$$

1. El índice de la partición de cálculo, **b**
2. La línea de frecuencia más baja en la partición, **ωlow_b**
3. La línea de frecuencia más alta en la partición, **ωhigh_b**
4. El valor medio de Bark de la partición, **bval_b** [12]
5. El límite inferior de **SNR** en la partición que controla los efectos de no-enmascarado de estéreo, **minval_b**
6. El valor de ruido tonal de enmascarado (**tone masking noise**) (en dB) para la partición, **TMN_b**.

Para cada tasa de muestreo existe el valor más grande de “b”, **b_{máx}**, igual al índice más alto.[12]

2.3.6 Convolución de la energía *particionada* y la impredecibilidad con la función de Spreading.

$$ecb_b = \sum_{bb=1}^{b_{\max}} e_{bb} \cdot sprdng(bval_{bb}, bval_b) \quad \text{Ec. 3.31}$$

$$ct_b = \sum_{bb=1}^{b_{\max}} c_{bb} \cdot sprdng(bval_{bb}, bval_b) \quad \text{Ec. 3.32}$$

Dado que **ct_b** está ponderado por la energía de la señal, éste debe ser normalizado a **cb_b**. [12]

$$\mathbf{Cb}_b = \mathbf{ct}_b / \mathbf{ecb}_b \quad \text{Ec. 3.33}$$

Al mismo tiempo, debido a la naturaleza no normalizada de la función **spreading**, **ecb_b** debería ser renormalizado y, entonces calcular la energía normalizada, **En_b**. [12]

$$\mathbf{En}_b = \mathbf{ecb}_b * \mathbf{rnorm}_b \quad \text{Ec. 3.34}$$

Donde el coeficiente de normalización, **rnorm_b** es:

$$Rnorm_b = \left(\sum_{bb=0}^{b_{\max}} sprdng(bval_{bb}, bval_b) \right)^{-1} \quad \text{Ec. 3.35}$$

2.3.7 Convertir cb_b a tb_b , el índice de tonalidad

$$Tb_b = -0,299 - 0,43 \ln(cb_b) \quad \text{Ec. 3.36}$$

Cada tb_b está limitado en su rango a: $0 < tb_b < 1$ [12]

2.3.8 Cálculo de la SNR requerida en cada partición.

$$NMT_b = 5,5 \text{ dB para todos los } b \quad \text{Ec. 3.37}$$

NMT_b es el valor del ruido tonal de enmascaramiento (en dB) para la partición.

La relación señal ruido requerida, SNR_b es:

$$SNR_b = \text{MAX}[\text{minval}_b, (tb_b * TNM_b + (1-tb_b)*NMT_b)] \quad \text{Ec. 3.38}$$

Donde la función MAX(a,b) retorna el valor menos negativo entre a y b.[12]

2.3.9 Cálculo de la relación de potencia.

La relación de potencia, bc_b , es:[12]

$$bc_b = 10^{\frac{-SNR_b}{10}} \quad \text{Ec. 3.39}$$

2.3.10 Cálculo del umbral de energía actual, nb_b . [12]

$$nb_b = en_b * bc_b \quad \text{Ec. 3.40}$$

2.3.11 Esparcir (spread) el umbral de energía sobre las líneas de FFT, obteniendo nb_w . [12]

$$nb_w = \frac{nb_b}{\omega_{high_b} - \omega_{low_b} + 1} \quad \text{Ec. 3.41}$$

2.3.12 Inclusión de los umbrales absolutos, obteniendo el umbral de energía de audibilidad final, thr_w : [12]

$$thr_w = \text{MAX}(nb_w, absth_r_w) \quad \text{Ec. 3.42}$$

Los valores de $absth_r_w$ (en dB) se encuentran en la tablas D.4 (Tablas de Umbral Absoluto, ver anexos) y son relativos al nivel que una señal senoidal de amplitud igual a $\pm 1/2$ Isb (least significant bit) tiene en la FFT utilizada para el

cálculo del umbral. Los valores en dB deben ser convertidos en el dominio de la energía después de considerar la normalización de la FFT actualmente usada.[12]

2.3.13 Control de Pre-eco

Para Capa III, el control de pre-eco tiene lugar en este punto. El control utilizado se describe en la parte de la especificación del codificador Capa III. Este paso se omite para los Capas I y II.[12]

2.3.14 Cálculo de las relaciones señal-máscara, SMR_n .

La tabla D.5 “Tabla de la partición del codificador para Capa I y II” (ver anexos) muestra:

- (a) El índice n de la partición del codificador.
- (b) El índice más bajo ω_{low_n} de la partición del codificador.
- (c) El índice más alto ω_{high_n} de la partición del codificador.[12]

$width_n$ Es el índice del ancho, tomando valores: $width_n=1$ para una banda de factores de escala (**SCF band**) psicoacústicamente angosta y $width_n=0$ para una banda de factores de escala (**SCF band**) psicoacústicamente ancha. Una banda de factores de escala psicoacústicamente angosta es aquella cuyo ancho es menor que 1/3 de la banda crítica, aproximadamente.[12]

La energía en el **SCF band**, $epart_n$ es:

$$epart_n = \sum_{\omega=\omega_{low_n}}^{\omega_{high_n}} rW^2 \quad \text{Ec. 3.43}$$

Entonces si $width_n = 1$, el nivel de ruido en el **SCF band**, $npart_n$ se calcula como:

$$npart_n = \sum_{\omega=\omega_{lowb_n}}^{\omega_{highb_n}} thrw \quad \text{Ec. 3.44}$$

si no,

$$npart_n = \text{MIN}[thr_{\omega_{lown}} , \dots, thr_{\omega_{highn}}] * [\omega_{highn} - \omega_{lown} + 1] \quad \text{Ec. 3.45}$$

Donde en este caso MIN(a,...,z) es una función que retorna el argumento positivo más pequeño de la lista de argumentos (a,...,z) .[12]

Las relaciones que se enviarán al codificador, SMR_n, se calculan como:

$$SMR_n = 10 \cdot \log_{10} \left(\frac{epart_n}{npart_n} \right) \quad [12] \quad \text{Ec. 3.46}$$

3. Modificaciones al modelo Psicoacústico II para su uso con Capa III [13]

Las adaptaciones que se le hacen al Modelo Psicoacústico 2 para su utilización con Capa III son las siguientes:

- El modelo se calcula 2 veces, en paralelo.
- Un proceso se hace con una longitud de desplazamiento (**iblen**) de 192 muestras (para ser usados en bloques cortos). Para esto, la longitud del bloque de la FFT se cambia a 256 y otros parámetros cambian acordemente.
- Se realiza otro proceso que se calcula con una longitud de desplazamiento (**iblen**) de 576 muestras (para bloques largos).
- Se realizan modificaciones en el cálculo de la impredecibilidad métrica del Modelo Psicoacústico 2. Estas son:

(a) Esta se calcula para las primeras 206 líneas espectrales, para las restantes esta se inicializa en el valor 0,4.

(b) Dentro de las primeras 206 líneas espectrales, para las primeras 6 líneas se utiliza una FFT larga (longitud de ventana =1024 **shiftlen**=576) y,

(c) Para las líneas espectrales 6 a 205 se utiliza una FFT corta (longitud de ventana=256 **shiftlen**=192.)

$$cw(w) = \begin{cases} cw_l & \text{para } 0 \leq w < 6 \\ cw_s & \text{para } 6 \leq w < 206 \text{ } w = 6,10,14\dots \\ 0,4 & \text{para } w \geq 206 \end{cases} \quad \text{Ec. 3.47}$$

Donde:

cw_l es la impredecibilidad calculada con la FFT larga,

cw_s es la impredecibilidad calculada desde el segundo bloque corto de los tres bloques cortos dentro de un gránulo.

w corresponde al número de línea espectral.

- Otros parámetros internos del modelo se modifican (la función “**spreading**” y parámetros de conversión de la impredecibilidad).

if j = i

tempy = 3,0(j-i)

else

tempy = 1,5(j-i)

end if

Ec. 3.48

- Sólo los valores de **sprdnf()** > 10⁻⁶ se utilizan. Los valores menores se inicializan en 0. Además:

conv1 = -0,299

conv2 = -0,43

Ec. 3.49

- El valor de **NMT** (Noise Masking Tone) se establece en 6db para todas las particiones donde se calcula el umbral.

- El parámetro **TMN** (Tone Masking Noise) se establece en 29 dB para todas las particiones.

- La entropía Psicoacústica se estima como la relación **thr/eb** donde **thr** es el umbral y **eb** es la energía.

$$pe = - \sum \left[cbwidth_k \cdot \log_{10} \left(\frac{thr_k}{(eb_k + 1)} \right) \right] \quad \text{Ec. 3.50}$$

donde **k** indexa las particiones de cálculo de umbral y **cbwidth** es el ancho de la partición de cálculo de umbral (ver tablas C.8)

- Se modifican las constantes en el cálculo de pre-ecos. (ver diagrama de bloques)

rpelev=2 y rpelev2=16

Ec. 3.51

El umbral no se extiende (*spread*) sobre las líneas de la FFT. Las particiones donde se calcula el umbral son convertidas directamente a **SCF bands** (bandas de factor de escala). Varias particiones conforman un **SCF band** y la 1^{ra} y última de las particiones, que conforman un **SCF band**, son ponderadas con constantes (ver tablas C.8, anexos). Dicha tabla contiene también el número de particiones que conforman cada **SCF band**, excluyendo la primera y la última que deben ser ponderadas en forma especial.

- Los parámetros **b_o** y **b_u** que se utilizan para la conversión de particiones donde se calcula el umbral a **SCF bands** son incluidos en la tablas C.8 en los anexos
- Para bloques cortos se utiliza una versión simplificada del cálculo del umbral (SNR constante). Estas constantes se pueden encontrar en las tablas C.7 en lo anexos

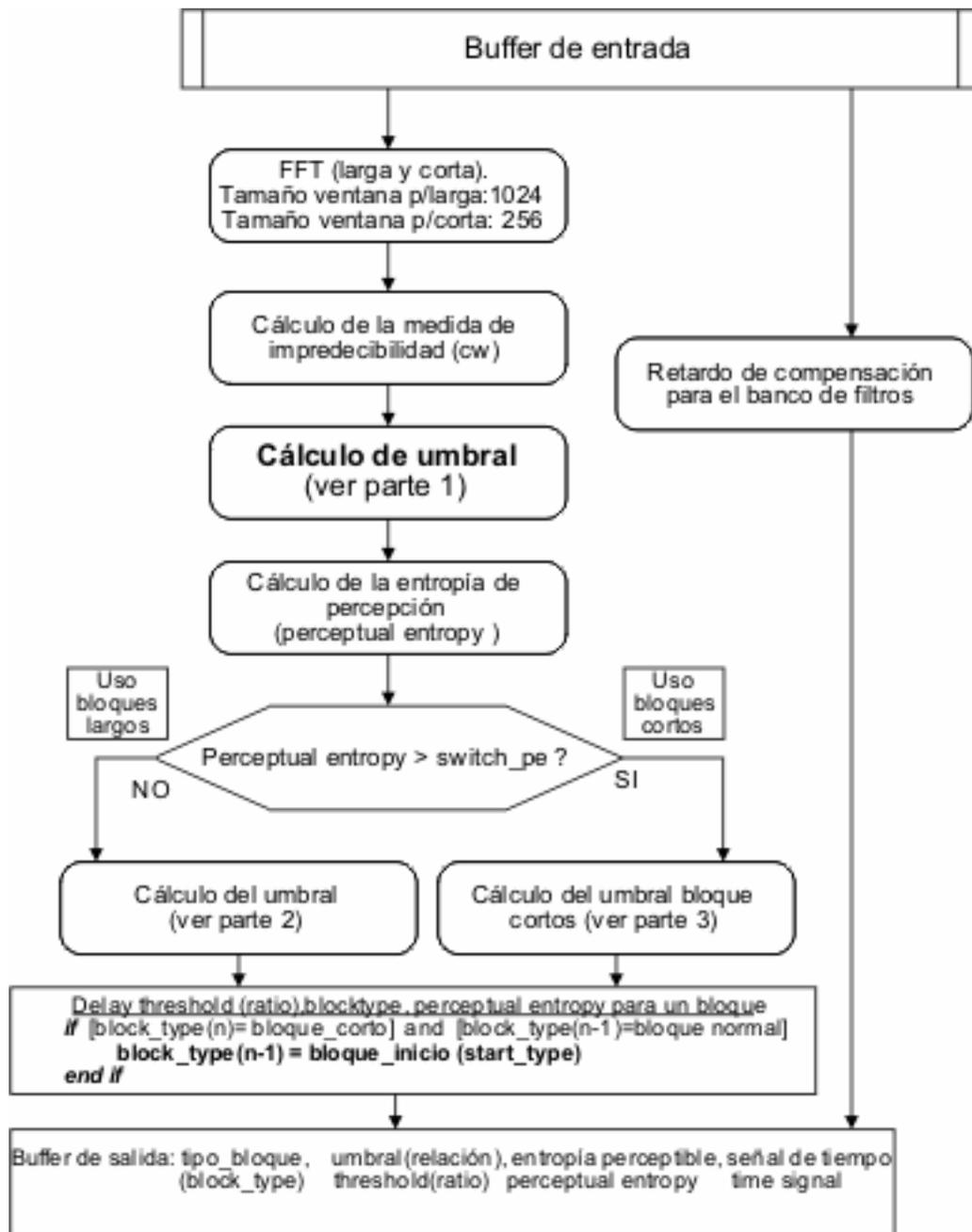


Figura 36. Diagrama de bloques del modelo psicoacústico 2: **Cálculo del Umbral**. Fuente: [13]

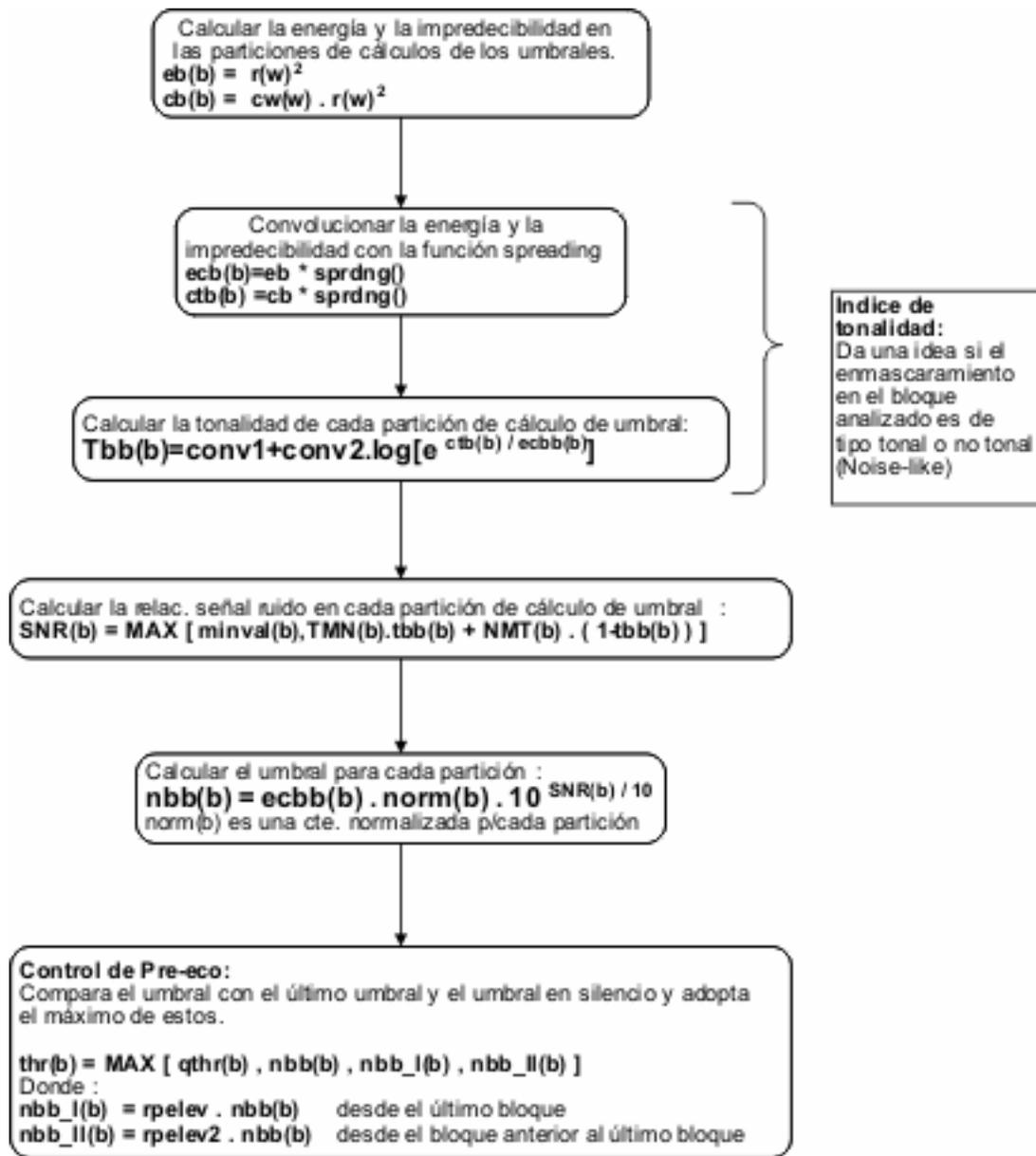


Figura 37. Diagrama de bloques del Modelo Psicoacústico 2: Cálculo del Umbral (Parte 1).

Fuente: [13]

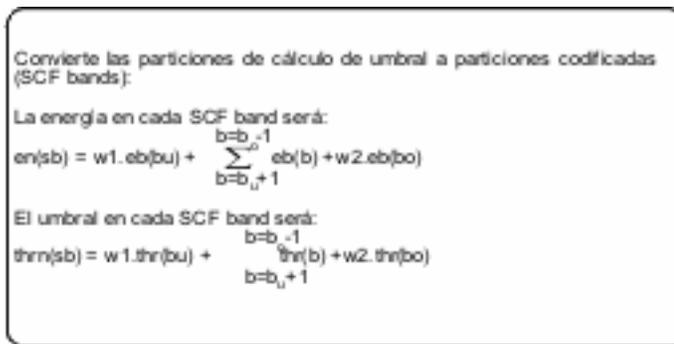


Figura 38. Diagrama de bloques del Modelo Psicoacústico 2: **Cálculo del Umbral (Parte 2).**
Fuente: [13]

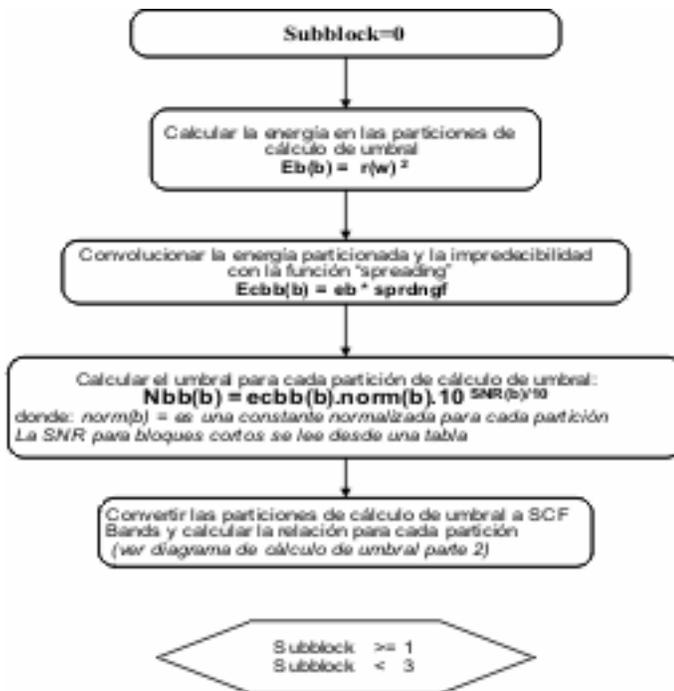


Figura 39. Diagrama de Bloques del Modelo Psicoacústico 2 Capa III. **Cálculo del umbral para bloques cortos. (Parte 3).** Fuente: [13]

CAPÍTULO IV

MEDICIONES DE CALIDAD

Una vez explicado el proceso de codificación perceptiva se debe comenzar a evaluar la calidad del audio obtenido. Las representaciones digitales de formas de ondas analógicas producen la introducción de alguna clase de distorsión la cual puede ser evaluada por medio de, básicamente, dos métodos: Mediciones perceptivas y Mediciones objetivas (No-perceptivas). Un resumen de tales métodos se encuentra en la siguiente tabla [7, 24].

Tabla 7. Diferentes mediciones de calidad de audio. Fuente: propia.

Mediciones en el dominio perceptivo	Mediciones en el dominio objetivo (no-perceptivo)	
	Mediciones en el dominio del tiempo	Mediciones en el dominio de la frecuencia
Modelos subjetivos	Signal-to-noise ratio (SNR)	Log-Likelihood ratio (LLR)
Mean Opinion Score (MOS)	Segmental signal-to-noise ratio (SNRseg)	Log-Area ratio (LAR)
Modelos objetivos	Czenakowski distance (CZD)	Itakura-Saito distance (ISD)
Bark Spectral Distortion (BSD)	Local Noise to Mask relation (local NMR)	COSH distance (COSH)
Modified Bark Spectral Distortion (MBSD)		Cepstral distance (CD)
Enhanced Modified Bark Spectral Distortion (EMBSD)		Short-Time Fourier-Radon Transform distance (STFRT)
Perceptual Evaluation Speech Quality (PESQ)		Spectral Phase Distortion (SP)
Perceptual Evaluation Audio Quality (PEAQ)		Spectral Phase-Magnitude Distortion (SPM)
Measuring Normalizing Block 1 (MNB1)		Total harmonic distorsion (THD)
Measuring Normalizing Block 2 (MNB2)		
Weighted Slope Spectral distance (WSS)		
Signal to Mask Relation (SMR)		

A continuación una breve descripción de las diferentes categorías ampliamente utilizadas.

1. Mediciones en el dominio perceptivo utilizando el puntaje medio de opinión subjetiva o *Mean Opinion Score* (MOS).
2. Técnicas de medición en el dominio no perceptivo
3. Mediciones en el dominio perceptivo utilizando modelos objetivos o también llamadas Técnicas de medición perceptiva.

1. Puntaje medio de opinión subjetiva (MOS).

Para este primer caso, el más popular de los métodos de evaluación es el ‘Puntaje de la opinión media’ **MOS** (Mean Opinion Score) donde de 20 a 50 individuos clasifican la calidad de los codificadores sobre una escala de calidad de 5 puntos, obteniendo una media entre todos los escuchas para la similitud en la percepción entre el audio comprimido y el original [7]. Con la entrada de la radiodifusión de audio comprimido y de grupos de audio MPEG, la ITU-R ha desarrollado un elaborado grupo de reglas para pruebas de audio. [9]

El resultado final de tales pruebas es un juicio promediado (MOS) establecido en la ITU-R con la norma P.800. Se usan dos escalas graduadas en 5 puntos. [7]

(a) Una para medir la calidad de la señal y

(b) Otra para medir sus diferencias respecto de la fuente. Un puntaje de 5 en la escala de diferencia indica que la señal reconstruida no puede ser diferenciada de la fuente (al oído humano), mientras que un puntaje de 4 indica que la señal reconstruida si bien se diferencia, no resulta molesta la diferencia. La escala de diferencia resulta extremadamente útil sobre todo si deben ser evaluados los codificadores con pequeñas distorsiones. [7] Para el valor 3, significa ligeramente molesto, 2 molesto y 1 demasiado molesto.

Los exámenes del formato ISO/MPEG han mostrado que los exámenes de *Triple estímulo*, *referencia oculta* y *doble escondido*, basados en tales evaluaciones (MOS) conducen a resultados fiables, no obstante se han detectado pequeñas diferencias en calidad. [7]

En estas pruebas tres señales A, B y C (*Triple estímulo*) son escuchadas por oyentes adiestrados siendo: A: siempre es la señal fuente; B y C: son la señal reconstruida y la fuente, pero tanto el oyente como el instructor de la prueba desconocen cual es una u otra (*doble escondido*) (*referencia oculta*). Los oyentes deben determinar si B o C son la fuente y en caso de poder distinguirlas calificar la diferencia (de 1 a 5). El lado negativo es que los resultados varían de acuerdo al panel de oyentes y al tiempo en que se realizan (por ser subjetivo) pudiendo los resultados variar mucho entre el promedio y determinados tipo de señal bajo prueba. [7]

Otra desventaja es que con el uso de equipos basados en la tecnología de audio compresión, tales como los reproductores portátiles con memoria, en sí mismo constituye un entrenamiento extensivo, con lo que se puede esperar que con el tiempo cualquiera se convierta en un experto oyente. [9] No obstante los exámenes realizados por MPEG y ITU-R con paneles adiestrados de oyentes condujeron a resultados similares y estables. [7] .



Figura 40. Auditorio utilizado por el Centro de Investigación de Comunicaciones de

Canadá. Fuente [16].

La siguiente gráfica muestra un ejemplo de la evaluación MOS para 7 distintas bandas de audio de prueba que se utilizaron para la comparación en el rendimiento entre Capa II y Capa III [7].

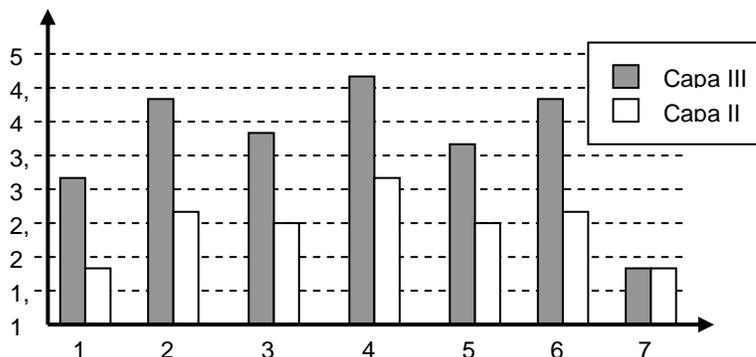


Figura 41. Ejemplo de MOS de la Capa II y Capa III (64 Kbps por canal). Fuente: [7]

Tomando en cuenta que las pruebas de audio tienen una duración de 30 minutos más unos 10 minutos para prepararlo y explicar la metodología, realizar la prueba a unas 50 personas insume aproximadamente unas 35 horas. Además de esto hay que considerar el tiempo que lleva procesar los datos y la generación de las secuencias de audio, por lo que es claro que este método requiere de un enorme esfuerzo en lo que se refiere a los detalles de ejecución y tiempo. Además, tales pruebas no pueden aplicarse dentro de lo práctico o un ambiente de ejecución en la vida diaria.

Por este motivo, para obtener un método de evaluación de audio comprimido por medio de MPEG se considerará otro método, que dependa únicamente de las variables intrínsecas del algoritmo de compresión.

2. Técnicas de medición en el dominio no perceptivo.

Para este caso, las medidas basadas en la percepción hacen uso de los umbrales de enmascaramiento derivados de la señal de ingreso, en orden de compararlos con el ruido de codificación actual del codificador. Por ejemplo, por un

simple criterio objetivo como lo es la relación señal ruido (SNR) como medida de la similitud entre dos formas de ondas (la original y la reconstruida) [7]. Resultados recientes han demostrado que tales medidas dan muy bajas correlaciones entre los puntajes MOS y puntajes objetivos.[7]

Otro conjunto de parámetros, incluyendo relaciones de ruido a máscara locales (local NMR) o la distorsión armónica total (THD) han probado que son fácilmente implementables pero no lo suficientemente precisos, los que las convierte en aplicaciones totalmente inadecuadas para el diseño del codificador y su evaluación.[23]

En muchas pruebas de escucha, Capa III ha demostrado su elevado rendimiento. En noviembre de 1991, tuvo lugar en Hannover la primera de las pruebas de implementación de Capa III en los estudios de la NDR. En éstos, Capa III probó ser notablemente superior a Capa II a igual tasa de bits, superando además a ASPEC el cual, hasta entonces, ostentaba la mejor calidad en códecs de 64 Kbps por canal de audio.[7]

En realidad, Capa III resultó ser el único codec que proveyó de una aceptable calidad de sonido para la codificación independiente a 60 Kbits/seg. Para el test calificado como: “VOZ MASCULINA” el valor medio en la escala de diferencias del ITU-R fue de 4.4 (comparado con 2.4 en Capa II). En general, en tests con altos contenidos de pre-ecos (castañuelas o percusiones) el valor medido (según ITU-R), fue de 3.6 para Capa III (comparado con 2.05 para Capa II).

3. Técnicas de medición perceptiva.

Esta se realiza a través de un criterio complejo que sirve como medición objetiva de la similitud percibida, el cual contempla hechos acerca de la percepción de un oyente humano. [7] Esta será la técnica a implementar en este proyecto.

En 1.984 se iniciaron muchas aplicaciones de modelación Psicoacústica para la predicción de la calidad de los códecs y la audibilidad de cierto tipo de hardware

relacionado. Mientras los niveles de rendimiento no fueron suficientes para lograr masificar estas aplicaciones y las pruebas de escucha directa se iban haciendo imprácticas, la técnica de medición perceptiva ha progresado al punto en el que son muy útiles como complemento a las pruebas de escucha y pueden llegar a reemplazarlas en muchos casos. El Grupo de Tarea 10/4 de la ITU-R trabajó por largo tiempo en la estandarización de una técnica medida perceptiva y produjo una recomendación para un sistema llamado PEAQ (Perceptual Evaluation of Audio Quality). La recomendación define un sistema multi-modos basado en la colaboración de laboratorios líderes en el tratamiento de mediciones perceptivas.[9]

El diseño de métodos de medición perceptiva basados en la percepción humana comienza en la década de los 80', gracias al trabajo de Zwicker, Schröder, Brandeburgo y otros. El método fue el resultado de las investigaciones realizadas en los laboratorios de Canadá, Holanda, Francia y Alemania. El primer algoritmo que fue llevado a un sistema de medida real fue el NMR (Noise to Mask Ratio) en 1.989. Se realizaron muchos otros algoritmos que evaluaban la calidad del ancho de banda en los códecs de audio, debido al auge en el uso que éstos tuvieron en la radiodifusión antes de que se introdujeran en las telecomunicaciones. En 1.996 se estandarizó el algoritmo PSQM (Perceptual Speech Quality Measure) como Rec.P.861, sólo para evaluar audio de voz sobre ancho de banda telefónico (300-3400 Hz). PEAQ se estandarizó en la ITU-R como Rec. BS.1387 para pruebas de ancho de banda de 20 a 20.000 Hz.[14, 15,16]

Cuando se comparan todas las características relevantes de éstos algoritmos, éstas pueden ser resumidas en el siguiente diagrama de bloques. La diferencia más significativa entre ellos es la forma en la que modelan la percepción del oído humano; por lo demás mantienen una alta similitud entre sus componentes. [14]

Esta estructura consiste de 2 entradas, una para la señal de referencia (sin procesar) y otra para la señal bajo prueba, dentro de un modelo perceptivo. Ésta

señal de referencia. En este proyecto será la originada por un compresor de audio perceptivo. En un primer paso de procesamiento es modelado el oído periférico, como un “Modelo perceptivo o “Modelo de oído”. Esta implementación varía enormemente entre los diferentes algoritmos construidos. Algoritmos como PAQM o NMR mantienen un ancho de banda en su algoritmo para evaluar calidad de voz, en cambio **PEAQ utiliza el más exacto, preciso y detallado modelo perceptivo implementado hasta ahora.**[14]

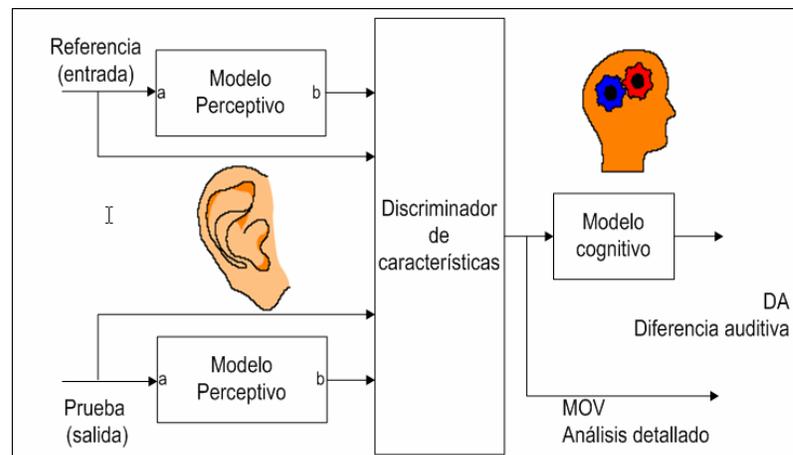


Figura 42. Modelo perceptivo de la plataforma de evaluación de calidad propuesta. Fuente: propia

El modelo psicoacústico empleado en este método produce un número de variables basadas en las comparaciones entre ambas señales. Estas variables serán utilizadas para predecir la calidad subjetiva que deberá ser asignada a la señal procesada, tal como si un oyente la estuviera sometiendo a prueba. La medida de calidad objetiva ha sido calibrada utilizando resultados de las pruebas de escucha realizadas, siguiendo la metodología recomendada por la ITU. [16]

El modelo psicoacústico utiliza la transformada discreta de Fourier para transformar la señal a una representación tiempo-frecuencia. Una versión avanzada utiliza un banco de filtros. Los datos de la TDF son mapeados desde la escala de frecuencia a la escala de tonos, el equivalente psicoacústico de frecuencia. La energía

frecuencia a la escala de tonos, el equivalente psicoacústico de frecuencia. La energía de entrada es distribuida en la región adyacente al tono como una función al nivel de entrada.[16]

En el siguiente paso el algoritmo modela la distorsión presente en la señal bajo prueba comparando las salidas del modelo perceptivo. Las salidas obtenidas por este proceso son llamadas MOV (*Model Output Variables*) las cuales son útiles para un análisis detallado de la señal. El objetivo final consiste en derivar la medida de la calidad en un solo número que indica la audibilidad de las distorsiones presentes en la señal bajo prueba. Para lograrlo se realizan algunos procesamientos adicionales sobre el MOV. [14]

El modelo cognitivo compara las representaciones internas y calcula variables escalares que acumula con el tiempo la actividad del modelo psicoacústico. Se toman diferencias importantes entre las representaciones de frecuencia y tono de la señal de muestra y original. En el dominio de la frecuencia, son medidos los anchos de banda espectrales de ambas señales y se determina el error en la estructura armónica. En el dominio del tono, se miden los errores calculados por la diferencia entre las modulaciones de la señal de excitación y la magnitud de la excitación y la excitación derivada de la señal de error calculada en el dominio de la frecuencia. La medida de calidad se basará en 11 variables en el desarrollo de esta plataforma. [16]

3.1. Modelo de variables de salida (*Model Output Variables* MOV)

El proceso de comparación, basado en el modelo cognitivo, es dividido en varias fases. Cada una da como resultado uno o más MOV, por ejemplo, descriptores de varios procesos cognitivos. La calidad final toma en cuenta todos los MOV y es representado en un solo número, el DA (Diferencia Auditiva).[17]

Tabla 8. Descripción de variables en el Modelo de salida (MOV) del sistema PEAQ básico. Fuente: [17]

Model Output Variable (MOV)	Descripción
WinModDiff	Diferencia promediada en la ventana de modulación (envolvente) entre la señal de referencia y la señal bajo prueba
AvgModDiff1	Promedio de diferencia en modulación
AvgModDiff2	Promedio de diferencia en modulación con énfasis o modulaciones introducidas y cambios en la modulación donde la referencia contenga poca o ninguna modulación.
RmsNoiseLoud	Valor RMS del promedio de nivel de ruido con énfasis sobre los componentes introducidos.
BandwidthRef	Ancho de banda de la señal de referencia
BandwidthTest	Ancho de banda de la señal bajo prueba
TotNMR	Logaritmo del promedio de la relación total ruido a máscara.
RelDistFrames	Fracción relativa de cuadros para los cuales al menos una banda de frecuencia contiene una componente significativa de ruido.
MFPD	Máximo de probabilidad de detección después del filtro pasa-bajo.
ADB	Average Distorted Block (por cuadro), logaritmo de la relación entre la distorsión total y el número total de cuadros distorsionados severamente.
EHS	Errores en el tiempo de la estructura armónica.

Por ejemplo, la medición de la calidad de audio perceptible (PEAQ) se ha aplicado a señales de audio en el ITU-R en pruebas de transmisión de sonido digital obteniéndose una **correlación de 0,98 (98%)** y una **desviación estándar de 0,17 entre las evaluaciones MOS y las mediciones con técnicas perceptivas.**[7] Por tal razón la evaluación de audio en este proyecto se basará en una técnica perceptiva modificada del PEAQ.

Para lograr una implementación acorde con lo esperado en las evaluaciones MOS que son utilizadas como referencia y que por razones prácticas ya descritas no serán evaluadas en este proyecto, se estimará una relación de señal a máscara (SMR) por medio de la evaluación de tres características fundamentales definidas en el próximo capítulo y de las que se extraerán la información suministrada por la implementación del modelo perceptivo del PEAQ.

Fueron considerados otros métodos de evaluación, incluso métodos en el dominio no-perceptivo como por ejemplo, la distancia Ikatura-Saito (ISD) o relación señal a ruido segmentada (*SNRseg*) pero no se implementaron por ser difíciles de interpretar.

CAPÍTULO V

PLATAFORMA DE EVALUACIÓN DE CALIDAD DE AUDIO

1. Introducción

La codificación de audio digital y su transmisión involucra el compromiso entre cuatro parámetros: complejidad, retardo, rata de bits y la calidad percibida del audio decodificado. Las tres primeras de ellas pueden ser cuantificadas de manera directa pero la calidad percibida es más difícil de cuantificar. En tal sentido el *MOS* es la medida más confiable hasta ahora practicada pero con las desventajas de ser económicamente costoso, de consumir mucho tiempo en aplicación y de muy complejo análisis de resultados. Por tal razón, aunque avances importantes en mediciones de calidad audio se soportan sobre test subjetivos, se han realizado numerosos intentos para complementarlos con estimadores en el dominio objetivo (no-perceptivo).

Tal es el caso de los parámetros SNR o SNRseg que pueden proveer de indicadores de calidad de audio en sistemas que introducen diferencias entre la señal original y la codificada. Desafortunadamente estos métodos muestran muy poca correlación con los resultados de audio percibido [25]. Por lo tanto la estimación objetiva de la calidad percibida en audio comprimido ha abierto múltiples posibilidades de estudio debiendo evaluar incluso errores en bits o pérdida total de ciertos cuadros de audio de la trama completa.

Importantes laboratorios de investigación sobre audio a nivel mundial han intentado una aproximación basada en la percepción para crear estimadores que “escuchen” las señales digitales de audio comprimido en la misma forma que lo hace

un ser humano. El más importante de estos estudios es el PEAQ, antes mencionado. Este capítulo concentrará la atención en hallar los parámetros de evaluación perceptiva como una manera de elaborar un estudio de la relación señal a máscara (SMR), comparando la señal original $X(t)$ con la señal degradada $Y(t)$, la cual ha pasado por el compresor perceptivo.

Ambas señales estarán en formato WAV, lo que significa que una vez que la señal original de prueba ha pasado de ser una señal digital sin compresión a una con compresión resultando en un archivo MP3, se vuelve a digitalizar para su correcta comparación. Este último proceso no compromete la calidad del audio comprimido.

El primer paso del evaluador perceptivo es computar los posibles retardos entre las señales. Luego el programa las compara ya alineadas en tiempo bajo el modelo perceptivo. La clave de este proceso es la transformación de ambas señales a una representación interna análoga a la representación psicoacústica del sistema auditivo humano, al igual que lo realiza el compresor estudiado. Para esto toma en cuenta la frecuencia percibida y su volumen. A través de este proceso tres parámetros principales serán los que se combinen para definir el SMR y producir así un MOS objetivo en el dominio perceptivo.

Las características principales que se tomarán en consideración son: (a) estimación de retraso entre la señal original y la que ha sido comprimida, que se denominará degradada; (b) la transformación perceptiva y (c) la apreciación de diferencia o distancia perceptiva entre ambas señales. Este conjunto se logra en varias etapas: alineación en tiempo, alineación en nivel para calibrarlo al nivel de escucha, mapeo en frecuencia y tiempo, análisis en el dominio de la frecuencia por medio de la FFT y un escalamiento en volumen. [26]

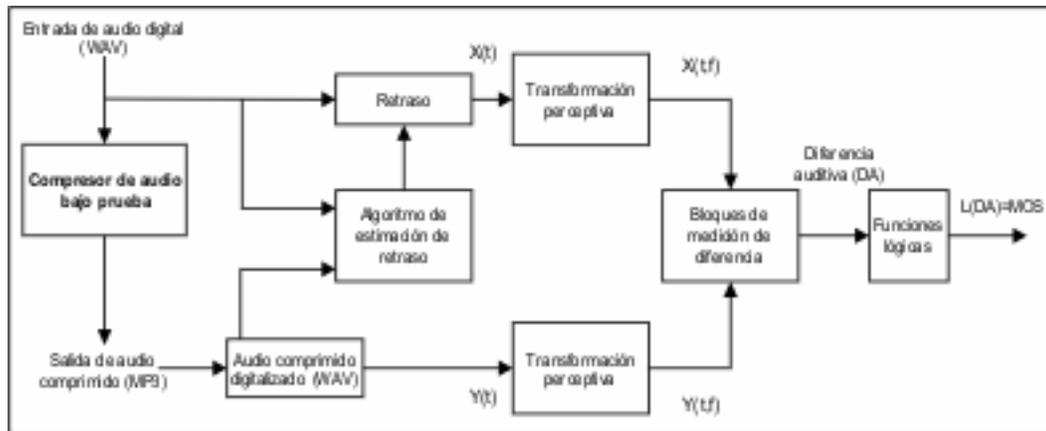


Figura 43. Esquema general de la plataforma de evaluación de calidad audio comprimido.

Fuente: propia

2. Estimación del retraso entre señales de audio.

La plataforma propuesta compara la señal original $X(t)$ con la señal degradada $Y(t)$, la cual ha pasado por el compresor perceptivo. La salida es una predicción de la calidad percibida que le sería dada a $Y(t)$ por oyentes en una prueba de escucha subjetiva. Esto posibilitará la evaluación de otros compresores con una misma plataforma.

Inicialmente se debe computar el retardo entre ambas señales, que son colocadas en vectores contenedores de las muestras de ambas señales. Se realiza un minucioso alineamiento en el tiempo ya que los modelos perceptivos son sensibles a los *offsets* de tiempo, así que es necesario calcular con exactitud el retraso de la muestra. Ambos deberán ser ajustados a un mismo tamaño para luego alinearlos y usar una representación de las señales en el oído humano, tomando en cuenta la frecuencia percibida (Bark) y el volumen (Sones) (Ver anexos, Caracterización del oído).

A fin de obtener la sincronización adecuada se ha creado un algoritmo, recurriendo a la correlación cruzada (*cross-correlation*) entre las señal original y degradada. El punto en el cual las señales poseen una máxima correlación indica el retraso entre ambas señales.

Luego se asume que la señal degradada Y es más larga que la original X, dado el retraso que introduce la compresión perceptiva, por tal razón se ha creado una subrutina que permite colocar ambos vectores a una misma longitud, cortando las muestras sobrantes de retraso.

Esto se computa de la siguiente manera:

- El máximo de correlación en la potencia 0,125 se usa como medida de seguridad en la alineación de cada cuadro. El índice de los máximos produce el retraso estimado en cada cuadro.
- Se realiza un histograma con los estimados de retraso calibrados contra la medida de seguridad. El índice del máximo en el histograma es combinado con el retraso previamente calculado lo cual produce el retraso final estimado.

3. Transformación perceptiva.

El objetivo de la transformación perceptiva es imitar el oído humano de tal forma que es retenida únicamente la información que es perceptivamente relevante. Como ya se ha estudiado en el Capítulo II, la escala de percepción de frecuencia del oído no es uniforme sobre la escala de Hertz y el volumen percibido está relacionado con la intensidad de la señal en escala logarítmica. Estas son las propiedades más importantes que serán modeladas.

Los laboratorios de audio han concluido una simple y efectiva transformación que es construida desde una secuencia de pasos actualmente establecida. La transformación perceptiva es aplicada en la representación del dominio de la frecuencia de las señales de audio, las cuales serán divididas en ventanas, multiplicadas por la ventana de Hamming y entonces transformadas al dominio de la frecuencia usando la FFT. La implementación que se llevará a cabo en este proyecto está basada en una muestra de 44100 muestras/s en cuadros de 1.408 muestras (32 ms) y 50% de solapamiento, “enventanados” por medio de Hamming y cros-correlacionados entre la señal original y la degradada, después de que se ha realizado el alineamiento por envolvente

La resolución no uniforme de frecuencias dentro del oído es tratada por medio del uso de la escala psicoacústica: la escala de frecuencias de Bark. (ver Capítulo II, 8.1) Esta relación se puede ver en la figura 17. La escala de Barks es aproximadamente lineal hasta los 500 Hz. Además existen distintos modelos para la percepción del volumen en función de la intensidad de la señal. En este proyecto se ha escogido la conversión logarítmica para convertir la intensidad de la señal en volumen percibido.

4. Medición de diferencias entre señales.

La medida de la diferencia o distancia entre dos señales de audio percibido se ha establecido gracias la estructura jerárquica del PEAQ antes mencionado. Primero se realiza una medición de tiempo luego una medición en frecuencia. Cada uno de estos bloques toma la entrada original y la transformada perceptivamente, la cual se ha llamado degradada, y devuelve los índices de medidas MOV establecidos en el capítulo anterior.

4.1 El bloque de medida en tiempo.

Este integra sobre la escala de frecuencia ambas señales y mide la diferencia entre ellas y normaliza la salida en múltiples tiempos.

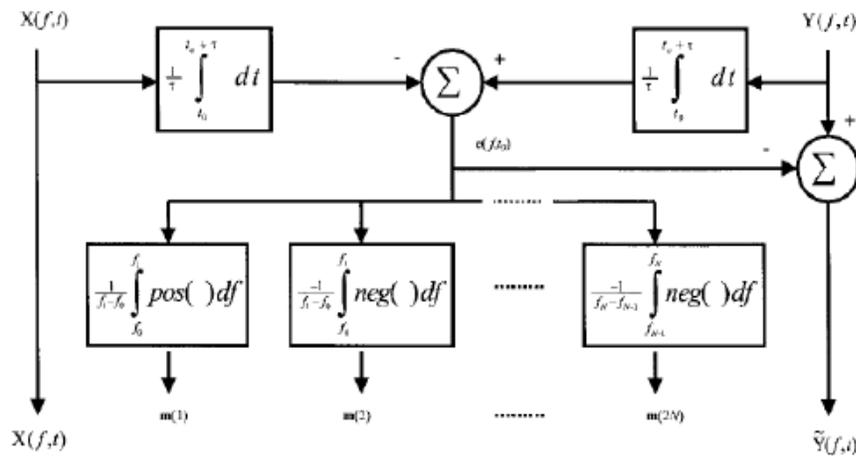


Figura 44. Esquema de medición en tiempo. Fuente [21]

4.2 El bloque de medida en frecuencia

Este integra sobre la escala de tiempo ambas señales y mide la diferencia entre ellas y normaliza la salida en múltiples frecuencias.

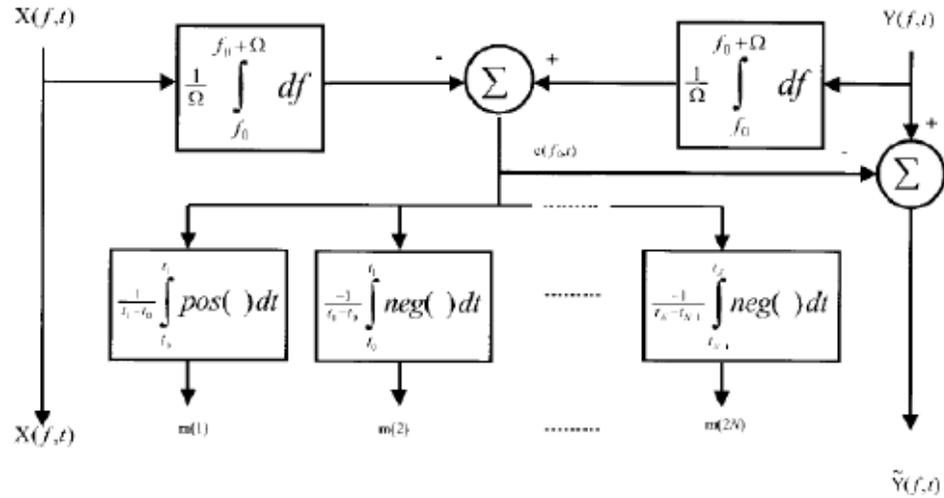


Figura 45. Esquema de medición en frecuencia. Fuente:[21]

Se trabajarán empíricamente ambos modelos, de tal forma que ambas estructuras jerárquicas, que funcionan desde grandes escalas de tiempo y frecuencia serán progresivamente ajustadas a escalas más pequeñas que se acomodan más al patrón de conducta de un escucha humano frente a la reacción de su oído a diferencias y cambios de tono.

Ambas subrutinas de evaluación, junto con la alineación en tiempo conforman el corazón del programa. Estas estructuras de evaluación se aplican a ambas señales, tal como se ve en el siguiente gráfico:

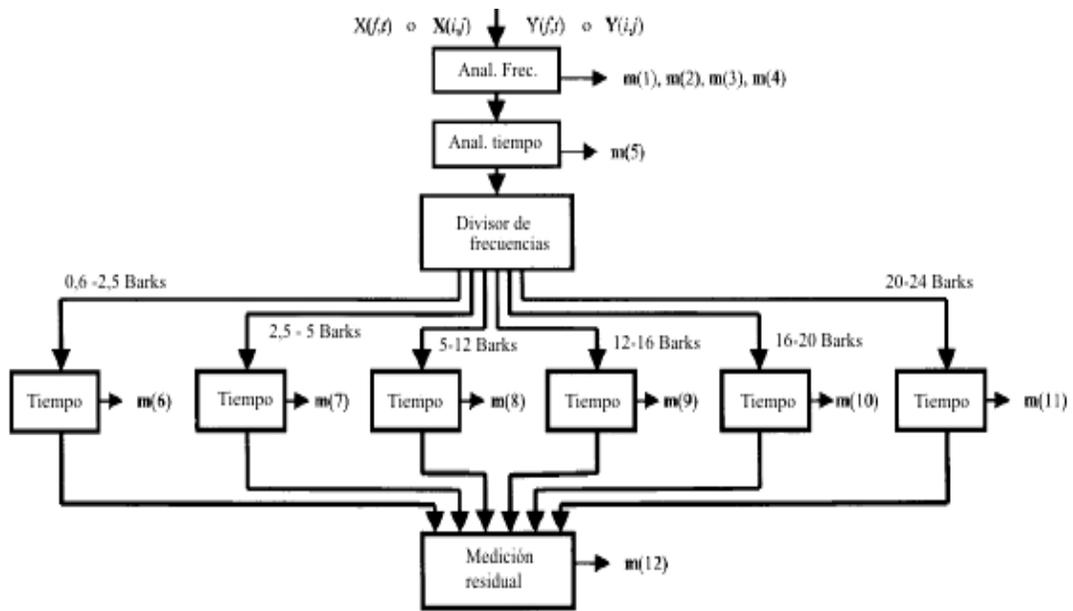


Figura 46. Estructura de evaluación en tiempo y frecuencia de la plataforma diseñada.

Fuente: propia.

Inicialmente son extraídas cuatro medidas las cuales cubren el ancho de banda desde 0 a 20.000 Hz y son guardadas en el vector m . Luego se dividen en escalas de Barks, adaptadas a la escala perceptiva del oído humano, hasta aproximadamente 24 Barks. Finalmente se realiza una medición residual.

Básicamente, a través de estas estructuras se realiza una descomposición de las muestras de audio dinámicamente tal como lo haría en tiempo y frecuencia el oído. Los parámetros resultantes de esta descomposición son linealmente combinados para formar una medición de la diferencia entre ambas señales. Cada medida m tomada a diferentes frecuencias es pesada con diferentes coeficientes ya que el oído no percibe cada frecuencia de la misma forma. Este grupo de valores m está dado en el anexo 10, donde se muestra la completa implementación de la plataforma en Matlab. El valor que arrojan estos resultados conforma una combinación lineal llamada **Diferencia Auditiva (DA)** :

$$DA = \overline{w.m} \quad \text{Ec.5.1}$$

Donde w posee los 11 valores de la estructura estudiada (ver Capítulo IV, tabla 8). En la práctica, cuando ambas señales son idénticas todas las medidas son cero y DA es cero. Cuando las señales mantienen diferencias perceptivas DA se incrementa. Para este proyecto, el uso del valor de DA en el desarrollo de la plataforma de evaluación de calidad de audio percibido se utiliza bajo la prueba de ACR (Absolute Category Rating) o Rateo de categoría absoluto, el cual produce la escala de MOS desde 1 a 5. Para eso se ha desarrollado empíricamente la siguiente ecuación logística cuyas asíntotas están entre 0 y 1:

$$L(DA) = \frac{1}{1 + e^{DA-4,6877}} \quad \text{Ec. 5.2}$$

Con lo cual L decrece en función de DA. La función realiza la compresión de los extremos valores altos o bajos y es aproximadamente lineal en el rango intermedio de calidad.

Hasta aquí se describe el uso de las mediciones para estimar la calidad del audio percibido. Por otra parte no se puede asumir, a través de este algoritmo, que se emule directamente la capacidad de percepción auditiva ni el juicio humano, los cuales son de una complejidad tal que escaparía de los alcances de este proyecto. Esta plataforma sirve para emular la respuesta de un oyente más no da una respuesta de cómo ese oyente podría llegar a determinar una conclusión sobre la calidad.

La implementación directa de la plataforma se halla en el apéndice bajo el código Matlab, junto a la implementación del compresor diseñado siguiendo las instrucciones de la Norma ISO/IEC 11172 – 3 Audio.

5. Resultados del uso de la plataforma de evaluación propuesta.

5.1 Codificadores utilizados

Se propone evaluar el compresor diseñado en conjunto con tres compresores de uso comercial, a saber:

- FreeRip, versión libre para plataformas Windows/XP. Su licencia es *shareware*. Permite compresiones entre formatos WAV, MP3 y OGG Vorbis. Ejecuta el programa con un máximo de bitrate de 128 kbps para compresión de WAV a MP3. (Será reconocido como codificador 1).
- AudioConvert, Utiliza tasas de transferencia comprendidas entre 48 y 128 kbps, para canalización mono y estéreo. Además permite modificar señales en formato WAV, mediante filtrados en frecuencia y cambios en la tasa de muestreo. Permite la conversión entre cualquier par de formatos soportados. La versión es de evaluación por 14 días. (Será reconocido como codificador 2)
- Codificador implementado por medio del modelo psico-acústico de la Norma ISO/IEC 11172 – 3 Audio. (Se reconocerá como codificador 3).

Estos manejan la compresión de audio utilizando la codificación perceptiva. Así comparando el compresor diseñado con éstos se obtiene el objetivo intermedio, la obtención de los parámetros más significativos a modificar para una correcta compresión perceptiva. El objetivo final se logra evaluando el audio comprimido por los diferentes compresores para comprobar la fiabilidad y consistencia de la plataforma de evaluación de audio.

5.2 Audio seleccionado

Se trabajará en un computador Athlon 1,15 Ghz de 512 MB de RAM, con capacidades de instrucciones MMX, 3DNow! y SSE. Se realizaron las pruebas con muestras de audio básicas considerando que el desempeño de la plataforma no ha sido optimizada a nivel de instrucciones y no se ejecuta en lenguaje de bajo nivel, por lo que se sometió a pruebas menos intensas, con audio de máximo 2 segundos de

duración, siempre monofónicas de 44100 Hz, procurando encontrar en ellas múltiples tonos y cambios bruscos en las notas las cuales son difíciles de codificar perceptivamente. Al ser de corta duración se podrá establecer con mejor precisión las posibles fallas y/o mejoras.

Un total de 7 señales utilizadas se encuentran en el disco compacto de soporte de esta tesis. Los archivos son:

- 1) *blip*: Nota alta básica. Se estudia ataque de la nota.
- 2) *error*: Combinación de dos frecuencias principales. Se pretende estudiar como se comporta la plataforma al cambio entre ellas.
- 3) *mail*: Mismo objeto de estudio que la anterior. Frecuencias más bajas.
- 4) *sesion*: Combina una nota alta con una inmediata baja. Se pretende evaluar el preeco entre ambas notas.
- 5) *tapestry*: Estándar de voz femenina hablada, proporcionada por la ITU-R como estándar de evaluación en voz humana, cuyo rango está entre 0,5 y 5 kHz.
- 6) *tec_07*: Recorre el rango de frecuencias de manera pausada. Se utiliza para observar el comportamiento de la plataforma frente a notas puras.
- 7) *utopia*: Notas altas y complejas, introducen densidad sonora con poca distorsión. Es el sonido más complejo estudiado.

Cada archivo posee WAV original posee tres codificaciones distintas, en *bitrate* constante o *CBR*, con un codificador perceptivo diferente a dos diferentes tasas de transferencia, 96 y 128 kbps, generando 42 archivos diferentes. Los archivos poseen la siguiente nomenclatura: *Namegnrate*, donde:

Name: Nombre del archivo

g: Si es el audio original es “orig”, si no es “deg”

n: Número del codificador utilizado.

rate: rata de muestreo del codificador utilizado.

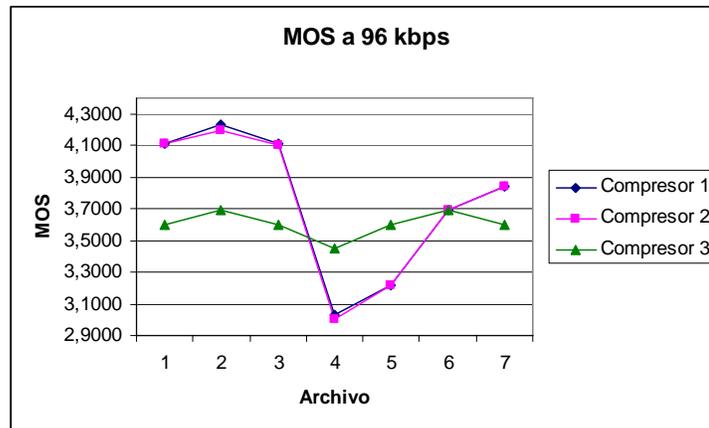
Por ejemplo: Errororig.wav corresponde al archivo de nombre “error” original y Errordeg296 es el mismo pero comprimido con AudioConvert a 96 kbps.

6. Resultados

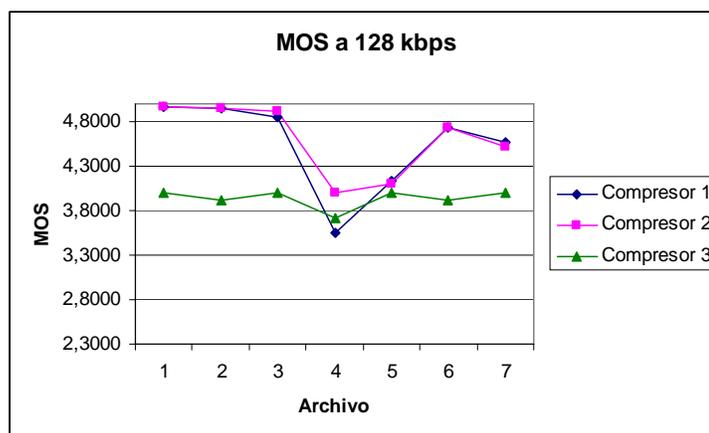
6.1 Respecto al modelo psico-acústico implementado.

El modelo psico-acústico codificador implementado ha funcionado efectivamente, comprimiendo el audio propuesto e incluso algunas formas mucho más complejas y de más larga duración. Su implementación ha seguido las normas básicas propuestas por la norma a pesar de que podrían ajustarse algunos parámetros para su correcta optimización.

Los resultados pueden verse claramente en las siguientes gráficas



Gráfica 1. Comparación de MOS obtenido por compresor a 96 kbps.



Gráfica 2. Comparación de MOS obtenido por compresor a 128 kbps.

El MOS obtenido para el compresor implementado es significativamente menor, comparado con los obtenidos de los productos comerciales. A pesar de esto su correcto desempeño se observa al modificar la calidad del bitrate ya que estando a 128 kbps el MOS promedio está por encima de los 3,8 puntos, en cambio para 96 kbps, se encuentra por debajo de 3,7 puntos.

No se observan incoherencias entre los resultados, sólo una disminución en la calidad obtenida para audio de voz, dado que son compresores especializados en un ancho de banda más amplio que el de la voz humana, acoplándose perfectamente a lo que sería la respuesta del oído.

6.2 Respecto a la plataforma de evaluación implementada.

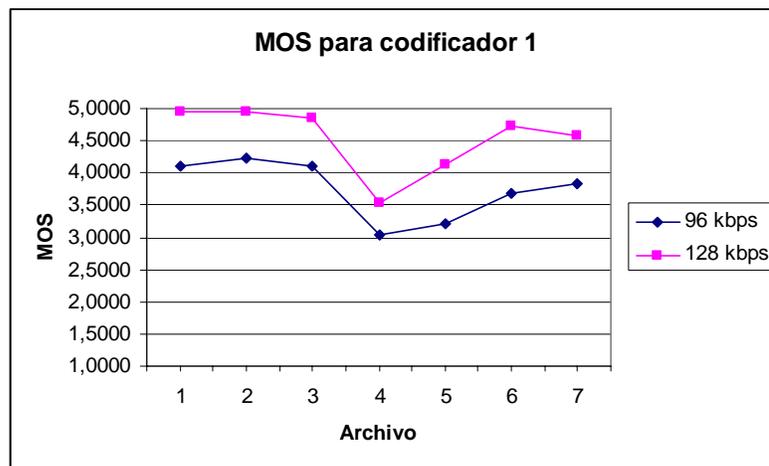
Inicialmente se esperaba que las muestras de audio finalmente estudiadas pudieran ser las sugeridas como estándares por la ITU-R, pero el sistema evaluador propuesto ha resultado de compleja implementación requiriendo largos tiempo de procesamiento. La plataforma demostró lentitud en la evaluación de audio complejo y de larga duración. Por tal razón se realizaron evaluaciones de audio sencillas, de baja complejidad, pero haciendo énfasis en puntos claves de evaluación perceptiva.

Los resultados para esta herramienta demostraron ser coherentes para los 3 diferentes compresores. A baja tasa de transferencia del archivo MP3 creado el MOS obtenido para cada compresor era más bajo que para un alto bitrate. La variación del MOS para un determinado bitrate entre los tres diferentes compresores no depende del compresor utilizado, comportándose estos de manera coherente, si no más bien, dependen del tipo de audio evaluado.

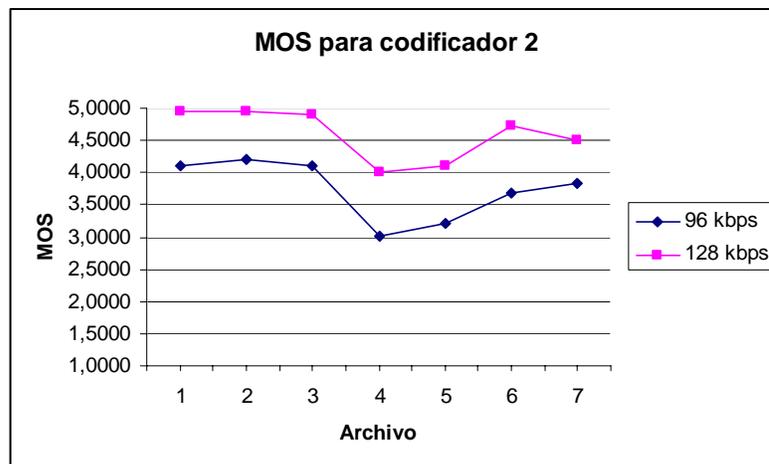
La plataforma ponderó con un MOS bajo la señal de voz, lo que indica que trabaja con mejor efectividad en un rango amplio de frecuencias que para evaluar un ancho de banda de 0 a 5 kHz, correspondiente a la voz. Los resultados para saltos en frecuencia o sonidos complejos fueron satisfactorios al poseer una buena correlación

entre los diferentes resultados obtenidos en cada compresor. Pruebas adicionales indicaron que si el archivo evaluado era el mismo, el MOS resultante era igual a 5, si eran completamente diferentes el MOS resultó ser 1.

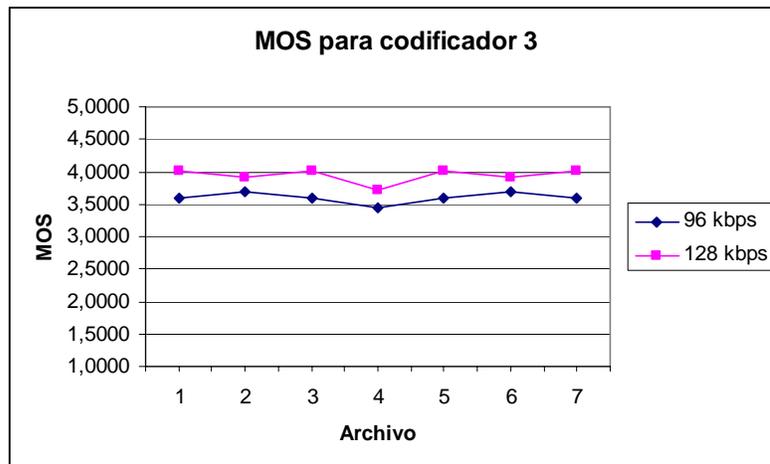
La subrutina de alineación en tiempo para la señal degradada resultó eficaz al permitir la correcta ponderación de la correlación entre las señales original y degradada en todos los casos estudiados. A continuación los gráficos con los resultados obtenidos:



Gráfica 3. MOS obtenido con codificador 1.



Gráfica 4. MOS obtenido con codificador 2.



Gráfica 5. MOS obtenido con codificador 3.

Se aprecia que ambos compresores comerciales poseen similares características de compresión y que el evaluador las interpreta correctamente. Para el compresor 3, implementado de forma básica, el MOS que genera es menor que los anteriores, pero no deja de ser coherente al valorar con mayor ponderación la más alta de tasa de transferencia, lo que significa que la plataforma está funcionando correctamente.

CONCLUSIONES

La calidad de alguna señal de audio que ha estado sujeta a algún tipo de procesamiento ha sido, hasta hace algunos años, determinada por impresiones subjetivas que un oyente normal percibe cuando escucha la señal procesada. Una medida puramente técnica puede dar un estimado muy crudo de la calidad de ese audio a menos que tome en consideración el modelo psicoacústico. Ninguna prueba en el dominio perceptivo o no-perceptivo sustituye en calidad a las pruebas subjetivas de MOS.

Las componentes de la señal que son asumidas imperceptibles no serán transmitidas o codificadas por lo que el ruido de codificación es espectralmente enmascarado de acuerdo a los umbrales de enmascaramiento de la señal de audio. Por tal razón los métodos tradicionales de medición de calidad tales como el SNR, no pueden evaluar estas distorsiones auditivas, por lo que son inútiles cuando se trata de evaluar el rendimiento de este tipo de codificadores.

Los codificadores psicoacústicos están contruidos sobre modelos y teoremas matemáticos que simulan la complejidad del sistema auditivo. Los actuales alcances computacionales y de modelación no pueden cubrir totalmente los fenómenos psicoacústicos. Aún así los codificadores pasan la prueba de calidad dependiendo del contexto en el que se les utilice.

Si algún codificador psicoacústico no está funcionando correctamente es debido a que no se está aplicando para su objetivo fundamental. Recuérdese que su fin es reducir drásticamente la irrelevancia en la señal digital de audio y aumentar la ganancia en la codificación. Los codificadores o compresores psicoacústicos deben ser usados de la misma manera que el oído humano. La complejidad evolutiva del oído se adapta al sonido percibido, lo que lo hace un sistema que *aprende* del medio,

sin tomar en cuenta el procesamiento posterior que realiza cada cerebro humano, lo que lo hace percibir mejor el sonido bajo cierta predisposición consciente. De esa misma manera debe utilizarse un codificador psicoacústico: cada uno tiene una limitación particular que debe ser respetada ya que no posee la capacidad de adaptación del oído.

El caso de este modelo psicoacústico implementado en MATLAB demuestra su mínimo funcionamiento, sin destacarse como el más óptimo. Como codificador de audio es significativamente inferior a los productos comerciales, lo que sugiere una optimización en trabajos futuros. El primer problema es el pre-eco, observado con el audio Nro.4, el cual generó un MOS bajo para cualquier compresor estudiado. Otro problema observado en los modelos psicoacústicos es la introducción de cambios drásticos en el *bit-allocation* frente a cambios bruscos de frecuencias, lo que genera silbidos repentinos o chasquidos en la reproducción. Para resolver ambos, se debe reducir el largo de la ventana de codificación, asemejándola a la del oído humano.

Los codificadores han sido verificados en funcionamiento gracias a la herramienta desarrollada. Es en extremo difícil conseguir plataformas de audio gratis o a costo moderado que permitan la comparación directa de la efectividad del evaluador estudiado. El codificador estudiado ha servido de instrumento para la correcta administración y escrutinio profundo sobre la aplicación de la psicoacústica en el campo de compresión audio digital. Su uso ha sido meramente informativo con lo cual se favorecen futuras mejorías.

Adicionalmente es difícil definir las pruebas que puedan cubrir todas las variedades de audio existentes, abarcar los compresores perceptivos más importantes de audio y además comparar los más significativos evaluadores de calidad de audio. Esto imposibilitaría llevar a buen término en corto plazo la ejecución de este proyecto.

Dentro del modelo psicoacústico de compresión, los módulos que deben ser probados más exhaustivamente son el módulo de cuantificación, la asignación de bits por código Huffman y la capa psicoacústica, los cuales son los más críticos y generan las diferencias en términos de calidad, bitrate y MIPS entre los codificadores. En este caso se han logrado definir pocas pruebas que permitan verificar la funcionalidad del vasto conjunto de componentes que conforman el modelo. Pero la funcionalidad del modelo psicoacústico y por ende de su cuantificador puede ser definido con exactitud sólo cualitativamente. Las pruebas aquí presentadas demuestran su correcta implementación y el adecuado funcionamiento del evaluador que se ha utilizado para definir la calidad en términos del bitrate.

No existe un método establecido en este proyecto para comparar el MOS obtenido por medio de los parámetros objetivos dentro del dominio perceptivo, el cual es objeto del presente estudio, y un MOS subjetivo, obtenido mediante la encuesta de un auditorio de expertos oyentes. Por lo que los índices acá obtenidos ofrecen un MOS equivalente dando una idea de cuan buena ha sido la compresión de las bandas de audio. De esta manera, al no estar dentro de los alcances de este proyecto la comparación de ambos MOS obtenidos, se han utilizado los parámetros objetivos que produce la norma PEAQ para evaluación objetiva de calidad de audio, con un 98% de correlación frente al MOS subjetivo.

Las diferencias auditivas en los casos estudiados no podrían ser percibidas auditivamente lo que enfoca la aplicación de la herramienta de evaluación a un uso de “afinación”, una vez que por medio de cualquier otro método más exigente se hayan definido las diferencias perceptivas de rigor en cada muestra de audio.

Las decisiones que toma el compresor al hacer *switching* entre el tamaño de una ventana y otra introducen un retraso. Éste es introducido inevitablemente en la señal comprimida y debe ser tomado en cuenta en el diseño de cualquier herramienta de evaluación de calidad de audio comprimido.

Para una correcta compresión y evaluación se debe priorizar una correcta ejecución de la división de las bandas críticas, tanto para el banco de filtros, en el caso del compresor, como para la generación de los MOVs, en el caso del evaluador. El concepto de dividir la señal en bandas de Barks y almacenar la energía asociada a cada una de ellas debe realizarse de manera detallada para que haya exactitud cuando se encuentren dos frecuencias iguales dentro de la misma subbanda. Esto se ha observado en la evaluación MOS del audio Nro. 6.

Los diferentes tamaños en el “enventanado” de cada compresor, el tipo de banco de filtro utilizado y los modelos psicoacústicos variantes hacen que varíen los resultados de evaluación. Aún así se observa coherencia en los resultados, generando sólo diferencias intrínsecas a la calidad del audio perceptible.

RECOMENDACIONES

En el caso del modelo psicoacústico planteado:

1. Se debe reducir el largo de la ventana de codificación, asemejándola a la del oído humano.
2. Implementar funciones de enmascaramiento que varíen con la frecuencia, volumen y tonalidad para asemejarlo a la respuesta adaptativa del oído humano.
3. Introducir una capacidad variable de bitrate (VBR) en la codificación, de tal forma que permita cambiar de ventanas con cuadros pequeños con están presentes las transitorias a cuadros más largos cuando el sonido es relativamente constante. Con eso mejoraría el MOS para la señal de voz. Además esto permitiría una reducción significativa de pre-ecos y la reverberación de la voz mientras mantiene una buena resolución en frecuencia en caso de ser necesaria.
4. Introducir evaluaciones para ambos canales estéreo, de tal forma que permitiría que los archivos de música y audio complejo sean codificados más eficientemente y ayudaría a comprender el enmascaramiento bi-auditivo.
5. Se debe mejorar la codificación correcta de las señales transitorias de audio las cuales introducen retrasos importantes en la codificación. Actualmente no se dispone del conocimiento adecuado para sugerir alguna solución.
6. Las pruebas deben realizarlas grupos de trabajo expertos en audio que puedan colocar parámetros fácilmente escalables a fin de ajustar con detalle las variables en conjunto del modelo psicoacústico planteado.

En el caso de la plataforma de audio:

1. Deberán plantearse estrategias para evaluar y comprimir audio de bajo *bitrate* (por debajo de los 16 kbps) ya que el audio se ve comprometido por la introducción del ruido de codificación a bajos niveles de muestreo.

2. Se debe reducir la complejidad computacional y optimizar el código implementado para lograr realizar evaluaciones de tramas completas de audio con algunos minutos de duración.
3. La aplicación debe ser llevada a cabo en lenguaje C el cual es más acorde para el manejo de grandes y complejos procesos computacionales.
4. La plataforma puede ser mejorada a fines de evaluar errores en canales de transmisión, pérdida de tramas, si el audio ha sido codificado a VBR o ruido ambiental introducido.
5. Encontrar una herramienta más compleja, de uso comercial, que pueda servir de comparación frente a la plataforma planteada y generar un método de evaluación más exhaustivo a fin de encontrar sus posibles fallas.
6. Comparar la plataforma desarrollada con una evaluación subjetiva MOS de riguroso análisis estadístico que permita definir su correcto funcionamiento y posibles mejoras.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Carline, Dylan; Edwards, Reuben; Coulton, Paul. *Psychoacoustic properties of multi-channel audio signals*. Universidad de Lancaster, Department of Communication Systems,
[en línea]. <<http://www.cms.livjm.ac.uk/pgnet2001/papers/DCarline.pdf> > [Consulta: 2.006]
- [2] Haykin, Simon. *Sistemas de Comunicación*, 1ra Ed. México: Editorial Limusa, S.A. de C.V., 2.002.
- [3] Ellis, Dan. *Audio Compression & Coding*. Columbia University. Department of Electrical Engineering. [en línea].
<<http://www.ee.columbia.edu/~dpwe/e6820/slides/E6820-L07-coding.pdf> > [Consulta: 2.006]
- [4] Haskell, Barry; Puri, Atul; Netravali, Arun. *Digital Video: An Introduction to MPEG-2*, (libro) 1ra Ed. EE. UU: Editorial Chapman and Hall, 1.997.
- [5] Wikipedia [en línea] <<http://es.wikipedia.org/wiki/Portada>> [Consulta: 2.006]
- [6] Maggiolo, Daniel [en línea].
<<http://www.eumus.edu.uy/docentes/maggiolo/acuapu/int.html> > [Consulta: 2.006]
- [7] Mangiaterra, Sergio. *Audio-mpeg*. Centro de Centro de Investigación de Comunicaciones de Canadá. [en línea]. <<http://www.crc.ca/> > [Consulta: 2.006]
- [8] Dr. Fernández, Luis. *Difusión y Multimedia*. Universidad Central de Venezuela. Clase 2. [en línea]
<(<http://neutron.ing.ucv.ve/Comunicaciones/Asignaturas/DifusionMultimedia/indice.htm>)> [Consulta: 2.006]

- [9] Brandenburg, Karlheinz. *MP3 and AAC Explained*. AES 17th International conference on High Quality Audio Coding. [en línea] <<http://www.aes.org>> [Consulta: 2.006]
- [10] *Masking and Perceptual Coding*, [en línea]
<(http://www.minidisc.org/MaskingPaper.html)> [Consulta: 2.006]
- [11] Garrigós, Rubén; Sandoval, José. *Compresión de video con el estándar mpeg: conceptos básicos*. [en línea]
<(http://www.aniret.org.mx/pdf/articulos/mpeg_basicos.pdf)> [Consulta: 2.006]
- [12] Norma ISO ISO/IEC 1117-3, Part 3: Audio, Anexo D.
- [13] Norma ISO ISO/IEC 1117-3, Part 3: Audio, Anexo C.
- [14] Opticom. *State of the art voice quality testing* [en línea].< www.opticom.de >
[Consulta: 2.006]
- [15] Beerends [Et al.] *Perceptual Evaluation of Speech Quality (PESQ), the new ITU standard for end-to-end speech quality assessment*. Part II – Psychoacoustic model.
- [16] Centro de Investigación de Comunicaciones de Canadá. [en línea].<
<http://www.crc.ca/en/html/aas/home/evaluation/evaluation> > [Consulta: 2.006]
- [17] Marija Salovarda [Et al.] *Comparison of audio codecs using PEAQ algorithm*. Facultad de Ingeniería Eléctrica y Computación. Dpto. de Electroacústica, Zagreb, Unska.
[en línea] <www.carnet.hr/CUC/cuc2004/program/radovi/b1_salovarda/b1_full.pdf>
[Consulta: 2.006]
- [18] Terhardt, E., *Calculating Virtual Pitch*. (Libro) Hearing Research, 1.979.
- [19] E. Zwicker, H. Fastl. *Psychoacoustics. Facts and models*. (Libro) Springer-Verlag,1.990.

[20] Painter, Ted; Spanias, Andreas. *Review of Algorithms for Perceptual Coding of Digital Audio Signals*. **EN:** Review of Algorithms for Perceptual Coding of Digital Audio Signals. Arizona State University. Department of Electrical Engineering. Telecommunications Research Center. April (2000).

[21] Kabal, P. *An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality*. **EN:** An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality. Department of Electrical & Computer Engineering, McGill University. (2002).

[22] Bayer, Stefan. *Mixing Perceptual Coded Audio*. **EN:** Mixing Perceptual Coded Audio Streams. Institute for Electronic Music and Acoustics Graz. University of Music and Dramatic Arts. Graz, Austria. (2001).

[23] Painter, Ted; Spanias, Andreas. *Perceptual Coding of Digital Audio*, **EN:** IEEE, Proceedings of the IEEE--Vol. 88, N° 4 (2000).

[24] Özera, Hamza [Et al.] *Steganalysis of Audio Based on Audio Quality Metrics*. **EN:** Steganalysis of Audio Based on Audio Quality Metrics. Department of Electrical and Electronics Engineering. (2000)

[25] S. Voran. *Objective Estimation of Perceived Speech Quality. Part II: Evaluation of the Measuring Normalizing Block Technique*. **EN:** IEEE transactions on speech and audio processing --Vol. 7, N° 4, pp. 385-390, (1999).

[26] Kabal, P. *An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality*. **EN:** An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality. Department of Electrical & Computer Engineering, McGill University. (2002).

[27] Efferlsberg, W. Steinmetz, R. *Audio Compression*. [en línea].
<<http://www.informatik.uni-mannheim.de/informatik/pi4/stud/veranstaltungen/ws200102/multimedia/mm2e-1-en.pdf>> [Consulta 2.006]

BIBLIOGRAFÍA

Baumgarte, Frank. Improved Audio Coding Using a Psychoacoustic model based on a Cochlear Filter Bank. IEEE transactions on speech and audio processing--Vol.10, N° 7, 2002.

Bayer,Stefan. Mixing Perceptual Coded Audio. Institute for Electronic Music and Acoustics Graz. University of Music and Dramatic Arts. Graz, Austria. 2001.

Beerends [Et al.] Perceptual Evaluation of Speech Quality (PESQ), the new ITU standard for end-to-end speech quality assessment. Part II – Psychoacoustic model. JAES Vol. 50 N°,10 pp. 755-764. 2002.

Brandenburg, Karlheinz. MP3 and AAC Explained. AES 17th International conference on High Quality Audio Coding [en línea]. <<http://www.aes.org> > [Consulta: 2006]

Carline, D., Edwards, R., Coulton, P. Psychoacoustic properties of multi-channel audio signals. Lancaster University. Department of Communication Systems [en línea]. <<http://www.cms.livjm.ac.uk/pgnet2001/papers/DCarline.pdf> > [Consulta: 2006]

Chang, Wen-Whei. Audio Coding Standards. National Chiao Tung University, Department of Computer Science and Information Engineering. Taiwán [en línea]. < <http://www.mp3-tech.org/programmer/docs/AudioCoding.pdf>> [Consulta: 2006]

Centro de Centro de Investigación de Comunicaciones de Canadá [en línea]. < <http://www.crc.ca/en/html/aas/home/evaluation/evaluation> > [Consulta: 2006]

Colome, C., Schmidmer, C., Treurniet, W.C. Perceptual-quality assessment for digital audio: PEAQ. Proceedings of the AES 17th. Internacional Conference, 1.999.

Efferlsberg, W.; Steinmetz, R. Audio Compression. [en línea]. <<http://www.informatik.uni-mannheim.de/informatik/pi4/stud/veranstaltungen/ws200102/>>

multimedia/mm2e-1-en.pdf> [Consulta 2006]

Ellis, Dan. Audio Compression & Coding. Columbia University. Department of Electrical Engineering [en línea].

<<http://www.ee.columbia.edu/~dpwe/e6820/slides/E6820-L07-coding.pdf> > [Consulta: 2006]

Fernández, Luis. Difusión y Multimedia. Universidad Central de Venezuela. Clase 2 [en línea].

<<http://neutron.ing.ucv.ve/Comunicaciones/Asignaturas/DifusionMultimedia/indice.htm>> [Consulta: 2006]

Garrigós, Rubén; Sandoval, José. Compresión de video con el estándar Mpeg: conceptos básicos [en línea]. <http://www.aniret.org.mx/pdf/articulos/mpeg_basicos.pdf> [Consulta: 2006]

Hasegawa-Johnson, Mark. Audio Engineering. University of Illinois at Urbana [en línea]. <<http://courses.ece.uiuc.edu/ece403/>> [Consulta: 2006]

Haskell, Barry; Puri, Atul; Netravali, Arun. Digital Video: An Introduction to MPEG-2, (libro) 1ra Ed. EE. UU: Editorial Chapman and Hall, 1.997.

Haykin, Simon. Sistemas de Comunicación, 1^{ra} Ed. México: Editorial Limusa, S.A. de C.V., 2.002.

Herre, Jürgen. Temporal noise shaping, quantization and coding methods in perceptual audio coding: A tutorial introduction. Fraunhofer Institute for Integrated Circuits FhG-IIS A, Erlangen, Germany. 2000

Kabal, P. An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality. Department of Electrical & Computer Engineering, McGill University. 2002.

Liu, C.M., Lee, W.J., Hong, R.S. A new criterion and 7 bit allocation method for current audio coding standards. Department of Computer Science and Information Engineering. University of National Chiao Tung University, Hsinchu, Taiwan, ROC. 2002 [en línea]. < http://www.mp3-tech.org/programmer/docs/DAFX02_Liu_Lee_Hong_bit_allocation.pdf > [Consulta: 2006]

Maggiolo, Daniel [en línea].
<<http://www.eumus.edu.uy/docentes/maggiolo/acuapu/int.html> > [Consulta: 2006]

Mangiaterra, Sergio. Audio-mpeg. Centro de Investigación de Comunicaciones de Canadá. [en línea]. <<http://www.crc.ca/> > [Consulta: 2006]

Marija Salovarda [Et al.] Comparison of audio codecs using PEAQ algorithm. Faculty of Electronic Engineering and Computing, Department of Electroacoustics. Zagreb, Unska [en línea].
<www.carnet.hr/CUC/cuc2004/program/radovi/b1_salovarda/b1_full.pdf> [Consulta: 2006]

Masking and Perceptual Coding [en línea].
<<http://www.minidisc.org/MaskingPaper.html>> [Consulta: 2006]

Norma ISO ISO/IEC 1117-3, Part 3: Audio, Anexo D.

Norma ISO ISO/IEC 1117-3, Part 3: Audio, Anexo C.

Opticom. State of the art voice quality testing [en línea].< www.opticom.de > [Consulta: 2006]

Özera, Hamza [Et al.] Steganalysis of Audio Based on Audio Quality Metrics. Department of Electrical and Electronics Engineering. The International Society for Optical Engineering. EEUU. 2000.

Painter, Ted; Spanias, Andreas. Perceptual Coding of Digital Audio, IEEE. Proceedings of the IEEE. Vol. 88, N° 4. 2000.

Painter, Ted; Spanias, Andreas. Review of Algorithms for Perceptual Coding of Digital Audio Signals. Arizona State University. Department of Electrical Engineering. Telecommunications Research Center. 2000.

Terhardt, E., Calculating Virtual Pitch. Hearing Research, Wengleinstr.7, D-81477 München, Germany. 1979.

Vakil, Sam. Gaussian Mixture Model Based Coding of Speech and Audio. McGill University, Department of Electrical & Computer Engineering. Montreal, Canada. 2004 [en línea]. <<http://www.comm.utoronto.ca/~vakil/main.pdf>> [Consulta: 2006]

Voran, S. Objective Estimation of Perceived Speech Quality. Part II: Evaluation of the Measuring Normalizing Block Technique. IEEE. Transactions on speech and audio processing .Vol. 7, N° 4, pp. 385-390. 1999.

Wikipedia [en línea] <<http://es.wikipedia.org/wiki/Portada>> [Consulta: 2006]

Zwicker, E.; Fastl, H.. Psychoacoustics. Facts and models. Springer, Berlín, 1990.

GLOSARIO

3DNow!	Nombre que recibe una extensión multimedia creada por AMD.
Adapatative bit allocation	La asignación de bits a las subbandas de manera variable en tiempo y frecuencia, de acuerdo al modelo psicoacústico.
Adapatative noise allocation	Es la asignación del ruido de codificación a las bandas de frecuencias de manera variable en tiempo y frecuencia de acuerdo al modelo psicoacústico.
Alias	Componente de señal espejada proveniente del muestreo con una tasa debajo de la de Nyquist.
Aliasing	En estadística, procesamiento digital de señales y áreas afines, el aliasing es el efecto indeseable que causa que señales continuas distintas se tornen indistinguibles cuando se les muestrea digitalmente. Cuando esto sucede, la señal original no puede ser reconstruida de forma unívoca a partir de la señal digital.
Análisis Filterbank	Banco de filtros del codificador que transforma una señal de audio PCM de banda ancha en un conjunto de muestras de subbandas submuestreadas.
Audio Access Unit	Para los Layers I y II una unidad de acceso de audio se define como la parte más pequeña del bitstream codificado que puede ser decodificado por sí mismo, donde decodificado significa: sonido completamente reconstruido. Para Layer III una unidad de acceso de audio representa la parte del bitstream que es decodificable con el uso de la información principal (main information) previamente adquirida.
Audio sequence	Series de tramas de audio NO interrumpidas en la que los siguientes parámetros no cambian: <ul style="list-style-type: none"> • ID. • Layer • Frecuencia de Muestreo Para Layers I y II: Índice de bitrate
Bark	Unidad de banda crítica. La escala Bark es un mapeo en la escala de frecuencias dentro del rango de audio que se corresponde de manera más cercana con la selectividad de frecuencias del oído humano sobre la banda audible.
Bitrate	Tasa de transferencia de bits. Tasa a la cual el bitstream comprimido es entregado desde el medio de almacenamiento a la entrada del decodificador
Bitstream	Arreglo de bits o trama de bits.
Broadcasting	La radiodifusión (en inglés, <i>broadcasting</i>) es la producción y difusión de señales radioeléctricas de audio y/o video a través de ondas o cable destinadas al público en general o bien a un sector del mismo.
Bound	La subbanda más baja a partir de la cual se utiliza la codificación de estéreo intenso
Cámara anaecoica	Sala especialmente diseñada para absorber el sonido que incide sobre las paredes, el suelo y el techo de la misma cámara, anulando los efectos de eco y reverberación del sonido. 1.

CBR	Constant bit rate. Rata de bit constante.
Codec	Sistema codificador-decodificador.
Coded audio bitstream	Es una representación codificada de una señal de audio como la especificada en la norma ISO 11172/3.
Critical band rate	Función psicoacústica de la frecuencia. A una dada frecuencia audible, ésta es proporcional al número de bandas críticas por debajo de esa frecuencia. La unidad de una banda crítica es el BARK.
Critical band	Es una medida psicoacústica en el dominio espectral la que se corresponde con la selectividad en frecuencia del oído humano. Esta selectividad se expresa en BARK.
Copyright	Derechos reservados por el autor.
Dual Channel Mode	Es un modo donde dos canales de audio con contenidos de programación independiente (por ejemplo: distintos idiomas) son codificados dentro de un bitstream. El proceso de codificación es el mismo que para el modo estéreo.
Emphasis	Filtrado aplicado a la señal de audio antes del almacenamiento o transmisión para mejorar la relación señal ruido en altas frecuencias.
Encoding Process	Proceso, no especificado en la Norma ISO 11172, que lee una secuencia de entrada de muestras de audio y produce un bitstream codificado válido a la salida como se especifica en la norma.
FFT	Fourier Fast Transform. Es un algoritmo, (no una transformada) o rutina de cálculo digital que calcula un espectro a partir de una forma de onda en tiempo.
Filterbank	Es un conjunto de filtros pasabandas que cubren el rango entero de audio.
Fixed segmentation	Es una subdivisión de la representación digital de una señal de audio en intervalos fijos de tiempo.
Forbidden	Indica que un determinado valor (al que se le aplica este término) nunca será usado. En general es para evitar similitudes con palabras utilizadas para sincronismo.
Free format	Cualquier bitrate distinto a los definidos que es menor que el máximo bitrate de cada layer.
Gránulos (Layer II)	Es un conjunto de 3 muestras de subbandas consecutivas (de las 32 subbandas) que se consideran juntas antes de la cuantificación. Corresponden a 36 muestras PCM.
Gránulos (Layer III)	576 líneas de frecuencias que llevan su propia información lateral (side information).
Hann Window	Es una función temporal que se aplica "muestra por muestra" a un bloque de muestras de audio antes de la transformada de Fourier.
Huffman Coding	En método específico de codificación entrópica.
Híbrid Filterbank	Es una combinación de un filtro subbanda y la MDCT (Modified Discrete Cosine Transform).
IEC	International Electrotechnical Comisión. Comisión Internacional Electrotecnia.
IMDCT	Inversa de la transformada modificada discreta del coseno.

Intensity Stereo	Es un método para explotar la irrelevancia y la redundancia en señales de audio estereofónicas basado en conservar solamente la energía de la envolvente de los canales derecho e izquierdo en altas frecuencias.
ISO	International Standards Organization. Organización Internacional de Normas.
ITU	Unión Internacional de Telecomunicaciones (UIT)
Joint Stereo Coding	Es un método que explota la irrelevancia o redundancia estereofónica
Layer	Uno de los niveles en la jerarquía de codificación del sistema de audio definido en la ISO/IEC 11172.
Mapping	Conversión de una señal de audio del dominio tiempo al dominio frecuencia por medio del filtrado subbanda y/o MDCT.
Masking	Enmascarado. Propiedad del sistema de audición humana por la cual una señal de audio no es percibida en presencia de otra señal de audio.
Masking Threshold	Una función del tiempo y la frecuencia debajo de la cual una señal de audio no puede ser percibida por el sistema de audición humana.
MDCT	Transformada Discreta del Coseno Modificada.
MIPS	Millones de instrucciones por segundo
MMX	MultiMedia eXtensions
MS stereo	Es un método para explotar la irrelevancia o redundancia en señales de audio estereofónicas basada en codificar la suma y diferencias de canales en vez de los canales derecho e izquierdo.
NDR	Norddeutscher Rundfunk (North German Broadcasting). Emisora de radio y televisión pública, establecida en Hamburg, Alemania.
Non tonal Component	Componente de tipo ruido de una señal de audio.
Octava	Rango de frecuencias entre dos notas que están separadas por una relación 2:1. El rango de frecuencias audibles por el oído humano es de 20 Hz a 20.000 Hz, el número de octavas que abarca este rango es de $\log_2(20000 / 20) = 9,97$ octavas.
Padding	Relleno. Es un método para ajustar la longitud promedio en el tiempo de una trama de audio a la duración de las muestras de audio PCM correspondientes, por medio de agregar, condicionalmente, un slot a la trama de audio.
PCM	Pulse Code Modulation. Modulación por codificación de pulsos.
Polyphase filterbank	Es un conjunto de filtros pasabandas de idéntico ancho de banda con una interrelación de fase especial, que permite un implementación eficiente del banco de filtros.
Psicoacoustic Model	Es un modelo matemático del comportamiento de enmascaramiento del sistema de audición humano.
Requantization	Re-cuantificación. Decodificación de las muestras de subbandas codificadas de manera de recuperar los valores

	cuantificados originalmente.
RMS	Root Mean Square, Raíz media cuadrada
Scalefactor Band	Es un conjunto de líneas de frecuencia en Layer III que son escaladas por un factor de escala.
Scalefactor Index	Código numérico que se asigna a un scalefactor.
Side Information	Información del bitstream necesario para el control del decodificador.
Slot	El slot es una parte elemental del bitstream. En Layer I consiste de 4 bytes y en layers II y III de 1 byte.
SPL	Sound Pressure Level. Nivel de presión Sonora.
Spreading function	Es una función que describe la propagación en frecuencia de la máscara.
SSE	(Streaming SIMD Extensions) extensión al grupo de instrucciones MMX para procesadores Pentium III,
Stereo Mode	Es un modo donde dos canales de audio, que forman un par estéreo, izquierdo y derecho, son codificados dentro de un bitstream. El proceso de codificación es el mismo que para el modo de doble canal.
Stuffing (bits), Stuffing (bytes)	bits o bytes que se incluyen en el bitstream comprimido que son descartados en el proceso de decodificación. Su función es incrementar la tasa de bits del bitstream.
Subband	Subdivisión de la banda de frecuencias de audio.
Subband filterbank	Es un conjunto de filtros de banda que cubren el rango completo de frecuencias de audio. En norma ISO/IEC 11172 el banco de filtros subbanda es un banco de filtros polifásicos.
Subband Samples	El banco de filtros dentro del codificador de audio genera una representación filtrada y submuestreada del conjunto de muestras de audio de la entrada. Estas muestras filtradas se denominan muestras de subbanda. Por cada 384 muestras consecutivas en el tiempo se generan 12 muestras de subbanda consecutivas (en el tiempo) en cada una de las 32 subbandas ($384/32=12$).
Syncword	Palabra código de 12 bits en el bitstream de audio que identifica el comienzo de una trama.
Synthesis Filterbank	Banco de filtros en el decodificador que reconstruye la señal de audio PCM a partir de las muestras de subbanda.
T1	El sistema del T-Portador. La tasa de transmisión es de 1,544 Mbps.
Tandem	Disposición en “tandem”, cuando se ubican dos elementos en serie, es decir, uno detrás de otro y que cumplen una función similar.
TDF	Transformada Discreta de Fourier.
Tonal Component	Una componente de la señal de audio asimilable a una senoide (y no a un ruido).
Transformada	Una transformada es la operación matemática que convierte una función de un dominio al otro, sin pérdida de información. Por ejemplo, la transformada de Fourier convierte una función de tiempo en una función de frecuencia.
Triplet	Es un conjunto de 3 muestras consecutivas de una subbanda.

	Un triplet de cada una de las 32 subbandas conforma un gránulo.
UIT-D	Sector de Desarrollo de las Telecomunicaciones.
UIT-R	Sector de Normalización de las Radiocomunicaciones (antes CCIR, Comité Consultivo Internacional de Radiocomunicaciones).
UIT-T	Sector de Normalización de las Telecomunicaciones (antes CCITT).
Variable length coding	Es un procedimiento reversible para la codificación que asigna códigos más cortos a los eventos más frecuentes y códigos más largos a los menos frecuentes.
VBR	Variable bit rate. Rata de bits variable.
Ventana	Es una función matemática usada con en el análisis y el procesamiento de señales para evitar las discontinuidades al principio y al final de los bloques analizados.
WAV	Archivo de audio con calidad de audio de un CD.
Window	(Ventana). Número del actual slot de tiempo en el caso de ser <code>block_type=2</code> , $0 \leq \text{window} \leq 2$.

[ANEXO 1]

EL OÍDO. MECANISMO PRINCIPAL DE PERCEPCIÓN SONORA.

1. Sistema Auditivo

La función del sistema auditivo es, esencialmente, transformar las variaciones de presión originadas por la propagación de las ondas sonoras en el aire en impulsos eléctricos (variaciones de potencial), información que los nervios acústicos transmiten al cerebro para la asignación de significados. El sistema auditivo se divide en: (a) Sistema auditivo periférico y (b) sistema auditivo central.

1.1 Sistema auditivo periférico.

El sistema auditivo periférico está compuesto por el oído externo, el oído medio y el oído interno.

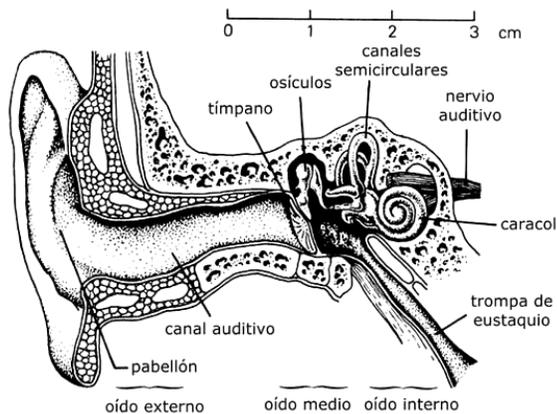


Figura 1: Sistema auditivo periférico. Fuente: Sistema Auditivo, Daniel Maggiolo

El sistema auditivo periférico cumple funciones en la percepción del sonido, esencialmente la transformación de las variaciones de presión sonora que llegan al tímpano en impulsos electroquímicos pero también desempeña una función importante en el sentido del equilibrio. [6]

1.1.1 Oído externo

El oído externo está compuesto por el pabellón, que concentra las ondas sonoras en el conducto, y el conducto auditivo externo que desemboca en el tímpano. La función del pabellón es la de concentrar las ondas sonoras en el conducto auditivo externo. [6]

Las vibraciones acústicas transmitidas por la cadena de huesecillos desde el tímpano llegan al caracol o cóclea a través de la ventana oval. Los huesecillos se comportan como una línea de transmisión dispersiva, de manera que la vibración transmitida se corresponde con la función de auto correlación de la presión acústica. Esto es típico de sistemas inerciales. [8]

La no linealidad de las funciones de transferencia del oído comienzan ya en el pabellón, ya que por sus características éste tiene una frecuencia de resonancia entre los 4.500 Hz y los 5.000 Hz. [6]

El canal auditivo externo tiene unos 2,7 cm de longitud y un diámetro promedio de 0,7 cm. Al comportarse como un tubo cerrado en el que oscila una columna de aire, la frecuencia de resonancia del canal es de alrededor de los 3.200 Hz. [6]

1.1.2 Oído medio

El oído medio está lleno de aire y está compuesto por el tímpano (que separa el oído externo del oído medio), los osículos (martillo, yunque y estribo, una cadena ósea denominada así a partir de sus formas) y la trompa de Eustaquio. [6]

El tímpano es una membrana que es puesta en movimiento por la onda (las variaciones de presión del aire) que la alcanza. Sólo una parte de la onda que llega al tímpano es absorbida, la otra es reflejada. Se llama impedancia acústica a esa tendencia del sistema auditivo a oponerse al paso del sonido. Su magnitud depende de

la masa y elasticidad del tímpano y de los osículos y la resistencia friccional que ofrecen. [6]

El centro del tímpano oscila como un cono asimétrico, al menos para frecuencias inferiores a los 2.400 Hz. Para frecuencias superiores a la indicada las vibraciones del tímpano ya no son tan simples, por lo que la transmisión al martillo es menos efectiva. [6]

Los osículos (martillo, yunque y estribo) tienen como función transmitir el movimiento del tímpano al oído interno a través de la membrana conocida como ventana oval. Dado que el oído interno está lleno de material linfático, mientras que el oído medio está lleno de aire, debe resolverse un desajuste de impedancias que se produce siempre que una onda pasa de un medio gaseoso a uno líquido. En el paso del aire al agua, en general sólo el 0,1% de la energía de la onda penetra en el agua, mientras que el 99,9% de la misma es reflejada. En el caso del oído ello significaría una pérdida de transmisión de unos 30 dB. [6]

El oído interno resuelve este desajuste de impedancias por dos vías complementarias. En primer lugar la disminución de la superficie en la que se concentra el movimiento. El tímpano tiene un área promedio de 69 mm^2 , pero el área vibrante efectiva es de unos 43 mm^2 . El pie del estribo, que empuja la ventana oval poniendo en movimiento el material linfático contenido en el oído interno, tiene un área de $3,2 \text{ mm}^2$. La presión (fuerza por unidad de superficie) se incrementa en consecuencia en unas 13,5 veces. [6]

Por otra parte el martillo y el yunque funcionan como un mecanismo de palanca y la relación entre ambos brazos de la palanca es de 1,31 : 1. La ganancia mecánica de este mecanismo de palanca es entonces de 1,3, lo que hace que el incremento total de la presión sea de unas 17,4 veces. El valor definitivo va a depender del área real de vibración del tímpano. Además, los valores pueden ser superiores para frecuencias entre los 2.000 Hz y los 5.000 Hz, debido a la resonancia del canal auditivo externo y a las frecuencias de resonancia características de los

conos asimétricos, como lo es el tímpano. Entre el oído externo y el tímpano se produce una amplificación de entre 5 dB y 10 dB en las frecuencias comprendidas entre los 2.000 Hz y los 5.000 Hz, lo que contribuye de manera fundamental para la zona de frecuencias en la que el sistema auditivo es más sensible. [6]

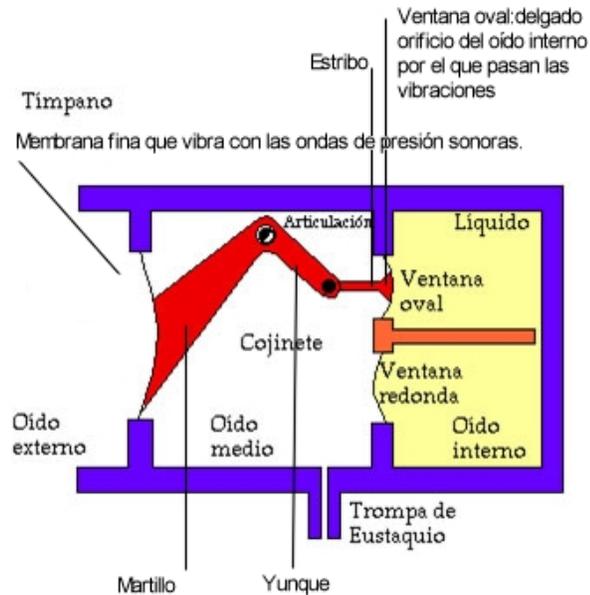


Figura 2. Diagrama del oído medio. Fuente: Sistema Auditivo, Daniel Maggiolo

Los músculos en el oído medio (el tensor del tímpano y el stapedius) pueden influir sobre la transmisión del sonido entre el oído medio y el interno. Como su nombre lo indica, el tensor del tímpano tensa la membrana timpánica aumentando su rigidez, produciendo en consecuencia una mayor resistencia a la oscilación al ser alcanzada por las variaciones de presión del aire. [6]

El stapedius separa el estribo de la ventana oval, reduciendo la eficacia en la transmisión del movimiento. En general responde como reflejo, en lo que se conoce como reflejo acústico. [6]

Ambos cumplen una función primordial de protección, especialmente frente

a sonidos de gran intensidad. Lamentablemente la acción de esos músculos no es instantánea de manera que no protegen al sistema auditivo ante sonidos repentinos de muy alta intensidad, como pueden ser los estallidos o impulsos. Por otra parte, se fatigan muy rápidamente de manera que pierden eficiencia cuando se encuentran expuestos por largo tiempo a sonidos de alta intensidad. [6]

La acción de estos músculos tienen el efecto de un filtro, por cuanto se ofrece una mayor resistencia a la transmisión de frecuencias menores (más graves), favoreciendo por consiguiente las frecuencias mayores (más agudas), que suelen ser portadoras de un mayor contenido de información útil para el ser humano, tanto en el habla como en situaciones de la vida cotidiana. [6]

También el aire que llena el oído medio es puesto en movimiento por la vibración del tímpano, de manera que las ondas llegan también al oído interno a través de otra membrana, la ventana redonda. No obstante la acción del aire sobre la ventana redonda es mínima en la transmisión de las ondas con respecto a la del estribo sobre la ventana oval. De hecho, ambas ventanas suelen moverse en sentidos opuestos, funcionando la ventana redonda como una suerte de amortiguadora de las ondas producidas dentro del oído interno. [6]

La trompa de Eustaquio comunica con la parte superior de la faringe y por su intermedio con el aire exterior. Una de sus funciones es mantener un equilibrio de presión a ambos lados del tímpano. [6]

1.1.3 Oído interno

Si en el oído externo se canaliza la energía acústica y en el oído medio se la transforma en energía mecánica transmitiéndola -y amplificándola- hasta el oído interno, es en éste en donde se realiza la definitiva transformación en impulsos eléctricos [6]. Las vibraciones son transmitidas por medio de la endolinfa hacia los pelos auditivos que flotan en dicho líquido y allí son convertidas en impulsos

nerviosos que son enviados al cerebro.[8]

El laberinto óseo es una cavidad en el hueso temporal que contiene el vestíbulo, los canales semicirculares y la cóclea (o caracol). Dentro del laberinto óseo se encuentra el laberinto membranoso, compuesto por el sáculo y el utrículo (dentro del vestíbulo), los ductos semicirculares y el ducto coclear. Este último es el único que cumple una función en la audición, mientras que los otros se desempeñan en nuestro sentido del equilibrio. [6]

El oído interno está inmerso en un fluido viscoso llamado endolinfa cuando se encuentra en el laberinto membranoso y perilinfina cuando separa los laberintos óseo y membranoso. [6]

La cóclea (o caracol) es un conducto casi circular enrollado en espiral (de ahí su nombre) unas 2,75 veces sobre sí mismo, de unos 35 mm de largo y unos 1,5 mm de diámetro como promedio. El ducto coclear divide a la cóclea en dos secciones, la rama vestibular y la rama timpánica. [6]

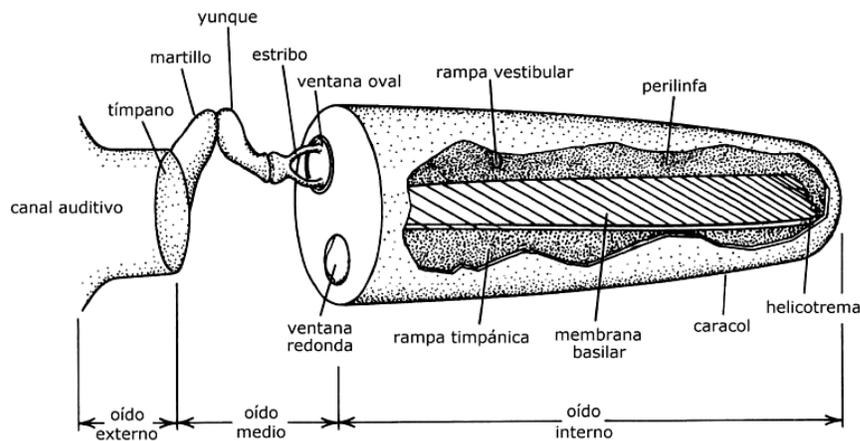


Figura 3. Esquema del sistema auditivo periférico con la cóclea desenrollada. Fuente: Sistema Auditivo, Daniel Maggiolo

La cóclea está dividida a lo largo por la membrana basilar y la membrana de Reissner. [6]

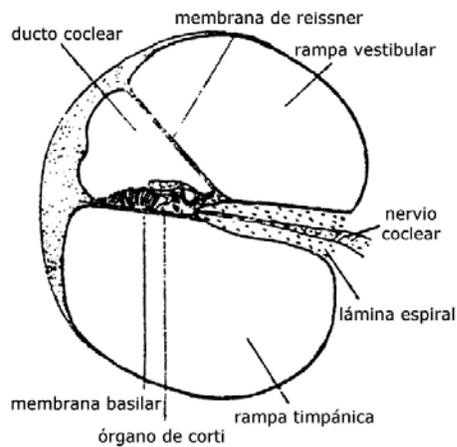


Figura 4. Corte de la cóclea. Fuente: Sistema Auditivo, Daniel Maggiolo

El movimiento de la membrana basilar afecta las células ciliares (también llamadas capilares o pilosas) del órgano de Corti que al ser estimuladas (deformadas) generan los impulsos eléctricos que las fibras nerviosas (nervios acústicos) transmiten al cerebro. Puede haber hasta cinco filas de células ciliares en el órgano de Corti, constando las más largas de unas 12.000 células en fila. [6]

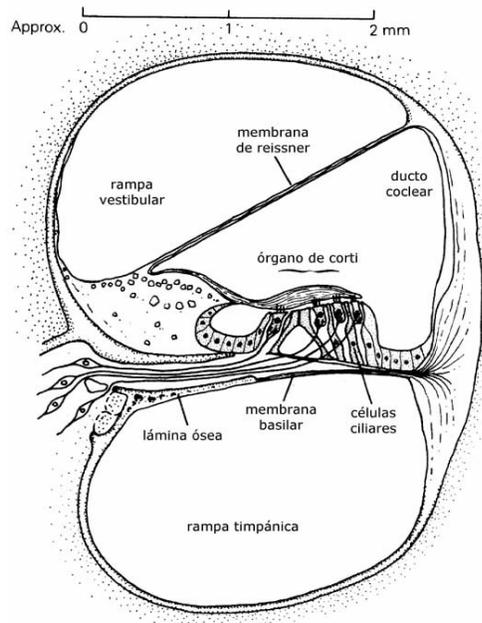


Figura 5: El órgano de Corti. Fuente: Sistema Auditivo, Daniel Maggiolo

La membrana basilar no llega hasta el final de la cóclea dejando un espacio para la intercomunicación del fluido entre la rama vestibular y la timpánica, llamado helicotrema que tiene aproximadamente unos $0,3 \text{ mm}^2$ de superficie. [6]

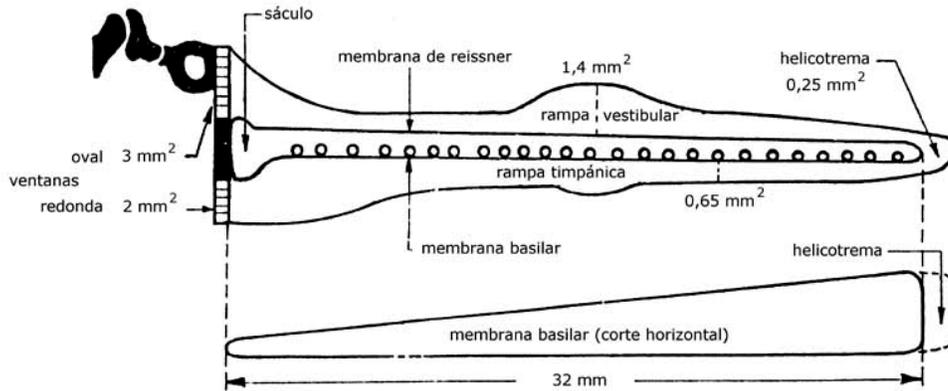


Figura 6. La membrana basilar. Fuente: Sistema Auditivo, Daniel Maggiolo

La membrana basilar se deforma como producto del movimiento del fluido linfático dentro de la cóclea. El punto de mayor amplitud de oscilación de la membrana basilar varía en función de la frecuencia del sonido que genera su movimiento, produciendo así la información necesaria para la percepción de la altura del sonido. Las frecuencias más altas son procesadas en el sector de la membrana basilar más cercano al oído medio y las más bajas en su sector más lejano (cerca del helicotrema). La cantidad de células ciliares estimuladas (deformadas) y la magnitud de dicha deformación determinarían la información acerca de la intensidad de ese sonido. [6]

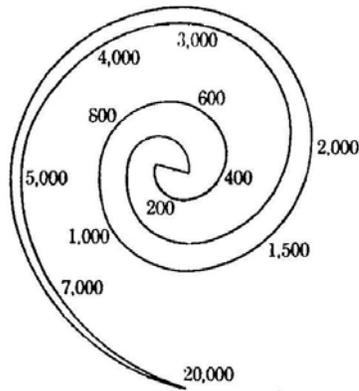


Figura 7: Ubicación de la zona de respuesta de frecuencias sobre la membrana basilar.

Fuente: Sistema Auditivo, Daniel Maggiolo

Los pelos acústicos se comportan como resonadores lo cual permiten descomponer los sonidos en una gran cantidad de sub-bandas, las cuales son enviadas independientemente al cerebro a través del nervio acústico. Sin embargo, las respuestas de frecuencia se solapan parcialmente por lo que un sonido de una frecuencia dada puede excitar a un grupo de pelos en frecuencias adyacentes.(Ver Cap.II.8)

A partir del movimiento de la membrana basilar que deforma las células ciliares del órgano de Corti se generarían patrones característicos de cada sonido que los nervios acústicos transmiten al cerebro para su procesamiento. [6]

Los impulsos nerviosos son de naturaleza electroquímica. El estímulo provocado por la excitación externa crea una reacción dentro de la célula nerviosa que trata de oponerse creando una barrera de potencial. Debido a esto la respuesta es no lineal, presentando efectos de saturación, lo cual puede aproximarse a una función logarítmica. Existe además un retardo de tiempo.[8]

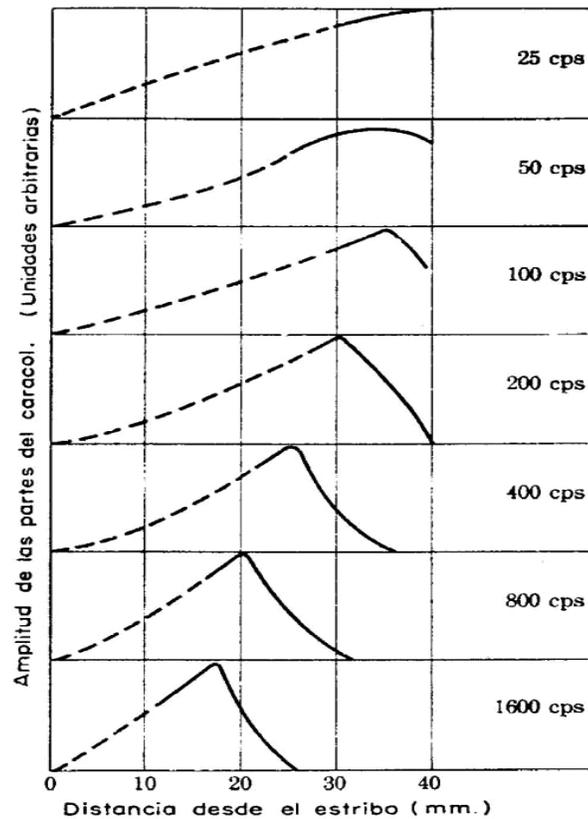


Figura 8. Esquema vibratorio de la membrana basilar. El punto de mayor oscilación depende de la frecuencia. Fuente: Sistema Auditivo, Daniel Maggiolo

1.1.4 Transmisión ósea

Además de a través del oído medio (el tímpano, los osículos), las ondas sonoras llegan al oído interno directamente por medio de la oscilación de los huesos del cráneo. [6]

Ello es fácilmente comprobable si colocamos un diapasón vibrando sobre el parietal o sobre el hueso mastoideo (detrás del pabellón). [6]

Dado que el oído interno se encuentra inserto en una cavidad del hueso temporal las oscilaciones del cráneo hacen entrar en oscilación directamente el fluido linfático, de una manera que no está totalmente clara aún. Lo que sí resulta evidente

es que cualquiera de las dos formas de transmisión de las ondas es igualmente efectiva, sirviendo la transmisión ósea como medio alternativo cuando hay enfermedades en el oído medio. [6]

La transmisión ósea es también la responsable de que el individuo auto-escuche su voz con un timbre distinto al que lo escucha el resto de las personas. [6]

1.2 Sistema auditivo central

El sistema auditivo central está formado por los nervios acústicos y los sectores del cerebro dedicados a la audición. Se trata también de la parte del sistema auditivo de la que menos se conoce. Esto es consecuencia del escaso conocimiento del cerebro y su funcionamiento en general. [6]

A menudo ignorado, el sistema auditivo central es fundamental en la audición, ya que es allí donde se procesa la información recibida y se le asignan significados a los sonidos percibidos, ya sea que pertenezcan a la música, al habla u otros. [6]

El nervio auditivo contendría alrededor de 30.000 neuronas y su función principal es la de transmitir los impulsos eléctricos al cerebro para su procesamiento. También parecen existir otras vías que conducen impulsos desde el cerebro hasta la cóclea. No se sabe mucho de estas neuronas descendentes, pero aparentemente servirían para ayudar a una especie de ajuste de sintonía fina en la selectividad de frecuencia de las células ciliares e incrementar las diferencias de tiempo, amplitud y frecuencia entre ambos oídos. [6]

2. El Cerebro

El cerebro es un órgano electroquímico y su conformación actual en el ser humano es el resultado de transformaciones sufridas a lo largo de millones de años de evolución. No obstante, es una de las partes del cuerpo humano sobre las cuales más

se ignora. En el cerebro hay miles de millones de neuronas, que son esencialmente similares a todas las demás células, pero que tienen la particularidad de recibir y transmitir impulsos eléctricos. [6]

Cada neurona está comunicada con decenas de miles de otras neuronas, conformando todas ellas una red (redes neurales) de intercomunicación sumamente complicada y las conexiones entre ellas, a medida que el individuo se desarrolla, son el producto de procesos de aprendizajes. Esta capacidad de cooperar (trabajar en redes) de millones de pequeñas unidades de procesamiento sería la causa de la alta eficacia y la potencia en el funcionamiento del cerebro. [6]

A partir de la deformación de las células ciliares en el órgano de Corti y a través de los nervios acústicos, el cerebro recibe patrones que contienen la información característica de cada sonido y los compara con otros almacenados en la memoria (la experiencia pasada) a efectos de identificarlos. [6]

Todas señales nerviosas son ensambladas por el cerebro. Los patrones recibidos desde el oído son comparados con otros que se hallan almacenados en el cerebro como si se tratase de una inmensa biblioteca buscando aquellos que presenten el máximo parecido, es decir, los que maximizan la correlación. [6]

Aparentemente, si el patrón recibido difiere de los patrones almacenados, el cerebro intentaría igualmente adaptarlo a alguno de los conocidos, al que más se le parezca. Esto es notable por ejemplo en la percepción de series armónicas. Si se recibe un número determinado de frecuencias aisladas, el cerebro intentará relacionarlas, identificándolas como parte de una serie armónica (aún cuando no lo sean), generando incluso la percepción de la altura determinada por su frecuencia fundamental, aunque ésta no esté físicamente presente y aunque la membrana basilar no esté oscilando en el punto correspondiente a dicha frecuencia. [8]

La memoria es una de las funciones más importantes del cerebro. Cada hecho a ser almacenado en la memoria es separado en partes y se guarda de manera asociativa (modelos asociativos) en diferentes conjuntos de neuronas interconectadas

entre sí, de manera que su ubicación física está distribuida a lo largo de diversas partes del cerebro.[8]

Si el patrón recibido no existe y no es posible encontrar alguno que se le parezca, el cerebro tendrá la opción de desecharlo o de almacenarlo (funciones de las memorias de corto, mediano y largo plazo) convirtiéndolo en un nuevo patrón de comparación.[8]

Aparentemente existirían en el cerebro al menos tres niveles diferenciados de procesamiento de los datos que transmiten los nervios acústicos. En un primer nivel el cerebro identificaría el lugar de procedencia del sonido (asociación de lugar, localización). En un segundo nivel el cerebro identificaría el sonido propiamente dicho, es decir, sus características tímbricas. Recién en un nivel posterior se determinarían las propiedades temporales de los sonidos, es decir su valor funcional a partir de su ubicación en el tiempo y su relación con otros sonidos que lo preceden y lo suceden, hecho de particular importancia en sistemas acústicos de comunicación como el habla (la lengua hablada) o la música.[8]

2.1 Hemisferios cerebrales

El cerebro está dividido en los hemisferios derecho e izquierdo. Todo parecería indicar que en el hemisferio izquierdo se localizan los centros que controlan el lenguaje y las funciones lógicas, mientras que en el derecho se concentran aquellas funciones no verbales, las actividades artísticas y las funciones emotivas. En el caso de la música el procesamiento se llevaría a cabo en el hemisferio derecho. [8]

2.1.1 Procesamiento cerebral de las percepciones acústicas

Como consecuencia del proceso de correlación el cerebro es capaz de caracterizar el sonido de acuerdo a varias variables subjetivas, las cuales guardan relación con las correspondientes variables físicas: el timbre, el tono y el volumen. [8]

[ANEXO 2]

CARACTERIZACIÓN DEL SONIDO

1. El Timbre

El timbre está asociado a la frecuencia fundamental de cada sonido. En el caso de que el sonido sea producido por un instrumento de cuerda la frecuencia fundamental depende del largo de la misma y de su masa. La voz humana es producida por las cuerdas vocales y es afectada por las dimensiones de la laringe. [8]

El timbre es el atributo que permite diferenciar dos sonidos con igual sonoridad, altura y duración [8]. En todo caso, se podría afirmar que el timbre es una característica propia de cada sonido, de alguna manera identificadora de la fuente sonora que lo produce. [6]

Hay diferentes grados de generalización en la consideración del timbre de una fuente sonora, por ejemplo, aquello que diferencia elementos de diferentes clases (por ejemplo, una guitarra de una flauta).[6]

Los principales factores que influyen en la determinación del timbre son:

- (a) La envolvente espectral, es decir, la intensidad relativa de los parciales;
- (b) La envolvente dinámica, en particular la conjunción de las envolventes dinámicas de cada uno de los parciales;
- (c) Los transitorios, que son parciales de muy corta duración que se generan en el ataque, pero también en la caída de un sonido. Ello hace que todos los sonidos tengan siempre una componente de ruido. [6]

El timbre es un fenómeno dinámico, quiere decir que varía en el tiempo. Esto se debe a la evolución de las envolventes dinámicas de cada uno de los parciales que hace que la envolvente espectral (es decir, la intensidad relativa de los parciales)

sea distinta en cada momento. La envolvente tímbrica es la superficie que generan las envolventes dinámicas de todos los parciales que componen ese sonido. [6]

2. El Tono

El tono (y en consecuencia, la tonalidad) está asociado al contenido armónico presente en la excitación acústica. Una misma nota musical emitida por dos instrumentos musicales difiere en su contenido armónico y puede ser reconocida por el cerebro. De igual manera, se puede reconocer la voz de personas conocidas si el canal utilizado tiene suficiente ancho de banda para permitir el paso de todas las armónicas. [8]

3. El Volumen

Finalmente, el volumen está asociado a la cantidad de energía acústica recogida por el oído en un intervalo de integración dado, en lugar de la potencia acústica instantánea. El oído es adaptativo por lo que este intervalo puede variar de 200 a 500 ms. Por esta razón los instrumentos de medición presentan una respuesta logarítmica y una ventana de integración. [8]

El volumen también se denomina sonoridad, como el atributo que permite ordenar sonidos en una escala del más fuerte al más débil.

3.1. Nivel de volumen

Una forma práctica de abordar el problema del volumen es medir el nivel de sonoridad, es decir, determinar cuándo un sonido es igual de fuerte que otro.

Las curvas de igual sonoridad, establecidas por primera vez por Munson y Fletcher en 1.930 (Fig.1) y recalculadas posteriormente por Robinson y Dadson (Fig.2), muestran la relación que debe existir entre las frecuencias e intensidades (o presión sonora) de

dos sonidos senoidales para ser percibidos igual de fuertes, es decir, con la misma sonoridad. [6]

Los sonidos senoidales contenidos a lo largo de cada curva tienen la misma sonoridad. Esta dependencia de la frecuencia estaría dada principalmente por las características de transferencia del oído externo y el medio. También debe notarse que a medida que aumenta el nivel de presión sonora las curvas se hacen más planas, es decir, la dependencia de la frecuencia es menor a medida que aumenta el nivel de presión sonora. [6]

El nivel de sonoridad de un sonido cualquiera (complejo) se determina comparando su sonoridad con la de un sonido senoidal. [6]

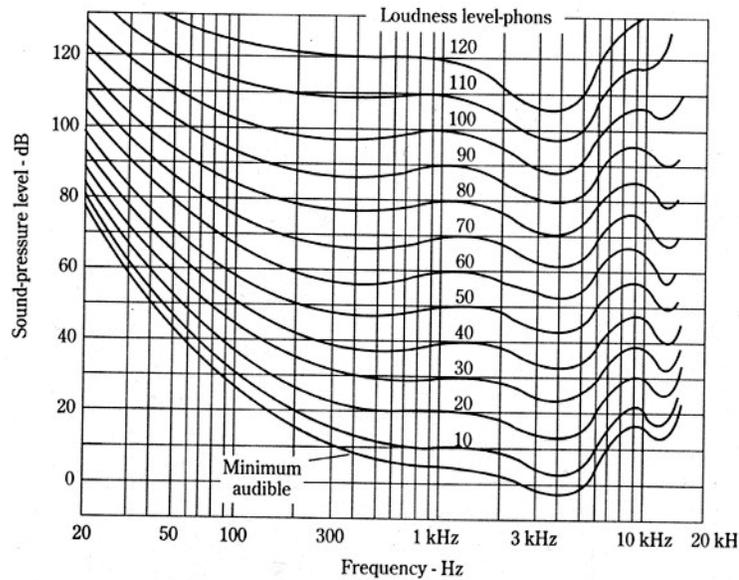


Figura 1. Curvas de igual sonoridad. Fuente: Munson y Fletcher, 1.930.

Para 1 kHz se ha definido que el nivel de presión sonora (en dB) corresponde al nivel de sonoridad (en fon = phon). Así 0 dB es igual a 0 fon y 120 dB es igual a 120 fon. Eso siempre para sonidos senoidales con frecuencias de 1 kHz. Obsérvese, por ejemplo, que un sonido senoidal con (aproximadamente) una frecuencia = 90 Hz y un nivel de intensidad = 40 dB sigue teniendo un nivel de sonoridad = 0 fon. [6]

Las curvas mostradas son válidas para el campo sonoro directo. El sistema auditivo no es sensible por igual a sonidos provenientes de diferentes direcciones. Esa dependencia de la dirección depende, a su vez, también de la frecuencia. Es por eso que las curvas de igual sonoridad no serán iguales en el campo sonoro directo y en el campo sonoro difuso, es decir, en una situación -que es la más usual- en la que el sonido venga de todas direcciones. [6]

La siguiente curva de la figura 2, muestra la corrección necesaria para que un sonido senoidal tenga igual sonoridad en el campo sonoro directo y en el difuso¹, en dependencia de la frecuencia de dicho sonido senoidal. [6]

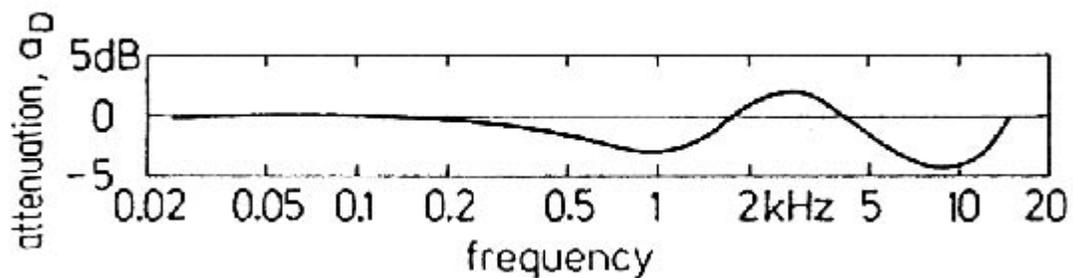


Figura 2. Corrección necesaria para que un sonido senoidal tenga igual sonoridad en el campo sonoro directo y en el difuso. Fuente: E. Zwicker, H. Fastl: Psychoacoustics. Facts and models.

Estas curvas tienen consecuencias directas en la reproducción de sonidos, dado que el balance (de alguna manera tímbrico) interno de los mismos varía según la intensidad con la cual el sonido es reproducido. Si se disminuye el nivel general del sonido, las componentes de frecuencia graves y las más agudas desaparecerán primero, producto de la característica de respuesta del sistema auditivo. [6]

¹ Si los sonidos provienen de diferentes direcciones se refiere al campo sonoro difuso, si viene de una sola, es campo sonoro directo.

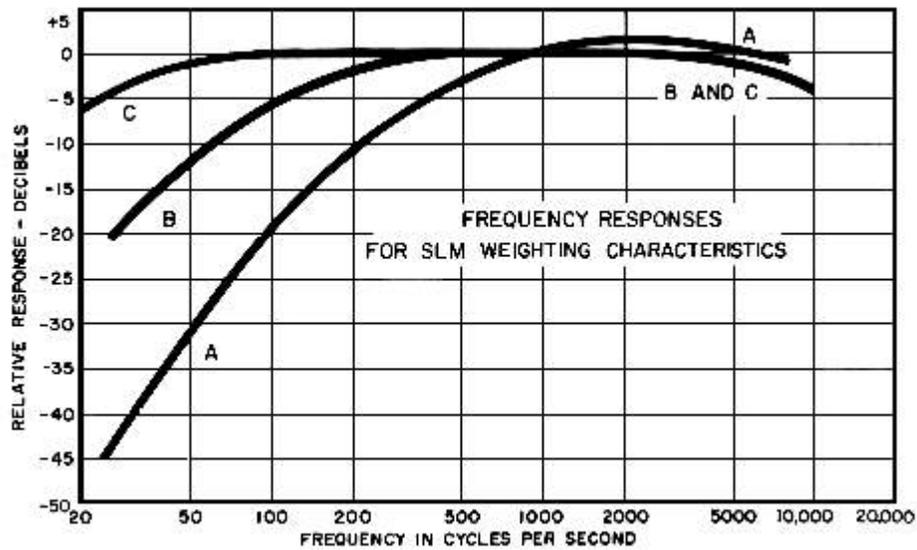


Figura 3. Curvas utilizadas para medidores de nivel de presión de sonido. Fuente: B. Truax: Handbook of Acoustic-Ecology

Estas curvas también fueron utilizadas para diseñar medidores de nivel de presión que respondan a las características del sistema auditivo, a partir de la introducción de filtros similares a la curva de respuesta del sistema auditivo. Se usan los filtros de ponderación con curvas A, B y C (que dan lugar a las escalas de decibeles dBA, dBB y dBC), donde las curvas A y luego la C son las más usadas. [6]

3.2 Escala de sonoridad

Dado que la escala de fons mide el nivel de sonoridad (y, en tanto está relacionada con una escala logarítmica) no es posible comparar los fons de dos sonidos para determinar cual es su relación real de sonoridad. El sone se propone como medida de sonoridad. El sone está definido arbitrariamente como la sonoridad de un sonido senoidal de 1 kHz con un nivel de presión sonora SPL=40dB. [6]

El SPL corresponde a un nivel de intensidad de presión de sonido de un estímulo acústico en decibelios, relativos a un nivel referencial (p_0) definido internacionalmente, esto es:[22]

$$L_{spl} = 20 \log_{10}(p/p_0) \quad \text{donde} \quad p_0 = 20 \cdot 10^{-5} \text{ Pa} \quad (1 \text{ Pa} = 1 \text{ N/m}^2) \quad (2.1)$$

Los experimentos han sugerido que la sonoridad percibida es una función exponencial de la intensidad física:

$$L = k \cdot I^{0,3} \quad (2.2)$$

Donde L es sonoridad, k una constante que depende del sujeto del experimento y de las unidades usadas e I intensidad. En una primera aproximación se puede afirmar que una duplicación de la sonoridad corresponde a un incremento de la intensidad en 10 dB. [6]

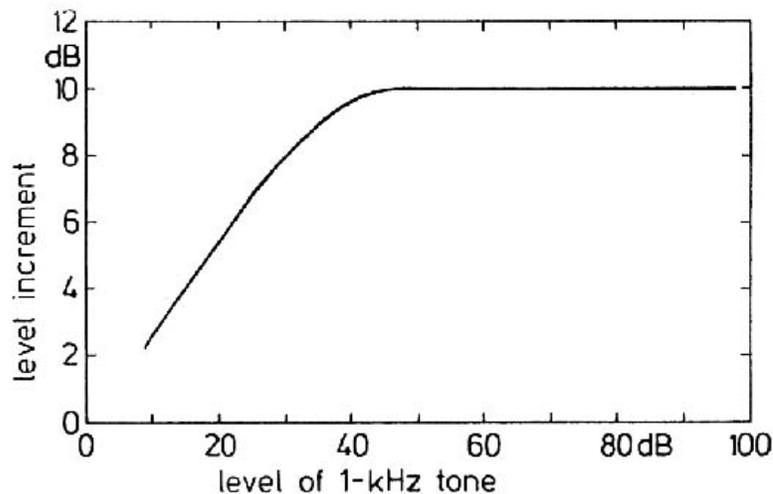


Figura 4. Sonoridad vs. Intensidad. Fuente: E. Zwicker, H. Fastl: Psychoacoustics. Facts and models

Esta relación es válida para sonidos con niveles de 40 dB o más, de manera que por ejemplo un sonido senoidal de 1 kHz con un SPL = 50 dB tendrá 2 sones, es decir, tendrá el doble de sonoridad que el mencionado anteriormente. Sin embargo, como se observa en la curva de la figura 4, para sonidos con niveles por debajo de los 40 dB la función planteada anteriormente no se cumple, y la sonoridad cambia más rápidamente con la variación de SPL. [6]

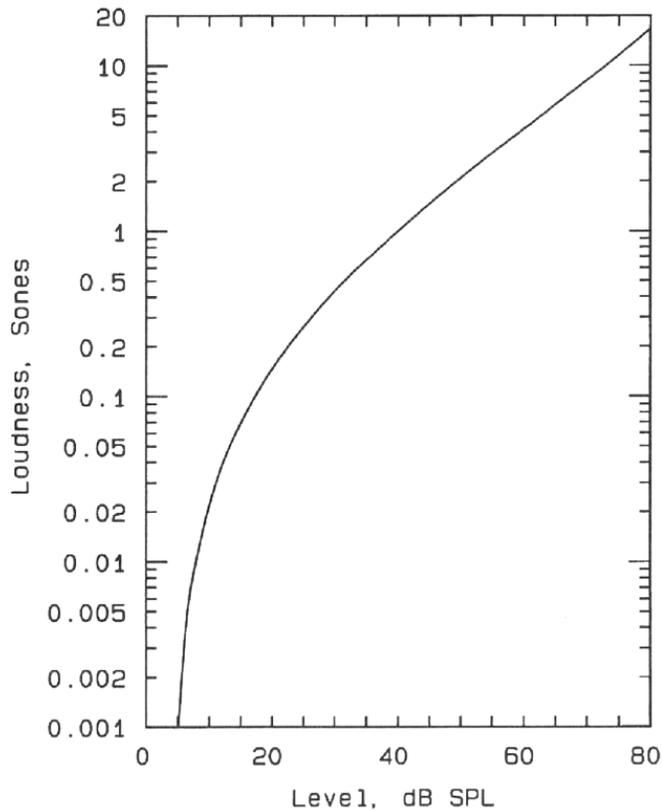


Figura 5. Sones vs. Nivel de presión a 1 kHz. Fuente: B.J.C. Moore: An Introduction to the Psychology of hearing.

La figura 5 muestra la relación existente entre la sonoridad en sones de un sonido senoidal de 1 kHz y su nivel de presión sonora o lo que es lo mismo, en este caso, su nivel de sonoridad en fones. [6]

3.3 Sonoridad y duración

Tanto los umbrales absolutos de los sonidos como la sonoridad dependen de la duración de los mismos. Para sonidos más largos que 500 ms el umbral parece ser independiente de la duración. Para sonidos más cortos que los 200 ms la intensidad necesaria para detectar un sonido crece de manera inversamente proporcional a la duración del sonido. [6]

Algunos experimentos han sugerido que el tiempo de integración del oído

disminuye con el aumento de la frecuencia, pero otros han mostrado que esto no es cierto. La sonoridad de los sonidos de corta duración podría depender también de la energía total, pero los experimentos han mostrado resultados bastante variados como para extraer conclusiones definitivas. [6]

Es bastante probable que el sistema auditivo integre en realidad la actividad neuronal y no la energía misma del estímulo. Para sonidos de mayor duración, el sistema auditivo tendría simplemente mayor cantidad de oportunidades de detectar el estímulo. [6]

Para duraciones de entre 15 - 150 ms el oído integraría la energía sonora a efectos de la detección del sonido, siendo la detectabilidad constante. Por consiguiente, podría decirse que el umbral dependería solamente de la cantidad total de energía contenida en el estímulo y no de la manera en la cual dicha energía está distribuida a lo largo del tiempo. No obstante, esto no es válido para duraciones mayores de los 150 ms o menores a los 15 ms. [6]

4. La resolución temporal

La resolución temporal del sistema auditivo es importante porque prácticamente todos los sonidos varían en el tiempo y porque la información, tanto en música como en el habla, está dada fundamentalmente por el ordenamiento temporal de los sonidos y por las transformaciones que se producen en el tiempo. [6]

Se ha estudiado la capacidad del sistema auditivo de detectar: (a) interrupciones en un sonido, (b) variaciones de la señal a lo largo del tiempo y (c) variaciones en la duración de los estímulos (umbrales diferenciales de duración).

No se hará referencia aquí a la problemática de la percepción del ordenamiento temporal de los sonidos y cómo ellos se combinan para generar significados. Si bien éste es un fenómeno fundamental en los casos de comunicación acústica (música o habla), merecen un capítulo especial aparte. [6]

La resolución temporal depende de :

- El análisis de los patrones temporales que ocurren dentro de cada canal de frecuencia
- La comparación de patrones temporales entre distintos canales de frecuencia.

Una dificultad en la medición de la resolución temporal del sistema auditivo humano radica en el hecho de que variaciones en los patrones temporales traen aparejadas variaciones en los patrones espectrales que también son usadas para discriminar la variación. [6]

Por ejemplo un sujeto puede discriminar un "click" de dos "clicks" separados por apenas unas decenas de microsegundos. El click y los dos clicks tienen diferencias espectrales especialmente detectables en las frecuencias altas. Si se incluye un ruido a efectos de enmascarar frecuencias mayores que los 10 kHz, el umbral aumenta significativamente, lo que indica que eran esas pistas (las variaciones espectrales) las que estaban siendo utilizadas primordialmente para detectar la variación general. [6]

5. Localización

La localización define la capacidad del individuo de determinar la ubicación de una fuente sonora en el espacio. [6]

La localización sólo es posible a partir de la audición binaural. Con un solo oído no es posible localizar fuentes sonoras. El sistema auditivo utiliza un conjunto de pistas para determinar la ubicación de la fuente sonora en el espacio. Por lo general se establecen tres planos característicos en los experimentos destinados a estudiar la localización por parte del ser humano. [6]

6. Percepción de altura

La altura es el atributo que permite ordenar sonidos en escalas de más grave a más agudo. [6]

Hay dos teorías que intentan explicar la percepción de altura: la Teoría del lugar y la temporal. La Teoría del lugar afirma que el oído interno realiza una especie de análisis espectral, correspondiendo la altura que se percibe al lugar en el cual se produce la mayor excitación de la membrana basilar. [6]

La Teoría temporal relaciona la altura percibida con patrones de tipo temporal en el estímulo, en particular su período. La actividad neuronal se realizaría en una relación con dichos patrones temporales. [6]

La Teoría del lugar, si bien ha sido bastante comprobada en su funcionamiento hasta ahora, no es del todo comprensible cuando se trata de sonidos complejos, dado que se excitan muchos puntos de la membrana basilar a la vez, y el punto de mayor excitación no tiene por qué corresponder a la frecuencia correspondiente a la altura que se percibe. La frecuencia fundamental de un sonido no tiene por qué ser el parcial con mayor amplitud (ni siquiera tiene que estar presente). [6]

7. Duración

Existe una duración objetiva, que es la duración de los sonidos posible de ser medida físicamente. La unidad usada suele ser el segundo. Pero existe también una duración subjetiva que es la duración en la que se perciben los sonidos. Suele usarse la unidad "dura" y se ha definido a 1 dura como la duración subjetiva de un sonido senoidal de 1 kHz, con 60 dB de SPL y 1 s de duración objetiva. Duplicando y reduciendo a la mitad se puede determinar la relación existente entre las duraciones objetivas y subjetivas. [6]

La figura 6 muestra la relación entre la duración objetiva y la subjetiva (en

escalas logarítmicas). Puede observarse que la relación de proporcionalidad (curva de 45°) se mantiene para sonidos de duración más larga, pero que a partir de los 100 ms de duración objetiva, la duración subjetiva aumenta un poco con respecto a la objetiva. [6]

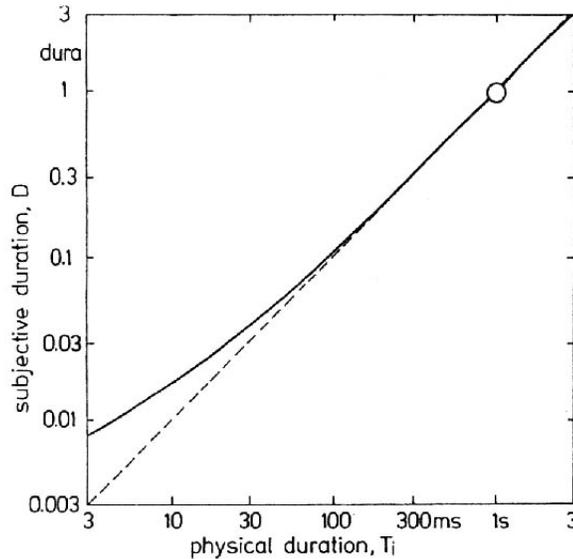


Figura 6. Relación entre la duración objetiva y subjetiva. Fuente: E. Zwicker, H. Fastl: Psychoacoustics. Facts and models

Dado que duraciones objetivas muy cortas pueden producir cambios drásticos en el espectro del sonido, se realizaron experimentos similares con ruido blanco, cuya duración puede reducirse hasta los 0,3 ms sin que se altere mayormente la forma espectral. Los resultados fueron similares a los obtenidos con sonidos senoidales, por lo que puede afirmarse que la relación entre duración objetiva y subjetiva es la que muestra la curva. [6]

7.1 Sonido y silencio

Los resultados se hacen más drásticos si se comparan sonidos con silencios. La duración objetiva y subjetiva de sonido y silencio coinciden si ambos tienen

duraciones mayores de 1 s, pero no es igual si los sonidos son más cortos. [6]

7.2 Discriminación de la duración

Los experimentos de discriminación de cambios en la duración de un sonido o de un intervalo de silencio rodeado por sonidos han mostrado que para duraciones de $T > 10$ ms el umbral dT aumenta con T . El umbral dT es bastante independiente de las propiedades espectrales de los sonidos. [6]

Esto es cierto tanto para el caso de sonidos, como para la discriminación de variaciones en la duración de un intervalo de silencio rodeado (precedido y sucedido) por sonido, siempre y cuando los sonidos sean idénticos a ambos lados del silencio. El umbral dT aumenta para niveles bajos de presión sonora y también cuando los sonidos alrededor del silencio tienen diferentes niveles o frecuencias. [6]

8. Bandas críticas

Puede entenderse la banda crítica como la mínima banda de frecuencias alrededor de una frecuencia determinada que activan la misma zona de la membrana basilar. [6]

Se puede considerar entonces al sistema auditivo periférico como un conjunto de filtros pasa-banda, con bandas superpuestas. El funcionamiento mismo de la membrana basilar sería el responsable de ello. Las propiedades generales del proceso de análisis de frecuencia realizado por el sistema auditivo tienen su base en el concepto de banda crítica. [6]

Si bien muchos experimentos se realizan considerando que los filtros auditivos tienen una forma rectangular, éstos tienen en realidad una forma Gausseana

como lo muestra la figura 7. Para niveles sonoros moderados el filtro auditivo es prácticamente simétrico (en una escala lineal de frecuencias). Para niveles sonoros altos, el lado de las frecuencias bajas del filtro se hace menos empinado que el de las frecuencias altas. [6]

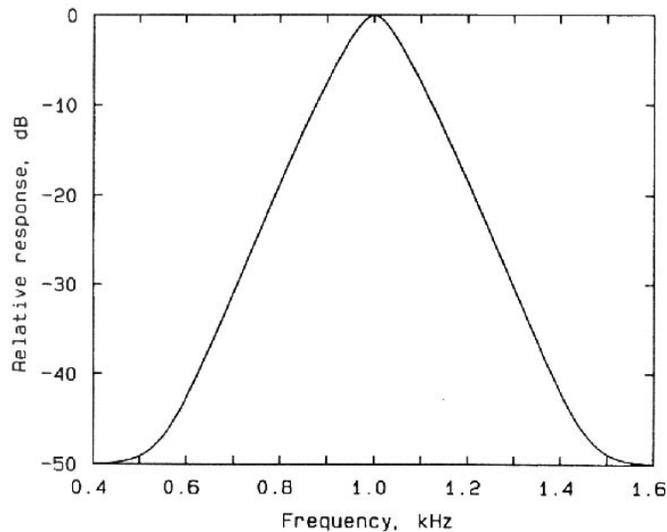


Figura 7. Forma de los filtros auditivos. Fuente: Fuente: Sistema Auditivo, Daniel Maggiolo

La explicación fisiológica del mecanismo de las bandas críticas, si bien todavía incierto, tiene su fundamento seguramente en la capacidad de resolución de la membrana basilar. Existen muchas similitudes entre la selectividad de frecuencia medida en la membrana basilar y la medida psicofísicamente. [6]

8.1 Escala de banda crítica

La figura 8 muestra el ancho de banda crítico en función de la frecuencia. La banda crítica es más o menos constante de unos 100 Hz hasta frecuencias de 500 Hz. (Recuérdese que el enmascaramiento de un sonido senoidal por ruido blanco muestra que el mismo es independiente de la frecuencia hasta los 500 Hz.). [6]

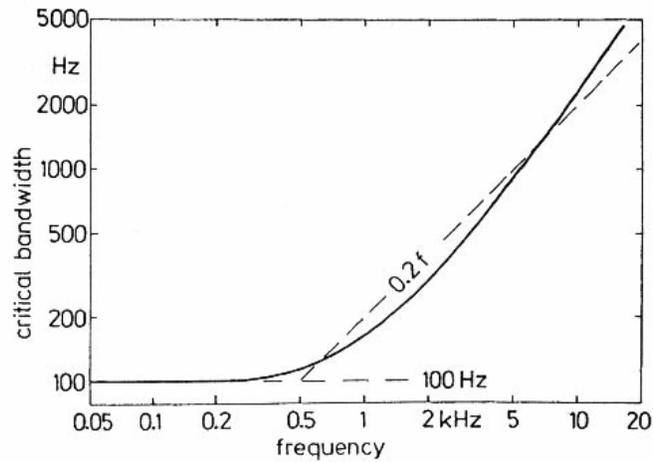


Figura 8. Ancho de banda crítico en función de la frecuencia. Fuente: E. Zwicker, H. Fastl: Psychoacoustics. Facts and models

Por encima de los 500 Hz puede asumirse una aproximación de $0,2 \cdot f$ (20%) de la frecuencia central para las bandas críticas [6]. Sigue la siguiente ecuación [19]:

$$BW_c(f) = 25 + 7 \left[1 + 1,4 \left(\frac{f}{1000} \right)^2 \right]^{0,69} \quad (\text{Hz}) \quad (2.1)$$

Tabla 1. Valores de las bandas críticas, su frecuencia central y las dos frecuencias de corte.

Fuente: Bandas Críticas, Daniel Maggiolo

Nº de banda crítica	Frec. central (Hz)	Frec. de corte inferior (Hz)	Frec. de corte superior (Hz)	Ancho de la BC(Hz)
1	50	0	100	100
2	150	100	200	100
3	250	200	300	100
4	350	300	400	100
5	450	400	510	110
6	570	510	630	120
7	700	630	770	140
8	840	770	920	150
9	1000	920	1080	160
10	1170	1080	1270	190
11	1370	1270	1480	210
12	1600	1480	1720	240
13	1850	1720	2000	280
14	2150	2000	2320	320
15	2500	2320	2700	380
16	2900	2700	3150	450

17	3400	3150	3700	550
18	4000	3700	4400	700
19	4800	4400	5300	900
20	5800	5300	6400	1100
21	7000	6400	7700	1300
22	8500	7700	9500	1800
23	10500	9500	12000	2500
24	13500	12000	15500	3500
25	19500	15500	20000	4500

La escala de bandas críticas muestra que nuestro sistema auditivo está dividido en 25 bandas críticas, cada una de aproximadamente un tercio de octava de dimensión. Estas pueden verse en la siguiente figura. [6]

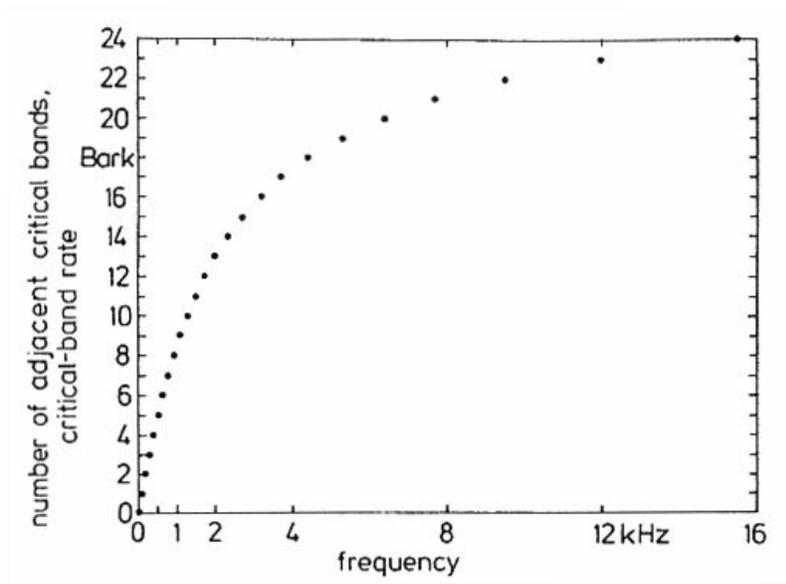


Figura 9. Espaciamiento entre las bandas críticas, Barks vs frecuencia. Fuente: E. Zwicker, H. Fastl: Psychoacoustics. Facts and models

Cada banda crítica correspondería a una distancia fija a lo largo de la membrana basilar, de aproximadamente 1,3 mm de longitud (independientemente de la frecuencia central) abarcando unas 150 células receptoras en el órgano de Corti, de un total de 3.600 células capilares que hay en línea entre el helicotrema y la ventana oval bajo la membrana basilar. A cada banda se le denomina unidad de Bark [6] y se rige por la siguiente ecuación, convirtiendo frecuencia (en Hz) a Barks [19]:

$$z(f) = 13 \arctan(0,00076f) + 3,5 \arctan\left[\left(\frac{f}{1000}\right)^2\right] \text{ (Bark)} \quad (2.2)$$

Tabla 2. Unidades de Bark y su correspondencia auditiva. Fuente: Bandas críticas. Daniel Maggiolo.

24 Bark	32 mm	640 pasos	2.400 mel	3.600 células capilares
1 Bark	1,3 mm	27 pasos	100 mel	150 células capilares
0,7 Bark	1 mm	20 pasos	75 mel	110 células capilares

8.2 Banda Crítica y discriminación de parciales en un sonido complejo

La discriminación de frecuencia tiene que ver con la capacidad del sistema auditivo de resolver componentes senoidales dentro de un sonido complejo. Se pueden detectar los primeros entre 5 y 8 parciales (armónicos) de un sonido complejo. A partir de allí los parciales caerían dentro de una misma banda crítica. Se ha establecido también que los músicos (con entrenamiento del oído) son capaces de realizar una detección más fina que quienes no poseen dicho entrenamiento, aún dentro de lo que sería una banda crítica. [6]

9. Límites de las percepciones acústicas. Conclusiones

Todas las variables estudiadas, conjuntamente con las características propias del aparato auditivo permiten definir como límites al funcionamiento del oído un ancho de banda de 20 a 20.000 Hz y una gama dinámica de aproximadamente 100 dB para dimensionar adecuadamente los equipos de comunicaciones.

Gran parte del mundo que envuelve al ser humano contiene una enorme cantidad de sonidos que actúan simultáneamente. El individuo automáticamente desempeña la tarea de distinguir cada uno de los sonidos atendiendo a aquel de mayor importancia, a menos que haya alguno que quiera escuchar pero no pueda; el individuo no tomará en cuenta todos y cada uno de los sonidos que no escuche. Con frecuencia es difícil escuchar un sonido cuando alguno más alto está presente. Este proceso parece intuitivo, pero en la Psicoacústica y en los niveles de percepción se

convierte en un sistema muy complejo.

El término para este proceso es Enmascaramiento y es probablemente el fenómeno más estudiado en la percepción auditiva.[10]

El enmascaramiento es definido por la ASA como la cantidad (o el proceso) por el cual el umbral de audibilidad para el sonido es elevado por la presencia de otro sonido (enmascarador) (B.C.J. Moore 1.982, p.74). El término fue originalmente prestado de los estudios sobre la visión, conociendo la falla para reconocer la presencia de un estímulo en la presencia de otro a un nivel normalmente adecuado para realzar la primera percepción (Schubert 1.978, p. 63).[10]

El enmascaramiento cae dentro de los estudios psicoacústicos que buscan determinar de qué manera la presencia de un sonido afecta la percepción de otro sonido. [6]

Se habla de enmascaramiento cuando un sonido impide la percepción de otro sonido, es decir, lo enmascara. Se produce una modificación (desplazamiento) del umbral de audibilidad en el sujeto. El enmascaramiento se produce, por ejemplo, cuando dos personas están conversando y el sonido del tráfico impide que una escuche total o parcialmente lo que está diciendo la otra. También puede darse en un conjunto musical, cuando la dinámica de un instrumento (o la suma de varios) impide percibir los sonidos que está produciendo otro instrumento musical. [6]

Si se aumenta de manera constante el nivel de un ruido (enmascarador) se podrá percibir también una transición continua entre un sonido de prueba audible y uno enmascarado. Es decir, existe también un enmascaramiento parcial, en el cual el nivel de percepción del sonido de prueba disminuye, pero no desaparece. [6]

Estos parámetros se estudiarán con profundidad en el capítulo destinado a la Psicoacústica.

[ANEXO 3]

DIGITALIZACIÓN DE AUDIO. NOCIONES BÁSICAS.

Las señales de audio transmitidas por aire de forma natural son ondas analógicas. La forma de las ondas de un sonido así como la percepción que de ésta se tiene mediante el oído son la base fundamental para trabajar sobre la compresión de audio. Un sonido está formado por un conjunto de ondas periódicas superpuestas de las que ya se ha establecido se perciben auditivamente por medio de el timbre (frecuencia), tono y volumen (conocido también como sonoridad). [8]

Para los fines de almacenamiento, procesamiento y transmisión en los sistemas digitales, la información asociada al audio es transformada a la representación digital. La clásica forma de lograrlo es por medio de la modulación por código de pulso (PCM). Consiste en tres pasos, que serán revisados brevemente: Muestreo, Cuantificación y Codificación. Dicha revisión definirá algunos patrones del oído humano para la correcta percepción del audio digital.

Dentro de los tres procesos inherentes a los sistemas digitales incursos, en la transmisión de la información se emplean dos recursos primarios: la potencia transmitida y el ancho de banda del canal. La primera corresponde a la potencia promedio de la señal transmitida. El ancho de banda del canal se define como la banda de frecuencias destinada a la transmisión de la señal. El objetivo es utilizar estos recursos de la forma más eficiente posible. Para el caso de la manipulación del audio digital se hace necesaria la manipulación efectiva de la limitación por ancho de banda. La limitación de potencia se restringe al consumo de los circuitos de cálculo de los codificadores y decodificadores y la potencia inherente a la transmisión de la señal digital.

1. Muestreo

Se describe en el dominio del tiempo. La señal analógica es muestreada periódicamente. Para cada intervalo de muestreo el valor analógico de la señal, que puede ser registrado como un nivel de voltaje, es marcado con un número real. Después del muestreo, la señal ya deja de ser continua y toma valores discretos en la dimensión temporal. La señal entonces es convertida en una secuencia correspondiente de muestras que suelen espaciarse uniformemente en el tiempo. [2]

Para lograr la reconstrucción de la señal analógica original sin pérdida se necesita una frecuencia mínima de muestreo. El mínimo de frecuencia de muestreo f_A está dado por el teorema de Nyquist (1.924): “Para los canales libres de ruido la frecuencia de muestreo f_A debe ser el doble de la más alta frecuencia que ocurre en la señal f_S ”, o también: “Una señal limitada en banda de energía finita, que no tiene componentes de frecuencia mayores que W hertz, puede recuperarse completamente a partir del conocimiento de sus muestras tomadas a una velocidad de $2W$ muestras por segundo”. [2] Esto es:

$$f_A = 2 \cdot f_S \quad \text{Ec.1}$$

En la práctica, una señal puede no estar limitada en banda así que puede darse cierto grado de submuestreo. Para evitar este efecto de traslape se utilizan dos medidas correctivas:

- (a) Antes del muestreo se emplea un filtro pasobajas de antitraslape para atenuar aquellas componentes de alta frecuencia de la señal que no son esenciales para la información que esta misma transporta.
- (b) La señal filtrada se muestrea a una frecuencia un poco mayor que la de Nyquist. [2]

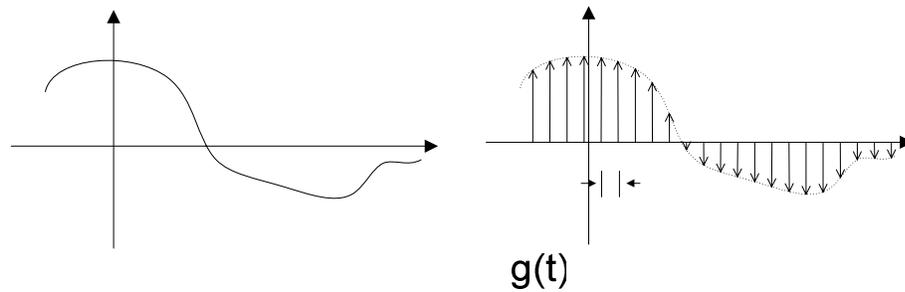


Figura 1. Proceso de muestreo. a) Señal analógica. b) Versión muestreada instantáneamente.

Fuente: Haykin, Simon. Sistemas de Comunicación.

2. Cuantificación

t

El rango de valores en la señal analógica es ⁰ subdividido en un número dado de valores discretos. Una señal continua tiene un intervalo continuo de amplitudes y por lo tanto sus muestras tienen un intervalo continuo de amplitud. En otras palabras, dentro del intervalo de amplitud finito de la señal se encuentra un número infinito de niveles de amplitud. De hecho, no es necesario captar las amplitudes exactas de las muestras. Cualquier sentido humano (el oído o el ojo), como receptor final, puede detectar únicamente diferencias de intensidad finitas¹, lo cual quiere decir que la señal continua original puede aproximarse mediante una señal construida a partir de amplitudes discretas elegidas de un conjunto disponible sobre una base de error mínimo[2].

La existencia de un número finito de niveles de amplitud discreta es una condición básica de la modulación por codificación de pulsos. Si se asigna con claridad los niveles de amplitud discretos con un espaciamiento bastante próximo, es posible lograr que la señal aproximada prácticamente no se distinga de la señal continua original [2].

¹ Ver enmascaramiento “por adelanto” o “progresivo”, Cap.I, el cual resulta de la acumulación de excitación neurológica la cual puede ocurrir hasta por unos 200 ms. Es decir, las neuronas guardan la energía inicial y no pueden recibir otra señal hasta después de que la hayan pasado, lo cual puede durar hasta 200 ms.

Con la cuantificación se logra darle valores de amplitud discretos a la señal analógica con infinitos niveles de amplitud, dentro de un conjunto finito de amplitudes posibles. En cada intervalo de tiempo se observa en que rango de voltaje se encuentra la señal y en función de esto se le asigna un nivel de voltaje a la salida tal y como se ilustra en la figura 2. El proceso de cuantificación es sin memoria e instantáneo [2]. En estas condiciones, como se muestra en la figura 3, la amplitud de la señal $x[n]$ se especifica por medio del índice n si se ubica dentro de una celda de partición J :

$$J_k : \{i_k < Q[x[n]] \leq i_{k+1}\} \quad k = 1, 2, \dots, L \quad \text{Ec. 2}$$

Donde L es el número total de niveles de amplitud utilizados en el cuantificador. Las amplitudes discretas i_k , en la entrada del cuantificador, reciben el nombre de niveles de decisión o umbrales de decisión. Las amplitudes discretas $Q[x[n]]$ reciben el nombre de niveles de representación o de reconstrucción y el espaciamiento entre dos niveles, es decir el tamaño de J_k se denomina cuanto o tamaño de paso.

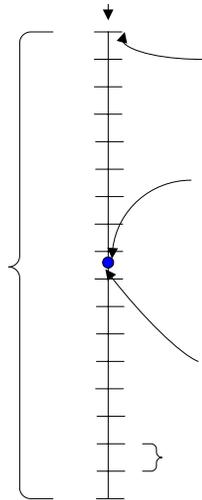


Figura 2. Esquema de la celda de cuantificación. Fuente: propia

2.1 Ruido de cuantificación

El uso de la cuantificación introduce un error que se define como la diferencia entre la señal de entrada $x[n]$ la señal de salida $Q[x[n]]$. El error recibe el nombre de ruido de cuantificación. Esto permite caracterizar la diferencia entre el valor instantáneo de la señal y la cuantificación por medio de una perturbación. El uso de un cuantificador siempre supone una señal de error. Por eso la señal cuantificada es distinta a la señal original.

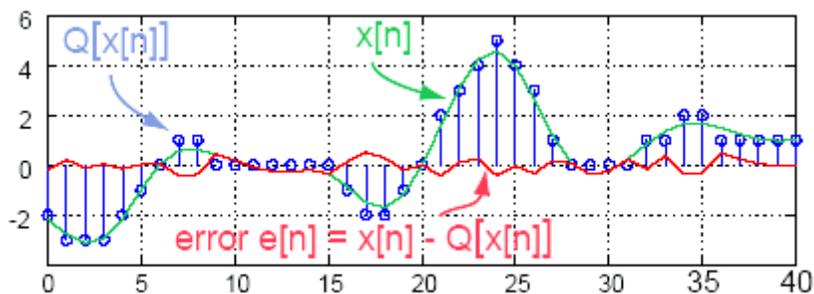


Figura 3. Representación de la onda analógica en niveles discretos. Fuente: Dan Ellis, Audio Compression & Coding

Siendo el error de cuantificación la variable aleatoria $e[n]$ del valor de muestreo n . Teniendo la entrada $x[n]$ media igual a cero (si no lo tiene, siempre se podrá eliminar restando la media de la entrada y sumándola de nuevo después de la cuantificación) [2], y suponiendo que el cuantificador es simétrico (como el peor caso), se concluye que la salida del cuantificador $Q[x[n]]$ y, en consecuencia, el error de cuantificación $e[n]$ también tendrán media cero. De tal modo, en una caracterización estadística parcial del cuantificador en términos de la relación señal a ruido de la salida (cuantificación) sólo se necesita determinar el valor cuadrático medio del error de cuantificación $e[n]$.

Este será el próximo paso a seguir para determinar finalmente el rango dinámico del oído, su sensibilidad a la relación señal a ruido y el número de bits

necesarios para lograr dicha medida.

Se considera entonces que la señal de entrada $x[n]$ de amplitud continua (de infinitos valores) en el intervalo $(-D_{\text{máx}}, D_{\text{máx}})$. La señal analógica tiene distribución uniforme en los niveles de tensión que podrían caracterizarla, por lo que el error asociado tendrá distribución uniforme producto de la señal de tensión real, en el rango J .

Si el cuantificador es simétrico, el tamaño del escalón o paso del cuantificador está dado por:

$$\Delta = \frac{2D_{\text{máx}}}{L} \quad \text{Ec. 3}$$

Para el cuantificador uniforme, el error de cuantificación $e[n]$ tendrá sus valores de las muestras delimitadas por $-\frac{\Delta}{2} \leq e[n] \leq \frac{\Delta}{2}$ (Ec. 4). Esto se traduce en que la función de probabilidad del error $e[n]$ viene dada por la figura 4.

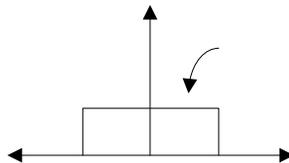


Figura 4. Distribución del error de cuantificación. Fuente: propia

Por lo que el valor esperado sea una de una potencia promedio cero. Como el tamaño de Δ es suficientemente pequeño, lo que implica J bastante grande, resulta razonable suponer que el error de cuantificación $e[n]$ es una variable aleatoria uniformemente distribuida. Y que el efecto de interferencia del ruido de

cuantificación en la entrada del cuantificador es similar al del ruido térmico². De tal forma queda, la función de densidad de probabilidad del error de cuantificación $e[n]$: [2]

$$f_{e[n]}(e) = \begin{cases} \frac{1}{\Delta}, & -\frac{\Delta}{2} \leq e[n] \leq \frac{\Delta}{2} \\ 0, & \text{en otro caso} \end{cases} \quad \text{Ec. 5}$$

Así la media del error de cuantificación es igual a cero, tal como se esperaba. Su varianza σ_e^2 es la misma que la del valor cuadrático medio:

$$\sigma_e^2 = E[e^2[n]] = \int_{-\Delta/2}^{\Delta/2} e^2[n] \cdot f_{e[n]}(e) de = \frac{\Delta^2}{12} \quad \text{Ec. 6}$$

lo cual depende proporcionalmente del valor del paso de cuantificación Δ . La relación señal a ruido viene dada por:

$$SNR = \frac{E([x^2[n]])}{E[e^2[n]]} = 12 \frac{E(x^2)}{\Delta^2} \quad \text{Ec. 7}$$

Por lo que si L crece Δ disminuye y la relación señal a ruido crece.

Para una señal uniformemente distribuida entre $(-D_{\text{máx}}, D_{\text{máx}})$ se tiene que

$$f_{x[n]}(x) = \frac{1}{2D_{\text{máx}}} \quad \text{Ec. 8}$$

Por otra parte para un cuantificador uniforme de M niveles se cumpliría que:

$$2 \cdot D_{\text{máx}} = L \cdot \Delta \quad \text{Ec. 9}$$

² El *ruido térmico* lo provoca el movimiento aleatorio de los electrones en conductores y dispositivos.

entonces,

$$E([x^2[n]]) = \int_{-D_{m\acute{a}x}}^{D_{m\acute{a}x}} [x^2[n]] \cdot f_{x[n]}(x) dx = 2 \int_{D_{m\acute{a}x}}^{D_{m\acute{a}x}} \frac{1}{2D_{m\acute{a}x}} x^2 dx =$$

$$\frac{1}{D_{m\acute{a}x}} \cdot \frac{D_{m\acute{a}x}^3}{3} = \frac{D_{m\acute{a}x}^2}{3} = \frac{L^2 \cdot \Delta^2}{12}$$
Ec. 10

Por lo que la relación señal a ruido queda, de (Ec.7 y Ec.10):

$$SNR = L^2$$
Ec. 11

Para fijar con valores digitales (p.e.: valores de voltaje) el valor L se utiliza la resolución binaria, representada con R , el cual muestra cuántos bits de resolución o detalle poseerá el valor de L .

$$L = 2^R$$
Ec. 12

Lo que significaría que una palabra de R bits por muestra y una cuantificación de paso Δ :

$$SNR = 2^{2R}$$
Ec. 13

O lo que es igual: $10\log(SNR) = 10\log(2^{2R}) = 20R\log 2$

$$\Rightarrow SNR(\text{dB}) \approx 6 \cdot R$$
Ec. 14

Esto significa que para cada bit adicional que se utilice de resolución de muestreo se producirá 6dB de ganancia en relación señal a ruido.

Entonces para lograr la gama dinámica del oído de 100 dB se necesitan de 16 bits de resolución digital. (16 bits → 96 dB). Esto significa que la gama del oído puede definirse con $L=65.536$ intervalos de cuantificación.

3. Codificación

Al finalizar el proceso de cuantificación, la señal se limita a un conjunto discreto de valores, un torrente de bits, que no es la forma más apropiada para el almacenamiento, procesamiento y/o transmisión. Al final del proceso se obtiene una palabra representando a un nivel de tensión.

Para lograr que la señal sea más robusta con respecto al ruido, la interferencia y otros efectos perturbadores, se requiere de un proceso de codificación para trasladar el conjunto discreto de valores de muestreo a una forma más apropiada de señal. A cada nivel de señal cuantificado se le asigna un código, basado en una cantidad n-aria de valores (binaria, ternaria, cuaternaria, etc.) [2]

Existen variantes de la modulación por códigos de pulso, tales como: DPCM, ADPCM, los cuales no serán tratados en este proyecto por su poca pertinencia.

En términos de un compromiso de ancho de banda incrementado por un desempeño mejorado ante el ruido, entra en juego la modulación digital por pulsos como una manera efectiva de lograr esta relación, en su forma básica de modulación por codificación de pulsos (PCM), en la que, como ya se ha visto, la señal analógica se representa en forma discreta tanto en tiempo como en amplitud. Con esta secuencia se logra que la señal se convierta en una secuencia de pulsos binarios codificados. Por lo tanto el efecto del ruido del canal en la salida del receptor puede reducirse hasta un nivel ignorable haciendo simplemente que la potencia promedio de la onda binaria transmitida PCM sea lo suficientemente grande en comparación con la correspondiente al ruido. [8]

3.1 Estándares de codificación

La codificación es útil si el receptor conoce el código. Los esfuerzos por la estandarización son importantes en este sentido. [4] La creciente demanda de redes de

comunicación de audio, la emergencia de novedosas aplicaciones multimedia y el desarrollo de procesadores cada vez más potentes y rápidos han definido las necesidades de estandarización en el mundo. Por ejemplo, la ITU tiene una enorme variedad de estándares de compresión y codificación, de dominio público, tal como se muestra en la siguiente tabla, lo que refleja las diferentes necesidades en lo que se refiere a una eficiente transmisión de la voz en los múltiples sistemas de comunicaciones.

Además de la ITU, está la European Telecom Standards Institute (ETSI), la International Standard Organization (ISO), Telecommunication Industry Association (TIA) en y la R&D Center for Radio Systems (RCR) en Japón.

Por ejemplo, ITU-T G.711 es un simple PCM de 8 bits con 64 kbps de codificación en calidad telefónica usando la ley de codificación A o μ . Esta recomendación representa el método más común de codificación de forma de onda utilizado alrededor del mundo. Mediante el uso de la recomendación G.711 se logra transmitir una señal de voz en forma digital y reconstruirla en el destino de una forma que, para el oído humano, es esencialmente idéntica a la original.

En PCM, cada una de las muestras enviadas al codificador es totalmente independiente una de las otras, por lo que PCM permite codificar forma de ondas totalmente arbitrarias cuya frecuencia máxima no exceda la mitad de la velocidad de muestreo. ITU-T G.722 utiliza codificación DPCM de calidad de audio monoaural AM para comprimir el audio en una tasa de 64 kbps. ITU-T G.728 emplea una tecnología de compresión mejor, la Codificación Lineal Predictiva (LPC) y cuantificación vectorial para reducir la tasa de bits a 16 kbps. La G.723 es capaz de obtener una razonable calidad de voz a tasas tan bajas como 5,3 kbps. La recomendación G.729, que puede codificar señales vocales con calidad de telefonía (o de telefonía interurbana) a 8 kbps.

Tabla 1. Estándares de codificación según la ITU-T. Fuente: ITU-T

Nombre ITU-T	Algoritmo de digitalización	Ancho de banda (Hz)	Tamaño de la palabra	Velocidad de muestreo (kHz)	Rata de bits (kbps)
G.711	PCM	200 a 3.200	8	8	64
G.721	ADPCM	200 a 3.200	4	8	32
G.722	DPCM	50 a 7.000	3,4 y5	16	48,56 y 64
G.723	MP-MLQ/ACELP	200 a 3.200	159 y 189	0,03333	5,3 y 6,3
G.726	ADPCM	200 a 3.200	2, 3 4 y 5	8	16,24,32 y 40
G.727	ADPCM	200 a 3.200	2 a 8	8	16 a 64
G.728	LD-CELP	200 a 3.200	10	1,6	16
G.729	CS-ACELP	200 a 3.200	80	0,1	8

En los estándares MPEG/audio (Estándar ISO) MPEG (Moving Pictures Expert Group) el audio ofrece altos índices de compresión, manteniendo la calidad del audio del arreglo (stream) original. Como se explicará con detalle más adelante, estos son algoritmos de compresión con pérdidas. MPEG se basa en la codificación perceptual apoyándose en los modelos psicoacústicos. Este tipo de compresión se basa en la característica de oído humano denominada enmascaramiento. Cuando una frecuencia posee una gran amplitud, el oído humano no puede escuchar las frecuencias de menor amplitud que se encuentran circundantes a ésta. Se dice que la de frecuencia de mayor amplitud enmascara a las de menor amplitud.

Dentro del enmascaramiento, todas las frecuencias que se encuentren de la gama de tiempo que dura el fenómeno de por debajo de una determinada amplitud, denominada umbral de enmascaramiento, no podrán ser captadas por el ser humano. Este umbral de enmascaramiento depende de la frecuencia y de la intensidad de la señal que lo provoca. La idea básica de la codificación por sub-bandas (SBC, Sub-Band Coding) es la de economizar ancho de banda eliminando la información de las frecuencias que van a resultar enmascaradas. No se utiliza ancho de banda para enviar sonidos que de todas maneras no podrán ser captados por el oído humano. El SBC divide la señal de audio en varias sub-bandas de frecuencia, aplicando a cada una de ellas una codificación similar al ADPCM. A cada sub-banda se le asigna de manera adaptativa un número de bits, el cual depende de las características de enmascaramiento que posean las muestras de las sub-bandas. La idea es manipular el

ruido de cuantificación de forma tal de que sea lo mayor posible (menor número de bits) pero que se encuentre por debajo del umbral de enmascaramiento (de forma tal de que no pueda ser apreciado).[8]

Para aquellas sub-bandas más afectadas por el enmascaramiento se le asignan menos bits (ruido de cuantificación se hace menos evidente) y para aquellas sub-bandas menos afectadas por el enmascaramiento se les asigna más bits. Es posible que para aquellas sub-bandas más afectadas no se les asigne ningún bit, de forma tal que las muestras pertenecientes a las mismas no sean codificadas. De esta manera, manipulando la relación entre el ruido de cuantificación y los umbrales de enmascaramiento y eliminando la información de las frecuencias que se ven enmascaradas, se logran compresiones importantes.

El SBC se hace más útil cuando se utiliza para codificar señales de audio de alta calidad (22 kHz), donde el fenómeno de enmascaramiento es más frecuente (y donde por tanto se pueden lograr mayores compresiones), que para codificar simplemente señales de voz (4 kHz). Esta técnica se utiliza para la codificación de audio de MPEG (estándar para la codificación de audio y vídeo).

Al utilizar PCM, el número de bits es constante para todas las muestras y, para tratar de mantener el ruido de cuantificación bajo, para digitalizar un señal de audio de 22 kHz, se le asigna usualmente 16 bits a cada muestra. Con esto se tiene que la señal de 22 kHz es muestreada 44.000 veces por segundo (Nyquist) y con 16 bits por muestra, da una velocidad de transmisión de 704 Kbps (que es el utilizado para un canal de audio para CD). Muchos de estos bits son desperdiciados ya que la información que ellos poseen no podrá ser percibida por el oído humano debido al enmascaramiento. Al utilizar SBC una asignación de bits variable, que depende de la relación entre el ruido de cuantificación y los umbrales de enmascaramiento, se puede lograr disminuir los 704 Kbps, para un canal de audio de CD, a 64 Kbps.[8]

Estos conceptos serán la base de los sistemas de compresión a estudiar en

este proyecto.

4. Aspectos finales. Ejemplos. Conclusiones

Como ejemplo ilustrativo se tiene al disco compacto de audio digital, que según el Libro Rojo³ cada pista de audio está almacenada utilizando una codificación PCM de 16 bits y muestreada a una tasa de 44.100 Hz, lo que significa que el arreglo de bits resultantes para una señal estereofónica es de $44.100 \times 16 \times 2 = 1.411.200 \text{ bps}$. Esta salida representa un ancho de banda de aproximadamente 1,5 MHz, más de 60 veces del que posee la señal analógica original.

Un canal de telefonía analógica, según la CCITT maneja un rango de frecuencia entre 300 y 3.400 Hz, lo que resulta en un ancho de banda de 3100 Hz. La máxima frecuencia a transmitir es de 3.400 Hz por lo que para evitar el traslape se calcula en base de 4.000 Hz. Con esta cifra, según el teorema de Nyquist, se deberían tomar 8.000 muestras por segundo. El mínimo número de niveles de cuantificación es determinado por la inteligibilidad del habla en el receptor y basado en las experiencias de las investigaciones de la CCITT se han escogido un mínimo de 256 niveles, lo que significaría una resolución de codificación a 8 bits, por lo que resultaría un torrente de 64.000 bps. Esta última cifra es la reconocida como el estándar para un canal PCM. Para la transmisión de radio digital se utilizan 32 kHz y 48 kHz en cintas de audio digital (DAT).[3]

En la figura 5, la calidad de la digitalización de un sonido viene dado por dos factores: la frecuencia y la resolución. La frecuencia hace referencia a la cantidad de muestras de la señal analógica que se toman por segundo. La resolución indica la calidad de la cuantificación, es decir, cuan preciso es el valor tomado respecto al valor real de la señal.

³ Libro rojo (Red Book) : Estándar para Discos Compactos de Audio. Contiene las especificaciones técnicas para todos los formatos de DC y DC-ROM

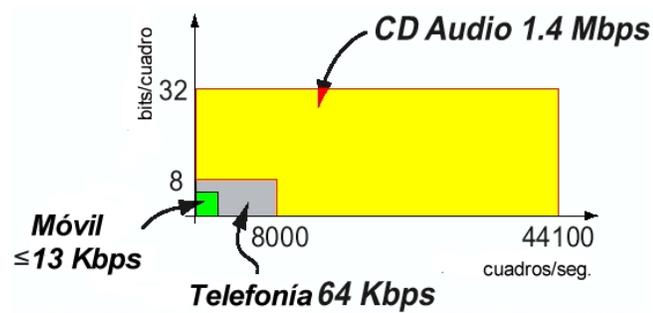


Figura 5. Uso de la frecuencia de muestreo versus la resolución en bits para la digitalización de audio. Fuente: Dan Ellis, Audio Compression & Coding

4.1 Virtudes, limitaciones y modificaciones de la PCM

Por las razones antes expuestas, la modulación por codificación de pulso (PCM) ha surgido como el esquema de modulación más favorable en la transmisión de señales que contiene información analógica como las de video y voz. Las ventajas de la PCM se pueden resumir del modo siguiente:

1. *Robustez* al ruido y la interferencia del canal de transmisión.
2. *Regeneración* eficiente de la señal codificada a lo largo de la trayectoria de transmisión.
3. *Intercambio* eficiente del ancho incrementado del canal por una relación señal a ruido mejorada, obedeciendo a una ley exponencial.
4. Un *formato uniforme* para la transmisión de diferentes tipos de señales de banda base y de allí su integración con otras formas de datos digitales con una red común.
5. *Facilidad* comparativa con la cual las fuentes de mensajes tal vez se eliminen o reincorporen en un sistema de multiplexado por división de tiempo.

6. Comunicación *segura* mediante el uso de esquemas de modulación o encriptado especiales.[2]

Tales ventajas se alcanzan a costa de una mayor complejidad del sistema e incrementar el ancho de banda utilizado para la transmisión. Si bien el uso de la PCM implica muchas operaciones complejas, en la actualidad todas ellas pueden ponerse en práctica con las nuevas tecnologías de procesadores de alta velocidad, incluidos hasta en los más pequeños dispositivos electrónicos personales y por supuesto en la gran capacidad de procesamiento de las computadoras personales.

Con respecto al ancho de banda, se admite que el aumento en el requisito del ancho de banda de la PCM quizá haya tenido una razón justificable en el pasado. Sin embargo, hoy en día no constituye una preocupación real por dos razones diferentes. En primer lugar, la disponibilidad creciente de canales de comunicación de banda ancha significa que el ancho de banda ya no es una restricción del sistema en la manera tradicional que solía serlo. La liberación de la restricción del ancho de banda ha sido posible por el despliegue de los satélites de comunicación para la transmisión y el uso cada vez mayor de fibras ópticas para la conexión de redes.

Actualmente, y por lo que concierne a las redes troncales, los cables de fibra óptica permiten la transmisión de señales digitales de vídeo y audio de una forma más fácil que en el pasado.

No obstante, todavía son numerosas las aplicaciones que requieren ahorro en el ancho de banda utilizado, siendo virtualmente obligatorio el uso de la compresión de señales en las aplicaciones vía satélite o mediante radioenlaces. Por otra parte, el costo de los medios de transmisión y las tarifas por su utilización juegan un papel importante en la elección de la velocidad de transmisión deseada. Ello explica como, incluso en el caso de transmisión por fibra óptica, puede ser útil la economía del ancho de banda.

La segunda razón estriba en que a través del uso *de técnicas complejas de compresión de datos* es en realidad posible eliminar la redundancia presente en forma inherente en una señal PCM y reducir de ese modo la tasa de bits de los datos transmitidos sin una degradación seria en el desempeño del sistema. La tremenda evolución en el desarrollo de la tecnología de señales digitales, en la fabricación de microprocesadores y en la capacidad de las memorias actuales, permite la ejecución en tiempo real de algoritmos extremadamente complejos, garantizando así que la calidad del audio (y vídeo) se vea apenas afectadas por la compresión.

En realidad, la mayor complejidad de procesamiento (y, por tanto, el mayor costo de la puesta en práctica) se intercambia por una tasa de bits reducida y, en consecuencia, un requerimiento de ancho de banda menor. Aquí estriba la importancia fundamental de la compresión de datos digitales.

En éste proyecto se introducirán primero conceptos generales sobre compresión y otros temas necesarios para el entendimiento de las técnicas utilizadas en los sistemas que se ajustan a las normas MPEG. Se tratará exclusivamente lo relacionado a la compresión del audio. Luego se expondrán temas más relacionados a la implementación de diferentes algoritmos necesarios para los códec MPEG .

[ANEXO 4]

SISTEMAS DE COMPRESIÓN

1. Definición

La compresión de las señales es el proceso mediante el que se obtiene la reducción del caudal de datos a transmitir y/o almacenar. (John Watkinson, 1.995)

Este proceso tiene que ser eficaz, esto es: la técnica empleada para la reducción debe permitir que el terminal receptor y/o reproductor presente la información de modo que, luego de expandirla, los observadores y/u oyentes no se encuentren en capacidad de notar diferencia alguna si tuvieran la oportunidad de comparar la señal original con la así procesada. Además, el proceso tiene que ser eficiente, es decir: ha de efectuarse en tiempo real y con un costo que justifique los beneficios pretendidos. El dominio elegido para este proceso es el digital, por las innumerables ventajas antes vistas que posee sobre el analógico.

Compresión, reducción de la tasa de bits y reducción de datos son todos términos que significan básicamente lo mismo en este contexto. En esencia, compresión significa que la misma información se transfiere usando una menor tasa de datos. Se debe indicar que en el lenguaje tradicional del audio, “compresión” significa un proceso en el cual se reduce el rango dinámico de la señal representativa del sonido. En el lenguaje digital la misma palabra significa que la tasa de bits se reduce, idealmente dejando sin cambios la dinámica de la señal. Por lo tanto en este contexto se adoptará el último significado.

La compresión puede ser utilizada de dos maneras diferentes: (a) en transmisión y (b) en almacenamiento.

La compresión utilizada en transmisión se resume en la figura 1. El compresor reduce la tasa de datos de la fuente. Los datos comprimidos atraviesan un canal de comunicación y luego son vueltos a la tasa original mediante el expansor. La relación entre la tasa de datos de la fuente y la tasa de datos del canal se llama factor de compresión. Se usa también el término ganancia de codificación. A veces a un compresor y un expansor en cascada se los denomina compansor. El compresor puede igualmente bien llamarse codificador y el expansor decodificador, en tales casos el par en tándem puede denominarse códec. En las comunicaciones, el costo de los enlaces de datos es, la mayoría de las veces, proporcional a la tasa de datos y por lo tanto es que surge una cierta presión económica para usar un factor alto de compresión. Sin embargo, se debe tener en mente que la implementación de un códec también tiene un costo que sube con el factor de compresión y en consecuencia será necesario llegar a un compromiso.

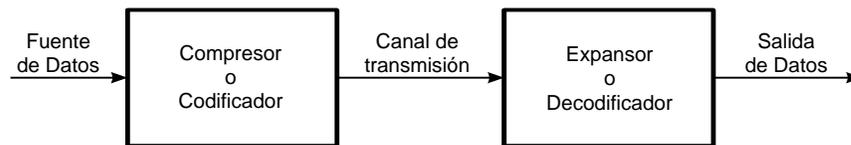


Figura 1. Esquema de compresión en la transmisión. Fuente:propia

Como se muestra en la figura 2, se utiliza un códec para ayudar en el almacenamiento de información. El uso de la compresión en aplicaciones de almacenamiento es sumamente poderoso.

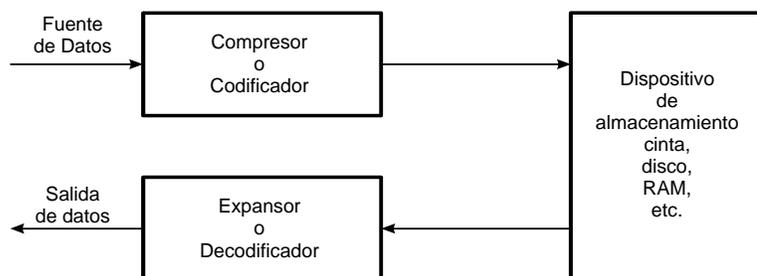


Figura 2. Esquema de Códec para almacenamiento de información. Fuente: propia

2. Clasificación

Las técnicas de compresión pueden dividirse en dos grandes familias, las técnicas sin pérdidas y las técnicas con pérdidas. Aunque hay muchas diferentes técnicas de compresión, todas ellas caen en una u otra de estas categorías. [11]

2.1 Técnicas de compresión sin pérdida

La compresión sin pérdida elimina la información redundante contenida en los datos de interés. Se dice que la compresión es sin pérdidas debido a que es completamente reversible, porque es posible reconstruir en forma exacta los datos originales. La compresión sin pérdidas también se conoce como compactación de datos. [2]

La compresión sin pérdida, sin embargo, permite solo una modesta reducción de tasa de bit. Los algoritmos de compresión/descompresión sin pérdida están indicados en aquellos casos en los que se deba garantizar que el flujo de datos original debe poderse recuperar exactamente tras el proceso de compresión y descompresión. En general estos casos se conocen como casos de datos "discretos" ya que son del tipo registros de bases de datos, hojas de cálculo, procesadores de texto e incluso imágenes donde la calidad es crítica.[11]

La compresión sin pérdida se implementa también como complemento a muchos protocolos de transmisión de información. Las empresas de telecomunicaciones utilizan este tipo de algoritmos para conseguir multiplexar un mayor número de líneas telefónicas o de datos por un mismo medio físico. Especialmente se realiza en el caso de información de audio analógico ya que en lo que se refiere a compresión de datos ésta se suele realizar en los propios equipos situados a los extremos de la línea.

Otro campo de aplicación es el almacenamiento de datos en medios físicos. Obviamente se desea que los datos restaurados sean idénticos a los originales así

como conseguir almacenar mayor información en el mismo espacio. [11]

2.2 Técnicas de compresión con pérdida.

La compresión con pérdidas implica la pérdida de información de una manera controlada: en consecuencia, la compresión quizá no sea por completo reversible. Sin embargo, este tipo de compresión es capaz de alcanzar una razón de compresión superior a la que se obtiene con los métodos sin pérdidas. [2]

La compresión con pérdida es irreversible, ya la señal de audio reconstruida perderá algunos de sus muestras originales. En general, una mayor compresión crea una mayor distorsión pero con técnicas cuidadosamente seleccionadas, esta distorsión puede hacerse bastante imperceptible al oído humano. Los códecs con pérdida no son apropiados para los datos de computadora, pero son habituales en aplicaciones de audio, puesto que ellos permiten un factor mayor de compresión que los códecs sin pérdida. [2]

La motivación de la existencia de los algoritmos con pérdidas es el aumento de las tasas de compresión a costa de pérdida de precisión. Otra motivación puede ser la disminución de los tiempos de compresión para el uso en sistemas de tiempo real. Muchos de estos algoritmos están diseñados teniendo en cuenta posibles implementaciones hardware mediante sistemas basados en DSP's. Hoy en día debido al creciente aumento de las prestaciones de los CPUs de propósito general así como la disminución en su precio hace que sea más factible enfocar los diseños a este tipo de procesadores. En general se debe tener en cuenta que existen archivos, especialmente aquellos que representan imágenes o sonidos, que pueden soportar pérdidas sin que realmente sean apreciadas por los sentidos humanos. Otra de las características de los algoritmos con pérdida es que suelen ser parametrizables pudiendo elegir distintas configuraciones para obtener distintos grados de calidad respecto a la tasa y tiempo de compresión/descompresión. [11]

En general los algoritmos con pérdida se componen de dos fases. En la primera de ellas se realiza un tratamiento digital de la señal consistente en transformar los datos al dominio de la frecuencia mediante transformadas rápidas de Fourier (FFTs) o transformaciones de coseno discretas (DCTs). Una vez obtenidos los datos transformados se eliminan o se recortan aquellos valores muy altos o muy bajos. Es en esta fase cuando se introducen los errores para facilitar la compresión. Acto seguido se comprimen mediante técnicas tradicionales sin pérdidas estos puntos de frecuencia que representan una imagen similar a la original. La función que se encarga de esa eliminación o recortado de datos se la suele denominar función de suavizado. Esta función es la que se parametriza buscando un nivel de suavizado que no comprometa la relación entre calidad y relación de compresión deseada.[11]

Los códecs con pérdida exitosos son aquellos en los cuales las diferencias se arreglan para que un perceptor humano los encuentre subjetivamente difícil de percibir. Así los codificadores de audio con pérdida deben ser basados en una comprensión de la percepción psicoacústica, los cuales se conocen como Codificadores Perceptivos.[11]

En la compresión (o codificación) perceptiva, un mayor factor de compresión requerirá que deban modelarse más precisamente los sentidos humanos. Los codificadores perceptiva tienen la ventaja de que pueden forzarse a operar a un factor fijo de compresión. Esto es conveniente para aplicaciones prácticas de transmisión y almacenamiento donde una tasa fija de datos es más fácil de manejar que una tasa variable. El resultado de emplear un factor fijo de compresión es que la calidad subjetiva puede variar con la “dificultad” del material de entrada al códec. [2]

Los códecs perceptivos no deben conectarse en cascada indiscriminadamente, en especial si usan algoritmos diferentes. Debido a que la señal reconstruida desde un códec perceptivo no es bit a bit precisa, tal códec no puede incluirse en cualquier sistema de prueba de tasa de bits erróneos ya que las diferencias introducidas por la compresión serían indistinguibles de los verdaderos

errores. [11]

Los códigos sin pérdida son menos comunes en audio (y en vídeo) donde los códigos perceptivos pueden utilizarse. Los códigos perceptivos frecuentemente obtienen una ganancia de codificación acortando las longitudes de las palabras de datos que representan la forma de onda de la señal. Esto aumenta el nivel de ruido. El truco está en asegurarse que el ruido resultante se pone a frecuencias donde los sentidos humanos son menos capaces de percibirlo. Como resultado, aunque la señal recibida sea apreciablemente diferente de los datos de la fuente, puede parecerle lo mismo al oyente. Lográndose así factores moderados de compresión. Como estos códigos confían en las características de la audición, ellos pueden solamente probarse en forma totalmente subjetiva. [11]

La idea de este proyecto es encontrar un método menos subjetivo, que dependa de las características intrínsecas de los algoritmos de construcción de la codificación perceptiva utilizada, aprovechando además los conocimientos sobre las cualidades psicoacústicas y su factible implementación en el lenguaje de programación MATLAB.

El factor de compresión de tales códigos puede establecerse a voluntad por elección de la longitud de las palabras de los datos comprimidos. Una compresión leve sería indetectable, pero con un mayor factor de compresión, las dificultades de codificación llegarían a ser perceptibles. [11]

Existen entonces dos principios: remover la información redundante o remover la información irrelevante.

2.2.1 Redundancia

La información redundante es implícita al concepto de residuo. Una señal o dato es redundante cuando su presencia, aunque perceptible, no provee un aporte a la información ya conocida. Por ejemplo, una señal con un ancho de banda limitado a

20 kHz. pero muestreada a 80 kHz. La información o muestra puede ser recuperada por medio de otra muestra por medio de interpolación. [3]

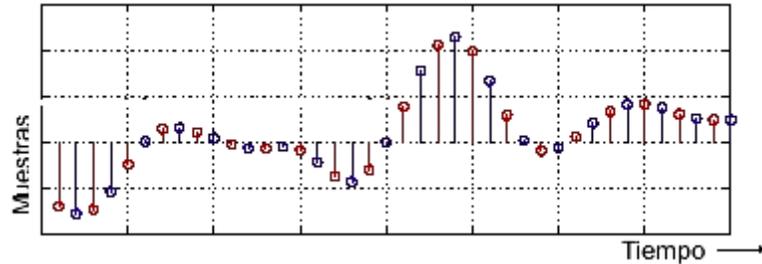


Figura 3. En una señal limitada por ancho de banda, las muestras en rojo pueden ser exactamente recuperadas interpolando las muestras azules. Fuente: Dan Ellis, Audio Compression & Coding

La mayoría de los sistemas de audio digital usan técnicas de modulación por codificación de pulsos (PCM) para digitalizar las señales de audio analógicas. Los sistemas de alta calidad, tales como los formatos CD's con PCM lineal de 16 bits, resultan en tasas de datos que exceden en mucho la tasa de información de la señal original. Por esta razón, los datos digitales PCM son altamente redundantes y generalmente con grandes anchos de banda.[7]

La redundancia también es denominada Redundancia Objetiva la cual es medible y cuantifica ciertas características numéricamente predecibles de las señales de audio, tales como la periodicidad de las formas de onda. [7]

2.2.2 Irrelevancia

La información irrelevante es única pero innecesaria. Por ejemplo, grabar en un micrófono que acepte un ancho de banda de 20 kHz una señal de 80 kHz. Irrelevante significa perceptivamente insignificante.[3]

La irrelevancia también se conoce como Redundancia Subjetiva, la que resulta de los fenómenos psicoacústicos de la audición humana. La eliminación de

redundancia es también citada como compresión estadística, mientras que la remoción de la irrelevancia es frecuentemente llamada codificación perceptiva. [7]

2.2.3 Objetivos de la compresión con pérdidas.

Teniendo en cuenta estas dos características, los distintos métodos de compresión se proponen eliminar del caudal de datos aquéllos que son irrelevantes y/o redundantes. Un buen algoritmo de compresión tiene dos metas. La primera es proveer un método más eficiente de representación del sonido (y/o de la imagen). La segunda meta es sacrificar aquellas características de la secuencia PCM para el caso del audio (y/o cuadro de vídeo) a las que los sentidos humanos son menos sensibles y sacrificar también aquella información de sucesos que son muy poco probables que aparezcan, es decir, en la codificación, solo se tendrán en cuenta los sucesos habituales. [7]

La figura 4 ilustra este principio en forma gráfica. Las curvas superiores en cada una de las muestras de música exhibe la respuesta de amplitud en el tiempo, mientras que las inferiores son el trazado de la amplitud vs. frecuencia de un breve muestreo de 32 ms. Nótese que la forma de onda (b) superior es altamente periódica y repetitiva y que el espectro es ponderado predominantemente hacia las bajas frecuencias debajo de los 4 kHz. Usando estas dos fundamentales propiedades, pueden idearse esquemas que examinen la forma de onda del audio o en el dominio del tiempo (la familiar respuesta amplitud en el tiempo) o dentro del dominio de la frecuencia, mediante por ejemplo la ejecución de la transformada de Fourier sobre la forma de onda más relevante.[7]

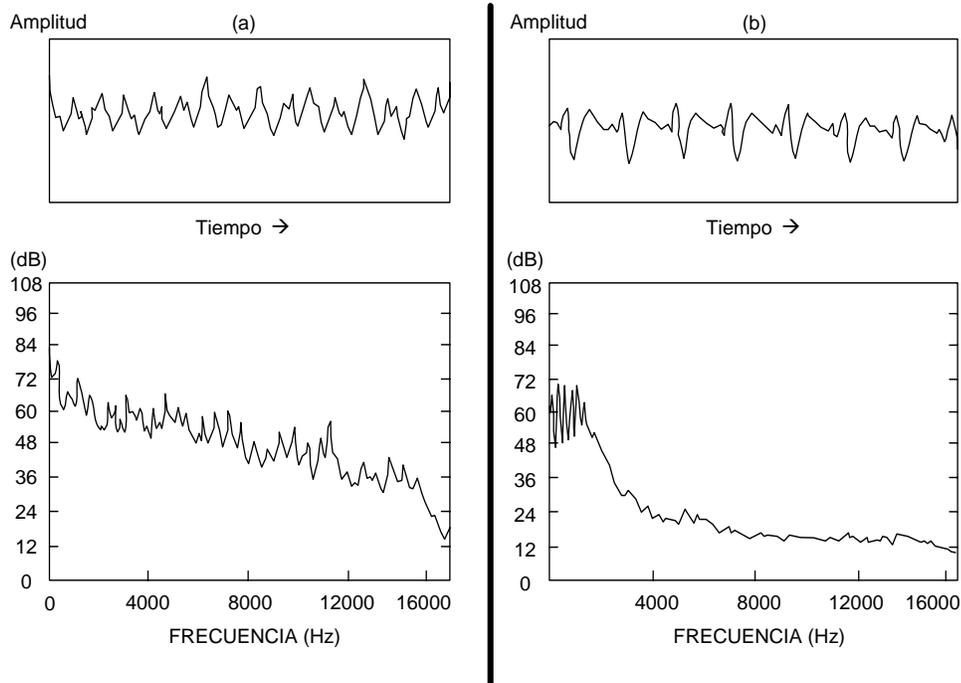


Figura 5. (a) Guitarra clásica y (b) trombón. Fuente: Sergio Mangiaterra, Audio-mpeg

3. Razones de las técnicas de compresión

Al entrar en este proyecto, el audio comenzó como un medio analógico. Las señales de audio han sido capturadas, almacenadas, procesadas y transmitido usando técnicas analógicas en las que una señal “forma” en sí misma el sonido.

Ahora ya se ha visto como el audio reúne otro tipo de información electrónica y las tecnologías utilizadas en comunicaciones son totalmente digitales. En vez de emplear formas de onda analógicas, las señales de audio se han digitalizado, lo cual significa que la información auditiva se muestrea en muchos puntos en el espacio y en el tiempo, y estos valores de muestreo se convierten en códigos digitales binarios (0's y 1's). Estos valores digitales, no son señales analógicas en sí mismas, son las que luego se almacenarán y/o transmitirán por cable o aire utilizando adecuados métodos de modulación. [7]

La transmisión, almacenamiento, y procesamiento de la señal de audio

digital requiere la manipulación de cuantiosos flujos de datos (mayor aún la compuesta de vídeo y audio). Ya se ha revisado como a esta velocidad de estos flujos de datos se le llama tasa de bits, la cual mide la cantidad de datos en bits por segundo. Cuánto más alta es la calidad de la señal de audio representada, más alta es la tasa de bits, lo cual implica un mayor desafío para los dispositivos de almacenamiento y para los canales de transmisión. [7]

Aunque la grabación de audio digital y la transmisión ofrecen muchas ventajas sonoras y operacionales sobre su contraparte analógica (inmunidad al ruido, mejor ancho de banda de audio y múltiples reproducciones sin degradación de la señal) esta claro que las señales de audio digital ocupan un extremadamente amplio ancho de banda de transmisión. Por lo tanto los sistemas que pueden reducir los requerimientos de espectro sin empobrecer la calidad son verdaderamente estimados. Ellos permiten que la transmisión digital sea práctica y económica. Con la amplia escala de aceptación de los discos compactos (DC), los beneficios del audio de alta fidelidad digital han atraído dramáticamente la atención de los usuarios y profesionales. La claridad y robustez del DC (con un ancho de banda de 1,5 Mhz) no han sido equiparadas por ninguno de los otros reproductores domésticos de la anterior tecnología. Para los medios de almacenamiento, esto no ha sido un factor limitante en el desarrollo de estaciones de trabajo (workstations) y/o de sistemas de reproducción al aire, pero ello puede ser un problema en aplicaciones de transmisión. [7]

Respondiendo a esta creciente demanda por mejorar la calidad del sonido, muchas compañías de audio y autoridades del *broadcasting* han determinado su intención por desarrollar los medios de audio digital, prometiendo un amplio rango de futuros servicios digitales. Para disfrutar del audio digital de uso extendido a lo largo de la industria del *broadcasting*, el ancho de banda de la señal digitalizada debe reducirse. En la actualidad se disponen de una variedad de sistemas innovadores que reducen el ancho de banda del audio digital con calidad de DC. Por numerosas razones, la transmisión de audio de alta calidad será la de mayor demanda y de aplicación persistente de tales técnicas de reducción de datos. [7]

En este proyecto se revisarán algunas técnicas que reducen la tasa de datos de las señales de audio PCM de 16 bits estándar tanto que ellas pueden transportarse sobre líneas de 64 Kbps, 96 Kbps, 128 Kbps o 256 Kbps (dependiendo de la frecuencia de muestreo y la relación de reducción de la tasa de bits), manteniendo el audio con calidad de CD como referencia. Esto permite su transmisión sobre ISDN o fracciones de canal T1. Las relaciones de compresión comunes varían entre 4:1 a tan altas como 12:1.

De lo expuesto se puede apreciar claramente que es necesario encontrar una manera de acomodar los cuantiosos flujos de datos a los medios de transmisión y almacenamiento disponibles. La solución a este problema no es otra más que la compresión de los datos representativos del audio, manteniendo lo mejor posible, de acuerdo a los requerimientos del servicio en particular, la calidad del sonido resultante.

Se puede afirmar entonces que existen dos razones fundamentales por lo cual se usan las técnicas de compresión:

1. Para posibilitar algunos procesos que serían impracticables sin ella.
2. Para realizar un proceso conocido en una forma más económica.

En conclusión el objetivo buscado es obtener una reducción del ancho de banda ocupado por las transmisiones de audio así como también disminuir la capacidad necesaria para el almacenamiento de dicha información, ya sea por cuestiones de practicidad o por cuestiones económicas. El resultado deberá ser una señal digital de audio comprimida, eliminando la irrelevancia de la señal por medio de la codificación perceptiva inherente a las características auditivas humanas. Para ello se hace necesario el Modelo Psicoacústico.

[ANEXO 5]

ESTÁNDAR DE CODIFICACIÓN DE AUDIO MPEG

A fin de normalizar todo lo referente a la codificación basada en la compresión, la ISO creó un subcomité en conjunto con la IEC denominado “Grupo de Expertos en Imágenes en Movimiento” MPEG (*Moving Pictures Experts Group*). Este grupo de trabajo ha desarrollado dos normas, una para el vídeo y audio digital de computadoras, identificada por la sigla MPEG-1 (publicada en noviembre de 1992) y otra para el vídeo y audio digitales empleados en la radiodifusión y en la distribución hilo-guiada de tales señales, bautizada con la denominación MPEG-2. En marzo de 1993 el MPEG publicó una propuesta de la norma MPEG-2, en la que se definió que el objetivo de calidad buscado era el nivel establecido en la Recomendación 601 de la ITU-R. En julio de 1993 se publicaron los requisitos establecidos por la Norma MPEG-2. La ISO trabajó en una nueva iniciativa para la codificación a tasa muy baja de bits de información de audio y de vídeo (MPEG-4) que se presentó en 1.999.

El MPEG no especifica una técnica única de compresión, sino un conjunto de las herramientas de compresión que pueden usarse en conformidad con reglas exactas de una sintaxis de compresión. Estas reglas exactas facilitan el intercambio de flujo de bits entre aplicaciones diferentes.

Tanto las señales de voz como audio son similares a causa de que para ambas la calidad del esquema de codificación se sustenta en las propiedades de la percepción auditiva humana. En el caso de las señales del habla, se tienen esquemas de codificación eficientes como el ADPCM debido a que se dispone de un modelo de producción del habla. Desafortunadamente no existe nada similar para las señales de audio. [2]

Es en este punto en el que resalta la importancia de el estándar de

codificación de audio MPEG, al igual que el ADPCM, el estándar de codificación de audio MPEG es un sistema de compresión con pérdida, aunque difiere de aquél en un importante aspecto práctico: el estándar MPEG es capaz de conseguir una compresión transparente y perceptualmente sin pérdidas de señales de audio estereofónicas a elevadas frecuencias de muestreo. En particular, las pruebas de audición subjetivas efectuadas por el comité de audio de la MPEG en condiciones de audición muy difíciles, han demostrado que incluso con una razón de compresión de 6 a 1, las señales de audio codificada y original son perceptualmente indistinguibles. [2]

Por ejemplo, el resultado final de utilizar el estándar MPEG en los dos canales de audio de un programa estereofónico, como el caso ya estudiado del CD de música, es que cada señal de audio digitalizada, que se obtiene a razón de 768 kbps por canal se logra comprimir a una velocidad tan baja como 16 kbps (48:1). De este modo el estándar de codificación de audio MPEG es adecuado para el almacenamiento de señales de audio en medios económicos o su transmisión sobre canales con ancho de banda limitado, mientras mantiene al mismo tiempo la calidad de percepción.

1. Generalidades del estándar ISO/MPEG.

El principal motivo del MPEG fue la codificación de video junto con la codificación de audio para el Almacenamiento en Medios Digitales (*Digital Storage Media*, DSM). El estándar MPEG Audio contempla tres niveles diferentes de codificación - decodificación de la señal de audio, de los cuales sólo los dos primeros están totalmente terminados. El otro es aplicable, y de hecho se utiliza habitualmente, pero sigue abiertos a ampliaciones. Estos tres niveles son: MPEG-1 MPEG-2 y MPEG-3. [7] [9]

1.1 MPEG-1 Norma ISO/IEC 11172 – 3 Audio: Codificación de imágenes en movimiento y audio asociado para medios de almacenamiento digital hasta 1,5 Mbps. Este es el nombre para la primera fase del trabajo de MPEG que comenzó en 1.988 y finalizó con la adopción del ISO/IEC IS 11172-3 en 1.992. Su frecuencia de muestreo

es de 32 Khz, 44,1 Khz y 48 Khz.[9]

1.2 MPEG-2: Codificación genérica de imágenes en movimiento e información de audio asociada. MPEG-2 denota la segunda fase del MPEG. Introdujo un gran número de conceptos dentro de la codificación de video MPEG para señales interlazadas de video. El área de aplicación principal fue la televisión digital.[7][9]

El estándar de audio MPEG-2 original, finalizado en 1.994, sólo contiene dos extensiones del MPEG-1:

- Compatible con tecnologías anteriores y permite mejoras hacia posibles futuras actualizaciones de codificación multicanal, incluyendo señales de 5.1 canales, conocido como sonido de cine. A este se le llama MPEG-2 BC (Backward Compatible) Norma ISO/IEC 13818-3. [7][9]
- Codificación a tasas más bajas de muestreo de frecuencias adicionales a los de la MPEG-1, tales como 16 Khz, 22,05 Khz y 24 Khz. Las nuevas bajas tasas de muestreo se logran con gran eficiencia. Llamado MPEG-2 BC LSF (Backward Compatible Lowsampling frequency) Norma ISO/IEC13818-3. [9]

Ambas extensiones se les reconoce generalmente como MPEG-2 EXT MPEG-1 (Extensión del MPEG1) y no introducen nuevos algoritmos de codificación sobre el audio MPEG-1.

Más tarde, en pruebas de verificación del MPEG-2 se comprobó que introduciendo un nuevo algoritmo de codificación y restándole compatibilidad “hacia atrás” para el MPEG-1, se lograba una significativa mejoría en la eficiencia de codificación, para el caso de los 5 canales. Como resultado de este trabajo se produjo una nueva definición, MPEG-2 NBC AAC (Non Backward Compatible Advanced Audio Coding) Norma ISO/IEC 13818-7. AAC se conoce como la segunda generación del esquema de codificación de audio para codificación general de señales de audio estéreo y multicanal. [7][9]

1.3 MPEG 2.5: Es un estándar privado, no abierto, del Instituto Franhoufer de

Circuitos Integrados el cual posee una tasa de muestreo de 8 Khz, 11,05 Khz y 12 Khz.[9]

1.4 MPEG-3: la planificación original contemplaba su aplicación a sistemas HDTV. Sin embargo, pronto se decidió que las herramientas desarrolladas para la codificación de video MPEG-2 contenían todo lo necesario para la HDTV, así que el desarrollo para el MPEG-3 fue finalmente incluido dentro de MPEG-2. Algunas veces a los MPEG-1/2 capa III (MP3) erróneamente se les denominan MPEG-3. [9]

1.5 MPEG-4: Codificación de objetos audiovisuales. Terminado y corregido a finales de 1.999, intenta ser el próximo mejor estándar en el mundo multimedia. Distinto a MPEG-1 y MPEG-2, el énfasis en MPEG-4 se basa en nuevas funcionalidades más que en una mejor eficiencia de compresión. Tanto terminales fijos como móviles, acceso a base de datos, comunicaciones y nuevos tipos de servicios interactivos serán las principales aplicaciones para MPEG-4. El nuevo estándar facilitará el crecimiento en la interacción y sobrepasará el hasta ahora separado mundo de la computación, la electrónica mediática de masas (TV y radio) y las telecomunicaciones. [9]

El audio MPEG-4 consiste en una familia de algoritmos codificadores de audio que abarcan el rango desde la codificación del habla de bajos bits (por debajo de 2 Kbps) hasta los codificadores de audio de alta calidad a 64 Kbps por canal e incluso superiores. La codificación general a media y alta tasa de bits es realizada por el AAC. Es conocido por la Norma ISO/IEC 14496-3. [7][9]

1.6 MPEG-7: Distinto a MPEG-1/2/4, este no define algoritmos de compresión. Aprobado en el 2.001 es un estándar para la representación de contenidos para información multimedia y lo inherente a su búsqueda, filtraje, administración y procesamiento. Se enfoca en la estandarización de una interface común para describir materiales multimedia, representando información acerca del contenido, no del contenido mismo (“bits acerca de los bits”). En este contexto, MPEG-7 dirige aspectos tales como facilitar interoperabilidad y globalización de las fuentes de datos y administración y flexibilidad de los datos. Es conocido por la Norma ISO/IEC

15938. [9]

En la tabla 1 de los anexos se presenta un resumen de los estándares.

A su vez, MPEG describe tres capas (capas) de codificación de audio denominados capa I, capa II y capa III. Del primero al tercero aumentan tanto la complejidad del codificador como la calidad del sonido a igual tasa de bits. MPEG define, para cada capa, el formato del arreglo de bits (*bitstream*) y el decodificador. Con vistas a admitir futuras mejoras no se define el codificador, pero señala las pautas para su construcción. En este proyecto se tomarán en cuenta dichas recomendaciones para la construcción efectiva del modelo psicoacústico a estudiar. Tanto MPEG-1 como MPEG-2 emplean esas tres capas, pero este último añade nuevas características.[7]

Las muestras de audio ingresan al codificador. El mapeo crea una representación filtrada y sub-muestreada de la tira de audio de entrada. Las muestras mapeadas pueden llamarse tanto muestras de sub-banda (como en capa I y II) o muestras de sub-bandas transformadas (como en capa III).

El modelo psicoacústico crea un conjunto de datos para controlar el cuantificador y la codificación. Estos datos son diferentes dependiendo de la implementación actual del codificador. Una posibilidad es el uso de una estimación del umbral de enmascaramiento (UE) para efectuar el control de la cuantificación. El cuantificador y el bloque de codificación crean un conjunto de símbolos codificados con las muestras de entrada mapeadas. Nuevamente, este bloque puede depender del sistema de codificación. El bloque de empaquetado de la trama ensambla la trama de bits actual desde la salida de datos de los otros bloques y le agrega información adicional, por ejemplo corrección de error, si es necesario.[7]

2 Capas (*Layers*)

Dependiendo de la aplicación, capas diferentes del sistema de codificación con creciente complejidad de codificación y mejor rendimiento pueden ser usadas. En

MPEG, el arreglo de bits PCM se subdivide en cuadros (*frames*). Cada cuadro corresponde a la señal de audio en cierto período de tiempo. La capa I contiene 384 muestras, las capas II y III, 1152 muestras. Las muestras representan valores extraídos de las 32 sub-bandas de frecuencia que se asemejan a las bandas críticas del oído. La siguiente gráfica muestra como se dividen las sub-bandas de las muestras de audio.

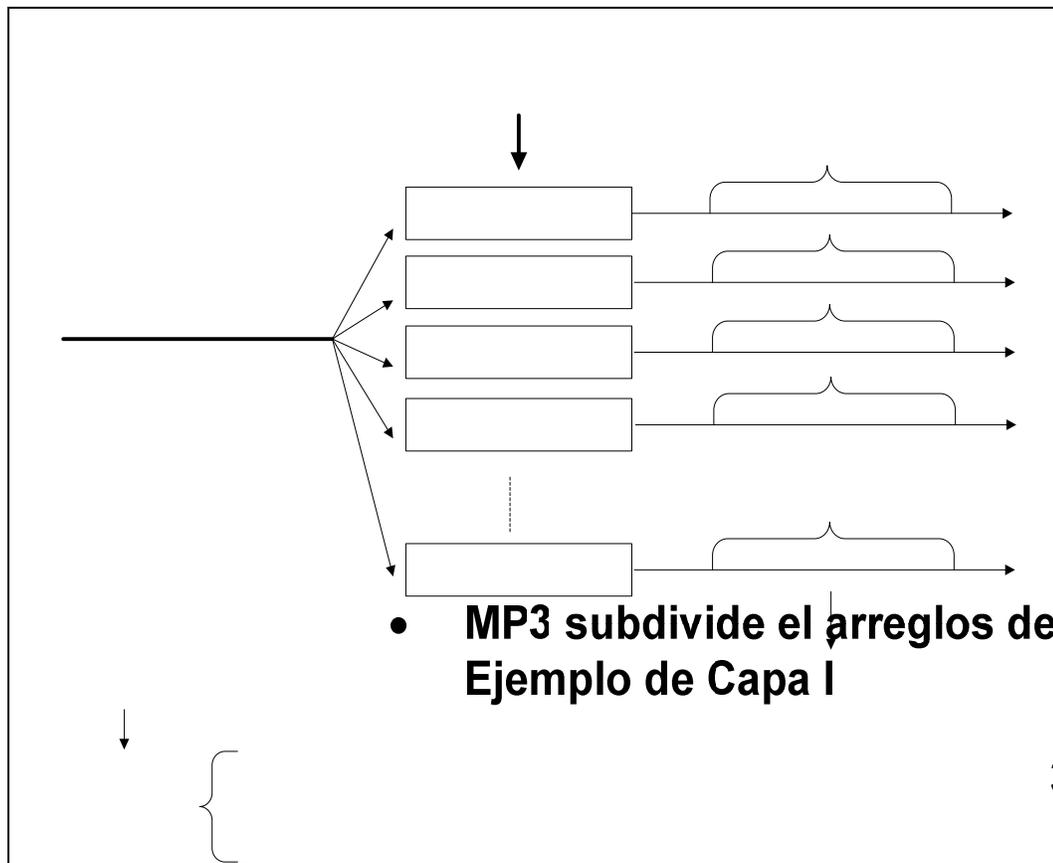


Figura 1. Cuadros de muestras en las diferentes capas MPEG. Fuente [27]

2.1 Capa I (Layer I).

El codificador para capa I usa bloques por sub-banda fijos de 12 muestras diezmadas. El mapeado tiempo-frecuencia se realiza con un banco de filtros polifase con 32 sub-bandas. Los filtros polifase consisten en un conjunto de filtros con el mismo ancho de banda con interrelaciones de fase especiales que ofrecen una

Entrada de muestras de audio

Filtro de s

Filtro de s

Filtro de s

Filtro de s

implementación eficiente del filtro sub-banda. Se denomina filtro sub-banda al que cubre todo el rango de frecuencias deseado. En general, los filtros polifase combinan una baja complejidad de computación con un diseño flexible y múltiples opciones de implementación. [7]

El modelo psicoacústico utiliza una FFT (*Fast Fourier Transform*) de 512 puntos para obtener información espectral detallada de la señal. El resultado de la aplicación de la FFT se utiliza para determinar los enmascaramientos en la señal, cada uno de los cuales produce un nivel de enmascaramiento, según la frecuencia, intensidad y tono. Para cada sub-banda, los niveles individuales se combinan y forman uno global, que se compara con el máximo nivel de señal en la banda, produciendo el SMR que se introduce en el cuantificador. [7]

El bloque de cuantificación y codificación examina las muestras de cada sub-banda, encuentra el máximo valor absoluto y lo cuantifica con 6 bits. Este valor es el factor de escala de la sub-banda. A continuación se determina la asignación de bits para cada sub-banda minimizando el NMR (*noise-to-mask ratio*) total. Es posible que algunas sub-bandas con un gran enmascaramiento terminen con cero bits, es decir, no se codificará ninguna muestra. Por último las muestras de sub-banda se cuantifican linealmente según el número de bits asignados a dicha sub-banda concreta. [7]

El trabajo del empaquetador de trama es sencillo. La trama, según la definición ISO, es la menor parte de la trama de bits decodificable por sí misma. Cada trama empieza con una cabecera para sincronización y diferenciación, así como 16 bits opcionales de CRC para detección y corrección de errores. Se emplean, para cada sub-banda, 4 bits para describir la asignación de bits y otros 6 para el factor de escala. El resto de bits en la trama se utilizan para la información de muestras, 384 en total, y con la opción de añadir cierta información adicional (ver figura 47). A 48 KHz, cada trama lleva 8 ms de sonido. [7]

Cada factor de escala está representado por 6 bits y es transmitido para cada

bloque de sub-banda a menos que la regla de asignación de bits indique que dicho bloque de sub-banda y su factor de escala no necesitan ser transmitidos. [7]

Por cada bloque de 12 muestras se calcula la SMR a través de la FFT de 512 puntos. Para cada sub-banda la asignación de bits selecciona un cuantificador uniforme de un conjunto de 15 cuantificadores con $M=2^m-1$ niveles ($m=0$ o $m=2\dots 15$ bits). 4 Bits se necesitan por bloque para la información de asignación de bits. [7]

En resumen: Incluye la división del mapeado básico de la señal de audio digital en 32 sub-bandas, segmentación para el formateo de los datos, modelo psicoacústico y cuantificación fija. El retraso mínimo teórico es de 19 ms. [7]

2.2. Capa 2 (*Layer II*).

El mapeado de tiempo-frecuencia es idéntico al de capa I. El codificador Capa 2 de la ISO/MPEG es básicamente similar al Capa 1 pero tiene una complejidad mayor con la que consigue un mejor rendimiento ya que la FFT es de orden 1024 con lo que se obtiene una resolución en frecuencia más fina para el cálculo de SMR global. El modelo psicoacústico es similar. En los demás aspectos, es idéntico. [7]

Si las características de la señal a codificar lo permiten, entonces tres muestras consecutivas (un *triplet*), pueden codificarse en una única palabra código (*codeword*) para su transmisión o almacenamiento. [7]

El factor de escala total se reduce por un factor de alrededor de 2; en cada sub-banda, bloques de 12 muestras se forman y factores de escala de 3 bloques adyacentes se calculan (lo cual implica $3 \times 12 \times 32 = 1.152$ muestras de entrada son tomadas en cuenta, ver figura 47) Dependiendo de sus valores relativos sólo uno, dos o los tres factores de escala son transmitidos. Solamente uno de los factores de escala tiene que ser transmitido si las diferencias son relativamente pequeñas y solamente el primero de los factores de escala adyacentes tiene que ser transmitido si el segundo tiene un valor más pequeño, como en el caso del pos-enmascaramiento puede ser aprovechado. En el caso de grandes cambios dinámicos en la señal todos los factores

de escala pueden tener que ser usados. El factor de escala seleccionado o los factores de escala son nuevamente representados por 6 bits. El modelo de los factores de escala transmitidos será codificado con 2 bit/sub-banda en la información adyacente (*side information*), denominado información de factor de escala selecto (SCFSI). [7]

Se provee una cuantificación más fina con una resolución de amplitud de hasta 16 bits (lo que reduce el ruido de codificación). Por otro lado, el número de cuantificadores disponibles decrece con el incremento del índice de sub-banda, el cual mantiene pequeña la información lateral (*side information*). La codificación continúa como en capa I. [7]

El bloque de cuantificación y codificación también es similar, generando factores de escala de 6 bits para cada sub-banda. Sin embargo, las tramas de capa II son tres veces más largas que las de capa I, de forma que se concede a cada sub-banda tres factores de escala, y el codificador utiliza uno, dos o los tres, según la diferencia que haya entre ellos. La asignación de bits es similar a la de capa I. [7]

El formateador de trama: la definición ISO de trama es la misma que en el punto anterior. Utiliza la misma cabecera y estructura de CRC que el de capa I. El número de bits que utilizan para describir la asignación de bits varía con las sub-bandas: 4 bits para las inferiores, 3 para las medias y dos para las superiores, adecuándose a las bandas críticas. Los factores de escala se codifican junto a un número de dos bits que indica si se utilizan uno, dos o los tres. Las muestras de sub-banda se cuantifican y a continuación se asocian en grupos de tres, llamados gránulos. Cada uno se codifica con una palabra clave, lo que permite interceptar mucha más información redundante que en el de capa 1. Cada trama contiene, pues, 1152 muestras PCM. A 48 KHz. cada trama lleva 24 ms de sonido. [7]

En resumen: Incluye codificación adicional, factores de escala y diferente composición de trama. El retraso mínimo teórico es de 35 ms.

2.3 Capa 3 (Layer III).

El esquema III es substancialmente más complicado que los dos anteriores e

incluye una serie de mejoras. Su diagrama de flujos es conceptualmente semejante al visto para los otros dos esquemas, salvo que se realizan múltiples iteraciones para procesar los datos con el mayor nivel de calidad en un cierto tiempo, lo cual complica su diseño hasta el punto de que los diagramas ISO ocupan decenas de páginas.[7]

El mapeado de tiempo-frecuencia añade un nuevo banco de filtros, el DCT (*Discrete Cosine Transform*), que con el polifase forman el denominado filtro híbrido. Proporciona una resolución en frecuencia variable, 6x32 o 18x32 sub-bandas, ajustándose mucho mejor a las bandas críticas de las diferentes frecuencias. [7]

El modelo psicoacústico es una modificación del empleado en el esquema II, y utiliza un método denominado predicción polinómica. Incluye los efectos del enmascaramiento temporal. [7]

El bloque de cuantificación y codificación también emplea algoritmos muy sofisticados que permiten tramas de longitud variable. La gran diferencia con los otros dos esquemas es que la variable controlada es el ruido, a través de bucles iterativos que lo reducen al mínimo posible en cada paso. [7]

El formateador de trama: la definición de trama para este esquema según ISO varía respecto de la de los niveles anteriores: "mínima parte de la trama de bits decodificable mediante el uso de información principal adquirida previamente". Las tramas contienen información de 1152 muestras y empiezan con la misma cabecera de sincronización y diferenciación, pero la información perteneciente a una misma trama no se encuentra generalmente entre dos cabeceras. La longitud de la trama puede variarse en caso de necesidad. Además de tratar con esta información, el esquema III incluye codificación Huffman de longitud variable, un método de codificación entrópica que sin pérdida de información elimina redundancia. Los métodos de longitud variable se caracterizan, en general, por asignar palabras cortas a los eventos más frecuentes, dejando las largas para los más infrecuentes. [7]

En resumen: incluye incremento de la resolución en frecuencia, basado en el uso de un banco de filtros híbrido. Cuantificación no uniforme, segmentación

adaptativa y codificación entrópica de los valores cuantificados. El retraso mínimo teórico es de 59 ms.[7]

Tabla 1. Cuadro comparativo de las capas MPEG. Fuente: propia

Esquema	Objetivo (Kbps)	Compresión	Calidad de 64 Kbps (MOS)	Calidad de 128 Kbps (MOS)	Retardo (ms)
Capa I	192	4 a 1	-	-	19
Capa II	128	6 a 1	2,1 a 2,6	Más de 4	35
Capa III	64	12 a 1	3,6 a 3,8	Más de 4	59

La calidad viene dada del 1 al 5, siendo el 5 la superior. (Ver capítulo anexo, Sistemas de compresión). Pese a los números de la norma ISO, el retraso típico acostumbra a ser tres veces mayor en la práctica.[7]

3. La trama de audio

3.1 Trama (Capa I y II)

Parte del *bitstream* que es decodificable por sí mismo. En capa I contiene información de 384 muestras de audio mientras que en capa II de 1152. Comienza con una palabra de sincronismo y termina justo antes de la próxima palabra de sincronismo. Consiste en un número entero de slots (4 bytes p/capa I y 1 byte p/capa II). [7]

3.2 Trama (Capa III)

Parte de la trama de bits que es decodificable con el uso de la información principal (*main information*) previamente adquirida. Cada trama en capa III contiene información de 1152 muestras. Aunque la distancia entre dos tramas consecutivas es número entero de slots (1 byte p/capa III) la información de audio perteneciente a una trama no está generalmente contenida entre dos palabras de sincronismo. [7]

3.3 Cabecera (*Header*)

La cabecera es la parte del *bitstream* que contiene la sincronización e

información de estado. Se especifica en la figura 47. Dentro de la cabecera se encuentran los siguientes campos: [7]

- *Syncword* (Palabra de sincronismo) (12 bits): “1111 1111 1111”
- ID (1 bit): indica si la trama responde a la norma 11172 (ID=1) o no (ID=0)
- *Layer* (2 bits): de acuerdo a su contenido indica:

Tabla 2. Cuadro de valores de *Layer*. Fuente: [12]

<u>Layer</u>	
11	Capa I
10	Capa II
01	Capa III
00	Reservado

Para cambiar de capa se necesita un reset del decodificador de audio.

- *Protection_bit* (1 bit): Indica si se ha introducido redundancia en la trama de bits de audio para facilitar la detección y/o cancelación de errores:

1 : indica que NO se adicione redundancia 0 : si se adicionó redundancia [7]

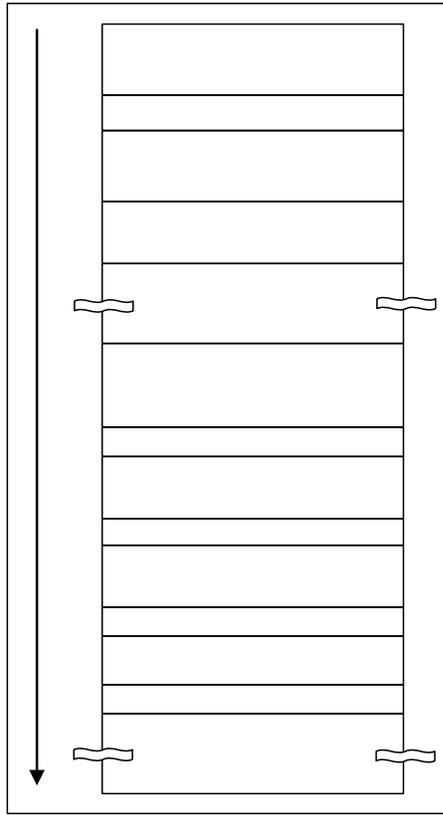


Figura 2. Diagrama de la cabecera (Header) de una trama: (Común a las 3 capas) Fuente:[12]

- *Bitrate_index* (4 bits): Indica la tasa de bits. Si el valor es 0000 indica que se transmite en formato libre lo que indica que la tasa de bits no necesita estar en la lista de *bitrates*. Formato fijo significa que la trama tiene N o N+1 slots dependiendo del valor del bit de relleno (*padding bit*). El *bitrate_index* es una entrada a la tabla de tasas de bits la cual varía para diferentes capas. [7]

Tabla 3. Índice de la variable *Bitrate_index*. Fuente:[12]

Bitrate_index	Capa I	Capa II	Capa III
0000	Free	Free	Free
0001	32	32	32
0010	64	48	40
0011	96	56	48
0100	128	64	56
0101	160	80	64
0110	192	96	80
0111	224	112	96
1000	256	128	112

Palabra

Capa

B
(Prot

Indice
E

Frecue
(sampl

Bit de re

Bit priv

Modo

Ex
(Modo

Co

Orig./Co

Énfasis

Cheq
redu
(CRC

1001	288	160	128
1010	320	192	160
1011	352	224	192
1100	384	256	224
1101	416	320	256
1110	448	384	320
1111	forbiden	forbiden	forbiden

El *bitrate_index* indica la tasa de bits total para el modo seleccionado (estéreo, joint_stereo, doble canal o canal único). A los efectos de proveer del menor retardo posible y complejidad el decodificador no necesita soportar tasas de bits continuamente variables cuando decodifica tramas de las capas I y II. [13]

La capa III soporta una tasa de bits variable por medio de conmutaciones en el *bitrate_index*. La conmutación del *bitrate_index* puede ser utilizada tanto para optimizar los requisitos de almacenamiento o para interpolar una tasa de bits que se encuentre entre dos tasas de bits contiguas en la tabla. Cuando se utiliza formato libre se requiere que la tasa de bits se mantenga fija. Por otra parte, al decodificador no se le exige que soporte tasas de bits superiores a 448, 384 y 320 Kbps para las capas I, II y III respectivamente cuando se transmite en formato libre. [13]

Para la capa II no todas las combinaciones de tasas de bits están disponibles en todos los modos. Esto se ve en la siguiente tabla:

Tabla 4. Modos permitidos según el *bitrate*. Fuente:[12]

Bitrate (kbits)	Modos permitidos
free	todos los modos
32	único canal
48	único canal
56	único canal
64	todos los modos
80	único canal
96	todos los modos
112	todos los modos
128	todos los modos
160	todos los modos
192	todos los modos
224	estéreo, est intenso, doble canal

256	estéreo, est. intenso, doble canal
320	estéreo, est. intenso, doble canal
384	estéreo, est. intenso, doble canal

- *Sampling_frequency* (2 bits): Indica la frecuencia de muestreo de acuerdo a la siguiente tabla:

Tabla 5. Frecuencia de muestreo. Fuente:[12]

Frec. de muestreo	Frec. especific. (kHz)
00	44.1
01	48
10	32
11	reservado

Se requiere un reset del decodificador para cambiar la tasa de muestreo.

- *Padding_bit* (1 bit): Si es =1 indica que la trama contiene un slot adicional para ajustar la tasa de bits a la frecuencia de muestreo. Si es 0 la trama sólo tiene N Slots. El relleno es necesario con una frecuencia de muestreo de 44.1 kHz y también puede ser necesario en formato libre. El relleno se deberá aplicar al *bitstream* de forma tal que la longitud acumulada de las tramas codificadas después de un cierto número de tramas no se desvíe en más de (+0, -1 slot) del siguiente valor calculado: [13]

$$Long. \text{ de trama acumulada} = \sum_{1ra \text{ trama}}^{trama \text{ actual}} \left(\frac{(\text{tamaño de la trama}) \cdot (\text{tasa de bits})}{frecuencia \text{ de muestreo}} \right) \quad \text{Ec.1}$$

Donde el tamaño de la trama es: 384 para capa I y 1.152 para capa II y III.

- *Private_bit* (1 bit): Para uso privado. No cumple función en la norma ISO 11172
- *Mode* (2 bits) Indica el modo de acuerdo a la tabla 7. Hay 4 diferentes modos posibles para MPEG 1. En capa III *joint_stereo* es *intensity stereo* y/o *ms_stereo*. *Joint Stereo* es uno de los modos del algoritmo de codificación de audio donde se explotan la redundancia e irrelevancia estereofónica (para mayor detalles ver Cap. VI El modo *Joint Stereo*). En modo *joint_stereo* el límite es determinado por la tabla de *mode_extension*. [13]

Tabla 6. Valores de *Mode*. Fuente:[12]

Mode	Modo especificado
00	Estéreo (canal derecho e izquierdo de audio en un <i>bitstream</i>)
01	<i>joint_st (intensity_st and/or ms_stereo)</i> (2 canales donde pueden explotarse redundancias e irrelevancias entre ambos)
10	canal dual (2 canales de audio independientes el uno del otro)
11	canal simple (monofónico)

- *Mode_extension* (2 bits): Estos bits se utilizan en el modo *joint_stereo*. En capa I y II indican que sub-bandas están en *intensity_stereo*. Todas las otras sub-bandas se codifican en estéreo. [13]

Tabla 7. Valores de *Mode_extension* Fuente:[12]

Mode_extension	Extensión
00	subbandas 4-31 en <i>intensity_st</i> , límite=4
01	subbandas 8-31 en <i>intensity_st</i> , límite=8
10	subbandas 12-31 en <i>intensity_st</i> , límite=12
11	subbandas 16-31 en <i>intensity_st</i> , límite=16

En capa III indican que tipo de método de codificación de *joint_stereo* se aplica. Los rangos de frecuencias sobre los que se aplican los modos *intensity_stereo* y *ms_stereo* están implícitos en el algoritmo. Mayores detalles se dan en el proceso de decodificación de capa III. [13]

Tabla 8. Modo para el *joint_stereo* según *Mode_extension*. Fuente:[12]

Mode_extension	intensity_st	ms_st
00	off	off
01	on	off
10	off	on
11	on	on

El modo estéreo se utiliza si los bits de modo (*mode_bits*) especifican “stereo” o si en los bits de modo (*mode_bits*) se especifica ‘joint_stereo’ y en *mode_extension* se especifica: *intensity_stereo* = off y *ms_stereo*= off. [13]

- *Copyright* (1 bit): ‘1’ indica protección de Copyright, ‘0’ que no está protegido el bitstream por el Copyright.

- *Original/copy* (1 bit): ‘0’ indica copia, ‘1’ indica que es original.

- *Emphasis*(1 bit): ‘1’ indica que se utiliza énfasis en la codificación, según tabla 9. [13]

Tabla 9. Valores de *Emphasis*. Fuente:[12]

Emphasis	Énfasis especificado
00	Ninguno
01	50/15 microsegundos
10	Reservado
11	CCITT J.17

- *RC_check*: Una palabra de 16 bits se utiliza en forma opcional dentro del *bitstream* codificado. Esta palabra contiene un código de redundancia cíclica que sirve para chequear los datos de la cabecera de la trama y determinar si se produjo error en la recepción de los datos correspondientes a ésta. [13]

Syntax	No. of bits	Mnemonic
<pre>audio sequence() { while (nextbits()==syncword) { frame() } }</pre>		

2.4.1.2 Audio frame

Syntax	No. of bits	Mnemonic
<pre>frame() { header() error_check() audio_data() ancillary_data() }</pre>		

2.4.1.3 Header

Syntax	No. of bits	Mnemonic
<pre>header() { syncword ID layer protection_bit bitrate_index sampling_frequency padding_bit private_bit mode mode_extension copyright original/copy emphasis }</pre>	<p>12</p> <p>1</p> <p>2</p> <p>1</p> <p>4</p> <p>2</p> <p>1</p> <p>1</p> <p>2</p> <p>2</p> <p>1</p> <p>1</p> <p>2</p>	<p>bslbf</p>

2.4.1.4 Error check

Syntax	No. of bits	Mnemonic
<pre>error_check() { if (protection_bit==0) crc_check }</pre>	16	rpchof

Figura 3. Estructura de la variable RC_check. Fuente:[12]

[ANEXO 6]

Tabla 1. Resumen de los estándares y su contenido.

MPEG-1
Information Technology—Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbps, ISO, Geneva, 1993.
ISO/IEC 11172-1, Part 1: Systems
ISO/IEC 11172-2, Part 2: Video
ISO/IEC 11172-3, Part 3: Audio
1. GENERAL NORMATIVE ELEMENTS
1.1 Scope
1.2 Organization of the Document
1.3 Normative References
2. TECHNICAL NORMATIVE ELEMENTS
2.1 Definitions
2.2 Symbols and Abbreviations
2.3 Method of Describing Bitstream Syntax
2.4 Requirements
2.4.1 Coding Structure and Parameters
2.4.2 Specification of the Coded Audio Bitstream Syntax
2.4.3 Semantics for the Audio Bitstream Syntax
2.4.4 The Audio Decoding Process
2.4.5 Compliance Requirements
3-Annex A (normative) Diagrams
3-Annex B (normative) Tables
3-Annex C (informative) The Encoding Process
3-Annex D (informative) Psychoacoustic Models
3-Annex E (informative) Bit Sensitivity to Errors
3-Annex F (informative) Error Concealment
3-Annex G (informative) Joint Stereo Coding
ISO/IEC 11172-4, Part 4: Conformance Testing
ISO/IEC 11172-5, Part 5: Software Simulation
MPEG-2
Information Technology—Generic Coding of Moving Pictures and Associated Audio Information (MPEG-2), ISO, Geneva, 1996. MPEG-2 was partially issued as an International Standard in collaboration with ITU-T SG15
ISO/IEC IS 13818-1 ITU-T H.222.0, Part 1: Systems
ISO/IEC IS 13818-2 ITU-T H.262), Part 2: Video
ISO/IEC IS 13818-3, Part 3: Audio
ISO/IEC IS 13818-4, Part 4: Compliance testing
ISO/IEC IS 13818-5, Part 5: Software simulation (Future TR)
ISO/IEC IS 13818-6, Part 6: Extensions for Digital Storage Media Command and Control (DSM-CC)
Part - 7 Extension for audio(NBC)
Part - 8 Extension for 10-bit video
Part - 9 Extension for realtime interface
ISO/IEC IS 13818-9, Part 9: Extension for RealTime Interface for Systems Decoders
Part -10 Extension for conformance testing
MPEG-4, versión 1
R. Koenen, ISO/IEC JTC1/SC29/WG11 N2323, MPEG-4 Overview (Dublin version), ISO, Geneva, July 1998.
MPEG-7
MPEG home page, http://www.cseit.it/mpeg
ISO/MPEG N2728, Applications for MPEG-7, MPEG Requirements Group, ISO, Geneva, March 1999, http://www.darmstadt.gmd.de/mobile/MPEG7/Documents/N2728.html .
ISO/MPEG N2729, MPEG-7 Context and Objectives, MPEG Requirements Group, Geneva, March 1999, http://www.darmstadt.gmd.de/mobile/MPEG7/Documents/N2729.html .

Para evitar confusión, las siguientes tablas siguen numeración de la Norma ISO/IEC 11172-3

Tabla 3-D.1a.: Frecuencias, Índice de bandas críticas, y Umbral Absoluto. Tabla válida para Capa I a una tasa de muestro de 32 kHz.

Index Number i	Frequency [Hz]	Crit. Band Rate [z]	Absolute Threshold [dB]
1	62,50	,617	33,44
2	125,00	1,232	19,20
3	187,50	1,842	13,87
4	250,00	2,445	11,01
5	312,50	3,037	9,20
6	375,00	3,618	7,94
7	437,50	4,185	7,00
8	500,00	4,736	6,28
9	562,50	5,272	5,70
10	625,00	5,789	5,21
11	687,50	6,289	4,80
12	750,00	6,770	4,45
13	812,50	7,233	4,14
14	875,00	7,677	3,86
15	937,50	8,103	3,61
16	1000,00	8,511	3,37
17	1062,50	8,901	3,15
18	1125,00	9,275	2,93
19	1187,50	9,632	2,73
20	1250,00	9,974	2,53
21	1312,50	10,301	2,32
22	1375,00	10,614	2,12
23	1437,50	10,913	1,92
24	1500,00	11,199	1,71
25	1562,50	11,474	1,49
26	1625,00	11,736	1,27
27	1687,50	11,988	1,04
28	1750,00	12,230	0,80
29	1812,50	12,461	0,55
30	1875,00	12,684	0,29
31	1937,50	12,898	0,02
32	2000,00	13,104	-0,25
33	2062,50	13,302	-0,54
34	2125,00	13,493	-0,83
35	2187,50	13,678	-1,12
36	2250,00	13,855	-1,43
37	2312,50	14,027	-1,73
38	2375,00	14,193	-2,04
39	2437,50	14,354	-2,34
40	2500,00	14,509	-2,64
41	2562,50	14,660	-2,93
42	2625,00	14,807	-3,22
43	2687,50	14,949	-3,49
44	2750,00	15,087	-3,74
45	2812,50	15,221	-3,98
46	2875,00	15,351	-4,20
47	2937,50	15,478	-4,40
48	3000,00	15,602	-4,57
49	3125,00	15,841	-4,82
50	3250,00	16,069	-4,96
51	3375,00	16,287	-4,97
52	3500,00	16,496	-4,86
53	3625,00	16,697	-4,63
54	3750,00	16,891	-4,29
55	3875,00	17,078	-3,87
56	4000,00	17,259	-3,39
57	4125,00	17,434	-2,86
58	4250,00	17,605	-2,31
59	4375,00	17,770	-1,77
60	4500,00	17,932	-1,24
61	4625,00	18,089	-0,74
62	4750,00	18,242	-0,29
63	4875,00	18,392	0,12
64	5000,00	18,539	0,48
65	5125,00	18,682	0,79
66	5250,00	18,823	1,06
67	5375,00	18,960	1,29
68	5500,00	19,095	1,49
69	5625,00	19,226	1,66
70	5750,00	19,356	1,81
71	5875,00	19,482	1,95
72	6000,00	19,606	2,08
73	6250,00	19,847	2,33
74	6500,00	20,079	2,59
75	6750,00	20,300	2,86
76	7000,00	20,513	3,17
77	7250,00	20,717	3,51
78	7500,00	20,912	3,89
79	7750,00	21,098	4,31
80	8000,00	21,275	4,79
81	8250,00	21,445	5,31
82	8500,00	21,606	5,88
83	8750,00	21,760	6,50
84	9000,00	21,906	7,19
85	9250,00	22,046	7,93
86	9500,00	22,178	8,75
87	9750,00	22,304	9,63
88	10000,00	22,424	10,58
89	10250,00	22,538	11,60
90	10500,00	22,646	12,71
91	10750,00	22,749	13,90
92	11000,00	22,847	15,18
93	11250,00	22,941	16,54
94	11500,00	23,030	18,01
95	11750,00	23,114	19,57
96	12000,00	23,195	21,23
97	12250,00	23,272	23,01
98	12500,00	23,345	24,90
99	12750,00	23,415	26,90
100	13000,00	23,482	29,03
101	13250,00	23,546	31,28
102	13500,00	23,607	33,67
103	13750,00	23,666	36,19
104	14000,00	23,722	38,86
105	14250,00	23,775	41,67
106	14500,00	23,827	44,63
107	14750,00	23,876	47,76
108	15000,00	23,923	51,04

Tabla 3-D.1b.: Frecuencias, Índice de bandas críticas, y Umbral Absoluto. Tabla válida para Capa I a una tasa de muestro de 44,1 kHz.

Index Number i	Frequency [Hz]	Crit. Band Rate [z]	Absolute Treshold [dB]
1	86,13	,850	25,87
2	172,27	1,694	14,85
3	258,40	2,525	10,72
4	344,53	3,337	8,50
5	430,66	4,124	7,10
6	516,80	4,882	6,11
7	602,93	5,608	5,37
8	689,06	6,301	4,79
9	775,20	6,959	4,32
10	861,33	7,581	3,92
11	947,46	8,169	3,57
12	1033,59	8,723	3,25
13	1119,73	9,244	2,95
14	1205,86	9,734	2,67
15	1291,99	10,195	2,39
16	1378,13	10,629	2,11
17	1464,26	11,037	1,83
18	1550,39	11,421	1,53
19	1636,52	11,783	1,23
20	1722,66	12,125	0,90
21	1808,79	12,448	0,56
22	1894,92	12,753	0,21
23	1981,05	13,042	-0,17
24	2067,19	13,317	-0,56
25	2153,32	13,578	-0,96
26	2239,45	13,826	-1,38
27	2325,59	14,062	-1,79
28	2411,72	14,288	-2,21
29	2497,85	14,504	-2,63
30	2583,98	14,711	3,03
31	2670,12	14,909	-3,41
32	2756,25	15,100	-3,77
33	2842,38	15,284	-4,09
34	2928,52	15,460	-4,37
35	3014,65	15,631	-4,60
36	3100,78	15,796	-4,78
37	3186,91	15,955	-4,91
38	3273,05	16,110	-4,97
39	3359,18	16,260	-4,98
40	3445,31	16,406	-4,92
41	3531,45	16,547	-4,81
42	3617,58	16,685	-4,65
43	3703,71	16,820	-4,43
44	3789,84	16,951	-4,17
45	3875,98	17,079	-3,87
46	3962,11	17,205	-3,54
47	4048,24	17,327	-3,19
48	4134,38	17,447	-2,82
49	4306,64	17,680	-2,06
50	4478,91	17,905	-1,32
51	4651,17	18,121	-0,64
52	4823,44	18,331	-0,04
53	4995,70	18,534	0,47
54	5167,97	18,731	0,89
55	5340,23	18,922	1,23
56	5512,50	19,108	1,51
57	5684,77	19,289	1,74

58	5857,03	19,464	1,93
59	6029,30	19,635	2,11
60	6201,56	19,801	2,28
61	6373,83	19,963	2,46
62	6546,09	20,120	2,63
63	6718,36	20,273	2,82
64	6890,63	20,421	3,03
65	7062,89	20,565	3,25
66	7235,16	20,705	3,49
67	7407,42	20,840	3,74
68	7579,69	20,972	4,02
69	7751,95	21,099	4,32
70	7924,22	21,222	4,64
71	8096,48	21,342	4,98
72	8268,75	21,457	5,35
73	8613,28	21,677	6,15
74	8957,81	21,882	7,07
75	9302,34	22,074	8,10
76	9646,88	22,253	9,25
77	9991,41	22,420	10,54
78	10335,94	22,576	11,97
79	10680,47	22,721	13,56
80	11025,00	22,857	15,31
81	11369,53	22,984	17,23
82	11714,06	23,102	19,34
83	12058,59	23,213	21,64
84	12403,13	23,317	24,15
85	12747,66	23,415	26,88
86	13092,19	23,506	29,84
87	13436,72	23,592	33,05
88	13781,25	23,673	36,52
89	14125,78	23,749	40,25
90	14470,31	23,821	44,27
91	14814,84	23,888	48,59
92	15159,38	23,952	53,22
93	15503,91	24,013	58,18
94	15848,44	24,070	63,49
95	16192,97	24,125	68,00
96	16537,50	24,176	68,00
97	16882,03	24,225	68,00
98	17226,56	24,271	68,00
99	17571,09	24,316	68,00
100	17915,63	24,358	68,00
101	18260,16	24,398	68,00
102	18604,69	24,436	68,00
103	18949,22	24,473	68,00
104	19293,75	24,508	68,00
105	19638,28	24,542	68,00
106	19982,81	24,574	68,00

Tabla 3-D.1c: Frecuencias, Índice de bandas críticas, y Umbral Absoluto. Tabla válida para Capa I a una tasa de muestro de 48 kHz.

Index Number i	Frequency [Hz]	Crit. Band Rate [z]	Absolute Treshold [dB]
1	93,75	,925	24,17
2	187,50	1,842	13,87
3	281,25	2,742	10,01
4	375,00	3,618	7,94
5	468,75	4,463	6,62
6	562,50	5,272	5,70
7	656,25	6,041	5,00
8	750,00	6,770	4,45
9	843,75	7,457	4,00
10	937,50	8,103	3,61
11	1031,25	8,708	3,26
12	1125,00	9,275	2,93
13	1218,75	9,805	2,63
14	1312,50	10,301	2,32
15	1406,25	10,765	2,02
16	1500,00	11,199	1,71
17	1593,75	11,606	1,38
18	1687,50	11,988	1,04
19	1781,25	12,347	0,67
20	1875,00	12,684	0,29
21	1968,75	13,002	-0,11
22	2062,50	13,302	-0,54
23	2156,25	13,586	-0,97
24	2250,00	13,855	-1,43
25	2343,75	14,111	-1,88
26	2437,50	14,354	-2,34
27	2531,25	14,585	-2,79
28	2625,00	14,807	-3,22
29	2718,75	15,018	-3,62
30	2812,50	15,221	-3,98
31	2906,25	15,415	-4,30
32	3000,00	15,602	-4,57
33	3093,75	15,783	-4,77
34	3187,50	15,956	-4,91
35	3281,25	16,124	-4,98
36	3375,00	16,287	-4,97
37	3468,75	16,445	-4,90
38	3562,50	16,598	-4,76
39	3656,25	16,746	-4,55
40	3750,00	16,891	-4,29
41	3843,75	17,032	-3,99
42	3937,50	17,169	-3,64
43	4031,25	17,303	-3,26
44	4125,00	17,434	-2,86
45	4218,75	17,563	-2,45
46	4312,50	17,688	-2,04
47	4406,25	17,811	-1,63
48	4500,00	17,932	-1,24
49	4687,50	18,166	-0,51
50	4875,00	18,392	0,12
51	5062,50	18,611	,64
52	5250,00	18,823	1,06
53	5437,50	19,028	1,39
54	5625,00	19,226	1,66
55	5812,50	19,419	1,88
56	6000,00	19,606	2,08
57	6187,50	19,788	2,27
58	6375,00	19,964	2,46

59	6562,50	20,135	2,65
60	6750,00	20,300	2,86
61	6937,50	20,461	3,09
62	7125,00	20,616	3,33
63	7312,50	20,766	3,60
64	7500,00	20,912	3,89
65	7687,50	21,052	4,20
66	7875,00	21,188	4,54
67	8062,50	21,318	4,91
68	8250,00	21,445	5,31
69	8437,50	21,567	5,73
70	8625,00	21,684	6,18
71	8812,50	21,797	6,67
72	9000,00	21,906	7,19
73	9375,00	22,113	8,33
74	9750,00	22,304	9,63
75	10125,00	22,482	11,08
76	10500,00	22,646	12,71
77	10875,00	22,799	14,53
78	11250,00	22,941	16,54
79	11625,00	23,072	18,77
80	12000,00	23,195	21,23
81	12375,00	23,309	23,94
82	12750,00	23,415	26,90
83	13125,00	23,515	30,14
84	13500,00	23,607	33,67
85	13875,00	23,694	37,51
86	14250,00	23,775	41,67
87	14625,00	23,852	46,17
88	15000,00	23,923	51,04
89	15375,00	23,991	56,29
90	15750,00	24,054	61,94
91	16125,00	24,114	68,00
92	16500,00	24,171	68,00
93	16875,00	24,224	68,00
94	17250,00	24,275	68,00
95	17625,00	24,322	68,00
96	18000,00	24,368	68,00
97	18375,00	24,411	68,00
98	18750,00	24,452	68,00
99	19125,00	24,491	68,00
100	19500,00	24,528	68,00
101	19875,00	24,564	68,00
102	20250,00	24,597	68,00

Tabla 3-D.1d.: Frecuencias, Índice de bandas críticas, y Umbral Absoluto. Tabla válida para Capa II a una tasa de muestro de 32 kHz.

Index Number i	Frequency [Hz]	Crit. Band Rate [z]	Absolute Threshold [dB]
1	31,25	,309	58,23
2	62,50	,617	33,44
3	93,75	,925	24,17
4	125,00	1,232	19,20
5	156,25	1,538	16,05
6	187,50	1,842	13,87
7	218,75	2,145	12,26
8	250,00	2,445	11,01
9	281,25	2,742	10,01
10	312,50	3,037	9,20
11	343,75	3,329	8,52
12	375,00	3,618	7,94
13	406,25	3,903	7,44
14	437,50	4,185	7,00
15	468,75	4,463	6,62
16	500,00	4,736	6,28
17	531,25	5,006	5,97
18	562,50	5,272	5,70
19	593,75	5,533	5,44
20	625,00	5,789	5,21
21	656,25	6,041	5,00
22	687,50	6,289	4,80
23	718,75	6,532	4,62
24	750,00	6,770	4,45
25	781,25	7,004	4,29
26	812,50	7,233	4,14
27	843,75	7,457	4,00
28	875,00	7,677	3,86
29	906,25	7,892	3,73
30	937,50	8,103	3,61
31	968,75	8,309	3,49
32	1000,00	8,511	3,37
33	1031,25	8,708	3,26
34	1062,50	8,901	3,15
35	1093,75	9,090	3,04
36	1125,00	9,275	2,93
37	1156,25	9,456	2,83
38	1187,50	9,632	2,73
39	1218,75	9,805	2,63
40	1250,00	9,974	2,53
41	1281,25	10,139	2,42
42	1312,50	10,301	2,32
43	1343,75	10,459	2,22
44	1375,00	10,614	2,12
45	1406,25	10,765	2,02
46	1437,50	10,913	1,92
47	1468,75	11,058	1,81
48	1500,00	11,199	1,71
49	1562,50	11,474	1,49
50	1625,00	11,736	1,27
51	1687,50	11,988	1,04
52	1750,00	12,230	0,80
53	1812,50	12,461	0,55
54	1875,00	12,684	0,29
55	1937,50	12,898	0,02
56	2000,00	13,104	-0,25
57	2062,50	13,302	-0,54
58	2125,00	13,493	-0,83
59	2187,50	13,678	-1,12
60	2250,00	13,855	-1,43
61	2312,50	14,027	-1,73
62	2375,00	14,193	-2,04
63	2437,50	14,354	-2,34
64	2500,00	14,509	-2,64
65	2562,50	14,660	-2,93
66	2625,00	14,807	-3,22
67	2687,50	14,949	-3,49
68	2750,00	15,087	-3,74
69	2812,50	15,221	-3,98
70	2875,00	15,351	-4,20
71	2937,50	15,478	-4,40
72	3000,00	15,602	-4,57
73	3125,00	15,841	-4,82
74	3250,00	16,069	-4,96
75	3375,00	16,287	-4,97
76	3500,00	16,496	-4,86
77	3625,00	16,697	-4,63
78	3750,00	16,891	-4,29
79	3875,00	17,078	-3,87
80	4000,00	17,259	-3,39
81	4125,00	17,434	-2,86
82	4250,00	17,605	-2,31
83	4375,00	17,770	-1,77
84	4500,00	17,932	-1,24
85	4625,00	18,089	-0,74
86	4750,00	18,242	-0,29
87	4875,00	18,392	0,12
88	5000,00	18,539	0,48
89	5125,00	18,682	0,79
90	5250,00	18,823	1,06
91	5375,00	18,960	1,29
92	5500,00	19,095	1,49
93	5625,00	19,226	1,66
94	5750,00	19,356	1,81
95	5875,00	19,482	1,95
96	6000,00	19,606	2,08
97	6250,00	19,847	2,33
98	6500,00	20,079	2,59
99	6750,00	20,300	2,86
100	7000,00	20,513	3,17
101	7250,00	20,717	3,51
102	7500,00	20,912	3,89
103	7750,00	21,098	4,31
104	8000,00	21,275	4,79
105	8250,00	21,445	5,31
106	8500,00	21,606	5,88
107	8750,00	21,760	6,50
108	9000,00	21,906	7,19
109	9250,00	22,046	7,93
110	9500,00	22,178	8,75
111	9750,00	22,304	9,63
112	10000,00	22,424	10,58
113	10250,00	22,538	11,60
114	10500,00	22,646	12,71
115	10750,00	22,749	13,90
116	11000,00	22,847	15,18
117	11250,00	22,941	16,54
118	11500,00	23,030	18,01
119	11750,00	23,114	19,57
120	12000,00	23,195	21,23
121	12250,00	23,272	23,01
122	12500,00	23,345	24,90
123	12750,00	23,415	26,90

124	13000,00	23,482	29,03	47	2024,12	13,181	-0,36
125	13250,00	23,546	31,28	48	2067,19	13,317	-0,56
126	13500,00	23,607	33,67	49	2153,32	13,578	-0,96
127	13750,00	23,666	36,19	50	2239,45	13,826	-1,38
128	14000,00	23,722	38,86	51	2325,59	14,062	-1,79
129	14250,00	23,775	41,67	52	2411,72	14,288	-2,21
130	14500,00	23,827	44,63	53	2497,85	14,504	-2,63
131	14750,00	23,876	47,76	54	2583,98	14,711	-3,03
132	15000,00	23,923	51,04	55	2670,12	14,909	-3,41

Tabla 3-D.1e.: Frecuencias, Índice de bandas críticas, y Umbral Absoluto. Tabla válida para Capa II a una rata de muestro de 44,1 kHz.

Index Number i	Frequency [Hz]	Crit. Band Rate [z]	Absolute Threshold [dB]
1	43,07	,425	45,05
2	86,13	,850	25,87
3	129,20	1,273	18,70
4	172,27	1,694	14,85
5	215,33	2,112	12,41
6	258,40	2,525	10,72
7	301,46	2,934	9,47
8	344,53	3,337	8,50
9	387,60	3,733	7,73
10	430,66	4,124	7,10
11	473,73	4,507	6,56
12	516,80	4,882	6,11
13	559,86	5,249	5,72
14	602,93	5,608	5,37
15	646,00	5,959	5,07
16	689,06	6,301	4,79
17	732,13	6,634	4,55
18	775,20	6,959	4,32
19	818,26	7,274	4,11
20	861,33	7,581	3,92
21	904,39	7,879	3,74
22	947,46	8,169	3,57
23	990,53	8,450	3,40
24	1033,59	8,723	3,25
25	1076,66	8,987	3,10
26	1119,73	9,244	2,95
27	1162,79	9,493	2,81
28	1205,86	9,734	2,67
29	1248,93	9,968	2,53
30	1291,99	10,195	2,39
31	1335,06	10,416	2,25
32	1378,13	10,629	2,11
33	1421,19	10,836	1,97
34	1464,26	11,037	1,83
35	1507,32	11,232	1,68
36	1550,39	11,421	1,53
37	1593,46	11,605	1,38
38	1636,52	11,783	1,23
39	1679,59	11,957	1,07
40	1722,66	12,125	0,90
41	1765,72	12,289	0,74
42	1808,79	12,448	0,56
43	1851,86	12,603	0,39
44	1894,92	12,753	0,21
45	1937,99	12,900	0,02
46	1981,05	13,042	-0,17
56	2756,25	15,100	-3,77
57	2842,38	15,284	-4,09
58	2928,52	15,460	-4,37
59	3014,65	15,631	-4,60
60	3100,78	15,796	-4,78
61	3186,91	15,955	-4,91
62	3273,05	16,110	-4,97
63	3359,18	16,260	-4,98
64	3445,31	16,406	-4,92
65	3531,45	16,547	-4,81
66	3617,58	16,685	-4,65
67	3703,71	16,820	-4,43
68	3789,84	16,951	-4,17
69	3875,98	17,079	-3,87
70	3962,11	17,205	-3,54
71	4048,24	17,327	-3,19
72	4134,38	17,447	-2,82
73	4306,64	17,680	-2,06
74	4478,91	17,905	-1,32
75	4651,17	18,121	-,64
76	4823,44	18,331	-,04
77	4995,70	18,534	,47
78	5167,97	18,731	,89
79	5340,23	18,922	1,23
80	5512,50	19,108	1,51
81	5684,77	19,289	1,74
82	5857,03	19,464	1,93
83	6029,30	19,635	2,11
84	6201,56	19,801	2,28
85	6373,83	19,963	2,46
86	6546,09	20,120	2,63
87	6718,36	20,273	2,82
88	6890,63	20,421	3,03
89	7062,89	20,565	3,25
90	7235,16	20,705	3,49
91	7407,42	20,840	3,74
92	7579,69	20,972	4,02
93	7751,95	21,099	4,32
94	7924,22	21,222	4,64
95	8096,48	21,342	4,98
96	8268,75	21,457	5,35
97	8613,28	21,677	6,15
98	8957,81	21,882	7,07
99	9302,34	22,074	8,10
100	9646,88	22,253	9,25
101	9991,41	22,420	10,54
102	10335,94	22,576	11,97
103	10680,47	22,721	13,56
104	11025,00	22,857	15,31
105	11369,53	22,984	17,23
106	11714,06	23,102	19,34
107	12058,59	23,213	21,64
108	12403,13	23,317	24,15
109	12747,66	23,415	26,88
110	13092,19	23,506	29,84
111	13436,72	23,592	33,05

112	13781,25	23,673	36,52	38	1781,25	12,347	0,67
113	14125,78	23,749	40,25	39	1828,13	12,518	0,49
114	14470,31	23,821	44,27	40	1875,00	12,684	0,29
115	14814,84	23,888	48,59	41	1921,88	12,845	0,09
116	15159,38	23,952	53,22	42	1968,75	13,002	-0,11
117	15503,91	24,013	58,18	43	2015,63	13,154	-0,32
118	15848,44	24,070	63,49	44	2062,50	13,302	-0,54
119	16192,97	24,125	68,00	45	2109,38	13,446	-0,75
120	16537,50	24,176	68,00	46	2156,25	13,586	-0,97
121	16882,03	24,225	68,00	47	2203,13	13,723	-1,20
122	17226,56	24,271	68,00	48	2250,00	13,855	-1,43
123	17571,09	24,316	68,00	49	2343,75	14,111	-1,88
124	17915,63	24,358	68,00	50	2437,50	14,354	-2,34
125	18260,16	24,398	68,00	51	2531,25	14,585	-2,79
126	18604,69	24,436	68,00	52	2625,00	14,807	-3,22
127	18949,22	24,473	68,00	53	2718,75	15,018	-3,62
128	19293,75	24,508	68,00	54	2812,50	15,221	-3,98
129	19638,28	24,542	68,00	55	2906,25	15,415	-4,30
130	19982,81	24,574	68,00	56	3000,00	15,602	-4,57

Tabla 3-D.1f: Frecuencias, Índice de bandas críticas, y Umbral Absoluto. Tabla válida para Capa II a una tasa de muestro de 48 kHz.

Index Number i	Frequency [Hz]	Crit. Band Rate [z]	Absolute Threshold [dB]				
1	46,88	,463	42,10	57	3093,75	15,783	-4,77
2	93,75	,925	24,17	58	3187,50	15,956	-4,91
3	140,63	1,385	17,47	59	3281,25	16,124	-4,98
4	187,50	1,842	13,87	60	3375,00	16,287	-4,97
5	234,38	2,295	11,60	61	3468,75	16,445	-4,90
6	281,25	2,742	10,01	62	3562,50	16,598	-4,76
7	328,13	3,184	8,84	63	3656,25	16,746	-4,55
8	375,00	3,618	7,94	64	3750,00	16,891	-4,29
9	421,88	4,045	7,22	65	3843,75	17,032	-3,99
10	468,75	4,463	6,62	66	3937,50	17,169	-3,64
11	515,63	4,872	6,12	67	4031,25	17,303	-3,26
12	562,50	5,272	5,70	68	4125,00	17,434	-2,86
13	609,38	5,661	5,33	69	4218,75	17,563	-2,45
14	656,25	6,041	5,00	70	4312,50	17,688	-2,04
15	703,13	6,411	4,71	71	4406,25	17,811	-1,63
16	750,00	6,770	4,45	72	4500,00	17,932	-1,24
17	796,88	7,119	4,21	73	4687,50	18,166	-,51
18	843,75	7,457	4,00	74	4875,00	18,392	,12
19	890,63	7,785	3,79	75	5062,50	18,611	,64
20	937,50	8,103	3,61	76	5250,00	18,823	1,06
21	984,38	8,410	3,43	77	5437,50	19,028	1,39
22	1031,25	8,708	3,26	78	5625,00	19,226	1,66
23	1078,13	8,996	3,09	79	5812,50	19,419	1,88
24	1125,00	9,275	2,93	80	6000,00	19,606	2,08
25	1171,88	9,544	2,78	81	6187,50	19,788	2,27
26	1218,75	9,805	2,63	82	6375,00	19,964	2,46
27	1265,63	10,057	2,47	83	6562,50	20,135	2,65
28	1312,50	10,301	2,32	84	6750,00	20,300	2,86
29	1359,38	10,537	2,17	85	6937,50	20,461	3,09
30	1406,25	10,765	2,02	86	7125,00	20,616	3,33
31	1453,13	10,986	1,86	87	7312,50	20,766	3,60
32	1500,00	11,199	1,71	88	7500,00	20,912	3,89
33	1546,88	11,406	1,55	89	7687,50	21,052	4,20
34	1593,75	11,606	1,38	90	7875,00	21,188	4,54
35	1640,63	11,800	1,21	91	8062,50	21,318	4,91
36	1687,50	11,988	1,04	92	8250,00	21,445	5,31
37	1734,38	12,170	0,86	93	8437,50	21,567	5,73
				94	8625,00	21,684	6,18
				95	8812,50	21,797	6,67
				96	9000,00	21,906	7,19
				97	9375,00	22,113	8,33
				98	9750,00	22,304	9,63
				99	10125,00	22,482	11,08
				100	10500,00	22,646	12,71
				101	10875,00	22,799	14,53
				102	11250,00	22,941	16,54

103	11625,00	23,072	18,77
104	12000,00	23,195	21,23
105	12375,00	23,309	23,94
106	12750,00	23,415	26,90
107	13125,00	23,515	30,14
108	13500,00	23,607	33,67
109	13875,00	23,694	37,51
110	14250,00	23,775	41,67
111	14625,00	23,852	46,17
112	15000,00	23,923	51,04
113	15375,00	23,991	56,29
114	15750,00	24,054	61,94
115	16125,00	24,114	68,00
116	16500,00	24,171	68,00
117	16875,00	24,224	68,00
118	17250,00	24,275	68,00
119	17625,00	24,322	68,00
120	18000,00	24,368	68,00
121	18375,00	24,411	68,00
122	18750,00	24,452	68,00
123	19125,00	24,491	68,00
124	19500,00	24,528	68,00
125	19875,00	24,564	68,00
126	20250,00	24,597	68,00

Tabla 3-D.2a. Límites de Banda Crítica.

Tabla válida para Capa I a una tasa de muestro de 32 kHz Las frecuencias representan el límite superior de cada banda crítica.

Nro.	Índice de tabla F&CB	Frecuencia [Hz]	Bark [z]
0	1	62,500	0,617
1	3	187,500	1,842
2	5	312,500	3,037
3	7	437,500	4,185
4	9	562,500	5,272
5	11	687,500	6,289
6	13	812,500	7,233
7	15	937,500	8,103
8	18	1.125,000	9,275
9	21	1.312,500	10,301
10	24	1.500,000	11,199
11	27	1.687,500	11,988
12	32	2.000,000	13,104
13	37	2.312,500	14,027
14	44	2.750,000	15,087
15	50	3.250,000	16,069
16	55	3.875,000	17,078
17	61	4.625,000	18,089
18	68	5.500,000	19,095
19	74	6.500,000	20,079
20	79	7.750,000	21,098
21	85	9.250,000	22,046
22	94	11.500,000	23,030
23	108	15.000,000	23,923

Tabla 3-D.2b. Límites de Banda Crítica.

Tabla válida para Capa I a una tasa de muestro de 44,1 kHz Las frecuencias representan el límite superior de cada banda crítica.

Nro.	Índice de tabla F&CB	Frecuencia [Hz]	Bark [z]
0	1	86,133	0,850
1	2	172,266	1,694
2	3	258,398	2,525
3	5	430,664	4,124
4	6	516,797	4,882
5	8	689,063	6,301
6	9	775,195	6,959
7	11	947,461	8,169
8	13	1.119,727	9,244
9	15	1.291,992	10,195
10	17	1.464,258	11,037
11	20	1.722,656	12,125
12	23	1.981,055	13,042
13	27	2.325,586	14,062
14	32	2.756,250	15,100
15	37	3.186,914	15,955
16	45	3.875,977	17,079
17	50	4.478,906	17,904
18	55	5.340,234	18,922
19	61	6.373,828	19,963
20	68	7.579,688	20,971
21	75	9.302,344	22,074

22	81	11.369,531	22.984
23	93	15.503,906	24.013
24	106	19.982,813	24.573

Tabla 3-D.2c. Límites de Banda Crítica.

Tabla válida para Capa I a una tasa de muestro de 48 kHz Las frecuencias representan el límite superior de cada banda crítica.

Nro.	Índice de tabla F&CB	Frecuencia [Hz]	Bark [z]
0	1	93,750	0,925
1	2	187,500	1,842
2	3	281,250	2,742
3	4	375,000	3,618
4	5	468,750	4,463
5	6	562,500	5,272
6	7	656,250	6,041
7	9	843,750	7,457
8	10	937,500	8,103
9	12	1.125,000	9,275
10	14	1.312,500	10,301
11	16	1.500,000	11,199
12	19	1.781,250	12,347
13	21	1.968,750	13,002
14	25	2.343,750	14,111
15	29	2.718,750	15,018
16	35	3.281,250	16,124
17	41	3.843,750	17,032
18	49	4.687,500	18,166
19	53	5.437,500	19,028
20	58	6.375,000	19,964
21	65	7.687,500	21,052
22	73	9.375,000	22,113
23	79	11.625,000	23,072
24	89	15.375,000	23,991
25	102	20.250,000	24,597

Tabla 3-D.2d. Límites de Banda Crítica.

Tabla válida para Capa II a una tasa de muestro de 32 kHz Las frecuencias representan el límite superior de cada banda crítica.

Nro.	Índice de tabla F&CB	Frecuencia [Hz]	Bark [z]
0	1	31,250	0,309
1	3	93,750	0,925
2	6	187,500	1,842
3	10	312,500	3,037
4	13	406,250	3,903
5	17	531,250	5,006
6	21	656,250	6,041
7	25	781,250	7,004
8	30	937,500	8,103
9	35	1.093,750	9,090
10	41	1.281,250	10,139
11	47	1.468,750	11,058
12	51	1.687,500	11,988

13	56	2.000,000	13,104
14	61	2.312,500	14,027
15	68	2.750,000	15,087
16	74	3.250,000	16,069
17	79	3.875,000	17,078
18	85	4.625,000	18,089
19	92	5.500,000	19,095
20	98	6.500,000	20,079
21	103	7.750,000	21,098
22	109	9.250,000	22,046
23	118	11.500,000	23,030
24	132	15.000,000	23,923

Tabla 3-D.2e. Límites de Banda Crítica.

Tabla válida para Capa II a una tasa de muestro de 44,1 kHz Las frecuencias representan el límite superior de cada banda

Nro.	Indice de tabla F&CB	Frecuencia [Hz]	Bark [z]
0	1	43,066	0,425
1	2	86,133	0,850
2	3	129,199	1,273
3	5	215,332	2,112
4	7	301,465	2,934
5	10	430,664	4,124
6	13	559,863	5,249
7	16	689,063	6,301
8	19	818,262	7,274
9	22	947,461	8,169
10	26	1.119,727	9,244
11	30	1.291,992	10,195
12	35	1.507,324	11,232
13	40	1.722,656	12,125
14	46	1.981,055	13,042
15	51	2.325,586	14,062
16	56	2.756,250	15,100
17	62	3.273,047	16,11
18	69	3.875,977	17,079
19	74	4.478,906	17,904
20	79	5.340,234	18,922
21	85	6.373,828	19,963
22	92	7.579,688	20,971
23	99	9.302,344	22,074
24	105	11.369,531	22,984
25	117	15.503,906	24,013
26	130	19.982,813	24,573

Tabla 3-D.2f. Límites de Banda Crítica.

Tabla válida para Capa II a una tasa de muestro de 48 kHz Las frecuencias representan el límite superior de cada banda

Nro.	Indice de tabla F&CB	Frecuencia [Hz]	Bark [z]
0	1	46,875	0,463
1	2	93,750	0,925
2	3	140,625	1,385
3	5	234,375	2,295
4	7	328,125	3,184
5	9	421,875	4,045
6	12	562,500	5,272
7	14	656,250	6,041
8	17	796,875	7,119
9	20	937,500	8,103
10	24	1.125,000	9,275
11	27	1.265,625	10,057
12	32	1.500,000	11,199
13	37	1.734,375	12,170
14	42	1.968,750	13,002
15	49	2.343,750	14,111
16	53	2.718,750	15,018
17	59	3.281,250	16,124
18	65	3.843,750	17,032
19	73	4.687,500	18,166
20	77	5.437,500	19,028
21	82	6.375,000	19,964
22	89	7.687,500	21,052
23	97	9.375,000	22,113
24	103	11.625,000	23,072
25	113	15.375,000	23,991
26	126	20.250,000	24,597

Tabla 3-D.3a. Cálculo de partición. Válida para la frecuencia de muestreo de 32 kHz.

Index	wlow	whigh	bval	minval	TMN
1	1	1	0,00	0,0	24,5
2	2	4	0,63	0,0	24,5
3	5	7	1,56	20,0	24,5
4	8	10	2,50	20,0	24,5
5	11	13	3,44	20,0	24,5
6	14	16	4,34	20,0	24,5
7	17	19	5,17	20,0	24,5
8	20	22	5,94	20,0	24,5
9	23	25	6,63	17,0	24,5
10	26	28	7,28	15,0	24,5
11	29	31	7,90	15,0	24,5
12	32	34	8,50	10,0	24,5
13	35	37	9,06	7,0	24,5
14	38	41	9,65	7,0	24,5
15	42	45	10,28	4,4	24,8
16	46	49	10,87	4,4	25,4
17	50	53	11,41	4,5	25,9
18	54	57	11,92	4,5	26,4
19	58	61	12,39	4,5	26,9
20	62	65	12,83	4,5	27,3
21	66	70	13,29	4,5	27,8
22	71	75	13,78	4,5	28,3
23	76	81	14,27	4,5	28,8
24	82	87	14,76	4,5	29,3
25	88	93	15,22	4,5	29,7
26	94	99	15,63	4,5	30,1
27	100	106	16,06	4,5	30,6
28	107	113	16,47	4,5	31,0
29	114	120	16,86	4,5	31,4
30	121	129	17,25	4,5	31,8
31	130	138	17,65	4,5	32,2
32	139	148	18,05	4,5	32,5
33	149	159	18,42	4,5	32,9
34	160	170	18,81	4,5	33,3
35	171	183	19,18	4,5	33,7
36	184	196	19,55	4,5	34,1
37	197	210	19,93	4,5	34,4
38	211	225	20,29	4,5	34,8
39	226	240	20,65	4,5	35,2
40	241	258	21,02	4,5	35,5
41	259	279	21,38	4,5	35,9
42	280	300	21,74	4,5	36,2
43	301	326	22,10	4,5	36,6
44	327	354	22,44	4,5	36,9
45	355	382	22,79	4,5	37,3
46	383	420	23,14	4,5	37,6
47	421	458	23,49	4,5	38,0
48	459	496	23,83	4,5	38,3
49	497	513	24,07	4,5	38,6

Tabla 3-D.3b. Cálculo de partición. Válida para la frecuencia de muestreo de 44,1 kHz.

Index	wlow	whigh	bval	minval	TMN
1	1	1	0,00	0,0	24,5
2	2	2	0,43	0,0	24,5
3	3	3	0,86	0,0	24,5
4	4	4	1,29	20,0	24,5
5	5	5	1,72	20,0	24,5
6	6	6	2,15	20,0	24,5
7	7	7	2,58	20,0	24,5
8	8	8	3,01	20,0	24,5
9	9	9	3,45	20,0	24,5
10	10	10	3,88	20,0	24,5
11	11	11	4,28	20,0	24,5
12	12	12	4,67	20,0	24,5
13	13	13	5,06	20,0	24,5
14	14	14	5,42	20,0	24,5
15	15	15	5,77	20,0	24,5
16	16	16	6,11	17,0	24,5
17	17	19	6,73	17,0	24,5
18	20	22	7,61	15,0	24,5
19	23	25	8,44	10,0	24,5
20	26	28	9,21	7,0	24,5
21	29	31	9,88	7,0	24,5
22	32	34	10,51	4,4	25,0
23	35	37	11,11	4,5	25,6
24	38	40	11,65	4,5	26,2
25	41	44	12,24	4,5	26,7
26	45	48	12,85	4,5	27,4
27	49	52	13,41	4,5	27,9
28	53	56	13,94	4,5	28,4
29	57	60	14,42	4,5	28,9
30	61	64	14,86	4,5	29,4
31	65	69	15,32	4,5	29,8
32	70	74	15,79	4,5	30,3
33	75	80	16,26	4,5	30,8
34	81	86	16,73	4,5	31,2
35	87	93	17,19	4,5	31,7
36	94	100	17,62	4,5	32,1
37	101	108	18,05	4,5	32,5
38	109	116	18,45	4,5	32,9
39	117	124	18,83	4,5	33,3
40	125	134	19,21	4,5	33,7
41	135	144	19,60	4,5	34,1
42	145	155	20,00	4,5	34,5
43	156	166	20,38	4,5	34,9
44	167	177	20,74	4,5	35,2
45	178	192	21,12	4,5	35,6
46	193	207	21,48	4,5	36,0
47	208	222	21,84	4,5	36,3
48	223	243	22,20	4,5	36,7
49	244	264	22,56	4,5	37,1
50	265	286	22,91	4,5	37,4
51	287	314	23,26	4,5	37,8
52	315	342	23,60	4,5	38,1
53	343	371	23,95	4,5	38,4
54	372	401	24,30	4,5	38,8
55	402	431	24,65	4,5	39,1
56	432	469	25,00	4,5	39,5
57	470	513	25,33	3,5	39,8

Tabla 3-D.3c. Cálculo de partición. Válida para la frecuencia de muestreo de 48 kHz.

Index	wlow	whigh	bval	minval	TMN
1	1	1	0,00	0,0	24,5
2	2	2	0,47	0,0	24,5
3	3	3	0,94	0,0	24,5
4	4	4	1,41	20,0	24,5
5	5	5	1,88	20,0	24,5
6	6	6	2,34	20,0	24,5
7	7	7	2,81	20,0	24,5
8	8	8	3,28	20,0	24,5
9	9	9	3,75	20,0	24,5
10	10	10	4,20	20,0	24,5
11	11	11	4,63	20,0	24,5
12	12	12	5,05	20,0	24,5
13	13	13	5,44	20,0	24,5
14	14	14	5,83	20,0	24,5
15	15	15	6,19	20,0	24,5
16	16	16	6,52	17,0	24,5
17	17	17	6,86	17,0	24,5
18	18	20	7,49	15,0	24,5
19	21	23	8,40	10,0	24,5
20	24	26	9,24	7,0	24,5
21	27	29	9,97	7,0	24,5
22	30	32	10,65	4,4	25,1
23	33	35	11,28	4,5	25,8
24	36	38	11,86	4,5	26,4
25	39	41	12,39	4,5	26,9
26	42	45	12,96	4,5	27,5
27	46	49	13,56	4,5	28,1
28	50	53	14,12	4,5	28,6
29	54	57	14,62	4,5	29,1
30	58	62	15,14	4,5	29,6
31	63	67	15,67	4,5	30,2
32	68	72	16,15	4,5	30,7
33	73	77	16,58	4,5	31,1
34	78	83	17,02	4,5	31,5
35	84	89	17,44	4,5	31,9
36	90	95	17,84	4,5	32,3
37	96	103	18,24	4,5	32,7
38	104	111	18,66	4,5	33,2
39	112	120	19,07	4,5	33,6
40	121	129	19,47	4,5	34,0
41	130	138	19,85	4,5	34,3
42	139	149	20,23	4,5	34,7
43	150	160	20,63	4,5	35,1
44	161	173	21,02	4,5	35,5
45	174	187	21,40	4,5	35,9
46	188	201	21,76	4,5	36,3
47	202	219	22,12	4,5	36,6
48	220	238	22,47	4,5	37,0
49	239	257	22,83	4,5	37,3
50	258	283	23,18	4,5	37,7
51	284	309	23,53	4,5	38,0
52	310	335	23,88	4,5	38,4
53	336	363	24,23	4,5	38,7
54	364	391	24,58	4,5	39,1
55	392	423	24,93	4,5	39,4
56	424	465	25,27	4,5	39,8
57	466	507	25,61	3,5	40,1
58	508	513	25,81	3,5	40,3

Tabla 3-D.4a. Tabla de Umbral Absoluto (Absolute Threshold). Válida para la frecuencia de muestreo de 32 kHz. Un valor de 0 dB representa un nivel en el cálculo del umbral absoluto de 96 dB por debajo de la energía de una onda senoidal de amplitud ± 32760 .

index (line)		absthr (dB)
lower	higher	
1	1	58,23
2	2	33,44
3	3	24,17
4	4	19,20
5	5	16,05
6	6	13,87
7	7	12,26
8	8	11,01
9	9	10,01
10	10	9,20
11	11	8,52
12	12	7,94
13	13	7,44
14	14	7,00
15	15	6,62
16	16	6,28
17	17	5,97
18	18	5,70
19	19	5,44
20	20	5,21
21	21	5,00
22	22	4,80
23	23	4,62
24	24	4,45
25	25	4,29
26	26	4,14
27	27	4,00
28	28	3,86
29	29	3,73
30	30	3,61
31	31	3,49
32	32	3,37
33	33	3,26
34	34	3,15
35	35	3,04
36	36	2,93
37	37	2,83
38	38	2,73
39	39	2,63
40	40	2,53
41	41	2,42
42	42	2,32
43	43	2,22
44	44	2,12
45	45	2,02
46	46	1,92
47	47	1,81
48	48	1,71
49	50	1,49
51	52	1,27
53	54	1,04
55	56	,80
57	57	,55
59	60	,29
61	62	,02
63	64	-,25

65	66	-,54
67	68	-,83
69	70	-1,12
71	72	-1,43
73	74	-1,73
75	76	-2,04
77	78	-2,34
79	80	-2,64
81	82	-2,93
83	84	-3,22
85	86	-3,49
87	88	-3,74
89	90	-3,98
91	92	-4,20
93	94	-4,40
95	96	-4,57
97	100	-4,82
101	104	-4,96
105	108	-4,97
109	112	-4,86
113	116	-4,63
117	120	-4,29
121	124	-3,87
125	128	-3,39
129	132	-2,86
133	136	-2,31
137	140	-1,77
141	144	-1,24
145	148	-,74
149	152	-,29
153	156	,12
157	160	,48
161	164	,79
165	168	1,06
169	172	1,29
173	176	1,49
177	180	1,66
181	184	1,81
185	188	1,95
189	192	2,08
193	200	2,33
201	208	2,59
209	216	2,86
217	224	3,17
225	232	3,51
233	240	3,89
241	248	4,31
249	256	4,79
257	264	5,31
265	272	5,88
273	280	6,50
281	288	7,19
289	296	7,93
297	304	8,75
305	312	9,63
313	320	10,58
321	328	11,60
329	336	12,71
337	344	13,90
345	352	15,18
353	360	16,54
361	368	18,01
369	376	19,57
377	384	21,23
385	392	23,01

393	400	24,90
401	408	26,90
409	416	29,03
417	424	31,28
425	432	33,67
433	440	36,19
441	448	38,86
449	456	41,67
457	464	44,63
465	472	47,76
473	480	51,03

Tabla 3-D.4b. Tabla de Umbral Absoluto (Absolute Threshold). Válida para la frecuencia de muestreo de 44,1 kHz. Un valor de 0 dB representa un nivel en el cálculo del umbral absoluto de 96 dB por debajo de la energía de una onda senoidal de amplitud ± 32760 .

index (line)	absthr	
	lower	higher (dB)
1	1	45,05
2	2	25,87
3	3	18,70
4	4	14,85
5	5	12,41
6	6	10,72
7	7	9,47
8	8	8,50
9	9	7,73
10	10	7,10
11	11	6,56
12	12	6,11
13	13	5,72
14	14	5,37
15	15	5,07
16	16	4,79
17	17	4,55
18	18	4,32
19	19	4,11
20	20	3,92
21	21	3,74
22	22	3,57
23	23	3,40
24	24	3,25
25	25	3,10
26	26	2,95
27	27	2,81
28	28	2,67
29	29	2,53
30	30	2,39
31	31	2,25
32	32	2,11
33	33	1,97
34	34	1,83
35	35	1,68
36	36	1,53
37	37	1,38
38	38	1,23
39	39	1,07
40	40	,90
41	41	,74

42	42	,56
43	43	,39
44	44	,21
45	45	,02
46	46	-,17
47	47	-,36
48	48	-,56
49	50	-,96
51	52	-1,37
53	54	-1,79
55	56	-2,21
57	58	-2,63
59	60	-3,03
61	62	-3,41
63	64	-3,77
65	66	-4,09
67	68	-4,37
69	70	-4,60
71	72	-4,78
73	74	-4,91
75	76	-4,97
77	78	-4,98
79	80	-4,92
81	82	-4,81
83	84	-4,65
85	86	-4,43
87	88	-4,17
89	90	-3,87
91	92	-3,54
93	94	-3,19
95	96	-2,82
97	100	-2,06
101	104	-1,33
105	108	-,64
109	112	-,04
113	116	,47
117	120	,89
121	124	1,23
125	128	1,51
129	132	1,74
133	136	1,93
137	140	2,11
141	144	2,28
145	148	2,45
149	152	2,63
153	156	2,82
157	160	3,03
161	164	3,25
165	168	3,49
169	172	3,74
173	176	4,02
177	180	4,32
181	184	4,64
185	188	4,98
189	192	5,35
193	200	6,15
201	208	7,07
209	216	8,10
217	224	9,25
225	232	10,54
233	240	11,97
241	248	13,56
249	256	15,30
257	264	17,23
265	272	19,33

273	280	21,64
281	288	24,15
289	296	26,88
297	304	29,84
305	312	33,04
313	320	36,51
321	328	40,24
329	336	44,26
337	344	48,58
345	352	53,21
353	360	58,17
361	368	63,48
369	376	69,13
377	384	69,13
385	392	69,13
393	400	69,13
401	408	69,13
409	416	69,13
417	424	69,13
425	432	69,13
433	440	69,13
441	448	69,13
449	456	69,13
457	464	69,13

Tabla 3-D.4c. Tabla de Umbral Absoluto (Absolute Threshold). Válida para la frecuencia de muestreo de 48 kHz. Un valor de 0 dB representa un nivel en el cálculo del umbral absoluto de 96 dB por debajo de la energía de una onda senoidal de amplitud ± 32760 .

index (line)		absthr (dB)
lower	higher	
1	1	42,10
2	2	24,17
3	3	17,47
4	4	13,87
5	5	11,60
6	6	10,01
7	7	8,84
8	8	7,94
9	9	7,22
10	10	6,62
11	11	6,12
12	12	5,70
13	13	5,33
14	14	5,00
15	15	4,71
16	16	4,45
17	17	4,21
18	18	4,00
19	19	3,79
20	20	3,61
21	21	3,43
22	22	3,26
23	23	3,09
24	24	2,93
25	25	2,78
26	26	2,63
27	27	2,47
28	28	2,32
29	29	2,17
30	30	2,02
31	31	1,86
32	32	1,71

33	33	1,55
34	34	1,38
35	35	1,21
36	36	1,04
37	37	,86
38	38	,67
39	39	,49
40	40	,29
41	41	,09
42	42	-,11
43	43	-,32
44	44	-,54
45	45	-,75
46	46	-,97
47	47	-1,20
48	48	-1,43
49	50	-1,88
51	52	-2,34
53	54	-2,79
55	56	-3,22
57	58	-3,62
59	60	-3,98
61	62	-4,30
63	64	-4,57
65	66	-4,77
67	68	-4,91
69	70	-4,98
71	72	-4,97
73	74	-4,90
75	76	-4,76
77	78	-4,55
79	80	-4,29
81	82	-3,99
83	84	-3,64
85	86	-3,26
87	88	-2,86
89	90	-2,45
91	92	-2,04
93	94	-1,63
95	96	-1,24
97	100	-,51
101	104	,12
105	108	,64
109	112	1,06
113	116	1,39
117	120	1,66
121	124	1,88
125	128	2,08
129	132	2,27
133	136	2,46
137	140	2,65
141	144	2,86
145	148	3,09
149	152	3,33
153	156	3,60
157	160	3,89
161	164	4,20
165	168	4,54
169	172	4,91
173	176	5,31
177	180	5,73
181	184	6,18
185	188	6,67
189	192	7,19
193	200	8,33

201	208	9,63
209	216	11,08
217	224	12,71
225	232	14,53
233	240	16,54
241	248	18,77
249	256	21,23
257	264	23,94
265	272	26,90
273	280	30,14
281	288	33,67
289	296	37,51
297	304	41,67
305	312	46,17
313	320	51,04
321	328	56,29
329	332	61,94
333	340	68,00
341	348	68,00
349	356	68,00
357	364	68,00
365	372	68,00
373	380	68,00
381	388	68,00
389	396	68,00
397	404	68,00
405	412	68,00
413	420	68,00
421	428	68,00

Tabla 3-D.5. Tabla del codificador de partición para Capa I y II

	Index		widthn
	w _{low} n+1	w _{high} n	
0	1	0	
1	17	0	
2	33	0	
3	49	0	
4	65	0	
5	81	0	
6	97	0	
7	113	0	
8	129	0	
9	145	0	
10	161	0	
11	177	0	
12	193	0	
13	209	1	
14	225	1	
15	241	1	
16	257	1	
17	273	1	
18	289	1	
19	305	1	
20	321	1	
21	337	1	
22	353	1	
23	369	1	
24	385	1	
25	401	1	

26	417	1
27	433	1
28	449	1
29	465	1
30	481	1
31	497	1
32	513	1

Tabla 3-C.1 Coeficientes Ci del Análisis de ventana.

C[0]= 0,000000000	C[1]= -0,000000477	C[2]= -0,000000477	C[3]= -0,000000477
C[4]= -0,000000477	C[5]= -0,000000477	C[6]= -0,000000477	C[7]= -0,000000954
C[8]= -0,000000954	C[9]= -0,000000954	C[10]= -0,000000954	C[11]= -0,000001431
C[12]= -0,000001431	C[13]= -0,000001907	C[14]= -0,000001907	C[15]= -0,000002384
C[16]= -0,000002384	C[17]= -0,000002861	C[18]= -0,000003338	C[19]= -0,000003338
C[20]= -0,000003815	C[21]= -0,000004292	C[22]= -0,000004768	C[23]= -0,000005245
C[24]= -0,000006199	C[25]= -0,000006676	C[26]= -0,000007629	C[27]= -0,000008106
C[28]= -0,000009060	C[29]= -0,000010014	C[30]= -0,000011444	C[31]= -0,000012398
C[32]= -0,000013828	C[33]= -0,000014782	C[34]= -0,000016689	C[35]= -0,000018120
C[36]= -0,000019550	C[37]= -0,000021458	C[38]= -0,000023365	C[39]= -0,000025272
C[40]= -0,000027657	C[41]= -0,000030041	C[42]= -0,000032425	C[43]= -0,000034809
C[44]= -0,000037670	C[45]= -0,000040531	C[46]= -0,000043392	C[47]= -0,000046253
C[48]= -0,000049591	C[49]= -0,000052929	C[50]= -0,000055790	C[51]= -0,000059605
C[52]= -0,000062943	C[53]= -0,000066280	C[54]= -0,000070095	C[55]= -0,000073433
C[56]= -0,000076771	C[57]= -0,000080585	C[58]= -0,000083923	C[59]= -0,000087261
C[60]= -0,000090599	C[61]= -0,000093460	C[62]= -0,000096321	C[63]= -0,000099182
C[64]= 0,000101566	C[65]= 0,000103951	C[66]= 0,000105858	C[67]= 0,000107288
C[68]= 0,000108242	C[69]= 0,000108719	C[70]= 0,000108719	C[71]= 0,000108242
C[72]= 0,000106812	C[73]= 0,000105381	C[74]= 0,000102520	C[75]= 0,000099182
C[76]= 0,000095367	C[77]= 0,000090122	C[78]= 0,000084400	C[79]= 0,000077724
C[80]= 0,000069618	C[81]= 0,000060558	C[82]= 0,000050545	C[83]= 0,000039577
C[84]= 0,000027180	C[85]= 0,000013828	C[86]= -0,000000954	C[87]= -0,000017166
C[88]= -0,000034332	C[89]= -0,000052929	C[90]= -0,000072956	C[91]= -0,000093937
C[92]= -0,000116348	C[93]= -0,000140190	C[94]= -0,000165462	C[95]= -0,000191212
C[96]= -0,000218868	C[97]= -0,000247478	C[98]= -0,000277042	C[99]= -0,000307560
C[100]= -0,000339031	C[101]= -0,000371456	C[102]= -0,000404358	C[103]= -0,000438213
C[104]= -0,000472546	C[105]= -0,000507355	C[106]= -0,000542164	C[107]= -0,000576973
C[108]= -0,000611782	C[109]= -0,000646591	C[110]= -0,000680923	C[111]= -0,000714302
C[112]= -0,000747204	C[113]= -0,000779152	C[114]= -0,000809669	C[115]= -0,000838757
C[116]= -0,000866413	C[117]= -0,000891685	C[118]= -0,000915051	C[119]= -0,000935555
C[120]= -0,000954151	C[121]= -0,000968933	C[122]= -0,000980854	C[123]= -0,000989437
C[124]= -0,000994205	C[125]= -0,000995159	C[126]= -0,000991821	C[127]= -0,000983715
C[128]= 0,000971317	C[129]= 0,000953674	C[130]= 0,000930786	C[131]= 0,000902653
C[132]= 0,000868797	C[133]= 0,000829220	C[134]= 0,000783920	C[135]= 0,000731945
C[136]= 0,000674248	C[137]= 0,000610352	C[138]= 0,000539303	C[139]= 0,000462532
C[140]= 0,000378609	C[141]= 0,000288486	C[142]= 0,000191689	C[143]= 0,000088215
C[144]= -0,000021458	C[145]= -0,000137329	C[146]= -0,000259876	C[147]= -0,000388145
C[148]= -0,000522137	C[149]= -0,000661850	C[150]= -0,000806808	C[151]= -0,000956535
C[152]= -0,001111031	C[153]= -0,001269817	C[154]= -0,001432419	C[155]= -0,001597881
C[156]= -0,001766682	C[157]= -0,001937389	C[158]= -0,002110004	C[159]= -0,002283096
C[160]= -0,002457142	C[161]= -0,002630711	C[162]= -0,002803326	C[163]= -0,002974033
C[164]= -0,003141880	C[165]= -0,003306866	C[166]= -0,003467083	C[167]= -0,003622532
C[168]= -0,003771782	C[169]= -0,003914356	C[170]= -0,004048824	C[171]= -0,004174709
C[172]= -0,004290581	C[173]= -0,004395962	C[174]= -0,004489899	C[175]= -0,004570484
C[176]= -0,004638195	C[177]= -0,004691124	C[178]= -0,004728317	C[179]= -0,004748821
C[180]= -0,004752159	C[181]= -0,004737377	C[182]= -0,004703045	C[183]= -0,004649162
C[184]= -0,004573822	C[185]= -0,004477024	C[186]= -0,004357815	C[187]= -0,004215240
C[188]= -0,004049301	C[189]= -0,003858566	C[190]= -0,003643036	C[191]= -0,003401756
C[192]= 0,003134727	C[193]= 0,002841473	C[194]= 0,002521515	C[195]= 0,002174854
C[196]= 0,001800537	C[197]= 0,001399517	C[198]= 0,000971317	C[199]= 0,000515938
C[200]= 0,000033379	C[201]= -0,000475883	C[202]= -0,001011848	C[203]= -0,001573563
C[204]= -0,002161503	C[205]= -0,002774239	C[206]= -0,003411293	C[207]= -0,004072189
C[208]= -0,004756451	C[209]= -0,005462170	C[210]= -0,006189346	C[211]= -0,006937027
C[212]= -0,007703304	C[213]= -0,008487225	C[214]= -0,009287834	C[215]= -0,010103703
C[216]= -0,010933399	C[217]= -0,011775017	C[218]= -0,012627602	C[219]= -0,013489246
C[220]= -0,014358521	C[221]= -0,015233517	C[222]= -0,016112804	C[223]= -0,016994476
C[224]= -0,017876148	C[225]= -0,018756866	C[226]= -0,019634247	C[227]= -0,020506859
C[228]= -0,021372318	C[229]= -0,022228718	C[230]= -0,023074150	C[231]= -0,023907185
C[232]= -0,024725437	C[233]= -0,025527000	C[234]= -0,026310921	C[235]= -0,027073860

C[236]=-0,027815342	C[237]=-0,028532982	C[238]=-0,029224873	C[239]=-0,029890060
C[240]=-0,030526638	C[241]=-0,031132698	C[242]=-0,031706810	C[243]=-0,032248020
C[244]=-0,032754898	C[245]=-0,033225536	C[246]=-0,033659935	C[247]=-0,034055710
C[248]=-0,034412861	C[249]=-0,034730434	C[250]=-0,035007000	C[251]=-0,035242081
C[252]=-0,035435200	C[253]=-0,035586357	C[254]=-0,035694122	C[255]=-0,035758972
C[256]=-0,035780907	C[257]=-0,035758972	C[258]=-0,035694122	C[259]=-0,035586357
C[260]=-0,035435200	C[261]=-0,035242081	C[262]=-0,035007000	C[263]=-0,034730434
C[264]=-0,034412861	C[265]=-0,034055710	C[266]=-0,033659935	C[267]=-0,033225536
C[268]=-0,032754898	C[269]=-0,032248020	C[270]=-0,031706810	C[271]=-0,031132698
C[272]=-0,030526638	C[273]=-0,029890060	C[274]=-0,029224873	C[275]=-0,028532982
C[276]=-0,027815342	C[277]=-0,027073860	C[278]=-0,026310921	C[279]=-0,025527000
C[280]=-0,024725437	C[281]=-0,023907185	C[282]=-0,023074150	C[283]=-0,022228718
C[284]=-0,021372318	C[285]=-0,020506859	C[286]=-0,019634247	C[287]=-0,018756866
C[288]=-0,017876148	C[289]=-0,016994476	C[290]=-0,016112804	C[291]=-0,015233517
C[292]=-0,014358521	C[293]=-0,013489246	C[294]=-0,012627602	C[295]=-0,011775017
C[296]=-0,010933399	C[297]=-0,010103703	C[298]=-0,009287834	C[299]=-0,008487225
C[300]=-0,007703304	C[301]=-0,006937027	C[302]=-0,006189346	C[303]=-0,005462170
C[304]=-0,004756451	C[305]=-0,004072189	C[306]=-0,003411293	C[307]=-0,002774239
C[308]=-0,002161503	C[309]=-0,001573563	C[310]=-0,001011848	C[311]=-0,000475883
C[312]=-0,000033379	C[313]=-0,000515938	C[314]=-0,0000971317	C[315]=-0,001399517
C[316]=-0,001800537	C[317]=-0,002174854	C[318]=-0,002521515	C[319]=-0,002841473
C[320]=-0,003134727	C[321]=-0,003401756	C[322]=-0,003643036	C[323]=-0,003858566
C[324]=-0,004049301	C[325]=-0,004215240	C[326]=-0,004357815	C[327]=-0,004477024
C[328]=-0,004573822	C[329]=-0,004649162	C[330]=-0,004703045	C[331]=-0,004737377
C[332]=-0,004752159	C[333]=-0,004748821	C[334]=-0,004728317	C[335]=-0,004691124
C[336]=-0,004638195	C[337]=-0,004570484	C[338]=-0,004489899	C[339]=-0,004395962
C[340]=-0,004290581	C[341]=-0,004174709	C[342]=-0,004048824	C[343]=-0,003914356
C[344]=-0,003771782	C[345]=-0,003622532	C[346]=-0,003467083	C[347]=-0,003306866
C[348]=-0,003141880	C[349]=-0,002974033	C[350]=-0,002803326	C[351]=-0,002630711
C[352]=-0,002457142	C[353]=-0,002283096	C[354]=-0,002110004	C[355]=-0,001937389
C[356]=-0,001766682	C[357]=-0,001597881	C[358]=-0,001432419	C[359]=-0,001269817
C[360]=-0,001111031	C[361]=-0,000956535	C[362]=-0,000806808	C[363]=-0,000661850
C[364]=-0,000522137	C[365]=-0,000388145	C[366]=-0,000259876	C[367]=-0,000137329
C[368]=-0,000021458	C[369]=-0,000088215	C[370]=-0,000191689	C[371]=-0,000288486
C[372]=-0,000378609	C[373]=-0,000462532	C[374]=-0,000539303	C[375]=-0,000610352
C[376]=-0,000674248	C[377]=-0,000731945	C[378]=-0,000783920	C[379]=-0,000829220
C[380]=-0,000868797	C[381]=-0,000902653	C[382]=-0,000930786	C[383]=-0,000953674
C[384]=-0,000971317	C[385]=-0,000983715	C[386]=-0,000991821	C[387]=-0,000995159
C[388]=-0,000994205	C[389]=-0,000989437	C[390]=-0,000980854	C[391]=-0,000968933
C[392]=-0,000954151	C[393]=-0,000935555	C[394]=-0,000915051	C[395]=-0,000891685
C[396]=-0,000866413	C[397]=-0,000838757	C[398]=-0,000809669	C[399]=-0,000779152
C[400]=-0,000747204	C[401]=-0,000714302	C[402]=-0,000680923	C[403]=-0,000646591
C[404]=-0,000611782	C[405]=-0,000576973	C[406]=-0,000542164	C[407]=-0,000507355
C[408]=-0,000472546	C[409]=-0,000438213	C[410]=-0,000404358	C[411]=-0,000371456
C[412]=-0,000339031	C[413]=-0,000307560	C[414]=-0,000277042	C[415]=-0,000247478
C[416]=-0,000218868	C[417]=-0,000191212	C[418]=-0,000165462	C[419]=-0,000140190
C[420]=-0,000116348	C[421]=-0,000093937	C[422]=-0,000072956	C[423]=-0,000052929
C[424]=-0,000034332	C[425]=-0,000017166	C[426]=-0,000000954	C[427]=-0,000013828
C[428]=-0,000027180	C[429]=-0,000039577	C[430]=-0,000050545	C[431]=-0,000060558
C[432]=-0,000069618	C[433]=-0,000077724	C[434]=-0,000084400	C[435]=-0,000090122
C[436]=-0,000095367	C[437]=-0,000099182	C[438]=-0,000102520	C[439]=-0,000105381
C[440]=-0,000106812	C[441]=-0,000108242	C[442]=-0,000108719	C[443]=-0,000108719
C[444]=-0,000108242	C[445]=-0,000107288	C[446]=-0,000105858	C[447]=-0,000103951
C[448]=-0,000101566	C[449]=-0,000099182	C[450]=-0,000096321	C[451]=-0,000093460
C[452]=-0,000090599	C[453]=-0,000087261	C[454]=-0,000083923	C[455]=-0,000080585
C[456]=-0,000076771	C[457]=-0,000073433	C[458]=-0,000070095	C[459]=-0,000066280
C[460]=-0,000062943	C[461]=-0,000059605	C[462]=-0,000055790	C[463]=-0,000052929
C[464]=-0,000049591	C[465]=-0,000046253	C[466]=-0,000043392	C[467]=-0,000040531
C[468]=-0,000037670	C[469]=-0,000034809	C[470]=-0,000032425	C[471]=-0,000030041
C[472]=-0,000027657	C[473]=-0,000025272	C[474]=-0,000023365	C[475]=-0,000021458
C[476]=-0,000019550	C[477]=-0,000018120	C[478]=-0,000016689	C[479]=-0,000014782
C[480]=-0,000013828	C[481]=-0,000012398	C[482]=-0,000011444	C[483]=-0,000010014
C[484]=-0,000009060	C[485]=-0,000008106	C[486]=-0,000007629	C[487]=-0,000006676
C[488]=-0,000006199	C[489]=-0,000005245	C[490]=-0,000004768	C[491]=-0,000004292
C[492]=-0,000003815	C[493]=-0,000003338	C[494]=-0,000003338	C[495]=-0,000002861
C[496]=-0,000002384	C[497]=-0,000002384	C[498]=-0,000001907	C[499]=-0,000001907
C[500]=-0,000001431	C[501]=-0,000001431	C[502]=-0,000000954	C[503]=-0,000000954
C[504]=-0,000000954	C[505]=-0,000000954	C[506]=-0,000000477	C[507]=-0,000000477
C[508]=-0,000000477	C[509]=-0,000000477	C[510]=-0,000000477	C[511]=-0,000000477

Tabla 3-C.2 Relaciones Señal a Ruido para Capa I

No. of steps	SNR (dB)
0	0,00
3	7,00
7	16,00
15	25,28
31	31,59
63	37,75
127	43,84
255	49,89
511	55,93
1023	61,96
2047	67,98
4095	74,01
8191	80,03
16383	86,05
32767	92,01

Tabla 3-C.3 Coeficientes de cuantificación para Capa I.

No. of steps	A	B
3	0,750000000	-0,250000000
7	0,875000000	-0,125000000
15	0,937500000	-0,062500000
31	0,968750000	-0,031250000
63	0,984375000	-0,015625000
127	0,992187500	-0,007812500
255	0,996093750	-0,003906250
511	0,998046875	-0,001953125
1023	0,999023438	-0,000976563
2047	0,999511719	-0,000488281
4095	0,999755859	-0,000244141
8191	0,999877930	-0,000122070
16383	0,999938965	-0,000061035
32767	0,999969482	-0,000030518

Tabla 3-C.4: Patrones de transmisión para Scale_factor en Capa II.

Class1	Class2	Transmission pattern	Select Information
1	1	1 2 3	0
1	2	1 2 2	3
1	3	1 2 2	3
1	4	1 3 3	3
1	5	1 2 3	0
2	1	1 1 3	1
2	2	1 1 1	2
2	3	1 1 1	2
2	4	4 4 4	2
2	5	1 1 3	1
3	1	1 1 1	2
3	2	1 1 1	2
3	3	1 1 1	2
3	4	3 3 3	2
3	5	1 1 3	1
4	1	2 2 2	2
4	2	2 2 2	2
4	3	2 2 2	2
4	4	3 3 3	2
4	5	1 2 3	0
5	1	1 2 3	0
5	2	1 2 2	3

5	3	1 2 2	3
5	4	1 3 3	3
5	5	1 2 3	0

Tabla 3-C.5: Relaciones Señal a Ruido para Capa II

No. of steps	SNR (dB)
0	0,00
3	7,00
5	11,00
7	16,00
9	20,84
15	25,28
31	31,59
63	37,75
127	43,84
255	49,89
511	55,93
1023	61,96
2047	67,98
4095	74,01
8191	80,03
16383	86,05
32767	92,01
65535	98,01

TABLA 3-C.6: Coeficientes de cuantificación para Capa II

No. of steps	A	B
3	0,750000000	-0,250000000
5	0,625000000	-0,375000000
7	0,875000000	-0,125000000
9	0,562500000	-0,437500000
15	0,937500000	-0,062500000
31	0,968750000	-0,031250000
63	0,984375000	-0,015625000
127	0,992187500	-0,007812500
255	0,996093750	-0,003906250
511	0,998046875	-0,001953125
1023	0,999023438	-0,000976563
2047	0,999511719	-0,000488281
4095	0,999755859	-0,000244141
8191	0,999877930	-0,000122070
16383	0,999938965	-0,000061035
32767	0,999969482	-0,000030518
65535	0,999984741	-0,000015259

Tablas 3-C.7: Cálculo de las particiones del umbral con los siguientes parámetros: **width, minval, threshold in quiet, norm and bval**

Tabla 3-C.7a: Frecuencia de muestreo= 48 kHz. Bloque largo.

no.	FFT-lines	minval	qthr	norm	bval
0	1	24,5	4,532	0,970	0,000
1	1	24,5	4,532	0,755	0,469
2	1	24,5	4,532	0,738	0,937
3	1	24,5	0,904	0,730	1,406
4	1	24,5	0,904	0,724	1,875
5	1	20	0,090	0,723	2,344
6	1	20	0,090	0,723	2,812
7	1	20	0,029	0,723	3,281
8	1	20	0,029	0,718	3,750
9	1	20	0,009	0,690	4,199
10	1	20	0,009	0,660	4,625
11	1	18	0,009	0,641	5,047
12	1	18	0,009	0,600	5,437
13	1	18	0,009	0,584	5,828
14	1	12	0,009	0,531	6,187
15	1	12	0,009	0,537	6,522
16	2	6	0,018	0,857	7,174
17	2	6	0,018	0,858	7,800
18	2	3	0,018	0,853	8,402
19	2	3	0,018	0,824	8,966
20	2	3	0,018	0,778	9,483
21	2	3	0,018	0,740	9,966
22	2	0	0,018	0,709	10,426
23	2	0	0,018	0,676	10,866
24	2	0	0,018	0,632	11,279
25	2	0	0,018	0,592	11,669
26	2	0	0,018	0,553	12,042
27	2	0	0,018	0,510	12,386
28	2	0	0,018	0,513	12,721
29	3	0	0,027	0,608	13,115
30	3	0	0,027	0,673	13,561
31	3	0	0,027	0,636	13,983
32	3	0	0,027	0,586	14,371
33	3	0	0,027	0,571	14,741
34	4	0	0,036	0,616	15,140
35	4	0	0,036	0,640	15,562
36	4	0	0,036	0,597	15,962
37	4	0	0,036	0,538	16,324
38	4	0	0,036	0,512	16,665
39	5	0	0,045	0,528	17,020
40	5	0	0,045	0,516	17,373
41	5	0	0,045	0,493	17,708
42	6	0	0,054	0,499	18,045
43	7	0	0,063	0,525	18,398
44	7	0	0,063	0,541	18,762
45	8	0	0,072	0,528	19,120
46	8	0	0,072	0,510	19,466
47	8	0	0,072	0,506	19,807
48	10	0	0,180	0,525	20,159
49	10	0	0,180	0,536	20,522
50	10	0	0,180	0,518	20,873
51	13	0	0,372	0,501	21,214
52	13	0	0,372	0,496	21,553
53	14	0	0,400	0,497	21,892
54	18	0	1,628	0,495	22,231
55	18	0	1,628	0,494	22,569
56	20	0	1,808	0,497	22,909
57	25	0	22,607	0,494	23,248

58	25	0	22,607	0,487	23,583
59	35	0	31,650	0,483	23,915
60	67	0	605,867	0,482	24,246
61	67	0	605,867	0,524	24,576

Tabla 3-C.7b: Frecuencia de muestreo= 44,1 kHz. Bloque largo.

no.	FFT-lines	minval	qthr	norm	bval
0	1	24,5	4,532	0,951	0,000
1	1	24,5	4,532	0,700	0,431
2	1	24,5	4,532	0,681	0,861
3	1	24,5	0,904	0,675	1,292
4	1	24,5	0,904	0,667	1,723
5	1	20	0,090	0,665	2,153
6	1	20	0,090	0,664	2,584
7	1	20	0,029	0,664	3,015
8	1	20	0,029	0,664	3,445
9	1	20	0,029	0,655	3,876
10	1	20	0,009	0,616	4,279
11	1	20	0,009	0,597	4,670
12	1	18	0,009	0,578	5,057
13	1	18	0,009	0,541	5,415
14	1	18	0,009	0,575	5,774
15	2	12	0,018	0,856	6,422
16	2	6	0,018	0,846	7,026
17	2	6	0,018	0,840	7,609
18	2	3	0,018	0,822	8,168
19	2	3	0,018	0,800	8,710
20	2	3	0,018	0,753	9,207
21	2	3	0,018	0,704	9,662
22	2	0	0,018	0,674	10,099
23	2	0	0,018	0,640	10,515
24	2	0	0,018	0,609	10,917
25	2	0	0,018	0,566	11,293
26	2	0	0,018	0,535	11,652
27	2	0	0,018	0,531	11,997
28	3	0	0,027	0,615	12,394
29	3	0	0,027	0,686	12,850
30	3	0	0,027	0,650	13,277
31	3	0	0,027	0,611	13,681
32	3	0	0,027	0,567	14,062
33	3	0	0,027	0,520	14,411
34	3	0	0,027	0,513	14,751
35	4	0	0,036	0,557	15,119
36	4	0	0,036	0,584	15,508
37	4	0	0,036	0,570	15,883
38	5	0	0,045	0,579	16,263
39	5	0	0,045	0,585	16,654
40	5	0	0,045	0,548	17,020
41	6	0	0,054	0,536	17,374
42	6	0	0,054	0,550	17,744
43	7	0	0,063	0,532	18,104
44	7	0	0,063	0,504	18,447
45	7	0	0,063	0,496	18,781
46	9	0	0,081	0,516	19,130
47	9	0	0,081	0,527	19,487
48	9	0	0,081	0,516	19,838
49	10	0	0,180	0,497	20,179
50	10	0	0,180	0,489	20,510
51	11	0	0,198	0,502	20,852
52	14	0	0,400	0,502	21,196
53	14	0	0,400	0,491	21,531

54	15	0	0,429	0,497	21,870
55	20	0	1,808	0,504	22,214
56	20	0	1,808	0,504	22,558
57	21	0	1,899	0,495	22,898
58	27	0	24,415	0,486	23,232
59	27	0	24,415	0,484	23,564
60	36	0	32,554	0,483	23,897
61	73	0	660,124	0,475	24,229
62	18	0	162,770	0,515	24,542

Tabla 3-C.7c: Frecuencia de muestreo= 32 kHz. Bloque largo.

no.	FFT-lines	minval	qthr	norm	bval
0	2	24,5	9,064	0,997	0,312
1	2	24,5	9,064	0,893	0,937
2	2	24,5	1,808	0,881	1,562
3	2	20	0,181	0,873	2,187
4	2	20	0,181	0,872	2,812
5	2	20	0,057	0,871	3,437
6	2	20	0,018	0,860	4,045
7	2	20	0,018	0,839	4,625
8	2	18	0,018	0,812	5,173
9	2	18	0,018	0,784	5,698
10	2	12	0,018	0,741	6,184
11	2	12	0,018	0,697	6,634
12	2	6	0,018	0,674	7,070
13	2	6	0,018	0,651	7,492
14	2	6	0,018	0,633	7,905
15	2	3	0,018	0,611	8,305
16	2	3	0,018	0,589	8,695
17	2	3	0,018	0,575	9,064
18	3	3	0,027	0,654	9,483
19	3	3	0,027	0,724	9,966
20	3	0	0,027	0,701	10,425
21	3	0	0,027	0,673	10,866
22	3	0	0,027	0,631	11,279
23	3	0	0,027	0,592	11,669
24	3	0	0,027	0,553	12,042
25	3	0	0,027	0,510	12,386
26	3	0	0,027	0,505	12,721
27	4	0	0,036	0,562	13,091
28	4	0	0,036	0,598	13,488
29	4	0	0,036	0,589	13,873
30	5	0	0,045	0,607	14,268
31	5	0	0,045	0,620	14,679
32	5	0	0,045	0,580	15,067
33	5	0	0,045	0,532	15,424
34	5	0	0,045	0,517	15,771
35	6	0	0,054	0,517	16,120
36	6	0	0,054	0,509	16,466
37	6	0	0,054	0,506	16,807
38	8	0	0,072	0,522	17,158
39	8	0	0,072	0,531	17,518
40	8	0	0,072	0,519	17,869
41	10	0	0,090	0,512	18,215
42	10	0	0,090	0,509	18,562
43	10	0	0,090	0,497	18,902
44	12	0	0,108	0,494	19,239
45	12	0	0,108	0,501	19,579
46	13	0	0,117	0,507	19,925
47	14	0	0,252	0,502	20,269
48	14	0	0,252	0,493	20,606
49	16	0	0,289	0,497	20,944
50	20	0	0,572	0,506	21,288

51	20	0	0,572	0,510	21,635
52	23	0	0,658	0,504	21,979
53	27	0	2,441	0,496	22,319
54	27	0	2,441	0,493	22,656
55	32	0	2,894	0,490	22,993
56	37	0	33,458	0,483	23,326
57	37	0	33,458	0,458	23,656
58	12	0	10,851	0,500	23,937

Tabla 3-C.7d: Frecuencia de muestreo= 48 kHz. Bloque corto.

no.	FFT-lines	qthr	norm	SNR (db)	bval
0	1	4,532	0,970	-8,240	0,000
1	1	0,904	0,755	-8,240	1,875
2	1	0,029	0,738	-8,240	3,750
3	1	0,009	0,730	-8,240	5,437
4	1	0,009	0,724	-8,240	6,857
5	1	0,009	0,723	-8,240	8,109
6	1	0,009	0,723	-8,240	9,237
7	1	0,009	0,723	-8,240	10,202
8	1	0,009	0,718	-8,240	11,083
9	1	0,009	0,690	-8,240	11,864
10	1	0,009	0,660	-7,447	12,553
11	1	0,009	0,641	-7,447	13,195
12	1	0,009	0,600	-7,447	13,781
13	1	0,009	0,584	-7,447	14,309
14	1	0,009	0,532	-7,447	14,803
15	1	0,009	0,537	-7,447	15,250
16	1	0,009	0,857	-7,447	15,667
17	1	0,009	0,858	-7,447	16,068
18	1	0,009	0,853	-7,447	16,409
19	2	0,018	0,824	-7,447	17,044
20	2	0,018	0,778	-6,990	17,607
21	2	0,018	0,740	-6,990	18,097
22	2	0,018	0,709	-6,990	18,528
23	2	0,018	0,676	-6,990	18,930
24	2	0,018	0,632	-6,990	19,295
25	2	0,018	0,592	-6,990	19,636
26	3	0,054	0,553	-6,990	20,038
27	3	0,054	0,510	-6,990	20,486
28	3	0,054	0,513	-6,990	20,900
29	4	0,114	0,608	-6,990	21,305
30	4	0,114	0,673	-6,020	21,722
31	5	0,452	0,637	-6,020	22,128
32	5	0,452	0,586	-6,020	22,512
33	5	0,452	0,571	-6,020	22,877
34	7	6,330	0,616	-5,229	23,241
35	7	6,330	0,640	-5,229	23,616
36	11	9,947	0,597	-5,229	23,974
37	17	153,727	0,538	-5,229	24,312

Tabla 3-C.7e: Frecuencia de muestreo= 44,1 kHz. Bloque corto.

no.	FFT-lines	qthr	norm	SNR (db)	bval
0	1	4,532	0,952	-8,240	0,000
1	1	0,904	0,700	-8,240	1,723
2	1	0,029	0,681	-8,240	3,445
3	1	0,009	0,675	-8,240	5,057
4	1	0,009	0,667	-8,240	6,422

5	1	0,009	0,665	-8,240	7,609
6	1	0,009	0,664	-8,240	8,710
7	1	0,009	0,664	-8,240	9,662
8	1	0,009	0,664	-8,240	10,515
9	1	0,009	0,655	-8,240	11,293
10	1	0,009	0,616	-7,447	12,009
11	1	0,009	0,597	-7,447	12,625
12	1	0,009	0,578	-7,447	13,210
13	1	0,009	0,541	-7,447	13,748
14	1	0,009	0,575	-7,447	14,241
15	1	0,009	0,856	-7,447	14,695
16	1	0,009	0,846	-7,447	15,125
17	1	0,009	0,840	-7,447	15,508
18	1	0,009	0,822	-7,447	15,891
19	2	0,018	0,800	-7,447	16,537
20	2	0,018	0,753	-6,990	17,112
21	2	0,018	0,704	-6,990	17,620
22	2	0,018	0,674	-6,990	18,073
23	2	0,018	0,640	-6,990	18,470
24	2	0,018	0,609	-6,990	18,849
25	3	0,027	0,566	-6,990	19,271
26	3	0,027	0,535	-6,990	19,741
27	3	0,054	0,531	-6,990	20,177
28	3	0,054	0,615	-6,990	20,576
29	3	0,054	0,686	-6,990	20,950
30	4	0,114	0,650	-6,020	21,316
31	4	0,114	0,612	-6,020	21,699
32	5	0,452	0,567	-6,020	22,078
33	5	0,452	0,520	-6,020	22,438
34	5	0,452	0,513	-5,229	22,782
35	7	6,330	0,557	-5,229	23,133
36	7	6,330	0,584	-5,229	23,484
37	7	6,330	0,570	-5,229	23,828
38	19	171,813	0,578	-4,559	24,173

Tabla 3-C.7f: Frecuencia de muestreo= 32 kHz. Bloque corto.

no.	FFT-lines	gthr	norm	SNR (db)	bval
0	1	4,532	0,997	-8,240	0,000
1	1	0,904	0,893	-8,240	1,250
2	1	0,090	0,881	-8,240	2,500
3	1	0,029	0,873	-8,240	3,750
4	1	0,009	0,872	-8,240	4,909
5	1	0,009	0,871	-8,240	5,958
6	1	0,009	0,860	-8,240	6,857
7	1	0,009	0,839	-8,240	7,700
8	1	0,009	0,812	-8,240	8,500
9	1	0,009	0,784	-8,240	9,237
10	1	0,009	0,741	-7,447	9,895
11	1	0,009	0,697	-7,447	10,500
12	1	0,009	0,674	-7,447	11,083
13	1	0,009	0,651	-7,447	11,604
14	1	0,009	0,633	-7,447	12,107
15	1	0,009	0,611	-7,447	12,554
16	1	0,009	0,589	-7,447	13,000
17	1	0,009	0,575	-7,447	13,391
18	1	0,009	0,654	-7,447	13,781
19	2	0,018	0,724	-7,447	14,474
20	2	0,018	0,701	-6,990	15,096
21	2	0,018	0,673	-6,990	15,667
22	2	0,018	0,631	-6,990	16,177
23	2	0,018	0,592	-6,990	16,636

24	2	0,018	0,553	-6,990	17,057
25	2	0,018	0,510	-6,990	17,429
26	2	0,018	0,506	-6,990	17,786
27	3	0,027	0,562	-6,990	18,177
28	3	0,027	0,598	-6,990	18,597
29	3	0,027	0,589	-6,990	18,994
30	3	0,027	0,607	-6,020	19,352
31	3	0,027	0,620	-6,020	19,693
32	4	0,072	0,580	-6,020	20,066
33	4	0,072	0,532	-6,020	20,461
34	4	0,072	0,517	-5,229	20,841
35	5	0,143	0,517	-5,229	21,201
36	5	0,143	0,509	-5,229	21,549
37	6	0,172	0,506	-5,229	21,911
38	7	0,633	0,522	-4,559	22,275
39	7	0,633	0,531	-4,559	22,625
40	8	0,723	0,519	-3,980	22,971
41	10	9,043	0,512	-3,980	23,321

Tablas 3-C.8: Tablas para convertir el cálculo de las particiones del umbral a bandas Scalefactors

Tabla 3-C.8a: Frecuencia de muestreo= 48 kHz. Bloque largo.

no. sb	cbw	bu	bo	w1	w2
0	3	0	4	1,000	0,056
1	3	4	7	0,944	0,611
2	4	7	11	0,389	0,167
3	3	11	14	0,833	0,722
4	3	14	17	0,278	0,639
5	2	17	19	0,361	0,417
6	3	19	22	0,583	0,083
7	2	22	24	0,917	0,750
8	3	24	27	0,250	0,417
9	3	27	30	0,583	0,648
10	3	30	33	0,352	0,611
11	3	33	36	0,389	0,625
12	4	36	40	0,375	0,144
13	3	40	43	0,856	0,389
14	3	43	46	0,611	0,160
15	3	46	49	0,840	0,217
16	3	49	52	0,783	0,184
17	2	52	54	0,816	0,886
18	3	54	57	0,114	0,313
19	2	57	59	0,687	0,452
20	1	59	60	0,548	0,908

Tabla 3-C.8b: Frecuencia de muestreo= 44,1 kHz. Bloque largo.

no. sb	cbw	bu	bo	w1	w2
0	3	0	4	1,000	0,056
1	3	4	7	0,944	0,611
2	4	7	11	0,389	0,167
3	3	11	14	0,833	0,722
4	3	14	17	0,278	0,139
5	1	17	18	0,861	0,917
6	3	18	21	0,083	0,583
7	3	21	24	0,417	0,250
8	3	24	27	0,750	0,805
9	3	27	30	0,194	0,574
10	3	30	33	0,426	0,537
11	3	33	36	0,463	0,819
12	4	36	40	0,180	0,100

13	3	40	43	0,900	0,468
14	3	43	46	0,532	0,623
15	3	46	49	0,376	0,450
16	3	49	52	0,550	0,552
17	3	52	55	0,448	0,403
18	2	55	57	0,597	0,643
19	2	57	59	0,357	0,722
20	2	59	61	0,278	0,960

Tabla 3-C.8c: Frecuencia de muestreo= 32 kHz. Bloque largo.

no. sb	cbw	bu	bo	w1	w2
0	1	0	2	1,000	0,528
1	2	2	4	0,472	0,305
2	2	4	6	0,694	0,083
3	1	6	7	0,917	0,861
4	2	7	9	0,139	0,639
5	2	9	11	0,361	0,417
6	3	11	14	0,583	0,083
7	2	14	16	0,917	0,750
8	3	16	19	0,250	0,870
9	3	19	22	0,130	0,833
10	4	22	26	0,167	0,389
11	4	26	30	0,611	0,478
12	4	30	34	0,522	0,033
13	3	34	37	0,967	0,917
14	4	37	41	0,083	0,617
15	3	41	44	0,383	0,995
16	4	44	48	0,005	0,274
17	3	48	51	0,726	0,480
18	3	51	54	0,519	0,261
19	2	54	56	0,739	0,884
20	2	56	58	0,116	1,000

Tabla 3-C.8d: Frecuencia de muestreo= 48 kHz. Bloque corto.

no. sb	cbw	bu	bo	w1	w2
0	2	0	3	1,000	0,167
1	2	3	5	0,833	0,833
2	3	5	8	0,167	0,500
3	3	8	11	0,500	0,167
4	4	11	15	0,833	0,167
5	4	15	19	0,833	0,583
6	3	19	22	0,417	0,917
7	4	22	26	0,083	0,944
8	4	26	30	0,055	0,042
9	2	30	32	0,958	0,567
10	3	32	35	0,433	0,167
11	2	35	37	0,833	0,618

Tabla 3-C.8e: Frecuencia de muestreo= 44,1 kHz. Bloque corto.

no. sb	cbw	bu	bo	w1	w2
0	2	0	3	1,000	0,167
1	2	3	5	0,833	0,833
2	3	5	8	0,167	0,500
3	3	8	11	0,500	0,167
4	4	11	15	0,833	0,167
5	5	15	20	0,833	0,250
6	3	20	23	0,750	0,583

7	4	23	27	0,417	0,055
8	3	27	30	0,944	0,375
9	3	30	33	0,625	0,300
10	3	33	36	0,700	0,167
11	2	36	38	0,833	1,000

Tabla 3-C.8f: Frecuencia de muestreo= 32 kHz. Bloque corto.

	no. sb	cbw	bu	bo	w1	w2
0	2	0	3	1,000	0,167	
1	2	3	5	0,833	0,833	
2	3	5	8	0,167	0,500	
3	3	8	11	0,500	0,167	
4	4	11	15	0,833	0,167	
5	5	15	20	0,833	0,250	
6	4	20	24	0,750	0,250	
7	5	24	29	0,750	0,055	
8	4	29	33	0,944	0,375	
9	4	33	37	0,625	0,472	
10	3	37	40	0,528	0,937	
11	1	40	41	0,062	1,000	

[ANEXO 7]

TRANSFORMADA DISCRETA DEL COSENO

La DCT (Discrete Cosine Transform) es un caso especial de la transformada discreta de Fourier en la que se han eliminado los componentes senoidales de los coeficientes dejando una sola componente cosenoidal. Esto resulta en la práctica bastante fácil. Espejando las muestras de entrada respecto a las ordenadas y aplicando la transformada discreta de Fourier sobre el bloque de doble longitud que el original se obtiene un set de muestras de la DCT. El efecto del espejado de las muestras de entrada es transformar el bloque en una función par en la que los coeficientes con senos son todos ceros. Este efecto se puede visualizar en la figura 2. En la práctica, el cálculo de los componentes senoidales no se realiza. Otra ventaja es que duplicando la longitud del bloque (al espejarlo) se duplica la resolución en frecuencia por lo que se obtiene el doble de los coeficientes. En realidad, la DCT produce tantos coeficientes útiles como muestras de entrada. Por otra parte, cuando se realiza el proceso inverso de la transformada, el bloque espejado se descarta obteniéndose el original.[7]

En MPEG-1 layer III se utiliza la Transformada Discreta del Coseno Modificada (MDCT). Esta consiste en utilizar una DCT en bloque que se solapan con un 50% de solapamiento. Esto genera el doble de los coeficientes dado que cada muestra aparece en dos sub-bloques. Luego se sub-muestra para obtener un muestreo crítico, lo que resulta en un potencial solapamiento (aliasing) en el dominio frecuencial. [7]

De todas formas, el solapamiento producido al transformarse el bloque (vía MDCT) se puede cancelar al realizarse la transformación inversa (IMDCT), dado que los productos de alias de la segunda mitad de un bloque transformado son de igual

magnitud pero de polaridad inversa a los de la primera mitad del bloque contiguo. [7]
Esto puede verse en la figura 1:

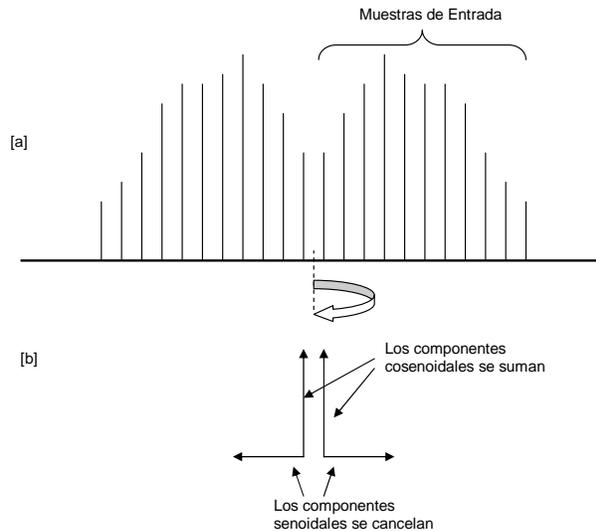


Figura 1. Gráfica de la transformada. Fuente: [7]

Las Transformadas

El proceso de convolución es demasiado extenso para representarlo sobre papel. Es mucho más fácil trabajar en el dominio de la frecuencia. Si una señal con un espectro, o un contenido de frecuencia pasa por un filtro de respuesta en frecuencia dado, el resultado será un espectro de salida que es simplemente el producto de ambos. Si se representan las respuestas en frecuencia en escalas logarítmicas (esto es, calibradas en decibelios), ambas pueden sumarse simplemente ya que la suma de logaritmos es lo mismo que su producto. La multiplicación de los espectros de las respuestas es un proceso mucho más simple que la convolución. [7]

Para poder pasar al dominio de la frecuencia o del espectro desde el dominio temporal o de la forma de onda, es necesario utilizar la transformada de Fourier o, en sistemas muestreados, la transformada discreta de Fourier (DFT). El análisis de Fourier sostiene que cualquier forma de onda puede ser reproducida

sumando un número arbitrario de ondas senoidales relacionadas armónicamente con distintas amplitudes y fases. [7]

La figura 2 muestra el modo de poder conseguir una onda cuadrada con armónicos. El espectro puede obtenerse trazando gráficamente la amplitud de los armónicos con relación a la frecuencia. Como puede verse, se obtiene un espectro que es una onda de caída, que pasa por el cero en todos los múltiplos pares de la fundamental. La forma del espectro es una curva senoidal x/x . Si una onda cuadrada tiene un espectro $(\text{sen } x)/x$, se deduce que un filtro con una respuesta impulsiva rectangular tiene un espectro $(\text{sen } x)/x$. [7]

Un filtro pasa-bajo tiene un espectro rectangular y éste tiene una respuesta impulsiva $(\text{sen } x)/x$. Estas características se conocen como par de transformada. En los pares de transformada, si un dominio tiene una forma del par, el otro dominio tendrá la otra forma. Así, una onda cuadrada tiene un espectro $(\text{sen } x)/x$ y un impulso $(\text{sen } x)/x$ tiene un espectro cuadrado. Un impulso de dominio temporal de una duración infinitamente corta presenta un espectro plano. Así, una forma de onda plana, sólo tiene cero en su espectro. Curiosamente, la transformada de una respuesta gaussiana sigue siendo gaussiana. La respuesta impulsiva de la óptica de un disco láser tiene una función $(\text{sen } x)^2/x^2$, y esto es responsable de la respuesta en frecuencia descendente triangular del lector láser. [7]

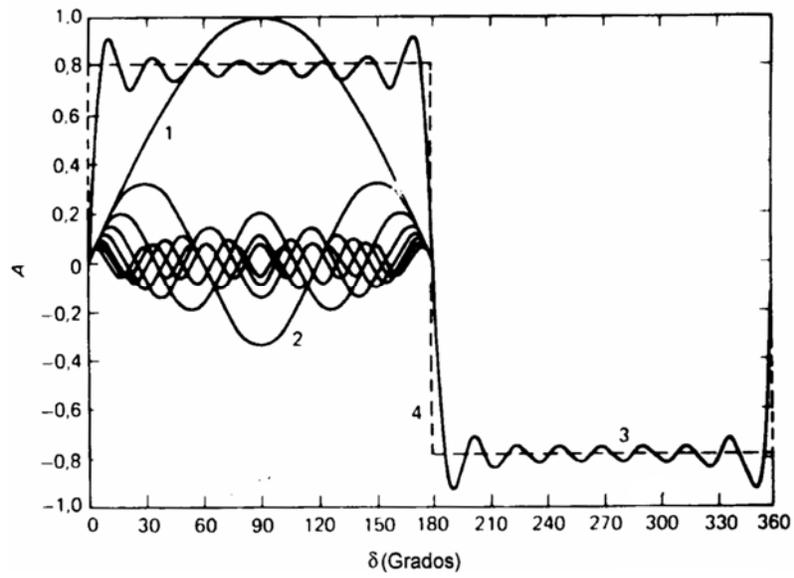


Figura 2. El análisis de Fourier de una onda cuadrada en fundamental y armónicos. A: amplitud, δ : fase de la onda fundamental en grados. 1, primer armónico (fundamental). 2, armónicos impares. 3, suma de armónicos. 4, onda cuadrada ideal. Fuente: [7]

[ANEXO 8]

CODIFICADORES HÍBRIDOS

En este trabajo se han expuesto brevemente los codificadores PCM o PAM de audio de la forma de onda, pero existen muchos otros como los codificadores híbridos.

En la codificación híbrida se combinan las técnicas de los codificadores de la forma de la onda con las de los modeladores de voz (vocoders) con el propósito de obtener una alta calidad de voz a bajas tasas de bits (inferiores a 8 kbps). En estos codificadores, las muestras de la señal de entrada se dividen en bloques de muestras (vectores) que son procesados como si fueran uno solo. Llevan a cabo una representación paramétrica de la señal de voz para tratar que la señal sintética se parezca lo más posible a la original.

También se les conoce como codificadores de análisis-por-síntesis. En el emisor se lleva a cabo un análisis que obtiene los parámetros de la señal para luego sintetizarla y conseguir el mayor parecido a la original.

Entre éstos están:

LD-CELP: Low Delay Code Excited Linear Prediction

CS-ACELP: Conjugate Structure Algebraic CELP

MP-MLQ: Multi Pulse Maximun Likelihood Quantization

DPCM: PCM Diferencial


```

% Obtiene la información del archivo WAV a codificar.
[PMA,Fs,bits] = wavread(archivo,1); % Sólo lee la primera muestra.
SIZ = wavread(archivo,'size'); % Obtiene el tamaño total del archivo.

% Si la frecuencia de muestreo no es 44100 Hz, se interrumpe la codificación.
if Fs ~= 44100
    fserror = ['Archivo WAV a ',int2str(Fs),' Hz no soportado.'];
    error(fserror)
end

% Si el archivo no es monofónico, se procesa el canal izquierdo.
if SIZ(2) == 2
    warning('El archivo WAV es estereofónico, se procesará el canal izquierdo.')
end

% Asigna la tasa de transferencia.
disp(' ')
disp('Las siguientes tasas de bits están disponibles (en Kbps):')
disp('96, 112, 128, 160, 192, 224, 256, ó 320')
tasa = input('Ingrese la tasa de bits deseada\n');
disp(' ')

% Inicialización de variables, de acuerdo con la tasa de bits escogida.
switch tasa
case 96
    bitrate_index = 7; % Índice de la tasa de bits.
    ajuste = 1; % Para ajustar el tamaño entero en bytes de la trama.
    ajgg = 11; % Para ajustar global_gain.
case 112
    bitrate_index = 8;
    ajuste = 2;
    ajgg = 10;
case 128
    bitrate_index = 9;
    ajuste = 3;
    ajgg = 9;
case 160
    bitrate_index = 10;
    ajuste = 1;
    ajgg = 9;
case 192
    bitrate_index = 11;
    ajuste = 3;
    ajgg = 8;
case 224
    bitrate_index = 12;
    ajuste = 1;
    ajgg = 8;
case 256
    bitrate_index = 13;
    ajuste = 3;
    ajgg = 7;
case 320
    bitrate_index = 14;
    ajuste = 3;
    ajgg = 7;
otherwise
    error(['La tasa de bits escogida (' , num2str(tasa),' Kbps) no es soportada.'])
end

% Asigna el tiempo de codificación.
max_rsize = fix(SIZ(1)/1152)*1152;
seg = max_rsize/44100;
tiempo = input(['¿Cuántos segundos desea codificar (mínimo 0.10449, máximo ',...
    num2str(seg),' seg.)?\n(ENTER para codificar todo el archivo).\n']);

% Obtiene el último valor del número de muestras del archivo WAV que sea múltiplo
% de 1152. En otras palabras, se obliga a que la cantidad de muestras del archivo
% WAV sea múltiplo de 1152. En este paso, se eliminan entre 1 y 1151 muestras PCM
% (son las últimas muestras del archivo WAV y, por lo tanto, no son procesadas).
if isempty(tiempo)
    rsize = max_rsize;
else
    rsize = fix(tiempo*44100/1152)*1152;
end

```

```

end
if rsize < 4608
    error('La cantidad de tiempo es insuficiente')
elseif rsize > max_rsize
    warning(['El archivo WAV sólo dura ', num2str(seg),...
            ' segundos, procesando todo el archivo.'])
    rsize = max_rsize;
end

disp('Okay, espere unos minutos...')

% Carga las tablas necesarias para el análisis psicoacústico.
[UA,MAP,UES] = Umbral_absoluto;
LBC = Limites_banda_critica;

% Carga las tablas necesarias para los códigos de Huffman.
% Tablas A y B para los cuádruplos count1, proporcionadas
% por el estándar ISO 11172-3.
tabla_0 = [1 4 4 5 4 6 5 6 4 5 5 6 6 6 6; % hlen.
           1 5 4 5 6 5 4 4 7 3 6 0 7 2 3 1]; % hcod en formato decimal.
tabla_1 = [ones(1,16)*4; % hlen.
           15:-1:0]; % hcod en formato decimal.

% Carga los coeficientes de la ventana del análisis (vector C).
load('Ci.mat')

% Tabla de las bandas del factor de escala (Para bloques largos y 44.1 kHz).
SFBT = [1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21;
        4 4 4 4 4 4 6 6 8 8 10 12 16 20 24 28 34 42 50 54 76;
        1 5 9 13 17 21 25 31 37 45 53 63 75 91 111 135 163 197 239 289 343;
        4 8 12 16 20 24 30 36 44 52 62 74 90 110 134 162 196 238 288 342 418];

% Tabla para scalefac_compress.
SFC = [0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4; % slen1.
       0 1 2 3 0 1 2 3 0 1 2 3 0 1 2 3 0 1 2 3; % slen2.
       0 1 2 3 5 5 6 7 8 8 9 10 4 11 12 13 14 14 14 15]; % scalefac_compress.

% Calcula el número promedio de bits disponibles para main_data, por gránulo
% (sin incluir el bit_reservoir).
% Inicializa las variables usadas para el control del formato.
init_mean_bits = fix(tasa*1000*(1152/44100)/2)-(84+ajuste);
bit_reservoir = 0;
padding_bit = 0;
rest = 0;

% Inicialización de variables para el formato del flujo de bits, de acuerdo con
% la tasa de bits escogida. Se inicializan:
% main_data_begin: Indica dónde empiezan los datos principales de la primera
trama.
% mbr: Máximo permitido para bit_reservoir.
% mmdb: Máximo permitido para main_data_begin.
% mdt: Cantidad máxima de bits para main_data, por trama.
main_data_begin = 0;
if tasa == 96 | 112 | 128 | 160
    mbr = init_mean_bits*2;
else
    mbr = 4088;
end
mmdb = mbr/8;
mdt = init_mean_bits*2;

% Ciclo Principal. Analiza grupos de 1152 muestras PCM.
% Cada grupo se convierte en una trama MP3.
for a = 1:1152:rsize,

    % Almacena los últimos 576 valores de la trama anterior (matriz S).
    U576 = S(19:36,:);
    S = []; % Reinicia la matriz del filtro subbanda para cada gránulo.
    MDCT = []; % Reinicia la matriz MDCT.
    system_const = 8; % Reinicia el valor de la constante del sistema.

    % Lee el valor de 1152 muestras PCM desde el archivo de audio WAV.
    ENT = wavread(archivo,[a a+1151]);

```

```

%%% ANÁLISIS PSICOACÚSTICO (Modelo I), Parte 1.

% Realiza un análisis FFT para calcular la densidad espectral de potencia.
F = Analisis_fft(ENT);

% Encuentra las componentes tonales (senoidales) y no-tonales (ruidosas)
% de la señal de audio.
[BS,LT,LNT] = Componentes_tonales(F,UA,MAP,LBC);

% Reduce las componentes enmascarantes: elimina todas las componentes
% enmascarantes irrelevantes.
[BSR,LTR,LNTR] = Reduccion(LT,LNT,BS,UA,MAP);

% Calcula los umbrales de enmascaramiento individual.
[UET,UENT] = Umbrales_enmasc_individual(F,LTR,LNTR,UA,MAP);

% Calcula el umbral de enmascaramiento global.
UEG = Umbral_enmasc_global(UES,UET,UENT);

% Determina el umbral de enmascaramiento mínimo en cada subbanda.
% Este paso no se realiza para la Capa III. Se puede usar para ampliar
% el codificador a las Capas I y/o II.
%UEM = Umbral_enmasc_minimo(UEG,MAP);

%%% FILTRADO SUBBANDA, Parte 1.

% Filtrado subbanda para el análisis. En la capa III se obtienen 36
% muestras subbanda consecutivas en el tiempo para cada una de las
% 32 subbandas (matriz S de 1152 valores).
for b = 1:32:1152,
    S = [S; Filtro_subbanda(ENT(b:b+31),C)];
end

%%% ANÁLISIS PSICOACÚSTICO (Modelo I), Parte 2.

% Los siguientes pasos no se realizan para la Capa III, por lo tanto, han sido
% desactivados. Pueden ser usados por cualquier persona interesada en el
proyecto,
% para ampliar el codificador a las Capas I y/o II.

% Calcula los factores de escala.
%FDE = Factores_escala(S);

% Determina el nivel de presión sonora en cada subbanda.
%NPS = Nivel_presion_sonora(F,FDE);

% Calcula la relación señal a máscara.
%SMR = NPS-UEM;

for gr = 1:2,

    %%% FILTRADO SUBBANDA, Parte 2.

    % Calcula la MDCT con 50% de solapamiento.
    for sb = 1:32,
        MDCT = [MDCT; Transf_discreta_coseno(S,U576,sb,gr)];
    end

    % Reduce el aliasing introducido por el 50% de solapamiento de la MDCT.
    XR = Aliasing(MDCT(1+576*(gr-1):576*gr));

    %%% CUANTIZACIÓN Y CODIFICACIÓN

    % Calcula los bits disponibles para cada gránulo (incluye el bit_reservoir).
    mean_bits = init_mean_bits + bit_reservoir;
    % Chequea la máxima cantidad de bits permitida para cada gránulo.
    if mean_bits > 4095
        mean_bits = 4095;
    end

    % Si no hay datos de audio, entonces devuelve ciertos valores por
    % defecto. Si se trata del primer gránulo, entonces se procesa el
    % segundo gránulo. Si está procesando el segundo gránulo, entonces
    % la ejecución continúa en la etapa de formato.

```

```

if XR == 0,
    scalefac_scale(gr) = 0;
    scalefac_l(:,gr) = zeros(21,1);
    IX(:,gr) = zeros(576,1);
    IS(:,gr) = zeros(576,1);
    count1table_select(gr) = 0;
    big_values(gr) = 0;
    region0_count(gr) = 0;
    region1_count(gr) = 0;
    table_select(gr,:) = zeros(1,3);
    preflag = 0;
    global_gain(gr) = 210 - 8; % Es decir, 210-system_const.
    scalefac_compress(gr) = 0;
    part2_3_length(gr) = 0;
    bit_reservoir = mean_bits;
    S = zeros(36,32);
    SHT = zeros(256,6,gr);
    slen1(gr) = 0;
    slen2(gr) = 0;
    rlb(gr,:) = zeros(1,3);
    fe(gr,:) = zeros(1,3);
    ff(gr,:) = zeros(1,3);
    count1(gr) = 0;
else
    % Inicialización de variables.
    scalefac_scale(gr) = 0;
    scalefac_l(:,gr) = zeros(21,1);

    % Calcula la distorsión permitida, según el modelo psicoacústico.
    xmin = Distorsion_permitida(SFBT,UA,UEG);

    % Ciclo interno. Chequea la tasa de bits. Si el vector IX requiere más
    % bits de los permitidos para ser codificado, repite el ciclo interno;
    % hasta que la cantidad disponible de bits sea suficiente para codificar
    % el espectro cuantizado (vector IX).
    [IX(:,gr),IS(:,gr),SHT(:, :,gr),overall_bitsum,count1table_select(gr),...
    big_values(gr),region0_count(gr),region1_count(gr),table_select(gr,:),...
        qqquant,quantanf,rlb(gr,:),count1(gr),fe(gr,:),ff(gr,:)] = ...
        Ciclo_interno(XR,system_const,SFBT);
    while mean_bits - 74 < overall_bitsum,
        system_const = system_const - 1;

    [IX(:,gr),IS(:,gr),SHT(:, :,gr),overall_bitsum,count1table_select(gr),...
    big_values(gr),region0_count(gr),region1_count(gr),table_select(gr,:),...
        qqquant,quantanf,rlb(gr,:),count1(gr),fe(gr,:),ff(gr,:)] = ...
        Ciclo_interno(XR,system_const,SFBT);
    end

    % Ciclo externo. Chequea la distorsión. Si la distorsión máxima
    % es excedida, vuelve a llamar el ciclo interno, de acuerdo con
    % los requerimientos del estándar internacional ISO/IEC 11172-3.
    % En este paso se incluyen las condiciones para terminar los ciclos;
    % si alguna de ellas se cumple, la ejecución de los ciclos se detiene
    % y los datos obtenidos hasta ese momento son usados para la etapa de
    % formato.
    [XFSF,preflag] = Ciclo_externo(SFBT,IX(:,gr),qqquant,quantanf,...
        scalefac_scale(gr));
    % Si alguna banda excede la distorsión permitida, entonces se chequean
    % las condiciones para terminar los ciclos. En el caso de que las
    % condiciones para terminación de los ciclos no se cumplan, entonces se
    % repite el ciclo interno.
    while length(find(xmin < XFSF)) > 0,
        % Chequea si todas las bandas del factor de escala ya han sido
amplificadas,
        % en cuyo caso se terminan los ciclos.
        if length(find(scalefac_l(:,gr) ~= 0)) == 21
            break
        end
        % Chequea el máximo de los factores de escala, teniendo en cuenta
        % el campo scalefac_scale.
        if (max(scalefac_l(1:11,gr))==15 | max(scalefac_l(12:21,gr))==7)...
            & scalefac_scale(gr)== 0

```

```

        % scalefac_scale se pone en '1' y scalefac_l se pone en '0'.
        scalefac_l(:,gr) = zeros(21,1);
        scalefac_scale(gr) = 1;
    end
    if (max(scalefac_l(1:11,gr))==15 | max(scalefac_l(12:21,gr))==7) ...
        & scalefac_scale(gr)==1
        % Se terminan los ciclos.
        break
    end
    % Si no se cumple ninguna condición para la terminación de los ciclos,
    % entonces se repite el ciclo interno.

[IX(:,gr),IS(:,gr),SHT(:,gr),overall_bitsum,count1table_select(gr),...
big_values(gr),region0_count(gr),region1_count(gr),table_select(gr,:),...
qquant,quantanf,rlb(gr,:),count1(gr),fe(gr,:),ff(gr,:)] = ...
Ciclo_interno(XR,system_const,SFBT);
while mean_bits - 74 < overall_bitsum,
    system_const = system_const - 1;

[IX(:,gr),IS(:,gr),SHT(:,gr),overall_bitsum,count1table_select(gr),...
big_values(gr),region0_count(gr),region1_count(gr),...
table_select(gr,:),qquant,quantanf,rlb(gr,:),count1(gr),...
fe(gr,:),ff(gr,:)] = Ciclo_interno(XR,system_const,SFBT);
end
[XFSF,preflag] = Ciclo_externo(SFBT,IX(:,gr),qquant,quantanf,...
scalefac_scale(gr));
end

% Calcula la ganancia global del sistema.
% Primero, calcula el vector IX sin realizar el redondeo (especialmente
% usado sólo para este cálculo).
% IXGG: IX para global_gain.
IXGG = (abs(XR)/(2^((qquant+quantanf)/4))) .^ 0.75 - 0.0946;
% Después, calcula la información del intervalo de cuantización (vector
iic)
% para cada una de las 576 muestras de audio del gránulo.
iic = zeros(576,1);
ifqstep = 2^(0.5*(1+scalefac_scale(gr)));
for h = SFBT,
    for q = h(3):h(4),
        iic(q) =
4*log2(abs(XR(q))*ifqstep^scalefac_l(h(1),gr)/(abs(IXGG(q)).^(4/3)));
    end
end

iic(419:576)=4.*log2(abs(XR(419:576)).*ifqstep./abs(IXGG(419:576)).^(4/3));
% Por último, calcula global_gain.
global_gain(gr) = round(mean(iic))+210-system_const-qquant-ajgg;

% Chequea el valor de global_gain (se debe tener en cuenta que este campo
% se escribe en la información secundaria, y consume 8 bits).
if global_gain(gr) < 0
    global_gain(gr) = 0;
end
if global_gain(gr) > 255
    global_gain(gr) = 255;
end

% Calcula la cantidad de bits necesaria para codificar los
% factores de escala.
if max(scalefac_l(1:11,gr)) ~= 0
    slen1(gr) = fix(log2(max(scalefac_l(1:11,gr))))+1;
else
    slen1(gr) = 0;
end
if max(scalefac_l(12:21,gr)) ~= 0
    slen2(gr) = fix(log2(max(scalefac_l(12:21,gr))))+1;
else
    slen2(gr) = 0;
end
scalefac_compress(gr) = SFC(3,4*slen1(gr)+slen2(gr)+1);

```

```

    % Corrige los valores de slen1 y slen2, para evitar una escritura
    % incorrecta del flujo de bits, ocasionada por valores de slen1 y
    % slen2 que no fueron incluidos en la tabla de scalefac_compress
    % proporcionada por el estándar ISO/IEC 11172-3.
    if scalefac_compress(gr) == 5
        slen1(gr) = 1;
        slen2(gr) = 1;
    end
    if scalefac_compress(gr) == 8
        slen1(gr) = 2;
        slen2(gr) = 1;
    end
    if scalefac_compress(gr) == 14
        slen1(gr) = 4;
        slen2(gr) = 2;
    end

    % Calcula la cantidad total de bits que usa el espectro cuantizado
    % (vector IX más los factores de escala), y se determina cuántos bits
    % quedan disponibles para el próximo gránulo.
    part2_3_length(gr) = overall_bitsum + slen1(gr)*11 + slen2(gr)*10;
    bit_reservoir = mean_bits - part2_3_length(gr);
end

end

%%% FORMATO DEL FLUJO DE BITS.

% Para cada trama, determina si es necesario activar padding_bit con el fin de
% ajustar la tasa de bits promedio. Antes, se calcula la cantidad de tramas.
frames = (a+1151)/1152;
if frames > 1
    dif = rem(144000*tasa,Fs);
    rest = rest - dif;
    if rest < 0
        padding_bit = 1;
        rest = rest + Fs;
    else
        padding_bit = 0;
    end
end

% Escribe el encabezado como una cadena binaria de texto (ASCII).
Encabezado(padding_bit,bitrate_index);

% Escribe la información secundaria como una cadena binaria de texto (ASCII).
Info_secundaria(main_data_begin,part2_3_length,big_values,...
    global_gain,scalefac_compress,table_select,region0_count,region1_count,...
    preflag,scalefac_scale,count1table_select);

% Escribe los códigos de Huffman y los factores de escala como una cadena
% binaria de texto (ASCII).
Datos_principales(IX,IS,SHT,slen1,slen2,big_values,table_select,rlb,...

fe,ff,tabla_0,tabla_1,count1,count1table_select,main_data_begin,padding_bit,mdt);

% Chequea el valor de bit_reservoir, de tal manera que nunca permite que
% main_data esté repartido entre más de 2 tramas. Además, redondea el valor
% de main_data_begin, de tal manera que sea múltiplo de 8 (main_data_begin
% se especifica en bytes).
if tasa == 320
    main_data_begin = 0;
    bit_reservoir = 0;
else
    if bit_reservoir > mbr
        main_data_begin = mmdb;
        bit_reservoir = mbr;
    else
        main_data_begin = fix(bit_reservoir/8);
        bit_reservoir = main_data_begin*8;
    end
end
end
end

```

```

% Transforma la cadena binaria de texto ASCII (bin_str) en un vector de valores
% binarios (unos y ceros) en formato decimal, y escribe este vector como un
% archivo
% binario, de acuerdo con los requerimientos del estándar ISO/IEC 11172-3 (usando
% el formato de máquina big-endian).
fid = fopen([archivo '.mp3'],'wb','b');
fwrite(fid,bin_str-48,'ubit1');
fclose(fid);

% Indica que el archivo MP3 ya ha sido creado.
disp([sprintf('\n'),'El archivo ',archivo,'.mp3 ha sido terminado.'])

clear Fs IXGG MAP SIZ PMA X a b bits dif fid gr h iic q rest ifqstep mdt
clear rsize sb v bit_reservoir bitrate_index init_mean_bits mean_bits mbr mmdb
ajgg
clear archivo ENT LBC SFBT SFC SHT U576 UA UES tabla_0 tabla_1 C ans bin_str tasa
clear tiempo seg max_rsize ajuste

```

```

function MDCTB = Aliasing(MDCT)
%ALIASING Reduce el aliasing introducido por la MDCT.
%
% MDCTB=ALIASING(MDCT)
% Retorna en el vector MDCTB de 576 valores el resultado de la reducción del
% aliasing introducido por el 50% de solapamiento usado al aplicar la MDCT o
% transformada discreta del coseno modificada.
%
% El vector MDCT contiene los 576 valores de la transformación obtenida a
% través de TRANSF_DISCRETA_COSENO.
%
%
MDCTB = [];

% Constantes para el cálculo mariposa, proporcionadas por el estándar
% ISO/IEC 11172-3 (vectores c, cs y ca).
% c = [-0.6; -0.535; -0.33; -0.185; -0.095; -0.041; -0.0142; -0.0037];
% cs = 1 ./ sqrt(1+c.^2);
% ca = c ./ sqrt(1+c.^2);
cs = [0.85749292571254; 0.88174199731771; 0.94962864910273; 0.98331459249179;
0.99551781606759; 0.99916055817815; 0.99989919524445; 0.99999315507028];
ca = [-0.51449575542753; -0.47173196856497; -0.31337745420390; -0.18191319961098;
-0.09457419252642; -0.04096558288530; -0.01419856857247; -0.00369997467376];

% Reducción del aliasing o solapamiento a través de 8 cálculos mariposa. La
% reducción se hace en grupos de 18 líneas de frecuencia (a las 16 primeras se
% les hace reducción y a las 2 últimas no), empezando en la muestra 11.
MDCTcs = zeros(8,2);
MDCTca = zeros(8,2);
for i = 11:18:551,

    % Las líneas de frecuencia se multiplican por cs y ca.
    MDCTcs = [[MDCT(i+7:-1:i) .* cs] [MDCT(i+8:i+15) .* cs]];
    MDCTca = [[MDCT(i+7:-1:i) .* ca] [MDCT(i+8:i+15) .* ca]];

    % El concepto mariposa involucra sumar y restar las líneas de frecuencia en
    % pares, entregando dos nuevas líneas de frecuencia por cada par. En total, se
    % reemplazan 16 líneas de frecuencia (vector MDCTBparcial).
    MDCTBparcial = [MDCTcs(8:-1:1,1)+MDCTca(8:-1:1,2); MDCTcs(:,2)-MDCTca(:,1)];

    % Concatenación del resultado anterior con los dos valores de MDCT a los que
    % no se les aplica la reducción del aliasing.
    MDCTB = [MDCTB; MDCTBparcial; MDCT(i+16:i+17)];

end

% Los primeros 10 y los últimos 8 valores de MDCT (que no son incluidos en el
% cálculo mariposa) se conservan para MDCTB.
MDCTB = [MDCT(1:10); MDCTB; MDCT(569:576)];

```

```

function F = Analisis_fft(ENT)
%ANALISIS_FFT Análisis Transformada Rápida de Fourier.
%
%   F=ANALISIS_FFT(ENT)
%   Calcula el espectro auditivo usando la Transformada rápida de Fourier.
%   El espectro (vector F) se expresa en decibeles, y sólo tiene 512 valores
%   debido a la simetría de la FFT. La cantidad de puntos de la transformada
%   es 1024 (Capa II). La variable de entrada (vector ENT) equivale a
%   1152 muestras de audio PCM. Con el fin de usar la FFT de 1024 puntos para
%   las 1152 muestras, se escogen las 1024 muestras centrales de las 1152.
%
%   A 1024 muestras centrales se les aplica una ventana de Hanning convencional
%   antes de calcular su FFT, para suavizar los extremos del intervalo de señal.

% Escoge las 1024 muestras centrales del vector ENT y las almacena en el vector s.
s = ENT(65:1088);

% Sólo se realizan los cálculos si el vector s no es un vector de ceros. De lo
% contrario, el espectro auditivo (vector F) es un vector de -INF.
if s ~= 0
    % Calcula la ventana de Hanning de 1024 puntos (vector h).
    h = sqrt(8/3) * hanning(1024);

    % Obtiene la densidad espectral de potencia (vector F) a través de la FFT.
    % F se reduce a la mitad de componentes, por la simetría de la FFT.
    F = max(20*log10(abs(fft(s .* h))/1024),-200);
    F = F(1:512);
else
    F = zeros(512,1)-200; % -200 dB corresponde a -INF.
end

end

function [XFSF,preflag] = Ciclo_externo(SFBT,IX,qquant,quantanf,scalefac_scale)
%CICLO_EXTERNO Realiza el ciclo para el control de la distorsión.
%
%   [XFSF,PREFLAG]=CICLO_EXTERNO(SFBT,IX,QQUANT,QUANTANF,SCALEFAC_SCALE)
%   XFSF es el vector de la distorsión en cada una de las bandas del factor de
%   escala. PREFLAG indica si la opción de preénfasis ha sido utilizada o no.
%
%   SFBT es la matriz de las bandas del factor de escala, obtenida en WAV2MP3.
%   IX es el vector de valores espectrales cuantizados, obtenido con
CICLO_INTERNO.
%   QQUANT y QUANTANF son los valores usados para el intervalo de cuantización,
%   obtenidos con CICLO_INTERNO. SCALEFAC_SCALE es el factor logarítmico de
%   cuantización para los factores de escala, obtenido en WAV2MP3.
%
%
global XR gr xmin scalefac_l

% La opción de preénfasis no se implementa.
preflag = 0;

% Calcula la distorsión en las bandas del factor de escala (vector XFSF), y luego
% amplifica las bandas del factor de escala que exceden el umbral de
enmascaramiento.
ifqstep = 2^(0.5*(1+scalefac_scale));
XFSF = zeros(1,21);
for sb = SFBT,
    for i = sb(3):sb(4),
        XFSF(sb(1)) = XFSF(sb(1))+(abs(XR(i))-
IX(i)^(4/3)*2^((qquant+quantanf)/4))^2/sb(2);
    end
    if xmin(sb(1)) < XFSF(sb(1))
        scalefac_l(sb(1),gr) = scalefac_l(sb(1),gr) + 1;
        xmin(sb(1)) = xmin(sb(1))*ifqstep^(2*scalefac_l(sb(1),gr));
        for i = sb(3):sb(4),
            XR(i) = XR(i)*ifqstep^scalefac_l(sb(1),gr);
        end
    end
end
end
end

```

```

function [IX,IS,SHT,overall_bitsum,count1table_select,big_values,region0_count,
...
    region1_count,table_select,qquant,quantanf,rlb,count1,fe,ff] = ...
    Ciclo_interno(XR,system_const,SFBT)
% CICLO_INTERNO Realiza el ciclo para el control de la tasa de bits.
%
% [IX,IS,SHT,OVERALL_BITSUM,COUNT1TABLE_SELECT,BIG_VALUES,REGION0_COUNT,...
% REGION1_COUNT,TABLE_SELECT,QQUANT,QUANTANF,RLB,COUNT1,FE,FF]=...
% CICLO_INTERNO(XR,SYSTEM_CONST,SFBT)
%
% IX es el vector de valores espectrales cuantizados (sin tener en cuenta el
% signo de XR). IS es el vector de valores espectrales cuantizados (teniendo
% en cuenta el signo de XR). SHT es la matriz que contiene las 3 tablas de
% Huffman seleccionadas para codificar las 3 subregiones de la región
% big_values. OVERALL_BITSUM representa la cantidad de bits que se usan para
% codificar los valores cuantizados. COUNT1TABLE_SELECT indica la tabla de
% Huffman escogida para codificar los cuádruplos de valores pertenecientes a
% la región count1. BIG_VALUES es la cantidad de valores espectrales, contados
% por pares, ubicados en las bajas frecuencias. REGION0_COUNT es la cantidad de
% bandas del factor de escala (disminuidas en 1) incluidas en region0.
% REGION1_COUNT es la cantidad de bandas del factor de escala (disminuidas
% en 1) incluidas en region1. TABLE_SELECT es el vector que incluye el número
% de las tablas de Huffman usadas para codificar region0, region1 y region2.
% QQUANT y QUANTANF son los valores usados para el intervalo de cuantización.
% El vector RLB indica cuántos bits deben usarse para codificar cada uno de los
% valores adicionales en cada una de las 3 subregiones en que se divide
big_values
% (el valor adicional es necesario si el valor máximo de la subregión es mayor
% que 15). COUNT1 es la cantidad de cuádruplos incluidos en la región count1
% (región intermedia del vector IX). El vector FE indica el principio de cada
% una de las 3 subregiones en que se divide big_values. El vector FF indica el
% final de cada una de las 3 subregiones en que se divide big_values.
%
% XR es el vector de muestras subbanda obtenido originalmente con ALIASING, y
% más tarde modificado en CICLO_EXTERNO. SYSTEM_CONST es la constante del
sistema,
% obtenida en WAV2MP3. SFBT es la matriz que proporciona las bandas del factor
de
% escala, obtenida en WAV2MP3.
%
%
% Inicialización de variables.
sfm = exp(sum(log(XR.^2))/576)/(sum(XR.^2)/576); % Calcula la planura espectral.
quantanf = system_const*log(sfm); % Selección del intervalo de cuantización.
qquant = 0; % Selección del intervalo de cuantización.
% Cuantiza los valores del vector XR. En IX se almacenan los valores
% absolutos y en IS se almacenan los valores signados.
IX = round((abs(XR)/(2^((qquant+quantanf)/4))) .^ 0.75 - 0.0946);
IS = sign(XR).*IX;
% Cuando todos los valores del vector IX son cero, se termina el Ciclo Interno.
if IX == 0
    count1table_select = 0;
    big_values = 0;
    region0_count = 0;
    region1_count = 0;
    table_select = zeros(1,3);
    overall_bitsum = 0;
    SHT = zeros(256,6);
    rlb = zeros(1,3);
    count1 = 0;
    fe = ones(1,3);
    ff = ones(1,3);
    return
else
    % Limita a (8191 + 15) el valor máximo de los componentes del vector IX.
    while max(IX) > 8206,
        qquant = qquant+1;
        IX = round((abs(XR)/(2^((qquant+quantanf)/4))) .^ 0.75 - 0.0946);
        IS = sign(XR).*IX;

```

```

end

% Calcula el número de pares de ceros (rzero) en el extremo superior
% del vector IX.
rzero = 0;
for i = 576:-1:1,
    if IX(i) == 0
        rzero = rzero + 0.5;
    else
        break
    end
end
rzero = fix(rzero);

% Calcula el número de cuádruplos de valores (count1) en el vector IX cuyo
valor
% absoluto no es mayor que uno, y que siguen luego de los valores rzero (en la
% parte intermedia del vector IX).
count1 = 0;
for i = 576-rzero*2:-1:1,
    if IX(i) == 0 | IX(i) == 1
        count1 = count1+0.25;
    else
        break
    end
end
count1 = fix(count1);

% Calcula el número de pares de valores (big_values) en el extremo inferior
% del vector IX.
big_values = (576-rzero*2-count1*4)/2;

% Almacena en bitsum_count1 el número de bits necesario para codificar los
% valores count1. Además, determina la tabla de Huffman más apropiada para
% dicha codificación (count1table_select).
count1table_0 = [1 4 4 5 4 6 5 6 4 5 5 6 5 6 6];
% Calcula los bits que se consumen con la tabla B.
bitsum_table1 = count1*4;
% Calcula los bits que se consumen con la tabla A.
bitsum_table0 = 0;
for k = big_values*2+1:4:big_values*2+count1*4,
    bitsum_table0 =
bitsum_table0+count1table_0(8*IX(k)+4*IX(k+1)+2*IX(k+2)+IX(k+3)+1);
end
% Calcula cuántos bits de signo se usan para la región count1.
count1_signbits = length(find(IX(big_values*2+1:576-rzero*2)));
% Calcula la cantidad total de bits que consume la región count1,
% usando la tabla de Huffman que usa menos bits para codificar los
% cuádruplos count1.
bitsum_count1 = min(bitsum_table0,bitsum_table1) + count1_signbits;
if bitsum_table0 < bitsum_table1
    % Escoge la tabla A, proporcionada por el estándar ISO 11172-3.
    count1table_select = 0;
else
    % Escoge la tabla B, proporcionada por el estándar ISO 11172-3.
    count1table_select = 1;
end

% Divide la región de los valores big_values en 3 subregiones, cuyos límites
% deben coincidir con los límites de las bandas del factor de escala incluidas
% en el tamaño de dicha región. Los valores de region0_count y region1_count
% obtenidos aquí, no están regidos por los requerimientos del estándar ISO
11172-3.
if big_values == 0
    bvscfb = 0;
else
    for i = SFBT,
        if big_values*2 <= i(4)
            break
        end
    end
    bvscfb = i(1);
end
region0_count = fix(bvscfb/2);

```

```

region2_count = fix(bvscfb/4);
region1_count = bvscfb - region0_count - region2_count;

% Encuentra el máximo valor cuantizado de cada subregión (max_region), el
% cual sirve para calcular linbits (vector rlb). Estos dos valores permiten
% seleccionar las tablas de Huffman apropiadas (matrices HCTN y HCT) para
% codificar cada una de las subregiones de big_values. En este paso, se puede
% dar la posibilidad de hasta tres (3) tablas de Huffman por subregión, para un
% caso máximo de nueve (9) tablas seleccionadas para big_values.
Posteriormente,
% se realiza la selección de la tabla más apropiada para cada subregión.
% Adicionalmente, se inicializa apropiadamente la variable bitsum_table, de
% acuerdo con la cantidad de subregiones en que se divide big_values.
rlb = zeros(1,3);
HCTN = zeros(3);
HCT = zeros(256,18);
if bvscfb > 3
    % Caso 1. La región big_values se ha dividido en tres (3) subregiones.
    % Máximo nueve (9), mínimo tres (3) tablas para big_values.
    fe = [1,SFBT(4,region0_count)+1,SFBT(4,region0_count+region1_count)+1];
    ff =
[SFBT(4,region0_count),SFBT(4,region0_count+region1_count),big_values*2];
    max_region0 = max(IX(1:ff(1)));
    max_region1 = max(IX(fe(2):ff(2)));
    max_region2 = max(IX(fe(3):ff(3)));
    if max_region0 > 15
        rlb(1) = ceil(log2(max_region0 - 14)); % Cálculo de linbits, region0.
    end
    if max_region1 > 15
        rlb(2) = ceil(log2(max_region1 - 14)); % Cálculo de linbits, region1.
    end
    if max_region2 > 15
        rlb(3) = ceil(log2(max_region2 - 14)); % Cálculo de linbits, region2.
    end
    % Selección de la(s) tabla(s) de Huffman por subregión.
    [HCTN(1,:),HCT(:,1:6)] = Huffman(rlb(1),max_region0);
    [HCTN(2,:),HCT(:,7:12)] = Huffman(rlb(2),max_region1);
    [HCTN(3,:),HCT(:,13:18)] = Huffman(rlb(3),max_region2);
    bitsum_table = ones(3)*inf;
elseif bvscfb == 1 | bvscfb == 2
    % Caso 2. La región big_values se ha dividido en una (1) subregión.
    % Máximo tres (3), mínimo una (1) tabla para big_values.
    fe = [1, 1, 1];
    ff = [1, big_values*2, 1];
    max_region1 = max(IX(1:ff(2)));
    if max_region1 > 15
        rlb(2) = ceil(log2(max_region1 - 14)); % Cálculo de linbits, region1.
    end
    % Selección de la(s) tabla(s) de Huffman por subregión.
    [HCTN(2,:),HCT(:,7:12)] = Huffman(rlb(2),max_region1);
    bitsum_table = [zeros(1,3); ones(1,3)*inf; zeros(1,3)];
elseif bvscfb == 3
    % Caso 3. La región big_values se ha dividido en dos (2) subregiones.
    % Máximo seis (6), mínimo dos (2) tablas para big_values.
    fe = [1, 5, 1];
    ff = [4, big_values*2, 1];
    max_region0 = max(IX(1:4));
    max_region1 = max(IX(5:ff(2)));
    if max_region0 > 15
        rlb(1) = ceil(log2(max_region0 - 14)); % Cálculo de linbits, region0.
    end
    if max_region1 > 15
        rlb(2) = ceil(log2(max_region1 - 14)); % Cálculo de linbits, region1.
    end
    % Selección de la(s) tabla(s) de Huffman por subregión.
    [HCTN(1,:),HCT(:,1:6)] = Huffman(rlb(1),max_region0);
    [HCTN(2,:),HCT(:,7:12)] = Huffman(rlb(2),max_region1);
    bitsum_table = [ones(2,3)*inf; zeros(1,3)];
elseif bvscfb == 0
    % Caso 4. La región big_values no existe, y por lo tanto, no hay
subdivisión.
    fe = [1, 1, 1];
    ff = [1, 1, 1];
    bitsum_table = zeros(3);

```

```

end

% Ajusta los valores de region0_count y region1_count, de acuerdo con los
% requerimientos del estándar ISO 11172-3.
if region0_count ~= 0
    region0_count = region0_count - 1;
end
if region1_count ~= 0
    region1_count = region1_count - 1;
end

% A partir de las tablas de Huffman seleccionadas anteriormente (mínimo 1,
% máximo 3 tablas para cada subregión), escoge aquella que usa menos bits
% en la codificación (table_select y la matriz SHT), y calcula los bits
% necesarios para codificar los valores en cada subregión. Por último,
% calcula overall_bitsum, que es la cantidad total de bits que se usan
% para codificar los valores cuantizados. Primero, se cargan los valores
% máximos de las tablas de Huffman.
maxth = [1 2 3 3 0 4 4 6 6 6 8 8 8 ones(1,19)*16];
% Después, se realiza el conteo de bits con todas las tablas de Huffman, y
% almacena el resultado en bitsum_table.
for j = 1:3,
    for t = 1:3,
        if HCTN(j,t)
            bs = 0;
            for k = fe(j):2:ff(j),
                bs = bs + HCT(maxth(HCTN(j,t)+1)*min(15,IX(k))+ ...
                    min(15,IX(k+1))+1,2*t+6*j-7);
            end
            bitsum_table(j,t) = bs;
        end
    end
    % Selecciona la tabla de Huffman que consume menos bits.
    [min_bitsum,min_bitsum_index] = min(bitsum_table(j,:));
    % Guarda el número de la tabla, en el vector table_select.
    table_select(j) = HCTN(j,min_bitsum_index);
    % Almacena en bitsum_region, la cantidad de bits que consume la subregión.
    bitsum_region(j) = min_bitsum;
    % Calcula la cantidad de bits que consumen los valores mayores que 14.
    rlbsum = length(find(IX(fe(j):ff(j)) > 14))*rlb(j);
    region_linbits_sum(j) = rlbsum;
    % Almacena en la matriz SHT, la tabla de Huffman seleccionada.
    SHT(:,2*j-1:2*j) = HCT(:,2*min_bitsum_index+6*j-7:2*min_bitsum_index+6*j-6);
end
% Calcula cuántos bits de signo se usan para la región big_values.
big_values_signbits = length(find(IX(1:big_values*2)));

% Calcula la cantidad de bits que consume el espectro cuantizado (vector IX).
overall_bitsum = sum(bitsum_region) + sum(region_linbits_sum) + ...
    big_values_signbits + bitsum_count1;
end

```

```

function [BS,LT,LNT] = Componentes_tonales(F,UA,MAP,LBC)
%COMPONENTES_TONALES Encuentra las componentes tonales y no-tonales de la señal
% de audio.
%
% [BS,LT,LNT]=COMPONENTES_TONALES(F,UA,MAP,LBC)
% La matriz LT lista las componentes tonales y la matriz LNT lista las
% componentes no-tonales; ambas matrices se componen de dos columnas, la
% primera proporciona el índice de la línea de frecuencia y la segunda, el
% nivel de presión sonora de esa línea. El vector BS proporciona las banderas
% para cada una de las 512 líneas de frecuencia, tal que:
% - Componente no examinada = 0.
% - Componente tonal = 1.
% - Componente no-tonal = 2.
% - Componente irrelevante = 3.
%
% F es el vector de densidad espectral de potencia normalizada, obtenido con
% ANALISIS_FFT. UA es la matriz de umbral absoluto y MAP es el vector de mapeo
% entre las líneas de frecuencia y su índice para la matriz UA; ambos obtenidos
% con UMBRAL_ABSOLUTO. LBC es la matriz que contiene los límites de las bandas

```

```

% críticas, obtenida con LIMITES_BANDA_CRITICA.

% Inicializa el vector de banderas BS para las 512 líneas de frecuencia.
BS = zeros(512,1);

% Determina la lista de los máximos locales (matriz LML) para las líneas de
% frecuencia. La primera columna de LML son los índices, y la segunda corresponde
% al nivel de presión sonora. El análisis sólo es necesario hacerlo para las
% líneas de frecuencia con índice entre 3 y 500, de acuerdo con los requerimientos
% del estándar ISO 11172-3, página 112 (mirar los intervalos de k usados para
% generar los intervalos de J, más adelante en este mismo archivo; éstos obligan a
% la exclusión de índices k menores que 2 y mayores que 501).
LML = [];
c = 1;
for k = 3:500,
    if (F(k)>F(k-1) & F(k)>=F(k+1))
        LML(c,1) = k;
        LML(c,2) = F(k);
        c = c+1;
    end
end

% Determina cuáles de los máximos locales son componentes tonales y calcula su
% nivel de presión sonora (columnas de la matriz LT). Además, asigna para cada
% una de las líneas de frecuencia analizadas su respectivo valor de bandera en
% el vector BS.
LT = [];
c = 1;
if not isempty(LML)
    for i = 1:length(LML(:,1)),
        k = LML(i,1);
        tonal = 1;

        % Determina el intervalo J que indica cuántas frecuencias adyacentes
        % deben ser examinadas, de acuerdo con la posición de la línea de
        % frecuencia, o sea, de acuerdo con el índice k.
        if (2<k & k<63)
            J = [-2,2];
        elseif (63<=k & k<127)
            J = [-3,-2,2,3];
        elseif (127<=k & k<255)
            J = [-6:-2,2:6];
        elseif (255<=k & k<=500)
            J = [-12:-2,2:12];
        else
            tonal = 0;
        end

        % Examina las frecuencias adyacentes de acuerdo con el intervalo definido
        % por J, y determina si el máximo local es definitivamente una componente
        % tonal.
        for j = J,
            tonal = tonal & (F(k)-F(k+j) >= 7);
        end

        % Si F(k) es realmente una componente tonal, entonces lo siguiente se lista:
        % - Índice k de la línea de frecuencia.
        % - Nivel de presión sonora.
        % - Bandera tonal.
        if tonal
            LT(c,1) = k;
            LT(c,2) = 10*log10(10^(F(k-1)/10)+10^(F(k)/10)+10^(F(k+1)/10));
            BS(k) = 1; % Bandera de componente tonal = 1.
            for j = [J,-1,1],
                BS(k+j) = 3; % Bandera de componente irrelevante = 3.
            end
            c = c+1;
        end
    end
end
end
end

```

```

function Info_secundaria(main_data_begin,part2_3_length,big_values,...
    global_gain,scalefac_compress,table_select,region0_count,region1_count,...
    preflag,scalefac_scale,count1table_select)

%INFO_SECUNDARIA Escritura de la información secundaria.
%
%   INFO_SECUNDARIA(MAIN_DATA_BEGIN,PART2_3_LENGTH,BIG_VALUES,...
%   GLOBAL_GAIN,SCALEFAC_COMPRESS,TABLE_SELECT,REGION0_COUNT,REGION1_COUNT,...
%   PREFLAG,SCALEFAC_SCALE,COUNT1TABLE_SELECT)
%
%   Realiza la escritura de la información secundaria para cada trama del
%   MP3
%
%   MAIN_DATA_BEGIN es el puntero que indica el comienzo de la información de
%   audio (main_data) de cada trama, obtenido en WAV2MP3. PART2_3_LENGTH es la
%   cantidad de bits usados para codificar los valores cuantizados, obtenido
%   en WAV2MP3. BIG_VALUES es la cantidad de valores espectrales, contados por
%   pares, ubicados en las bajas frecuencias, obtenido con CICLO_INTERNO.
%   GLOBAL_GAIN es la información del intervalo de cuantización, obtenida en
%   WAV2MP3. SCALEFAC_COMPRESS es un índice a una tabla que proporciona la
%   cantidad de bits que consume la codificación de los factores de escala,
%   obtenido en WAV2MP3. TABLE_SELECT es el vector que incluye el número de las
%   tablas de Huffman usadas para codificar region0, region1 y region2, obtenido
%   con CICLO_INTERNO. REGION0_COUNT es la cantidad de bandas del factor de
%   escala (disminuidas en 1) incluidas en region0, obtenida con CICLO_INTERNO.
%   REGION1_COUNT es la cantidad de bandas del factor de escala (disminuidas en 1)
%   incluidas en region1, obtenida con CICLO_INTERNO. PREFLAG es la bandera que
%   indica el uso de la opción de preénfasis, obtenida con CICLO_EXTERNO.
%   SCALEFAC_SCALE es el escalamiento para los factores de escala, obtenido en
%   WAV2MP3. COUNT1TABLE_SELECT indica la tabla de Huffman escogida para codificar
%   los cuádruplos de valores pertenecientes a la región count1, obtenida con
%   CICLO_INTERNO.
%
%   Ver también WAV2MP3, CICLO INTERNO, CICLO EXTERNO.

global bin_str

% Inicialización de variables.
private_bits = 0; % 5 bits que no se usan. Se ponen en '0'.
scfsi = 0; % Los factores de escala se transmiten para cada gránulo.
window_switching_flag = 0; % Sólo se usan bloques largos.

% Escritura de la información secundaria en la cadena binaria.
bin_str = [bin_str dec2bin(main_data_begin,9)];
bin_str = [bin_str dec2bin(private_bits,5)];
bin_str = [bin_str dec2bin(scfsi,4)];
for gr = 1:2,
    bin_str = [bin_str dec2bin(part2_3_length(gr),12)];
    bin_str = [bin_str dec2bin(big_values(gr),9)];
    bin_str = [bin_str dec2bin(global_gain(gr),8)];
    bin_str = [bin_str dec2bin(scalefac_compress(gr),4)];
    bin_str = [bin_str dec2bin(window_switching_flag,1)];
    for region = 1:3,
        bin_str = [bin_str dec2bin(table_select(gr,region),5)];
    end
    bin_str = [bin_str dec2bin(region0_count(gr),4)];
    bin_str = [bin_str dec2bin(region1_count(gr),3)];
    bin_str = [bin_str dec2bin(preflag,1)];
    bin_str = [bin_str dec2bin(scalefac_scale(gr),1)];

```

```

    bin_str = [bin_str dec2bin(count1table_select(gr),1)];
end

```

```

function LBC = Limites_banda_critica
%LIMITES_BANDA_CRITICA Límites de las Bandas Críticas.
%
% LBC=LIMITES_BANDA_CRITICA
% Proporciona, en la matriz LBC, los valores de la tabla "LÍMITES DE LAS BANDAS
% CRÍTICAS" tal y como está dada en el estándar ISO 11172-3; para la Capa II,
% y una frecuencia de muestreo de 44.1 kHz.
%
% Ver también UMBRAL_ABSOLUTO

% En la matriz LBC se almacena la tabla "Límites de las Bandas Críticas" para la
% Capa II a 44.1 kHz, proporcionada por el estándar ISO 11172-3.Las frecuencias
% corresponden al límite superior de cada banda crítica.

% Índice   |   Frecuencia   |   Tasa de Banda Crítica
%          |                |   (Hz)
% (z)
LBC = [
    1           43.066           0.425
    2           86.133           0.850
    3           129.199           1.273
    5           215.332           2.112
    7           301.465           2.934
   10           430.664           4.124
   13           559.863           5.249
   16           689.063           6.301
   19           818.262           7.274
   22           947.461           8.169
   26          1119.727           9.244
   30          1291.992          10.195
   35          1507.324          11.232
   40          1722.656          12.125
   46          1981.055          13.042
   51          2325.586          14.062
   56          2756.250          15.100
   62          3273.047          16.110
   69          3875.977          17.079
   74          4478.906          17.904
   79          5340.234          18.922
   85          6373.828          19.963
   92          7579.688          20.971
   99          9302.344          22.074
  105          11369.531          22.984
  117          15503.906          24.013
  130          19982.813          24.574
];

```

```

function NPS = Nivel_presion_sonora(F,FDE)
%NIVEL_PRESION_SONORA Nivel de presión sonora en cada subbanda.
%
% NPS=NIVEL_PRESION_SONORA(F,FDE)
% Determina el nivel de presión sonora para cada subbanda (vector NPS).
% El vector F es la densidad espectral de potencia normalizada.
% FDE es la matriz de factores de escala (3 por subbanda).
%
% Ver también ANALISIS_FFT, FACTORES_ESCALA

% Halla el máximo entre los 3 factores de escala de cada subbanda.
MAXFDE = max(FDE);

% Encuentra el valor máximo de las líneas de frecuencia para cada subbanda
% (componentes del vector XKI).
XKI = [];
for k = 1:16:512,
    Xk = max(F(k:k+15));

```

```

    XKI = [XKI; Xk];
end

% Calcula el nivel de presión sonora por subbanda (vector NPS).
for i = 1:32,
    NPS(i) = max(XKI(i), 20*log10(MAXFDE(i)*32768)-10);
End



---


function [BSR,LTR,LNTR] = Reduccion(LT,LNT,BS,UA,MAP)
%REDUCCION Reduce la cantidad de componentes enmascarantes tonales y no-tonales
% de la señal de audio
%
% [BSR,LTR,LNTR]=REDUCCION(LT,LNT,BS,UA,MAP)
% La matriz LTR lista las componentes tonales y la matriz LNTR lista las
% componentes no-tonales; ambas después de disminuir la cantidad de
% componentes enmascarantes. Las componentes que se encuentran por debajo
% del mínimo umbral auditivo, o que su distancia con respecto a otra componente
% es menor a medio Bark, son eliminadas. Las matrices LTR y LNTR se componen
% de dos columnas, la primera proporciona el índice de la línea de frecuencia
% y la segunda el nivel de presión sonora de esa línea. El vector BSR
% proporciona las banderas para cada una de las 512 líneas de frecuencia,
% después de realizar la disminución sobre LT y LNT, tal que:
% - Componente no examinada = 0.
% - Componente tonal = 1.
% - Componente no tonal = 2.
% - Componente irrelevante = 3.
%
% UA es la matriz de umbral absoluto y MAP es el vector de mapeo entre las
% líneas de frecuencia y su índice para la matriz UA; ambos obtenidos con
% UMBRAL_ABSOLUTO. LBC es la matriz de límites de las bandas críticas
% obtenida con LIMITES_BANDA_CRITICA. LT y LNT son las matrices de
% componentes tonales y de componentes no-tonales, y BS es el vector de
% banderas, todos obtenidos con COMPONENTES_TONALES.
%
% Ver también UMBRAL_ABSOLUTO, LIMITES_BANDA_CRITICA, COMPONENTES_TONALES.

BSR = BS;

% a) Caso no-tonal y caso tonal, parte I:
% Elimina las componentes tonales (filas de la matriz LT) o las componentes
% no-tonales (filas de la matriz LNT) en las cuales el nivel de presión sonora
% esté por debajo del mínimo umbral auditivo (tercera columna de la matriz
% UA); y fija en 3 su valor de bandera en el vector BSR.

% Primero, se analiza el caso no-tonal.
LNTR = [];
if not isempty(LNT)
    for i = 1:26,
        k = LNT(i,1);
        if (LNT(i,2) < UA(MAP(k),3))
            BSR(k) = 3;
        else
            LNTR = [LNTR; LNT(i,:)];
        end
    end
end

% Después, se analiza el caso tonal, parte I.
LTR = [];
if not isempty(LT)
    for i = 1:length(LT(:,1)),
        k = LT(i,1);
        if (LT(i,2) < UA(MAP(k),3))
            BSR(k) = 3;
        else
            LTR = [LTR; LT(i,:)];
        end
    end
end

% b) Caso tonal, parte II:
% Elimina dos o más componentes tonales (filas de la matriz LT) cuya distancia

```

```

% respecto a otra componente es menor que medio bark, y fija en 3 su valor de
% bandera en el vector BSR. Se elimina(n) la(s) componente(s) con menor nivel
% de presión sonora y se conserva la componente con mayor nivel.
if not isempty(LTR)
    i = 1;
    while (i < length(LTR(:,1))),
        k = LTR(i,1);
        sigk = LTR(i+1,1);
        if (UA(MAP(sigk),2) - UA(MAP(k),2) < 0.5)
            if (LTR(i,2) < LTR(i+1,2))
                % Elimina la componente con índice k.
                LTR = LTR([1:i-1,i+1:length(LTR(:,1))],:);
                BSR(k) = 3;
            else
                % Elimina la componente con índice k+1.
                LTR = LTR([1:i,i+2:length(LTR(:,1))],:);
                BSR(sigk) = 3;
            end
        end
        i = i+1;
    end
end
end

```

```

function MDCT = Transf_discreta_coseno(S,U576,sb,gr)
%TRANSF_DISCRETA_COSENO Calcula la transformada discreta del coseno modificada.
%
% MDCT=TRANSF_DISCRETA_COSENO(S,U576,SB,GR)
% Retorna los 18 valores (vector MDCT) correspondientes al resultado de aplicar
% la transformada discreta del coseno modificada, con 50% de solapamiento,
% sobre 36 muestras subbanda consecutivas en el tiempo
%
% S es la matriz de 1152 muestras subbanda obtenidas con FILTRO_SUBBANDA. La
% matriz U576 proporciona las últimas 576 muestras subbanda de la matriz S
% anterior, SB es el número de subbanda a procesar, GR es el gránulo actual,
% los cuales fueron obtenidos en WAV2MP3.
%
% Ver también FILTRO_SUBBANDA, WAV2MP3.

% En cada subbanda, los 18 valores de salida del gránulo anterior (almacenados en
% U576) y los 18 valores de salida del gránulo actual se ensamblan en un bloque
% de 36 muestras (vector xp). En caso de que el gránulo sea el segundo, para cada
% subbanda, la matriz S es procesada directamente.
if gr == 1
    xp = [U576(1:18,sb); S(1:18,sb)];
else
    xp = S(1:36,sb);
end

% Calcula el vector z de 36 valores para el tipo de bloque NORMAL.
z = xp' .* sin((pi/36)*(0.5:35.5));

% Calcula el vector MDCT de 18 valores para bloques largos.
MDCT = zeros(18,1);
fMDCT = pi/72*(19:2:89);
for i = 1:18,
    MDCT(i) = sum(z.*cos((2*i-1)*fMDCT));
end

```

```

function [UA,MAP,UES] = Umbral_absoluto
%UMBRAL_ABSOLUTO Mínimo Umbral Auditivo.
%
% [UA,MAP,UES]=UMBRAL_ABSOLUTO
% Proporciona en la matriz UA, los valores de:
% - Índices de líneas de frecuencia (columna 1)
% - Tasa de Banda Crítica (columna 2)
% - Umbral Absoluto (columna 3)
% para la capa II con una frecuencia de muestreo de 44.1 kHz. En el vector MAP
se
% proporciona el mapeo entre las líneas de frecuencia y su índice para la matriz
UA.

```

```
% El vector UES contiene solamente el Umbral Absoluto, también llamado Umbral en
% Silencio.
```

```
% Almacena en la matriz TablaUA la tabla denominada "Frecuencias, Tasas de Banda
% Crítica y Umbral Absoluto" para la Capa II a 44.1 kHz, proporcionada por el
% estándar ISO 11172-3.
```

```
% -----
```

```
% Frecuencia | Tasa de Banda Crítica | Umbral Absoluto
% (Hz) (dB) (z)
```

```
TablaUA = [
43.07 0.425 45.05
86.13 0.850 25.87
129.20 1.273 18.70
172.27 1.694 14.85
215.33 2.112 12.41
258.40 2.525 10.72
301.46 2.934 9.47
344.53 3.337 8.50
387.60 3.733 7.73
430.66 4.124 7.10
473.73 4.507 6.56
516.80 4.882 6.11
559.86 5.249 5.72
602.93 5.608 5.37
646.00 5.959 5.07
689.06 6.301 4.79
732.13 6.634 4.55
775.20 6.959 4.32
818.26 7.274 4.11
861.33 7.581 3.92
904.39 7.879 3.74
947.46 8.169 3.57
947.46 8.450 3.40
1033.59 8.723 3.25
1076.66 8.987 3.10
1119.73 9.244 2.95
1162.79 9.493 2.81
1205.86 9.734 2.67
1248.93 9.968 2.53
1291.99 10.195 2.39
1335.06 10.416 2.25
1378.13 10.629 2.11
1421.19 10.836 1.97
1464.26 11.037 1.83
1507.32 11.232 1.68
1550.39 11.421 1.53
1593.46 11.605 1.38
1636.52 11.783 1.23
1679.59 11.957 1.07
1722.66 12.125 0.90
1765.72 12.289 0.74
1808.79 12.448 0.56
1851.86 12.603 0.39
1894.92 12.753 0.21
1937.99 12.900 0.02
1981.05 13.042 -0.17
2024.12 13.181 -0.36
2067.19 13.317 -0.56
2153.32 13.578 -0.96
2239.45 13.826 -1.38
2325.59 14.062 -1.79
2411.72 14.288 -2.21
2497.85 14.504 -2.63
2583.98 14.711 -3.03
2670.12 14.909 -3.41
2756.25 15.100 -3.77
2842.38 15.284 -4.09
2928.52 15.460 -4.37
3014.65 15.631 -4.60
3100.78 15.796 -4.78
3186.91 15.955 -4.91
3273.05 16.110 -4.97
3359.18 16.260 -4.98
```

3445.31	16.406	-4.92
3531.45	16.547	-4.81
3617.58	16.685	-4.65
3703.71	16.820	-4.43
3789.84	16.951	-4.17
3875.98	17.079	-3.87
3962.11	17.205	-3.54
4048.24	17.327	-3.19
4134.38	17.447	-2.82
4306.64	17.680	-2.06
4478.91	17.905	-1.32
4651.17	18.121	-0.64
4823.44	18.331	-0.04
4995.70	18.534	0.47
5167.97	18.731	0.89
5340.23	18.922	1.23
5512.50	19.108	1.51
5684.77	19.289	1.74
5857.03	19.464	1.93
6029.30	19.635	2.11
6201.56	19.801	2.28
6373.83	19.963	2.46
6546.09	20.120	2.63
6718.36	20.273	2.82
6890.63	20.421	3.03
7062.89	20.565	3.25
7235.16	20.705	3.49
7407.42	20.840	3.74
7579.69	20.972	4.02
7751.95	21.099	4.32
7924.22	21.222	4.64
8096.48	21.342	4.98
8268.75	21.457	5.35
8613.28	21.677	6.15
8957.81	21.882	7.07
9302.34	22.074	8.10
9646.88	22.253	9.25
9991.41	22.420	10.54
10335.94	22.576	11.97
10680.47	22.721	13.56
11025.00	22.857	15.31
11369.53	22.984	17.23
11714.06	23.102	19.34
12058.59	23.213	21.64
12403.13	23.317	24.15
12747.66	23.415	26.88
13092.19	23.506	29.84
13436.72	23.592	33.05
13781.25	23.673	36.52
14125.78	23.749	40.25
14470.31	23.821	44.27
14814.84	23.888	48.59
15159.38	23.952	53.22
15503.91	24.013	58.18
15848.44	24.070	63.49
16192.97	24.125	68.00
16537.50	24.176	68.00
16882.03	24.225	68.00
17226.56	24.271	68.00
17571.09	24.316	68.00
17915.63	24.358	68.00
18260.16	24.398	68.00
18604.69	24.436	68.00
18949.22	24.473	68.00
19293.75	24.508	68.00
19638.28	24.542	68.00
19982.81	24.574	68.00

];

% Convierte la primera columna de la matriz TablaUA de 130 valores de frecuencia
% a sus equivalentes 130 índices de líneas de frecuencia.

UA = [round(TablaUA(:,1)/44100*1024) TablaUA(:,2:3)];

% Almacena en el vector MAP el resultado del mapeo entre 512 líneas de frecuencia

```

% y sus 130 índices para la matriz UA. Primero, se mapea el valor inicial.
MAP(1) = 1;

% Después, se mapean los valores finales.
MAP(UA(130,1):512) = 130;

% Por último, se mapean todos los demás valores.
for i = 2:129,
    MAP(UA(i,1):UA(i+1,1)) = i;
end

% Como la mínima tasa de bits es 96 Kbps, se disminuyen en 12 dB los valores
% del Umbral Absoluto, almacenados en la tercera columna de la matriz UA.
UA(:,3) = UA(:,3)-12;

% Almacena en el vector UES la tercera columna de la matriz UA que contiene
% los valores de Umbral en Silencio o Umbral Absoluto para la Capa II.
UES = UA(:,3);

```

```

function UEG = Umbral_enmasc_global(UES,UET,UENT)
%UMBRAL_ENMASC_GLOBAL Calcula el umbral de enmascaramiento global.
%
% UEG=UMBRAL_ENMASC_GLOBAL(UES,UET,UENT)
% Calcula el umbral de enmascaramiento global (vector UEG) para el subconjunto
% de líneas de frecuencia definidas en la tabla denominada "Frecuencia, Tasa
% de Banda Crítica y Umbral Absoluto" para la Capa II a 44.1 kHz, proporcionada
% por el estándar ISO 11172-3. Éste es la suma (en la escala normal de amplitud
% cuadrada del espectro) de los umbrales de enmascaramiento individual y del
% umbral absoluto o umbral en silencio
%
% La matriz UET es el efecto enmascarante de las componentes tonales y la matriz
% UENT es el efecto enmascarante de las componentes no tonales, ambas obtenidas
% con UMBRALES_ENMASC_INDIVIDUAL. El vector UES es el umbral en silencio,
% proporcionado por UMBRAL_ABSOLUTO.
%
% Ver también UMBRALES_ENMASC_INDIVIDUAL, UMBRAL_ABSOLUTO.

% El umbral de enmascaramiento global (vector UEG) se calcula para el subconjunto
% de frecuencias definidas en la tabla denominada "Frecuencia, Tasa de Banda
% Crítica y Umbral Absoluto" para la Capa II a 44.1 kHz, proporcionada por el
% estándar ISO 11172-3. Éste es la suma de las potencias de los correspondientes
% umbrales de enmascaramiento individual (matrices UET y UENT) y el umbral en
% silencio (vector UES).
if not(isempty(UET))
    m = length(UET(:,1));
end
if not(isempty(UENT))
    n = length(UENT(:,1));
end

for i = 1:130

    % Para el umbral en silencio.
    t = 10^(UES(i)/10);

    % Contribución de las componentes tonales.
    if not(isempty(UET))
        for j = 1:m,
            t = t+10^(UET(j,i)/10);
        end
    end

    % Contribución de las componentes no tonales.
    if not(isempty(UENT))
        for j = 1:n,
            t = t+10^(UENT(j,i)/10);
        end
    end

    % Umbral de enmascaramiento global.
    UEG(i) = 10*log10(t);

```

end

```
function UEG = Umbral_enmasc_global(UES,UET,UENT)
%UMBRAL_ENMASC_GLOBAL Calcula el umbral de enmascaramiento global.
%
% UEG=UMBRAL_ENMASC_GLOBAL(UES,UET,UENT)
% Calcula el umbral de enmascaramiento global (vector UEG) para el subconjunto
% de líneas de frecuencia definidas en la tabla denominada "Frecuencia, Tasa
% de Banda Crítica y Umbral Absoluto" para la Capa II a 44.1 kHz, proporcionada
% por el estándar ISO 11172-3. Éste es la suma (en la escala normal de amplitud
% cuadrada del espectro) de los umbrales de enmascaramiento individual y del
% umbral absoluto o umbral en silencio
%
% La matriz UET es el efecto enmascarante de las componentes tonales y la matriz
% UENT es el efecto enmascarante de las componentes no tonales, ambas obtenidas
% con UMBRALES_ENMASC_INDIVIDUAL. El vector UES es el umbral en silencio,
% proporcionado por UMBRAL_ABSOLUTO.
%
% Ver también UMBRALES_ENMASC_INDIVIDUAL, UMBRAL_ABSOLUTO.

% El umbral de enmascaramiento global (vector UEG) se calcula para el subconjunto
% de frecuencias definidas en la tabla denominada "Frecuencia, Tasa de Banda
% Crítica y Umbral Absoluto" para la Capa II a 44.1 kHz, proporcionada por el
% estándar ISO 11172-3. Éste es la suma de las potencias de los correspondientes
% umbrales de enmascaramiento individual (matrices UET y UENT) y el umbral en
% silencio (vector UES).
if not(isempty(UET))
    m = length(UET(:,1));
end
if not(isempty(UENT))
    n = length(UENT(:,1));
end

for i = 1:130

    % Para el umbral en silencio.
    t = 10^(UES(i)/10);

    % Contribución de las componentes tonales.
    if not(isempty(UET))
        for j = 1:m,
            t = t+10^(UET(j,i)/10);
        end
    end

    % Contribución de las componentes no tonales.
    if not(isempty(UENT))
        for j = 1:n,
            t = t+10^(UENT(j,i)/10);
        end
    end

    % Umbral de enmascaramiento global.
    UEG(i) = 10*log10(t);

end
```

```
function UEM = Umbral_enmasc_minimo(UEG,MAP)
%UMBRAL_ENMASC_MINIMO Calcula el umbral de enmascaramiento mínimo.
%
% UEM=UMBRAL_ENMASC_MINIMO(UEG,MAP)
% Retorna en el vector UEM el valor mínimo del umbral de enmascaramiento
% global para cada una de las 32 subbandas
%
% El vector UEG es el umbral de enmascaramiento global obtenido con
% UMBRAL_ENMASC_GLOBAL. El vector MAP es el mapeo entre las líneas de
% frecuencia y su índice para la matriz UA en la función UMBRAL_ABSOLUTO.
```

```

%
% Ver también UMBRAL_ENMASC_GLOBAL, UMBRAL_ABSOLUTO.

% Extrae del umbral de enmascaramiento global (vector UEG), su valor mínimo
% en cada subbanda (vector UEM).
for n = 1:32,
    UEM(n) = min(UEG(MAP((n-1)*16+1):MAP((n-1)*16+16)));
end



---



function [UET,UENT] = Umbrales_enmasc_individual(F,LTR,LNTR,UA,MAP)
%UMBRALES_ENMASC_INDIVIDUAL Calcula los umbrales de enmascaramiento individual.
%
% [UET,UENT]=UMBRALES_ENMASC_INDIVIDUAL(F,LTR,LNTR,UA,MAP)
% Calcula el efecto enmascarante de las componentes tonales (matriz UET) y
% de las componentes no tonales (matriz UENT) sobre las líneas de frecuencia
% vecinas. Para esto, el nivel de presión sonora de cada componente enmascarante
% es sumado con su índice de enmascaramiento y con su función de
enmascaramiento.
%
% F es el vector de densidad espectral de potencia normalizada, obtenido con
% ANALISIS_FFT. UA es la matriz de umbral absoluto y MAP es el vector de mapeo
% entre las líneas de frecuencia y su índice para la matriz UA; ambos obtenidos
% con UMBRAL_ABSOLUTO. La matriz LTR es la lista reducida de las componentes
% tonales y la matriz LNTR es la lista reducida de las componentes no-tonales,
% ambas obtenidas con REDUCCION.
%
% Los umbrales de enmascaramiento individual para las componentes tonales
% y no tonales se fijan en -INF, ya que la función de enmascaramiento tiene
% atenuación infinita más allá de -3 y de +8 Barks, o sea, la componente no
% tiene efecto enmascarante sobre frecuencias más allá de aquellos rangos.
if isempty(LTR)
    UET = [];
else
    UET = zeros(length(LTR(:,1)),130) - 200; % -200 dB equivale a -INF.
end
if isempty(LNTR)
    UENT = [];
else
    UENT = zeros(length(LNTR(:,1)),130) - 200; % -200 dB equivale a -INF.
end

% Sólo un subconjunto de las muestras son consideradas para el futuro cálculo
% del umbral de enmascaramiento global. El número de estas muestras depende
% de la tasa de muestreo y de la capa de codificación. Toda la información
% necesaria está en la matriz UA, la cual contiene las frecuencias, las tasas
% de banda crítica y el umbral absoluto.
if not(isempty(LTR)) & not(isempty(LNTR))
    for i = 1:130,
        zi = UA(i,2); % Tasa de banda crítica de la frecuencia considerada.
        if not(isempty(LTR))

            % Para las componentes tonales.
            for k = 1:length(LTR(:,1)),
                j = LTR(k,1);
                zj = UA(MAP(j),2); % Tasa de banda crítica de la componente
enmascarante.
                dz = zi-zj; % Distancia en Barks hasta la componente enmascarante.

                % Como la componente tonal tiene atenuación infinita más allá de -3 y
                % de +8 Barks, entonces los cálculos sólo se realizan para el rango
                % dz = {-3...+8}.
                if (dz>=-3 & dz<8)

                    % Índice de Enmascaramiento.
                    avtm = -1.525-0.275*zj-4.5;

                    % Función de Enmascaramiento.
                    if (-3<=dz & dz<-1)
                        vf = 17*(dz+1)-(0.4*F(j)+6);
                    elseif (-1<=dz & dz<0)

```

```

        vf = (0.4*F(j)+6)*dz;
    elseif (0<=dz & dz<1)
        vf = -17*dz;
    elseif (1<=dz & dz<8)
        vf = -(dz-1)*(17-0.15*F(j))-17;
    end

    % Umbral de enmascaramiento, componentes tonales.
    UET(k,i) = F(j)+avtm+vf;
end
end
end

% Para las componentes no tonales.
if not(isempty(LNTR))
    for k = 1:length(LNTR(:,1)),
        j = LNTR(k,1);
        zj = UA(MAP(j),2); % Tasa de banda crítica de la componente
enmascarante.
        dz = zi-zj; % Distancia en Barks hasta la componente enmascarante.

        % Como la componente no-tonal tiene atenuación infinita más allá
        % de -3 y de +8 Barks, entonces los cálculos sólo se realizan para
        % el rango dz = {-3...+8}.
        if (dz>=-3 & dz<8)

            % Índice de Enmascaramiento.
            avnm = -1.525-0.175*zj-0.5;

            % Función de Enmascaramiento.
            if (-3<=dz & dz<-1)
                vf = 17*(dz+1)-(0.4*F(j)+6);
            elseif (-1<=dz & dz<0)
                vf = (0.4*F(j)+6)*dz;
            elseif (0<=dz & dz<1)
                vf = -17*dz;
            elseif (1<=dz & dz<8)
                vf = -(dz-1)*(17-0.15*F(j))-17;
            end

            % Umbral de enmascaramiento, componente no-tonal.
            UENT(k,i) = F(j)+avnm+vf;
        end
    end
end
end
end
end
end

```

[ANEXO 10]

CÓDIGO MATLAB DE PLATAFORMA DE EVALUACIÓN DE AUDIO

```
function [DA,L] = LAST_MOS(Audorg, Auddeg)
% A. PREPARA SEÑALES PARA ANALISIS
% Señal original en vector x , degradada en y. SOLO WAV MONO
%INTRODUCIR PRIMERO SEÑAL ORIGINAL DESPUES DEGRADADA
[x, Fs, kbitx]=wavread(Audorg);
[y, Fs, kbity]=wavread(Auddeg);
%A.1 Algoritmo requiere que x y y esten sincronizados a la misma
longitud.
%Se debe sincronizar ambas .
%Se toma el mínimo valor de longitud de ambas.
nx = length(x);
ny = length(y);
n1 = min(nx,ny);
k = min(nx,ny);
[XCF,Lags] = crosscorr(x, y, k-1);
%XCF y Lags son vectores de 2k-1 elementos. Lags contiene los indices
%de la correlación entre -(k-1) y +(k-1), delay guarda retraso.
M=max(XCF)
for r=1:(2*k-1)
if XCF(r)>=M
delay=Lags(r);
end
end
% XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
%Retraso entre la señal orig y degradada se encuentra en
%la máxima crosscorrelacion
%y "busca" a x por las muestras retardadas
nx = length(x);
y(1:delay)=[]; % Se sincroniza removiendo las primeras muestras
% Si y es mas largo que x
oldny = length(y);
y((nx+1):oldny)=[]; %Ahora x y y tienen el mismo tamaño
ny = length(y);
% XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
%
%A.2 Graficas de muestras desde el comienzo y final de ambas señales
samples=10000; t=1:1:samples; figure (2); subplot(4,1,1); plot(x);
axis ([0 samples 1.1*min(x) 1.1*max(x)]);
title('Comienzo original');
xlabel('muestras'); subplot(4,1,2); plot(y);
axis([0 samples 1.1*min(y) 1.1*max(y)]);
title('Comienzo degradada'); xlabel('muestras'); subplot(4,1,3);
plot(x);
axis ([nx-samples nx 1.1*min(x) 1.1*max(x)]);
title('Fin original'); xlabel('muestras'); subplot(4,1,4); plot(y);
axis([ny-samples ny 1.1*min(y) 1.1*max(y)]);
title('Fin degradada'); xlabel('muestras');

%A.3 REMOCION del valor promedio de x y y
mean_x=mean(x); %valor promedio de x
mean_y=mean(y); %valor promedio de y
%El valor promedio es removido de x y y
for i=1:n1
x(i)=x(i)-mean_x;
```

```

y(i)=y(i)-mean_y;
end

% x y y son normalizados a nivel RMS
rms_x=norm(x)/sqrt(n1); %rms de x
rms_y=norm(y)/sqrt(n1); %rms de y
%Normalización
for i=1:n1
x(i)=x(i)/rms_x;
y(i)=y(i)/rms_y;
end
%
%
% B. TRANSFORMACION AL DOMINIO DE LA FRECUENCIA
%Las señales se dividen en series de ventanas de 1408 muestras c/u
%con 50% de solapamiento. Algunas muestras al final seran descartadas
%xbufl1 ybufl1 incluyen solo ventanas completas, zx zy puede incluir,
%si existen,muestras al final de la ventana llena.
[xbufl1,zx]=buffer(x,128,64);
[ybufl1,zy]=buffer(y,128,64);
frames=size(xbufl1,2); %numero de cuadros
n2=frames;
% Enventanado en Hamming
for j=1:n2
for i=1:128
xbufl(i,j)=xbufl1(i,j)*(0.54-0.46*cos((2*pi*(i-1))/127));
ybufl(i,j)=ybufl1(i,j)*(0.54-0.46*cos((2*pi*(i-1))/127));
end
end
%
%
%
%B.2 FFT para xbuf y ybuf
xfft=abs(fft(xbuf,128));
yfft=abs(fft(ybuf,128));
%Únicamente serán guardadas las magnitudes en raíz de las muestras 1-65
%de cada ventana. Serán guardadas en matrices X y Y de 65*n2
%(n2=numero de ventanas)
for j=1:n2
for i=1:65
X(i,j)=xfft(i,j)^2;
Y(i,j)=yfft(i,j)^2;
end
end
%
%
%
% C. SELECCION DE LA VENTANA
%Solo se usaran las ventanas de X y Y que exceden el umbral de energia
% para el cálculo de los vectores de diferencia de energia entre ventanas
xenergy=zeros(1,n2);
yenergy=zeros(1,n2);
for j=1:n2
for i=1:65
xenergy(j)=xenergy(j)+X(i,j);
yenergy(j)=yenergy(j)+Y(i,j);
end
end
%umbrales
xthreshold=10^(-1.5)*max(xenergy);
ythreshold=10^(-3.5)*max(yenergy);
%
%
%
%Se retienen las ventanas donde se reconocen ambos niveles de energia

```

```

%Se eliminan de X y Y las que contienen una o mas muestras iguales a cero
%Para detectar 0, 0 significa que no hay ceros en la ventana
% 1 que si hay un cero, Sirve para X y Y.
iszero=zeros(1,n2);
for j=1:n2
for i=1:65
if ((X(i,j)==0) | (Y(i,j)==0))
    iszero(j) = 1;
end
end
end
j=1; %Ventana de anteriores matrices X y Y
k=1; %Ventana de nuevas matrices
% Remoción de ventanas que no cumplen con requisitos antes expuestos
while j <= n2
if ((xenergy(j)>=xthreshold)&(yenergy(j)>=ythreshold)& (iszero(j) == 0))
for i=1:65
X1(i,k)=X(i,j);
Y1(i,k)=Y(i,j);
end
k=k+1;
j=j+1;
else
    j=j+1;
end
end
n3=size(X1,2); %Numero de ventanas despues de removerse sobrantes
if (n3>0)
% Matrices que van a quedar
X=X1;
Y=Y1;
end;
%
%
%
%
% D. APROXIMACION DE VOLUMEN PERCIBIDO
%Se lleva a dB
for j=1:n3
for i=1:65
X(i,j)=10*log10(X(i,j));
Y(i,j)=10*log10(Y(i,j));
end
end
%
%
%
%
% E.CORE DEL PROGRAMA.
%Medición de diferencia de frecuencias por ventana
f1=zeros(1,65);
f2=zeros(1,65);
for i=1:65
meanX=0;
meanY=0;
for j=1:n3
meanX=meanX+X(i,j);
meanY=meanY+Y(i,j);
end
meanX=meanX/n3;
meanY=meanY/n3;
%Determina f1
f1(i)= meanY-meanX;
end
%Normalización de Y

```

```

Ysum=Ysum*h;
Xsum=Xsum*h;
t1(j)=Ysum-Xsum;
end
% Normalizacion de Y
for j=1:n3
for i=g(k):(g(k+1)-1)
Y(i,j)=Y(i,j)-t1(j);
end
end
% Guardado de la parte positiva de la medición
sum=0;
for j=1:n3
if (t1(j)>0)
sum=sum+t1(j);
end
end
m(5+k)=sum/n3;
end
% F.3 Medición final de sobrantes
for j=1:n3
for i=1:65
t2(i,j)=Y(i,j)-X(i,j);
end
end
% Guardado de la parte positiva de medicion de sobrantes. Deteccion de
%estructura armónica
sum=0;
for j=1:n3
for i=2:65
if (t2(i,j)>0)
sum=sum+t2(i,j);
end
end
end
m(12)=sum/(n3*64); % Errores en el tiempo de la estructura armonica
%
%
%
% H. Combinaciones lineales y funciones logicas
% wi es vector ponderacion para las estructuras 1 y 2. a1 b1 a2 b2 seran
%parametros logicos
w1=[0.0034 -0.0650 -0.1304 0.1352 0.5931 0.2040 0.5577 0.1008 0.0627
0.0052 0.0107 1.1037];
% Vector w2 empirico, Vector con coeficientes de prueba
w2=[0.0000 -0.0837 -0.1199 0.1260 0.1660 0.6387 0.2195 0.0122 1.5544
0.0954 0.1720 1.1127]; VECTOR DE PRUEBA
a1=1.0000;
b1=-4.6877;
a2=1.0000;
b2=-3.0613;
% VALOR PRINCIPAL DE DIFERENCIA DA
DA=w1*m;
DA
%SALIDA FINAL DEL ALGORITMO, CON LA VARIABLE L
L=1/(1+exp(a1*DA+b1));
L
% CALCULO FINAL DEL MOS
MOS=5*L*1.009203940
end

```