

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA ELÉCTRICA Y
ELECTRÓNICA

ESTUDIO, ANÁLISIS E IMPLEMENTACIÓN EN SOFTWARE DE
SIMULACIÓN MATLAB DEL ALGORITMO MP3 PARA
COMPRESIÓN DE AUDIO A TRAVÉS DEL USO DE TÉCNICAS
DIGITALES PARA EL PROCESAMIENTO DE SEÑALES

PROYECTO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN
ELECTRÓNICA Y TELECOMUNICACIONES

BYRON PATRICIO DUQUE PILATAXI

bpdp4986@hotmail.com

CHRISTIAN JOSÉ TIPANTUÑA TENELEMA

ctipantunia@hotmail.com

DIRECTOR: DR. GUALBERTO HIDALGO HIDROVO

ghidalgo@mailfie.epn.edu.ec

Quito, Noviembre 2010

DECLARACIÓN

Nosotros, Byron Patricio Duque Pilataxi y Christian José Tipantuña Tenelema, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que hemos consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedemos nuestros derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

Byron Patricio Duque Pilataxi

Christian José Tipantuña Tenelema

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Byron Patricio Duque Pilataxi y Christian José Tipantuña Tenelema, bajo mi supervisión.

DR. GUALBERTO HIDALGO HIDROVO
DIRECTOR DEL PROYECTO

AGRADECIMIENTO

A Dios, por darme la vida y permitirme volver a la vida, guiarme e iluminarme, permitiéndome dar un paso tan importante.

A mi madre, de todo corazón, por ser la persona que ha estado junto a mí compartiendo mis alegrías y tristezas y quien con su sacrificio y apoyo incondicional ha hecho de mí una buena persona. Gracias mamita, por darme todo tu amor, estar conmigo en todo momento y brindarme la fortaleza necesaria para aferrarme a la vida.

A mi hermano, por ser mi mejor amigo, por apoyarme en todo momento, brindarme los mejores consejos, soportar mi carácter y permitirme compartir su vida y ser en ella alguien tan especial.

A mi familia y a las personas que me consideran como parte de su familia, les quiero dar las gracias por siempre estar pendiente de mí, por darme su apoyo constante en todo momento. A quien ha sido para mí como un padre, por brindarme su apoyo.

A mi amigo y compañero de tesis, con quien hemos pasado duros momentos durante la realización del proyecto, hasta ahora poder ver realizado nuestro proyecto.

A la universidad, por permitirme conocer a tantas personas valiosas y a todos mis profesores durante la carrera por transmitirme sus conocimientos, en especial, al Dr. Gualberto Hidalgo, con quien ha sido un honor haber trabajado durante todo el proyecto.

A mis amigos y compañeros, les doy las gracias infinitamente, por su apoyo y por estar junto a mí durante todo este tiempo y hacer de mi paso por la universidad algo inolvidable.

Son tantas personas a las que tengo que agradecer por estar conmigo en los momentos más difíciles de mi vida

A todos, les doy las gracias de todo corazón.

Byron P. Duque P.

AGRADECIMIENTO

A Dios

Por absolutamente todo lo que me ha dado.

A mis padres

Con los que siempre he contado y espero seguir contando el resto de mi vida; por darme su amor, su apoyo y ayuda incondicional en todo momento, procurando lo mejor para mí y facilitándome las cosas en la medida de lo posible; por sus sacrificios y esfuerzos, por tratar de hacer de mí un hombre de bien y darme una educación de calidad. Por su ánimo, paciencia, por haber soportado mi desorden y en varias ocasiones mi mal carácter.

A mis hermanos

Quienes siempre me han dado su amor y apoyo incluso cuando hemos estado peleados, por hacerme pensar que soy capaz de lograr mucho más de lo que yo realmente pienso, por creer que soy mucho más de lo que realmente soy, por ser tolerantes, por soportar mis bromas y hasta mi mal genio.

A todo el resto de mi familia, a mis tíos, primos y a mi abuelita quienes han estado conmigo en la buenas y malas, se han preocupado por mí y me han deseado siempre lo mejor.

A mi amigo y compañero de tesis por su ayuda en la elaboración del proyecto. A los profesores de mi facultad quienes han sido una inspiración profesional y quienes me han transmitido sus conocimientos, en especial al Dr. Gualberto Hidalgo por su apoyo y ayuda incondicional durante la realización del proyecto, también al Ing. Marcelo Lazzati por su pronta y desinteresada ayuda.

A todos mis amigos y compañeros.

A todos los mencionados anteriormente, de corazón.

Muchas gracias.

Christian J. Tipantuña T.

DEDICATORIA

Este trabajo va dedicado con todo el amor a mi madre Graciela y a mi hermano Andrés por ser una parte fundamental en mi vida y apoyarme en todo momento.

Los quiero mucho.

Byron P. Duque P.

DEDICATORIA

Dedicado con mucho amor y cariño a mis padres Marina y José y a mis hermanos Viviana y Saúl quienes son un pilar fundamental en mi vida.

Christian J. Típantuña T.

CONTENIDO

RESUMEN.....	XXII
PRESENTACIÓN.....	XXIV
CAPÍTULO 1.....	1
ESTUDIO DEL SONIDO	1
1.1 EL SONIDO.....	1
1.1.1 INTRODUCCIÓN	1
1.1.2 DEFINICIÓN.....	2
1.1.3 FÍSICA DEL SONIDO	3
1.1.3.1 Ondas de sonido.....	4
1.1.3.2 Medición del sonido	4
1.1.3.3 Velocidad del sonido.....	5
1.1.3.4 Longitud de onda del sonido	5
1.1.3.5 Amplitud	6
1.1.3.5.1 Intensidad.....	8
1.1.3.6 Frecuencia	9
1.1.3.6.1 Tono	11
1.1.3.6.2 Armónicas	12
1.1.3.7 Fase	12
1.1.3.8 Forma de onda.....	13
1.1.3.8.1 Representación gráfica	14
1.1.4 FUENTES SONORAS	15
1.2 EL SISTEMA FONATORIO HUMANO	16
1.2.1 PROPIEDADES DEL HABLA	16
1.2.2 CLASIFICACIÓN DE LAS VOCES	18
1.2.2.1 Clasificación sexual	18
1.2.2.1.1 Clasificación según tesitura	19
1.3 PSICOACÚSTICA.....	20
1.3.1 DEFINICIÓN.....	20
1.3.2 EL SISTEMA AUDITIVO HUMANO.....	20
1.3.2.1 El oído humano	21
1.3.2.1.1 Oído externo	22
1.3.2.1.2 Oído medio.....	23
1.3.2.1.3 Oído interno	25
1.3.2.2 Bandas críticas del sistema auditivo humano	26
1.3.2.2.1 Escala de Barks	27
1.3.2.3 Enmascaramiento	28
1.3.2.3.1 Enmascaramiento en frecuencia	29
1.3.2.3.2 Enmascaramiento temporal.....	30
1.4 AUDIO DIGITAL.....	31

1.4.1	INTRODUCCIÓN	31
1.4.2	AUDIO ANALÓGICO	32
1.4.2.1	Señales analógicas.....	32
1.4.2.2	Audio analógico	33
1.4.2.2.1	Historia	33
1.4.2.2.2	Definición.....	33
1.4.3	DIGITALIZACIÓN DE SEÑALES ANALÓGICAS	34
1.4.3.1	Muestreo	34
1.4.3.1.1	Teorema del muestreo.....	35
1.4.3.1.2	Aliasing	36
1.4.3.1.3	Jitter.....	37
1.4.3.1.4	Cuantización	38
1.4.3.1.5	Ruido de cuantización.....	38
1.4.3.1.6	Dithering	40
1.4.3.1.7	Cuantización uniforme	40
1.4.3.1.8	Cuantización no uniforme	41
1.4.3.1.9	Cuantización Ley A	42
1.4.3.1.10	Cuantización Ley μ	43
1.4.3.2	Codificación.....	43
1.4.3.3	Ventajas y desventajas de señales y sistemas digitales	44
1.4.3.3.1	Ventajas	44
1.4.3.3.2	Desventajas	45
1.4.4	CARACTERÍSTICAS DEL AUDIO DIGITAL	45
1.4.4.1	La computadora en sistemas de audio digital.....	46
1.4.4.2	Hardware para audio digital.....	47
1.4.4.2.1	Funcionamiento de la tarjeta de sonido	48
1.4.4.3	Parámetros de audio digital.....	49
1.4.4.4	Calidad del audio digital.....	50
1.4.5	FORMATOS DE FICHERO	51
1.4.5.1	Formatos de fichero autodestructivos	51
1.4.5.2	Formatos de fichero sin cabecera o tipo "raw"	52
	REFERENCIAS BIBLIOGRÁFICAS	53

CAPÍTULO 2. 54

COMPRESIÓN DE AUDIO 54

2.1	COMPRESIÓN DE LA INFORMACIÓN	54
2.1.1	INTRODUCCIÓN	54
2.1.2	NECESIDAD PARA LA COMPRESIÓN.....	55
2.1.3	COMPRESIÓN DE DATOS	56
2.1.3.1	Utilización de la compresión	57
2.1.3.2	Técnicas de compresión.....	58
2.1.3.2.1	Compresión física y lógica.....	58
2.1.3.2.2	Compresión simétrica y asimétrica.....	59
2.1.3.2.3	Codificación de fuente y de entropía.....	59
2.1.3.2.4	Compresión con y sin pérdida de información	60

2.1.3.2.5 Codificación no adaptativa, semiadaptativa y adaptativa	60
2.2 COMPRESIÓN LOSSY Y LOSSLESS	61
2.2.1 COMPRESIÓN CON PÉRDIDA DE INFORMACIÓN (LOSSY)	61
2.2.1.1 Codificación diferencial.....	62
2.2.1.2 Codificación por transformada.....	62
2.2.1.3 Cuantización vectorial.....	63
2.2.2 COMPRESIÓN SIN PÉRDIDA DE INFORMACIÓN (LOSSLESS).....	63
2.2.2.1 Codificación estadística	64
2.2.2.1.1 Codificación Huffman.....	65
2.2.2.2 Basados en diccionario.....	66
2.3 COMPRESIÓN DE AUDIO	67
2.3.1 INTRODUCCIÓN	67
2.3.1.1 Características de la compresión de audio	68
2.3.1.2 Códec de audio	68
2.3.1.2.1 Codificadores perceptuales	68
2.3.1.2.2 Codificadores paramétricos	69
2.3.1.2.3 Vcoders	69
2.3.1.2.4 Codificadores de forma de onda	69
2.3.1.2.5 Codificadores híbridos	70
2.3.1.3 Parámetros de los códecs de audio	70
2.3.2 COMPRESIÓN DE MÚSICA	72
2.3.2.1 Mínimo umbral auditivo.....	72
2.4 ALGORITMOS DE COMPRESIÓN DE AUDIO	73
2.4.1 INTRODUCCIÓN	73
2.4.2 CODIFICACIÓN PERCEPTUAL DE AUDIO	73
2.4.3 CODIFICACIÓN DE SUB-BANDAS	75
2.4.4 COMPRESIÓN DE AUDIO ADPCM.....	77
2.4.5 FORMATO DE AUDIO WAV	78
2.4.5.1 Características de un archivo WAV.....	79
2.4.6 FLAC	80
2.4.7 CODIFICACIÓN DE AUDIO EN EL ESTÁNDAR MPEG.....	81
2.4.7.1 Estándar MPEG-1.....	84
2.4.7.2 Estándar MPEG-2.....	85
2.4.7.3 Estándar MPEG-4.....	87
2.4.7.4 AAC.....	88
2.4.8 ESTUDIO COMPARATIVO DE LAS CARACTERÍSTICAS DE LOS ALGORITMOS DE COMPRESIÓN DE AUDIO	89
REFERENCIAS BIBLIOGRÁFICAS	91

CAPÍTULO 3. 92

SISTEMAS DE ALTA FIDELIDAD 92

3.1 INTRODUCCIÓN.....	92
3.2 ALTA FIDELIDAD	92
3.2.1 PREÁMBULO.....	92
3.2.2 HISTORIA.....	93

3.2.3 GENERALIDADES.....	94
3.2.3.1 Hi-Fi y apariencia de realismo	95
3.2.3.2 Modularidad	96
3.2.4 ASPECTOS DE LA ALTA FIDELIDAD	97
3.2.5 ALTA FIDELIDAD COMO NORMA DE CALIDAD	98
3.2.6 EQUIPOS DE ALTA FIDELIDAD.....	99
3.2.6.1 Equipos modernos	99
3.3 COMPARATIVA CON EL FORMATO MP3	100
REFERENCIAS BIBLIOGRÁFICAS	103

CAPÍTULO 4. 104

FORMATO MP3 104

4.1 INTRODUCCIÓN.....	104
4.1.1 HISTORIA.....	104
4.1.1.1 Cronología del mp3.....	105
4.1.2 DEFINICIÓN.....	106
4.2 ASPECTOS PRINCIPALES DEL ESTÁNDAR ISO/IEC 11172-3 (MPEG-1 CAPA DE AUDIO III).....	108
4.2.1 CODIFICACIÓN	109
4.2.1.1 Codificación MPEG-1 para la capa 3	110
4.2.1.2 Análisis psicoacústico	112
4.2.1.2.1 Alineación en tiempo.....	113
4.2.1.2.2 Representación espectral	113
4.2.1.2.3 Componentes tonales y no tonales	114
4.2.1.2.4 Estimación del índice de tonalidad	114
4.2.1.2.5 Función de dispersión.....	115
4.2.1.2.6 Umbral de enmascaramiento individual	115
4.2.1.2.7 Umbral de enmascaramiento global.....	116
4.2.1.2.8 Pre - eco.....	117
4.2.1.2.9 Umbral de enmascaramiento mínimo.....	118
4.2.1.2.10 Relaciones señal a máscara	118
4.2.1.3 Banco de filtros híbridos conmutados	119
4.2.1.3.1 Filtro pasa-altos.....	120
4.2.1.3.2 Banco de filtros polifásicos	120
4.2.1.3.3 Transformada discreta del coseno modificada (MDCT)	124
4.2.1.4 Repartición de ruido.....	126
4.2.1.4.1 Ciclo interno (rate control loop).....	127
4.2.1.4.2 Ciclo externo (distortion control loop).....	127
4.2.1.5 Flujo de bits MP3 válido.....	128
4.2.1.6 Cuantización no uniforme	129
4.2.1.7 Codificación Huffman (codificación entrópica)	129
4.2.1.8 Reserva de bits	131
4.2.1.9 Modos de funcionamiento.....	133
4.2.1.9.1 Codificación joint stereo.....	133
4.2.2 DECODIFICACIÓN	134

4.3	FORMATO DE LAS TRAMAS MP3	134
4.3.1	ENCABEZADO DE TRAMAS	136
4.3.2	CHEQUEO DE ERRORES	140
4.3.3	INFORMACIÓN SECUNDARIA.....	142
4.3.4	DATOS PRINCIPALES	146
4.4	CURIOSIDADES DEL FORMATO MP3	149
4.5	INNOVACIONES EN EL FORMATO MP3.....	151
4.5.1	FORMATO MP3 DE ALTA CALIDAD.....	151
4.5.2	MP3PRO	152
4.5.3	FORMATO MT9	153
	REFERENCIAS BIBLIOGRÁFICAS	154

CAPÍTULO 5. 156

DISEÑO E IMPLEMENTACIÓN EN MATLAB DEL ALGORITMO MP3 PARA COMPRESIÓN DE AUDIO 156

5.1	INTRODUCCIÓN	156
5.2	SOFTWARE DE SIMULACIÓN MATLAB.....	157
5.2.1	DESCRIPCIÓN GENERAL.....	157
5.2.2	CARACTERÍSTICAS	157
5.2.3	GUIDE (GRAPHICAL USER INTERFACE DEVELOPMENT ENVIROMENT)	158
5.3	CRITERIOS DE DISEÑO.....	158
5.4	IMPLEMENTACIÓN DEL ALGORITMO DE CODIFICACIÓN MP3 EN MATLAB...	159
5.4.1	DIAGRAMAS DE FLUJO	160
5.4.1.1	Proceso de adquisición de la señal de audio de Microsoft *.wav	160
5.4.1.2	Codificación del archivo *.wav a un archivo de audio en formato comprimido *.mp3.....	161
5.4.1.2.1	Selección de parámetros necesarios para la codificación	161
5.4.1.2.2	Codificación.....	161
5.4.1.2.3	Banco de filtros polifásico	163
5.4.1.2.4	Análisis FFT	163
5.4.1.2.5	Transformada discreta del coseno modificada (MDCT)	164
5.4.1.2.6	Cuantización y codificación	164
5.4.1.2.7	Formato del flujo de bits MP3 válido	167
5.4.1.3	Proceso de adquisición de la señal de audio en formato comprimido *.mp3	168
5.4.1.4	Comparación entre archivo de audio original *.wav y archivo en formato de audio comprimido *.mp3.....	169
5.4.2	IMPLEMENTACIÓN DEL BANCO DE FILTROS	170
5.4.2.1	Filtro subbanda polifásico	170
5.4.3	IMPLEMENTACIÓN DEL MODELO PSICOACÚSTICO	176
5.4.3.1	Análisis FFT	176
5.4.3.2	Componentes tonales y no tonales	177
5.4.3.3	Reducción de componentes enmascarantes	180

5.4.3.4	Umbral de enmascaramiento individual	181
5.4.3.5	Umbral de enmascaramiento global.....	181
5.4.4	IMPLEMENTACIÓN DE LA TRANSFORMADA DISCRETA DEL COSENO MODIFICADA (MDCT).....	181
5.4.5	IMPLEMENTACIÓN DE LA ETAPA DE CUANTIZACIÓN Y CODIFICACIÓN	182
5.4.6	IMPLEMENTACIÓN DEL FORMATO DE LA TRAMA MP3	184
	REFERENCIAS BIBLIOGRÁFICAS	186

CAPÍTULO 6. 187

PRUEBAS DE FUNCIONAMIENTO Y RESULTADOS 187

6.1	INTRODUCCIÓN	187
6.2	PRUEBAS DE FUNCIONAMIENTO DEL CODIFICADOR	188
6.2.1	INTRODUCCIÓN	188
6.2.1.1	Especificaciones de hardware.....	188
6.2.1.2	Requerimientos de software	188
6.2.2	FUNCIONAMIENTO DEL CODIFICADOR.....	189
6.2.3	CARACTERÍSTICAS DE LOS ARCHIVOS CODIFICADOS	198
6.2.3.1	Archivos de audio	198
6.2.3.1.1	Ritmo Folklore	198
6.2.3.1.2	Ritmo Ranchera	199
6.2.3.1.3	Ritmo Rock.....	200
6.2.3.1.4	Ritmo Balada.....	200
6.2.3.1.5	Ritmo Instrumental.....	201
6.2.3.2	Análisis de las características del codificador.....	202
6.2.4	COMPORTAMIENTO DEL CODIFICADOR.....	204
6.2.4.1	Tiempo del proceso de codificación vs. Tasa de bits	205
6.2.4.2	Tamaño del archivo codificado vs. Tasa de bits	205
6.2.4.3	Radio de compresión vs. Tasa de bits	206
6.2.4.4	Tamaño del archivo comprimido vs. Duración del archivo sin comprimir	207
6.2.4.5	Tamaño del archivo comprimido vs. Tamaño del archivo sin comprimir	208
6.3	ANÁLISIS DE RESULTADOS DE LOS ARCHIVOS CODIFICADOS	208
6.3.1	PROGRAMAS UTILIZADOS	208
6.3.1.1	Cool Edit Pro 2.0	209
6.3.1.2	GoldWave 5.2	210
6.3.2	PROPIEDADES Y REPRODUCCIÓN DE LOS ARCHIVOS CODIFICADOS.....	211
6.3.2.1	Propiedades de los archivos codificados	211
6.3.2.2	Reproducción de los archivos codificados	213
6.3.3	ANÁLISIS DE LAS CARACTERÍSTICAS DE LOS ARCHIVOS CODIFICADOS	214
6.3.3.1	Análisis de los archivos sin comprimir.....	215
6.3.3.1.1	Análisis del archivo original en el dominio del tiempo.....	215
6.3.3.1.2	Análisis del archivo original en el dominio de la frecuencia.....	215

6.3.3.2	Análisis de los archivos comprimidos.....	218
6.3.3.2.1	Análisis del archivo codificado en el dominio del tiempo	218
6.3.3.2.2	Análisis del archivo codificado en el dominio de la frecuencia	218
	REFERENCIAS BIBLIOGRÁFICAS	220

CAPÍTULO 7. 221

CONCLUSIONES Y RECOMENDACIONES 221

7.1	CONCLUSIONES	221
7.2	RECOMENDACIONES.....	224

ANEXOS

ANEXO A A-1

MANUAL DE USUARIO A-1

A.1	DESCRIPCIÓN DEL PROGRAMA.....	A-1
A.2.1	REQUERIMIENTO DE SOFTWARE Y DE HARWARE.....	A-2
A.1.1.1	Especificaciones de hardware	A-2
A.1.1.2	Especificaciones de software	A-2
A.2.2	ACCESO AL PROGRAMA	A-3
A.2	EJECUCIÓN DEL PROGRAMA	A-11
A.2.1	VENTANA MPEG1_LayerIII	A-11
A.2.2	VENTANA Codificador_MP3	A-12
A.2.3	VENTANA Archivo_MP3.....	A-19
A.2.4	VENTANA Comparación_Wav_Mp3	A-23
A.3	MENÚS DEL CODIFICADOR.....	A-24
A.4	RESOLUCIÓN DE PROBLEMAS.....	A-33
A.5	RECOMENDACIONES.....	A-35

ANEXO B B-1

INSTALACIÓN DEL TOOLBOX MP3 EN MATLAB..... B-1

B.1	DESCRIPCIÓN DE LA INSTALACIÓN.....	B-1
B.2	INSTALACIÓN DEL TOOLBOX MP3	B-2
B.2.1	ALMACENAMIENTO DEL TOOLBOX.....	B-2
B.2.2	ADICIÓN DE LA RUTA DE ACCESO AL TOOLBOX	B-2
B.2.3	COMPROBACIÓN DE LA INSTALACIÓN DEL TOOLBOX	B-5

LISTA DE FIGURAS

Figura 1. 1 Rarefacción y compresión en una onda sonora.....	4
Figura 1. 2 Rangos y umbrales de audición	7
Figura 1. 3 Intensidad vs. Distancia	8
Figura 1. 4 Frecuencia del sonido	10
Figura 1. 5 Comparación de la gama de frecuencias que escuchan	11
Figura 1. 6 Frecuencia fundamental (f) y armónicas (2f, 3f)	12
Figura 1. 7 Dos ondas sinusoidales en diferente relación de fase.....	13
Figura 1. 8 Representación gráfica del espectro del sonido	14
Figura 1. 9 Onda de sonido emitida por un diapasón.....	15
Figura 1. 10 Clasificación de la voz por tesitura	19
Figura 1. 11 El oído humano	21
Figura 1. 12 Respuesta de frecuencia del canal auditivo.....	22
Figura 1. 13 Izquierda, los tres huesitos del oído medio. Derecha, su mecánica	24
Figura 1. 14 Onda viajera en la membrana basilar	25
Figura 1. 15 Esquema de las bandas críticas del sistema auditivo humano	26
Figura 1. 16 Enmascaramiento	29
Figura 1. 17 Enmascaramiento por un tono de 1 KHz	29
Figura 1. 18 Enmascaramiento temporal.....	30
Figura 1. 19 Diferencia entre sonido y audio	32
Figura 1. 20 Representación de una señal analógica (señal eléctrica)	32
Figura 1. 21 Audio analógico.....	33
Figura 1. 22 Muestreo de una señal analógica.....	35
Figura 1. 23 Aliasing en el dominio de la frecuencia	37
Figura 1. 24 Aliasing en el dominio del tiempo	37
Figura 1. 25 Cuantización con tres bits.....	38
Figura 1. 26 Cuantización y ruido de cuantización	39
Figura 1. 27 Izquierda, señal sin dither. Derecha, señal con dither	40
Figura 1. 28 Cuantización uniforme	40
Figura 1. 29 Cuantización no uniforme	41
Figura 1. 30 Cuantizador ley A.....	42
Figura 1. 31 Ejemplo de codificación de una señal	44
Figura 1. 32 La computadora en sistemas de audio digital	46
Figura 1. 33 Tarjeta de sonido	48
Figura 2. 1: Compresión utilizada en transmisión.....	57
Figura 2. 2: Compresión utilizada en almacenamiento	58
Figura 2. 3: Ejemplo de codificación Huffman	66
Figura 2. 4: Mínimo umbral auditivo (ambiente silencioso)	73
Figura 2. 5: Esquema de un codificador perceptual de audio	74
Figura 2. 6: Diagrama de bloques de un codificador de sub-bandas.....	76
Figura 2. 7: Diagrama de bloques de un codificador ADPCM.....	77

Figura 3. 1: iPod Híbrido.....	101
Figura 3. 2: iPod Shuffle de 2 GB	102
Figura 4. 1: Creadores del formato MP3,.....	104
Figura 4. 2: Codificador según la norma ISO/IEC 11172-3.....	109
Figura 4. 3: Diagrama de bloques de un codificador MP3	111
Figura 4. 4: Banco de filtros polifásicos	120
Figura 4. 5: División en subbandas de las muestras de audio.....	123
Figura 4. 6: Diagrama de bloques de las operaciones de la MDCT	124
Figura 4. 7: Ventanas de datos usadas durante el proceso MP3	126
Figura 4. 8: Codificación Huffman.....	130
Figura 4. 9: Algoritmo Huffman	131
Figura 4. 10: Ejemplo de uso de la reserva de bits (bit reservoir).....	131
Figura 4. 11: Decodificador según la norma ISO 11172-3.....	134
Figura 4. 12: Formato de la trama MP3	135
Figura 4. 13: Encabezado de la trama MP3	136
Figura 4. 14: Formato de la información secundaria	142
Figura 4. 15: Información secundaria para cada gránulo	143
Figura 4. 16: Campos incluidos en los datos principales	146
Figura 4. 17: Ejemplo ilustrativo	148
Figura 4. 18: Curiosos reproductores mp3	151
Figura 5. 1: Diagrama de flujo para la adquisición de la señal de audio de Microsoft *.wav	160
Figura 5. 2: Diagrama de flujo para la selección de los parámetros necesarios para la codificación	161
Figura 5. 3: Diagrama de flujo para la codificación del archivo *.wav a un archivo de audio en formato comprimido *.mp3	162
Figura 5. 4: Diagrama de flujo para la etapa correspondiente al banco de filtros	163
Figura 5. 5: Diagrama de flujo para el análisis FFT del Modelo psicoacústico I.....	163
Figura 5. 6: Diagrama de flujo para la transformada discreta del coseno modificada MDCT.....	164
Figura 5. 7: Diagrama de flujo del lazo de iteración para la cuantización no uniforme ...	165
Figura 5. 8: Diagrama de flujo del lazo de iteración para el ciclo externo	166
Figura 5. 9: Diagrama de flujo del lazo de iteración para el ciclo interno	166
Figura 5. 10: Diagrama de flujo para el formato de la trama MP3 y obtención del flujo de bits MP3 válido bajo el estándar internacional ISO/IEC 11172-3	167
Figura 5. 11: Diagrama de flujo para la lectura del archivo de audio comprimido *.mp3	168
Figura 5. 12: Diagrama de flujo para la comparación del archivo .wav y el archivo de audio comprimido .mp3	169
Figura 5. 13: Artefactos de audio presentes en la señal de audio codificada a 160 Kbps	173

Figura 5. 14: Artefactos de audio presentes en la señal de audio codificada a 320 Kbps	173
Figura 5. 15: a) Ventana de análisis recomendada en el estándar b) Ventana modificada en el diseño e implementada para velocidades superiores a 128 Kbps	174
Figura 5. 16: Eliminación de los artefactos de audio presentes en la señal de audio codificada a 160 Kbps	175
Figura 5. 17: Eliminación de los artefactos de audio presentes en la señal de audio codificada a 320 Kbps	175
Figura 6. 1: Primera ventana del codificador (Ventana de la carátula)	189
Figura 6. 2: Segunda ventana del codificador, ventana "Codificador_MP3"	190
Figura 6. 3: Opción para abrir archivo sin comprimir *.wav	190
Figura 6. 4: Selección y apertura del archivo *.wav	190
Figura 6. 5: Señales de archivo de audio de entrada, indicaciones y parámetros de codificación	191
Figura 6. 6: Elección de la tasa de bits	191
Figura 6. 7: Parámetros mostrados después de seleccionar la tasa de bits	192
Figura 6. 8: Ingreso del tiempo que se desea codificar del archivo original (no codificado)	192
Figura 6. 9: Ejecución de la codificación	193
Figura 6. 10: Finalización del proceso de codificación	193
Figura 6. 11: Activación del botón "Ver MP3" para activar la ventana "Archivo_MP3"	194
Figura 6. 12: Ventana "Archivo_MP3"	195
Figura 6. 13: Opción para abrir archivo codificado *.MP3	195
Figura 6. 14: Selección y apertura del archivo del archivo codificado (*.mp3)	195
Figura 6. 15: Características del archivo codificado	196
Figura 6. 16: Resumen de la codificación	196
Figura 6. 17: Comparación de las características de los archivos sin codificación (izquierda) y con codificación (derecha)	197
Figura 6. 18: Tiempo del proceso de codificación vs. Tasa de bits	205
Figura 6. 19: Tiempo del archivo codificado vs. Tasa de bits	205
Figura 6. 20: Radio de compresión vs. Tasa de bits	206
Figura 6. 21: Tamaño del archivo comprimido vs. Duración del archivo sin comprimir... 207	
Figura 6. 22: Tamaño del archivo comprimido vs. Tamaño del archivo sin comprimir 208	
Figura 6. 23: Ventana principal de Cool Edit Pro 2.0	209
Figura 6. 24: Ventana principal de GoldWave 5.2	210
Figura 6. 25: Características de archivo de audio codificado (Explorador de Windows) 211	
Figura 6. 26: Ventana principal de Cool Edit Pro 2.0 y venta de propiedades del audio 212	
Figura 6. 27: Características del archivo de audio codificado (Programa: Cool Edit Pro 2.0)..... 212	
Figura 6. 28: Reproducción del archivo de comprimido (Programa: Cool Edit Pro 2.0).. 213	
Figura 6. 29: Reproducción del archivo de comprimido (Programa: GoldWave 5.2) 214	
Figura 6. 30: Archivo de audio original en el dominio del tiempo..... 215	
Figura 6. 31: Opción para activar la ventana del análisis de frecuencia..... 215	
Figura 6. 32: Análisis en frecuencia del archivo audio original (en 2,199 segundos)..... 216	

Figura 6. 33: Análisis en frecuencia del archivo audio en varios instantes de tiempo	216
Figura 6. 34: Análisis de frecuencia del archivo audio obtenido con el codificador y comparando con el análisis de frecuencia obtenido con Cool Edit Pro 2.0	217
Figura 6. 35: Archivo de audio codificado en el dominio del tiempo	218
Figura 6. 36: Análisis en frecuencia del archivo de audio codificado obtenido con el codificador y comparando con el análisis de frecuencia obtenido con Cool Edit Pro 2.0.....	219

Figura A. 1: Archivos del Codificador MPEG-1 Layer III	A-1
Figura A. 2: Procesador de la computadora utilizada para correr el programa	A-2
Figura A. 3: Especificaciones de software para el codificador MPEG-1 Layer III.....	A-3
Figura A. 4: Archivos *.fig y *.m de MatLab	A-3
Figura A. 5: Ventana principal de Matlab y comando para acceder a las herramientas de la interfaz gráfica de usuario	A-4
Figura A. 6: Ventana de inicio rápido de la interfaz gráfica de usuario.....	A-5
Figura A. 7: Búsqueda de la interfaz gráfica de interés.....	A-5
Figura A. 8: Archivos *.fig correspondientes a las interfaces gráficas	A-6
Figura A. 9: Archivo *.fig correspondiente a la interfaz gráfica	A-6
Figura A. 10: Actualización del directorio y ruta de acceso al archivo	A-7
Figura A. 11: Acceso a la ventana de carátula del codificador mediante la utilización del ambiente gráfico	A-7
Figura A. 12: Apertura de archivos	A-8
Figura A. 13: Búsqueda y selección del archivo *.m perteneciente al codificador	A-8
Figura A. 14: Apertura del archivo MPEG1_LayerIII.m	A-9
Figura A. 15: Actualización del directorio y ruta de acceso al archivo	A-9
Figura A. 16: Acceso a la ventana de carátula del codificador mediante la utilización del editor de archivos .m	A-10
Figura A. 17: Archivo MPEG1_LayerIII.m.....	A-10
Figura A. 18: Carátula del codificador.....	A-11
Figura A. 19: Mensaje al seleccionar el botón “SALIR”	A-12
Figura A. 20: Segunda ventana del codificador, ventana “Codificador_MP3”	A-12
Figura A. 21: Opción para abrir archivo sin comprimir *.wav	A-13
Figura A. 22: Selección y apertura del archivo *.wav	A-13
Figura A. 23: Procesamiento de archivo de audio de entrada	A-14
Figura A. 24: Botones de reproducción multimedia.....	A-14
Figura A. 25: Señales de archivo de audio de entrada, indicaciones y parámetros de codificación	A-15
Figura A. 26: Elección de la tasa de bits.....	A-16
Figura A. 27: Parámetros mostrados después de seleccionar la tasa de bits	A-16
Figura A. 28: Ingreso del tiempo de codificación.....	A-17
Figura A. 29: Ejecución de la codificación	A-17
Figura A. 30: Finalización del proceso de codificación.....	A-18
Figura A. 31: Activación del botón “Ver MP3” para activar la ventana “Archivo_MP3” .	A-19
Figura A. 32: Tercera ventana del codificador, ventana “Archivo_MP3”	A-20
Figura A. 33: Opción para abrir archivo codificado *.MP3	A-20

Figura A. 34: Selección y apertura del archivo del archivo codificado (*.mp3).....	A-21
Figura A. 35: Características del archivo codificado	A-21
Figura A. 36: Características del archivo codificado	A-22
Figura A. 37: Resumen de la codificación	A-22
Figura A. 38: Comparación de las características de los archivos sin codificación (izquierda) y con codificación (derecha)	A-23
Figura A. 39: Menú “Abrir” y opción “Archivos .WAV”	A-24
Figura A. 40: Menú “Abrir” y opción “Archivos .MP3”	A-24
Figura A. 41: Menú “Estándar ISO/IEC 11172-3” y opciones correspondientes	A-25
Figura A. 42: Menú “Guardar Imágenes” y opciones correspondientes	A-25
Figura A. 43: Almacenamiento de la imagen de la señal correspondiente	A-26
Figura A. 44: Menú “Créditos” y submenú “Autores”	A-26
Figura A. 45: Menú “Créditos” y submenú “Autores”	A-27
Figura A. 46: Biografía resumida de autores	A-27
Figura A. 47: Menú “ADICIONAL*” y opción “Efectos de audio”	A-28
Figura A. 48: Ventana principal de los efectos de audio	A-28
Figura A. 49: Apertura de archivos *.WAV o *.MP3 para aplicar efectos de audio	A-28
Figura A. 50: Apertura del archivo de audio para aplicar el efecto deseado	A-29
Figura A. 51: Procesamiento del archivo original (archivo sin efectos)	A-29
Figura A. 52: Efectos de audio disponibles	A-30
Figura A. 53: Archivos de audio con efectos	A-31
Figura A. 54: Efectos: eco, inversión temporal y variación de la velocidad (de izquierda a derecha).....	A-31
Figura A. 55: Efectos:	A-32
Figura A. 56: Utilización del ecualizador	A-32
Figura A. 57: Grabación del archivo de audio con efectos	A-33
Figura A. 58: Error en la frecuencia de muestreo	A-34
Figura A. 59: Error al ingresar un tiempo muy pequeño	A-34
Figura A. 60: Error al ingresar un tiempo muy grande	A-34
Figura A. 61: Mensaje de error al introducir un valor no numérico (Ejemplo: letra “A”)	A-35
Figura A. 62: Uso de recursos (CPU y memoria RAM)	A-35
Figura B. 1: Almacenamiento del toolbox en el directorio correspondiente	B-2
Figura B. 2: Direccionamiento de la ruta donde se encuentra el toolbox	B-3
Figura B. 3: Adición de carpetas y subcarpetas pertenecientes al toolbox	B-3
Figura B. 4: Direccionamiento del toolbox requerido	B-4
Figura B. 5: Almacenamiento y actualización de cambios	B-4
Figura B. 6: Comprobación del funcionamiento del toolbox	B-5

LISTA DE TABLAS

Tabla 1. 1 Niveles de presión sonora en decibeles e intensidad sonora en watts	7
Tabla 1. 2 Niveles de intensidad de la voz humana	18
Tabla 1. 3 Frecuencias de la voz por tesitura	19
Tabla 1. 4 Escala de Barks, para estimación de las bandas críticas	28
Tabla 1. 5 Espacio en disco para la grabación de audio digital	50
Tabla 1. 6 Formatos de fichero autodescriptivos.....	52
Tabla 2. 1: Símbolos y sus respectivas frecuencias (Codificación Huffman)	65
Tabla 2. 2: Características técnicas de códecs de audio	71
Tabla 2. 3: Ejemplos de códecs de audio con y sin pérdida	72
Tabla 2. 4: Tamaños del archivo WAV	79
Tabla 2. 5: Características de los estándares MPEG.....	82
Tabla 2. 6: Partes del estándar MPEG-1	84
Tabla 2. 7: Partes del estándar MPEG-2.....	85
Tabla 2. 8: Cuadro comparativo de las características de los algoritmos de compresión de audio	89
Tabla 3. 1: Deformación de un sistema de audio	97
Tabla 3. 2: Características del iPod Shuffle de 2 GB	102
Tabla 4. 1: Razón de compresión de acuerdo al ancho de banda, modo, bitrate y radio de compresión	119
Tabla 4. 2: Valores posibles para los bits 18 y 17 para indicar cuál capa de MPEG-1 se utiliza.....	136
Tabla 4. 3: Codificación para la tasa de bits	137
Tabla 4. 4: Codificación para la frecuencia de muestreo	137
Tabla 4. 5: Codificación para el modo de canal.....	138
Tabla 4. 6: Codificación de la extensión al modo	140
Tabla 4. 7: Codificación de la información de énfasis	140
Tabla 5. 1: Coeficientes C_i de la ventana de análisis según el estándar.....	172
Tabla 5. 2: Frecuencias, Tasas de Bandas Críticas y Umbral Absoluto	179
Tabla 5. 3: Límites de las bandas críticas.....	180
Tabla 6. 1: Características del archivo monofónico sin codificación, ritmo Folklore.....	198
Tabla 6. 2: Características de los archivos codificados, ritmo Folklore	199
Tabla 6. 3: Características del archivo monofónico sin codificación, ritmo Ranchera.....	199
Tabla 6. 4: Características de los archivos codificados, ritmo Ranchera	199

Tabla 6. 5: Características del archivo monofónico sin codificación, ritmo Rock	200
Tabla 6. 6: Características de los archivos codificados, ritmo Rock	200
Tabla 6. 7: Características del archivo monofónico sin codificación, ritmo Balada	200
Tabla 6. 8: Características de los archivos codificados, ritmo Balada	201
Tabla 6. 9: Características del archivo monofónico sin codificación, ritmo Instrumental	201
Tabla 6. 10: Características de los archivos codificados, ritmo Instrumental	201
Tabla 6. 11: Características del archivo original (no codificado), ritmo balada	204
Tabla 6. 12: Características de los archivos codificados (comprimidos), ritmo balada ...	204
Tabla 6. 13: Características del archivo original (no codificado).....	206
Tabla 6. 14: Características de los archivos codificados a 128 [Kbps]	207

RESUMEN

El presente proyecto de titulación se constituye como un referente técnico para el desarrollo de temas orientados al procesamiento digital de señales y especialmente a la compresión de audio, ya que, en la actualidad casi todos los sistemas emplean formatos de audio comprimidos, sin pérdidas significativas en la calidad.

El enfoque de este proyecto es el siguiente: en los primeros 4 capítulos se presenta una descripción teórica de los temas directamente involucrados con el proyecto.

En el primer capítulo se realiza un estudio del sonido, donde se incluyen aspectos como su definición, física y las fuentes que lo generan. Se presentan también aspectos relacionados con el sistema fonatorio, la psicoacústica y el funcionamiento del sistema auditivo humano.

Además, este capítulo contiene un análisis del audio digital y la digitalización de señales analógicas, diferenciando el sonido del audio y presentando características tanto del audio analógico como del audio digital.

En el segundo capítulo se describe el proceso de compresión de audio, justificando la necesidad para la compresión, su utilidad y las técnicas desarrolladas para comprimir, de las cuales, se detallan la compresión con y sin pérdida de información. Dentro de la compresión con pérdida de información (lossy) se presenta la codificación por transformada, mientras tanto, en la compresión sin pérdida de información (lossless) se presenta la codificación estadística con la codificación Huffman.

Además, este capítulo presenta un análisis de la codificación perceptual, la codificación de sub-bandas y de los algoritmos de compresión de audio más utilizados como: WAVE Audio Format, ADPCM Audio Compression, FLAC, MPEG-4 Audio Lossless Coding (ALS) y Advanced Audio Coding (AAC).

En el tercer capítulo se hace referencia a los sistemas de alta fidelidad y su relación con los formatos de audio comprimidos, haciendo énfasis a la relación con el formato de audio MP3.

En el cuarto capítulo se muestra al detalle el formato MP3 basado en los aspectos principales del estándar ISO/IEC 11172-3 (MPEG-1 Capa 3), donde se incluyen aspectos como la codificación y el formato de las tramas MP3. Se presentan también curiosidades e innovaciones que ha tenido el formato de audio en la actualidad.

En el capítulo cinco se discuten los parámetros de diseño y se describe de manera específica como fueron implementados los diferentes bloques que componen el codificador MP3 según la norma ISO/IEC 11172-3 para compresión de audio en software de simulación Matlab, mediante el uso de técnicas digitales para el procesamiento de señales y la utilización de la interfaz gráfica GUI (Graphical User Interface) de Matlab para crear un ambiente amigable al usuario.

En el capítulo seis se indican las pruebas realizadas al codificador MP3 implementado en Matlab para demostrar la funcionalidad del software desarrollado, así como, las pruebas mediante la comparación de archivos comprimidos con archivos originales en relación al tamaño del archivo y sus características en tiempo y frecuencia, con la ayuda de software comerciales, como el Cool Edit Pro 2.0 y Gold Wave v5.52.

En el capítulo siete se presentan las conclusiones y recomendaciones del proyecto de titulación.

Finalmente como complemento se anexan materiales de apoyo relacionados con la implementación del codificador MP3 basados en la norma ISO/IEC 11172-3 y un manual de usuario del codificador MP3 implementado en Matlab.

PRESENTACIÓN

En la actualidad los sistemas multimedia están teniendo una gran importancia y una amplia variedad de dispositivos incluyen entre sus características la reproducción de audio. Cada vez más, estos dispositivos admiten una mayor variedad de formatos y soportes de almacenamiento como discos duros, CDs, DVDs, memorias Flash, etc.

Muchos de los dispositivos poseen una capacidad de almacenamiento limitado lo que hace necesario la utilización de formatos comprimidos que permitan almacenar estos elementos multimedia con una calidad aceptable y en muchos casos sin pérdida aparente de ésta.

De la necesidad de obtener un formato de audio comprimido de alta calidad surgió el algoritmo MP3 para compresión de audio.

El presente proyecto de titulación tiene como objetivo principal desarrollar e implementar un codificador MP3 en software de simulación Matlab, completamente detallado y considerando los aspectos técnicos del formato de audio comprimido MP3, bajo el estándar ISO/IEC 11172-3, que sea capaz de comprimir un archivo de audio en formato WAV de Microsoft, en un archivo de audio en formato MP3 (MPEG-1 Capa III), con tasas de transferencia entre 96 a 320 Kbps y una frecuencia de muestreo de 44.100 Hz, pretendiendo lograr con esto una reducción significativa de la cantidad de bits para obtener la máxima transferencia (almacenamiento) de la información con un mínimo número de bits posible, lo que se traduce en un ahorro en el ancho de banda y la optimización del uso del espectro; considerando que al reproducir el archivo de audio comprimido, la diferencia en comparación con el archivo de audio original sea mínima. La implementación del modelo se lo hace en un ambiente amigable al usuario utilizando la interfaz gráfica GUI (Graphical User Interface) de Matlab, permitiendo crear una interacción entre el usuario y el codificador MP3 implementado en Matlab.

CAPÍTULO 1.

ESTUDIO DEL SONIDO

1.1 EL SONIDO

1.1.1 INTRODUCCIÓN

Mucho de lo que aprendemos de nuestro mundo nos llega a través de nuestro sentido del oído. El oír es importante no solamente para aprender del mundo, sino también para comunicarse con otros humanos, y con los animales. Podemos identificar a personas por el sonido único de su voz o por el sonido de sus pisadas. Podemos identificar a los animales por sus sonidos: el canto de los pájaros, el rugido de los leones, el zumbido de los insectos, etc.

El sonido es una parte importante de nuestra vida. Es uno de los primeros estímulos a que reaccionan los recién nacidos, y su presencia o ausencia nos forma y afecta durante nuestra vida.

Los sonidos son creados por vibraciones. Cuando un objeto vibra, se ponen en movimiento moléculas de aire. Este movimiento se parece a ondas. Las ondas se alejan de la fuente de vibración como las pequeñas olas en una laguna cuando se tira una piedra al agua. Las vibraciones que se producen en el aire se captan por nuestro oído y se escuchan como sonidos. Cuando hablamos, las cuerdas vocales vibran, creando así ondas sonoras que, al llegar a nuestro oído, se escuchan como habla.

El hombre, así como la gran mayoría de las especies animales (y probablemente algunas vegetales), ha desarrollado la capacidad de usar el sonido para recibir y transmitir información. El sentido del oído ofrece ventajas con respecto a los otros sentidos. No necesita conexión directa como la vista, ni contacto físico como el

tacto o el gusto, permanece activo durante el sueño y es mucho más rápido que el olfato a los fines de detectar situaciones de riesgo. Por otra parte, es el único sentido que actúa de una forma analítica, es decir, es capaz de descomponer el sonido en sus componentes (frecuencias) individuales, lo cual le permite distinguir unos sonidos en presencia de otros. Además, la voz humana es única en su habilidad de expresar ideas abstractas.

El silencio también es un medio expresivo. No es sólo la ausencia de sonido. Desde el punto de vista comunicativo, es una forma del sonido.

1.1.2 DEFINICIÓN

Para la mayoría de nosotros, el sonido es un fenómeno muy familiar, porque lo oímos todo el tiempo. No obstante, cuando nosotros intentamos definir el sonido, nos encontramos que podemos acercarnos a éste concepto desde dos puntos de vista diferentes, y terminamos con dos definiciones [2], como sigue:

Una definición científica: *El sonido es una perturbación física en un medio elástico (gaseoso, líquido o sólido). Se propaga en el medio como una onda de presión por el movimiento de átomos o moléculas.*

Una definición psicoacústica: *El sonido es la sensación detectada por nuestros oídos e interpretada por nuestro cerebro de una cierta manera.*

Como vemos, los rasgos comunes en las definiciones refieren a:

1. Un elemento que vibra (fuente generadora del sonido)
2. Un medio transmisor de las ondas
3. Un estímulo sobre el sentido auditivo (receptor u oyente del sonido)

Los dos primeros elementos se refieren al sonido como fenómeno físico. El sonido como fenómeno físico nos interesa en nuestro proyecto para comprender lo que hacemos cuando manipulamos sonido digital. El tercero podemos analizarlo

desde el punto de vista fisiológico, psicológico y comunicacional, y es el que nos interesa particularmente en el proceso de elaborar comunicación multimedia.

De las definiciones también se desprende que *no toda vibración del aire* puede ser considerada sonido: sólo aquella capaz de estimular el sentido auditivo. Existen umbrales superiores e inferiores de frecuencia, por arriba y por abajo de los cuales las vibraciones no pueden ser registradas por el oído.

Nosotros normalmente oímos el sonido cuando se propaga a través del aire y golpea el diafragma en nuestros oídos. Sin embargo, el sonido puede propagarse en muchos medios de comunicación diferentes. Los buenos aisladores de sonido son raros, y el mejor aislador es el vacío, donde no hay ninguna partícula para vibrar y propagar la perturbación, es decir, no existe un medio cuya presión cambie.

1.1.3 FÍSICA DEL SONIDO

Para que haya sonido es preciso que un cuerpo material vibre, que haya un soporte material que propague esas vibraciones y, por último, que las mismas sean capaces de impresionar los nervios auditivos del oído. Por lo tanto, el sonido también puede ser considerado una onda, aunque su frecuencia puede cambiar todo el tiempo.

Las ondas de sonido son longitudinales, el desplazamiento de las partículas es paralelo a la dirección de propagación de la onda y tiene un carácter similar al de las perturbaciones longitudinales en un resorte. Cuando esta onda alcanza alguna superficie (como el tímpano del oído humano o la membrana de un micrófono), produce una vibración en dicha superficie por resonancia. De esta forma, la energía acústica es transferida desde la fuente al receptor manteniendo las características de la vibración original. En contraste, las ondas electromagnéticas y ondas del océano son ondas transversales, sus ondulaciones son perpendiculares a la dirección de la onda.

1.1.3.1 Ondas de sonido

El sonido es una señal producida por una fuente en vibración. Esta vibración perturba las moléculas de aire adyacentes a la fuente en sincronismo con la vibración, creando zonas donde la presión del aire es menor a la presión atmosférica (rarefacción o enrarecimiento) y zonas donde la presión del aire es mayor a la presión atmosférica (compresión). Estas zonas de rarefacción y compresión, generan una onda de sonido la cual viaja a través del aire.

Las ondas en general poseen ciertas propiedades comunes. Las ondas transportan información de un lugar a otro y también transportan energía. Las ondas son parametrizables, lo que quiere decir que pueden describirse de acuerdo a algunos pocos parámetros.

Los cuatro parámetros más comunes son: amplitud, periodo, fase y forma de onda. Existen algunos otros parámetros tales como frecuencia, espectro o intensidad que pueden derivarse de los parámetros mencionados anteriormente [6]. En la figura 1.1 se muestran algunos de ellos.

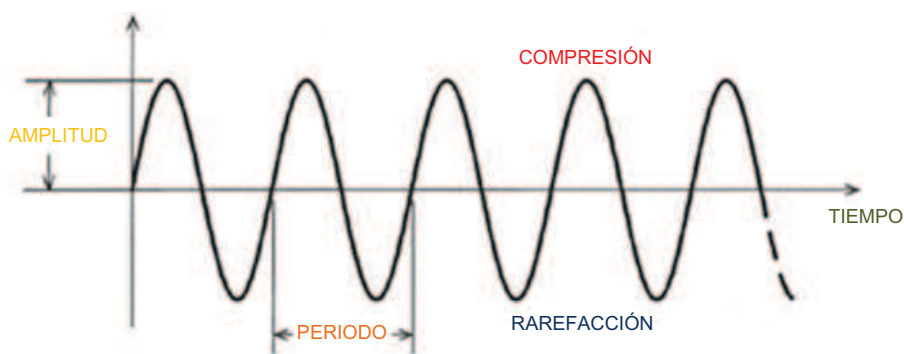


Figura 1. 1 Rarefacción y compresión en una onda sonora

1.1.3.2 Medición del sonido

El sonido en su dimensión física es medible o cuantificable. Existen parámetros que pueden ser medidos en forma precisa como su intensidad y otros pueden ser estimados con mayor o menor precisión como la frecuencia o su forma de onda.

El sonido fundamentalmente es una onda de presión, y la presión (P) puede medirse de la siguiente forma:

$$P = F/a \quad (\text{Ec. 1.1})$$

Donde, F es una fuerza ejercida y a es una unidad de área determinada. La presión corresponde a la fuerza ejercida en forma perpendicular a una superficie dividida por el área de ésta superficie.

Las ondas de sonido transportan energía. La energía es una medida abstracta de mucha utilidad en la física, dado que se define de tal forma que en un sistema cerrado la energía siempre es constante, principio que se conoce como la conservación de la energía. La energía se mide en unidades de masa por velocidad al cuadrado y usualmente se mide en [Joules]. Una importante medida del sonido es su potencia, la cual corresponde a la energía total por unidad de tiempo y se mide en [Joules/seg].

1.1.3.3 Velocidad del sonido

El sonido necesita de un medio para propagarse. El sonido puede viajar a través de objetos sólidos, líquidos o gases. Su velocidad es proporcional a la densidad del medio por el cual viaja. En aire, al nivel del mar y a 20° Celsius (68° Fahrenheit), la velocidad del sonido es 343,8 m/s (metros por segundo), y en agua, es de 1.500 m/s. La velocidad del sonido es independiente de su intensidad y su intensidad decae con la distancia en forma inversamente proporcional. Las frecuencias altas se propagan más rápidamente que las bajas frecuencias.

1.1.3.4 Longitud de onda del sonido

La longitud de onda (λ) es la distancia recorrida por la onda sonora. El rango perceptible por la audición humana varía entre menos de 2 cm (una pulgada) y

aproximadamente los 17 metros (56 pies). La longitud de onda es inversamente proporcional a la frecuencia del sonido.

$$\lambda = c/f \quad (\text{Ec. 1.2})$$

Donde, c es la velocidad del sonido (343,8 m/s) y f es la frecuencia.

1.1.3.5 Amplitud

La amplitud de un sonido corresponde a la magnitud del cambio, sea este positivo o negativo, de la presión atmosférica causado por la compresión y rarefacción de las ondas acústicas. Esta cantidad es un indicador de la magnitud de energía acústica de un sonido y es el factor que determina que tan fuerte se percibe un sonido. Por lo tanto, la amplitud refiere a la altura de la onda sonora y significa la intensidad o volumen del sonido. Amplitud cero equivale a silencio, amplitudes pequeñas a sonidos leves y amplitudes grandes a sonidos fuertes o intensos.

La amplitud es la magnitud de la presión de aire medida en Pascales (Pa), y el volumen es la percepción de esa amplitud por el oído humano, algunas veces es llamado también sonoridad. El volumen del sonido depende de la amplitud de onda y ésta varía según el valor de la presión que lo genera.

El oído de una persona sana puede detectar una presión sonora cuya amplitud sea 20 millonésimas de Pascal (20 μPa), lo que representa unas 5.000 millones de veces menor que la presión atmosférica normal [2]. Sorprendentemente el oído humano puede tolerar presiones sonoras superiores a un millón de veces más altas. Si midiéramos el sonido en Pa utilizaríamos unas cantidades enormes e inimaginables. Para evitar esto se utiliza otra escala como es el decibelio (dB). El decibelio no es una unidad de medida absoluta (no tiene unidades físicas). Es una relación logarítmica entre una cantidad medida y un nivel de referencia acordado (20 μPa). Esto permite que la escala de decibelios se aproxime mucho mejor a la percepción humana de sonoridad relativa que la escala en Pa, ya que, el oído reacciona a un cambio logarítmico de nivel de presión sonora.

En términos de nivel de presión sonora, los sonidos audibles van desde el umbral de audición de 0 dB hasta el umbral del dolor que está alrededor de 130 dB. El cambio más pequeño que podemos percibir corresponde a 3 dB y cada 10 dB el sonido parece, subjetivamente, que se dobla (físicamente se dobla cada 6dB). Los sonidos superiores a 110 dB producen sensación dolorosa y la exposición permanente a esos niveles provoca la disminución de la capacidad auditiva. En la figura 1.2 se pueden apreciar los rangos y umbrales de audición.

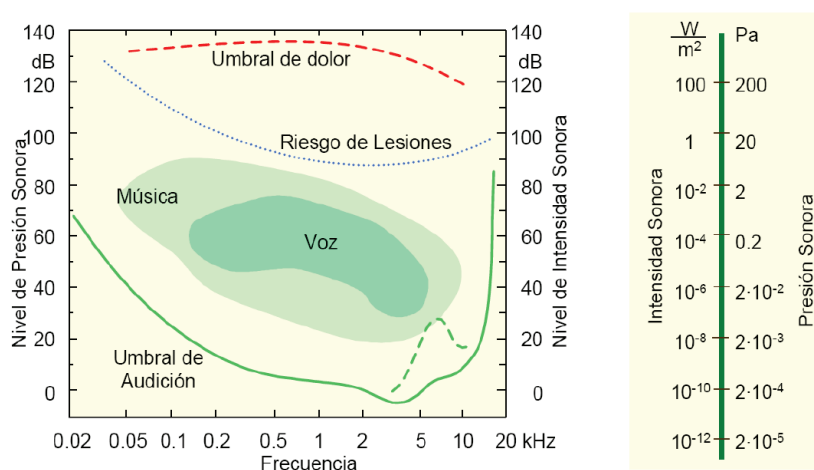


Figura 1. 2 Rangos y umbrales de audición

Nivel de presión sonora (dB)	Intensidad sonora (W/m^2)	Sonido
195	25×10^6 a 40×10^6	Cohete Saturno
170	10^5	Jet
160	10^4	Disparo de arma de fuego
150	10^3	Perforación instantánea del tímpano
140	10^2	Avión a reacción (a 30m de distancia)
130	10	Orquesta de 75 instrumentos
120	1	Taladro neumático, avión despegando
110	10^{-1}	Remachadora
100	10^{-2}	Automóvil de carrera
90	10^{-3}	Tren subterráneo
80	10^{-4}	Tráfico pesado
70	10^{-5}	Canto
60	10^{-6}	Tráfico normal
50	10^{-7}	Conversación normal
40	10^{-8}	Canto de pájaros
30	10^{-9}	Casa tranquila
20	10^{-10}	Conversación en voz baja
10	10^{-11}	Ruido de hojas
0	10^{-12}	Estudio de grabación en silencio

Tabla 1. 1 Niveles de presión sonora en decibeles e intensidad sonora en watts [4]

1.1.3.5.1 Intensidad

La intensidad del sonido caracteriza la razón a la cual la energía es entregada en la sensación audible asociada con la amplitud. Suponiendo una fuente puntual que irradia energía uniforme en todas las direcciones, entonces la presión sonora varía en forma inversamente proporcional a la distancia medida desde la fuente y la intensidad cambia en forma inversamente proporcional al cuadrado de la distancia, como se puede apreciar en la figura 1.3. Si esta distancia es r , entonces se tiene que:

$$I = \mathcal{P} / 4\pi r^2 \quad (\text{Ec. 1.3})$$

Donde, \mathcal{P} es la potencia sonora.

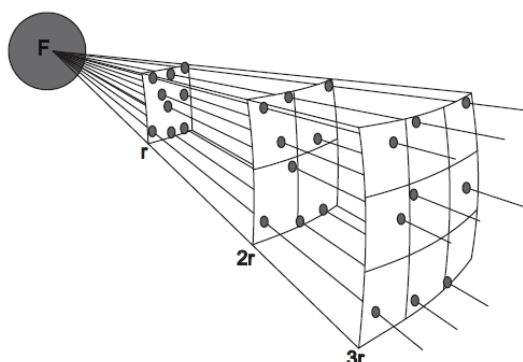


Figura 1. 3 Intensidad vs. Distancia

La intensidad se mide en $[W/m^2]$. La pérdida de intensidad al incrementarse la distancia es de $\text{dB} = 20\log(r_1/r_2)$. Al doblarse la distancia, se experimenta una pérdida en intensidad de 6 dB.

La presión e intensidad se relacionan a través de la siguiente ecuación:

$$P = \sqrt{I\beta c} \quad (\text{Ec. 1.4})$$

Donde β equivale a densidad del medio y c es la velocidad del sonido en el aire.

El sistema auditivo humano puede detectar un inmenso rango de intensidades desde 10^{-12} [W/m²] a 1 [W/m²]. En términos prácticos, medir la intensidad de sonido en [W/m²] resulta inmanejable debido a su enorme rango, por lo que una escala logarítmica de medición de intensidad resulta mucho más apropiada. La intensidad (sound level) del sonido se mide en decibeles o decibelios (dB).

$$1dB = 10 \log_{10}(I_1/I_0) \quad (\text{Ec. 1.5})$$

I_0 se escoge típicamente como 10^{-12} [W/m²]. Un incremento de 10 dB equivale a un incremento de la intensidad del sonido de un orden de magnitud. Un incremento de 3dB equivale a doblar la intensidad y un incremento de 1dB representa un 25% de incremento en la intensidad. La intensidad es proporcional al cuadrado de la presión.

$$1dB = 20 \log_{10}(P_1/P_0) \quad (\text{Ec. 1.6})$$

Donde, $P_0 = 2 \times 10^{-5}$ [Pa].

Esta medida se conoce como dB SPL (sound pressure level). 0 dB se escoge para el umbral de audición, el sonido más tenue que puede ser detectado. En los equipos de audio suele usarse el dBVU (volume unit), donde 0 dB corresponde al máximo nivel de audio posible sin tener distorsión (clipping). El área sobre los 0 dB en este caso se conoce como headroom.

1.1.3.6 Frecuencia

En el caso de una onda periódica, el patrón que se repite corresponde a un ciclo. La duración de cada uno de los ciclos de una onda se conoce como periodo. La tasa a la cual los ciclos de una onda periódica se repiten se conoce como frecuencia y por lo general se mide en ciclos por segundo o Hertz (Hz). Por lo tanto, la frecuencia es igual al número de ciclos de la onda sonora en la unidad de tiempo. Matemáticamente, la frecuencia es el inverso del periodo, por lo tanto un periodo de 1 ms (milisegundos) tiene una frecuencia de 1.000 Hz.

$$f = 1/T \quad (\text{Ec. 1.7})$$

Donde, f es la frecuencia y T es el período.

En términos de su contenido de frecuencia, un sonido puede poseer lo que se denomina una frecuencia fundamental, comúnmente denotada por f_0 , que usualmente corresponde a la frecuencia más baja y de mayor amplitud presente en el espectro. Es la frecuencia fundamental de una onda la que determina en gran medida su altura musical, la cual es una medida perceptual. Las señales aperiódicas no poseen una frecuencia fundamental fácilmente determinable, dado que nada se repite en forma periódica. La estimación de la f_0 para señales complejas es en sí un problema bastante complicado, además, la ausencia de una fundamental hace que este tipo de ondas se perciba musicalmente como ruido.

Los sonidos comprendidos entre 20 Hz a 20.000 Hz son los percibidos por el oído humano y se denominan *frecuencias audibles*. Los sonidos inferiores a 20 Hz se denominan *infrasonidos* y los superiores a 20.000 Hz *ultrasonidos*. Se distinguen tres gamas de frecuencias en función del oído humano [7], como se puede apreciar en la figura 1.4.

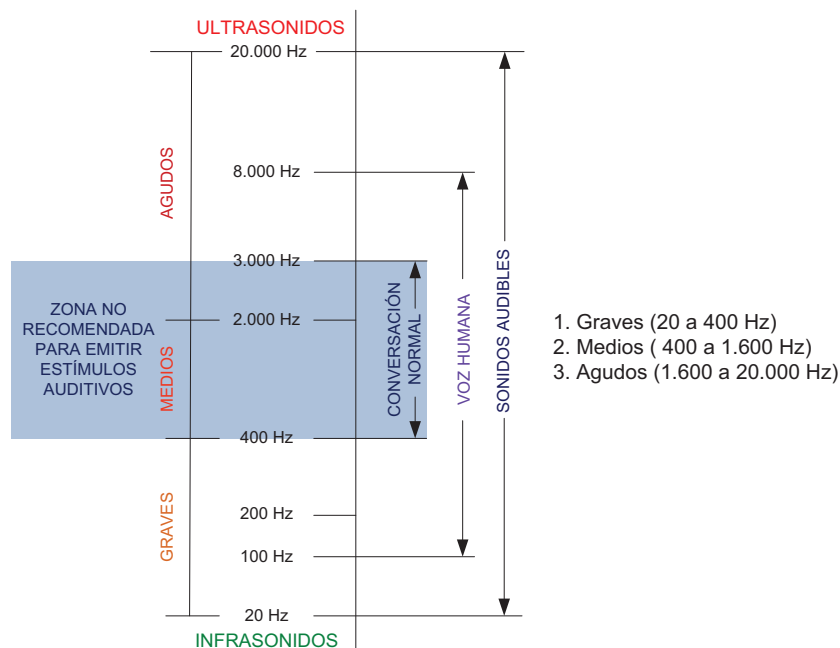


Figura 1. 4 Frecuencia del sonido

Muchos animales oyen una gama de frecuencias más amplia que los seres humanos. Por ejemplo, los silbatos para perros vibran a una frecuencia alta, que los seres humanos no son capaces de detectar; mientras que ciertas evidencias sugieren que los delfines y las ballenas se comunican con frecuencias fuera del alcance del oído humano, como se observa en la figura 1.5.

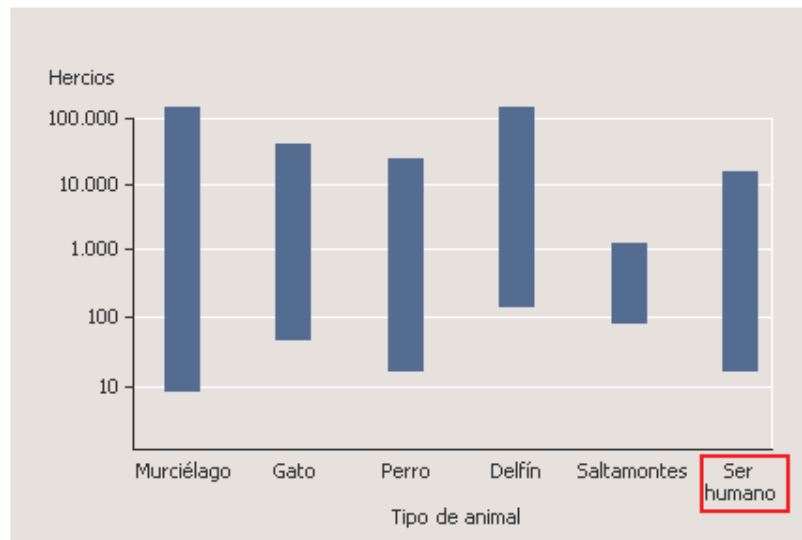


Figura 1. 5 Comparación de la gama de frecuencias que escuchan los animales, y el ser humano

1.1.3.6.1 Tono

Llamado también nota; el tono está relacionado con la percepción y depende de la frecuencia de la onda. La frecuencia y el tono están relacionados, pero no de manera lineal. La frecuencia es una entidad física y por tanto puede ser medida de forma objetiva por diferentes medios; por el contrario el tono de un sonido es un fenómeno totalmente subjetivo y por tanto no es posible medirlo de forma objetiva. Normalmente cuando se aumenta la frecuencia de un sonido, su tono también sube, sin embargo esto no se da de forma lineal, o sea no se corresponde la subida del valor de la frecuencia con la percepción de la subida de tono.

El oído humano puede diferenciar entre un sonido entre los 100 y los 300 Hz, pero no entre un sonido entre 10 KHz y los 10,2 KHz.

1.1.3.6.2 Armónicas

La frecuencia principal a la que vibra un instrumento es llamada frecuencia fundamental y la reconocemos por el tono. Los instrumentos musicales y la voz producen frecuencias fundamentales y sobretonos de las frecuencias fundamentales. Estos sobretonos son denominados *armónicas* [7].

Una nota musical tocada en un violín no suena igual a la misma nota tocada en un piano. Esto se debe a que sus ondas no tienen una única forma de onda senoidal pura sino que pueden verse como la suma de muchas ondas senoidales de distintas frecuencias. Esto se conoce en física como síntesis (obtención del sonido) por transformada de Fourier (FFT) que descompone una forma de onda en la suma de múltiples ondas senoidales, como se muestra en la figura 1.6.

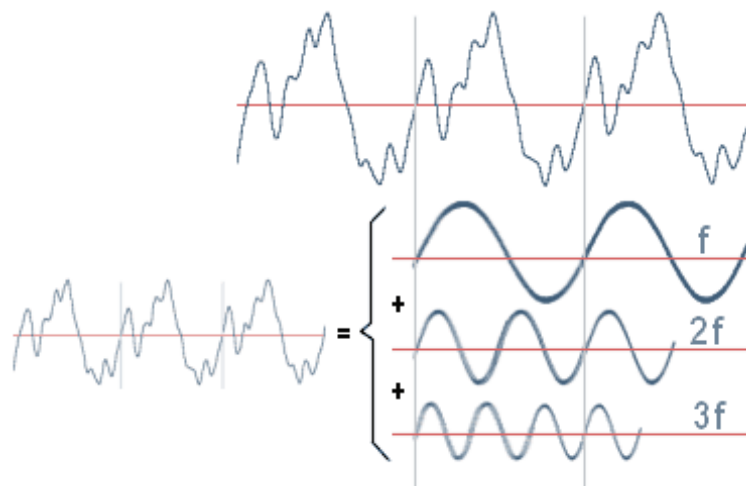


Figura 1. 6 Frecuencia fundamental (f) y armónicas ($2f$, $3f$)

1.1.3.7 Fase

La fase es simplemente el desfase o corrimiento de una señal respecto a un punto de referencia, el que se determina en forma arbitraria.

La fase se mide en radianes, en el rango $[0; 2\pi]$ o en grados, en el rango $[0; 360]$.

La fase de un sonido aislado no altera en nada su percepción. La fase adquiere importancia cuando dos o más sonidos se mezclan entre sí. Dos sonidos pueden ser idénticos, pero estar desfasados entre sí, lo que implica que un sonido comenzó antes que el otro. Al interactuar, el resultado percibido puede cambiar radicalmente dependiente del grado de desfase entre ellos. Si el desfase es 0, o bien 2π , los sonidos al mezclarse se suman, y como las zonas de rarefacción y compresión de ambos sonidos coinciden, como resultado se obtiene el mismo sonido pero amplificado. Si el desfase es de π o 180 grados, significa que las zonas de rarefacción de un sonido coinciden con las zonas de compresión del otro, y al mezclarse, los sonidos se anulan completamente. El resultado es que no se percibe nada. Esto no es un problema perceptual, es un fenómeno puramente físico.

En la figura 1.7 se muestran dos ondas sinusoidales en diferente relación de fase.

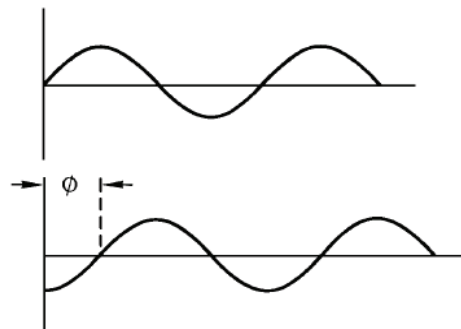


Figura 1. 7 Dos ondas sinusoidales en diferente relación de fase

1.1.3.8 Forma de onda

El patrón de variaciones de presión producido por una fuente de acuerdo al tiempo se conoce como la *forma de onda*. La forma de onda determina en gran medida la cualidad del sonido. Un factor importante de considerar en un sonido es su periodicidad. La forma de onda de un sonido determina y está determinada por su contenido de frecuencias o espectro. La forma de onda se asocia comúnmente con lo que se denomina timbre, cualidad perceptual que le otorga identidad al sonido. Es la forma de onda la que permite diferenciar, por ejemplo, el sonido de una trompeta del de un violín.

1.1.3.8.1 Representación gráfica

Existen diversas formas de representar el sonido en forma gráfica. La forma de representación más utilizada se basa en un diagrama de *amplitud versus tiempo*. Un sonido también puede ser representado por su espectro, mediante un gráfico *amplitud versus frecuencia*. Este gráfico muestra las amplitudes de cada componente de frecuencia contenida en el sonido. Por lo general, siempre es posible pasar de una representación en el tiempo a una representación en la frecuencia y viceversa, mediante la transformada de Fourier. No obstante, es importante destacar que el diagrama de tiempo no contiene información alguna sobre el contenido de frecuencias del sonido y el espectro no contiene información de tipo temporal.

Una representación intermedia es lo que se llama el *sonograma*. Un sonograma consiste básicamente en un eje tridimensional donde se grafica la *magnitud del espectro de un sonido versus tiempo*. Esto se logra mediante la subdivisión de la señal de audio en varias pequeñas ventanas de tiempo, usualmente traslapadas entre sí. En cada una de éstas ventanas temporales, se estima el espectro mediante lo que se denomina la transformada de Fourier de tiempo corto. De esta forma es posible determinar cómo va cambiando el contenido de frecuencias del sonido en el tiempo. Si bien el sonograma es muy útil, la información que entrega es altamente dependiente de los parámetros que se utilicen para su cálculo, como el tipo de ventana, el tamaño de cada ventana y el porcentaje de traslape, entre otros. En la figura 1.8 se muestra representado el sonograma.

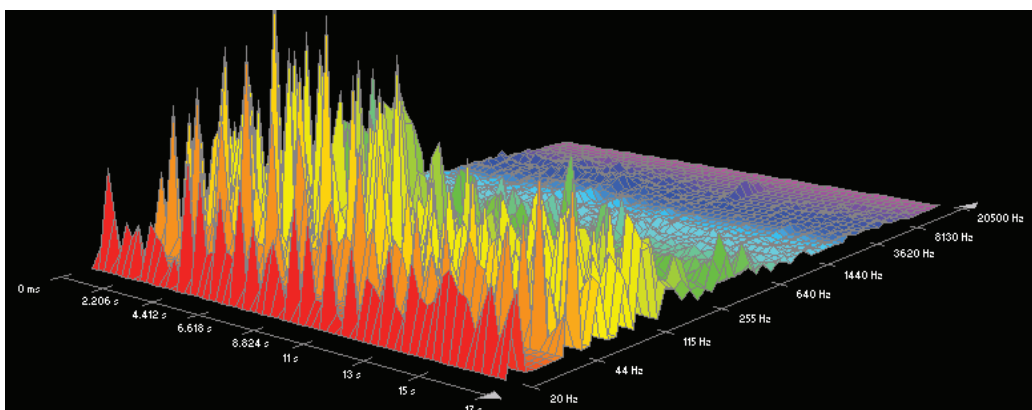


Figura 1. 8 Representación gráfica del espectro del sonido

1.1.4 FUENTES SONORAS

El sonido puede ser producido por distintos tipos de fuentes y procesos [6]:

- a) *Cuerpos en vibración.* Un ejemplo de este tipo de fuentes es un diapasón, el cual al ponerse en vibración genera un cierto tipo de onda sonora, como se muestra en la figura 1.9. Al estar la fuente vibrando, causa un desplazamiento en el aire cercano, lo que produce cambios locales en la presión de aire. Estas fluctuaciones de presión viajan en forma de una onda. Los cuerpos en vibración son las fuentes sonoras más comunes.

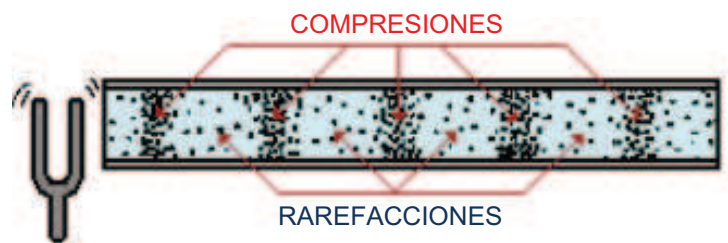


Figura 1. 9 Onda de sonido emitida por un diapasón

- b) *Cambios en flujos de aire.* Un ejemplo de este tipo de fuentes es lo que sucede cuando hablamos. Las cuerdas vocales se abren y cierran en forma alternada, produciendo cambios en la tasa del flujo de aire, lo que a su vez se traduce en una onda sonora. Este mismo principio se aplica a los instrumentos de viento como el clarinete u oboe. Otro ejemplo de este tipo de fuentes es una sirena, la cual produce sonido a través de una placa rotatoria la cual bloquea en forma alternada el flujo proveniente de un compresor de aire.
- c) *Fuentes de calor.* Una chispa eléctrica produce un sonido, tal como lo produce un trueno. En estos casos, el sonido se produce por un brusco cambio en la temperatura, el cual produce una veloz expansión del aire circundante.
- d) *Flujo supersónico.* En el caso de un avión supersónico se producen ondas de choque que fuerzan al aire a viajar más rápido que la velocidad del sonido.

1.2 EL SISTEMA FONATORIO HUMANO

La voz humana fue definida por Platón como “*un impacto del aire que llega por los oídos al alma*”. La voz es el sustrato en el que se apoya el método de comunicación habitual del ser humano, con el que se transmite la cultura, con el que se expresan los sentimientos y las emociones.

Para que la comunicación tenga lugar, el hablante debe producir una señal de voz en forma de una onda de presión sonora, la cual viaja desde su boca hasta el oído del oyente. Aunque la mayoría de ondas de presión sonora son originadas desde la boca, el sonido también puede proceder de los orificios nasales, la garganta y las mejillas.

Las señales del habla están compuestas de una secuencia de sonidos, los cuales sirven como una representación simbólica de los pensamientos que el hablante desea transmitir al oyente. Así el orden de estos sonidos es gobernado por reglas asociadas al lenguaje. El estudio científico del lenguaje y la manera en la cual estas reglas son utilizadas en la comunicación humana se denomina *lingüística*. En cambio la ciencia que estudia las características de la producción de sonidos humanos sobretodo en aspectos como descripción, clasificación y transcripción es llamada *fonética*.

1.2.1 PROPIEDADES DEL HABLA

El habla humana es una forma de onda continua con una frecuencia fundamental en el rango de 100 a 400 Hz. (El promedio está cerca de los 100 Hz para los hombres y 200 Hz para las mujeres.) A distancias de múltiplos enteros de la frecuencia fundamental, se encuentran una serie de armónicos cambiantes a los que podemos llamar "formantes" y que están determinados por las características de resonancia de la región vocal. Los "formantes" crean los variados sonidos vocales y las transiciones entre ellos. Los sonidos consonantes, que son impulsivos y/o ruidosos, tienen lugar en el rango de los 2 KHz a los 9 KHz.

La potencia sonora del habla es llevada por las vocales, cuya duración promedio es de 30 a 300 milisegundos. La inteligibilidad es dada por las consonantes, cuya duración promedio es de 10 a 100 milisegundos de duración y pueden ser unos 27 dB más débiles en amplitud que las vocales.

El tono del sonido es controlado variando la forma del tracto vocal (en la mayoría moviendo la lengua) y por el movimiento de los labios. La intensidad (volumen) es controlada por la variación de la cantidad de aire enviado desde los pulmones. Los pulmones operan lentamente y el tracto vocal cambia su forma lentamente, por eso, el tono y el volumen del habla varían de manera lenta.

El movimiento de la glotis y el tracto vocal dan lugar a diferentes tipos de sonido los tres principales tipos de sonido son los siguientes:

- Sonidos sonoros: estos son sonidos que nosotros realizamos cuando hablamos. Las cuerdas vocales vibran, las cuales abren y cierran la glotis, enviando así los pulsos de aire a variadas presiones al tracto, donde son transformadas en ondas de sonido. Las frecuencias de la voz humana están generalmente en el rango de 500 Hz hasta alrededor de 2 KHz. Esto es equivalente a periodos de tiempo de 2 ms a 20 ms, y para una computadora tales períodos son muy largos. Así, los sonidos sonoros tienen una periodicidad larga y esta es la llave para una buena compresión del habla.
- Sonidos sordos: estos son sonidos emitidos y pueden ser escuchados, pero no son parte del habla. Tal sonido es el resultado de sostener la glotis, abrirla y forzar al aire a través de la contracción en el tracto vocal. Cuando un sonido sordo es muestreado las muestras manifiestan poca correlación y son aleatorias o casi aleatorias.
- Sonidos explosivos: estos sonidos se originan cuando la glotis se cierra, los pulmones transmiten aire a presión sobre la glotis produciendo que la misma se abra inesperadamente y el resultado es un sonido de chasquido.

1.2.2 CLASIFICACIÓN DE LAS VOCES

1.2.2.1 Clasificación sexual

En base a esta clasificación, se puede enumerar cuatro tipos de voces:

Voz de mujer: laringe y cuerdas vocales más reducidas que en el hombre, por lo que el sonido es aproximadamente una octava¹ superior que en voces masculinas.

Voz de hombre: laringe mayor, por lo que el sonido es más grave.

En la tabla 1.2 se puede observar los diferentes niveles sonoros y de intensidad de voces masculinas y femeninas en diversas circunstancias.

Emisión	Intensidad (w/m²)	Nivel sonoro (dB)
Nivel mínimo de la voz humana	10^{-10}	20
Mujer conversando en voz baja	$3,16 \times 10^{-10}$	25
Hombre conversando en voz baja	10^{-9}	30
Mujer conversando en voz normal	10^{-7}	50
Hombre conversando en voz normal	$3,16 \times 10^{-7}$	55
Mujer hablando en público	10^{-6}	60
Hombre hablando en público	$3,16 \times 10^{-6}$	65
Mujer hablando esforzándose	10^{-5}	70
Hombre hablando esforzándose	$3,16 \times 10^{-5}$	75
Mujer cantando	10^{-4}	80
Hombre cantando	$3,16 \times 10^{-4}$	85
Nivel máximo de la voz humana	10^{-3}	90

Tabla 1. 2 Niveles de intensidad de la voz humana

Voz infantil: laringe más corta que en hombres y mujeres, por lo que el sonido es muy agudo.

Voz asexual: voz del niño castrado antes de la pubertad. Dicha castración de los órganos sexuales evita el crecimiento y desarrollo de la laringe. En el siglo XVI, al

¹ En música, una octava es el intervalo que separa dos sonidos cuyas frecuencias fundamentales tienen una relación de dos a uno.

no estar admitidas las mujeres en las iglesias como cantantes, son sustituidas por niños o castrados. En los siglos subsiguientes estas voces, que en general poseen mayor extensión y potencia que las de las cantantes, obtienen un éxito enorme sobre todo en Italia entre el público de teatro de ópera, pues las mujeres tampoco solían cantar en la escena. En 1770, el Papa Clemente XIV autoriza el canto de las mujeres en la iglesia y se prohíbe la castración. Los castrados desaparecen a principios del siglo XX.

1.2.2.1.1 Clasificación según tesitura

La clasificación de la voz por tesitura se define, como aquella, que clasifica la voz por su amplitud tonal. Constituye el conjunto de notas que puede emitir una determinada persona. Un sentido de interpretar la tesitura, es el que sitúa el conjunto de sonidos, en los que la voz se adapta mejor, la parte de la gama vocal, en que el cantante se siente cómodo, sin ningún tipo de fatiga. En base a la clasificación de la voz por tesitura se definen algunos términos tanto para hombres y mujeres como se observa en la figura 1.10.

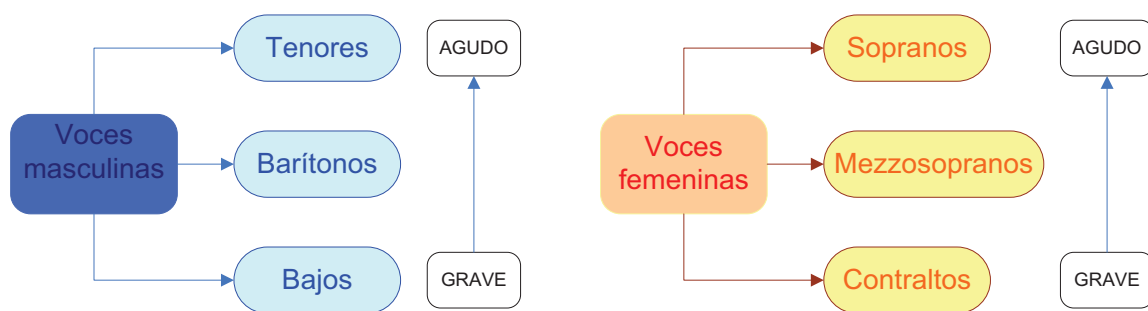


Figura 1. 10 Clasificación de la voz por tesitura

Voz	Extensión (Hz)	Tesitura
Soprano	247-1056	SI ₃ - DO ₆
Mezzosoprano	220-900	LA ₃ - SIb ₅
Contralto	176-840	FA ₃ - LA _{b5}
Tenor	132-528	DO ₃ - DO ₅
Barítono	110-440	LA ₂ - LA ₄
Bajo	82-396	MI ₂ - SOL ₄

Tabla 1. 3 Frecuencias de la voz por tesitura

1.3 PSICOACÚSTICA

1.3.1 DEFINICIÓN

La psicoacústica es la ciencia que estudia la percepción de los sonidos [6]. Estudia la relación entre las propiedades físicas del sonido y la interpretación que hace de ellas el cerebro.

Una vez que una onda sonora proveniente del mundo físico ingresa al sistema auditivo humano, se suceden una serie de reacciones en forma casi instantánea que producen como resultado una representación mental de lo escuchado, que no corresponde exactamente a lo que sucede en el mundo físico. Las variables físicas del sonido no constituyen una representación fidedigna de lo que ocurre en el mundo perceptual. Por ejemplo, en ciertas situaciones hay sonidos que bloquean a otros sonidos, incluso si éstos ocurren en forma asincrónica². Este fenómeno se conoce como enmascaramiento. En estos casos, si bien todas las ondas sonoras en juego existen en el mundo físico, al presentarse todas juntas algunas de ellas simplemente no se perciben, a pesar de que si cada uno de estos sonidos se presentaran por separado, sí se percibirían.

1.3.2 EL SISTEMA AUDITIVO HUMANO

La generación de sensaciones auditivas en el ser humano es un proceso extraordinariamente complejo, el cual se desarrolla en tres etapas básicas:

- a) Captación y procesamiento mecánico de las ondas sonoras.
- b) Conversión de la señal acústica (mecánica) en impulsos nerviosos, y transmisión de dichos impulsos hasta los centros sensoriales del cerebro.
- c) Procesamiento neural de la información codificada en forma de impulsos nerviosos.

² Forma asincrónica, cuando no existe coincidencia temporal.

La captación, procesamiento y transducción de los estímulos sonoros se llevan a cabo en el oído propiamente dicho, mientras que la etapa de procesamiento neural, en la cual se producen las diversas sensaciones auditivas, se encuentra ubicada en el cerebro. Así pues, se pueden distinguir dos regiones o partes del sistema auditivo: la región periférica, en la cual los estímulos sonoros conservan su carácter original de ondas mecánicas hasta el momento de su conversión en señales electroquímicas, y la región central, en la cual se transforman dichas señales en sensaciones. En la región central también intervienen procesos cognitivos, mediante los cuales se asigna un contexto y un significado a los sonidos; es decir, permiten reconocer una palabra o determinar que un sonido dado corresponde a un violín o a un piano.

1.3.2.1 El oído humano

Sin duda uno de los componentes más importantes del sistema auditivo humano es el oído. El oído se encarga de convertir las ondas de presión acústica en impulsos nerviosos que le permiten al cerebro crear una representación mental de la sensación auditiva. El oído humano está dividido en tres partes principales: oído externo, oído medio y oído interno, como se muestra a continuación en la figura 1.11.

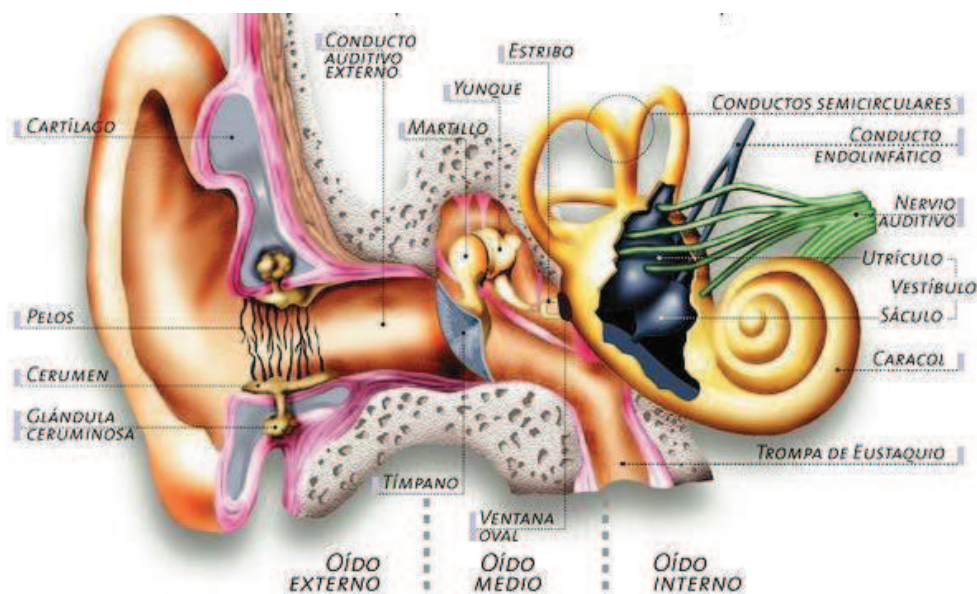


Figura 1. 11 El oído humano

El funcionamiento general del oído es el siguiente: cuando una onda sonora llega al oído viaja desde el pabellón auricular o pinna hasta el tímpano, a través del canal auditivo. El sonido es modificado en términos de su contenido de frecuencias por el oído externo. El tímpano se encarga de traspasar el patrón de vibraciones de presión hacia el oído medio al hacer contacto con tres diminutos huesecillos, que a su vez traspasan la vibración hacia la ventana oval, lo que causa una onda de propagación del líquido contenido al interior de la cóclea, estimulando las células ciliares de la membrana basilar, la que está conectada a un gran número de terminales nerviosos que envían a su vez señales eléctricas al cerebro. De esta manera, el cerebro puede recibir la información proveniente de la onda sonora para su posterior procesamiento.

1.3.2.1.1 Oído externo

El oído externo está constituido por el *pabellón auricular* o *pinna*, el *canal auditivo* y el *tímpano*. El pabellón recoge las ondas sonoras y las conduce hacia el canal auditivo mediante reflexiones y difracciones. El efecto del pabellón es atenuar y enfatizar cierto contenido de frecuencias y juega un rol fundamental en la localización de sonidos. El canal auditivo mide alrededor de 2,5 cm y actúa como un resonador para el rango de frecuencias entre 1.000 y 4.000 Hz, siendo el máximo alrededor de 3.000 Hz. El tímpano vibra en respuesta al sonido y transmite esta vibración de presión en forma de vibración mecánica hacia el oído medio.

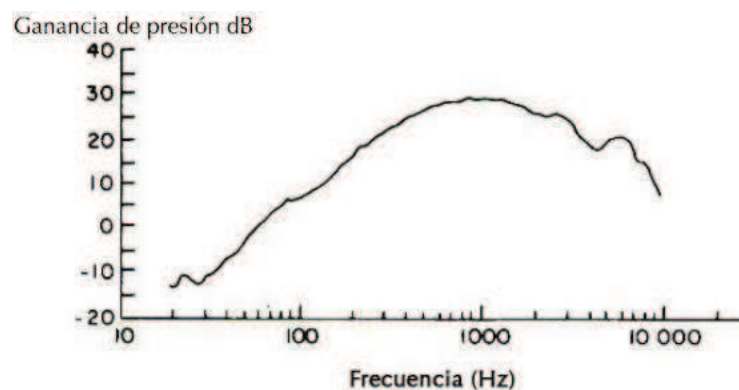


Figura 1. 12 Respuesta de frecuencia del canal auditivo

En la figura 1.12, claramente se puede observar que el canal auditivo enfatiza el rango desde 1 a 4.000 Hz. Esto corresponde al rango de la voz humana hablada. El rol que cumple el canal auditivo, por lo tanto, es el de optimizar la señal acústica del tal forma de resaltar la voz humana.

1.3.2.1.2 Oído medio

Está ubicado en la caja timpánica, y lo integran el *tímpano*, los *huesecillos* u *osículos*, y la *trompa de Eustaquio*. El tímpano es una membrana elástica, semitransparente y algo cónica, que comunica el canal auditivo externo con la caja timpánica.

El oído medio actúa como un transductor de vibración. Su rol es amplificar la vibración de presión mediante un sistema mecánico. El tímpano recibe las vibraciones del aire y las comunica a los huesecillos. Esto se hace mediante tres huesitos llamados martillo, yunque y estribo que se pueden apreciar en la figura 1.13. Están sostenidos en su lugar por una serie de pequeños ligamentos y músculos. La finalidad de esta cadena es convertir vibraciones de gran amplitud y poca presión, como las hay en el tímpano, en vibraciones de pequeña amplitud y mayor presión, requeridas en el líquido que llena el oído interno. Esta función es asimilable, por consiguiente, a una palanca mecánica. A causa del efecto palanca, las vibraciones del estribo son de menor amplitud pero de mayor fuerza. La ganancia mecánica de esta palanca es de 1,3, lo que significa que la fuerza que el estribo ejerce sobre la ventana oval es 1,3 veces mayor que la que ejerce el tímpano sobre el martillo. A este efecto de palanca se agrega la gran diferencia de áreas entre el tímpano (0,6 cm²) y la ventana oval (0,04 cm²), lo cual implica que la relación entre las presiones en el tímpano y en la ventana oval es del orden de:

$$\frac{P_{\text{ventana oval}}}{P_{\text{tímpano}}} = \frac{1,3 \times 0,6}{0,04} \cong 20 \quad (\text{Ec. 1.8})$$

Esta diferencia de presiones es necesaria ya que en el tímpano existe una impedancia acústica mucho menor que en el oído interno, ya que éste último

contiene agua. El conjunto actúa, por consiguiente, como un ingenioso adaptador de impedancias acústicas.

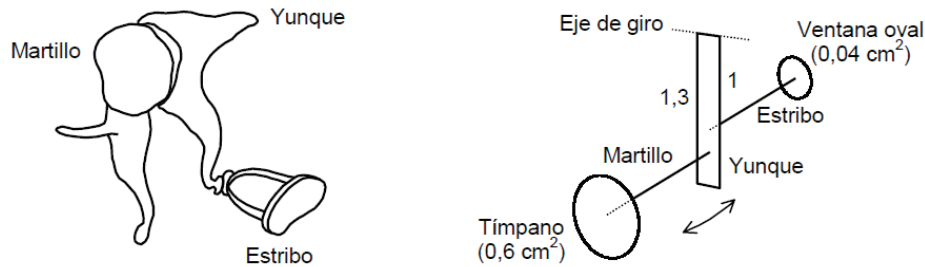


Figura 1. 13 Izquierda, los tres huesitos del oído medio. Derecha, su mecánica

Los músculos, además de la función de sostén de la cadena osicular, sirven de protección del oído interno frente a sonidos intensos. Cuando penetra en el oído un ruido muy intenso, se produce la contracción refleja de estos músculos, rigidizando la cadena, que pierde entonces su eficiencia mecánica, y la energía es disipada antes de alcanzar el oído interno. Esta protección sólo es efectiva, sin embargo, para sonidos de más de 500 ms de duración.

La trompa de Eustaquio es un pequeño conducto que comunica la caja timpánica con la laringe. Su función es la de igualar la presión del oído medio con la presión atmosférica. Normalmente, permanece cerrada, abriéndose en forma refleja durante la acción de tragar o de bostezar. Si permaneciera siempre abierta, el tímpano vibraría con una amplitud muy pequeña, ya que el movimiento del tímpano es el resultado de una diferencia de presión a uno y otro lado de éste.

Por consiguiente, para vibrar en concordancia con las variaciones de presión sonora es preciso que la presión dentro de la caja timpánica se mantenga constante. Si, en cambio, la trompa de Eustaquio estuviera cerrada permanentemente (o no existiera), al variar la presión atmosférica se produciría una diferencia de presiones medias que curvaría el tímpano hacia afuera o hacia adentro, reduciendo notablemente la respuesta auditiva, particularmente para los sonidos agudos (altas frecuencias).

1.3.2.1.3 Oído interno

El oído interno representa el final de la cadena de procesamiento mecánico del sonido, y en él se llevan a cabo tres funciones primordiales: filtraje de la señal sonora, transducción y generación probabilística de impulsos nerviosos.

El oído interno consiste básicamente de la cóclea [1], ya que los canales semicirculares, si bien se encuentran allí, no tienen rol alguno en la audición. La cóclea es un pasaje angosto, con forma de caracol, lleno de un líquido llamado perilinfa, largo y enrollado 3,5 veces sobre sí mismo. El diámetro de este pasaje es de 2 mm y su largo es 35 mm. En este lugar se encuentran las células pilosas que por respuesta a las vibraciones venidas desde un estímulo externo, transmiten las sensaciones de sonido al cerebro. El interior del conducto está dividido en sentido longitudinal por la membrana basilar y la membrana de Reissner, las cuales forman tres compartimientos o escalas.

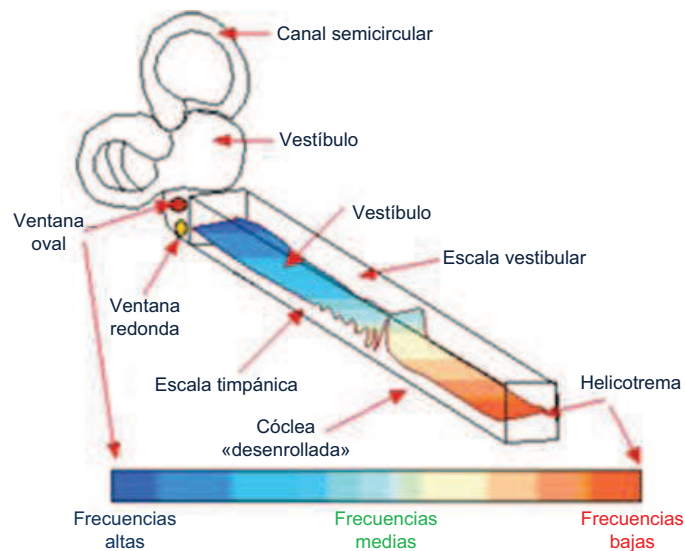


Figura 1. 14 Onda viajera en la membrana basilar [5]

La membrana basilar es una estructura cuyo espesor y rigidez no es constante, está cerca de la ventana oval, la membrana es gruesa y rígida, pero a medida que se acerca hacia el vértice de la cóclea se vuelve más delgada y flexible. La rigidez decae casi exponencialmente con la distancia a la ventana oval; esta variación de la rigidez en función de la posición afecta la velocidad de propagación de las

ondas sonoras a lo largo de ella, y es responsable en gran medida de un fenómeno muy importante: *la selectividad en frecuencia del oído interno*. Esto confiere al oído interno una cualidad analítica que es de fundamental importancia en la discriminación tonal del sonido, especialmente para los sonidos de frecuencias superiores a los 1.000 Hz.

1.3.2.2 Bandas críticas del sistema auditivo humano

Una característica fundamental del sistema auditivo humano es su capacidad de resolución de frecuencia e intensidad [6]. Al momento de estudiar éste aspecto de nuestra audición es fundamental el concepto de banda crítica. Una forma de entender el funcionamiento del sistema auditivo es suponer que contiene una serie o banco de filtros pasa banda sobrelapados conocidos como filtros auditivos³. Estos filtros se producen a lo largo de la membrana basilar y tienen como función aumentar la resolución de frecuencia de la cóclea y así incrementar la habilidad de discriminar entre distintos sonidos. Este banco de filtros no sigue una configuración lineal, y el ancho de banda y morfología de cada filtro depende de su frecuencia central. El ancho de banda de cada filtro auditivo se denomina banda crítica⁴.

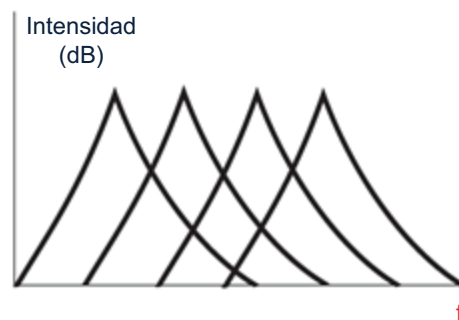


Figura 1. 15 Esquema de las bandas críticas del sistema auditivo humano

Las bandas críticas, esquematizadas en la figura 1.15, son rangos de frecuencia dentro de los cuales un sonido bloquea o enmascara la percepción de otro sonido.

³ Fletcher (1940) citado en Moore, 1998

⁴ Fletcher (1940) citado en Gelfand, 2004

Las bandas críticas conceptualmente están ligadas a lo que sucede en la membrana basilar, ya que una onda que estimula la membrana basilar perturba la membrana dentro de una pequeña área más allá del punto de primer contacto, excitando a los nervios de toda el área vecina. Por lo tanto, las frecuencias cercanas a la frecuencia original tienen mucho efecto sobre la sensación de intensidad del sonido. La intensidad percibida no es afectada, en cambio, en la presencia de sonidos fuera de la banda crítica. Es importante destacar aquí que el concepto de banda crítica es una construcción teórica y no algo físicamente comprobado.

1.3.2.2.1 Escala de Barks

Existe una escala de medición de las bandas críticas llamada la *escala de Barks*⁵, la cual se detalla en la tabla 1.4. La escala tiene un rango del 1 al 24 y corresponde a las primeras veinticuatro bandas críticas del sistema auditivo [6]. Esta escala tiene relación con la escala mel⁶.

El Bark es la unidad de frecuencia perceptual; específicamente, un Bark mide la tasa de banda crítica, o sea, una banda crítica tiene un ancho de un Bark. La escala de Barks relaciona la frecuencia absoluta (en Hz) con las frecuencias medidas perceptualmente (el caso de las bandas críticas).

Usando el Bark, un sonido en el dominio de la frecuencia puede ser convertido a sonido en el dominio psicoacústico. De esta manera, un tono puro (representado por una componente en el dominio de la frecuencia) puede ser representado como una curva de enmascaramiento psicoacústico. La relación entre un Bark y un Hz está dada por:

$$Bark = 13 \arctan(0.00076f) + 3,5 \arctan\left(\left(\frac{f}{7500}\right)^2\right) \quad (\text{Ec. 1.9})$$

⁵ En honor al físico alemán Georg Heinrich Barkhausen.

⁶ El nombre mel deriva de melodía, como una forma de explicitar que se trata de una escala basada en comparaciones entre alturas.

Banda crítica (Bark)	Frecuencia central (Hz)	Ancho de banda (Hz)	Frecuencia mínima (Hz)	Frecuencia máxima (Hz)
1	50	-	-	100
2	150	100	100	200
3	250	100	200	300
4	350	100	300	400
5	450	110	400	510
6	570	120	510	630
7	700	140	630	770
8	840	150	770	920
9	1000	160	920	1080
10	1170	190	1080	1270
11	1370	210	1270	1480
12	1600	240	1480	1720
13	1850	280	1720	2000
14	2150	320	2000	2320
15	2500	380	2320	2700
16	2900	450	2700	3150
17	3400	550	3150	3700
18	4000	700	3700	4400
19	4800	900	4400	5300
20	5800	1100	5300	6400
21	7000	1300	6400	7700
22	8500	1800	7700	9500
23	10500	2500	9500	12000
24	13500	3500	12000	15500
25	18775	6550	15500	22050

Tabla 1. 4 Escala de Barks, para estimación de las bandas críticas del sistema auditivo [6]

1.3.2.3 Enmascaramiento

El enmascaramiento ocurre cuando la presencia de un sonido, llamado máscara, hace inaudible otro sonido que sería perfectamente audible en la ausencia de la máscara. El sistema auditivo no es capaz de diferenciar dos sonidos al interior de una banda crítica. Basta con que exista algo de energía al interior de una banda crítica para que ésta se active y el sistema auditivo perciba actividad en esa banda. Si existe más de un sonido o se incrementa la energía al interior del filtro, esto no cambia la información desde el punto de vista del sistema auditivo.

Entonces, si un sonido se encuentra al interior de una banda crítica de otro sonido y si su amplitud no es lo suficientemente grande como para traspasar el umbral de

dicha banda y activar otra banda crítica cercana, se produce el fenómeno denominado enmascaramiento, que se muestra en la figura 1.16. La codificación perceptual de audio, se basa fuertemente en éste fenómeno para reducir la cantidad de información necesaria para almacenar y reproducir una señal sonora.

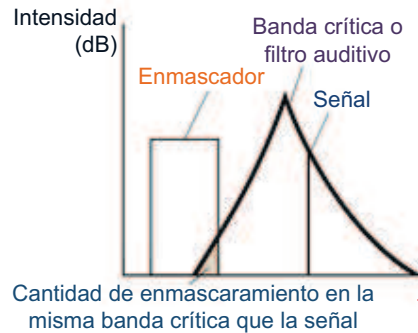


Figura 1. 16 Enmascaramiento

Las bandas críticas pueden medirse en función del enmascaramiento que producen. El ancho de banda de una máscara constituida por ruido blanco se ensancha continuamente y se mide el nivel de enmascaramiento que produce respecto a una señal de prueba. Una vez que se alcanza el punto sobre el cual el enmascaramiento no cambia significativamente al agrandarse el ancho de banda, se está en la presencia de los límites de la banda crítica en medición.

1.3.2.3.1 Enmascaramiento en frecuencia

El enmascaramiento en frecuencia está relacionado con la sensibilidad tonal del oído. Su estructura interna, divide el rango de frecuencias audibles en 24 bandas críticas. En la figura 1.17 se muestra el enmascaramiento producido por un tono de 1 KHz.

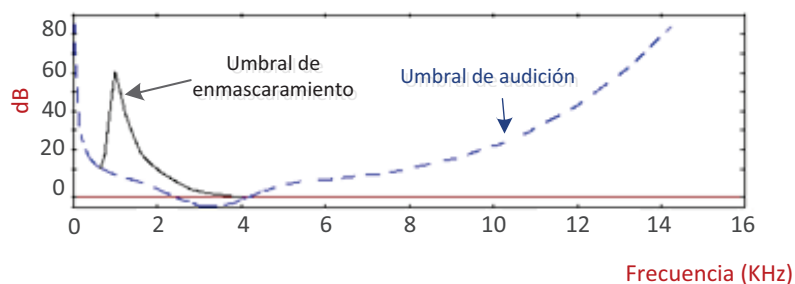


Figura 1. 17 Enmascaramiento por un tono de 1 KHz

Funciona de manera que un sonido en determinada frecuencia puede enmascarar o disminuir el nivel de otro sonido en las frecuencias adyacentes, siempre y cuando el nivel del sonido enmascarante sea más alto (un sonido más intenso, más fuerte) que el nivel del sonido adyacente.

1.3.2.3.2 Enmascaramiento temporal

El enmascaramiento temporal del oído está relacionado con la inercia del oído interno. Cualquier sonido audible, produce una excitación mecánica del oído interno, prolongando artificialmente la sensación auditiva hasta que la situación de equilibrio se recupera completamente. Se presenta cuando un tono suave está muy cercano en el dominio del tiempo (unos cuantos milisegundos) a un tono fuerte. Si se está escuchando un tono suave y aparece un tono fuerte, el tono suave será enmascarado por el tono fuerte, antes de que el tono fuerte efectivamente aparezca (pre-enmascaramiento). Posteriormente, cuando el tono fuerte desaparece, el oído necesita un pequeño intervalo de tiempo (entre 50 y 300 ms) para que se pueda seguir escuchando el tono suave (post-enmascaramiento), ilustrado en la figura 1.18.

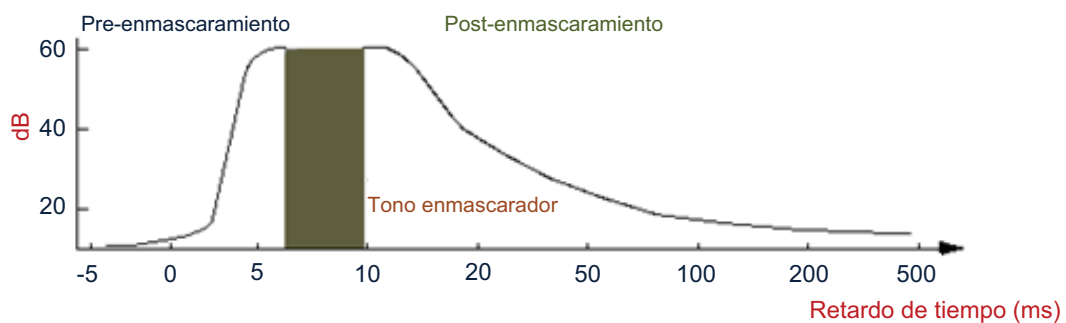


Figura 1. 18 Enmascaramiento temporal

El pre-enmascaramiento es un fenómeno inesperado, pues pareciera implicar que el sistema auditivo es no causal: una señal puede enmascarar a otra antes de ser aplicada. Sin embargo, es posible justificar la existencia del pre-enmascaramiento si se piensa que cualquier sensación sonora no se produce instantáneamente, sino que se requiere de un cierto tiempo para que se origine dicha sensación.

Con el post-enmascaramiento no hay problemas; pero el pre-enmascaramiento sugiere que un tono será enmascarado por otro tono, antes de que el tono enmascarador realmente aparezca. Para este fenómeno, se han presentado dos explicaciones:

- a) El cerebro integra el sonido sobre un período de tiempo, y procesa la información por ráfagas en la corteza auditiva, o
- b) Simplemente, el cerebro procesa los sonidos fuertes más rápido que los sonidos suaves.

En un sonido cualquiera, se presentan ambos tipos de enmascaramiento. El enmascaramiento en frecuencia es mucho más importante que el enmascaramiento temporal; aunque en ciertos dispositivos para compresión de audio se tiene en cuenta ambos tipos de enmascaramiento, con lo cual se logra mejor compresión de datos.

1.4 AUDIO DIGITAL

1.4.1 INTRODUCCIÓN

Antes de tratar aspectos como el audio analógico y el audio digital resulta necesario conocer cuál es la diferencia entre el sonido y el audio.

El sonido es una variación de presión acústica ejercida sobre las partículas de un medio (en nuestro caso: el aire). Mientras que el audio es una representación eléctrica del sonido. Actualmente se utilizan dos tipos de tecnologías aplicadas al audio: el "audio analógico" y el "audio digital".

Tanto el sonido, una onda de presión sonora, y el audio (analógico) una señal eléctrica continua pueden observarse en la figura 1.19.

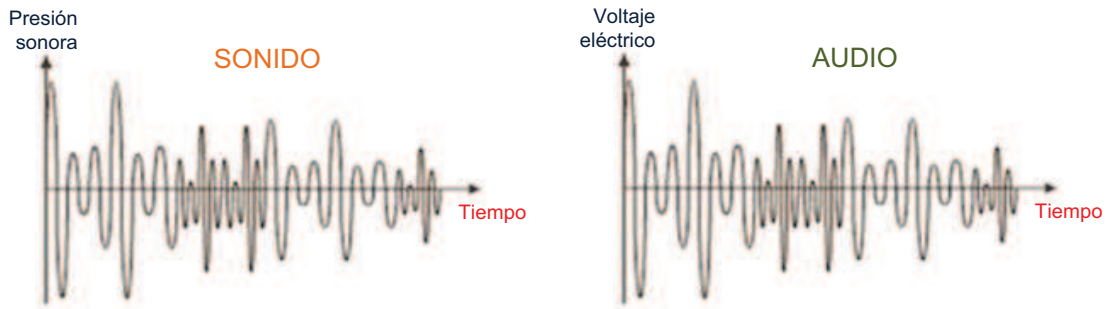


Figura 1. 19 Diferencia entre sonido y audio

1.4.2 AUDIO ANALÓGICO

1.4.2.1 Señales analógicas

Las señales analógicas son aquellas en las que la información está representada por señales continuas, así, una señal continua o análoga puede tomar un número infinito de valores entre dos máximos.

Las señales analógicas se llaman así porque son "análogas" a la forma de la señal original. Es decir, si se compara la señal, ésta equivaldría a la señal resultante en su forma.

En la figura 1.20 se representa una señal de voltaje analógica. El margen de valores que toma la función se denomina margen dinámico y suele estar acotado entre un valor mínimo y uno máximo que dependen de las limitaciones físicas de los sistemas de transducción.



Figura 1. 20 Representación de una señal analógica (señal eléctrica)

1.4.2.2 Audio analógico

1.4.2.2.1 Historia

La primera persona en dar una aplicación pública al sonido tratado con medios eléctricos fue Alexander Graham Bell, quien en 1876 patentó un equipo para transmitir audio: el teléfono.

Ese mismo año Emile Berliner, inventó el primer micrófono utilizado como transmisor de voz y la Bell Telephone Company adquirió la patente. Un año después, el 4 de diciembre de 1877, el inventor Thomas Alva Edison construye el primer fonógrafo de cilindro cubierto con papel de aluminio y allí realizó la primera grabación de la voz humana. Justo ese año, Ernst Siemens patenta el primer altavoz. Mientras que Edison continuaba trabajando en el desarrollo de la grabación en cilindros, Berliner inventó el disco plano y patentó el gramófono con dicha tecnología; con ello se hizo posible en adelante reproducir el sonido de una manera práctica.

1.4.2.2.2 Definición

El denominado "audio analógico" es la representación eléctrica del sonido como se observa en la figura 1.21, en la cual las variaciones de presión sonora son representadas por las variaciones de voltaje. La señal de audio analógica es una señal de carácter continuo, es decir, si se pretendiera obtener el valor intermedio entre dos puntos cualesquiera de la señal, dicho punto se podría obtener.



Figura 1. 21 Audio analógico

1.4.3 DIGITALIZACIÓN DE SEÑALES ANALÓGICAS

Las señales denominadas digitales son aquellas señales en las que la información está representada por valores discretos, implican un conjunto discreto de valores tanto en tiempo como en amplitud. Una señal digital puede tomar ciertos valores entre dos máximos, es decir, tomar un número finito de niveles o estados entre un máximo y un mínimo.

Una señal continua es convertida en una señal digital mediante un proceso denominado digitalización, el cual es un proceso que depende de varias operaciones como: muestreo, cuantización y codificación. Además de estas tres operaciones se considera una operación más, la misma que se puede añadir dependiendo del tipo de señales a procesar; esta etapa es la del filtrado. El filtrado puede resultar importante en ciertos casos, así, un ejemplo ilustrativo de una señal con rango de frecuencias limitado por su propia naturaleza sería el de los sistemas de audio, cuyo rango de frecuencias queda limitado a menos de 20.000 Hz, pues sobre los 20 KHz se supera los límites de percepción del oído humano.

1.4.3.1 Muestreo

El muestreo (en inglés, *sampling*) consiste en tomar muestras de la señal continua cada cierto tiempo a intervalos constantes. El muestreo consiste en retener en la salida de un circuito denominado *sample & hold*, al valor analógico de voltaje de entrada durante un lapso de tiempo denominado periodo de muestreo (T_s). La separación de los impulsos en el dominio del tiempo se conoce como periodo de muestreo y la separación en el dominio de la frecuencia se conoce como frecuencia o tasa de muestreo.

$$f_s = 1/T_s \quad (\text{Ec. 1.10})$$

Donde, f_s es la frecuencia de muestreo y representa el número de muestras por segundo, y tiene una vinculación con la respuesta de frecuencia del audio.

Cuanto más rápidos son los cambios temporales que experimenta la señal, más elevada debe ser la frecuencia de muestreo a fin de evitar que se produzca una pérdida de información significativa. En una forma más precisa, el muestreo consiste en multiplicar una señal analógica por un tren de impulsos unitarios discretos, tal como se muestra en la figura 1.22.

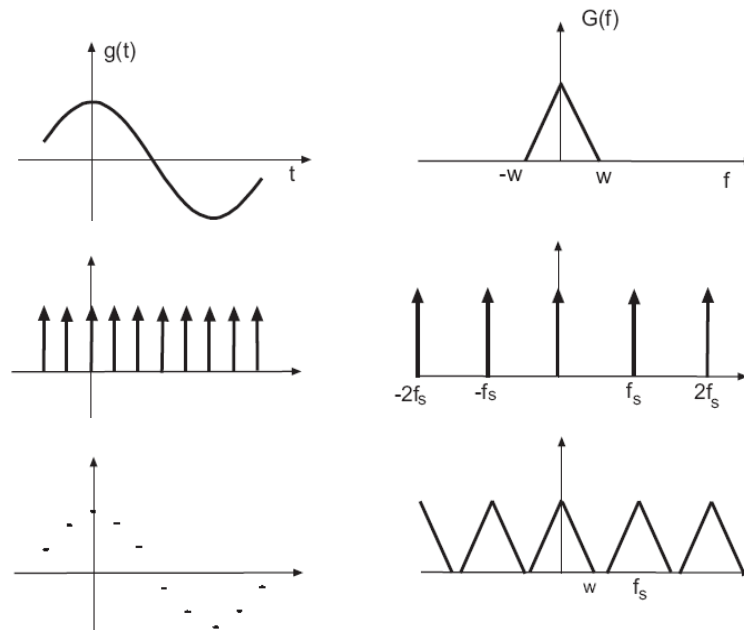


Figura 1. 22 Muestreo de una señal analógica

1.4.3.1.1 Teorema del muestreo

Para que la señal original pueda ser correctamente analizada en frecuencia, según el denominado teorema de Nyquist, es psicoacústicamente necesario muestrear la señal analógica con una frecuencia de muestreo mayor o igual a 2 veces la máxima frecuencia contenida en la señal. Así se tiene:

$$f_s \geq 2f_m \quad (\text{Ec. 1.11})$$

A esta frecuencia de muestreo que cumple con el teorema de Nyquist se la suele denominar frecuencia o tasa de Nyquist. A mayor valor en la tasa de Nyquist, no se gana nada en la determinación de las componentes de frecuencia. En cambio, si interesa analizar o recuperar la información de la forma de onda se debe al menos muestrear a una $f_s = 10f_m$.

Cuando el sonido se digitaliza en alta fidelidad, por lo tanto, debe ser muestreado en un poco más de la tasa de Nyquist de $2 \times 22.000 = 44.000$ Hz. Esta es la razón por la cual el sonido digital de alta calidad se basa en 44.100 Hz. Cualquier frecuencia de muestreo inferior a esta tasa resulta en distorsión, mientras tasas más altas no producen mejoras en la reconstrucción (reproducción) del sonido. Se puede considerar la tasa de muestreo de 44.100 Hz como un filtro pasa bajo, ya que elimina de forma efectiva todas las frecuencias por encima de 22.000 Hz.

En términos informáticos, una mayor frecuencia de muestreo requiere una mayor resolución (número de bits).

La máxima frecuencia de sampleo (muestreo) utilizada en sistemas de audio actualmente es de 192 KHz.

Las frecuencias de muestreo estándar son:

- 8.000 Hz para transmisión de voz (teléfono)
- 11.025 Hz para transmisión de voz (mayor Calidad)
- 22.050 Hz para transmisión de música
- 32.000 Hz para la radiodifusión sonora
- 44.100 Hz para la producción musical, transmisión de música (alta calidad CD)
- 48.000 Hz para la post-producción de audio para imagen
- 96.000 Hz en DVD
- 192.000 Hz en DVD Audio

1.4.3.1.2 *Aliasing*

Si la tasa de muestreo no es suficiente, las réplicas del espectro se traslaparán entre sí y no será posible reconstruir el espectro original al filtrar con el filtro pasa bajos. Este fenómeno se conoce como *aliasing*, aliasión o solapamiento, que se muestra en la figura 1.23 y en la figura 1.24.

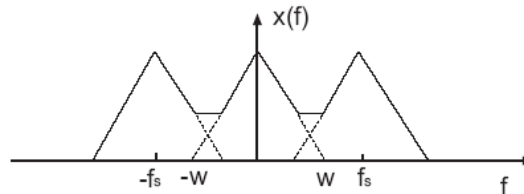


Figura 1. 23 Aliasing en el dominio de la frecuencia

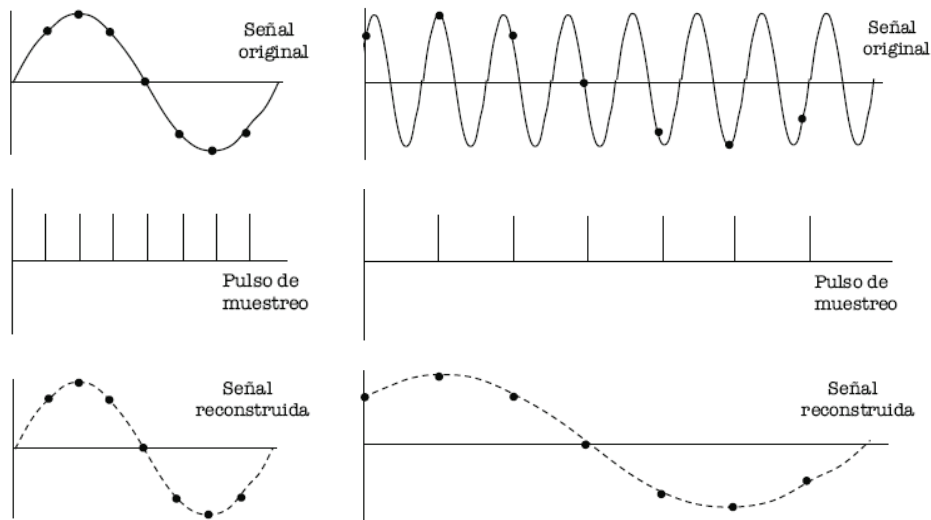


Figura 1. 24 Aliasing en el dominio del tiempo

Claramente la señal reconstruida en el caso submuestreado se encuentra distorsionada respecto a la original. Si se trata de una señal estocástica y no se tiene conocimiento de su frecuencia máxima (f_m), se debe limitar dicha f_m por medio de un filtro (denominado anti-aliasing).

Para eliminar el aliasing, los sistemas de digitalización incluyen filtros pasa bajos, que eliminan todas las frecuencias que sobrepasan la frecuencia que corresponde a la mitad de la frecuencia de muestreo elegida en la señal de entrada. Es decir, todas las frecuencias que queden por encima de la mitad de la frecuencia de muestreo seleccionada son eliminadas.

1.4.3.1.3 Jitter

El jitter se produce cuando las muestras que deberían llegar cada cierto intervalo, llegan a destiempo. El jitter afecta a la reproducción de las frecuencias y está en el orden de 1 a 100 nanosegundos.

1.4.3.1.4 Cuantización

La cuantización también conocida como cuantificación, profundidad de bits, o resolución, es el proceso que tiene como finalidad discretizar el dominio de la amplitud de una señal muestreada. Una vez muestreada la señal, a cada valor de amplitud único por periodo de muestreo se le asigna un código binario (de no coincidir exactamente el valor de amplitud original con los valores discretos en la escala de amplitud, se aproximará el valor al más cercano en la escala).

Los valores continuos de la señal son aproximados a 2^n niveles de amplitud cuantizados, donde n corresponde al número de bits disponible. La resolución de la señal, por consiguiente, tendrá relación con el número de niveles que se tenga para codificar. En el caso del compact disc o CD, se utilizan 16 bits para representar la amplitud. Esto significa que hay $2^{16} = 65.536$ niveles distintos para representar la amplitud.

En caso de considerar una cuantización de tres bits, la señal que se visualizaría sería la de la figura 1.25.

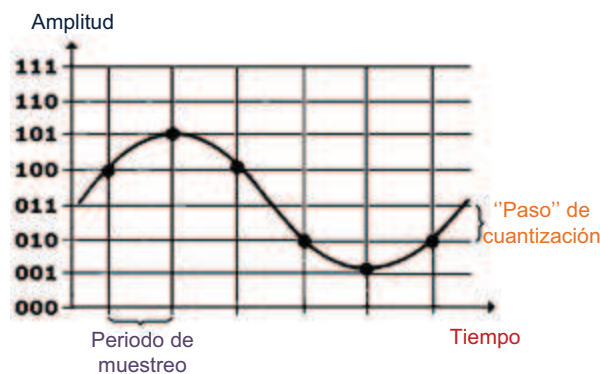


Figura 1. 25 Cuantización con tres bits

1.4.3.1.5 Ruido de cuantización

La figura 1.26 muestra el proceso de discretización de la amplitud de la señal e indica cómo se introduce un error entre el valor real de la señal analógica y el valor con que se codificará la muestra una vez digitalizada.

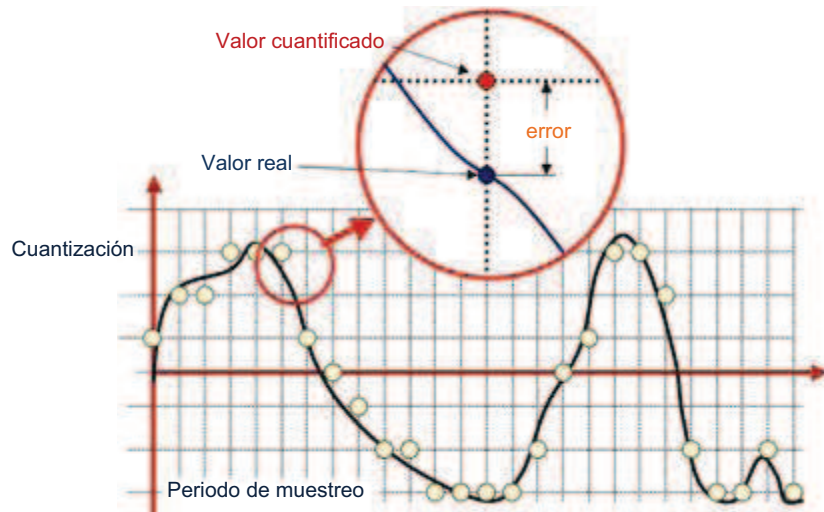


Figura 1. 26 Cuantización y ruido de cuantización

El proceso de digitalización introduce por tanto un error aleatorio en la amplitud de la señal que es equivalente a la adición de una componente de ruido. En efecto, se puede suponer que el valor cuantificado corresponde al de la señal original más un ruido “virtual”, que se ha superpuesto con la señal, dando lugar al valor que realmente se adquiere. El ruido puede ser tanto positivo como negativo y su valor máximo es igual a la mitad del paso de cuantificación.

Debido al error o ruido de cuantización se tiene la relación señal a ruido de cuantificación SNR_q:

$$SNR_q = \frac{E[x^2(n)]}{E[(x(n) - x_q(n))^2]} \quad (\text{Ec. 1.12})$$

Donde, $E[\]$ = valor esperado

$x(n)$ = señal de entrada al cuantizador, que corresponde a la señal que sale del muestreador, esto es: $x(n) = x_s(t)$

$x_q(n)$ = señal de salida del cuantizador

El error de cuantización $e(n)$ corresponde por tanto a la diferencia de valores entre la amplitud de la señal actual (muestra) y su correspondiente cuantificada:

$$e(n) = x_q(n) - x(n) \quad (\text{Ec. 1.13})$$

1.4.3.1.6 Dithering

En el mundo del audio digital, aplicar “dither” significa mezclarle “ruido” de manera controlada a una señal.

De ser una onda sinusoidal perfecta, por ejemplo, al ser cuantizada se convierte en una señal a pasos, que en vez de ser una onda con curvas perfectas, tiene forma de escalera. Esto se debe al error de cuantización y generalmente, hace que el audio suene mal por haber perdido resolución. Añadir entonces un poco de ruido blanco a ésta señal hará más “suave” la forma escalonada de una señal cuantizada creando un sonido mucho más natural. Más allá de los 24 bits de cuantización en una señal, ya no es necesario aplicar dithering.

En la figura 1.27 se puede apreciar la diferencia entre una señal sin dither y con dither.

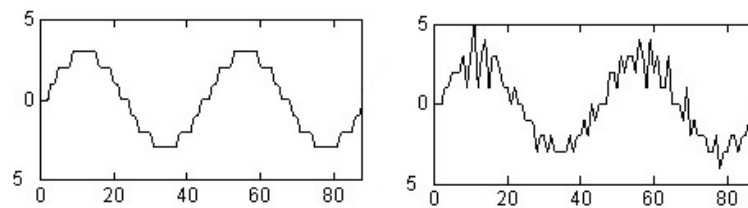


Figura 1. 27 Izquierda, señal sin dither. Derecha, señal con dither

1.4.3.1.7 Cuantización uniforme

Se tiene una cuantización uniforme cuando se establecen intervalos de cuantización iguales de forma uniforme para todo el margen de amplitudes de la señal como se puede apreciar en la figura 1.28.

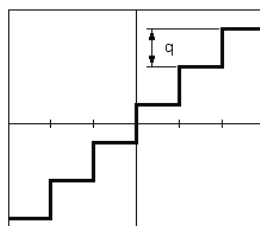


Figura 1. 28 Cuantización uniforme

Cuando las señales de entrada son pequeñas pueden resultar diferencias relativamente grandes entre el valor de una muestra y su nivel de cuantización asignado. Siendo "q" el tamaño del intervalo de cuantización, el error de cuantización $e(n)$ puede tomar valores dentro del rango:

$$-\frac{q}{2} \leq e(n) \leq \frac{q}{2} \quad (\text{Ec. 1.14})$$

Además, para cuantización uniforme la relación señal a ruido de cuantización es:

$$SNRq = \frac{V^2}{q^2/12} \quad (\text{Ec. 1.15})$$

$$SNRq = 10 \log \frac{V^2}{q^2/12} = 10,8 + 20 \log \frac{V}{q} \quad (\text{Ec. 1.16})$$

Donde, V = valor rms de la señal de entrada
 q = tamaño del intervalo de cuantización

1.4.3.1.8 Cuantización no uniforme

A fin de aumentar la relación señal a ruido de cuantización, es necesario tener niveles de cuantización más cercanos para niveles de señal de menor amplitud y niveles de cuantización más espaciados para niveles de señal de mayor amplitud, a un proceso de esta naturaleza se denomina *cuantización no uniforme* y se aprecia en la figura 1.29.

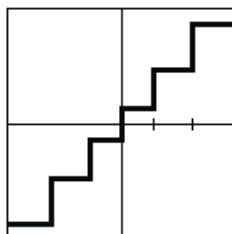


Figura 1. 29 Cuantización no uniforme

Un cuantizador no uniforme usualmente se obtiene pasando la señal a través de dispositivos no lineales (*compander*), seguido por un cuantizador uniforme.

Así es muy usual utilizar compansores logarítmicos. En telefonía digital, la UIT-T (según recomendación G.711) especifica dos esquemas de compansión logarítmica, denominados Ley A y Ley μ .

1.4.3.1.9 Cuantización Ley A

Usado en Europa, África y Sudamérica (Ecuador). Utiliza un cuantizador no uniforme, es decir, para valores de señal inferiores a cierto umbral utiliza una función lineal; y, por encima de ese margen se usa una función no lineal (logarítmica). La característica de transferencia consta de 13 segmentos, cada uno de ellos con 16 niveles, dando como resultado 256 niveles en total, 128 de polaridad positiva y 128 de polaridad negativa, como se muestra en la figura 1.30.

Se emplean 8 bits para representar los niveles, de modo que el codificador maneja palabras de 8 bits.

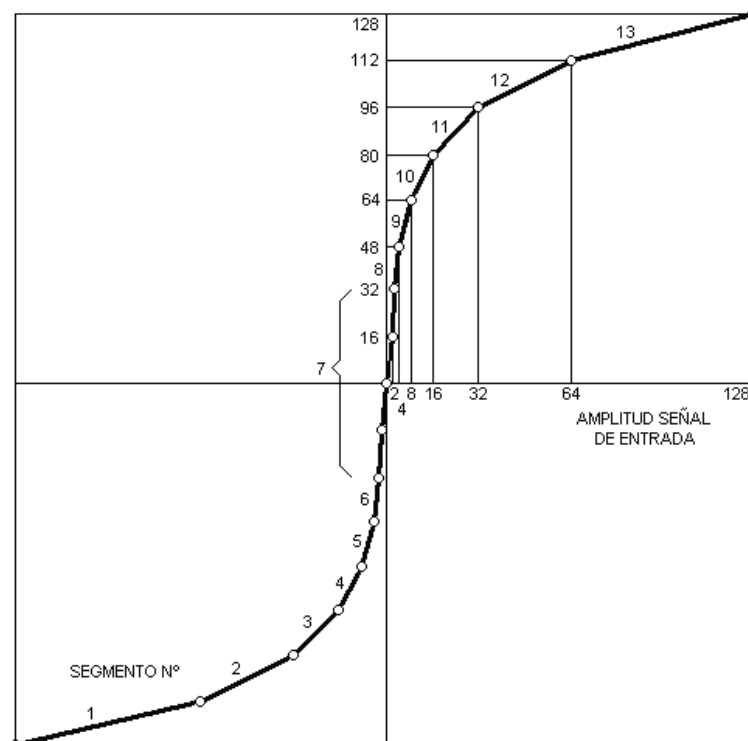


Figura 1. 30 Cuantizador ley A

Siendo x la señal que ingresa al cuantizador, entonces la señal de salida es:

$$F_A(x) = \begin{cases} \operatorname{sgn}(x) \frac{A|x|}{1+\ln x} & 0 \leq |x| \leq \frac{1}{A} \\ \operatorname{sgn}(x) \frac{1+\ln(A|x|)}{1+\ln(A)} & \frac{1}{A} \leq |x| \leq 1 \end{cases} \quad (\text{Ec. 1.17})$$

Donde, $A =$ coeficiente numérico igual a 87,6

$\operatorname{Sgn}(x) =$ función signo de x

La palabra PCM de ocho bits (b1,b2,b3,b4,b5,b6,b7,b8) que entrega el codificador representa: b1 polaridad (1L positiva, 0L negativa). Los tres bits restantes b2, b3, b4 especifican el segmento en cuestión y, los últimos cuatro bits b5, b6, b7 y b8 discriminan el intervalo de cuantificación dentro del segmento especificado.

1.4.3.1.10 Cuantización Ley μ

Usado en los EEUU, Japón y Canadá. Tiene una característica de transferencia de 15 segmentos, que representa una función logarítmica para todo el margen de valores de una señal a comprimir. Maneja 256 intervalos de cuantificación y consecuentemente el codificador usa palabras de 8 bits.

Si x es la señal que ingresa al cuantizador, la salida del mismo será:

$$F_A(x) = \operatorname{sgn}(x) \frac{\ln(1+\mu|x|)}{\ln(1+\mu)}, \quad \text{donde} \quad \operatorname{sgn}(x) \begin{cases} +1 & x \geq 0 \\ 0 & x = 0 \\ -1 & x \leq 0 \end{cases} \quad (\text{Ec. 1.18})$$

Donde, el coeficiente numérico $\mu = 255$

La ley A es inferior a la ley μ , en términos de calidad para pequeñas señales.

1.4.3.2 Codificación

La codificación consiste en traducir los valores obtenidos durante la cuantificación al código binario. Hay que tener presente que el código binario es el más utilizado, pero también existen otros tipos de códigos que también son utilizados.

La codificación asigna una secuencia de bits a cada valor analógico discreto. La longitud de la secuencia de bits es función del número de niveles analógicos empleados en la cuantificación. Así la señal analógica va a quedar transformada en un tren de impulsos digital (sucesión de ceros y unos). En la figura 1.31 se representa esquemáticamente el proceso de codificación de una señal. A cada muestra se le asigna el código binario correspondiente que representará el nivel de amplitud de la señal.

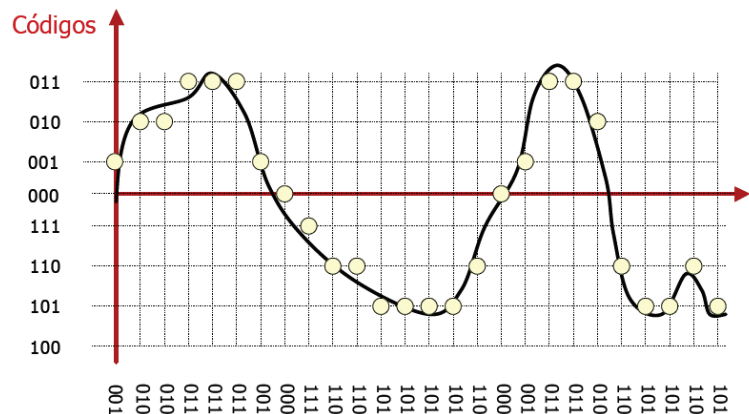


Figura 1. 31 Ejemplo de codificación de una señal

1.4.3.3 Ventajas y desventajas de señales y sistemas digitales

1.4.3.3.1 Ventajas

- Menores costos de procesamiento.
- Inmunidad al ruido.
- Capacidad de aplicar métodos de corrección de errores o recuperación de la señal.
- Posibilidad de encriptar la información.
- Procesamiento y tratamiento de señales de gran versatilidad.
- Los archivos de audio digitales se guardan y procesan más fácilmente que los archivos de audio analógico.
- Posibilidad de almacenar grandes cantidades de datos en diferentes soportes.
- Las grabaciones no se deterioran con el paso del tiempo como sucede con las cintas analógicas.

- Permite realizar regrabaciones sucesivas sin que se pierda ningún dato y, por tanto, calidad.
- Permite la compresión para reducir la capacidad de almacenamiento.
- Facilita la edición visual de las imágenes y del sonido en un ordenador o computadora personal, utilizando programas apropiados.

1.4.3.3.2 Desventajas

- Para su transmisión requiere un mayor ancho de banda en comparación con una señal analógica.
- La conversión A/D (Analógico/Digital) y D/A (Digital/Analógico) introduce ruido de cuantización.
- La sincronización entre los relojes de un transmisor inalámbrico digital y el receptor debe ser precisa.
- Las transmisiones de señales digitales son incompatibles con las instalaciones existentes para transmisiones analógicas.

1.4.4 CARACTERÍSTICAS DEL AUDIO DIGITAL

Tanto como una imagen puede ser digitalizada y dividida en píxeles, donde cada píxel (picture element) es un número, el sonido también puede ser digitalizado y dividido en números. De esta manera el denominado "audio digital" es una representación discreta (no continua) del sonido. El audio digital basa su funcionamiento en dos propiedades básicas: la *frecuencia de muestreo* y la *cuantización*. Una señal de audio digital, consiste básicamente en una secuencia discreta de números, el audio digital es la representación de señales sonoras mediante un conjunto de datos binarios.

Las computadoras actuales permiten, en general, manejar audio digital. Es decir, permiten almacenar, procesar y reproducir sonidos de diversos tipos.

1.4.4.1 La computadora en sistemas de audio digital

Como se conoce, cualquier tipo de información que sea manejado por una computadora es almacenado en ésta mediante ceros y unos (información digital). Esos ceros y unos pueden representar texto, imágenes, sonido o cualquier otro tipo de elemento factible de ser procesado digitalmente. Dado que los computadores son binarios, es necesario digitalizar el audio para que el computador pueda leer la información sonora. En otras palabras, un computador solo puede procesar audio digital. Esto no significa que el computador no pueda interactuar con audio analógico.

Para digitalizar el sonido se pueden utilizar diversos dispositivos, pero la solución más económica, en el caso de una PC (personal computer), generalmente consiste en utilizar una tarjeta de sonido (ver figura 1.33). Una computadora puede generar y tratar sonidos de varias formas diferentes.

La figura 1.32 muestra las tres formas más usuales del uso del computador en sistemas de audio.

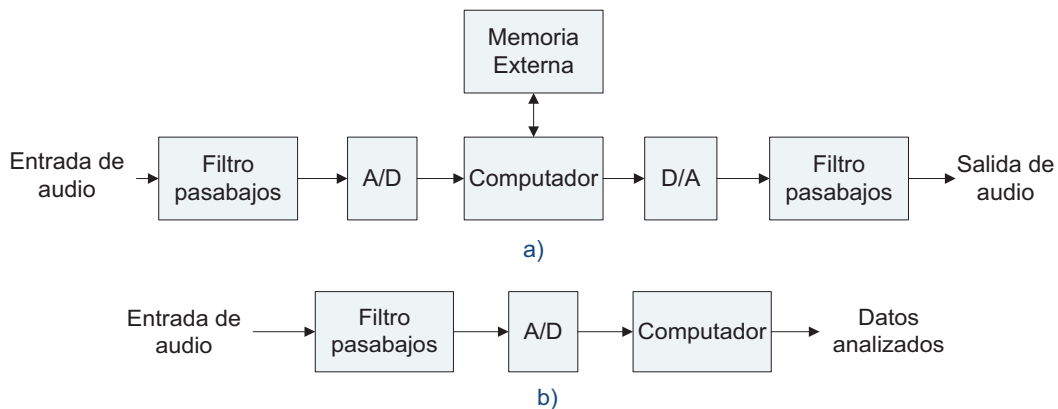


Figura 1. 32 La computadora en sistemas de audio digital

El caso 1.32 a) corresponde al proceso de grabación o procesamiento digital. El audio analógico es captado por un transductor, por ejemplo un micrófono, que convierte las variaciones de presión del aire (energía acústica) en una señal eléctrica, que consiste básicamente en variaciones de voltaje. Esta señal es luego filtrada por un filtro pasa bajos con el fin de eliminar componentes de frecuencia

no deseadas y eliminar la posibilidad de aliasing. Una vez filtrada la señal, ésta es digitalizada (muestreada) por un conversor analógico/digital. A la salida del conversor se obtiene audio digital, que puede ser procesado o guardado en la memoria externa por la computadora.

Un ejemplo de procesamiento de la señal podría ser aplicar reverberación al sonido original o mezclarlo con otro previamente almacenado en la memoria externa. Una vez realizado el proceso de digitalización, el audio es introducido esta vez en un conversor digital/analógico y pasado por un filtro pasa bajos. La salida de este filtro produce audio analógico que podría ser amplificado y enviado nuevamente al aire mediante los parlantes.

La grabación digital tiene varias ventajas por sobre su similar analógica. Dado que, la información digitalizada contiene solo números y no una señal analógica, tiene una calidad superior ya que no se deteriora con el tiempo ni depende de variables como la temperatura ambiente, presión atmosférica, viscosidad del aire o el ruido ambiental. Además, una copia digital del archivo de audio original digital es exactamente idéntica y fiel.

El caso 1.32 b) corresponde al análisis de señales. Cuando la computadora opera como un analizador de señales, la computadora toma una señal digital y matemáticamente determina sus características. Por ejemplo, un análisis computacional puede revelar información importante sobre las propiedades acústicas de un determinado sonido.

1.4.4.2 Hardware para audio digital

El dispositivo principal empleado en la actualidad que permite la adquisición, reproducción y procesamiento del audio tanto analógico como digital es la tarjeta de sonido, la misma que se observa en la figura 1.33; y que tiene ciertas funciones básicas las mismas que son:

- Reproducción de archivos de sonido desde el disco duro o lectores externos.
- Captura y almacenamiento de audio procedente de una fuente externa.
- Síntesis y generación de sonidos a partir de información (MIDI).
- Procesamiento de sonidos almacenados en el disco duro.

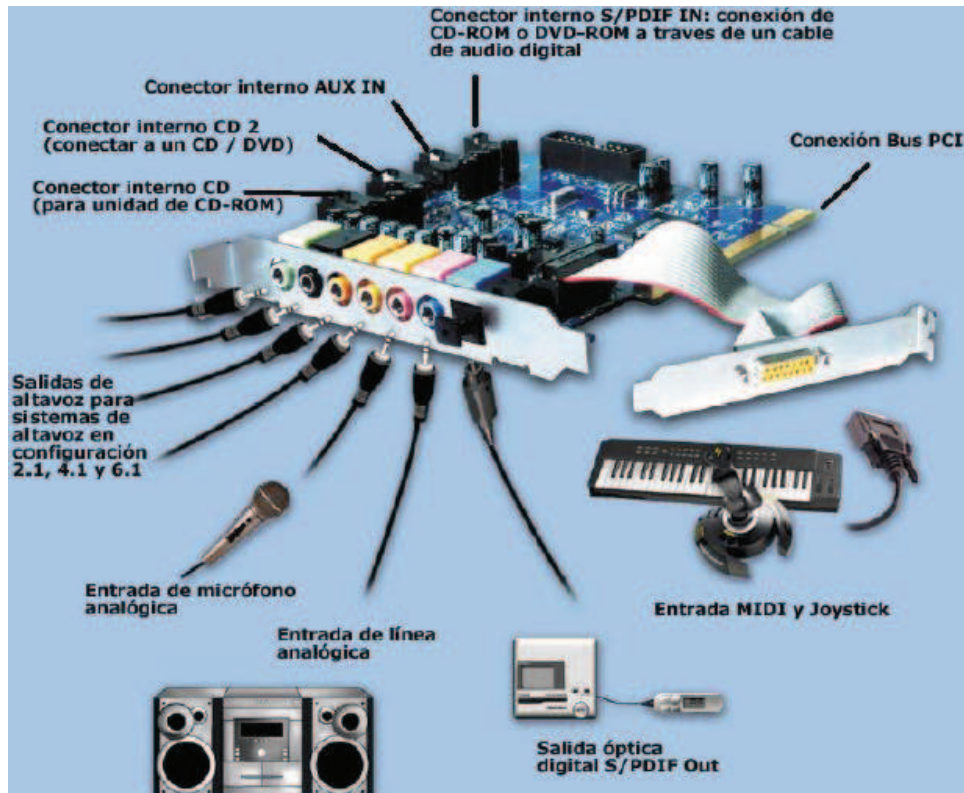


Figura 1. 33 Tarjeta de sonido

1.4.4.2.1 Funcionamiento de la tarjeta de sonido

El núcleo de cualquier tarjeta de sonido está formado por tres subsistemas:

- Convertidor analógico/digital (CAD).
- Procesador digital de señales (DSP).
- Convertidor digital/analógico (CDA).

Los elementos de interfaz son:

- Entrada de micrófono.

- Entrada de línea, permite introducir señales procedentes de otras fuentes de audio (reproductores de cinta, receptores de radio, etc.).
- Salida de audio, monofónico, estéreo o multicanal.
- Algunas tarjetas proporcionan entradas y salidas digitales que permiten introducir la información de audio directamente en formato digital.

1.4.4.3 Parámetros de audio digital

Los parámetros básicos para describir la secuencia de muestras que representa el sonido son:

El número de canales: 1 canal para monofónico, 2 canales para estéreo, 4 para el sonido cuadrafónico, etc. Como regla general, las muestras de audio multicanal suelen organizarse en tramas. Una trama es una secuencia de tantas muestras como canales, correspondiendo cada una a un canal. En este sentido el número de muestras por segundo coincide con el número de tramas por segundo.

En estéreo, el canal izquierdo suele ser el primero. Cada canal del audio puede contener una información independiente a los demás canales. Se suele almacenar el sonido en más de un canal para así tener más información del mismo. Cada canal podría contener la grabación de un mismo sonido realizada con distintos micrófonos, o simplemente efectos distintos que se le quieran añadir al sonido, ya que todos los canales pueden ser reproducidos al mismo tiempo.

Tasa de muestreo: es el número de muestras tomadas por segundo en cada canal.

Número de bits por muestra: habitualmente son 8 o 16 bits.

Teniendo en cuenta los diversos parámetros del audio digital, el espacio requerido en disco duro para la grabación de información de audio digital a distintas frecuencias de muestreo, cuantizaciones y cantidad de canales de audio es:

Número de pistas y duración	16 bits a 44,1 KHz	16 bits a 48 KHz	24 bits a 44,1 KHz	24 bits a 48 KHz	16 bits a 88,2 KHz	16 bits a 96 KHz	24 bits a 88,2 KHz	24 bits a 96 KHz
1 pista mono 1 minuto	5 MB	5,5 MB	7,5 MB	8,2 MB	10 MB	11 MB	15 MB	16,4 MB
1 pista estéreo 1 minuto	10 MB	11 MB	15 MB	16,5 MB	20 MB	22 MB	30 MB	33 MB
1 pista estéreo 60 minutos	606 MB	660 MB	908 MB	989 MB	1,2 GB	1,3 GB	1,8 GB	2 GB
32 pistas mono 5 minutos	808 MB	880 MB	1,2 GB	1,3 GB	1,6 GB	1,8 GB	2,4 GB	2,6 GB
32 pistas mono 60 minutos	9,4 GB	10,4 GB	14 GB	15,5 GB	18,9 GB	20,6 GB	28 GB	31 GB

MB = Megabyte (1 MB = 10⁶ bytes)

Tabla 1. 5 Espacio en disco duro para la grabación de audio digital

1.4.4.4 Calidad del audio digital

La calidad del audio digital depende principalmente de los parámetros con los que la señal de sonido ha sido adquirida, pero no son los únicos parámetros importantes para determinar la calidad.

Una forma de estimar la calidad del audio digital es analizar la diferencia entre el sonido original y el sonido reproducido a partir de su representación digital. Para los sistemas de audio ésta medida estará determinada por el número de bits por muestra y la tasa de muestreo.

La calidad que percibe el oyente está determinada por la respuesta del oído humano a las ondas sonoras, es decir, si el audio digital es de buena calidad el oyente no percibe muchas de las posibles diferencias. Lógicamente si las señales son muy parecidas, el oído no las podrá diferenciar, pero también pueden ser muy distintas y ser percibidas como la señal original. Por lo tanto, parece más apropiada la evaluación de la calidad de un sistema digital mediante parámetros de sensibilidad del oído humano y pruebas específicas con oyentes especializados.

1.4.5 FORMATOS DE FICHERO

Los formatos de fichero indican la estructura con la que el audio es almacenado. En los comienzos del audio digital aparecieron multiplicidad de formatos de audio y cada sistema determinaba el formato que utilizaba, con el tiempo el conjunto de formatos usados se redujo mediante la aparición de formatos cada vez más flexibles y eficientes, algunos formatos son usados de forma masiva, mientras que otros tienen usos muy reducidos. Los formatos de fichero no tienen por qué coincidir con las características del reproductor. En general un mismo formato de fichero permite contener diversas codificaciones, tasas de muestreo, etc. En este sentido se distingue entre dos grupos de formatos de ficheros de audio:

- a) Formatos autodescriptivos: contienen de forma explícita los parámetros del dispositivo y la codificación en algún punto del fichero.
- b) Formatos sin cabecera o tipo raw: los parámetros del dispositivo y codificación empleada son fijos.

1.4.5.1 Formatos de fichero autodescriptivos

Los formatos de fichero autodescriptivos, suelen permitir la elección entre varias codificaciones, de entre las cuales se especifica la utilizada en una cabecera. La cabecera suele comenzar por lo que se conoce como un número mágico, que no es más que un valor fijo que permite identificar el fichero como un fichero del formato buscado. Esta cabecera suele contener la tasa de muestreo, el número de bits por muestra, si las muestras tienen signo o no, si se colocan en formato little-endian⁷ (último el bit menos significativo LSB) o big-endian⁸ (último el bit más significativo MSB), y otro tipo de información como la descripción del sonido que contiene o notas de copyright.

⁷ Formato Little Endian (LSB MSB), se almacena primero el dato más pequeño; el byte de menor peso se almacena en la dirección más baja de memoria y el byte de mayor peso en la más alta.

⁸ Formato Big Endian (MSB LSB), se almacena primero el dato más grande; el byte de mayor peso se almacena en la dirección más baja de memoria y el byte de menor peso en la más alta.

Existe un subconjunto de estos ficheros formados por aquellos que en lugar de tener una cabecera con la información de codificación, organizan el fichero en bloques de datos y bloques de información de codificación, intercalando unos y otros. Estos ficheros permiten la utilización de diversas codificaciones de datos a lo largo de un mismo fichero.

La tabla 1.6 muestra una relación de algunos de los formatos de fichero de audio autodescriptivos más habituales e indica algunos de los parámetros que permiten modificar.

EXTENSIÓN	NOMBRE	ORIGEN	PARÁMETROS MODIFICABLES				COMENTARIOS
			Tasa	Canales	Codific.	Otra info.	
.au, .snd		NeXT, SUN	x	x	x	x ⁽¹⁾	
.aif, .aiff	AIFF	Apple, SGI	x	x	x ⁽²⁾	x	
.aif, .aiff	AIFC	Apple, SGI	x	x	x ⁽²⁾	x	AIFF con compresión
.iff	IFF/8SVX	AMIGA	x	x		x	Info. de instrumento, 8 bits
.mp2, .mp3	MPEG	MPEG	x	x	x		
.ra	Real Audio	Real Networks	x	x	x		
.sf	IRCAM		x	x	x	x	
.smp		Turtle Beach					16 bits/1 canal, bucles
.voc		SoundBlaster	x				8 bits/1canal, puede detectar el silencio
.wav	WAVE	Microsoft	x	x	x ⁽²⁾	x	
.wve		Psion					8kbps, 1 canal, ley- A, 8 bits
(ninguna)	HCOM	Mac	x				8 bits/1 canal, comp. Huffman

(1) Una cadena de información

(2) Sólo longitud de muestra

Tabla 1. 6 Formatos de fichero autodescriptivos [3]

1.4.5.2 Formatos de fichero sin cabecera o tipo “raw”

Estos formatos definen un único esquema de codificación y no permiten la variación de los parámetros salvo, en algunos casos, la tasa de muestreo. De hecho, muchas veces no se puede conocer de ninguna forma la tasa de muestreo empleada a menos que se escuche el sonido.

Estos formatos son menos importantes que los autodescriptivos, por ser menos flexibles. Hoy en día están prácticamente en desuso, aunque en el pasado fueron los primeros en aparecer.

REFERENCIAS BIBLIOGRÁFICAS

CAPITULO 1: ESTUDIO DEL SONIDO

LIBROS:

- [1] Guyton, Arthur C., y John E. Hall. (2002). Tratado de Fisiología médica. Capítulo 52, El sentido de la audición. Décimo primera edición. México: McGraw-Hill, Interamericana.
- [2] Salomon, D. (2007). Data Compression. Chapter 7, Audio Compression. Fourth Edition. London: Springer-Verlag.
- [3] López, A. Formatos de Audio Digital. Ingeniería de Ondas. España.

INTERNET:

- [4] Asinsten, J. El sonido. Ministerio de Educación Ciencia y Tecnología.
<http://coleccion.educ.ar/coleccion/CD13/contenidos/materiales/archivos/sonido.pdf>
- [5] Bruscianelli, C. Electroacústica. Universidad Simón Bolívar, Departamento de Electrónica.
<http://www.labc.usb.ve/EC4514/AUDIO/Indice.html>
- [6] Cádiz, R. (2008). Introducción a la Música Computacional. Centro de Investigación en Tecnologías de Audio, Instituto de Música, Pontificia Universidad Católica de Chile.
<http://www.rodriocadiz.com/imc/>
- [7] ESPOL. (2009). Procesamiento de Audio y Video. Facultad de Ingeniería en Electricidad y Computación, ESPOL OCW.
<http://www.ocw.espol.edu.ec/facultad-de-ingenieria-en-electricidad-y-computacion/procesamiento-de-audio-y-video-1/course-schedule>

CAPÍTULO 2.

COMPRESIÓN DE AUDIO

2.1 COMPRESIÓN DE LA INFORMACIÓN

2.1.1 INTRODUCCIÓN

Con la llegada de las computadoras personales de bajo costo en los años 1980 y 1990 llegaron las aplicaciones multimedia, donde texto, imágenes, videos y audio se almacenan en el computador, y se pueden cargar, descargar, editar, y reproducir. En la actualidad es impensable la información multimedia en formato sin compresión. Principalmente, porque día a día son más las aplicaciones que requieren de un amplio rango de calidad y funcionamiento de acuerdo a los requerimientos de los distintos usuarios. Afortunadamente, un gran número de investigaciones durante las últimas décadas han conducido a muchas técnicas y algoritmos de compresión que hacen factible la transmisión de multimedia. Es el resultado de más de 50 años de investigación.

Muchas de las técnicas de compresión actuales, que nacieron a inicios de los años noventa, compiten entre sí o son complementarias. Los métodos de compresión modernos están basados en un compromiso entre criterios como: que la calidad de los datos codificados, y luego decodificados, sea tan buena como sea posible; que se logren tasas de compresión tan altas como se pueda; y una adecuada relación costo-beneficio.

La tendencia creciente hacia la estandarización de formatos de compresión, han hecho posible una compatibilidad casi universal entre distintos sistemas de comunicación. El paso de la era industrial a la sociedad de la información, se sustenta en los avances en redes, computadoras y telecomunicaciones.

La existencia de diferentes tipos de archivos con los cuales trabajamos diariamente y en ocasiones, de manera simultánea; han hecho que los algoritmos de compresión se desarrollen en correspondencia al tipo de archivos y de acuerdo a determinadas relaciones de compresión requeridas. Así por ejemplo para archivos de texto, audio, imágenes y video existen los algoritmos correspondientes. En general se puede decir que el texto no ocupa mucho espacio en el computador, así por ejemplo un libro promedio, consiste de un millón de caracteres y se puede almacenar sin comprimir en alrededor de 1 MB, ya que cada carácter de texto ocupa un byte. En contraste, las imágenes ocupan mucho más espacio, dando otro significado a la frase "una imagen vale más que mil palabras". Dependiendo del número de colores utilizados en una imagen, un solo píxel ocupa entre un bit y tres bytes. En cuanto a la voz, audio y video se consideran medidas cuantitativas de distorsión perceptiva para así obtener una compresión eficiente.

2.1.2 NECESIDAD PARA LA COMPRESIÓN

Actualmente existen dispositivos con una gran capacidad de almacenamiento (discos duros, DVDs, memorias flash, etc.) pudiendo almacenar en ellos, muchos datos de gran tamaño. Sin embargo, la complejidad y volumen de los datos aumenta en forma casi paralela al aumento de las capacidades de estos medios de almacenamiento, por ello hay que comprimirlos.

Imágenes, video y audio sin comprimir pueden implicar una enorme cantidad de datos y ello puede originar problemas en la transmisión, almacenamiento y procesamiento; por ejemplo una señal de audio sin comprimir, con calidad telefónica, muestreada a 8 KHz y utilizando 8 bits/muestra, implica una tasa de transmisión de 64Kbps. Una foto de 35 mm a una resolución de 2000 x 2000 ocuparía 10 MB. Una señal de audio estéreo sin comprimir y con calidad CD, muestreada a 44,1 KHz utilizando 16 bits/muestra, demanda 1.4112 Mbps. Señales de video para TV y para PC sin comprimir requieren tasas en el orden de las centenas de Mbps, para video de alta definición (HDTV) se requiere tasas en el orden de las unidades de Gbps.

La compresión permite reducir el número de bits a transmitirse, con una menor tasa de bits se consigue una disminución en el tiempo de transmisión y ahorro en ancho de banda. Un archivo comprimido ocupa menos espacio de memoria. La compresión afronta entonces el problema de la reducción de la cantidad de datos, con lo cual se logra disminuir las tasas de datos, sin pérdidas significativas en la calidad. Además el ahorro de recursos como: ancho de banda, espacio de almacenamiento y tiempo; se ven vinculados directamente con ahorro en costos.

Una razón importante en cuanto a la necesidad de compresión se refiere, aparece con las redes de comunicaciones, la Internet es un ejemplo de ello. Esta razón es el tiempo (que también se traduce en dinero) que es necesario para acceder o descargar archivos de gran tamaño por la red. Por tanto, si esos datos están comprimidos se tardarán menos (y se gastará menos dinero) en enviarlos o recibirlos (por correo electrónico, web, ftp⁹ o cualquier otro protocolo de transferencia de datos).

2.1.3 COMPRESIÓN DE DATOS

Comprimir significa reducir el tamaño de algo. El objetivo principal de cualquier esquema de compresión es describir la misma información con un conjunto de datos de menor magnitud.

La compresión de datos se define como el proceso de reducir la cantidad de datos necesarios para representar eficazmente una información, la compresión es la búsqueda para obtener la máxima transferencia (almacenamiento) de información con un mínimo número de bits posible. Si la compresión es eficiente, el resultado ocupará menor espacio que los símbolos originales.

El método de compresión depende intrínsecamente del tipo de datos que se van a comprimir, no se comprime una imagen del mismo modo que un archivo de audio.

⁹ FTP = File Transfer Protocol (protocolo de transferencia de archivos), es un protocolo de la capa aplicación (capa 4) del modelo TCP/IP basado en la arquitectura cliente-servidor utilizado para descargar o enviar archivos.

Cuando se habla de compresión, es útil mencionar algunos términos como los siguientes:

- Relación de compresión (CR): También denominada radio de compresión o factor de compresión, mide cuánto comprime un algoritmo de compresión. La CR en forma de relación n : 1 viene dada por:

$$CR = \frac{\text{número de bits sin compresión}}{\text{número de bits con compresión}} \quad (\text{Ec. 2.1})$$

El número de bits sin compresión representa el tamaño original del archivo (TO) mientras que el número de bits con compresión representa el tamaño comprimido (TC). Por esta razón, expresado como porcentaje la CR viene dada por:

$$CR = \frac{TO-TC}{TO} \times 100 [\%] \quad (\text{Ec. 2.2})$$

Por ejemplo si $CR = 65.536 / 16.384 = 4$; representaría una relación de 4:1.

- Factor de Mérito: Es el inverso de la relación de compresión

$$FR = \frac{1}{CR} \quad (\text{Ec. 2.3})$$

2.1.3.1 Utilización de la compresión

La compresión puede ser utilizada principalmente en transmisión y almacenamiento. La compresión utilizada en transmisión se ilustra a continuación en la figura 2.1.

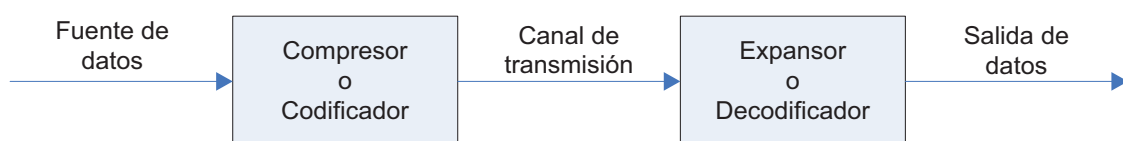


Figura 2. 1: Compresión utilizada en transmisión

Se puede observar en la figura 2.1 que el compresor reduce la tasa de datos de la fuente de información¹⁰, los datos comprimidos atraviesan un canal de comunicación y luego son retornados a la tasa original mediante el expansor. En ocasiones a un compresor y un expansor en cascada se los denomina *compansor*. Al compresor se lo puede llamar codificador y al expansor decodificador en cuáles casos el par puede denominarse *códec*.

El uso de la compresión en aplicaciones de almacenamiento es sumamente poderoso, la figura 2.2 muestra el uso de un códec para almacenamiento de información. Con la ayuda de los diferentes algoritmos de compresión, una mayor cantidad de archivos pueden ser almacenados en diferentes dispositivos de almacenamiento.

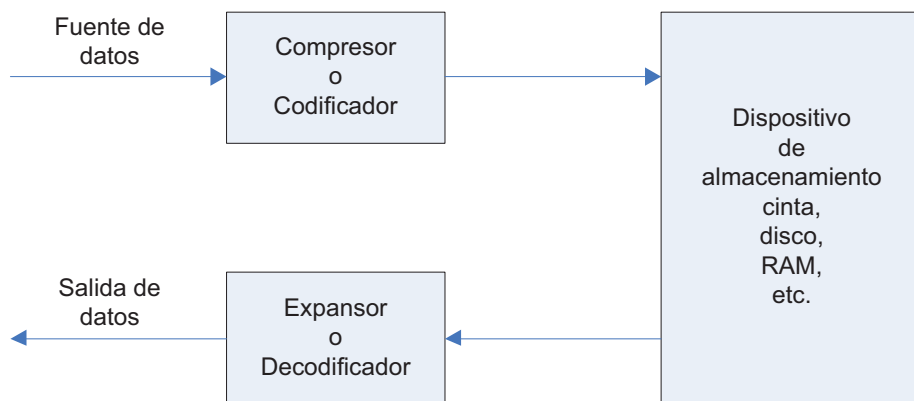


Figura 2. 2: Compresión utilizada en almacenamiento

2.1.3.2 Técnicas de compresión

2.1.3.2.1 Compresión física y lógica

La compresión física actúa directamente sobre los datos; por lo tanto, es cuestión de almacenar los datos repetidos de un patrón de bits a otro. La compresión lógica, por otro lado, se lleva a cabo por razonamiento lógico al sustituir esta información por información equivalente.

¹⁰ Fuente de información: Es aquella que produce un conjunto finito de posibles mensajes si es discreta y mensajes definidos sobre un rango continuo de valores si es continua.

2.1.3.2.2 *Compresión simétrica y asimétrica*

En el caso de la compresión simétrica, se utiliza el mismo método para comprimir y para descomprimir los datos, cada operación requiere la misma cantidad de trabajo. En general, se utiliza este tipo de compresión en la transmisión de datos.

La compresión asimétrica requiere más trabajo para una de las dos operaciones, compresión o descompresión, es frecuente buscar algoritmos para los cuales la compresión es más lenta que la descompresión.

Los algoritmos que realizan la compresión de datos con mayor rapidez que la descompresión pueden ser necesarios cuando se trabaja con archivos de datos a los cuales se accede con muy poca frecuencia (por razones de seguridad, por ejemplo), ya que esto crea archivos compactos.

2.1.3.2.3 *Codificación de fuente y de entropía*

La codificación de fuente tiene las siguientes características:

- Codifica los datos basándose en las características y propiedades de éstos.
- Suelen tolerar pérdidas en la codificación (lossy códecs) que perceptualmente pasan inadvertidas para el usuario.
- Son codificadores de propósito específico.
- De manera general obtienen mayores beneficios que los codificadores basados en la entropía.
- Suelen ser técnicas de compresión con pérdidas.
- Se obtienen tasas de compresión elevadas.

Ejemplos: Codificación Diferencial, Codificación por Transformada (Transformada Discreta del Coseno Modificada - MDCT), etc.

La codificación de entropía¹¹ tiene las siguientes características:

- Codifica los datos sin necesidad de conocer la naturaleza de éstos.
- Es una codificación de propósito general (para todo tipo de datos).
- Son técnicas de compresión sin pérdidas.

Ejemplos: Compresores estadísticos (Huffman, Aritméticos, etc.)

2.1.3.2.4 Compresión con y sin pérdida de información

Existe varias formas de clasificar los diferentes algoritmos y esquemas de compresión en forma general se puede mencionar dos grandes grupos: la compresión con pérdida de información o entropía (reducción de entropía) *lossy* en inglés y la compresión sin pérdida de información o entropía (con pérdida de redundancia) o *lossless* en inglés. A la compresión con y sin pérdida de información se las tratará con detalle, más adelante en este capítulo.

2.1.3.2.5 Codificación no adaptativa, semiadaptativa y adaptativa

Los codificadores no adaptativos están basados en diccionarios¹² para un tipo específico de datos. La repetición de letras en un archivo de texto, por ejemplo, depende del idioma en el que ese texto esté escrito. Un codificador semiadaptativo crea un diccionario según los datos que va a comprimir, es decir, crea el diccionario mientras analiza el archivo y después lo comprime.

Al contrario, un codificador adaptativo se adapta a los datos que va a comprimir, no parte de un diccionario ya preparado para un tipo de datos determinado.

¹¹ Entropía $H(S)$: Es la cantidad media de información por símbolo emitida por una fuente de información de memoria nula, es decir aquella que emite símbolos estadísticamente independientes.

¹² El modelo basado en diccionario usa un código simple para reemplazar cadenas de símbolos, si una cadena concuerda, un indicador en el diccionario puede salir en lugar del código del símbolo.

2.2 COMPRESIÓN LOSSY Y LOSSLESS

2.2.1 COMPRESIÓN CON PÉRDIDA DE INFORMACIÓN (LOSSY)

La compresión con pérdida de información se usa en aplicaciones donde se pueden tolerar diferencias entre los datos originales y comprimidos: voz, música, imágenes y video. Por ejemplo, para imágenes uno de los formatos más conocidos es el JPEG, para video los formatos: MPEG-1, MPEG-2 y MPEG-4 son muy conocidos. En cuanto a audio se refiere existen formatos tanto para el habla y diálogo como también para lo que es música; en cuanto a música los formatos: AAC (usado por Apple Computer), ADPCM, ATRAC, Dolby AC-3, MP2, MP3, WMA son muy utilizados, para habla y diálogo existen varios formatos entre algunos de ellos están: CELP, G.711 (PCM), G.726 (ADPCM).

En la compresión con pérdida de información se introducen medidas de distorsión perceptiva y se fijan los umbrales tolerables, a menudo el observador humano no aprecia las diferencias. Al descomprimir la información, no recuperamos exactamente la original, en el proceso de descompresión se pierde cierta información que no se considera fundamental. De esta forma pareciera que una técnica de compresión que perteneciera a esta categoría difícilmente podría tener una aplicación en la vida real; pero no es así, la explosión de la multimedia y el tráfico de imágenes, audio y videos en la Internet, presente hoy en día se ha dado gracias a los esfuerzos investigativos en este campo.

La pérdida de datos asociada a las técnicas con pérdida se ve compensada por su alta relación de compresión (en imágenes se pueden alcanzar razones de 25:1 sin una pérdida notable de la calidad y en audio se pueden alcanzar razones de 12:1 con el formato MP3). La compresión con pérdida, a diferencia de la compresión sin pérdida, elimina información para lograr la mejor relación de compresión posible mientras mantiene un resultado que es lo más cercano posible a los datos originales.

Los archivos ejecutables, por ejemplo, no pueden comprimirse mediante este método, porque necesitan especialmente preservar su integridad para poder ejecutarse. De hecho, es inconcebible reconstruir un programa omitiendo y después agregando bits.

2.2.1.1 Codificación diferencial

La codificación diferencial se fundamenta en la correlación existente entre muestras sucesivas de la fuente, basándose en las diferencias entre dos símbolos consecutivos, no se codifica la muestra si no la diferencia entre muestras. Ciertos tipos de datos tienen la propiedad de similitud entre símbolos consecutivos: señal de audio, vídeo, imágenes, etc. Esto permite codificar con pocos bits las diferencias. Dentro de este tipo de codificadores se tiene:

- DPCM (PCM Diferencial)
- DM (Modulación Delta)
- Codificación Predictiva

2.2.1.2 Codificación por transformada

Consiste en una codificación por bloques. La señal de entrada se transforma a un dominio diferente, se traslada del dominio del tiempo al dominio de la transformada, y se codifican los coeficientes de la transformación, se evalúa el contenido energético de los coeficientes de la transformada y se transmiten únicamente aquellos coeficientes codificados de mayor aporte energético. En el receptor, el decodificador calcula la transformada inversa para obtener la señal original reconstruida. La transformación más usada es la transformada discreta coseno (sus siglas en inglés, DCT).

La codificación por transformada se utiliza en la codificación de señales de banda ancha de imagen y audio. Sin embargo, no se usa mucho en codificación de voz debido a su complejidad.

Dentro de las transformadas de mayor uso se tiene:

- KL (Karhunen Loevé) o de componentes principales.
- Fourier.
- Hadamard.
- Haar.
- Coseno.
- Ondilla (wavelet).

2.2.1.3 Cuantización vectorial

Opera sobre bloques de datos, en lugar de sobre datos aislados, un bloque de datos se puede ver como un vector (siendo N la dimensión de los vectores), de ahí el nombre.

Se utiliza un libro de códigos conocido como registro de códigos (RC), que usa una tabla, code-book en inglés; que contiene los posibles vectores a la salida de un cuantificador; cada vector del registro de códigos se llama vector de código y su correspondiente índice en la tabla es la palabra código. Se debe definir el número de bits del registro de códigos, la dimensión de los vectores de entrada, y los valores de los vectores de códigos a fin de disminuir el error entre la señal cuantificada y su original. Cuantificar un vector de entrada consiste en determinar su vecino más cercano dentro del registro de códigos, para ello hay que calcular su distancia a todos los vectores del registro de códigos y escoger el de menor distancia.

2.2.2 COMPRESIÓN SIN PÉRDIDA DE INFORMACIÓN (LOSSLESS)

La compresión sin pérdida se usa en aplicaciones donde no es tolerable ninguna diferencia entre los datos originales y comprimidos: textos, programas, imágenes críticas (radiografías, teledetección).

Como ejemplos de los compresores sin pérdida, para archivos se tiene: RAR, Gzip, Bzip, zip, 7z; en imágenes: PNG, RLE; en formatos de audio: FLAC, WAV. En video es muy raro, suelen ser utilizados para captura.

Empleando compresión sin pérdidas los datos originales pueden ser recuperados en un 100% después de aplicar sobre los datos comprimidos un algoritmo compatible de descompresión. Si bien los algoritmos de compresión sin pérdida tienen la propiedad de disminuir el tamaño y conservar la totalidad de la información, la compresión sin pérdida logra un índice de compresión muy bajo (aproximadamente 2:1), comparada con la compresión con pérdida. Esta forma de compresión se caracteriza porque la tasa de compresión que proporciona está limitada por la entropía (redundancia de datos) de la señal original. Es decir, existe un límite teórico de compresión para los compresores sin pérdida, este límite fue enunciado por Shannon en 1948. En su primer teorema Shannon menciona que la longitud media de un código¹³ (L) siempre es mayor o igual a la entropía de la fuente $[H(s)]$, de lo que se concluye que el código más corto que puede existir para codificar la salida de una fuente de mensajes está limitado por la misma fuente.

2.2.2.1 Codificación estadística

La codificación estadística se basa en asignar códigos cortos a los datos que aparecen más frecuentemente, y asignar códigos más largos a los datos más infrecuentes. De este modo, el número promedio de bits por dato codificado es menor. Los compresores estadísticos utilizan las propiedades estadísticas de la fuente para mejorar la codificación.

Entre los compresores estadísticos se tienen los siguientes: Huffman, Shannon-Fano, Predictores (adaptivo) y Codificación Aritmética.

¹³ Longitud media de un código (L): constituye el número promedio de símbolos del alfabeto código (conjunto de símbolos de otro alfabeto diferente al de la fuente) que utiliza un código para codificar un símbolo de la fuente. La longitud media constituye el número de símbolos código que en promedio conforma una palabra código.

Como en la implementación del codificador mp3 en Matlab se emplea el algoritmo de codificación de Huffman, referido al estándar ISO/IEC 11172-3, éste será descrito de forma breve a continuación.

2.2.2.1.1 Codificación Huffman

Una técnica que consiste en asignarles códigos de bits más cortos a los datos que mayor frecuencia de aparición tienen y códigos más largos a los que aparecen con menos regularidad. Es una técnica que ofrece altos ratios de compresión.

Para realizar el proceso se construye un árbol binario de abajo hacia arriba, se comienza teniendo todos los símbolos del archivo junto con sus frecuencias. Se escogen los 2 datos con menor frecuencia y se los une. El padre de ambos datos será la suma de sus frecuencias. Luego se toma este árbol con 2 hijos y se le agrega la menor frecuencia siguiente. Se repite el proceso hasta que se llega al dato con mayor frecuencia.

Para obtener el código de Huffman (serie de bits que representan un dato) se recorre el árbol desde el nodo hijo hasta la raíz del árbol. Dependiendo de la rama que se tome se agrega un 0 o un 1 al código de Huffman. Si es rama derecha se asigna un 1 y para la rama izquierda un 0.

Una característica particular de estos códigos es que dada una serie de bits es posible determinar si se debe continuar leyendo el código o se debe detener. Por ejemplo: no habrá un código 001 y otro 00101, porque los 3 primeros bits se repiten. Se tiene uno u otro, de forma que no se repitan los primeros bits.

Ejemplo:

Símbolo	A	B	C	D	E
Frecuencia	15	7	6	6	5

Tabla 2. 1: Símbolos y sus respectivas frecuencias (Codificación Huffman)

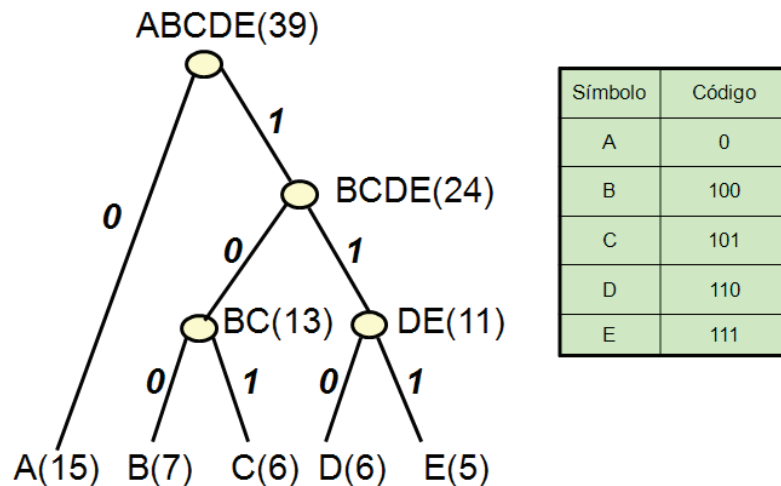


Figura 2. 3: Ejemplo de codificación Huffman

2.2.2.2 Basados en diccionario

Se construye un diccionario basado en la entrada procesada hasta ese momento, ese diccionario contiene cadenas de mensajes identificados por un índice. Estos compresores no requieren un conocimiento de la probabilidad de aparición de cada símbolo. Por ello serán útiles en aquellas aplicaciones donde no sea posible conocer las probabilidades de los símbolos. Se basan en que algunas cadenas de símbolos se repiten frecuentemente. Las cadenas procesadas hasta el momento pasan a ser parte de un diccionario indexado, el cual puede definirse de forma explícita o implícita. Durante el proceso de compresión, si la cadena que se está procesando, coincide con una entrada en el diccionario, la salida emitida por el compresor es el índice correspondiente a dicha cadena; índices que generalmente son códigos de menor longitud que la cadena a la que hacen referencia.

Existen gran variedad de algoritmos basados en el uso de diccionarios, la diferencia fundamental entre todos ellos es la manera en que cada uno crea y gestiona el diccionario, entre los compresores más conocidos se tiene: RLE (Run Length Encoding) y Lempel Ziv.

2.3 COMPRESIÓN DE AUDIO

2.3.1 INTRODUCCIÓN

Los requisitos de almacenamiento del audio son más pequeños que los de las imágenes o películas, pero más grandes que los de texto. Esta es la razón porque la compresión de audio ha llegado a ser importante y ha sido objeto de mucha investigación y experimentación en toda la década de 1990. Hay características importantes en la compresión de audio, una tiene que ver con la pérdida de información y otra con el requerimiento de una rápida decodificación. El texto debe ser de compresión sin pérdidas, pero las imágenes y el audio pueden perder mayor cantidad de datos sin una degradación notable de la calidad. En la mayoría de los casos en la compresión de audio se aprovechan de las propiedades que presentan las fuentes de sonido y las limitaciones físicas del oído humano. Las características de una señal de audio y los distintos tipos de calidad requerida influyen en la utilización de una u otra técnica de compresión de audio.

Las secuencias de audio forman parte de las aplicaciones multimedia. El estudio de la codificación y compresión se puede enfocar en función de la aplicación. En aplicaciones interactivas (audio o video conferencia) se emplean códecs¹⁴ simétricos, mientras que en aplicaciones de difusión y reproducción de medios (TV digital, audio Hi-Fi, DVD, etc.) se emplean códecs asimétricos

Cuando se habla del audio se puede hacer referencia a dos grandes categorías una tiene que ver con la música y la otra tiene que ver con la voz.

En el presente proyecto de titulación se hará mayor énfasis al aspecto de la compresión de música y no al de la compresión de voz ya que el objetivo general es implementar un codificador mp3 en Matlab para comprimir un archivo de música *.wav a *.mp3.

¹⁴ Como se mencionó anteriormente en este capítulo el compresor puede bien llamarse codificador y el expansor decodificador en cuáles casos el par puede denominarse códec.

2.3.1.1 Características de la compresión de audio

Las técnicas de compresión de audio, en general, se pueden clasificar en dos categorías básicas: sin pérdidas y con pérdidas. Aunque hay muchas diferentes técnicas de compresión, todas ellas caen en una u otra de estas categorías. A menudo, el audio se almacena en formato comprimido y debe ser descomprimido en tiempo real cuando el usuario quiere escucharlo; este es el por qué la mayoría de los métodos de compresión de audio son asimétricos. El codificador puede ser lento, pero el decodificador tiene que ser rápido.

En compresión del audio la técnica más simple para reducir el volumen de datos en una señal digital es la reducción de la frecuencia de muestreo o de la cantidad de bits de cuantización. Si bien es un método bastante rudimentario y poco eficiente, todavía puede ser utilizado en algunas aplicaciones. Solamente hay que tener en cuenta que reduciendo la frecuencia de muestreo se reduce el ancho de banda que es posible procesar, y reduciendo los bits de cuantización disminuye la relación señal a ruido de cuantización SNR_q (Ec. 1.12) introduciéndose ruido en la señal.

2.3.1.2 Códec de audio

Un códec de audio, es un códec que incluye un conjunto de algoritmos que permiten codificar y decodificar los datos auditivos, lo cual significa reducir la cantidad de bits que ocupa el fichero de audio. Un códec puede ser implementado en software, hardware o una combinación de ambos. Actualmente existen una gran variedad de códecs de audio. Aunque cada vez son más complejos y añaden características adicionales, dependiendo del método de codificación o reducción de la tasa de bits se pueden clasificar en:

2.3.1.2.1 *Codificadores perceptuales*

Los codificadores perceptuales aprovechan las limitaciones en la percepción del sistema auditivo humano (umbral de audición, enmascaramiento temporal y/o

frecuencial) para codificar el flujo de datos. Las distintas versiones de MPEG-1, siendo MP3 (MPEG-1 Layer 3) la más conocida, utilizan este método de codificación.

La codificación perceptual se tratará con detalle más adelante en este capítulo.

2.3.1.2.2 Codificadores paramétricos

Los codificadores paramétricos se basan en que el audio y la voz se pueden representar y sintetizar con tonos aislados, patrones armónicos (representados con sinusoides) y componentes ruidosas. Con parámetros como la amplitud, la frecuencia fundamental o los componentes espectrales se requieren pocos bits para representar el audio o la voz.

2.3.1.2.3 Vocoders

Los vocoders son codificadores paramétricos específicos para la codificación de la voz. Estos analizan la señal de voz correspondiente a un segmento temporal considerado estacionario para extraer los parámetros del modelo y la excitación.

2.3.1.2.4 Codificadores de forma de onda

Los codificadores de forma de onda se basan en el estudio de la señal, de forma que intentan reproducir la forma de la señal de entrada. Generalmente se diseñan para ser independientes de la señal, de manera que se utilizan para codificar una gran variedad de señales.

Estos codificadores aprovechan la redundancia de la señal, y a partir de una predicción lineal permiten codificar la señal auditiva; de esta forma se consiguen tasas de compresión elevadas cuando las señales son muy redundantes y prácticamente nulas cuando no es así.

2.3.1.2.5 Codificadores híbridos

Los codificadores híbridos, también conocidos como *codificadores de análisis por síntesis* combinan las técnicas de los codificadores de forma de onda con los vocoders, esencialmente son codificadores de voz. Entre los codificadores híbridos se encuentran: RELP, MPC, CELP, VSELP, RPE-LTP.

2.3.1.3 Parámetros de los códecs de audio

Los códecs de audio se caracterizan por los siguientes parámetros:

- Número de canales: Dependiendo el número de señales de audio simultáneas que contiene el flujo de datos (stream en inglés). Un códec puede ser monofónico (1 canal), estéreo (2 canales) o multicanal como 5.1 (seis canales) o 7.1 (ocho canales).
- Frecuencia de muestreo: Determina la calidad percibida, por lo tanto cuanto más alta sea la frecuencia de muestreo mayor será la fidelidad del audio obtenido respecto al original. La máxima frecuencia de muestreo utilizada en sistemas de audio actualmente es de 192 KHz.
- Número de bits por muestra: Determina la precisión con la que se reproduce la señal original y el rango dinámico de la misma. Se suelen utilizar 8 bits por muestra (para un rango dinámico de hasta 45 dB), 16 bits por muestra (para un rango dinámico de hasta 90 dB como el formato CD) o 24 bits por muestra (para un rango dinámico desde 109 hasta 120 dB). El más común es 16 bits.
- Tipo de compresión: Un códec de audio puede ser de compresión con pérdidas (lossy) o sin pérdidas (lossless).
- Tasa de bits: Determina el número de bits de información necesarios por unidad de tiempo. La tasa de bits no se puede deducir de los parámetros anteriores puesto que la compresión puede ser con pérdidas o sin pérdidas.

Además, la tasa de bits puede ser constante (CBR: Constant Bit Rate), variable (VBR: Variable Bit Rate) o media (ABR: Average Bit Rate). En audio se utiliza tasa de bits variable (VBR), puesto que es más eficiente que CBR cuando hay silencios o segmentos donde la complejidad es baja y se pueden almacenar con menos bits.

Códec	AAC	FLAC	MP3
Tipo de compresión	Con pérdidas, Híbrido	Sin pérdidas	Con pérdidas
Frecuencia de muestreo	8 KHz a 192 KHz	1 KHz a 1.04857 MHz	8, 11.025, 12, 16, 22.05, 24, 32, 44.1, 48 KHz
Tasa de bits	8 - 529 Kbps (estéreo)	Variable	8, 16, 24, 32, 40, 48, 56, 64, 80, 96, 112, 128, 160, 192, 224, 256, 320 Kbps
Bits por muestra	Cualquiera	4, 8, 16, 24, 32	Cualquiera
CBR	Si	No	Si
VBR	Si	Si	Si
Multicanal	hasta 28 canales	hasta 8 canales	No

Tabla 2. 2: Características técnicas de códecs de audio

- **Estandarización:** Hay dos organizaciones que dominan la estandarización de los codificadores de audio. La UIT-T (Unión Internacional de Telecomunicaciones) dentro del conjunto de normas H.320 y H.323 donde se definen los estándares para videoconferencia y telefonía IP G.7xx y la ISO/IEC (Organización Internacional de Estándares y Comisión Electrotécnica Internacional) principalmente con los estándares MPEG (Moving Picture Experts Group).

Sin pérdidas	Con pérdidas
ALAC (Apple Lossless)	AAC (Advanced Audio Coding)
DST (Direct Stream Transfer)	aacPlus
FLAC (Free Lossless Audio Codec)	AC3 (Dolby Digital A/52)
LA (Lossless Audio)	ADPCM
LPAC (Lossless Predictive Audio Codec)	ADX (Videojuegos)
LTAC (Lossless Transform Audio Codec)	ATRAC (Adaptive Transform Acoustic Coding)
MLP (Meridial Lossless Packing)	DRA
Monkey's Audio (APE)	DTS (Digital Theater Systems)
MPEG-4 ALS	MP1 (MPEG-1 audio layer-1)
MPEG-4 SLS	MP2 (MPEG-1 audio layer-2)
OptimFROG	MP3 (MPEG-1 audio layer-3)
QDesign	mp3PRO

RealAudio Lossless	Musepack
RKAU	Ogg Vorbis
Shorten (SHN)	Perceptual Audio Coding
TTA (True Audio)	RTA (Real Time Audio Códec)
WavPack	TwinVQ
WMA lossless (Windows Media Audio Lossless)	Siren
	WMA (Windows Media Audio)

Tabla 2. 3: Ejemplos de códecs de audio con y sin pérdida

2.3.2 COMPRESIÓN DE MÚSICA

Para la compresión de música se desea normalmente que la fuente este muestreada preferiblemente a 44,1 KHz y 16 bits/muestra. El objetivo es reducir al mínimo el tamaño, pero preservando una calidad casi idéntica a la de la fuente original. En la actualidad se puede conseguir una calidad muy buena entre los 64 y 128 Kbps. A estos códecs se los conoce como “códecs perceptuales” ya que se basan en las limitaciones del oído humano para realizar su trabajo.

Para lograr la compresión se remueve toda la información presente en la señal que no es detectada por el oído humano. Para saber qué se puede eliminar se recurre al conocimiento de la psicoacústica, esto es de la percepción humana del sonido.

2.3.2.1 Mínimo umbral auditivo

Este umbral, también conocido como umbral absoluto, corresponde al sonido de intensidad más débil que se puede escuchar en un ambiente silencioso. El mínimo umbral auditivo no tiene un comportamiento lineal; se representa por una curva de Intensidad (dB) contra Frecuencia (Hz), que posee niveles mínimos entre 2 y 5 KHz, los cuales corresponden a la parte más sensitiva del oído humano.

Por lo tanto, en los sistemas de compresión de audio que sacan provecho de la psicoacústica, no es necesario codificar los sonidos situados bajo este umbral (el área por debajo de la curva en la figura 2.4), ya que éstos no serán percibidos.

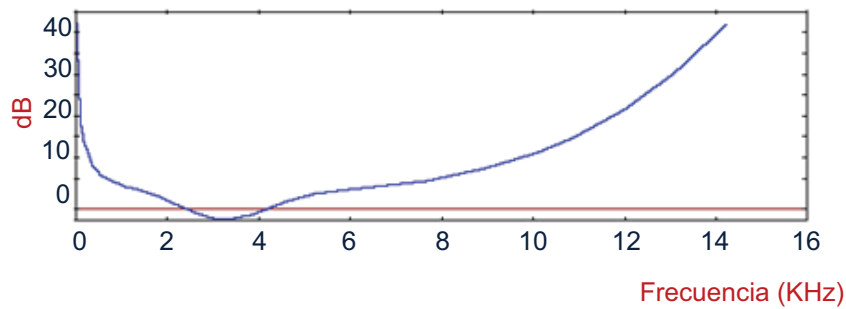


Figura 2. 4: Mínimo umbral auditivo (ambiente silencioso)

2.4 ALGORITMOS DE COMPRESIÓN DE AUDIO

2.4.1 INTRODUCCIÓN

La compresión es un proceso que intenta alcanzar una representación más compacta de la señal digital, mediante la eliminación de la redundancia existente en dicha señal. El objetivo es minimizar la cantidad de bits necesarios para su transmisión o almacenamiento, preservando la calidad de la señal.

Las señales de audio son interesantes desde el punto de vista de la compresión, debido a que contienen gran cantidad de información redundante que en gran parte puede eliminarse antes de su transmisión o almacenamiento.

2.4.2 CODIFICACIÓN PERCEPTUAL DE AUDIO

La codificación perceptual de audio consiste, de modo general, en un método para reducir la cantidad requerida de datos para representar una señal de audio digital [13]. Este método inevitablemente genera pérdidas en términos de calidad, introduciendo una cierta cantidad de ruido que podría perfectamente caer dentro del rango de la audición humana si se analiza en forma aislada. Sin embargo, la codificación perceptual está diseñada de tal manera que el ruido generado por el codificador cae fuera de los límites de audición humana en presencia de la señal original. Esta distinción es muy importante, ya que los algoritmos de compresión basados en percepción, a diferencia de los esquemas puramente numéricos o

algorítmicos, tales como Ley μ o ADPCM, se aprovechan de las limitaciones del sistema auditivo humano.

La idea fundamental de la codificación perceptual de audio, es que la presencia de ciertos estímulos auditivos puede influenciar la habilidad del cerebro humano para percibir otros estímulos. En palabras más simples, este tipo de algoritmos se basa fuertemente en el fenómeno de enmascaramiento. Un codificador perceptual por lo tanto, no codifica aquellas componentes de la señal de audio que se verán enmascaradas por otras, ahorrando de esta manera una considerable cantidad de datos perceptualmente redundantes e innecesarios.

Hoy en día existen numerosos esquemas de compresión basados en esta premisa, siendo el más conocido el MPEG-1 Capa 3, comúnmente conocido como MP3.

El esquema de un codificador perceptual de audio se muestra a continuación en la figura 2.5.

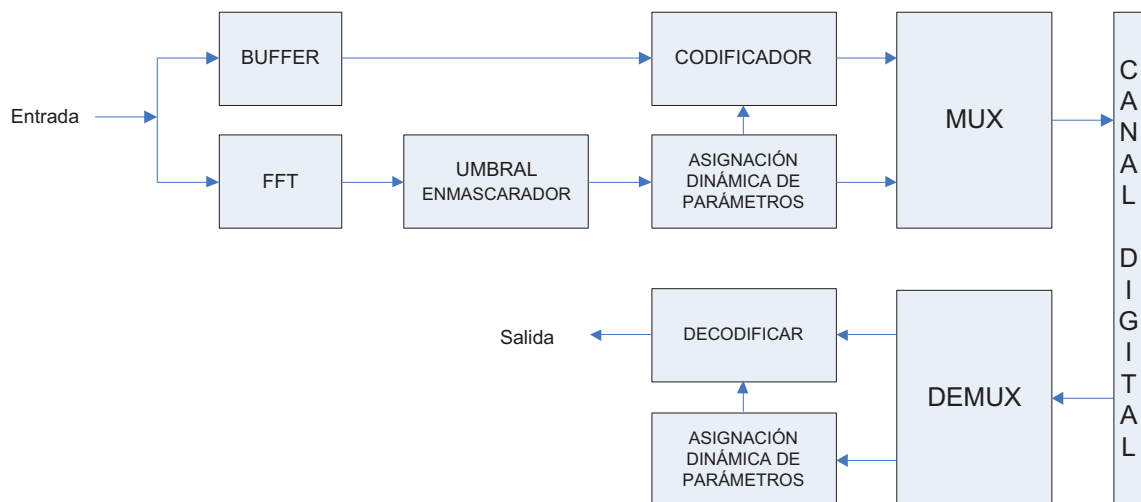


Figura 2. 5: Esquema de un codificador perceptual de audio

Como se puede apreciar en el esquema, en el codificador, la señal de entrada se descompone en múltiples bandas de frecuencia. De esta forma, los datos de cada banda se pueden procesar de manera independiente y cada banda puede ser representada con un grado variable de resolución. La idea es asignar una menor

resolución en aquellas bandas de frecuencia que pueden ser representadas con una menor cantidad de información, debido principalmente al enmascaramiento.

Cuando la resolución se reduce en alguna banda en particular, crece el ruido de cuantización, en esa zona de frecuencias. La idea es cambiar el nivel de cuantización de esa banda de manera de satisfacer la tasa de bits objetivo, manteniendo la mayor cantidad de detalles posible. El codificador está constantemente analizando la señal de entrada y toma decisiones acerca de qué zonas del espectro se ven enmascaradas y por lo tanto, al ser inaudibles, pueden descartarse de la señal y así disminuir la resolución. Para decodificar se aplica una transformada inversa de manera de combinar las bandas y restaurar la señal original. En el caso en que la resolución de una banda no se vea reducida, el proceso es ideal y sin pérdida.

La efectividad de un codificador perceptual depende de que tan bien puede modelar las limitaciones perceptuales del sistema auditivo humano, pero también depende de si dispone del ancho de banda necesario para contener todo el detalle sonoro que los seres humanos somos capaces de percibir.

2.4.3 CODIFICACIÓN DE SUB-BANDAS

La codificación de sub-bandas o SBC (sub-band coding) es un método potente y flexible para codificar señales de audio eficientemente. A diferencia de los métodos específicos para ciertas fuentes, el SBC puede codificar cualquier señal de audio sin importar su origen, ya sea voz, música o sonido de tipos variados. El principio básico del SBC es la limitación del ancho de banda por descarte de información en frecuencias enmascaradas. El resultado simplemente no es el mismo que el original, pero si el proceso se realiza correctamente, el oído humano no percibe la diferencia.

En la compresión de audio el primer proceso al que se somete la señal digital no comprimida de entrada, es a una codificación en sub-bandas, la cual consiste en filtrar la señal mediante un banco de filtros que descomponen la banda total en

una serie de sub-bandas o ventanas y, a continuación cada sub-banda se codifica adaptativamente [14], aplicando un modelo psicoacústico que analiza tanto las bandas, como la señal y determina los niveles de enmascaramiento utilizando los datos psicoacústicos que dispone. Considerando estos niveles de enmascaramiento se cuantizan y codifican las muestras de cada banda, si en una frecuencia dentro de la banda hay una componente por debajo de dicho nivel, se desecha. Si lo supera, se calculan los bits necesarios para cuantizarla y se codifica. Por último se agrupan los datos según el estándar correspondiente que estén utilizando codificador y decodificador, de manera que éste pueda descifrar los bits que le llegan de aquél y recomponer la señal. Si la señal de sonido se hace pasar por un banco de filtros, cuyos anchos de banda son contiguos en todo el espectro de la señal y no se traslapan, las sub-bandas resultantes pueden recombinarse para recuperar la señal original sin distorsión perceptible.

En el caso de la codificación de audio MPEG se emplean 32 filtros, cada uno con 512 derivaciones. El número de bits utilizado en el proceso de codificación, en general es diferente para la señal de cada sub-banda, en que la cuantización se realiza con un criterio perceptual. Al codificar individualmente la señal de cada sub-banda, el ruido de cuantización queda confinado sólo a esa sub-banda. Los flujos binarios de salida de cada codificador se multiplexan para su transmisión o procesado posterior. En el receptor se realiza primero un demultiplexado seguido por la decodificación de las señales de cada sub-banda para recuperar la señal original.

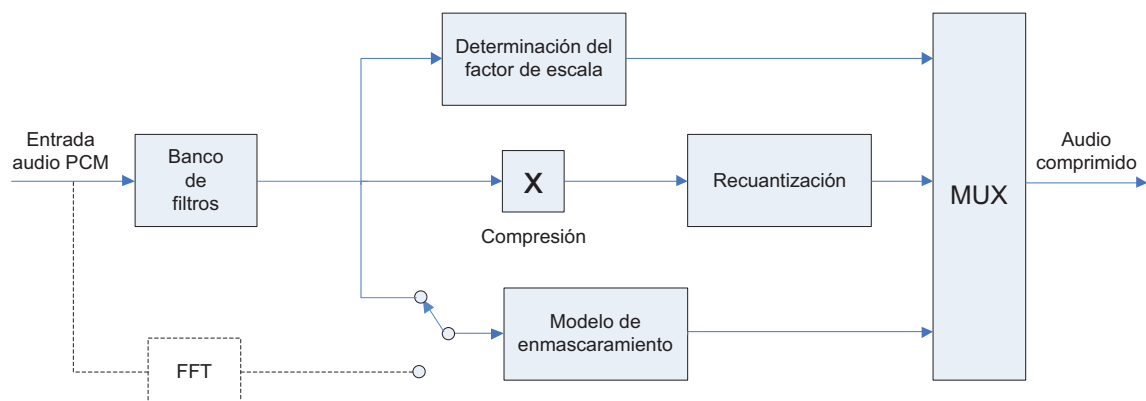


Figura 2. 6: Diagrama de bloques de un codificador de sub-bandas [15]

2.4.4 COMPRESIÓN DE AUDIO ADPCM

ADPCM (*Adaptive Differential Pulse Code Modulation*), reduce la muestra y cuantiza adaptativamente¹⁵. El muestreo se realiza a 8 KHz, con muestras de 8 bits a 64 Kbps. Utiliza diferencias de 4 bits, con una tasa de bits final de 32 Kbps.

Los códecs ADPCM son codificadores de forma de onda, los cuales en vez de cuantificar la señal directamente, como los codificadores PCM, cuantifican la diferencia entre la señal y una predicción hecha a partir de la señal, por lo que se trata de una codificación diferencial. Para este caso se usa la codificación diferencial o predictiva.

En la figura 2.7 se ilustra el diagrama de bloques de un codificador ADPCM.



Figura 2. 7: Diagrama de bloques de un codificador ADPCM

Antes de la digitalización se coge la señal analógica y se divide en bandas de frecuencia gracias a los filtros QMF¹⁶ (Quadrature Mirror Filter), obteniendo sub-bandas de señal. Cada sub-banda es tratada de modo distinto utilizando las propiedades de DPCM, es decir, se lleva a cabo el proceso de muestreo, cuantización del error de predicción y finalmente se codifica.

¹⁵ Cuantización adaptativa: Usa pasos más largos para codificar diferencias entre muestras muy distintas en magnitud (de alta frecuencia) y pasos más pequeños para muestras que son similares (bajas frecuencias).

¹⁶ En procesamiento digital de señales, un filtro espejo en cuadratura (Quadrature Mirror Filter, QMF) es un filtro que divide la señal de entrada en dos bandas que posteriormente suelen ser submuestreadas por un factor 2. Además, ambas bandas (superior e inferior) de frecuencias se intercambian entre sí. Es decir, las frecuencias bajas se codifican como frecuencias altas y viceversa. Las frecuencias en la banda de transición se codifican en las bandas inferior y superior con diferentes amplitudes.

Una vez que se obtiene la sucesión de bits (bitstream) de cada sub-banda, se multiplexan los resultados y ya se puede proceder a almacenar los datos o bien transmitirlos. El decodificador tiene que realizar el proceso inverso, es decir, demultiplexar y decodificar cada sub-banda del bitstream.

Algunas de las técnicas ADPCM se utilizan en las comunicaciones de voz sobre IP. ADPCM también fue utilizado por IMA (Interactive Multimedia Association)¹⁷ para el desarrollo del códec de audio ADPCM conocido como DVI, IMA ADPCM o DVI4, en la década de 1990.

ADPCM se articula en los estándares CCITT G.721, CCITT G.723 y en el CCITT G.726, que reemplazó a los dos anteriores definiendo estándares para 16, 24, 32 y 40 Kbps (que corresponden a tamaños de muestra de 2, 3, 4 y 5 bits respectivamente).

2.4.5 FORMATO DE AUDIO WAV

La palabra WAV es una abreviatura inglesa de wave (onda) y se refiere a la forma que tiene la representación gráfica típica del sonido. Lo flexible de este formato lo hace muy usado para el tratamiento del sonido pues puede ser comprimido y grabado en distintas calidades y tamaños, desde 11.025 Hz, 22.050 Hz a 44.100 Hz.

Aunque los archivos .wav pueden tener un excelente sonido comparable al del CD (16 bits y 44,1 KHz, estéreo), el tamaño necesario para esa calidad es demasiado grande (especialmente para los usuarios de Internet) una canción convertida a .wav puede ocupar fácilmente entre 20 y 30 MB. La opción más pequeña es grabar a 4 bits/muestra y a una frecuencia de muestreo de 11.025 Hz, lo más bajo

¹⁷ La Asociación de Multimedia Interactiva (IMA) fue una asociación de la industria que desarrolló un conjunto de algoritmos de audio. El más importante es el algoritmo ADPCM que está en uso por Apple y Microsoft. La Asociación de Multimedia Interactiva dejó de funcionar alrededor de 1998.

posible, el problema es la baja calidad del sonido, los ruidos, la estática e incluso cortes en el sonido; por esta razón casi siempre se usa para muestras de sonido. La ventaja más grande es la de su compatibilidad para convertirse en varios formatos por medio del software adecuado, un ejemplo de ello es pasar de .wav a .mp3.

2.4.5.1 Características de un archivo WAV

La descripción de las características de un archivo WAV es relativa. Las cualidades superiores se utilizan para archivos de música, mientras tanto, las intermedias para voz y efectos sonoros. Como referencia, la música de los discos compactos (CD's) está grabada a 16 bits por muestra y una frecuencia de muestreo de 44,1 KHz. La música suena muy mal con sólo 8 bits de resolución, y se pierden matices cuando se graban voces.

La elección entre archivos monofónicos o estéreo (la relación de tamaños es de 2:1) depende no sólo del uso del archivo, sino de la captura de los sonidos: si se graba con un micrófono monoaural, no tiene mucho sentido utilizar un archivo estéreo. Tampoco para efectos de sonido, salvo que se utilice los canales para crear efectos de desplazamiento o distancia entre las fuentes sonoras. Los archivos .wav admiten tres frecuencias de muestreo (11.025 Hz, 22.050 Hz y 44.100 Hz); puede asignar 8 o 16 bits de resolución (bits por muestra), y pueden usar uno o dos canales: mono o estéreo.

Bits por muestra	Frecuencia de muestreo	Canales	Tamaño/minuto
16	44.100 Hz	Estéreo	10 MB
16	44.100 Hz	Mono	5 MB
16	22.050 Hz	Estéreo	5 MB
8	44.100 Hz	Estéreo	5 MB
8	44.100 Hz	Mono	2,5 MB
8	22.500 Hz	Estéreo	2,5 MB
8	22.500 Hz	Mono	1,25 MB
8	11.000 Hz	Mono	0,63 MB

Tabla 2. 4: Tamaños del archivo WAV

2.4.6 FLAC

FLAC significa códec de audio libre sin pérdida (Free Lossless Audio Códec). FLAC es libre lo que significa que, a diferencia de MP3, AAC, Ogg¹⁸ u otros, no existe pérdida de información desde la fuente de audio, convirtiéndose en un formato ideal para archivos de audio de muy alta calidad, ya que permite reconstruir el audio original en su totalidad. FLAC admite cualquier resolución PCM entre 4 y 32 bits por muestra, y cualquier frecuencia de muestreo (sample rate) desde 1 a 65.535 Hz, en incrementos de 1 Hz.

FLAC está diseñado para comprimir audio. Debido a ello, los archivos resultantes son reproducibles y útiles, además de ser más pequeños que si se hubiera aplicado directamente al archivo PCM un algoritmo de compresión estadística (como ZIP). Los algoritmos con pérdida pueden comprimir a más de 1/10 del tamaño inicial, descartando información; FLAC, en su lugar, usa la predicción lineal para convertir las muestras en series de pequeños números no correlativos (conocido como "*residuos*"), que se almacenan eficientemente usando la codificación Golomb-Rice¹⁹. Además de esto, para aprovechar los silencios (donde los valores numéricos presentan mucha repetición) usa codificación RLE (Run-Length Encoding) para muestras idénticas.

FLAC se ha convertido en uno de los formatos preferidos para la venta de música por Internet, al igual que Monkey's Audio²⁰ que funciona de manera similar.

¹⁸ Ogg es un formato de archivo contenedor multimedia, desarrollado por la Fundación Xiph.org y es el formato nativo para los códecs multimedia que también desarrolla Xiph.org. El formato es libre de patentes y abierto al igual que toda la tecnología de Xiph.org, diseñado para dar un alto grado de eficiencia en el "streaming" y la compresión de archivos.

¹⁹ La codificación Golomb es un tipo de codificación de entropía inventada por Solomon W. Golomb que es óptima para alfabetos que siguen una distribución geométrica, lo que significa que los valores bajos son mucho más comunes que los altos.

²⁰ Monkey's Audio es un formato de fichero para comprimir información de audio. Siendo un formato de compresión sin pérdida, Monkey's Audio no elimina información del flujo de audio.

Además es usado en el intercambio de canciones por la red, como alternativa al MP3, cuando se desea obtener una mayor reducción del tamaño que en un archivo WAV-PCM, y no perder calidad de sonido. También es el formato ideal para realizar copias de seguridad de CD's, ya que permite reproducir exactamente la información del original, y recuperarla en caso de problemas con este material.

Los archivos FLAC tienen una extensión .flac; son perfectamente reproducibles con algunos reproductores, incluso en computadoras antiguas, ya que una de las características del formato, es que los archivos decodifiquen en modo sencillo. Estos archivos son de velocidad de bits variable, ya que no todas las partes de una misma canción son igualmente comprimibles. FLAC se populariza gracias a la banda ancha y a los discos duros de mayor capacidad. Por supuesto, el principal inconveniente del uso del FLAC radica en el espacio que ocupa un archivo en este formato. A pesar de que sus creadores defiendan que en tareas de compresión y descompresión de los archivos es el más rápido de su categoría, y a que, según sus datos, consiga una reducción de alrededor del 50% del espacio de la canción original, el resultado es que 'pesa' bastante más que los MP3.

Así, sí se toma como referencia una canción comprimida en MP3 de 5 minutos, su tamaño variará entre los 4,6 Megabytes y los 11,5 Megabytes, en función de la tasa de kilobits por segundo (lo que se conoce como "bitrate") con la que se codifique (entre 128 kilobits por segundo, el mínimo para conseguir un sonido similar al CD, y 320 Kbps, el máximo de calidad de este formato). Sin embargo, el mismo archivo comprimido en FLAC supera los 35 Megabytes.

2.4.7 CODIFICACIÓN DE AUDIO EN EL ESTÁNDAR MPEG

El estándar de compresión de audio MPEG define tres capas de calidad subjetiva y complejidad crecientes. Soporta frecuencias de muestreo a 32, 44,1 y 48 KHz. A 16 bits por muestra, el audio no comprimido produce un caudal binario de alrededor de 1,5 Mbps.

Nombre	Estándar	Características
MPEG-1	ISO/IEC 11172-3	Proporciona codificación de un canal (mono) o dos canales (estéreo o mono dual) con frecuencias de muestreo de 32, 44,1 y 48 KHz. Las tasas de bits (<i>bit rates</i>) predefinidas son: <ul style="list-style-type: none"> • Layer I: De 32 a 448 Kbps • Layer II: De 32 a 384 Kbps • Layer III: De 32 a 320 Kbps
MPEG-2 AAC	ISO/IEC 13818-7	Es un estándar de codificación audio de muy alta calidad de hasta 48 canales con frecuencias de muestreo entre 8 y 96 KHz con capacidades multicanal, multilinguaje y multiprograma. Trabaja a <i>bitrates</i> desde 8 Kbps para señal monofónica de voz hasta más de 160 Kbps/chanal para codificación de muy alta calidad.
MPEG-4	ISO/IEC 14496-3	<ul style="list-style-type: none"> • Codificación y composición de objetos audio tanto naturales como sintetizados. • Escalabilidad en el <i>bitrate</i>. • Escalabilidad en la complejidad de los codificadores y decodificadores. • Audio Estructurado: Lenguaje universal para la síntesis de sonido. • TTSI: Un interfaz para la conversión de texto a voz.

Tabla 2. 5: Características de los estándares MPEG

Después de la compresión, las tasas de bits para canales monofónicos se sitúan entre 32 y 192 Kbps y los estereofónicos, entre 128 y 324 Kbps. Al igual que otros sistemas de compresión de audio, MPEG aprovecha las características psicoacústicas del sistema auditivo humano, principalmente el enmascaramiento simultáneo y el temporal. El enmascaramiento simultáneo se basa en el hecho de que una señal de audio de cierta amplitud y frecuencia, puede enmascarar a otras señales de frecuencias cercanas y menor amplitud. El enmascaramiento temporal, por otra parte, se basa en el hecho de que el oído no percibe señales inmediatamente anteriores o posteriores a una señal enmascaradora.

La primera acción en el proceso de compresión MPEG es la de codificación en sub-bandas, segmentando el audio digital en ventanas de 384 muestras. Las capas I y II de MPEG utilizan un banco de filtros para descomponer cada ventana en 32 sub-bandas, cada una con un ancho de banda de aproximadamente 750 Hz, para una frecuencia de muestreo de 48 KHz. Cada sub-banda es diezmada

de modo que la tasa de muestreo por sub-banda es de 1,5 KHz con 12 muestras por ventana. Por otra parte, a la entrada de audio digital se le aplica una transformada rápida de Fourier (FFT) con el fin de determinar su espectro y calcular un umbral global de enmascaramiento para cada sub-banda; con ello, se elige un cuantizador uniforme que produce la mínima distorsión a la tasa binaria requerida.

Para cada sub-banda se amplifica el nivel de la señal para obtener niveles normalizados máximos. La ganancia necesaria es constante para cada bloque y se transmite un factor de escala con cada bloque, en cada sub-banda, a fin de invertir el proceso en el decodificador.

Uno de los problemas introducidos por la cuantificación son los *pre-ecos*, que pueden ocurrir cuando un sonido percusivo agudo va precedido de silencio. Al reconstruir la señal, los errores debidos a la cuantización tienden a distribuirse sobre el bloque de muestras causando una distorsión audible antes de la señal real. En una ventana de 8 ms el enmascaramiento temporal no puede suprimir los pre-ecos por completo.

En la capa III de MPEG el control de pre-ecos es una parte importante y, para ello, en esa capa se agrega una descomposición de las sub-bandas mediante la transformada discreta del coseno modificada (MDCT), con el fin de conseguir una división de frecuencia mucho más fina. En la capa III se agrega cuantización no uniforme, en el sentido de que las señales de mayor nivel pueden enmascarar errores de cuantización mayores. Además se incluye codificación por entropía y conmutación dinámica de ventanas, con lo que se obtiene mejor resolución temporal y mejor control de los pre-ecos. La máxima tasa binaria que soporta es de 384 Kbps.

Las tres capas soportan tasas binarias tan bajas como 32 Kbps, de ellas, la capa I es la más simple de las tres y proporciona audio de buena calidad a 192 Kbps. La máxima tasa binaria que soporta es de 448 Kbps. La capa II utiliza una FFT de mayor resolución que la capa I, cuantificación más fina y una forma más eficiente

de transmitir los factores de escala para las sub-bandas. El modelo psicoacústico empleado en esta capa es más complejo que el utilizado en la capa I. Proporciona audio con calidad de CD a 128 Kbps por canal y la máxima tasa binaria que soporta es de 384 Kbps.

2.4.7.1 Estándar MPEG-1

El estándar internacional ISO/IEC 11172, más conocido como MPEG-1 (Codificación de imágenes en movimiento y el audio asociado para medios de almacenamiento digital a una tasa cercana a 1,5 Mbps) está dividido en las siguientes partes:

Parte	Descripción
Parte 1	Sistema (Multiplexación y control para sincronización del video, el audio y la información secundaria)
Parte 2	Codificación del video
Parte 3	Codificación del audio
Parte 4	Pruebas del sistema
Parte 5	Simulación por software

Tabla 2. 6: Partes del estándar MPEG-1

- **Parte 1, Sistema**: se refiere al problema de combinar uno o más flujos de datos provenientes de la parte de audio o video, incluyendo información de temporización y sincronización para lograr un flujo único de datos, apropiado para ser almacenado o transmitido. Los archivos audio de la capa I utilizan típicamente la extensión *.mp1* o a veces *.m1a*.
- **Parte 2, Video**: especifica una representación codificada que puede ser usada para comprimir secuencias de video (tanto de 625 como de 525 líneas de frecuencia). Los archivos de audio de la capa II utilizan típicamente la extensión *.mp2* o a veces *.m2a*.
- **Parte 3, Audio**: describe la representación codificada que puede usarse para comprimir secuencias de audio (en modo monofónico, o en estéreo). Los archivos de audio de la capa III utilizan la extensión *.mp3*.

- **Parte 4, Pruebas:** trata del diseño y de las pruebas que verifican si el flujo de bits y los decodificadores cumplen con los requerimientos y especificaciones dadas en las partes 1, 2 y 3.
- **Parte 5, Simulación:** no se trata de un estándar, sino de un reporte técnico; proporciona implementación completa por software de las tres primeras partes del estándar MPEG-1.

2.4.7.2 Estándar MPEG-2

Numerado de manera formal como ISO/IEC MPEG 13818 (Codificación genérica para información de imágenes en movimiento y el audio asociado), fue dado a conocer en 1994, tan sólo un año después del primer estándar, y se encuentra constituido por más partes que el estándar anterior.

Parte	Descripción
Parte 1	Sistema (Multiplexación y control para sincronización del audio y video)
Parte 2	Codificación del video
Parte 3	Codificación del audio (Compatible con el audio MPEG-1)
Parte 4	Pruebas del sistema
Parte 5	Reportes técnicos
Parte 6	DSM-CC: <i>Digital Storage Media-Command and Control</i> , Medios de Almacenamiento Digital-Comando y Control
Parte 7	AAC: <i>Advanced Audio Coding</i> , Codificación Avanzada de Audio (No compatible con el audio MPEG-1)
Parte 8	Fue abandonada cuando se comprobó que no había interés de la industria. Intentó codificar el video cuando las muestras de entrada son 10 bits
Parte 9	RTI: <i>Real Time Interface</i> , Interface en Tiempo Real
Parte 10	Pruebas del DSM-CC

Tabla 2. 7: Partes del estándar MPEG-2

El estándar en su primera fase fue una simple extensión en las capacidades de compresión y codificación del primer estándar, mientras que en la segunda fase se desarrollaron nuevos algoritmos que implicaban diferentes métodos para realizar la compresión tanto de video como de audio, pero en este punto se sacrificó la compatibilidad con el estándar anterior en algunas áreas.

- **Parte 1, Sistema:** es similar al primer estándar, realiza la misma función, pero aquí la realiza de dos maneras diferentes, el flujo de programa y el flujo de transporte, cada uno optimizado para diferentes conjuntos de aplicaciones. El flujo de programa se pensó para ambientes libres de errores y es apropiado para aplicaciones que involucren procesamiento por software, mientras que el flujo de transporte se usa para almacenamiento o transmisión en medios que presenten ruido y pérdidas.
- **Parte 2, Video:** mejora las capacidades para compresión de video del estándar anterior MPEG-1.
- **Parte 3, Audio:** además de agregar capacidad multicanal, también proporciona capacidad para múltiples idiomas, con respecto a la parte 3 del estándar MPEG-1.
- **Parte 4, Pruebas y parte 5, Reporte:** corresponden con las partes 4 y 5 del estándar MPEG-1.
- **Parte 6, Medios de Almacenamiento Digital, Comando y Control:** especifica un conjunto de protocolos que proporcionan las funciones y operaciones de control para manejar los flujos de bits MPEG-1 y MPEG-2.
- **Parte 7, AAC:** se trata de un algoritmo de codificación de audio multicanal, incompatible con la parte 3 de MPEG-1 y MPEG-2.
- **Parte 9, RTI:** da especificaciones acerca de la interface en tiempo real para los decodificadores del flujo de transporte (correspondiente a la parte 1) y se puede adaptar para su uso en todas las redes que transporten flujos de este tipo.
- **Parte 10, Pruebas del DSM-CC:** considera las pruebas que deben ser realizadas para verificar el correcto funcionamiento de la parte 6 (DSM-CC, Medios de Almacenamiento Digital - Comando y Control).

2.4.7.3 Estándar MPEG-4

Este estándar consta de dos versiones. El estándar *MPEG-4 Versión 1* se terminó en octubre de 1998 y se publicó en 1999. La *segunda versión* se terminó en diciembre de 1999 y se publicó en el 2000.

MPEG-4 es una fase que engloba todos los esquemas vistos para MPEG-1 y MPEG-2, define estructuras de más alto nivel por las cuales se permite el control y la combinación de elementos de audio procedentes tanto de fuentes sonoras digitalizadas (p.ej. por medio de los formatos MPEG-1 y MPEG-2, codificación CELP, etc.) como sintetizadas (voz o sonidos musicales, tanto simples como complejos). Es por lo tanto un estándar que combina diferentes fuentes de audio en un mismo *bitstream*²¹ con especial énfasis en aplicaciones multicanal, soporte multilinguaje, tasa binaria variable y recuperación de errores.

La mayor novedad a los formatos de audio que incorpora MPEG-4 está asociada a codificaciones de voz con tasas binarias extra bajas. El estándar especifica el uso de codificadores paramétricos de tasa binaria extra baja para la codificación de voz, que consiguen tasas binarias comprendidas entre 2 y 8 Kbps. Los codificadores paramétricos establecen un modelo de señal que ajustan al sonido a codificar. La información transmitida son los parámetros del modelo que mejor lo ajustan a la señal.

Teóricamente, el MPEG-4 permite desde un ancho de banda muy bajo (telefonía móvil) hasta la televisión en alta definición (HDTV). Por supuesto, los dispositivos actuales no soportan la reproducción de todo el rango de especificaciones pero, con el tiempo, se presentarán nuevos equipos en el mercado.

Las extensiones utilizadas en archivos que contienen datos en el formato .mp4 son: .mp4, extensión oficial para audio, video y contenidos avanzados; .m4a, sólo para archivos de audio; .m4p, FairPlay archivos protegidos.

²¹ El término bitstream hace referencia a una sucesión de bits.

2.4.7.4 AAC

AAC representa un formato de codificación de audio avanzado (Advanced Audio Coding), basado en el algoritmo de compresión con pérdida, un proceso por el que se eliminan algunos de los datos de audio para poder obtener el mayor grado de compresión posible, resultando en un archivo de salida que suena lo más parecido posible al original.

El formato AAC corresponde al estándar internacional “ISO/IEC 13818-7” como una extensión de MPEG-2: un estándar creado por MPEG. Debido a su excepcional rendimiento y calidad, la codificación de audio avanzada se encuentra en el núcleo del MPEG-4, 3GPP y 3GPP2²², y es el códec de audio de elección para Internet, conexiones inalámbricas y de radio difusión digital. AAC no es compatible con MPEG-1.

El AAC utiliza una tasa de bits variable (VBR), un método de codificación que adapta el número de bits utilizados por segundo para codificar datos de audio, en función de la complejidad de la transmisión de audio en un momento determinado. La frecuencia de muestreo de AAC es de 8 Hz a 96 KHz, el número de canales es de 1 a 48. AAC es un algoritmo de codificación de banda ancha de audio que tiene un rendimiento superior al del MP3, que produce una mejor calidad en archivos pequeños y requiere menos recursos del sistema para codificar y decodificar. Este códec está orientado a usos de banda ancha y se basa en la eliminación de redundancias de la señal acústica, así como en la compresión mediante la transformada discreta del coseno modificada (MDCT)²³, muy parecido al MP3. Los archivos de audio con formato AAC utilizan las extensiones .m4a, .m4b, .m4p, .m4v, .m4r, .3gp, .mp4 y .aac.

²² .3gp, .3g2 son extensiones utilizadas en archivos que contienen datos en el formato *.mp4 utilizados por la telefonía móvil 3G.

²³ La Transformada Discreta del Coseno Modificada (MDCT, por sus siglas en inglés, Modified Discrete Cosine), es una transformada lineal ortogonal solapada en la cual se usan ventanas con un solapamiento del 50%; basada en la idea de la cancelación del aliasing del dominio de tiempo (TDAC, Time Domain Aliasing Cancellation). La MDCT se utiliza típicamente en codificadores con una ventana longitud de 512 muestras, y 256 muestras nuevas para cada bloque.

2.4.8 ESTUDIO COMPARATIVO DE LAS CARACTERÍSTICAS DE LOS ALGORITMOS DE COMPRESIÓN DE AUDIO

A continuación se muestra una tabla comparativa entre los algoritmos de compresión de audio desarrollados en este capítulo, en cuanto a sus características técnicas se refiere.

Formato de compresión de audio	Algoritmo de compresión	Frecuencia de muestreo	Tasa de bits	Bits por muestra
WAVE Audio Format	LPCM (Modulación por impulsos lineales)	11,025; 22,050 y 44,1 KHz	Variable 1,411 Kbps	8 y 16 bits
ADPCM Audio Compression	Adaptativo (Modulación por codificación de impulsos diferenciales adaptativos)	8; 11,025; 22,050 y 44,1 KHz	16; 24; 32 y 40 Kbps	2, 3, 4 y 5 bits
FLAC	Lossless	1 KHz a 1,04857 MHz	Variable	4, 8, 16, 24 y 32 bits
MPEG-4 Audio Lossless Coding (ALS)	Lossless	192 KHz a 384 KHz	Variable	Cualquiera
Advanced Audio Coding. (AAC)	MDCT, Filtro Híbrido Subbanda	8 KHz a 192 KHz	8 a 529 Kbps (estéreo)	Cualquiera
MPEG-1 Layer 3 (MP3)	MDCT, Filtro Híbrido Subbanda	8; 11,025; 12; 16; 22,05; 24; 32; 44,1 y 48 KHz	8; 16; 24; 32; 40; 48; 56; 64; 80; 96; 112; 128; 160; 192; 224; 256 y 320 Kbps	Cualquiera

Tabla 2. 8: Cuadro comparativo de las características de los algoritmos de compresión de audio

La calidad de un archivo MP3 viene dada por el tiempo de codificación y el tamaño del archivo codificado que se quiera obtener, para obtener un archivo de audio con calidad digital la tasa de transferencia de bits o bitrate recomendada es de 128 Kbps, con una frecuencia de muestreo de 44.100 Hz.

El formato MP3 es el más utilizado y el más popular por tres razones: capacidad de compresión, calidad de compresión y facilidad de distribución.

La capacidad de compresión tiene relación con el hecho de que el formato MP3 logra reducir hasta 12 veces el tamaño de un archivo de audio original (sin compresión).

El algoritmo de compresión de audio MP3 logra un equilibrio en cuanto a la calidad de compresión, ya que, logra reducir el tamaño de un archivo de audio original con una mínima pérdida de calidad, por el mismo hecho de ser un algoritmo de compresión con pérdidas (lossy), el cual elimina información para lograr la mejor relación de compresión.

Con relación a su facilidad de distribución el formato MP3 tiene amplia ventaja con respecto a otros formatos de compresión de audio, alcanzando una gran popularidad en la Internet, tanto que en la actualidad se asocia al MP3 como sinónimo de música en Internet inclusive en conexiones de bajo ancho de banda, otro factor es la aparición de la radio digital y el servicio de audio streaming²⁴.

²⁴ Servicio de audio streaming: es una tecnología que permite transmitir cualquier fuente de audio a los oyentes a través de Internet de una forma simple y transparente. Una computadora se encarga de transmitir el audio a los servidores de super-hosting (servicio que provee a los usuarios de Internet un sistema para poder almacenar información) y estos a su vez, distribuyen el audio a los oyentes conectados en ese momento. Todo este proceso se realiza prácticamente en tiempo real, es decir, los oyentes escucharán lo que en ese momento se esté transmitiendo tal como lo harían con una radio convencional.

REFERENCIAS BIBLIOGRÁFICAS

CAPITULO 2: COMPRESIÓN DE AUDIO

LIBROS:

- [8] Salomon, D. (2007). Data Compression. Chapter 7, Audio Compression. Fourth Edition. London: Springer-Verlag.
- [9] Jiménez, M. (2007), Teoría de Información y Codificación, Capítulos 2 y 3, Escuela Politécnica Nacional, Quito.

INTERNET:

- [10] Suárez F. (2009), Compresión de la Información Multimedia, Universidad de Oviedo, Área de Arquitectura y Tecnología de Computadores, <http://www.atc.uniovi.es/teleco/5tm/archives/3comp.pdf>
- [11] Marín R., Compresión de la información, Fundamentos de la Compresión de Información, Dpto. de Ingeniería de la Información y las Comunicaciones, Facultad de Informática, Universidad de Murcia, <http://perseo.dif.um.es/aike/roque/index.htm>
- [12] ESPOL. (2009). Procesamiento de Audio y Video. Facultad de Ingeniería en Electricidad y Computación, ESPOL OCW. <http://www.ocw.espol.edu.ec/facultad-de-ingenieria-en-electricidad-y-computacion/procesamiento-de-audio-y-video-1/course-schedule>
- [13] Cádiz, R. (2008). Introducción a la Música Computacional. Centro de Investigación en Tecnologías de Audio, Instituto de Música, Pontificia Universidad Católica de Chile. <http://www.rodriгодadiz.com/imc/>
- [14] Pérez, C. (2005). Introducción a la Compresión de Audio. Dpto. de Ingeniería de Comunicaciones, Universidad de Cantabria. <http://personales.unican.es/perezvr/pdf/Introducci%C3%B3n%20a%20la%20compresi%C3%B3n%20de%20audio.pdf>
- [15] Cuello, Freddy F., y Juan Carlos Rueda. Compresión de datos de audio. http://tav.net/multimedia/sonido/compresion_datos/index.htm

CAPÍTULO 3.

SISTEMAS DE ALTA FIDELIDAD

3.1 INTRODUCCIÓN

En este capítulo se va a tratar acerca de la alta fidelidad y de la relación con los formatos de audio comprimidos haciendo énfasis a la relación con el formato de audio MP3.

Para poder entender la vinculación de los formatos de audio comprimidos con los sistemas de alta fidelidad, primero hay que definir qué es un sistema de alta fidelidad. Como el mundo de la alta fidelidad es un mundo muy interesante pero a la vez muy extenso en este capítulo se va a cubrir aspectos básicos y generales que permitan tener una idea clara de lo que es la alta fidelidad.

3.2 ALTA FIDELIDAD

3.2.1 PREÁMBULO

Alta fidelidad o HiFi (High Fidelity) es un término usado para describir la capacidad de un sistema de reproducir el sonido con un alto grado de realismo, la alta fidelidad pretende que los ruidos y la distorsión sean mínimos. Al hablar de alta fidelidad se hace mención de la reproducción y grabación de información de audio, con la mayor precisión posible, comparado con su versión original.

Es un error pensar que el sonido profesional es lo mejor, y no es así, simplemente es diferente y tiene una aplicación distinta. Igual que un equipo doméstico, no

satisface las expectativas en una sonorización²⁵ al aire libre, un equipo profesional no es idóneo para una sala de audición de dimensiones normales, ni es capaz de reproducir tantos detalles y sutilezas que un equipo de alta fidelidad.

Habitualmente se habla de cuatro mundos en cuanto a la alta fidelidad, los mismos que son: cine en casa, estudios de grabación, sonorización de exteriores, y car audio, un estudio más profundo de los 4 mundos de la alta fidelidad involucra un análisis complejo, el mismo que es tratado por ingenieros o expertos en el campo del sonido y de la acústica.

3.2.2 HISTORIA

La alta fidelidad, aparece en Estados Unidos tras la segunda guerra mundial, acuñado por los técnicos que buscaban diseñar sistemas de grabación, reproducción, amplificación y altavoces, con la mínima distorsión posible, con la mejor relación señal/ruido alcanzable y capaces de mostrar la totalidad de las frecuencias que el oído humano puede percibir como sonido, en general de 20 Hz a 20.000 Hz.

Así, en ese afán de mejorar la fidelidad, se dieron distintos avances, desde los rollos de cera de los primeros fonógrafos, los discos de 78 rpm²⁶ de acetato, los discos de 33 1/3 rpm Long Play (LP) microsurco de vinilo, la radio FM, con mayor ancho de banda de audio y menor susceptibilidad a la interferencia de la señal, mejores diseños de amplificadores, con mayor atención a la respuesta de frecuencia²⁷ y potencia de salida de mucha mayor capacidad y muchos otros avances.

²⁵ Sonorización: denominada también refuerzo sonoro, consiste en la amplificación del audio empleando grandes y sofisticados sistemas de audio en escenarios de gran envergadura.

²⁶ rpm: revoluciones por minuto

²⁷ Respuesta de frecuencia: la respuesta de frecuencia es un parámetro que describe las frecuencias que puede grabar o reproducir un dispositivo.

Sin olvidar nunca que el fin último de todos los desarrollos es recrear, a domicilio, la ilusión de una música que alguien interpretó en otro sitio y momento, en el fondo trata de la capacidad de un equipo de sonido para recrear de forma creíble y emotiva la música que en su momento fue grabada.

En los años 50 el término "alta fidelidad" empezó a ser usado por fabricantes de equipos de audio como un término de marketing para describir discos y equipos que pretendían suministrar una reproducción fiel del sonido.

Al período entre finales de los años 50 y principios de los años 60, se lo conoció como "La Edad de Oro de Hi-Fi", cuando los fabricantes de equipos de válvulas crearon algunos de sus modelos más sofisticados, justo antes de que aparezcan los equipos de estado sólido (equipos que utilizan transistores).

3.2.3 GENERALIDADES

La alta fidelidad es deudora en su desarrollo de otras aplicaciones en las que el sonido es parte fundamental, como el teléfono, la radio o el cine, que ya estaban muy perfeccionadas a principios de los años 40. Tanto las válvulas de vacío, como la posibilidad de coleccionar y almacenar sonido en todo su rango de frecuencias, ya existía a finales de los años 30, y algunos de ellos, como las válvulas que hoy siguen vigentes, datan de finales de los años 20. Así pues se ve que la alta fidelidad apareció antes que el sonido estéreo y éste es muy anterior al multicanal.

La alta fidelidad trata de lograr que en la casa por ejemplo, acondicionando la sala adecuadamente; escogiendo los altavoces apropiados, seleccionando la amplificación y empleando la fuente de sonido adecuada en función de las grabaciones de las que se disponga para escuchar, se pueda sentir que realmente hay personas frente al público haciendo música. Y esto, es lo que, para diferenciarlo de la masificada HiFi, se conoce como "High End" o búsqueda de la excelencia en el sonido reproducido.

El problema es que no existe un sonido perfecto absoluto para todos, cada uno según su percepción y su subjetividad puede encontrar su sonido ideal. El poder encontrar un sonido ideal puede también requerir un poco de adiestramiento del sistema auditivo, lo cual no es una tarea fácil pero no imposible. Ingenieros en sonido y muchos expertos, con la experiencia, pueden mostrar un buen adiestramiento del sistema auditivo pudiendo saber con mucha precisión la diferencia, de parámetros como: frecuencia de muestreo, profundidad de bits y formato de los diversos archivos de audio.

3.2.3.1 Hi-Fi y apariencia de realismo

En un principio la alta fidelidad se limitaba a la reproducción de sonido monofónico, la aproximación realista a lo que un oyente experimentaría en una sala de conciertos era limitada. Los investigadores se dieron cuenta tempranamente de que la manera ideal para tener una experiencia musical reproducida por un equipo de audio era a través de múltiples canales de transmisión, pero la tecnología no estaba disponible en ese momento. Por ejemplo, se descubrió que una representación realista de la separación entre intérpretes en una orquesta desde una posición de escucha ideal en la sala de concierto requeriría por lo menos tres altavoces para los canales frontales.

Para la reproducción de la reverberación²⁸, por lo menos se requerirían dos altavoces detrás o a los lados del oyente. El sonido estereofónico suministró una solución parcial al problema de crear alguna apariencia de la ilusión de intérpretes tocando en una orquesta.

Un intento para suministrar la reproducción de la reverberación se probó en 1970 a través del sonido cuadrafónico, pero, otra vez, la tecnología en ese momento era insuficiente para la tarea. Los consumidores no querían pagar el costo adicional requerido en dinero y espacio para un mejoramiento marginal en el

²⁸ Reverberación: fenómeno acústico que consiste en la intensificación del sonido a causa de múltiples reflexiones que sufre antes de llegar al oído.

realismo. Con el incremento de la popularidad del cine en casa²⁹, sin embargo, los sistemas de reproducción multicanal se volvieron asequibles, y los consumidores se dispusieron a tolerar los seis a ocho canales de los equipos de cine en casa.

Además de realismo espacial, la reproducción de la música debe ser subjetivamente libre de ruido para lograr realismo. El disco compacto (CD), prevé al menos 90 decibeles de rango dinámico, que es tanto como la mayoría de las personas pueden tolerar en una media de sala de estar. Esto requiere, por lo tanto, el equipo de reproducción para ofrecer una relación señal/ruido de al menos 90 decibeles, además el equipo de audio debe ser capaz de reproducir frecuencias suficientemente altas y bajas para darnos un resultado realista.

3.2.3.2 Modularidad

Un sistema de alta fidelidad puede ser un sistema integrado (todo en uno) o puede estar constituido de varios módulos, entre los módulos se pueden tener: amplificadores, tocadiscos, digital media players, caseteras (ahora en desuso, casi ni se fabrican), reproductores de audio digital, reproductores de DVD o Blue-Ray que reproducen una amplia variedad de discos incluyendo CD's, grabadoras de CD, grabadores MiniDisc, grabadoras de video (ecualizadores, procesadores de señal, y cajas de parlantes (con tweeters, midranges y subwoofers³⁰).

Esta modularidad permite al melómano (persona fanática de la música) o audiófilo (persona amante del audio de buena calidad), gastar como mucho o poco, lo que quiera en un componente que se adapte a sus necesidades específicas. Un sistema construido a partir de módulos, permite el uso parcial del resto del

²⁹ Cine en casa: Si el sistema de alta fidelidad incluye componentes como un reproductor DVD, un proyector, o una televisión, entonces es llamado frecuentemente como cine en casa.

³⁰ Tweeter: altavoz reproductor de ultra-altos o agudos, cubren frecuencias de audio desde 4.000 a 20.000 HZ.

Midrange: altavoz de medios-bajos, cubre rango desde 400 a 4.000 HZ.

Woofer: altavoz de ultra-bajos o graves reproduce frecuencias de audio desde 400 a 40 HZ.

sistema en el caso que alguno falle o se descomponga. Una reparación de un sistema integrado, sin embargo, significa la ausencia total de uso del sistema.

3.2.4 ASPECTOS DE LA ALTA FIDELIDAD

Los sonidos reales son muy complejos, la mezcla de sonidos de varios instrumentos tocados al mismo tiempo dan a la onda una forma muy irregular.

El audio original pasa por tantos equipos que es difícil que sea una réplica fiel del original. Cada componente es deficiente en la perfecta duplicación por una cantidad pequeña y medible. Si el total excede el 15%, es decir, si el sonido reproducido difiere un 15% del original, dicho sonido no puede ser llamado alta fidelidad.

A continuación en la tabla 3.1 algunas cantidades típicas de deformación en un sistema de audio:

Dispositivo	Tolerancia
Micrófono	2 %
Amplificador	1 %
Grabador	2 %
Duplicador ³¹	1 %
Fonocaptor ³²	2,5 %
Parlante	3,5 %
TOTAL	12%

Tabla 3. 1: Deformación de un sistema de audio

El sistema de la tabla 3.1 es sin duda un sistema muy bueno ya que no excede el 15%.

³¹ Duplicador: dispositivo que permite dividir la señal de audio, trabaja como un splitter.

³² Fonocaptor: aparato que, aplicado a un disco de gramófono, permite reproducir electrónicamente las vibraciones inscritas en el disco.

La razón para que el oído pueda aceptar estas altas cifras es que el oído humano se engaña fácilmente. Sin embargo, una vez acostumbrado a una buena reproducción del sonido, no se satisfará luego con menos.

3.2.5 ALTA FIDELIDAD COMO NORMA DE CALIDAD

La alta fidelidad como norma de calidad significa que la reproducción del audio es muy fiel al original. El término alta fidelidad, se aplica normalmente a todo sistema doméstico de razonable calidad. En un intento por normar la alta fidelidad en 1973, el Deutsches Institut für Normung (DIN: Instituto Alemán de Normalización) creó la norma DIN 45500; la norma estableció requerimientos mínimos de las medidas de: respuesta de frecuencia, distorsión, ruido y otros parámetros; logrando algún reconocimiento de algunas revistas de audio.

Cumpliendo esa norma se han fabricado equipos de música de una calidad muy razonable, pero este estándar no es el único ya que existen y se fabrican muchos equipos y sistemas de audio bajo otros estándares de Hi-Fi, estándares propietarios. Como la norma DIN 45500 no es la única norma, no resulta necesario mencionar las características técnicas citadas en esta norma; ya que por razones de marketing la mayoría de equipos se publicitan como de alta fidelidad por cubrir el rango de frecuencias que va desde 20 Hz a 20 KHz.

La norma DIN 45500 fue bien intencionada, pero cumplió sólo con éxito limitado normar la definición de alta fidelidad. En los comienzos la misma norma garantizaba que los equipos cumplieran con lo especificado en ésta, por lo tanto solo los equipos que habían pasado las pruebas podían llevar el nombre de Hi-Fi, pero un tiempo después muchos otros fabricantes comenzaron a poner ese nombre en sus productos sin que esta norma lo avale y así la norma 45500 fue perdiendo popularidad.

En la década de 1990, habiendo mejorado la tecnología, se acuñó otro término para definir otro estándar con niveles cualitativos más altos, el "Hi-End" ; el audio High End es un concepto que se inventó para los transistores incorporando

componentes de mayor calidad y estética, como conectores bañados en oro, etc.; más o menos se entiende como una alta fidelidad más exquisita, más purista, más alejada de lo comercial, más exclusiva, con una tecnología de más alto nivel, y en definitiva mucho más cara.

3.2.6 EQUIPOS DE ALTA FIDELIDAD

La calidad de un sistema de grabación y reproducción de alta fidelidad se mide a partir de la capacidad de éste de reproducir la señal de audio de manera más fiel a la realidad, es decir, con el máximo parecido respecto a la señal original. Eso significa por ejemplo, conservar la naturaleza y la perfecta inteligibilidad de la voz o las propiedades tímbricas de los distintos instrumentos.

Entre las características más importantes que ha de tener un sistema para ser considerado de alta fidelidad, hay que destacar tanto la respuesta en frecuencia, que ha de ser suficientemente uniforme y amplia en el campo de las frecuencias audibles, como la ausencia de distorsión de señal, que se evidencia cuando se introduce algún componente que no estaba presente en la señal original y produce una especie de ruido de fondo. Además la fidelidad está condicionada por las características ambientales donde tiene lugar la audición.

Lo cierto es que un buen sistema de alta fidelidad crea las condiciones potenciales para una reproducción del audio satisfactoria. De todas formas no es conveniente exagerar respecto a los resultados en torno a la perfección de cada componente del equipo, ya que siempre pueden existir nuevos elementos que intervengan de manera determinante y que conviertan en mediocre la reproducción efectuada con el equipo técnicamente más sofisticado y costoso.

3.2.6.1 Equipos modernos

Equipos modernos de alta fidelidad de audio digital generalmente incluyen fuentes de señal tales como: reproductores de CD, cinta de audio digital (DAT) y Digital

Audio Broadcasting (DAB), sintonizadores de radio HD, un amplificador, un preamplificador, y altavoces. Algunos equipos modernos de alta fidelidad pueden ser conectados digitalmente, utilizando hilos de fibra óptica y pueden tener puertos seriales universales (USB) y soporte de fidelidad inalámbrica (WiFi).

3.3 COMPARATIVA CON EL FORMATO MP3

Una amplia franja de la población, escucha música en formato MP3 bajada gratuitamente a través de la Internet, directamente en sus ordenadores o reproductores portátiles sin preocuparse mucho por la calidad del sonido. En la actualidad existen servidores de música que contienen archivos de música en formato comprimido y en la mayoría de los casos en formato MP3.

Con mayores avances tecnológicos en temas como la compresión y en especial con la aparición del formato MP3 se puede decir que calidad sonora empeoró a criterio de expertos en audio y audiófilos. Hace un tiempo existía mucha gente a la cual le parecía una barbaridad escuchar música comprimida en MP3 ya que con la utilización de los formatos comprimidos como el MP3 se pierde parte del sonido original; pero en la actualidad ir contra los formatos de audio comprimido y en especial contra el MP3, sería como remar contra la corriente.

Muchos sectores detractores de los formatos comprimidos argumentan que la compresión de audio provoca la pérdida de la emoción musical y se preguntan por qué en la era de los mayores avances tecnológicos la calidad sonora empeoró como nunca. Es muy lógico que tanto, los formatos de audio sin pérdida o con pérdida sufran la degradación de la fidelidad, ya que el audio reproducido en las computadoras actuales o dispositivos multimedia es un audio digital. Al pasar las señales analógicas por el proceso de la digitalización las señales originales sufren cambios que pueden ser muy notorios, dependiendo del procesamiento de la señal y de los algoritmos de compresión.

Es cierto que los formatos de compresión han degradado la calidad musical, por ello por más éxito y expansión que tenga el MP3, siempre habrán puristas dedicados a escuchar sólo en vinilo o en cinta y con equipos valvulares. Ya sea por aspectos psicológicos, subjetivos o por nostalgia de la época de antaño, varias empresas como Apple, por motivos comerciales principalmente, han dedicado parte de sus esfuerzos a producir equipos híbridos, como el que se muestra en la figura 3.1, para de esta manera satisfacer a un sector de la población que nunca se ha olvidado de los sistemas de audio antiguos.



Figura 3. 1: iPod Híbrido

Los fetichistas del vinilo siempre hablan sobre la música de antaño, suelen mencionar que aquella música fue hecha por gente en una habitación, no por una computadora. Tanto los archivos comprimidos, como el CD son el equivalente a escuchar una copia de esta música, aunque si uno no escuchara el original quizá nunca se daría cuenta.

En cuanto a los formatos comprimidos se hace mucho énfasis en el formato de música grabada en MP3 ya que a pesar de que existen innumerables códecs de audio el formato MP3 es uno de los que más sobresale.

Parecería que al hablar de MP3 y de alta fidelidad, se habla de palabras muy opuestas, pero en la actualidad no es así, ya que por marketing y por motivos en su mayoría comerciales, gran parte de los equipos comerciales son denominados

de alta fidelidad si cubren el rango de frecuencias de 20 Hz a 20 KHz y muy poco se preocupan que el equipo reproduzca formatos comprimidos, es más, la reproducción de los formatos comprimidos es una de las características más publicitada de los dispositivos multimedia y sistemas de audio, como ejemplo de esto se presenta la figura 3.2 con sus características más importantes, resaltando la posibilidad de reproducir archivos de formato MP3.



Figura 3. 2: iPod Shuffle de 2 GB

Equipo (Ipod)	Principales características técnicas
Apple iPod Shuffle 2GB 3era generación MP3	Unidad flash de 2 GB
	Reproduce hasta 500 o 1.000 canciones
	Formatos AAC, MP3 , WAV

Tabla 3. 2: Características del iPod Shuffle de 2 GB

Finalmente hay que tener en cuenta que la calidad final no sólo depende del formato utilizado, sino también del equipo de audio, los altavoces, la forma del mobiliario y el entorno.

Es innegable que un archivo no comprimido suena mejor que un archivo en formato MP3 y que a criterio de muchos expertos los sistemas de audio de los años 50 y 60 suenan mejor que los de ahora, pero hoy en día la calidad de audio de un mp3 puede ser aceptable y hasta más que suficiente para el usuario promedio. Además el formato se debe elegir de acuerdo a las necesidades, posibilidades y objetivos.

REFERENCIAS BIBLIOGRÁFICAS

CAPITULO 3: SISTEMAS DE ALTA FIDELIDAD

LIBROS:

- [16] Lazzati, M., (2009), Electroacústica III, Universidad de las Américas - Quito, Ingeniería en Sonido y Acústica.

INTERNET:

- [17] Como Funcionan, (2009), Altas fidelidades, Como Funcionan.com,
<http://www.como-funcionan.com>
- [18] JVC, (2009), Componentes Hi-Fi, Barcelona-España.
<http://www.jvc.es>
- [19] Pro Audio, (2008), Revista Pro Audio Show,
<http://issuu.com/revistaacm/docs/webabril08>
- [20] Doctor ProAudio, (2009), Biblioteca Hi-Fi,
<http://www.doctorproaudio.com>
- [21] Philips, (2007), Koninklijke Philips Electronics N.V, version 2.0,
<http://www.philips.com>
- [22] Chiacchio, D. (2009), Audio Hi-Fi y Hi-End, Audiohiend,
<http://www.audiohiend.com.ar>
- [23] Audiotest.org, (2010), Pro Audio Recording.
<http://www.audiotest.org>
- [24] Meyer Sound, (2009), High Fidelity Case Study, Houson-Texas.
<http://www.meyersound.com/spanish/products/mseries/milo/news.php>

CAPÍTULO 4.

FORMATO MP3

4.1 INTRODUCCIÓN

4.1.1 HISTORIA

A finales de los años '80, se convirtió en necesidad potenciar la digitalización del sonido más allá de los formatos existentes. Una de las prioridades era crear un sistema de compresión, ya que el formato WAV que hasta ese momento se utilizaba, requería mucho espacio en disco, algo que dificultaba su almacenamiento y transporte.

La tecnología del formato de audio MP3³³ fue desarrollada en Alemania por Karlheinz Brandenburg, Harald Popp y Bernhard Grill, tres científicos del instituto tecnológico de Fraunhofer en Ilemenau en el año 1986. Luego en 1992 el Moving Picture Experts Group (MPEG) aprobó oficialmente la tecnología. En ese país, ganó el premio a la innovación tecnológica "Future Prize" del gobierno en el año 2000.



Figura 4. 1: Creadores del formato MP3,

Bernhard Grill, Karlheinz Brandenburg y Harald Popp (de izquierda a derecha)

³³ Nota legal: el formato MP3 es un formato propietario del Instituto Fraunhofer y ha sido utilizado sólo con fines académicos en el presente proyecto de titulación.

Brandenburg, describió él mismo la forma en que se incorporó al proyecto: “A comienzos de los ochenta, en la época en que se digitalizó la red telefónica en Alemania, el profesor Dieter Seitzer, de la Universidad de Erlangen, tuvo la idea de transmitir por la línea telefónica RDSI de 64 kilobits por segundo algo más que la voz. Entonces comenzó a investigar un pequeño grupo en el que yo entré para hacer mi tesis de fin de carrera.”

Su planteamiento era que para transmitir toda la información de un CD por la línea telefónica se necesitaba hacerlo a 1,5 megabits por segundo, pero sólo se podía transmitir 64 kilobits por segundo. Brandenburg pensó en que si se comprimía hasta doce veces la información, se podría transmitir por una sola línea telefónica. Lo primero que desarrolló fue un códec, un programa que podía comprimir y descomprimir audios, manteniendo en gran parte la calidad del sonido.

Y de ahí al gran salto, al MP3 (las siglas o el diminutivo de MPEG-1 Audio Layer 3 o MPEG-1 Capa de Audio 3).

“La posibilidad de intercambiar música por Internet la vimos desde el principio y se la ofrecimos a la industria... La respuesta que nos dieron entonces hoy les debe de dar vergüenza: ‘Pero si nadie tiene Internet en casa’. Entonces ya teníamos la forma de codificarlo que permitía la venta por Internet, pero no supieron ver el potencial y nos mandaron a casa.” [36]

Por supuesto que no fue fácil hacer que este formato se volviera popular. En sus inicios era visto como un demonio, por las discográficas por supuesto, quienes trataban de cerrarle las puertas de cualquier manera mostrando una imagen irreal de lo que realmente era. Pero de nada les sirvió esos intentos. Hoy MP3 es el formato de la música online.

4.1.1.1 Cronología del mp3

En 1987, el Instituto Fraunhofer en Alemania inició una investigación bajo el nombre de proyecto Eureka EU147, Digital Audio Broadcasting.

En el mes de enero de 1988, se creó el Grupo de Expertos de Imágenes en Movimiento o MPEG como una subcomisión de la Organización Internacional de Estandarización (ISO).

En abril de 1989, Fraunhofer obtuvo la patente alemana para el MP3.

En el año 1992, el algoritmo de codificación de audio de Fraunhofer y de Dieter Seitzer fue integrado en MPEG-1.

En el año 1993, el estándar MPEG-1 fue publicado.

En 1994, fue desarrollado MPEG-2 y publicado un año después.

En el mes de noviembre de 1996, fueron concedidas las patentes del MP3 para los Estados Unidos.

En septiembre de 1998, Fraunhofer empezó a hacer valer sus derechos de patente haciendo que todos los desarrolladores de codificadores MP3 se vieran obligados a pagar una cuota de licencia a Fraunhofer.

En febrero de 1999, una empresa disquera llamada SubPop comenzó a distribuir las pistas de música en formato MP3.

En el mismo año aparecieron los reproductores de mp3 portátiles, los cuales se han masificado, siendo un estandarte del género el iPod³⁴ de Apple.

4.1.2 DEFINICIÓN

MP3 es un formato de datos que debe su nombre a un algoritmo de codificación llamado MPEG 1 Layer 3, el cual, a su vez, es un sistema de compresión de audio que permite almacenar sonido con una calidad similar a la de un CD. Este formato se utiliza para comprimir formatos de audio normales (WAV o CD audio) en una relación de 1:12. En la práctica, permite almacenar el equivalente a 12 CD-ROM de álbumes de música en el espacio de un solo CD, es decir, unas 150 canciones aproximadamente. Es más, el formato MP3 casi no altera la calidad del sonido para el oído humano.

³⁴ iPod es una marca de reproductores multimedia portátiles diseñados y comercializados por Apple Inc., el iPod puede reproducir archivos MP3, WAV, AAC/M4A, AIFF y Apple Lossless. El iPod de 5ª generación también es capaz de reproducir archivos de video en formatos H.264 y MPEG-4.

El sistema de codificación que utiliza MP3 es un algoritmo de compresión con pérdidas. Es decir, el sonido original y el que obtenemos posteriormente no son idénticos. Técnicamente hablando es un sistema de compresión aplicado al audio, cuya invención se basa en un estudio relacionado con el oído humano. Para crear mp3, los científicos analizaron cómo el oído humano y el cerebro perciben el sonido. Esta técnica de codificación, de alguna manera, “engaña” el oído eliminando las partes menos esenciales de un archivo de música. Por ejemplo, si dos notas son muy similares, o si un tono alto y bajo ocurren exactamente al mismo tiempo, el cerebro percibe sólo a uno de ellos; entonces el algoritmo mp3 selecciona la señal más importante y desecha la otra. Los investigadores recortaron, además, las frecuencias muy altas y muy bajas que no son percibidas por el oído. De esta manera, la pérdida es inaudible y el tamaño del archivo se reduce 12 veces en relación a su original.

La principal ventaja evidente fue el reducido tamaño y buena calidad en comparación con el resto de los formatos en ese momento. Para establecer comparaciones, un CD de música de tracks (pistas) es grabado a 44,1 KHz y tiene 16 bits con dos canales (estéreo), esto equivale a 172 KB. Por lo tanto una canción típica de 4'30" ocupa 46.512 KB o unos 45 MB. Bajar archivos tan grandes en las velocidades de Internet de hace unos años no era tarea fácil; además en un CD normal no entrarían más de 14 o 15 canciones.

El MP3 es, de lejos, el formato de audio más popular; prácticamente todos los aparatos reproductores soportan el formato de audio comprimido MP3 como "el denominador común" para la más fácil interoperabilidad entre aplicaciones diferentes. A raíz del descubrimiento del mp3, empresas como Apple Computers, Real Networks o Sony han desarrollado, cada uno, tecnologías propias de reproducción. Es inevitable, en la actualidad, relacionar el mp3 con la descarga de archivos a través de las redes peer-to-peer (P2P)³⁵.

³⁵ Una red peer-to-peer (P2P) o red de pares, es una red de computadoras en la que todos o algunos aspectos de ésta funcionan sin clientes ni servidores fijos, sino una serie de nodos que se comportan como iguales entre sí. Es decir, actúan simultáneamente como clientes y servidores respecto a los demás nodos de la red.

4.2 ASPECTOS PRINCIPALES DEL ESTÁNDAR ISO/IEC 11172-3 (MPEG-1 CAPA DE AUDIO III)

La parte concerniente al audio del estándar ISO/IEC 11172 es la parte 3, la misma que es designada como estándar ISO/IEC 11172-3.

La primera edición del estándar ISO/IEC 11172-3 fue publicada en inglés, en Ginebra (Suiza) el 1 de agosto de 1993, posteriormente se hace una corrección al estándar el 15 de abril de 1995, así mismo, consecuentemente la organización conjunta ISO/IEC realiza dos revisiones adicionales al estándar una en el año 2000 y otra en el año 2002.

De manera general se puede decir que el estándar internacional ISO/IEC 11172-3 posee extensa información tecnológica en cuanto al audio y lo que al procesamiento de la señal se refiere; el estándar provee aspectos tanto informativos como normativos. El estándar contiene aspectos generales, elementos técnicos, para un mejor entendimiento de la norma, y varios anexos los cuales contienen a su vez variada y muy explícita información incluyendo conceptos, algoritmos y diagramas para un mejor entendimiento de los diversos procesos que están vinculados al estándar internacional.

Los elementos generales, los elementos técnicos y los anexos A y B son de carácter normativo. En tanto los otros anexos (del C al G) son de carácter informativo, es decir, no obligan a seguir al pie de la letra uno u otro procedimiento, por el contrario brindan la libertad y visto de cierta forma dan pautas o sugerencias para realizar cualquier trabajo o tarea; por ejemplo el estándar ISO/IEC 11172-3 no tiene definido ningún modelo psicoacústico pero a su vez provee dos modelos, uno más complicado que el otro, y que pueden ser utilizados indistintamente según el trabajo, los recursos disponibles o la conveniencia.

Como la información que provee el estándar es muy interesante pero a la vez muy extensa, en este capítulo se pretende hacer un resumen de los aspectos más

relevantes y los que estén más ligados al proceso de la codificación con el objetivo de obtener archivos .mp3 válidos.

Gran parte de la información técnica está contenida en los anexos los mismos que de manera general son descritos a continuación:

- Anexo A (normativo) - Diagramas
- Anexo B (normativo) - Tablas
- Anexo C (informativo) - Proceso de codificación
- Anexo D (informativo) - Modelos psicoacústicos
- Anexo E (informativo) - Sensibilidad de los bits individuales a errores
- Anexo F (informativo) - Cancelación de errores
- Anexo G (informativo) - Codificación Joint Stereo

4.2.1 CODIFICACIÓN

El codificador procesa la señal de audio digital y produce el bitstream empaquetado para su almacenamiento y/o transmisión. El algoritmo de codificación no está determinado, y puede utilizar enmascaramiento, cuantización variable y escalada. Sin embargo, debe ajustarse a las especificaciones del decodificador.

La figura 4.2 ilustra el esquema del codificador según la norma ISO/IEC 11172-3.

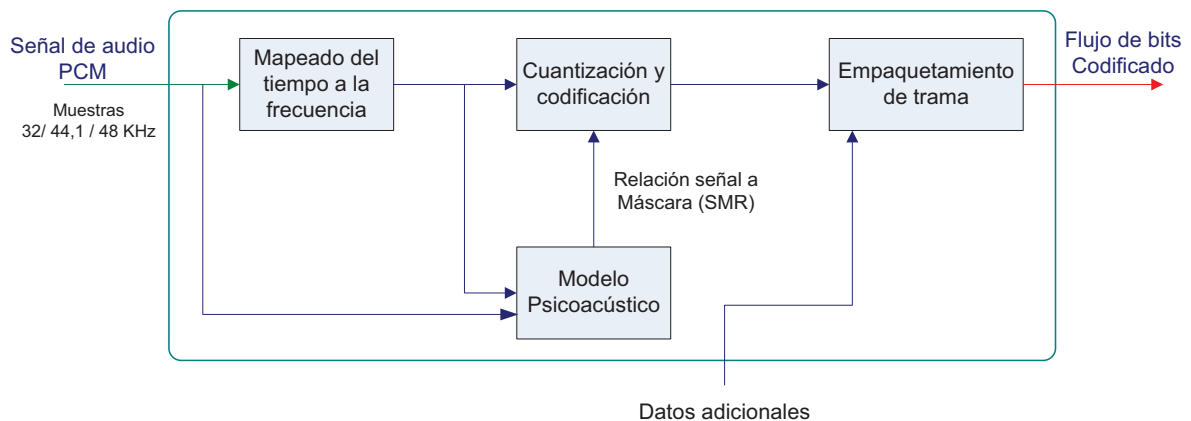


Figura 4. 2: Codificador según la norma ISO/IEC 11172-3

Las muestras se introducen en el codificador y a continuación el mapeador crea una representación filtrada y submuestreada de la señal de entrada. Las muestras mapeadas se denominan muestras de subbanda transformadas. El modelo psicoacústico crea una serie de datos (dependiendo de la implementación del codificador) que sirven para controlar la cuantización y codificación. Este último bloque crea a su vez su propia serie de datos, de nuevo dependiendo de la implementación. Por último, el bloque de empaquetamiento de trama se encarga de agrupar como corresponde todos los datos, pudiendo añadir algunos más, llamados datos adicionales, como por ejemplo CRC o información del usuario.

4.2.1.1 Codificación MPEG-1 para la capa 3

Para comprender detalladamente el proceso de codificación descrito en el estándar ISO/IEC 11172-3, se describe a continuación su funcionamiento.

El mapeado de tiempo-frecuencia para la codificación MPEG-1 capa 3 (MP3) añade un nuevo banco de filtros, procesando las salidas filtradas con una Transformada Discreta del Coseno Modificada (MDCT), estos dos bloques conforman el denominado *filtro híbrido*, el cual proporciona una resolución en frecuencia variable, 6 x 32 o 18 x 32 subbandas, ajustándose mucho mejor a las bandas críticas de las diferentes frecuencias. A diferencia del banco de filtros polifásicos, sin cuantización, la MDCT no presenta pérdidas. La MDCT también subdivide las salidas de subbanda en frecuencia para incrementar la resolución y así poder dividir el audio en bandas que se ajustan mejor a las bandas críticas del oído. Una vez que los componentes de subbanda son subdivididos en frecuencia, el codificador cancela parte del aliasing causado por el banco de filtros polifásicos.

El modelo psicoacústico utiliza un método denominado predicción polinómica, incluyendo los efectos totales del enmascaramiento tanto en frecuencia como en el tiempo.

El bloque de cuantización y codificación también emplea algoritmos muy sofisticados que permiten tramas de longitud variable, donde, la variable

controlada es el ruido, a través de bucles iterativos que lo reducen al mínimo posible en cada paso.

En el formateador de trama³⁶, 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. El empaquetado de trama incluye el uso de una reserva de bits (*bit reservoir*), que hace posible emplear más bits en partes de la señal que lo necesiten. La longitud de la trama puede variarse en caso de necesidad. También, permite alta calidad en el audio a tasas tan bajas como 64 Kbps.

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.

En la figura 4.3 se muestra un diagrama en bloques más detallado del codificador MP3.

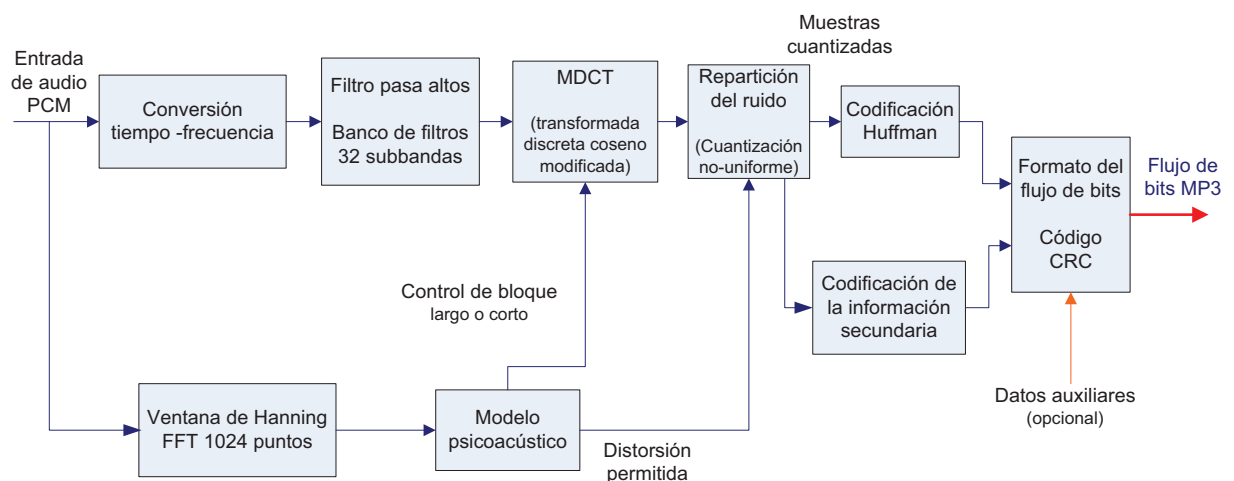


Figura 4. 3: Diagrama de bloques de un codificador MP3

³⁶ La definición de trama para este esquema según ISO nos dice que: "Trama es la mínima parte del bitstream decodificable mediante el uso de información principal adquirida previamente".

En forma resumida el proceso de codificación es el siguiente: el flujo de audio a la entrada pasa a través de un banco de filtros que divide la señal en múltiples subbandas. En forma paralela al filtrado se realiza el análisis psicoacústico que determina el ruido (distorsión permitida) en cada subbanda. La etapa "Repartición de ruido" usa las distorsiones permitidas para decidir cómo dividir el número total de bits de código disponibles.

Por último, las muestras codificadas mediante Huffman junto con la información secundaria son convertidas en un flujo de bits MP3 válido.

4.2.1.2 Análisis psicoacústico

El estándar ISO/IEC 11172-3 (que describe la manera de generar flujos de audio MP3 válidos), proporciona dos modelos psicoacústicos; el primer modelo psicoacústico es menos complejo que el segundo y simplifica mucho los cálculos. Ambos modelos trabajan para cualquiera de las capas, aunque requieren adaptaciones específicas para el esquema de la capa 3. Existe considerable libertad en la implementación del modelo psicoacústico; la precisión que se requiera del modelo es dependiente de la aplicación y de la tasa de bits que se quiere lograr. Para bajos niveles de compresión, donde hay un número generoso de bits para realizar la codificación, el modelo psicoacústico puede ser completamente omitido, en cuyo caso, sólo se calcula la SNR³⁷ más baja, y con este valor se realiza el proceso de repartición de ruido para la subbanda.

El modelo psicoacústico II que se usa en la Capa 3 tiene mejoras adicionales que se adaptan mejor a las propiedades del oído humano. Primero el modelo convierte el audio al dominio espectral, usando una FFT de 1024 puntos para conseguir una buena resolución de frecuencia y poder calcular correctamente los umbrales de enmascaramiento. Antes de la FFT, se aplica una ventana de Hanning convencional para evitar las discontinuidades en los extremos de la señal. La salida de la FFT se usa primero para analizar qué tipo de señal está

³⁷ SNR: Relación señal – ruido.

siendo procesada: una señal estacionaria hace que el modelo escoja bloques largos, y una señal con muchos transitorios da como resultado bloques cortos. El tipo de bloque se usa luego en la parte MDCT del algoritmo. Después de esto, el modelo psicoacústico calcula el mínimo umbral de enmascaramiento para cada subbanda. Estos valores de umbral se usan luego para calcular la distorsión permitida. El modelo pasa entonces las distorsiones permitidas a la sección "Repartición de ruido" en el codificador para uso posterior.

Algunas de las funciones del modelo psicoacústico se presentan a continuación:

4.2.1.2.1 Alineación en tiempo

Se debe tener en cuenta que cuando se hace la evaluación psicoacústica, los datos de audio que son enviados al modelo deben ser concurrentes con los datos de audio a ser codificados.

El modelo psicoacústico debe tener en cuenta el retardo de los datos al pasar por el banco de filtros y aplicar un desplazamiento adicional, de tal manera que los datos relevantes queden centrados en la ventana del análisis psicoacústico.

4.2.1.2.2 Representación espectral

El modelo psicoacústico realiza una conversión del tiempo a la frecuencia totalmente independiente del mapeo realizado por el banco de filtros porque necesita una mejor resolución en frecuencia para calcular con gran precisión los umbrales de enmascaramiento. Ambos modelos usan una transformada de Fourier para realizar el mapeo.

El modelo I usa una FFT de 1024 puntos para la capa 3. El análisis psicoacústico para la capa 3 se realiza sobre 1152 muestras, así que la FFT de 1024 puntos no proporciona cobertura total. Idealmente, la FFT debería cubrir todas las 1152

muestras; aunque 1024 puntos es un compromiso razonable ya que las muestras que se omiten, no tienen mayor impacto en el análisis psicoacústico.

El modelo II usa una FFT de 1024 puntos para la capa 3, el modelo ejecuta dos cálculos psicoacústicos de 1024 puntos. El primer cálculo se encarga de las 576 muestras iniciales, y el segundo cálculo se realiza sobre las últimas 576 muestras. El modelo II combina los resultados de ambos cálculos, de tal manera que el resultado total implique la selección del umbral de enmascaramiento de ruido (Noise Masking Treshold) más bajo en cada subbanda. Para simplificar los cálculos, ambos modelos procesan los valores espectrales en unidades perceptuales, denominadas Barks (Ec. 1.9).

4.2.1.2.3 Componentes tonales y no tonales

Ambos modelos identifican y separan las componentes tonales y las componentes de ruido en la señal de audio. Esto se debe a que cada componente presenta un tipo de enmascaramiento diferente.

El modelo I identifica las componentes tonales, basado en los picos locales del espectro de potencias. Después de procesar todas las componentes tonales, el modelo concentra los valores espectrales restantes en una única componente no tonal por banda crítica. El índice de frecuencia de cada una de estas componentes no tonales es el valor más cercano a la media geométrica de la banda crítica a la cual pertenece cada componente no tonal.

El modelo II realmente nunca separa las componentes tonales ni las no tonales, sino que calcula un índice de tonalidad en función de la frecuencia.

4.2.1.2.4 Estimación del índice de tonalidad

Este índice mide el comportamiento que presenta cada tipo de componente. El modelo II usa este índice para interpolar entre valores puros de TMN y valores

puros de NMT³⁸. El índice de tonalidad se basa en una predicción mediante una extrapolación lineal de los últimos dos cálculos, para predecir los valores de la componente que está siendo procesada. Las componentes tonales son más predecibles y, por lo tanto, tienen índices de tonalidad más altos. Este método de discriminación es mejor que el usado por el modelo I.

4.2.1.2.5 *Función de dispersión*

La capacidad de enmascarar de una componente determinada se distribuye por toda la banda crítica que la rodea. Ambos modelos determinan el umbral de enmascaramiento de ruido para ambos tipos de componentes; para lograr esto, el modelo I compara con un enmascaramiento determinado empíricamente, mientras que el modelo II aplica la función de dispersión descrita en la siguiente ecuación:

$$THN = E_N - 1.45 - B \quad (\text{Ec. 4.1})$$

Donde, THN = umbral de enmascaramiento del ruido
 E_N = nivel de energía enmascaradora de las bandas de ruido
 B = es el número de bandas críticas

En las aplicaciones de la capa 3, solo se toman en cuenta aquellos valores de la función de dispersión mayores a 60 dB.

4.2.1.2.6 *Umbral de enmascaramiento individual*

Para poder calcular el umbral de enmascaramiento global, el modelo I debe calcular primero los umbrales de enmascaramiento que cada componente tonal o no tonal genera sobre la señal de audio (llamados "Umbrales de enmascaramiento individuales"). Debe tenerse en cuenta que antes de esto se realiza un proceso conocido como "Decimation of maskers" (diezmado en la

³⁸ El enmascaramiento simultáneo o en frecuencia tiene tres tipos: Ruido enmascara tono (NMT), Tono enmascara ruido (TMN) y Ruido enmascara ruido (NMN).

cantidad de componentes enmascarantes). Este proceso consiste en escoger únicamente las componentes tonales y no tonales que verdaderamente enmascaran el sonido (cuya magnitud y distancia en Barks debe ser apropiada), desechando el resto de componentes calculadas en el paso anterior.

Después de realizada esta elección, el modelo I calcula el efecto de enmascaramiento que cada componente enmascaradora (tonal o no tonal) tiene sobre las líneas de frecuencia adyacentes a ella. Este análisis sólo es necesario hacerlo para las líneas de frecuencia que se encuentran entre -3 y +8 Barks a partir de la componente enmascaradora. O sea, el análisis abarca todas las líneas de frecuencia que se encuentren tres bandas críticas a la izquierda (hacia las bajas frecuencias), y ocho bandas críticas a la derecha (hacia las altas frecuencias) de la componente enmascaradora. Esto se debe a que el efecto de enmascaramiento de la componente tonal o no tonal que está siendo analizada (por más intensidad que ésta tenga) es demasiado tenue por fuera de este rango.

Como el modelo II nunca separa las componentes no tonales y tonales, sino que calcula el índice de tonalidad (en función de la frecuencia) que presenta cada componente enmascaradora, entonces no es necesario hacer el cálculo de los umbrales de enmascaramiento individuales.

4.2.1.2.7 Umbral de enmascaramiento global

Ambos modelos psicoacústicos incluyen un umbral de enmascaramiento absoluto, el cual ha sido determinado empíricamente: el mínimo umbral auditivo en un ambiente silencioso. Se debe recordar que éste es la intensidad del sonido más débil que se puede escuchar cuando no hay más sonidos presentes.

Usando el modelo I, este umbral absoluto se combina con los umbrales individuales calculados en el paso anterior para determinar el umbral de enmascaramiento global sobre toda la banda de audio.

El modelo II no calcula el umbral de enmascaramiento global, sino que trabaja todos los datos dentro de cada subbanda, de acuerdo con el índice de tonalidad que tenga cada componente enmascaradora en esa subbanda.

4.2.1.2.8 *Pre - eco*

Los efectos de pre-ecos son muy comunes cuando se trabaja con esquemas perceptuales de codificación de audio que usan alta resolución en frecuencia. Para entender el origen de los pre-ecos, consideremos el diagrama simplificado del decodificador de un sistema de codificación perceptual de la figura 4.11.

Las líneas de frecuencia reconstruidas son combinadas por el filtro síntesis, que consiste en una matriz de modulación y una ventana de síntesis. El error de cuantización introducido por el codificador puede verse como una señal agregada a las líneas de frecuencia originales, con un intervalo en el tiempo que es igual a la duración de la ventana de síntesis. Por esto, los errores de la reconstrucción se esparcen por toda la duración de la ventana. Si la señal de audio presenta un incremento abrupto de energía, el error de cuantización también se incrementa. Si ese pico de energía ocurre dentro de la ventana de síntesis, el error se esparcirá dentro de la ventana de síntesis completa, precediendo en el tiempo la causa real de su existencia. Si dicha señal pre-ruido se extiende más allá del período de pre-enmascaramiento del oído humano, se vuelve audible y se llama pre-eco.

La capa 3 incorpora varios pasos para reducir el pre-eco. Primero, el modelo psicoacústico de la capa 3 contiene modificaciones que detectan las condiciones de pre-eco. Segundo, la capa 3 puede pedir prestados codebits de la reserva de bits para reducir el ruido de cuantización cuando las condiciones de pre-eco se presentan.

Por último el codificador puede cambiar a un tamaño de bloque MDCT más pequeño para reducir el tiempo de ventana efectivo.

4.2.1.2.9 *Umbral de enmascaramiento mínimo*

Ambos modelos psicoacústicos seleccionan el mínimo umbral de enmascaramiento en cada subbanda.

Con el modelo I, para encontrar el umbral de enmascaramiento mínimo en cada subbanda, simplemente se extrae el mínimo valor del espectro global incluido entre las dos frecuencias límites de cada subbanda, o sea, el valor extraído del umbral global debe ser el valor mínimo de enmascaramiento en la subbanda. Este método se comporta bien para las subbandas más bajas donde la subbanda es estrecha con respecto a las bandas críticas, pero se vuelve inadecuado para las subbandas altas porque una banda crítica en esta frecuencia se distribuye sobre varias subbandas. Esta imprecisión se incrementa todavía más, debido a que el modelo I concentra todas las componentes no tonales, dentro de cada banda crítica, en un único valor para una sola frecuencia.

El modelo II selecciona el mínimo de todos los umbrales de enmascaramiento en cada subbanda sólo para regiones de frecuencia donde el ancho de la subbanda es amplio comparado con el ancho de la banda crítica. Si el ancho de la subbanda es estrecho en comparación con el ancho de la banda crítica, el modelo realiza un promedio entre todos los umbrales de enmascaramiento en esa subbanda. El modelo II es más preciso para las subbandas altas, ya que éste no concentra las componentes de ruido.

4.2.1.2.10 *Relaciones señal a máscara*

Los dos modelos calculan la relación señal a máscara, SMR, como la relación entre la energía de la señal en la subbanda (para la Capa 3, un grupo de bandas) y el mínimo umbral de enmascaramiento para esa subbanda. En el modelo psicoacústico de la Capa 3, el valor que se entrega no es la SMR, sino un valor equivalente llamado "Distorsión permitida" o "Ruido permitido". Este valor determina cuál es la cantidad máxima de ruido de cuantización que se permite en el bloque "Repartición de ruido".

4.2.1.3 Banco de filtros híbridos conmutados

El banco de filtros usado en MPEG-1 Capa 3 pertenece a la clase de bancos de filtros híbridos. Estos son construidos poniendo en cascada dos bancos de filtros diferentes, primero un banco de filtros polifásico y segundo, un banco de filtros con la transformada discreta del coseno modificada (MDCT).

Dependiendo del bitrate al que se vaya a codificar, este proceso eliminará más o menos datos siguiendo el modelo psicoacústico hasta lograr la compresión necesaria.

Dentro del formato MP3 podemos comprimir con distinto ancho de banda, modo y bitrate obteniendo distintas calidades según para que vayamos a utilizar ese sonido, como se muestra en la tabla 4.1.

calidad del sonido	ancho de banda	Modo	bitrate	radio de compresión
sonido telefónico	2,5 KHz	Mono	8 Kbps	96:1
mejor que onda corta	4,5 KHz	Mono	16 Kbps	48:1
mejor que radio AM	7,5 KHz	Mono	32 Kbps	24:1
similar a radio FM	11 KHz	estéreo	56...64 Kbps	26...24:1
cercano al CD	15 KHz	estéreo	96 Kbps	16:1
CD	>15 KHz	estéreo	112..128 Kbps	14..12:1

Tabla 4. 1: Razón de compresión de acuerdo al ancho de banda, modo, bitrate y radio de compresión

En un disco compacto (CD) tenemos una velocidad de muestreo de 44,1 KHz, muestreados a 16 bits en estéreo lo cual produce aproximadamente 1400 Kbps ($44100 \times 16 \times 2$ bits por segundo). Codificándolo por ejemplo a un MP3 de 128 Kbps obtenemos una reducción en torno al 1/12 del espacio inicial. También se puede optar por compresiones a mayor bitrate llegando a 192 o incluso 256 Kbps. Pero el más popular es el de 128 Kbps con el que se consigue una calidad excelente con una compresión sobresaliente.

4.2.1.3.1 Filtro pasa-altos

El estándar ISO/IEC 11172-3 proporciona respuesta en frecuencia hasta el nivel de 0 Hz. Sin embargo, para ciertas aplicaciones, se puede incluir un filtro pasa altos a la entrada del codificador, con su frecuencia de corte ubicada entre 2 y 10 Hz. La aplicación de tal filtro evita el innecesario requerimiento de una alta tasa de bits para la subbanda más baja y aumenta la calidad total en el sonido.

4.2.1.3.2 Banco de filtros polifásicos

El banco de filtros polifásicos divide la señal de audio en 32 subbandas de frecuencias igualmente espaciadas como se ve en la figura 4.4.

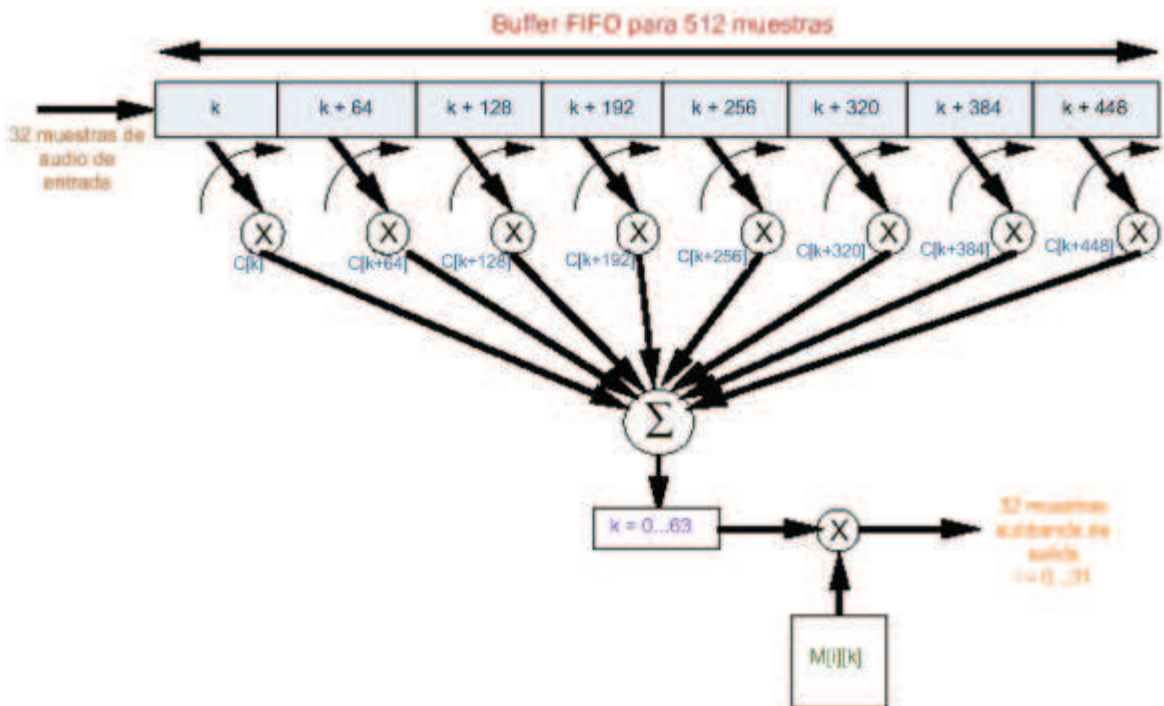


Figura 4. 4: Banco de filtros polifásicos [35]

Los filtros son relativamente simples y ofrecen una buena resolución temporal con una razonable resolución en frecuencia.

Para el diseño de estos filtros se deben de tener en cuenta algunas consideraciones:

Primero, la igualdad del ancho de las subbandas no representa adecuadamente las bandas críticas del oído. Muchos efectos psicoacústicos son consistentes si tomamos un escalado de frecuencias como el de las bandas críticas. Por ejemplo, la audición de una señal en presencia de una señal enmascaradora es diferente para señales que están dentro de una banda crítica que para señales que están más allá de una banda crítica. Para bajas frecuencias una subbanda abarca varias bandas críticas. En estas condiciones el número de bits de cuantización no puede ser explícitamente fijado por el ruido enmascarador disponible por cada una de las bandas críticas. En cambio, la banda crítica con el menor enmascaramiento de ruido nos da el número de bits de cuantización necesarios para toda la subbanda.

En segundo lugar, el banco de filtros polifásicos y su inversa son transformaciones con pérdidas. Incluso sin cuantización, la transformación inversa no puede recuperar perfectamente la señal original. Sin embargo, el error introducido por el banco de filtros polifásicos es pequeño e inaudible.

Finalmente, las bandas de dos filtros adyacentes tienen un gran solapamiento de frecuencia, esto es, un tono puede generar una salida en dos subbandas adyacentes.

La salida del filtro mostrado en la figura 4.4 es:

$$s_t(i) = \sum_{k=0}^{63} \sum_{j=0}^7 M[i][k] \times (C[k + 64j] \times x[k + 64j]) \quad (\text{Ec. 4.2})$$

Donde, i es el índice de la subbanda, los márgenes van desde 0 hasta 31.

$s_t(i)$ es la muestra de salida del filtro para la subbanda i en un tiempo t .

t es un entero, múltiplo de los 32 intervalos de muestras de audio.

$C[n]$ es uno de los 512 coeficientes de la ventana de análisis definida en el estándar.

$x[n]$ es la muestra de audio de entrada extraída de un buffer de 512 muestras.

$M[i][k] = \cos[(2i + 1) \times (k - 16) \times \pi/64]$ es la matriz de coeficientes de análisis.

La ecuación 4.2 está parcialmente optimizada para reducir el número de operaciones. Debido a que la función, que está entre paréntesis, es independiente del valor de i , y $M[i][k]$ es independiente de j , las 32 salidas de los filtros necesitan solo $512 + 32 \times 64 = 2560$ productos y $64 \times 7 + 32 \times 63 = 2464$ sumas, o aproximadamente 80 productos y sumas por cada salida. Sin embargo se puede mejorar sustancialmente el número de operaciones mediante una transformada discreta del coseno rápida (FDCT), o la FFT.

Podemos modificar la ecuación 4.2 utilizando la típica ecuación de convolución:

$$s_t[i] = \sum_{n=0}^{511} x[t - n] \times H_i[n] \quad (\text{Ec. 4.3})$$

Donde, $x[t]$ es una muestra de audio en un tiempo t

$H_i[n] = h[n] \cos[(2i + 1) \times (k - 16) \times \pi/64]$ con $h[n] = -C[n]$, si la parte entera de $(n/64)$ es impar, o $h[n] = C[n]$ en el caso contrario, para $n = 0$ hasta 511.

De esta forma cada subbanda del banco de filtros tiene su propio filtro pasa banda correspondiente a la respuesta al impulso $H_i[n]$. A pesar de que esta forma es muy conveniente para el análisis matemático, es claramente, una solución ineficiente para ser implementada. Una implementación directa de esta ecuación requiere $32 \times 512 = 16384$ multiplicaciones y $32 \times 511 = 16352$ sumas para obtener las 32 salidas de los filtros.

Podemos observar que en la ecuación 4.3 se modula la señal $h[n]$ mediante el producto con un coseno, de esta forma obtenemos un desplazamiento en frecuencia de la señal $h[n]$, es por este motivo que se denominan filtros

polifásicos. A pesar de que el banco de filtros polifásicos tiene pérdidas, los errores que de éste se obtienen son pequeños.

A la salida del banco de filtros polifásico, las muestras de audio se dividen por subbandas de la manera mostrada en la figura 4.5. Como se ve, cada subbanda aporta 12 muestras para un total de 384 muestras de audio, en la Capa 1, mientras que para la Capa 3, cada subbanda aporta 36 muestras de audio para un total de 1152 muestras subbanda por trama.

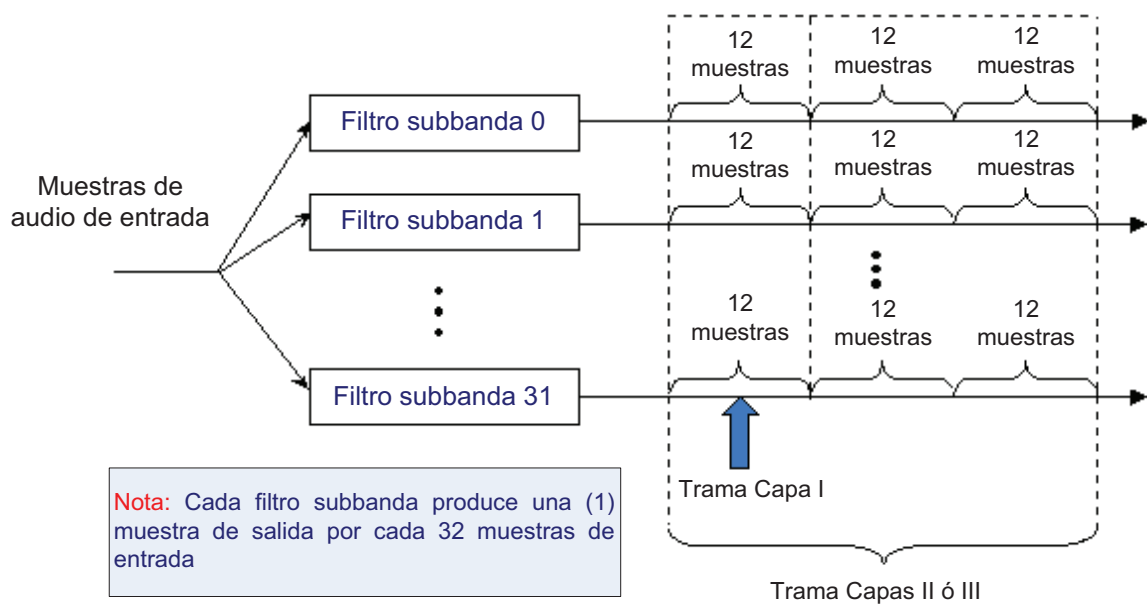


Figura 4. 5: División en subbandas de las muestras de audio

El oído tiene una limitada selectividad en frecuencia que varía desde menos de 100 Hz para las frecuencias más bajas hasta un poco más de 4 KHz, para las frecuencias más altas. El ancho de banda que proporcionan los filtros polifásicos es demasiado amplio para las bajas frecuencias, y demasiado estrecho para las altas frecuencias, así que el número de bits del cuantizador no se puede optimizar para la sensibilidad al ruido dentro de cada banda crítica. Debido a esto, lo mejor es que al espectro audible se le hagan particiones en bandas críticas, por medio de la transformada discreta del coseno modificada MDCT, que reflejen la selectividad en frecuencia del oído.

4.2.1.3.3 Transformada discreta del coseno modificada (MDCT)

La Capa 3 subdivide cada una de las 32 bandas (salidas del banco de filtros) mediante una transformada discreta coseno modificada de seis o dieciocho puntos (líneas de frecuencia) y 50% de solapamiento, con el fin de compensar la falta de precisión del banco de filtros, logrando subdividir la salida espectral en frecuencias que proporcionen mejor resolución con respecto a las bandas críticas, como se muestra en la figura 4.6.

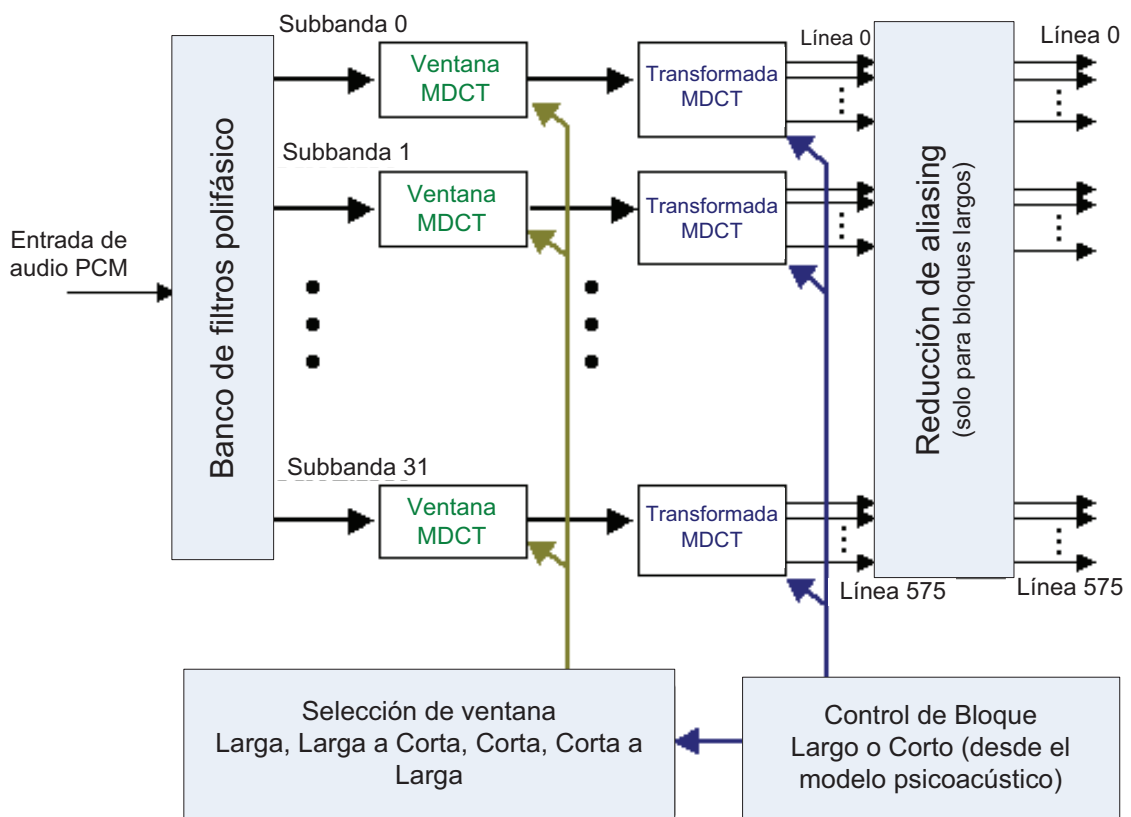


Figura 4. 6: Diagrama de bloques de las operaciones de la MDCT

Usando dieciocho puntos, el número máximo de componentes en frecuencia es: $32 \times 18 = 576$. Usando 6 puntos, la resolución en frecuencia (ancho de banda) es menor, pero la temporal es mayor, y se aplica en aquellas zonas en las que cabe esperar efectos de pre-eco (transiciones bruscas de silencios a altos niveles energéticos, como por ejemplo justo antes de un sonido de percusión). En estos casos se produce un transitorio con elevados errores de cuantización, debido a la saturación del cuantizador.

Al realizar la decodificación, el error se distribuye por toda la trama, ocasionando que las partes de silencio ya no sean silencio, sino que presenten parte de la energía de las otras regiones de la trama. Esto obliga al uso de ventanas MDCT temporales más pequeñas que limitan el efecto de pre-eco a un número menor de muestras, en comparación con el uso de ventanas grandes logrando de esta manera, reducir la distorsión. El pre-enmascaramiento temporal evita que la distorsión restante sea audible.

La Capa 3 tiene tres modos de bloque: dos modos donde las 32 salidas del banco de filtros pueden pasar a través de las ventanas y las transformadas MDCT, todas las salidas con la misma longitud de bloque. Y un modo de bloque mixto donde las dos bandas de frecuencia más bajas usan bloques largos y las 30 bandas superiores usan bloques cortos. La decisión del modo de bloque a ser usado recae sobre el modelo psicoacústico: si la señal presenta muchos transitorios se debe usar bloque corto, correspondiente a tres ventanas cortas; pero si la señal es más estacionaria, se debe usar bloque largo, correspondiente a una ventana larga. El cambio entre modos no es instantáneo; un bloque largo con una ventana de datos especializada (ventana larga a corta o, ventana corta a larga) proporciona el mecanismo de transición entre modos.

Si se ejecuta la MDCT sobre cualquiera de las ventanas largas (NORMAL, START, o STOP), se producirán 18 líneas de frecuencia debido al 50% de solapamiento. Cuando se usan las tres ventanas cortas se producirán 3 grupos, cada grupo con 6 líneas de frecuencia que pertenecen a diferentes intervalos de tiempo. El proceso de la transformada MDCT sobre cualquier tipo de bloque producirá, entonces, 576 líneas de frecuencia referidas como "Gránulo" (subdivisión de una trama).

En la figura 4.7 se muestran los cuatro tipos de ventana que se usan durante el proceso de codificación MP3:

- a) NORMAL,
- b) transición de ventana larga a corta (START),

- c) tres ventanas cortas (SHORT), y
- d) transición de ventana corta a larga (STOP).

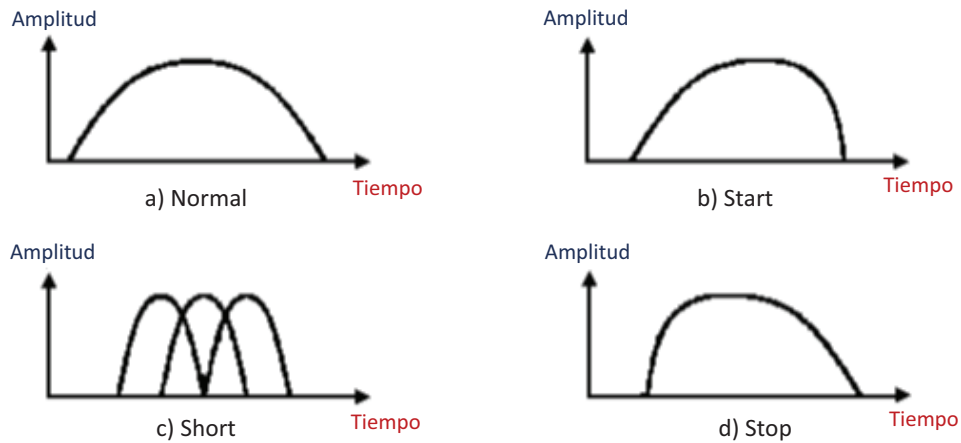


Figura 4. 7: Ventanas de datos usadas durante el proceso MP3

En resumen, el proceso que se ha mostrado es: 576 muestras PCM de entrada se convierten en 576 muestras subbanda. El solapamiento, antes de la MDCT, ocasiona que esta cantidad se duplique: en este punto son 1152 muestras subbanda, las cuales finalmente producen 576 coeficientes MDCT (líneas de frecuencia) de salida.

Antes de continuar, se realiza la reducción del aliasing introducido por el filtro análisis. Este proceso se realiza aquí, para obtener una reducción en la cantidad de información a ser codificada y transmitida.

4.2.1.4 Repartición de ruido

Mientras las capas 1 y 2 usan repartición de bits el codificador de la capa 3 usa repartición de ruido. La repartición de bits³⁹ únicamente aproxima la cantidad de ruido causado por la cuantización, mientras que la repartición de ruido verdaderamente calcula el ruido.

³⁹ En el proceso de repartición de bits, estos son asignados a las subbandas de acuerdo al nivel de enmascaramiento que resultó de aplicar el modelo psicoacústico. El objetivo de la repartición de bits es minimizar el máximo de las relaciones máscara – ruido (NMR), el máximo tomado sobre todos los canales y todas las subbandas. Donde: $NMR (dB) = SNR (dB) - SMR (dB)$.

La repartición se hace en un ciclo de iteración que consiste de un ciclo interno y uno externo.

4.2.1.4.1 Ciclo interno (rate control loop)

El ciclo interno realiza la cuantización no uniforme y escoge un determinado paso de cuantización, cuantiza los valores espectrales, y a estos datos cuantizados se les aplica codificación de Huffman. Si el número de bits resultante de la codificación excede el número de bits disponible para codificar un bloque de datos dado, de acuerdo con el bitrate escogido, se realiza una corrección ajustando la ganancia global. De este modo se logra tener un paso de cuantización más grande, dando así valores cuantizados más pequeños, entonces el ciclo comienza otra vez con un nuevo intervalo de cuantización, ejecutando la cuantización y la codificación de Huffman otra vez.

El ciclo termina cuando los valores cuantizados que han sido codificados con Huffman usan menor o igual número de bits que la máxima cantidad de bits permitida.

4.2.1.4.2 Ciclo externo (distortion control loop).

Para moldear el ruido de cuantización de acuerdo al umbral de enmascaramiento, factores de escala son aplicados a cada banda de factor de escala, que se adecúe a los límites de frecuencia variable del umbral de enmascaramiento.

El sistema comienza con un factor por defecto de 1,0 para cada banda. Si el ruido de cuantización en una determinada banda excede el umbral de enmascaramiento (ruido permitido) obtenido del modelo psicoacústico, el factor de escala para esta banda se ajusta para reducir el ruido de cuantización.

Ahora el ciclo externo se encarga de verificar si el factor de escala para cada subbanda tiene más distorsión que la permitida (ruido en la señal codificada),

comparando cada banda del factor de escala (*scalefactor band*) con los datos previamente calculados en el análisis psicoacústico. Si cualquiera de las bandas del factor de escala tiene más ruido que el máximo permitido, el ciclo amplifica esa banda de factor de escala, decrementa el tamaño del paso del cuantizador para las mismas y ejecuta ambos ciclos (el interno y el externo) de nuevo.

El ciclo externo termina cuando una de las siguientes condiciones se cumple:

- Ninguna de las bandas del factor de escala tiene una distorsión mayor a la permitida.
- La próxima iteración amplificaría cualquiera de las bandas por encima del valor máximo permitido.
- Todas las bandas han sido amplificadas al menos una vez.

Ya que el ciclo consume mucho tiempo, una aplicación en tiempo real debe tener en cuenta una cuarta condición, que detenga el ciclo evitando que la codificación se ejecute fuera de tiempo.

Dado que lograr un ruido de cuantización menor requiere un número de pasos de cuantización mayor y por tanto una bitrate más alta, el ciclo interno debe repetirse cada vez que se usen nuevos factores de escala. En otras palabras, el ciclo interno se anida dentro del ciclo externo. El ciclo externo es ejecutado hasta que el ruido real (calculado a partir de la resta entre los valores espectrales originales y los valores espectrales cuantizados) está por debajo del umbral de enmascaramiento para cada factor de escala (p.ej. banda crítica).

4.2.1.5 Flujo de bits MP3 válido

El último bloque en el proceso de codificación MP3 es el encargado de producir un flujo de bits MP3 válido. Este bloque almacena el audio codificado y algunos datos adicionales en tramas, donde cada trama contiene información de 1152 muestras de audio.

Una trama es un bloque de datos con su propio encabezado e información de audio junto con el chequeo de errores y los datos auxiliares, estos dos últimos campos son opcionales.

El encabezado describe, entre otros, cuál capa, tasa de bits y frecuencia de muestreo se están usando para el audio codificado. Los datos codificados con Huffman y su información secundaria están localizados en la parte de los datos de audio, donde la información secundaria dice qué tipo de bloque, tablas de Huffman y factores de ganancia deben ser usados.

4.2.1.6 Cuantización no uniforme

El cuantizador eleva su entrada a la $3/4$ potencia antes de cuantizar, de esta manera se busca tener una mayor consistencia de los valores de SNR (relación señal/ruido) sobre el rango de cuantización. El decodificador realiza el proceso inverso.

La ecuación completa para el cuantizador es:

$$x_q[i] = Rnd \left[\left(\frac{x[i]}{2^{\Delta/4}} \right)^{3/4} - 0.0946 \right] \quad (\text{Ec. 4.4})$$

Donde, Δ es el paso de cuantización
 $x[i]$ es la señal a cuantizar

El máximo valor de cuantización permitido tiene como fin limitar el tamaño de las tablas usadas para la búsqueda del decodificador.

4.2.1.7 Codificación Huffman (codificación entrópica)

El formato MP3 emplea la técnica del algoritmo de Huffman que es un algoritmo de codificación (no de compresión), que tiene su efecto al final del proceso de

c) Repetir los pasos 1 y 2 hasta que se incluyan todos los símbolos.



Figura 4. 9: Algoritmo Huffman

Cuando se realiza la decodificación, la tabla de probabilidad debe ser recuperada primero. Para saber cuándo cada representación de un símbolo termina simplemente se debe seguir el árbol desde la raíz hasta encontrarnos en un símbolo. Esto es posible porque ninguna codificación es un subconjunto de otra (prefijo de codificación).

4.2.1.8 Reserva de bits

En la Capa 3, las tramas no son siempre totalmente independientes, debido al posible uso del *bit reservoir*, que es una especie de buffer, las tramas son a menudo dependientes unas de otras. En el peor caso, se pueden necesitar hasta nueve tramas, antes de poder realizar la decodificación de una sola.

La figura 4.10 muestra con un ejemplo este concepto.

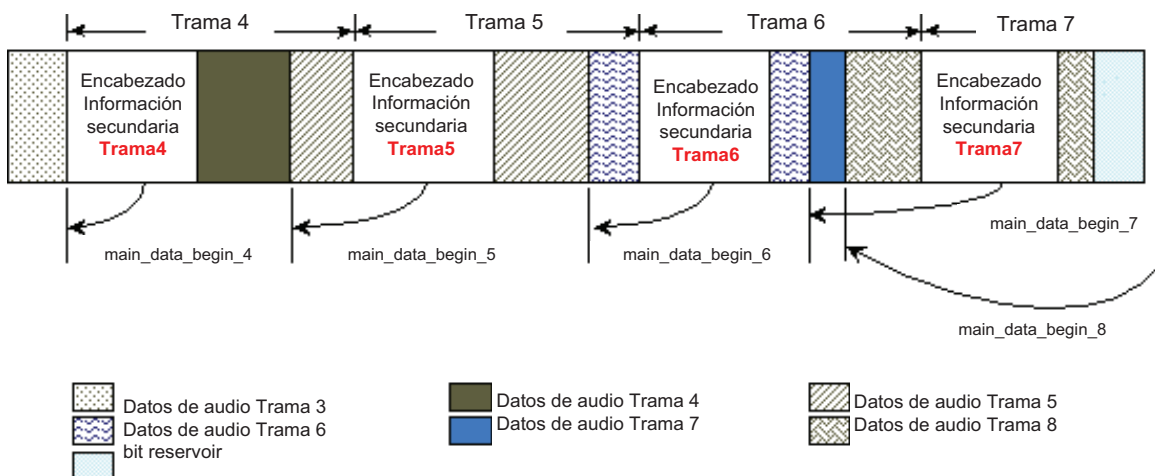


Figura 4. 10: Ejemplo de uso de la reserva de bits (bit reservoir)

El *main_data_begin* es un puntero (de nueve bits) de ajuste negativo, incluido dentro de la información secundaria, que apunta a la posición del byte de inicio de la información de audio dentro de cada trama. Por ejemplo, *main_data_begin_4* es igual a cero, indicando que los datos de audio empiezan inmediatamente después de la información secundaria. Para indicar que el audio de la trama 5 se inicia en la trama 4, se especifica *main_data_begin_5* como un ajuste negativo que indica el desplazamiento en bytes hacia la izquierda para encontrar el primer dato de audio de la trama 5.

En el ejemplo se ve como cada trama permite el uso del bit reservoir. En el caso de la trama 7, el proceso empieza codificando la información de audio de su propia trama, como los datos requieren muy pocos bits, y la trama 6 tenía espacio disponible, entonces todos los datos de audio de la trama 7 se incluyen en la trama 6, pero la trama 6 sigue con espacio para bit reservoir, que se usa para datos de la trama 8; por lo que gracias al bit reservoir, la trama 6 incluye los datos de audio de tres tramas: las tramas 6, 7 y 8. El audio de la trama 8 se reparte entre las tramas 6 y 7; sin embargo, éste no alcanza a ocupar todo el espacio disponible en la trama 7, así que el bit reservoir de la trama 7 se usa para la trama 9, y así sucesivamente, teniendo en cuenta que los datos de audio de una determinada trama no pueden estar desplazados más de nueve tramas.

Este caso puede ocurrir en una señal de audio MPEG-1 estéreo, si la frecuencia de muestreo es 48 KHz y la tasa de transferencia deseada es 32 Kbps. En este caso, cada trama consume 768 bits, donde 304 bits (32 bits para el encabezado, 16 bits para el chequeo de errores, 256 bits para la información secundaria) son fijos. Por lo tanto, quedan 464 bits disponibles para los datos codificados con Huffman, y debido a que el valor de *main_data_begin* puede apuntar máximo 511 bytes (4088 bits) hacia atrás, entonces es posible que *main_data_begin* apunte sobre más de ocho tramas (no se cuenta ninguno de los bits usados para el encabezado y la información secundaria de ninguna trama).

También es importante mencionar que el bit reservoir sólo puede originarse de tramas que ya han sido codificadas; para este buffer no es posible usar tramas

para las que todavía no se haya hecho la repartición de los bits disponibles (repartición de ruido).

4.2.1.9 Modos de funcionamiento

Hay cuatro modos de funcionamiento para el esquema III de MPEG-1 (MP3):

- **Single Channel** o canal único: una señal en un bitstream.
- **Dual Channel** o canal doble: dos señales independientes en un mismo bitstream.
- **Stereo**: como el anterior, perteneciendo las señales al canal izquierdo y derecho de una señal estéreo original.
- **Joint Stereo**: como el anterior, aprovechando ciertas características del estéreo como irrelevancia y redundancia de datos para reducir la tasa de bits.

4.2.1.9.1 Codificación joint stereo

La mayoría de los sistemas de alta fidelidad utilizan un "boomer" (que reproduce los graves). Sin embargo, para el oído humano parece que el audio no proviene del boomer sino de los otros altavoces. Por debajo de una determinada frecuencia, el oído humano no puede distinguir de dónde provienen los sonidos.

El formato MP3 puede, opcionalmente, tomar ventaja de este fenómeno utilizando el modo *joint stereo*. El modo joint stereo aprovecha las semejanzas de audio entre los canales izquierdo y derecho, descartando la información idéntica que se encuentra en ambos canales. Esto significa que ciertas frecuencias se graban en modo monofónico, pero incluyen datos adicionales para que suenen como una configuración de varios altavoces.

En muchos casos un resultado satisfactorio es adquirido con este proceso. La mayoría de los códecs utilizan el modo joint stereo cuando codifican archivos de audio a una tasa de bits de 128 Kbps.

4.2.2 DECODIFICACIÓN

El decodificador debe procesar el bitstream para reconstruir la señal de audio digital. La figura 4.11 ilustra el esquema simplificado del decodificador.

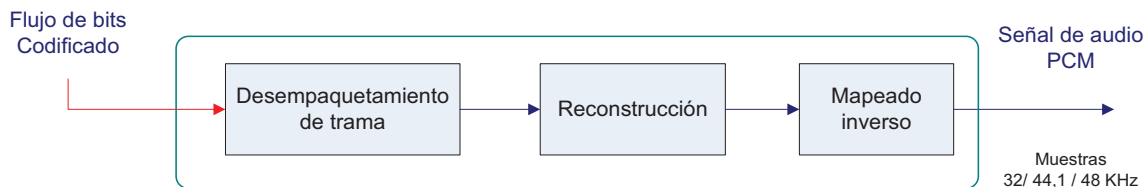


Figura 4. 11: Decodificador según la norma ISO 11172-3

Los datos del bitstream son desempaquetados para recuperar las diversas partes de la información. El bloque de reconstrucción recompone la versión cuantizada de la serie de muestras mapeadas. El mapeador inverso transforma estas muestras de nuevo a PCM (Pulse-Code Modulation).

4.3 FORMATO DE LAS TRAMAS MP3

La trama MP3 está constituida por las muestras de audio y por la información secundaria. Esta última sirve de control, además de proporcionar información del archivo.

Para la capa 3, la trama está compuesta por *1152 muestras de audio más la información de la trama*. Adicionalmente, cada trama es subdividida en dos partes (conocidas como gránulos), cada una con 576 muestras. Dado que la tasa de bits determina el tamaño de cada muestra, el aumento de la tasa de bits también aumentará el tamaño de la trama.

El formato que tiene cada trama MP3 se muestra en la figura 4.12, en la cual se puede ver el encabezado de trama que posee 32 bits (cuatro bytes) de longitud; los primeros 12 bits siempre se ponen en uno '1', se llaman "FRAME SYNC", y se usan para sincronización de la trama.

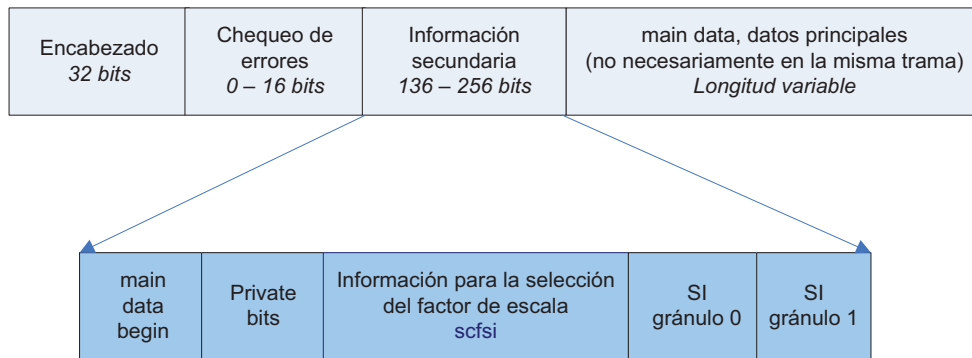


Figura 4. 12: Formato de la trama MP3

Las tramas pueden tener opcionalmente un CRC para chequeo de errores. Su longitud es de 16 bits, y si existe, se pone después del encabezado. Volviendo a calcular el CRC se puede comprobar si la trama ha sido alterada durante la transmisión del flujo de bits de audio MP3.

A continuación sigue la información secundaria (*Side Information*) que indica cómo se realizó la codificación, y por lo tanto, cómo debe realizarse la decodificación.

En el último bloque vienen los datos de audio (*main data*), repartidos entre dos gránulos. Dentro de la información secundaria, que usa 136 bits en modo monofónico y 256 bits en los otros modos, se incluye el *main_data_begin*. Los bits privados están a disposición del usuario. Después viene la información que indica cuál combinación de factores de escala se está usando (*scfsi*, *scalefactor selection information*). Los últimos dos sub-bloques corresponden a la información secundaria (*Side Info*, *SI*) para cada uno de los dos gránulos (sub-tramas) en los que se divide una trama.

El último bloque, main data, es el que lleva la información de audio; las muestras MDCT codificadas con Huffman, repartidas entre dos gránulos. Cada gránulo contiene información de 576 muestras de audio (exactamente la mitad de la información total de la trama). Además, en este mismo bloque se incluyen los factores de escala de la trama y la información auxiliar, siendo esta última opcional.

4.3.1 ENCABEZADO DE TRAMAS

No existe un encabezado principal de archivo en el formato de audio MPEG. En éste el encabezado es individual para cada trama (fragmento de archivo). Cuando se quiere leer información de un archivo MP3, usualmente es suficiente encontrar la primera trama, leer su encabezado y asumir que las otras tramas son iguales. Pero éste no es siempre el caso; por ejemplo, existen algunos archivos con bitrates variables, donde cada trama posee su propia bitrate. Esto se hace con el fin de mantener constante la calidad del sonido durante todo el archivo. Otro método usado para mantener constante la calidad de sonido es emplear más bits (con ayuda del buffer *bit reservoir*) en las partes donde se necesite.

El encabezado de la trama tiene la siguiente presentación, figura 4.13, con las posiciones para cada uno de los 32 bits.

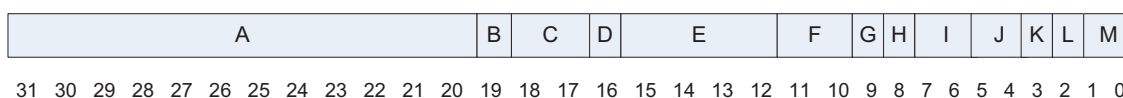


Figura 4. 13: Encabezado de la trama MP3

A: **Syncword**. Con 12 bits de longitud, todos en uno '1111 1111 1111' para identificar el comienzo de la trama.

B: **ID**. Un bit usado para identificación del audio. Siempre en uno '1', para indicar que se trata de audio MPEG-1.

C: **Layer**. Dos bits usados para la descripción de la capa. Para identificar cuál esquema fue usado durante la codificación del audio, ver tabla 4.2.

00	Reservado
01	Capa III
10	Capa II
11	Capa I

Tabla 4. 2: Valores posibles para los bits 18 y 17 para indicar cuál capa de MPEG-1 se utiliza

D: **protection_bit**. Un bit de protección. Si está en cero '0' indica que la trama está protegida por un código de redundancia cíclica para detección de errores. En la mayoría de los archivos MP3 este bit es uno '1' para indicar que no hay redundancia.

E: **bit_rate_index**. Cuatro bits para proporcionar el índice de la tasa de bits, de acuerdo con la tabla 4.3. Además indica la tasa de bits total independientemente del modo utilizado (stereo, joint stereo, dual channel o single channel).

Código bit_rate_index	Tasa de bits Capa 3
0000	formato libre
0001	32 Kbps
0010	40 Kbps
0011	48 Kbps
0100	56 Kbps
0101	64 Kbps
0110	80 Kbps
0111	96 Kbps
1000	112 Kbps
1001	128 Kbps
1010	160 Kbps
1011	192 Kbps
1100	224 Kbps
1101	256 Kbps
1110	320 Kbps

Nota: Si la trama usa formato libre (una tasa de bits diferente a las listadas), la tasa debe permanecer constante, y debe ser menor a la máxima tasa de bits permitida que es 320 Kbps.

Tabla 4. 3: Codificación para la tasa de bits

F: **sampling_frequency**. Dos bits que indican la frecuencia de muestreo, de acuerdo a la siguiente tabla.

Código	Frecuencia de muestreo
00	44,1 KHz
01	48 KHz
10	32 KHz
11	Reservado

Tabla 4. 4: Codificación para la frecuencia de muestreo

G: **padding_bit**. Un bit usado para relleno. Únicamente se usa para frecuencias de 44,1 KHz. Si se usan tramas de 417 bytes de largo no se logra la tasa de transferencia de 128 Kbps. Para solucionar esto, éste bit se pone en uno '1' y se agrega un byte extra al final de esas tramas para así obtener 128 Kbps.

H: **private_bit**. Un bit para uso privado. Generalmente no se usa.

I: **mode**. Dos bits que indican el modo de canal, tal y como se muestra en la tabla 4.5.

Código	Modo de canal
00	Stereo
01	Joint Stereo
10	Dual Channel (2 canales monofónicos independientes) (<i>intensity_stereo</i> y/o <i>ms_stereo</i>)
11	Single Channel (1 canal monofónico)

Tabla 4. 5: Codificación para el modo de canal

El modo *Stereo* indica que el canal comparte bits, pero no usa codificación Joint Stereo.

En el modo *Joint Stereo* sí se saca provecho de la correlación existente entre los dos canales para representar más eficientemente la señal.

El modo *Dual Channel* está conformado por dos canales monofónicos totalmente independientes (cada uno es un archivo de audio diferente); cada canal usa exactamente media tasa de bits del archivo. La mayoría de los decodificadores los procesan como estéreo, pero no es siempre el caso.

El modo *Single Channel* consiste en un único canal de audio.

J: **mode_extension**. Dos bits indicando extensión al modo; sólo se usa en modo Joint Stereo. La extensión al modo se usa para información que no es de ninguna utilidad en el efecto estéreo. Estos bits se determinan dinámicamente por un

codificador en el modo Joint Stereo, y este modo puede cambiar entre tramas, o incluso se puede dejar de usar en algunas tramas.

En la Capa 3, estos dos bits indican qué tipo de codificación Joint Stereo se está usando, intensidad estéreo o estéreo M/S (*intensity_stereo* o *ms_stereo*).

Estéreo M/S se refiere a transmitir los canales normalizados Middle/Side (Suma/Diferencia) de los canales izquierdo y derecho en lugar de los habituales Izquierdo/Derecho. En el lado del codificador los canales habituales se reemplazan usando la ecuación 4.5.

$$M_i = \frac{\sqrt{2}}{2}(L_i + R_i) \quad \text{y} \quad S_i = \frac{\sqrt{2}}{2}(L_i - R_i) \quad (\text{Ec. 4.5})$$

Donde, M_i es el canal normalizado Middle (Suma)

S_i es el canal normalizado Side (Diferencia)

L_i es el canal Izquierdo

R_i es el canal Derecho

Los valores M_i se transmiten por el canal izquierdo y los valores S_i se transmiten por el canal derecho.

En el lado del decodificador los canales izquierdo y derecho se reconstruyen de la siguiente forma:

$$L_i = \frac{M_i + S_i}{\sqrt{2}} \quad \text{y} \quad R_i = \frac{M_i - S_i}{\sqrt{2}} \quad (\text{Ec. 4.6})$$

Intensidad estéreo (*intensity_stereo*) se refiere a retener en las frecuencias superiores a 2 KHz sólo la envolvente de los canales izquierdo y derecho.

El código indica que tipo de extensión al modo se está usando, como se indica a continuación en la tabla 4.6.

Código	<i>intensity_stereo</i>	<i>ms_stereo</i>
00	No	no
01	Si	no
10	No	si
11	Si	si

Tabla 4. 6: Codificación de la extensión al modo

K: **copyright**. Un bit usado para copyright. Tiene el mismo significado que el bit de copyright en CD's y cintas DAT⁴⁰ (Digital Audio Tape), indica que es ilegal copiar el contenido del archivo, ya que está protegido por derechos de autor cuando este bit está en '1'. Mientras que si este bit es igual a cero '0' no hay derechos de autor en el flujo de bits codificados.

L: **original/copy**. Un bit usado para indicar si se trata de un medio original, si el bit está colocado en uno '1'. En cero '0' indica que es una copia del medio original.

M: **emphasis**. Dos bits usados para información del énfasis. Le indica al decodificador que el sonido debe ser "re-ecualizado" después de una supresión de ruido tipo Dolby. Generalmente no se utiliza.

Código	<i>Énfasis</i>
00	No énfasis
01	50/15 [µs] de énfasis
10	Reservado
11	CCITT J.17

Tabla 4. 7: Codificación de la información de énfasis

4.3.2 CHEQUEO DE ERRORES

Si el bit de protección en el encabezado es igual a cero '0', se incluye un CRC de 16 bits después del encabezado. Si el bit de protección está en uno '1', no hay

⁴⁰ Las cintas DAT se crearon como formato de audio con calidad CD, pero en 1998 SONY y HP definieron el estándar para el almacenamiento de datos digitales, aplicando esta tecnología al almacenamiento de datos. Las cintas DAT se encuentran en formato DDS y DataDAT. Las frecuencias de muestreo que puede utilizar el DAT son 32, 44,1 y 48 KHz. La velocidad de transferencia de datos del DAT son 2,77 Mbps.

chequeo de errores y estos bits pueden ser usados para los datos de audio. El método para detección de errores que se utiliza es CRC-16, cuyo polinomio generador es:

$$CRC - 16 = x^{16} + x^{15} + x^2 + 1 \quad (\text{Ec. 4.7})$$

Los códigos de redundancia cíclica forman una subclase de los códigos bloque usados en esquemas de transmisión de bloques (frames) y se los utiliza para la detección de errores. Son códigos cíclicos porque el desplazamiento cíclico de una palabra código válida genera otra palabra válida. Es así que, dado un bloque de m bits de información, el transmisor genera una secuencia de k bits de control, conocida como la FCS (Frame Check Sequence); tal que la trama resultante consistente de $(m+k)$ bits, sea exactamente divisible para una secuencia denominada *secuencia generadora* de $(k+1)$ bits [25].

El algoritmo para calcular la suma de verificación en un CRC-16, es el siguiente:

- Se anexan 16 bits cero al final de la trama, para que ahora contenga $m+16$ bits la trama resultante.
- Se divide la cadena de bits correspondiente al polinomio generador $x^{16} + x^{15} + x^2 + 1 \Rightarrow$ '11000000000000101' (secuencia generadora) entre la trama resultante usando una división módulo 2.
- Se resta el residuo (que siempre es de 16 bits o menos bits) a la cadena de bits correspondiente a la trama resultante de $m+16$ bits usando una resta módulo 2. El resultado es la trama con suma de verificación que va a transmitirse.

En recepción se divide la trama con suma de verificación recibida para la secuencia generadora y:

- Si el residuo = 0 \Rightarrow \nexists error en la transmisión
- Si el residuo \neq 0 \Rightarrow \exists error en la transmisión

4.3.3 INFORMACIÓN SECUNDARIA

Éste campo consta de 17 bytes para el modo monofónico, y de 32 bytes en cualquier otro modo.

La información que contiene, consiste de cuatro partes: el puntero *main_data_begin*, información secundaria para ambos gránulos (*private_bits* y *scfsi*), información secundaria para el gránulo 0, e información secundaria del gránulo 1, como se muestra en la figura 4.14.

main_data_begin (9)	private_bits (5,3)	scfsi (4,8)	SI gránulo 0 (59,118)	SI gránulo 1 (59,118)
------------------------	-----------------------	----------------	--------------------------	--------------------------

Figura 4. 14: Formato de la información secundaria

main_data_begin: el campo *main_data* es el bloque que contiene los datos de audio, repartidos entre dos gránulos y no está necesariamente localizado justo después de la información secundaria.

main_data_begin es un puntero que usa 9 bits, indicando la localización donde está el primer byte del *main_data* de la trama actual. La localización está especificada como un desplazamiento negativo en bytes desde el encabezado actual (bytes a la izquierda, antes del primer bit del encabezado).

La información secundaria (SI) común a ambos gránulos se muestra a continuación:

private_bits: El número de *private_bits* para la información secundaria depende del número de canales (5 para mono y 3 para estéreo). El número de bits reservados para *private_bits* es definido por el usuario.

scfsi: El campo *scfsi* (información para selección del factor de escala) determina si los factores de escala se envían para cada gránulo, o si son comunes para ambos gránulos, por canal.

Se transmiten cuatro bits por canal, cada bit perteneciente a un grupo de bandas del factor de escala diferente. Un cero '0' para un grupo específico de bandas del factor de escala, indica que los factores de escala para ese grupo en particular, se transmiten para cada gránulo. Un uno '1' indica que se usan los mismos factores de escala para ambos grupos; por lo tanto, sólo se transmiten los factores de escala correspondientes al grupo de bandas del primer gránulo.

Después de la información secundaria para ambos gránulos, sigue la información secundaria para cada gránulo, como se muestra en la figura 4.15.

part2_3_length (12,24)	big_values (9,18)	global_gain (8,16)	scalefac_compress (4,8)	window_switching_flag (1,2)
---------------------------	----------------------	-----------------------	----------------------------	--------------------------------

a)

block_type (2,4)	mixed_block_flag (1,2)	table_select (10,20)	subblock_gain (9,18)
---------------------	---------------------------	-------------------------	-------------------------

b)

table_select (15,30)	region0_count (4,8)	region1_count (3,6)
-------------------------	------------------------	------------------------

c)

preflag (1,2)	scalefac_scale (1,2)	count1table_select (1,2)
------------------	-------------------------	-----------------------------

d)

Figura 4. 15: Información secundaria para cada gránulo

En el caso de bloques largos, la información secundaria para cada gránulo es:

part2_3_length: Denota el número de bits que son usados en *main_data* para los factores de escala y los datos codificados con Huffman. Se usan 12 bits en modo mono y 24 bits en los otros modos. Como la cantidad de bits usados para la información secundaria es constante, *part2_3_length* puede usarse para calcular el comienzo del próximo gránulo.

big_values: Después de la cuantización, las 576 muestras MDCT cuantizadas están organizadas en un orden determinado (de menor a mayor frecuencia).

Luego, estos valores se dividen en tres particiones consecutivas: ***rzero***, ***count1*** y ***big_values***.

La primera partición, *rzero*, se localiza en las altas frecuencias y consiste en pares de ceros.

La segunda partición, *count1*, consiste de cuádruplos cuyo valor es -1, 0 ó +1.

La última partición, *big_values*, se localiza en las bajas frecuencias extendiéndose hasta el nivel de frecuencia de 0 Hz y se compone de pares de valores restringidos a una amplitud máxima absoluta de 8206 ($8191+15$, el cual es el máximo valor cuantizado permitido). El campo *big_values* indica la cantidad de pares cuantizados que pertenecen a esta partición. Nueve bits se usan para *big_values* en modo monofónico y 18 bits en los otros modos.

global_gain: Contiene información acerca del intervalo usado en el cuantizador, donde la cuantización se hace logarítmicamente. La variable *global_gain* usa 8 bits en modo mono y 16 bits para los otros modos.

scalefac_compress: Es una variable de 4 bits en modo monofónico y de 8 bits en los otros modos, transmitida para cada gránulo, la cual determina el número de bits usados para la transmisión de los factores de escala. Cada gránulo se divide en 12 ó 21 bandas del factor de escala dependiendo del tipo de ventana que se esté usando.

Estas bandas del factor de escala se dividen de nuevo en dos grupos (0-10 y 11-20 para ventanas largas; 0-5 y 6-11 en el caso de ventanas cortas).

La variable *scalefac_compress* se usa como índice a una tabla proporcionada en el estándar ISO/IEC 11172-3 [30], la cual retorna dos variables llamadas "*slen1*" y "*slen2*", que indican la cantidad de bits usados para los factores de escala del primer y segundo grupo de bandas, respectivamente.

window_switching_flag: Un bit por canal que señala si una ventana diferente del tipo NORMAL se está usando. Este valor determina los siguientes 22 bits en la información secundaria: si está en uno '1', se añaden los bits de la figura 4.15 (b); si está en cero '0', se añaden los bits de la figura 4.15 (c).

table_select: Habilita el uso de 32 diferentes tablas para el código de Huffman [30], dependiendo de las estadísticas de la señal. Se usan 15 bits por canal (5 bits por región) para indicar cuáles de las 32 tablas han sido seleccionadas.

region0_count: Para mejorar el desempeño en la codificación, la partición *big_values* se subdivide en tres regiones llamadas *region0*, *region1* y *region2*.

Cada región se codifica con una de las 32 tablas de Huffman (seleccionada con *table_select*). La variable *region0_count* especifica el límite entre la *region0* y *region1*. Esta variable de 4 bits en modo monofónico especifica la cantidad de bandas del factor de escala incluidas en esta *region0*, pero disminuidas en 1.

$$region0_count = \text{bandas del factor de escala en } region0 - 1 \quad (\text{Ec. 4.8})$$

region1_count: Especifica el límite entre la *region1* y *region2*. Esta variable de 3 bits por canal indica las bandas del factor de escala incluidas en *region1*, disminuidas en 1.

$$region1_count = \text{bandas del factor de escala en } region1 - 1 \quad (\text{Ec. 4.9})$$

preflag: Un bit por canal, indicando que se usó preénfasis (es decir, amplificación adicional en las altas frecuencias). Este valor apunta a una tabla en el estándar ISO/IEC 11172-3, cuyos 21 valores son sumados a los factores de escala. Para bloques cortos, no se usa preénfasis.

scalefac_scale: Los factores de escala están cuantizados de manera logarítmica con un intervalo de 2 ó $(2)^{1/2}$, dependiendo del valor de *scalefac_scale*, que usa 1 bit por canal.

count1table_select: Esta variable, que usa 1 bit por canal, indica cuál de dos posibles tablas de Huffman [30] fue usada para codificar la partición count1.

En el caso de bloques cortos, la información secundaria sólo cambia en las variables mostradas en la figura 4.15 (c), las cuales son reemplazadas por aquellas de la figura 4.15 (b). Las otras variables mostradas en la figura 4.15 no cambian.

block_type: Indica el tipo de ventana que se usa en un gránulo particular. La variable *block_type* usa 2 bits por canal.

mixed_block_flag: Esta variable, que usa 1 bit por canal, indica que se usan diferentes tipos de ventana en las bajas y en las altas frecuencias. Si esta variable está en uno '1', las dos subbandas más bajas usan ventana NORMAL, y las 30 subbandas restantes usan el tipo de ventana especificado por *block_type*.

table_select: En este caso, *table_select* usa 10 bits por canal, debido a que, para bloques cortos, la partición *big_values* sólo se subdivide en dos regiones.

subblock_gain: Habilita una ganancia por un factor de 4 para un sub-bloque particular. Esta variable se usa sólo con bloques tipo 2 (ventanas cortas). Esta variable usa 3 bits por canal.

4.3.4 DATOS PRINCIPALES

En esta parte del flujo de bits de la Capa 3, están incluidos los campos mostrados en la figura 4.16:

Factores de escala <i>longitud variable</i>	Código de Huffman <i>longitud variable</i>	Datos auxiliares <i>longitud variable</i>
--	---	--

Figura 4. 16: Campos incluidos en los datos principales

Factores de escala: Éstos se usan para colorear el ruido⁴¹ de cuantización. Los factores de escala se transmiten para cada grupo de líneas de frecuencia (bandas del factor de escala) de cada gránulo, dependiendo del valor de *scfsi* para ese grupo particular de líneas de frecuencia. La cantidad de factores de escala realmente transmitidos, también depende de *block_type*, *window_switching_flag* y *mixed_block_type*.

Los factores de escala consumen entre 0 y 74 bits.

Código de Huffman: Las líneas de frecuencia de cada gránulo se dividen en tres particiones (*rzero*, *count1* y *big_values*).

La partición *rzero* no se codifica, ya que sólo contiene valores iguales a cero.

La partición *count1* contiene cuádruplos de valores iguales a -1, 0 ó +1, que se codifican usando una de las 2 posibles tablas de Huffman, la cual ha sido especificada por *coun1table_select*. Para cada valor diferente de cero, se agrega un bit que indica el signo (cero '0' si es positivo).

La partición *big_values* fue subdividida en tres regiones, las cuales se codifican separadamente, usando una de 32 posibles tablas de Huffman (numeradas de 0 a 31, pero en realidad son 30, ya que las tablas 4 y 14 no existen), o sea, una tabla por región.

Dentro de la partición *big_values*, los pares de líneas de frecuencia con valor absoluto menor que 15, se codifican directamente. Para cada valor absoluto mayor o igual a 15, se agregan 1 ó 2 campos extras llamados "*linbitsx*" o "*linbitsy*" dependiendo de cuál es el valor del par (x, y) que es mayor o igual a 15. Este campo extra usa de 0 a 13 bits, dependiendo del parámetro "*linbits*", el cual se calcula con base en el valor máximo de la región, ver ecuación 4.9.

⁴¹ Cuando la gráfica de la densidad espectral de potencia (PSD, Power Spectral Density) no es plana, entonces, se dice que el ruido está "coloreado". Dependiendo de la forma que tenga la gráfica de la PSD del ruido se definen diferentes colores.

$$\text{Linbits} = \log_2(\text{m\u00e1ximo valor cuantizado} - 14) \quad (\text{Ec. 4.9})$$

Donde, se realiza un redondeo por exceso al resultado, es decir, el n\u00famero decimal se aproxima al entero mayor.

De nuevo, para cada valor diferente de cero, se agrega un bit de signo (cero '0' si es positivo).

Por ejemplo: As\u00famase, primero que la tabla de Huffman ya ha sido seleccionada, y tambi\u00e9n los siguientes valores:

Par de valores cuantizados $(x,y) = (0,15)$

M\u00e1ximo valor cuantizado de la regi\u00f3n = 1039

C\u00f3digo de Huffman para el par $(0,15) = '01101'$

Valor adicional para 'y' = $\text{linbitsy} = 15 - 15 = 0$

$\text{linbits} = \log_2(1039 - 14) \cong 10,0014 \Rightarrow \text{linbits} = 11$

$\text{linbitsy} = 15 - 15 = 0 = '00000000000'$

Codificaci\u00f3n del par $(0,15) = \text{Codificaci\u00f3n del par } (0,15) + \text{linbitsy}$

Codificaci\u00f3n del par $(0,15) = '01101'00000000000'$

Bits necesarios para codificar el par $(0,15) = 16$ bits

(x,y) 5 bits	linbitsx 0 bits	linbitsy 11 bits	signx 0 bits	signy 1 bits
Flujo de bits '01101' '00000000000' '0'				

Figura 4. 17: Ejemplo ilustrativo

En el caso de que 'x' tambi\u00e9n sea mayor que 14, se debe buscar el c\u00f3digo de Huffman para el par $(15,15)$, y adem\u00e1s, se debe codificar un valor adicional llamado "*linbitsx*", que indica la diferencia entre 15 (m\u00e1ximo valor de las tablas) y el valor verdadero de 'x'.

Adicionalmente, por cada valor diferente de cero se debe agregar un bit de signo (cero '0' si es positivo, uno '1' si es negativo). En el ejemplo, la cantidad total de

bits que se necesita para codificar el par es 17 bits, ya que se debe agregar un bit para indicar que 'y' es diferente de cero.

Datos auxiliares: Éstos son opcionales, y la cantidad de bits repartidos para este campo, se define por el usuario.

4.4 CURIOSIDADES DEL FORMATO MP3

Las investigaciones sobre el formato MP3 empezaron en 1987 y a solo 2 días de lanzar la primera versión del MP3 en 1991 el proyecto estuvo a punto de morir, ya que en las modificaciones la codificación no funcionaba correctamente. Ya en 1992 fue adoptado el estándar MPEG y junto a éste el MP3, pero la capacidad de los procesadores en los ordenadores de la época no proporcionaban la potencia necesaria. Hoy 18 años después, el MP3 es más que un algoritmo, es el símbolo del fenómeno sociológico del compartir en internet, llegando a convertirse en la palabra más empleada dentro de los buscadores web.

Para la reproducción de archivos MP3 existen multitud de programas específicos en su mayoría shareware⁴², aunque algunos son incluso freeware⁴³.

Un año después de que el formato MP3 fue desarrollado en el Instituto Fraunhofer en Alemania por concepto de patentes ingresaban 1,2 millones de euros. Diez años más tarde esta cantidad alcanzó los 26,1 millones y sigue en aumento.

Si bien los primeros codificadores⁴⁴ MP3 estuvieron disponibles desde 1994, al principio eran pocos los que poseían este tipo de software y el hardware suficiente

⁴² Se denomina *shareware* a una modalidad de distribución de software, en la que el usuario puede evaluar de forma gratuita el producto, pero con limitaciones en el tiempo de uso o en algunas de las formas de uso o con restricciones en las capacidades finales.

⁴³ El término *freeware* define un tipo de software que se distribuye sin costo, disponible para su uso y por tiempo ilimitado, siendo una variante gratuita del shareware, en el que la meta es lograr que un usuario pruebe el producto durante un tiempo limitado (con funcionalidad limitada o no), y si le satisface, pague por él, habilitando toda su funcionalidad.

para utilizarlo con eficiencia; los primeros intercambios de archivos MP3 se hicieron a través de FTP (File Transfer Protocol), un servicio de internet no demasiado popular en esa época.

En Mayo de 1999, un estudiante universitario de la Northeastern University de Boston llamado Shawn Fanning presentó en la web un servicio gratuito de intercambio de archivos MP3 llamado *Napster*, basado en la tecnología "Peer-To-Peer" (P2P). Este sistema permite a los usuarios conectarse entre sí y formar un gigantesco depósito de MP3 en el que todos pueden buscar cualquier tipo de música disponible y bajarla a su disco.

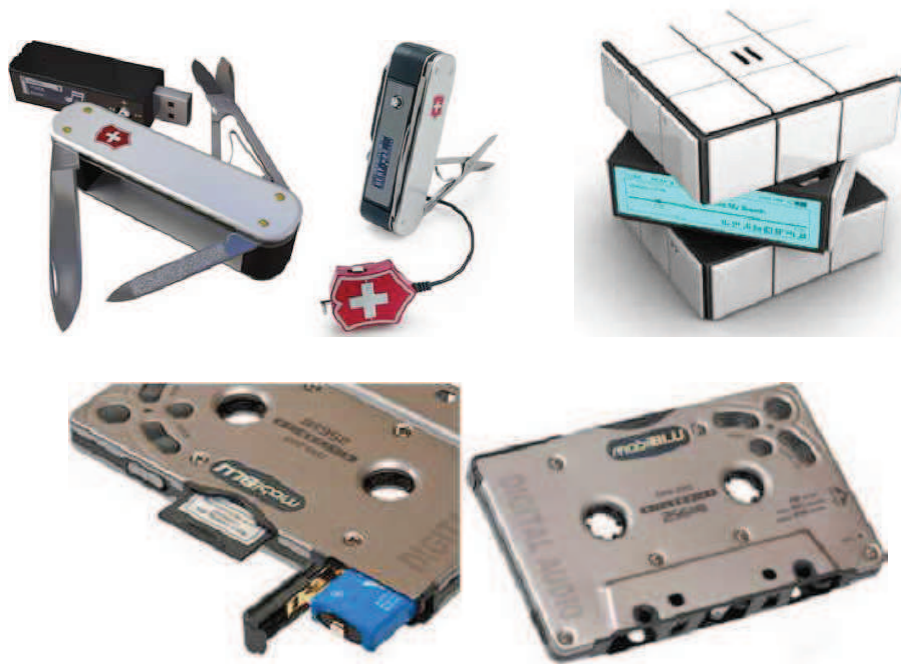
Al principio el formato MP3 tenía el limitante de funcionar en las computadoras por medio de un software reproductor de MP3, pero gracias a la creciente popularidad que este nuevo formato ha presentado ya existen en el mercado cientos de productos portátiles para la reproducción de estos archivos.

En la figura 4.18 se muestra una recopilación de los reproductores MP3 más curiosos y originales que se pueden encontrar.



a) *Tenis para correr MP3, se puede hacer ejercicio mientras se disfruta de la música favorita, ya sea con o sin audífonos.*

⁴⁴ Codificadores son programas compresores que codifican los datos de acuerdo a unas rutinas de codificación específicas y que se usan principalmente para convertir un fichero WAV a un archivo MP3, cuyo tamaño es notablemente más reducido.



- b) *Cassette normal MP3, se introduce una tarjeta SD o una memoria flash, y funciona como un cassette normal.*



- c) *Reproductor de música MP3 considerado como el más caro del mundo, está valorado en 25 mil euros, realizado en oro de 18 kilates y con 63 diamantes incrustados de 1 Kilate.*

Figura 4. 18: Curiosos reproductores mp3 [38]

4.5 INNOVACIONES EN EL FORMATO MP3

4.5.1 FORMATO MP3 DE ALTA CALIDAD

MP3HD, o de alta definición, es un nuevo formato de compresión de audio que permite comprimir matemáticamente el audio sin causar pérdida en la calidad

(lossless) que mantiene la misma extensión .mp3 para sus archivos y que es compatible con los reproductores de MP3 clásicos.

El formato MP3HD maneja tasas de transferencia entre 500 y 900 Kbps, según el tipo de música. El tamaño final de los archivos es aproximadamente cuatro veces mayor que el de un MP3 clásico, pero con la ventaja de ofrecer una mejor calidad.

4.5.2 MP3PRO

MP3PRO es la combinación del estándar MP3 y una nueva tecnología de codificación de audio llamada Spectral Band Replication (SBR), que mejora la eficiencia y fidelidad de los códecs de audio existentes.

SBR fue desarrollado por Coding Technologies y aplicado en conjunto con Thomson Multimedia y Fraunhofer IIS.

Los archivos MP3PRO tienen la extensión .mp3 al igual que los MP3 convencionales, aunque internamente son diferentes. Para bitrates bajos, el rendimiento del formato MP3PRO es significativamente superior al del MP3. Un archivo MP3PRO a 64 Kbps suena mejor que un MP3 a 96 Kbps, por otra parte la calidad de un MP3PRO a 96 Kbps (máximo bitrate constante de los encoders MP3PRO actuales) es similar a la de un MP3 a 128 Kbps ocupando mucho menos espacio en disco y ofreciendo una calidad de audio muy buena para el usuario promedio.

El formato MP3PRO no reemplaza al MP3 original sino que está pensado para ser utilizado en bitrates bajos (32 Kbps a 96 Kbps) ya que es en este espectro donde es mucho mejor que MP3 y es por este motivo especialmente apto para ser usado en reproductores portátiles que usualmente tienen una memoria limitada. Por otro lado para bitrates medios y altos (160 Kbps a 320 Kbps) y consecuentemente muy alta calidad de sonido el formato MP3 es superior a MP3PRO.

Al codificar audio con MP3PRO se producen dos componentes, la parte MP3 propiamente dicha para las frecuencias bajas y la parte SBR o "PRO" para las frecuencias altas. Como la parte PRO tiene muy pocos requerimientos de memoria, la codificación se puede realizar de forma compatible con el formato MP3, lo que permite utilizar reproductores mp3 para escuchar archivos MP3PRO, ignorando la parte PRO.

4.5.3 FORMATO MT9

Este nuevo formato también conocido como Music 2.0, desarrollado por la compañía coreana ETRI (Electronics and Telecommunications Research), trae un nuevo planteamiento a la música digital.

El formato MT9 está estructurado en 6 canales independientes, que contienen la voz, los coros, el piano, las guitarras, el bajo y la percusión, cada uno independiente de los demás pero que son reproducidos simultáneamente en perfecta armonía. Este revolucionario formato permite a los usuarios modular el sonido de cada canción o pieza musical a su gusto y según sus preferencias, permitiéndole subir y bajar (incluso silenciar) el volumen de cada una de estas partes de la canción, pudiendo así disfrutar de solo la percusión, eliminar la voz para usar las canciones en el karaoke, subir las guitarras para captarlas mejor, etc.

La única desventaja, hasta el momento es el traspaso de formato desde MP3 a MT9, ya que para la conversión de una canción en el formato MT9 es necesario remasterizar digitalmente la música original con el fin de conseguir la separación en capas de las pistas de instrumentos.

Su formato de archivo es .mt9 y fue presentado a la Motion Picture Experts Group (MPEG), que es responsable de la certificación de todos los formatos de audio y vídeo MPEG. Actualmente se está considerando hacer del MT9 un formato oficial que pueda ser soportado por cualquier dispositivo.

REFERENCIAS BIBLIOGRÁFICAS

CAPITULO 4: FORMATO MP3

LIBROS:

- [25] Jiménez, M., (2007), Teoría de Información y Codificación, Capítulo 6, Escuela Politécnica Nacional, Quito.
- [26] Tanenbaum, Andrew S., (2003), Redes de computadoras, Prentice-Hall, Pearson Educación, Capítulo 3, Cuarta edición, México.
- [27] Stallings, William, (2000), Comunicaciones y Redes de Computadoras, Prentice-Hall, Sexta edición.
- [28] Lazzati, M., (2009). Electroacústica III, Universidad de las Américas - Quito, Ingeniería en Sonido y Acústica.
- [29] ISO/IEC, (2002). MPEG Audio specification (ISO/IEC 11172-3)⁴⁵ (2002), CD 11172-3 CODING OF MOVING PICTURES AND ASSOCIATED AUDIO FOR DIGITAL STORAGE MEDIA AT UP TO ABOUT 1.5 MBIT/s Part 3 AUDIO, General normative elements, Technical normative elements.
- [30] ISO/IEC, (2002). MPEG Audio specification (ISO/IEC 11172-3) (2002), Annex A (normative) Diagrams, Annex B (normative) Tables, Annex C (informative) The Encoding Process, Annex D (informative) Psychoacoustic Models, Annex E (informative) Bit Sensitivity to Errors, Annex F (informative) Error Concealment, Annex G (informative) Joint Stereo Coding.

⁴⁵ Nota legal: el estándar ISO/IEC 11172-3 tiene copyright siendo propiedad de la Organización Internacional para la Estandarización (ISO, por sus siglas en inglés: International Organization for Standardization) y ha sido utilizado sólo con fines académicos en el presente proyecto de titulación.

Nota legal: el MP3 es un formato propietario del Instituto Fraunhofer y ha sido utilizado sólo con fines académicos en el presente proyecto de titulación.

- [31] Ramapriya Rangachar, (2001). ANALYSIS AND IMPROVEMENT OF THE MPEG-1 AUDIO LAYER III ALGORITHM AT LOW BIT-RATES, Arizona State University.
- [32] Joebert S. Jacaba, (2001). AUDIO COMPRESSION USING MODIFIED DISCRETE COSINE TRANSFORM: THE MP3 CODING STANDARD, Department of Mathematics, College of Science the University of the Philippines Diliman, Quezon City.
- [33] Sieler M., Sperschneider R., (1997). MPEG-Layer3, Bitstream Syntax and Decoding.
- [34] Bourlot, C., Vollmann, A., (2000). Reproductor Portátil De Audio Digital Comprimido, Universidad Católica de Córdoba, Argentina.

PAPERS:

- [35] Pan, Davis. A tutorial on MPEG/Audio compression, IEEE Multimedia Journal, Vol. 2 No. 2 (Summer 1995).

INTERNET:

- [36] MasterMagazine (2004). Biografía de Karlheinz Brandenburg / MP3.
<http://www.mastermagazine.info/articulo/11696.php>
- [37] Tecnologiabit, (2010). Nuevos MP3 | MP3HD, MP3Pro, MP3D, MP3 SX y MP3 Surround, Argentina.
<http://www.meyersound.com/spanish/products/mseries/milo/news.php>
- [38] Dogguie.com, (2010). Curiosos reproductores mp3.
<http://www.dogguie.com/curiosos-reproductores-mp3/>
- [39] Predrag Supurovic, (1999). MPEG Audio Compression Basics.
http://www.mpgedit.org/mpgedit/mpeg_format/mpeghdr.htm
- [40] Livingstone, D., (1999). Historia y Reproductores de MP3.
<http://www.pc-doctor.com.mx/Radio%20Formula/temas/Historia%20del%20MP3.htm>
- [41] Nogales, A., (2002). MP3.
<http://www.iesleonardo.info/ele/tra/CURSO%202001-2002/t02-mp3/>

CAPÍTULO 5.

DISEÑO E IMPLEMENTACIÓN EN MATLAB DEL ALGORITMO MP3 PARA COMPRESIÓN DE AUDIO

5.1 INTRODUCCIÓN

El propósito de este capítulo es el diseño y la implementación en software de simulación Matlab del algoritmo MP3 para compresión de audio, basados en el estándar ISO/IEC 11172-3 propiedad de la Organización Internacional para la Estandarización (ISO, por sus siglas en inglés: International Organization for Standardization), cabe señalar que el MP3 es un formato propietario del Instituto Fraunhofer. Con esto, se pretende lograr una reducción significativa de la cantidad de bits considerando que al reproducir el archivo de audio comprimido, la diferencia en comparación con el audio original sea mínima

El proyecto se basa en el estándar ISO/IEC 11172-3 ya que el mismo describe la manera de generar archivos de audio MP3 válidos.

El software de simulación escogido es el Matlab debido a que permite realizar cálculos numéricos rápidos y con alta precisión, la programación se realiza mediante un lenguaje de alto nivel, con la ayuda de la extensa biblioteca de funciones que permite el uso de técnicas digitales para el procesamiento de señales, además, incluye un soporte básico para diseño de interfaz gráfica que permitirá realizar la implementación del modelo en un ambiente amigable al usuario, permitiendo crear una interacción entre el usuario y el codificador MP3 implementado.

En la implementación del algoritmo MP3 se hace una descripción completamente detallada de la secuencia en que se realiza la codificación de un archivo de audio .wav a un archivo de audio comprimido .mp3.

5.2 SOFTWARE DE SIMULACIÓN MATLAB

5.2.1 DESCRIPCIÓN GENERAL

El software utilizado para la implementación del codificador MP3 es Matlab®⁴⁶ (Matrix Laboratory) un producto de la empresa The Mathworks Inc., empresa fundada en 1984 y que cuenta actualmente con más de 1 millón de usuarios registrados en más de 175 países.

Matlab es un programa que permite realizar cálculos numéricos con vectores y matrices, puede también trabajar con números escalares tanto reales como complejos, con cadenas de caracteres y con otras estructuras de información más complejas. Una de las capacidades más atractivas es la de realizar una amplia variedad de gráficos en dos y tres dimensiones.

Matlab constituye una de las herramientas líderes en el área de ingeniería, dispone de un extenso conjunto de comandos y funciones que realizan tareas específicas, además dispone de un código básico y más de 30 librerías especializadas llamadas “toolboxes” las cuales trabajan en áreas especiales como las comunicaciones, el procesamiento de señales, etc.

Matlab, es un lenguaje de programación de alto nivel que permite realizar interfaces gráficas de usuario y también permite tener interfaces con otros lenguajes de programación.

5.2.2 CARACTERÍSTICAS

El software de simulación Matlab empleado en la implementación del codificador MP3 tiene las siguientes características:

⁴⁶ Nota legal: MATLAB es un software propietario de The Mathworks Inc. y ha sido utilizado sólo con fines académicos en el presente proyecto de titulación.

- Versión: 7.8.0.347 - Release: R2009a; versión para PC: 32 bits (win32)
- Toolboxes empleados: Matlab Toolbox™, Signal Processing Toolbox™, Communications Toolbox™, Mp3 Toolbox™, Guide Toolbox™
- Compatible con Windows XP, Windows Vista y Windows 7

5.2.3 GUIDE (GRAPHICAL USER INTERFACE DEVELOPMENT ENVIRONMENT)

GUIDE es un entorno de programación visual disponible en Matlab para realizar y ejecutar programas que necesiten ingreso continuo de datos. Tiene las características básicas de todos los programas visuales como Visual Basic o Visual C++.

Una aplicación GUIDE consta de dos archivos: *.m* y *.fig*. El archivo *.m* es el que contiene el código con las correspondencias de los botones de control de la interfaz y el archivo *.fig* contiene los elementos gráficos. Cada vez que se adicione un nuevo elemento en la interfaz gráfica, se genera automáticamente código en el archivo *.m*.

5.3 CRITERIOS DE DISEÑO

Para la implementación del codificador MP3 en Matlab se considera realizar varios archivos *.m*, archivos *.fig* y archivos *.mat* relacionados entre sí, que permitan obtener al final un archivo de audio comprimido mp3 válido bajo el estándar ISO/IEC 11172-3.

Ya que la simulación se realiza con una computadora, es necesario discretizar la señal analógica. Por esta razón, solo se toman valores de la señal analógica cada cierto tiempo denominado periodo de muestreo (el inverso de la frecuencia de muestreo). Una vez discretizada la señal de audio se introducirán algunos efectos (estos efectos están completamente detallados en el código fuente) que permitan procesar la señal digital de audio.

Con respecto a la utilización de la herramienta gráfica de Matlab Guide, el programa está distribuido en varias interfaces gráficas interactivas con el usuario, en total el programa está constituido por cinco ventanas principales y dos ventanas adicionales. Tanto los elementos de la interfaces gráficas como las interfaces en sí mismo van apareciendo secuencialmente mientras se va procesando la señal de audio.

El programa implementado permite convertir un archivo de audio de formato Windows PCM (.wav) con frecuencia de muestreo de 44.100 Hz en modo estéreo o monofónico, a un archivo de audio con formato comprimido MPEG-1 Layer III (.mp3) monofónico, con una tasa de bits (bitrate) variable que puede ser escogida por el usuario desde los 96 Kbps a 320 Kbps, con una frecuencia de muestreo de 44.100 Hz.

5.4 IMPLEMENTACIÓN DEL ALGORITMO DE CODIFICACIÓN MP3 EN MATLAB

La implementación del algoritmo de codificación MP3 está realizada en base a las siguientes etapas:

- Banco de filtros
- Modelo psicoacústico
- Transformada discreta del coseno modificada (MDCT)
- Cuantización y codificación
- Formato de la trama MP3

En la implementación del algoritmo se ha creído conveniente tratar estas cinco etapas principales en varios archivos .m en Matlab, en total el algoritmo de compresión de audio MP3 implementado, consta de 41 archivos distribuidos de la siguiente manera: 30 archivos *.m, 7 archivos *.fig y 4 archivos *.mat, pensados para trabajar en conjunto, que permitan mostrar en detalle cómo opera el algoritmo de codificación MP3 hasta obtener un archivo de audio comprimido válido con extensión *.mp3.

5.4.1 DIAGRAMAS DE FLUJO

5.4.1.1 Proceso de adquisición de la señal de audio de Microsoft *.wav

En la figura 5.1 se puede apreciar el diagrama correspondiente al proceso de adquisición de la señal de audio de Microsoft (*.wav) muestreada a 44.100 Hz en modo estéreo o monofónico.

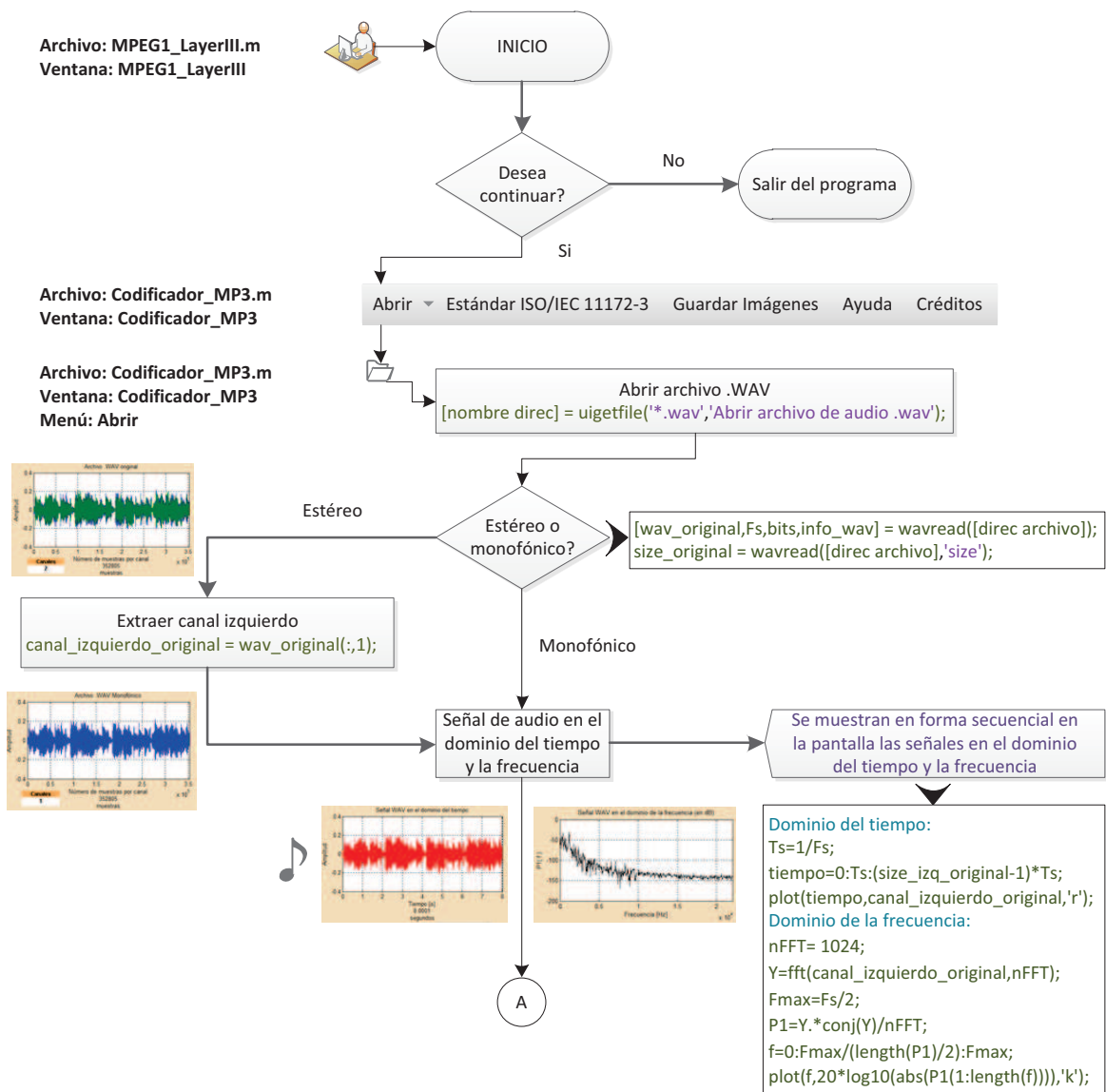


Figura 5. 1: Diagrama de flujo para la adquisición de la señal de audio de Microsoft *.wav

5.4.1.2 Codificación del archivo *.wav a un archivo de audio en formato comprimido *.mp3

5.4.1.2.1 Selección de parámetros necesarios para la codificación

Una vez adquirida la señal de audio en formato .wav, se procede a escoger los parámetros como la tasa de bits (bitrate) en un rango de 96 Kbps a 320 Kbps y el ingreso del tiempo de codificación en segundos, (cuánto tiempo se quiere codificar del archivo original) entre un tiempo mínimo correspondiente a procesar 4 tramas MP3 como mínimo y un tiempo máximo debido a la aplicación de la FFT (transformada rápida de Fourier) de 1024 puntos, sobre las 1152 muestras que comprenden una trama MP3.

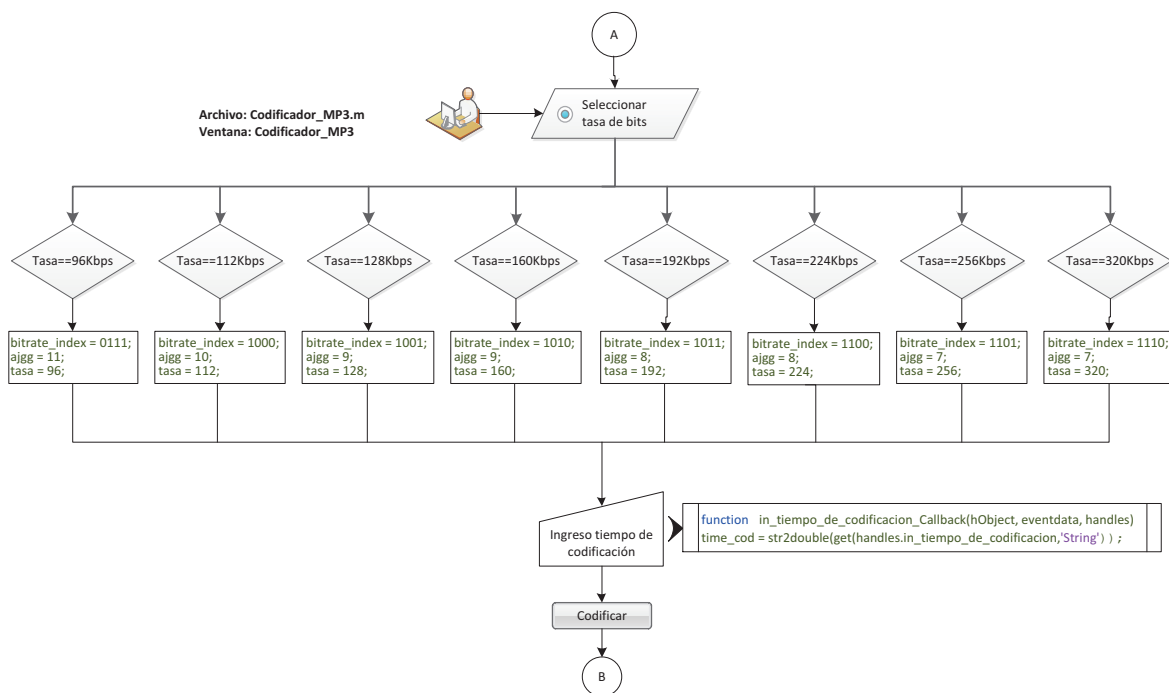


Figura 5. 2: Diagrama de flujo para la selección de los parámetros necesarios para la codificación

5.4.1.2.2 Codificación

Una vez seleccionada la tasa de bits e introducido el tiempo que se quiere codificar de la señal, se procede a ejecutar el algoritmo de compresión de audio MP3 según el estándar internacional ISO/IEC 11172-3, para ello se sigue el diagrama de bloques provisto en la figura 5.3.

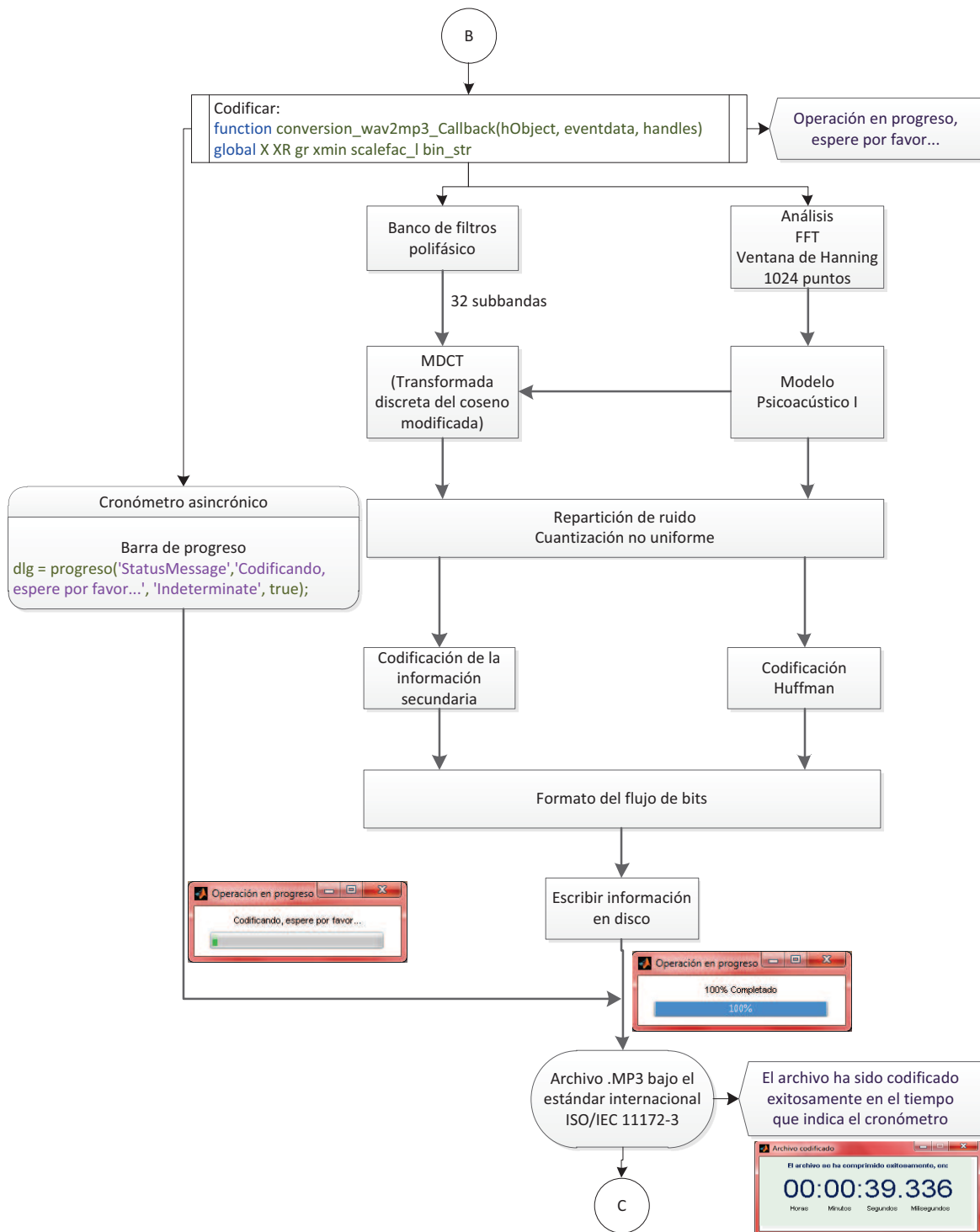


Figura 5. 3: Diagrama de flujo para la codificación del archivo *.wav a un archivo de audio en formato comprimido *.mp3

Una vez realizada la codificación se puede visualizar y reproducir el archivo de audio comprimido en formato MP3, el diagrama correspondiente se puede observar en la figura 5.11.

5.4.1.2.3 Banco de filtros polifásico

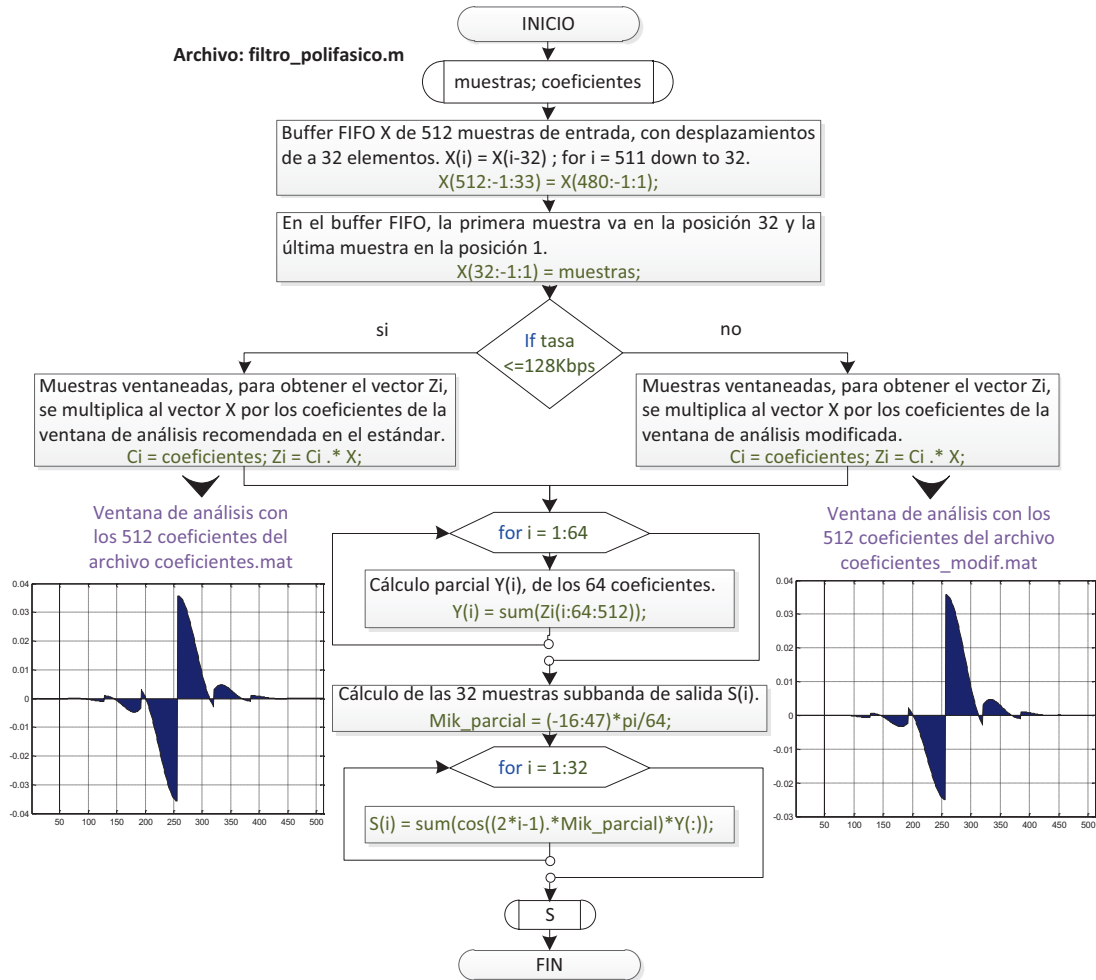


Figura 5. 4: Diagrama de flujo para la etapa correspondiente al banco de filtros [45]

5.4.1.2.4 Análisis FFT

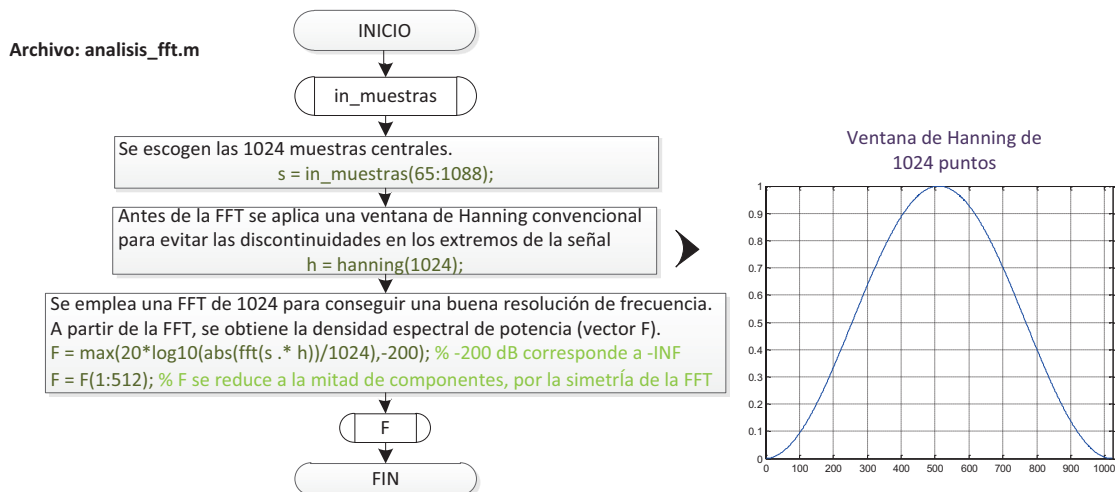


Figura 5. 5: Diagrama de flujo para el análisis FFT del Modelo psicoacústico I

5.4.1.2.5 Transformada discreta del coseno modificada (MDCT)

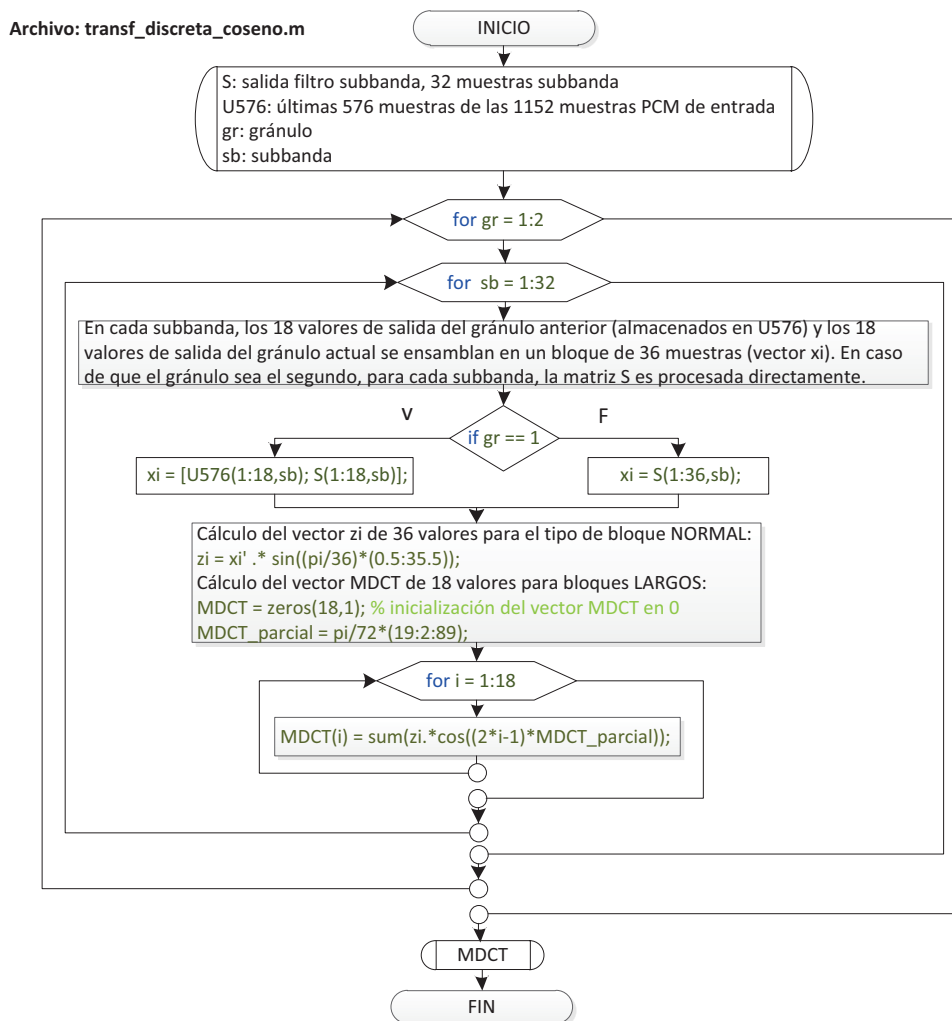


Figura 5. 6: Diagrama de flujo para la transformada discreta del coseno modificada MDCT [45]

5.4.1.2.6 Cuantización y codificación

La etapa correspondiente a la cuantización y codificación es la más compleja del estándar y debe ser seguida de acuerdo a lo especificado en el estándar respetando sus pasos.

Para ello, a continuación las figuras 5.7, 5.8 y 5.9 describen de manera simplificada la manera en que se realiza la cuantización no uniforme y la codificación Huffman; se presentan los 3 lazos de iteración necesarios en la cuantización.

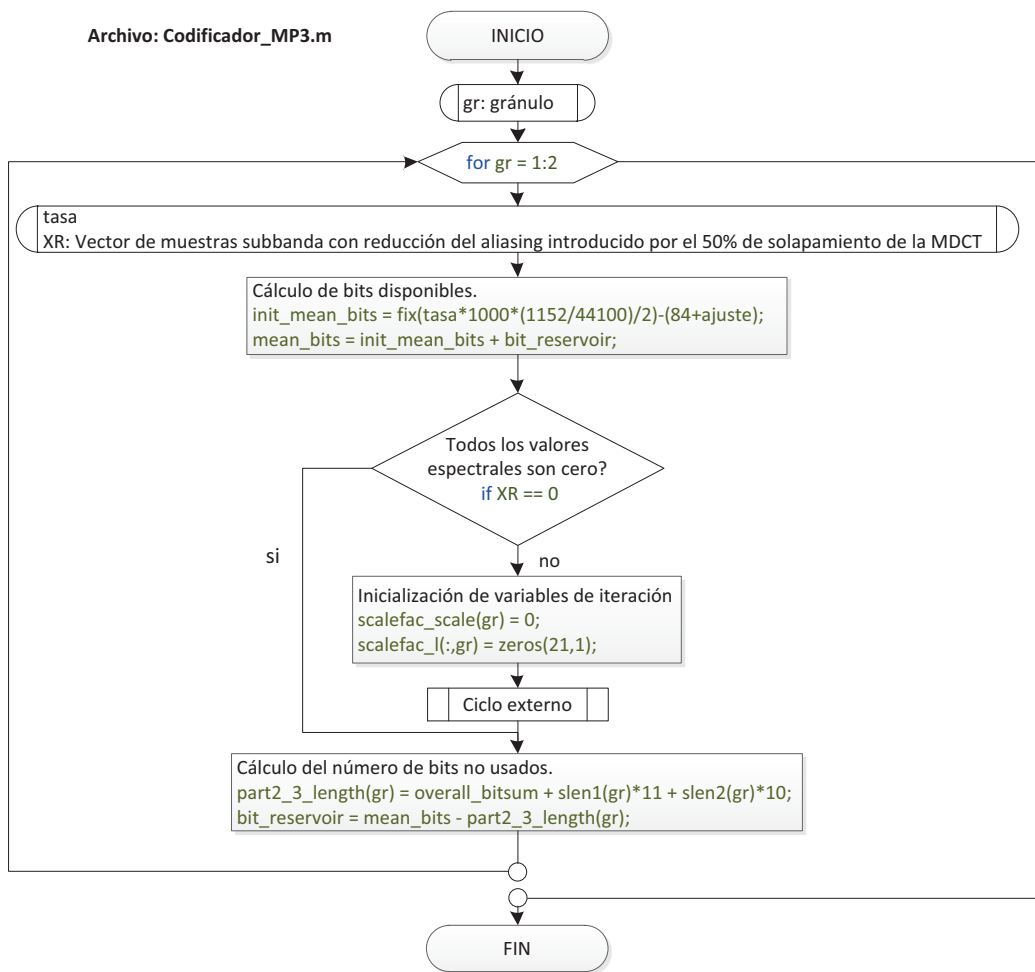
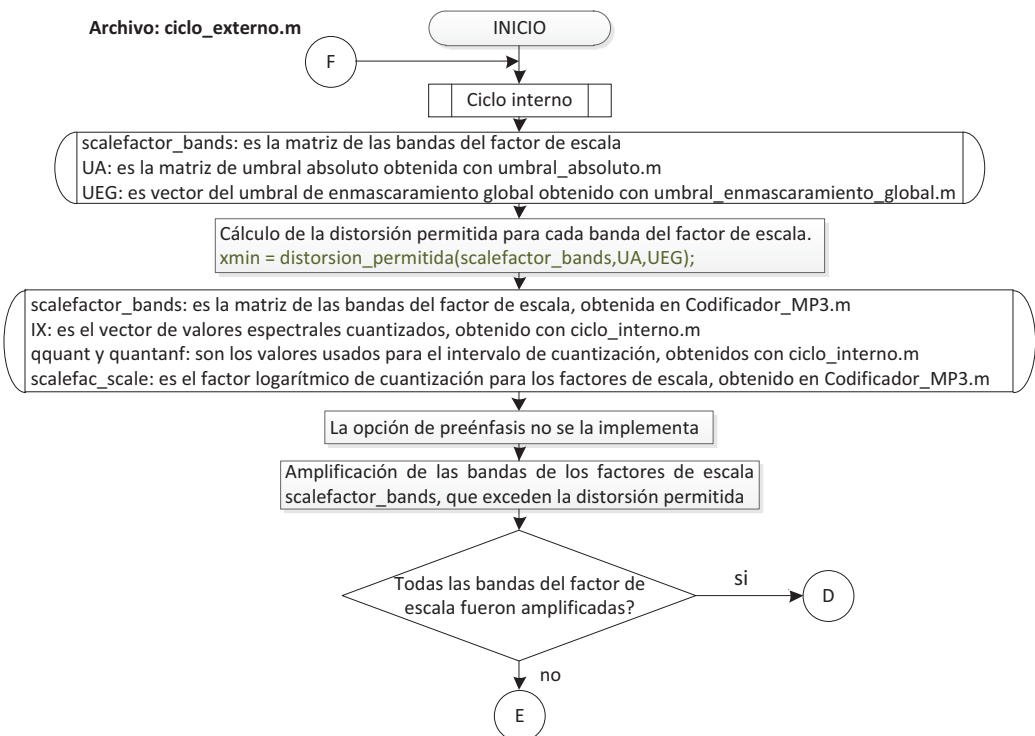


Figura 5. 7: Diagrama de flujo del lazo de iteración para la cuantización no uniforme [43]



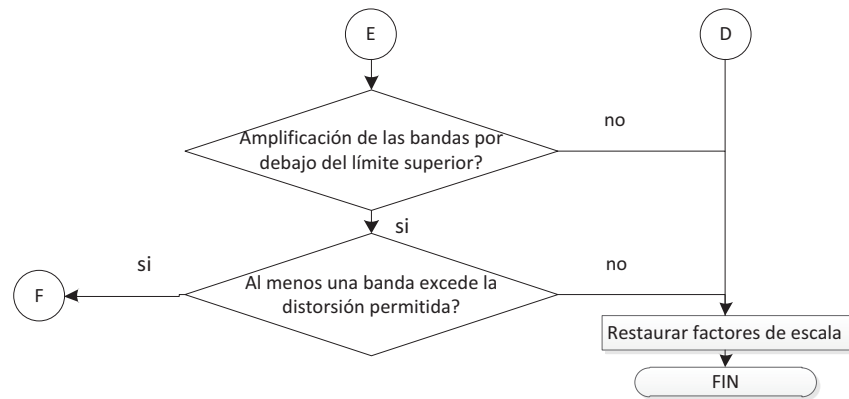


Figura 5. 8: Diagrama de flujo del lazo de iteración para el ciclo externo [43]

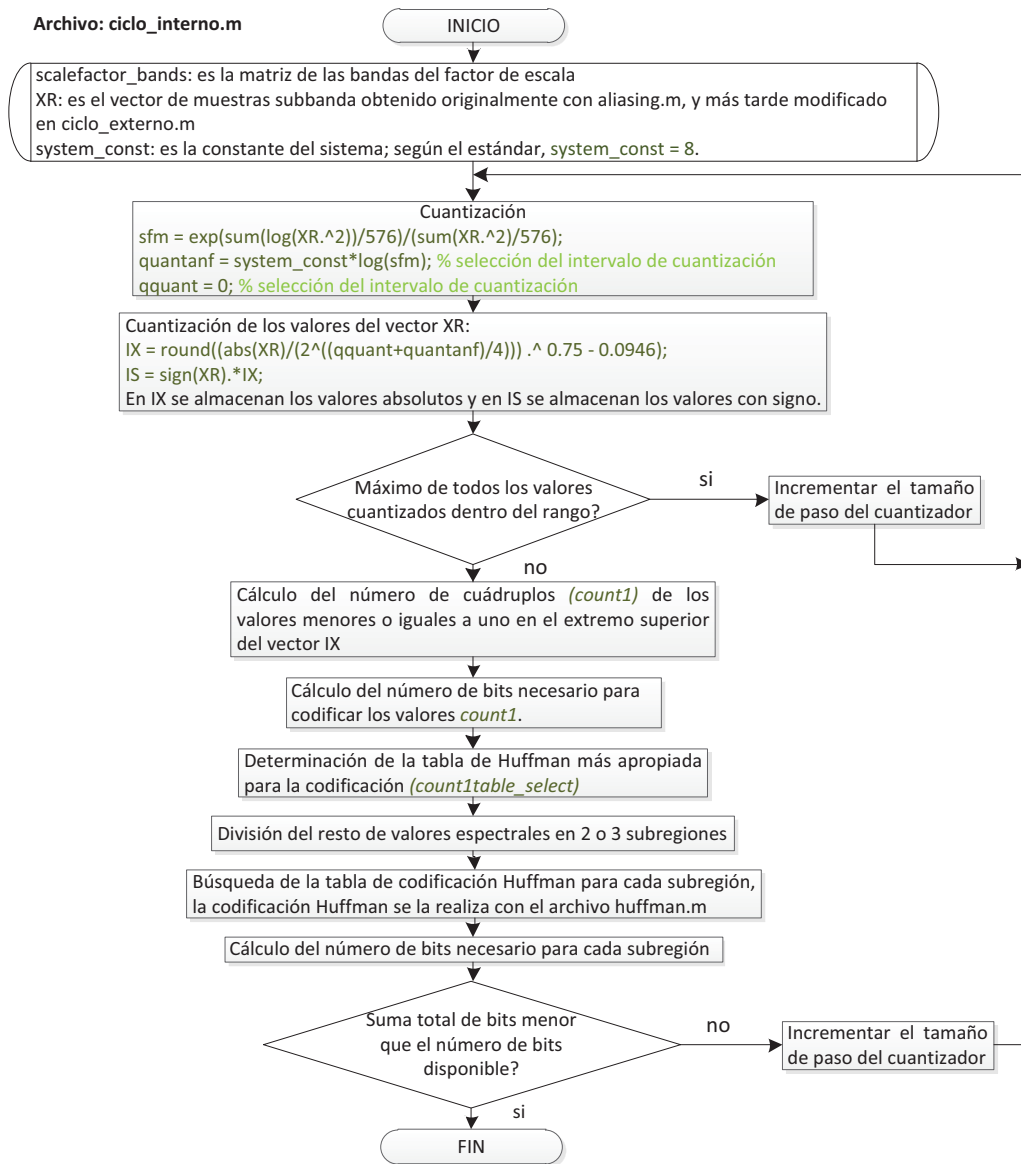


Figura 5. 9: Diagrama de flujo del lazo de iteración para el ciclo interno [43]

5.4.1.2.7 Formato del flujo de bits MP3 válido

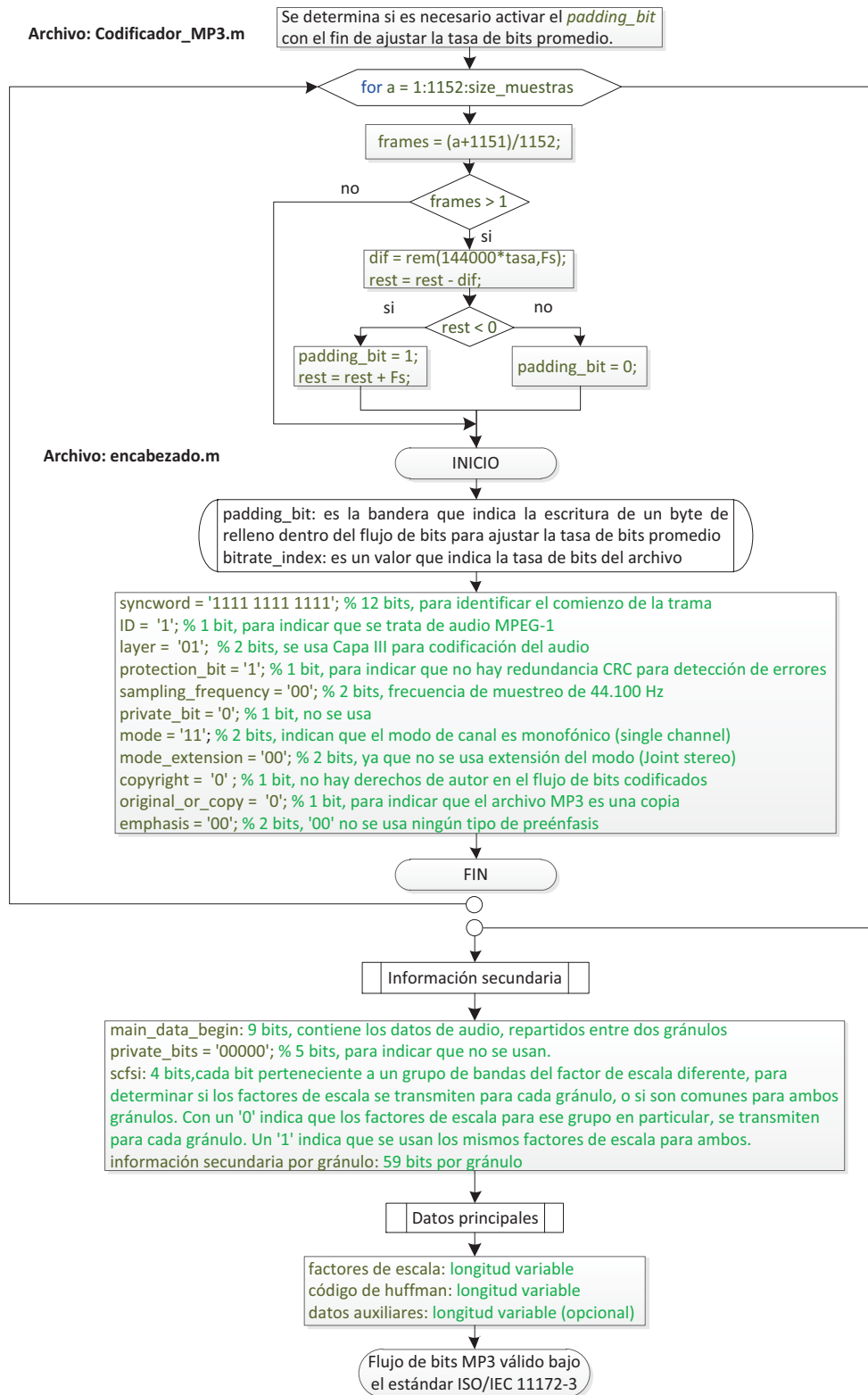


Figura 5. 10: Diagrama de flujo para el formato de la trama MP3 y obtención del flujo de bits MP3 válido bajo el estándar internacional ISO/IEC 11172-3 [43]

5.4.1.3 Proceso de adquisición de la señal de audio en formato comprimido *.mp3

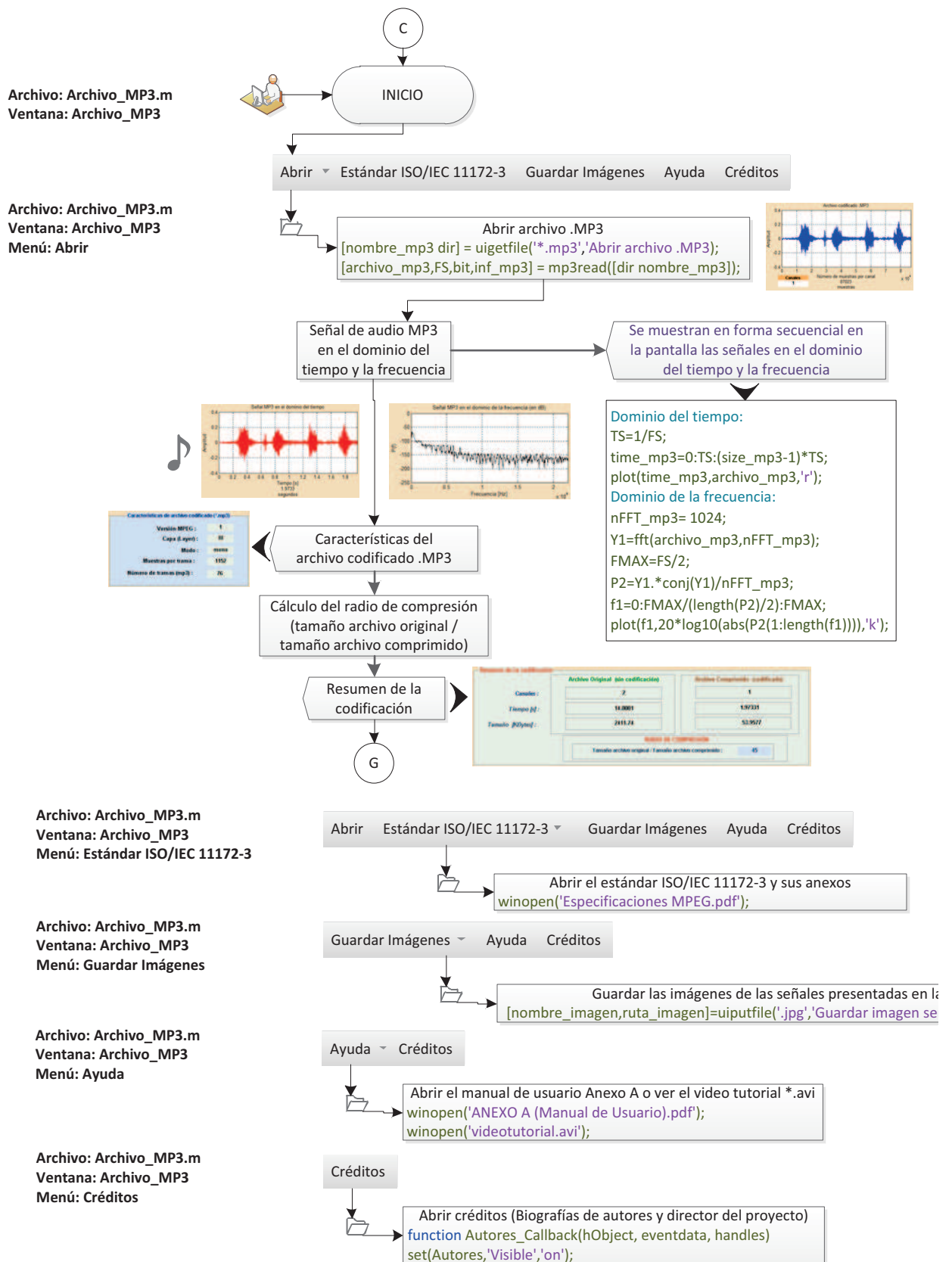


Figura 5. 11: Diagrama de flujo para la lectura del archivo de audio comprimido *.mp3

5.4.1.4 Comparación entre archivo de audio original *.wav y archivo en formato de audio comprimido *.mp3

Para poder comparar el archivo de audio original con el archivo de audio comprimido, se creyó conveniente crear otra ventana adicional que permita visualizar y comparar ambos archivos tanto en el dominio del tiempo como en el dominio de frecuencia.

El diagrama correspondiente a la comparación de los archivos .wav y .mp3 se puede apreciar en la figura 5.12, a continuación.

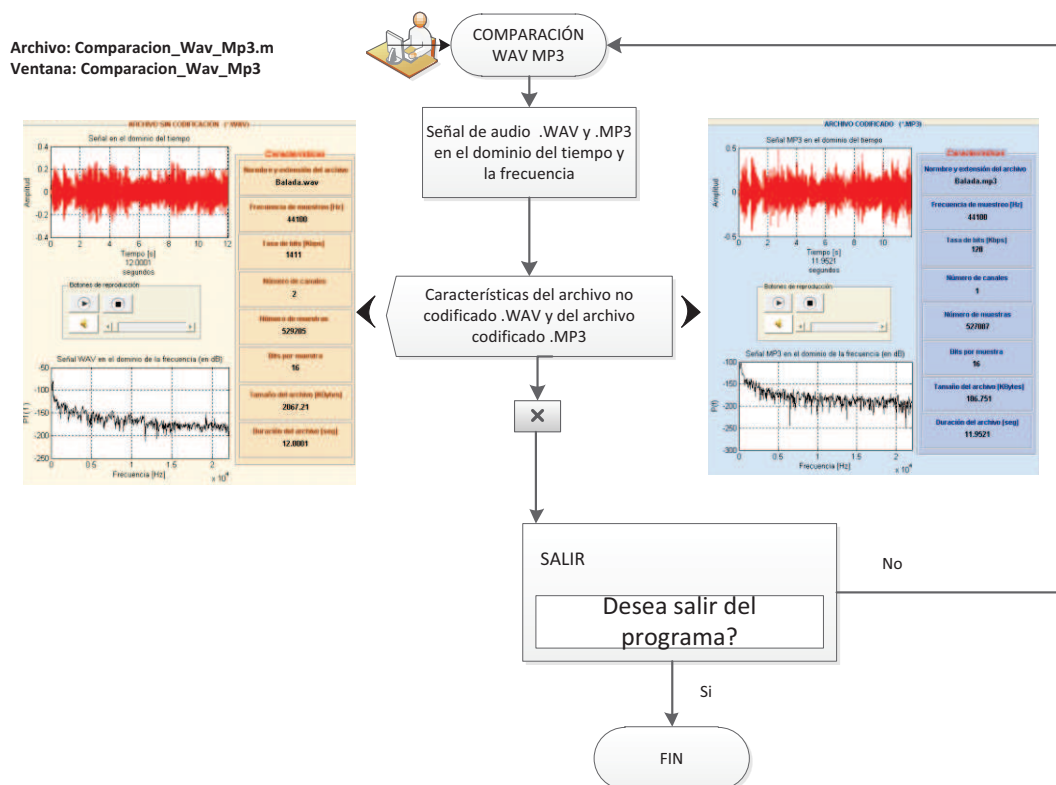


Figura 5. 12: Diagrama de flujo para la comparación del archivo .wav y el archivo de audio comprimido .mp3

Una vez realizados los diagramas de flujo correspondientes a la implementación del algoritmo MP3 para compresión de audio en Matlab, se explica cada una de las etapas en cuanto a su implementación y los archivos *.m y *.mat que comprenden cada etapa, pero que juntos permiten ejecutar paso a paso el algoritmo de compresión de audio MP3 para obtener al final un archivo de audio comprimido .mp3 válido.

5.4.2 IMPLEMENTACIÓN DEL BANCO DE FILTROS

Los archivos de Matlab que comprenden la etapa del banco de filtros son:

- filtro_polifasico.m
- coeficientes.mat y coeficientes_modif.mat

La primera etapa correspondiente al banco de filtros, consiste en dividir la señal de audio PCM muestreada a $F_s = 44.100$ Hz en 32 subbandas igualmente espaciadas a $1.378,125$ Hz ($F_s/32$).

5.4.2.1 Filtro subbanda polifásico

Para la implementación del filtro subbanda polifásico en el algoritmo MP3 se efectúan los siguientes pasos:

1. Se toman 32 muestras de audio de entrada.
2. En un buffer FIFO (vector) de 512 muestras se almacenan las muestras de audio de entrada. Las 32 muestras de audio son cambiadas, la primera muestra va en la posición 32 y la última muestra en la posición 1.
3. Se aplica una ventana de análisis recomendada en el estándar ISO/IEC 11172-3 al vector con las 512 muestras, mediante un producto entre el vector y los 512 coeficientes de la ventana de análisis.

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]= -0.000971317	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 5. 1: Coeficientes C_i de la ventana de análisis según el estándar [42]

Al realizar la implementación de la ventana de análisis recomendada por el estándar, se observó que a partir de los 160 Kbps hasta 320 Kbps empiezan a originarse picos no deseados, denominados como *artefactos de audio*⁴⁷ los cuales distorsionan la señal de audio codificada.

⁴⁷ Artefactos de audio: son las inconsistencias en el espectro de frecuencias causadas por las normas de compresión de un algoritmo de codificación, en éste caso el algoritmo MP3. Los artefactos de audio pueden ser reconocidos al reproducir el audio digital y escuchar un crujido.

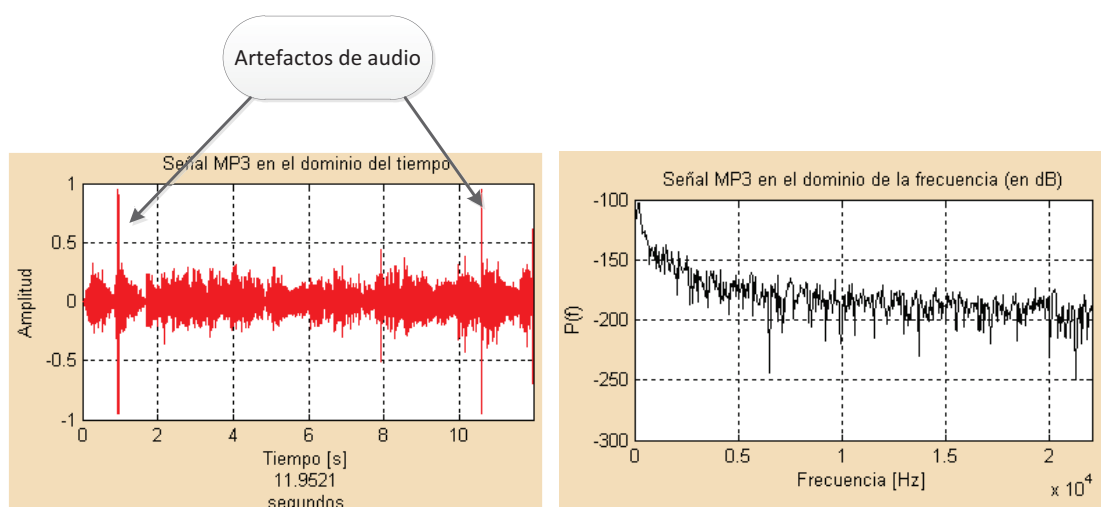


Figura 5. 13: Artefactos de audio presentes en la señal de audio codificada a 160 Kbps

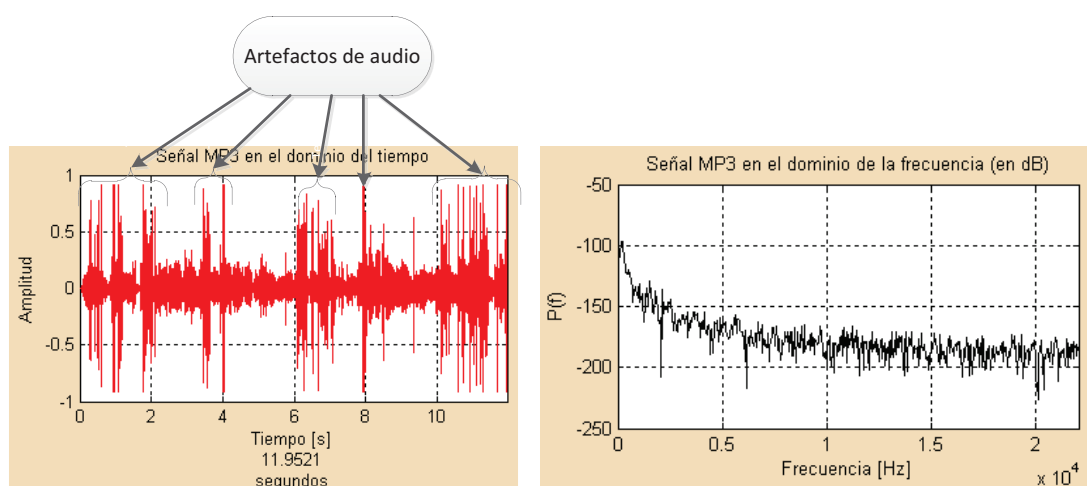


Figura 5. 14: Artefactos de audio presentes en la señal de audio codificada a 320 Kbps

Como se puede ver en las figuras 5.13 y 5.14 a mayores tasas de bits la presencia de los artefactos de audio es mayor, esto conlleva a tener un mayor número de “crujidos” en la señal al momento de reproducirla, siendo un fenómeno molesto no deseado.

Tras un exhaustivo análisis y verificando en detalle la programación empleada mediante la ayuda del depurador de Matlab, la existencia de los artefactos de audio se puede justificar por las siguientes razones:

- El algoritmo de codificación para la capa III es muy complejo, en varias etapas del programa se necesita realizar procesos iterativos, y al trabajar con tasas de bits altas, se realiza mayor procesamiento y por lo tanto se puede decir que a partir de los 160 [Kbps] se produce un efecto de *overflow* el cual origina los artefactos de audio.
- El diseño y la implementación del algoritmo MP3 en Matlab han sido basados en la información del estándar internacional ISO/IEC 11172-3, y a pesar de que dicha información se ha seguido al pie de la letra, puede que algún detalle se haya omitido en la publicación del estándar al ser éste de tipo propietario.

Tras el análisis correspondiente se logró determinar que modificando los coeficientes de la ventana de análisis recomendada en el estándar, tal y como se muestra en la figura 5.15, los artefactos de audio no deseados desaparecen y esto se evidencia al reproducir las señales en el dominio del tiempo de las figuras 5.16 y 5.17.

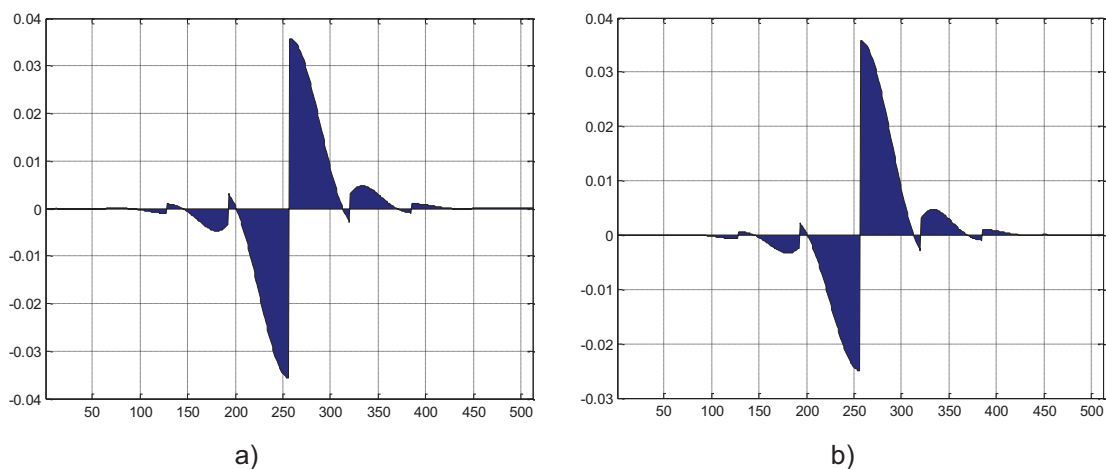


Figura 5. 15: a) Ventana de análisis recomendada en el estándar b) Ventana modificada en el diseño e implementada para velocidades superiores a 128 Kbps

Es posible modificar la ventana de análisis debido a que la misma se encuentra en el Anexo C del estándar que es de carácter informativo, es decir, no es mandatorio la implementación tal como está la ventana dada, debido a que la misma es obtenida a través de pruebas y es una recomendación en el estándar.

Los coeficientes modificados de la ventana de análisis son los primeros 256 valores multiplicados por un factor de 0,7; los restantes 256 coeficientes se conservan tal como se muestra en la tabla 5.1.

En las figuras 5.16 y 5.17 se muestran los resultados obtenidos al solucionar el problema de los artefactos de audio, cabe destacar que en el diseño y la implementación del algoritmo se busca la reducción significativa del tamaño (en bytes) del archivo de audio con la mínima pérdida en la calidad.

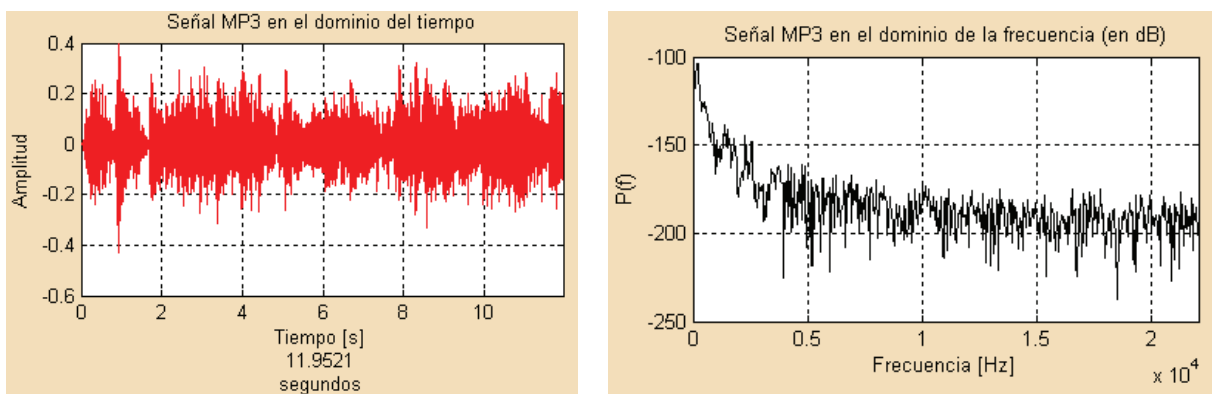


Figura 5. 16: Eliminación de los artefactos de audio presentes en la señal de audio codificada a 160 Kbps

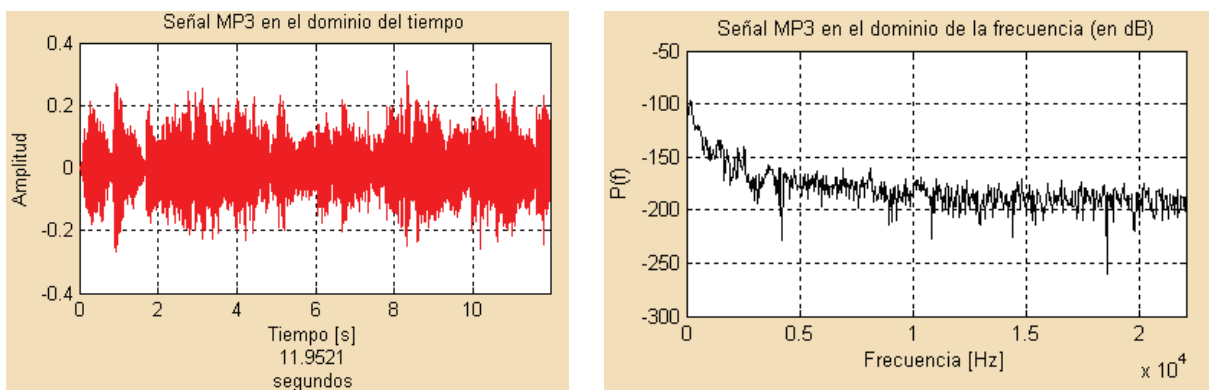


Figura 5. 17: Eliminación de los artefactos de audio presentes en la señal de audio codificada a 320 Kbps

4. Por último se obtienen las 32 muestras subbanda de salida (cada filtro subbanda produce una muestra subbanda de salida por cada 32 muestras de audio de entrada). En la capa 3 de MPEG-1, cada subbanda aporta 36 muestras de audio para un total de 1152 muestras subbanda por trama.

5.4.3 IMPLEMENTACIÓN DEL MODELO PSICOACÚSTICO

El estándar ISO/IEC 11172-3 deja a libertad de elección el modelo psicoacústico a implementar, es por ello que, el modelo psicoacústico elegido en el proyecto e implementado en Matlab es el modelo I detallado en el capítulo 4, se escogió este modelo para simplificar los cálculos y comprender mejor el algoritmo.

Los archivos de Matlab que comprenden la etapa del modelo psicoacústico I son:

- analisis_fft.m
- componentes.m
- umbral_absoluto.m
- limites_banda_critica.m
- reducción_componentes.m
- umbral_enmascaramiento_individual.m
- umbral_enmascaramiento_global.m

En esta etapa es necesario determinar en cada subbanda el máximo nivel de señal y el mínimo umbral de enmascaramiento.

El mínimo umbral de enmascaramiento se deriva de la FFT de la señal de entrada PCM. La FFT en paralelo con el filtro subbanda compensan la falta de selectividad espectral obtenida en las bajas frecuencias por el banco de filtros subbanda.

5.4.3.1 Análisis FFT

El modelo I, el cual se está utilizando en la implementación del codificador, usa una FFT de 1024 puntos para la capa 3.

Al utilizar la FFT de 1024 puntos para las 1152 muestras PCM de entrada, se escogen las 1024 muestras centrales de las 1152 muestras, a las cuales se les aplica una ventana de Hanning convencional antes de calcular su FFT, para evitar las discontinuidades en los extremos de la señal.

Se emplea una FFT de 1024 para conseguir una buena resolución de frecuencia y poder calcular correctamente los umbrales de enmascaramiento.

A partir de la FFT, se obtiene la densidad espectral de potencia (F).

5.4.3.2 Componentes tonales y no tonales

Una vez obtenida la señal en el dominio de la frecuencia, se identifica y separa las componentes tonales y no tonales presentes en la señal de audio, para ello, se determina una lista de los máximos (picos) locales para las líneas de frecuencia, de acuerdo con los requerimientos del estándar ISO/IEC 11172-3, el análisis sólo es necesario hacerlo para las líneas de frecuencia con índice k entre 3 y 500.

Si se cumplen las dos condiciones de que $F(k) > F(k-1)$ & $F(k) \geq F(k+1)$, una línea de frecuencia es etiquetada como máximo local, una vez identificados los máximos locales, hay que diferenciarlos entre componentes tonales y no tonales y determinar su nivel de presión sonora, en este punto hacemos uso de la función `umbral_absoluto.m` que contiene una tabla de "*Frecuencias, Tasas de Bandas Críticas y Umbral Absoluto*", no es necesario codificar los sonidos situados bajo el umbral absoluto ya que éstos no serán percibidos.

Frecuencia Hz	Tasa de Banda Crítica Barks	Umbral Absoluto dB
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

Tabla 5. 2: Frecuencias, Tasas de Bandas Críticas y Umbral Absoluto [42]

Mientras que la función `limites_banda_critica.m` que contiene una tabla con los "*Límites de las Bandas Críticas*", destacando que las frecuencias corresponden al límite superior de cada banda crítica (ver tabla 1.4).

Índice	Frecuencia	Tasa de Banda Crítica
<i>i</i>	Hz	Barks
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	24.573

Tabla 5. 3: Límites de las bandas críticas [42]

5.4.3.3 Reducción de componentes enmascarantes

Determinadas las componentes tonales y no tonales, se realiza la reducción de componentes enmascarantes para lo cual se procede de la siguiente manera:

1. Las componentes que se encuentran por debajo del mínimo umbral auditivo, o que su distancia con respecto a otra componente es menor a 0,5 Barks, son eliminadas.
2. Se elimina(n) la(s) componente(s) con menor nivel de presión sonora y se conserva la componente con mayor nivel.

5.4.3.4 Umbral de enmascaramiento individual

Los umbrales de enmascaramiento individual para las componentes tonales y no tonales se fijan en $-\infty$, ya que la función de enmascaramiento tiene atenuación infinita más allá de -3 y de $+8$ Barks, es decir, la componente no tiene efecto enmascarante sobre frecuencias más allá de aquellos rangos.

Es por ello, que éste análisis sólo es necesario hacerlo para las líneas de frecuencia que se encuentren entre -3 y $+8$ Barks a partir de la componente enmascaradora.

Sólo un subconjunto de las muestras son consideradas para el futuro cálculo del umbral de enmascaramiento global.

5.4.3.5 Umbral de enmascaramiento global

El umbral de enmascaramiento global se obtiene de la suma de las potencias correspondientes de los umbrales de enmascaramiento individual y del umbral absoluto o umbral en silencio.

5.4.4 IMPLEMENTACIÓN DE LA TRANSFORMADA DISCRETA DEL COSENO MODIFICADA (MDCT)

Los archivos de Matlab que comprenden la etapa de la transformada discreta del coseno modificada son:

- `transf_discreta_coseno.m`
- `aliasing.m`

En esta etapa se emplea la transformada discreta del coseno modificada para compensar la falta de precisión del banco de filtros, logrando subdividir la salida

espectral en frecuencias que proporcionen mejor resolución con respecto a las bandas críticas.

En el proceso de la transformada MDCT implementado en Matlab se producen 576 líneas de frecuencia referidas como "Gránulo" (subdivisión de una trama).

El archivo `aliasing.m` realiza la reducción del aliasing introducido por el 50% de solapamiento de la MDCT, permitiendo obtener una reducción en la cantidad de información a ser codificada y transmitida.

5.4.5 IMPLEMENTACIÓN DE LA ETAPA DE CUANTIZACIÓN Y CODIFICACIÓN

La etapa correspondiente a la cuantización y codificación dentro del algoritmo de compresión de audio MP3, la componen los siguientes archivos de Matlab:

- `distorsion_permitida.m`
- `ciclo_interno.m`
- `ciclo_externo.m`
- `huffman.m`

El fundamento de la compresión de audio MPEG es la cuantización, es por ello que en la implementación de esta etapa se consideró realizar la cuantización en tres etapas y siguiendo los lineamientos del estándar ISO/IEC 11172-3.

El archivo `distorsion_permitida.m` nos permite calcular la distorsión admitida para cada una de las 21 bandas del factor de escala necesarias en los lazos iterativos de la cuantización, para ello se utiliza los niveles de umbral absoluto dados en la tabla 5.2 y el umbral de enmascaramiento global obtenido del modelo psicoacústico I.

Una vez verificada la distorsión en cada banda del factor de escala y almacenada la información en un vector se procede a ejecutar el lazo de iteración principal, en este lazo se realizan los siguientes pasos:

- Cálculo del número de bits disponibles para cada gránulo (subdivisión de una trama).
- Se inicializan todas las variables de iteración.
- Si no hay datos de audio, se devuelven ciertos valores por defecto, proporcionados en el estándar.
- En caso de haber datos, se comparan los valores espectrales, si estos valores son todos igual a cero, se calcula el número de bits no usados y se retorna al lazo. En caso de ser diferente de cero todos los valores espectrales, se ejecuta el lazo de iteración externo mediante el archivo `ciclo_externo.m`

En el lazo de iteración externo o denominado en el estándar como *ciclo externo* se llevan a cabo las siguientes acciones:

- Primero se lleva a cabo el lazo de iteración interno mediante el archivo `ciclo_interno.m`
- Se utiliza el vector con la distorsión permitida para cada subbanda del factor de escala calculado mediante el archivo `distorsion_permitida.m`
- Se guardan los factores de escala.
- En la implementación del algoritmo, no se emplea la opción de *preénfasis*.
- Se amplifican las bandas del factor de escala cuyo valor supera a la distorsión permitida
- Se verifica si todas las bandas del factor de escala han sido amplificadas, en caso de no serlo, se amplifican todas las bandas por debajo del límite superior, una vez amplificadas, se verifica si por lo menos existe una banda con más de la distorsión permitida, en caso de existir alguna banda se vuelve a ejecutar el lazo de iteración externo, caso contrario se restauran los valores de los factores de escala.

El lazo de iteración interno o denominado en el estándar como *ciclo interno* comprende los siguientes pasos:

- Se realiza la cuantización, para ello se emplea un cuantizador no uniforme, según recomendación del estándar. El cuantizador eleva los valores a ser cuantizados a $\frac{3}{4}$ de potencia antes de la cuantización, esto provee una relación señal a ruido SNR más estable.
- Se escoge un determinado paso de cuantización, para cuantizar los valores espectrales, y a estos datos cuantizados se les aplica codificación de Huffman.
- Si la suma total de los bits utilizados para la codificación Huffman no es menor que el número de bits disponible, se incrementa el tamaño del paso de cuantización y se vuelve a ejecutar el lazo de iteración.
- El ciclo interno termina cuando los valores cuantizados que han sido codificados con Huffman usan menor o igual número de bits que la máxima cantidad de bits disponible.

El archivo `huffman.m` contiene todas las tablas recomendadas en el estándar, y como la codificación Huffman es una parte normativa del estándar, se implementaron todas las tablas proporcionadas en el mismo.

5.4.6 IMPLEMENTACIÓN DEL FORMATO DE LA TRAMA MP3

La trama MP3 está constituida por el encabezado, los datos principales y la información secundaria, para la implementación de esta etapa se emplean los siguientes archivos de Matlab:

- `encabezado.m`
- `datos_principales.m`
- `info_secundaria.m`

El archivo encabezado.m permite escribir los 32 bits correspondientes al header de la trama MP3 dentro de una cadena binaria.

Al inicio del encabezado se emplea una palabra de sincronismo de 12 bits, todos '1111 1111 1111' para identificar el comienzo de la trama, a continuación se tiene un bit de identificación del audio en éste caso en '1' para indicar que se trata de audio MPEG-1, luego se tienen 2 bits, en '01', para indicar que se usa la Capa III de MPEG-1 para la codificación del audio.

Como el objetivo del proyecto de titulación no es la implementación de un software comercial, sino más bien de un material académico de apoyo para futuras investigaciones en cuanto a procesamiento de audio digital se refiere, no se realiza la detección de errores, es por ello que el bit de protección es puesto en '1', para indicar que no hay redundancia CRC para detección de errores. A continuación se tienen 2 bits, en '00' para indicar que la frecuencia de muestreo es de 44.100 Hz. El bit de uso privado es puesto en '0' para indicar que no se usa, para reducir el tiempo de procesamiento, se decidió procesar y obtener archivos codificados monofónicos, es por ello que los 2 bits correspondientes al modo son puestos en '11', e indican que el modo de canal es monofónico, como no se emplea un extensión del modo los 2 bits correspondientes están en '00'. Como el presente proyecto está basado en un estándar internacional se ha considerado no poner derechos de autor al flujo de bits codificados, es decir, el bit de copyright está en '0', para indicar que el archivo MP3 es una copia. La opción de preénfasis no se utiliza, es por ello que los 2 bits, son puestos en '00'.

El archivo info_secundaria.m consta de 17 bytes por ser de modo monofónico, almacenados en una cadena binaria que consiste de cuatro partes: el puntero *main_data_begin*, información secundaria para ambos gránulos (*private_bits* y *scfsi*), información secundaria para el gránulo 0, e información secundaria del gránulo 1.

El archivo datos_principales.m contiene la información relacionada a los factores de escala y la codificación Huffman, referidos al estándar.

REFERENCIAS BIBLIOGRÁFICAS

CAPITULO 5: DISEÑO E IMPLEMENTACIÓN EN MATLAB DEL ALGORITMO MP3 PARA COMPRESIÓN DE AUDIO

LIBROS:

- [42] Joebert S. Jacaba, (2001). AUDIO COMPRESSION USING MODIFIED DISCRETE COSINE TRANSFORM: THE MP3 CODING STANDARD, Department of Mathematics, College of Science the University of the Philippines Diliman, Quezon City.
- [43] Salomon, D. (2007). Data Compression. Chapter 7, Audio Compression. Fourth Edition. London: Springer-Verlag.
- [44] ISO/IEC, (2002). MPEG Audio specification (ISO/IEC 11172-3) (2002), CD 11172-3 CODING OF MOVING PICTURES AND ASSOCIATED AUDIO FOR DIGITAL STORAGE MEDIA AT UP TO ABOUT 1.5 MBIT/s Part 3 AUDIO, General normative elements, Technical normative elements.
- [45] ISO/IEC, (2002). MPEG Audio specification (ISO/IEC 11172-3) (2002), Annex A (normative) Diagrams, Annex B (normative) Tables, Annex C (informative) The Encoding Process, Annex D (informative) Psychoacoustic Models, Annex E (informative) Bit Sensitivity to Errors, Annex F (informative) Error Concealment, Annex G (informative) Joint Stereo Coding.

PAPERS:

- [46] Pan, Davis. A tutorial on MPEG/Audio compression, IEEE Multimedia Journal, Vol. 2 No. 2 (Summer 1995).

INTERNET:

- [47] MATLAB, The MathWorks, Inc., (2010), Mathworks Homepage.
<http://www.mathworks.com>
- [48] MATPIC, (2010). Manual de Interfaz Gráfica de Usuario en Matlab.
<http://www.matpic.com/>

CAPÍTULO 6.

PRUEBAS DE FUNCIONAMIENTO Y RESULTADOS

6.1 INTRODUCCIÓN

Las pruebas de funcionamiento se dividen en dos partes, la primera parte de las pruebas consiste en el análisis del comportamiento y funcionamiento del codificador y la segunda parte consiste en el análisis de los archivos codificados.

Para efectuar las pruebas de funcionamiento del codificador y analizar el comportamiento del mismo, se ha procedido a codificar varios archivos de audio de diferente duración y diferente ritmo. Al codificar los distintos archivos se obtienen diferentes características producto de la codificación, dicha información junto con la información de los archivos no codificados (originales) sirven para elaborar tablas las cuales muestran en forma cuantitativa el comportamiento del codificador, además, con la información recolectada para elaborar las tablas se realizan diagramas que nos indican de forma más didáctica el comportamiento y tendencia del codificador.

Para la segunda parte de las pruebas, es decir, para el análisis de los archivos codificados, se comparan las señales obtenidas con el codificador, con las señales que se obtienen empleando programas de audio comerciales (Cool Edit Pro 2.0 y GoldWave 5.2), con la ayuda de estos programas se comprueban las características del archivo codificado, en el dominio del tiempo y en el dominio de la frecuencia.

Además utilizando los programas mencionados, se comprueba la veracidad y validez del archivo e incluso se visualiza la reproducibilidad (que el archivo se pueda abrir y reproducir) del archivo comprimido.

6.2 PRUEBAS DE FUNCIONAMIENTO DEL CODIFICADOR

6.2.1 INTRODUCCIÓN

6.2.1.1 Especificaciones de hardware

El programa ha sido elaborado, diseñado y ejecutado en una computadora notebook Compaq Presario C706LA, de las siguientes características:

- Procesador: AMD Turion (tm) 64, Mobile Technology MK38, 2.21 [GHz].
- RAM: 1024 MB, DDR2 667 [MHz].
- Disco duro: 120 GB, enhanced IDE 5400 RPM SATA.
- Tarjeta gráfica: Nvidia GeForce Go 1600, 256 MB.

Debido a que el codificador demanda una alta utilización de recursos del CPU, procesador y memoria (ver anexo A) en el momento de la codificación, para poder tener un rendimiento aceptable es recomendable correr el programa en una computadora que por lo menos conste de un procesador Pentium IV y 512 MB de RAM.

6.2.1.2 Requerimientos de software

El codificador ha sido desarrollado en software de simulación MatLab 7.8 (R2009a), corriendo en un sistema operativo Windows XP SP3 de 32 bits. En cuanto a los toolboxes o librerías empleados, el codificador necesita los siguientes toolboxes: Matlab Toolbox™, Signal Processing Toolbox™, Communications Toolbox™, MP3 Toolbox™ y Guide Toolbox™. Como el programa ha sido desarrollado en un ambiente de Microsoft Windows, el codificador es compatible con Windows XP, Windows Vista y Windows 7.

Para la verificación de los resultados (archivos codificados) se emplean dos programas especializados en el tratamiento y procesamiento del audio, estos programas son: Cool Edit 2.0 y GoldWave 5.2.

6.2.2 FUNCIONAMIENTO DEL CODIFICADOR

El ejemplo que se indica a continuación no pretende explicar de manera detallada la utilización y el funcionamiento del codificador, al contrario se hace una breve y sencilla descripción de la utilización del codificador, esto se lo hace con el objetivo de mostrar cómo y de dónde se obtiene la información utilizada en la elaboración de las diferentes tablas de las características del codificador, las mismas que serán mostradas posteriormente. Para entender de mejor manera y en detalle el funcionamiento y utilización del codificador se recomienda ver el Anexo A.

Una vez que se ha tenido acceso al codificador, la primera ventana que se visualiza en la pantalla es la ventana “MPEG1_LayerIII” (ventana de la carátula).



Figura 6. 1: Primera ventana del codificador (Ventana de la carátula)

Al seleccionar la opción “CONTINUAR” de la ventana “MPEG1_LayerIII” se tiene acceso a la ventana “Codificador_MP3” como se observa en la figura 6.2

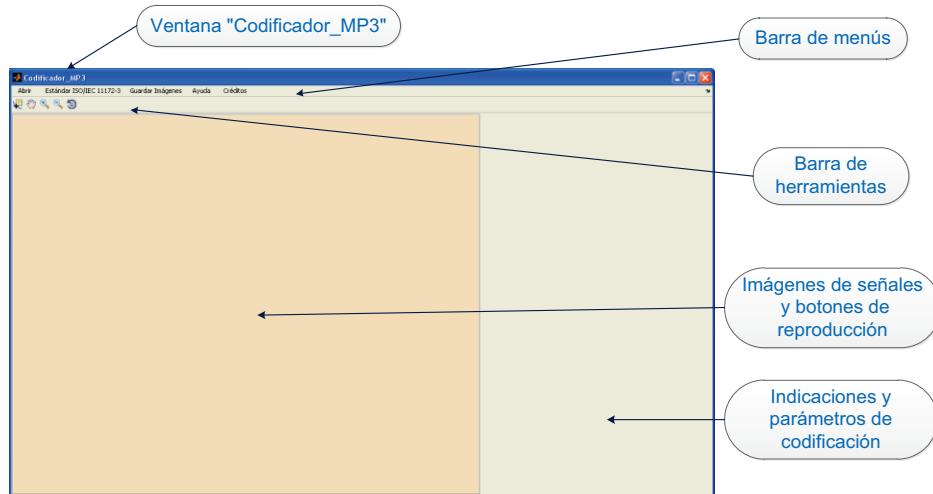


Figura 6. 2: Segunda ventana del codificador, ventana “Codificador_MP3”

Como se observa en la figura 6.2, inicialmente la ventana “Codificador_MP3” aparece vacía, porque aún no se ha seleccionado ningún archivo de audio y no se ha procesado señal alguna. Para poder abrir un archivo de audio deseado, de la barra de menús se selecciona el menú “Abrir” y luego la opción “Archivos .WAV”.

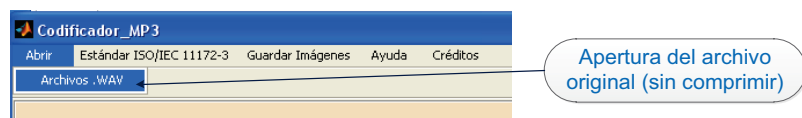


Figura 6. 3: Opción para abrir archivo sin comprimir *.wav

Luego de seleccionar la opción “Archivos .WAV” del menú “Abrir” se despliega una ventana de búsqueda, la misma que permite ubicar el directorio, seleccionar y abrir el archivo de audio con el que se requiera trabajar, tal como se indica en la figura 6.4.

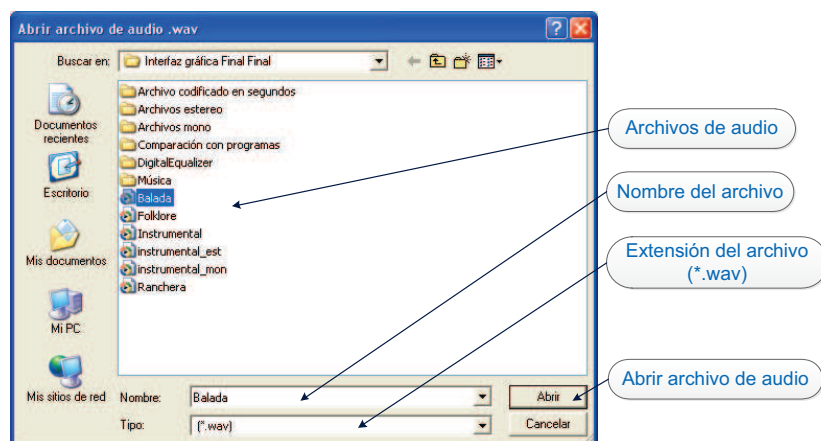


Figura 6. 4: Selección y apertura del archivo *.wav

Al abrir el archivo de audio, empieza el procesamiento de la señal, y progresivamente van apareciendo las imágenes de las señales procesadas; al final la ventana lucirá como se muestra en la figura 6.5.

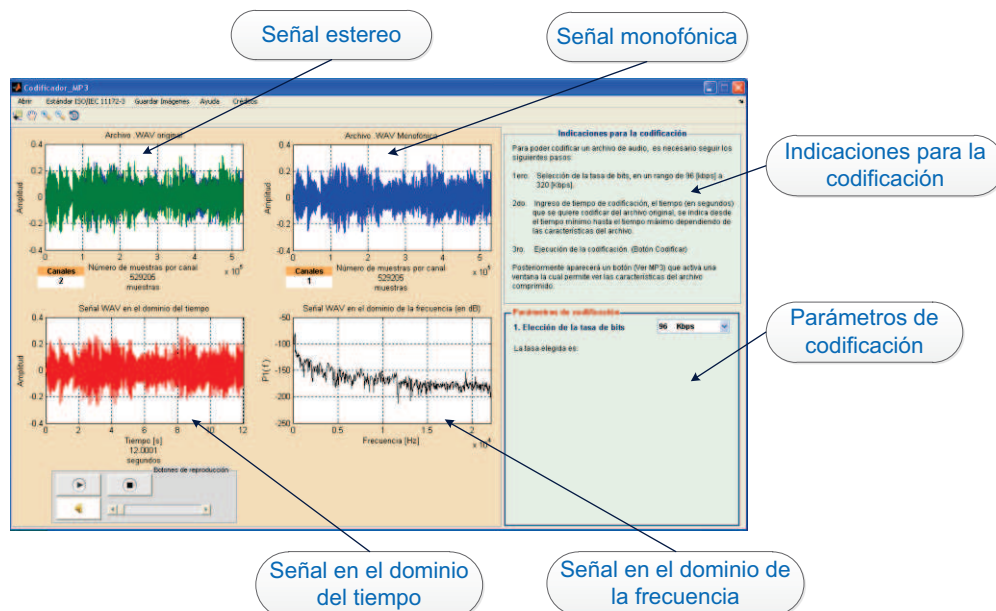


Figura 6. 5: Señales de archivo de audio de entrada, indicaciones y parámetros de codificación

Hasta este punto, la información que será utilizada posteriormente es la de las imágenes de las señales en el dominio del tiempo y de la frecuencia.

El paso posterior para continuar con la utilización del codificador, es el ingreso de los parámetros de codificación. El primer parámetro a seleccionar es la tasa de bits, en un rango de 96 Kbps a 320 Kbps, en el caso del ejemplo en particular la elección ha sido de 128 Kbps.

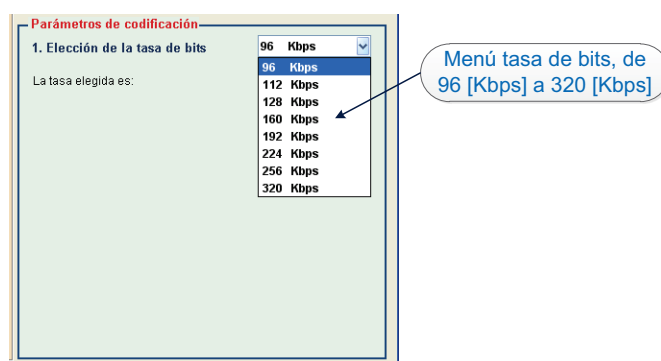


Figura 6. 6: Elección de la tasa de bits

Como se muestra en la figura 6.7, al seleccionar la tasa de bits no solo se activa una caja de texto de edición, la cual permite el ingreso del tiempo requerido, sino que además se despliega información de los tiempos mínimos y máximos que pueden ser codificados. En el caso del ejemplo que se está indicando y para todos los demás archivos con los que se elaboran las tablas que describen el comportamiento del codificador, el tiempo utilizado es el máximo.

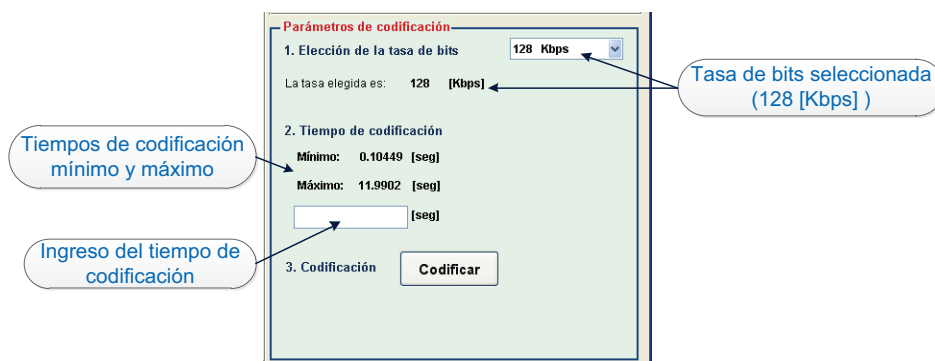


Figura 6. 7: Parámetros mostrados después de seleccionar la tasa de bits

Después de seleccionar la tasa de bits se ingresa el tiempo que se desea codificar, es decir cuantos segundos del archivo original (no codificado) van a ser codificados, como se mencionó para éste y los demás casos el tiempo ingresado será el máximo (se codificará todo el archivo original).

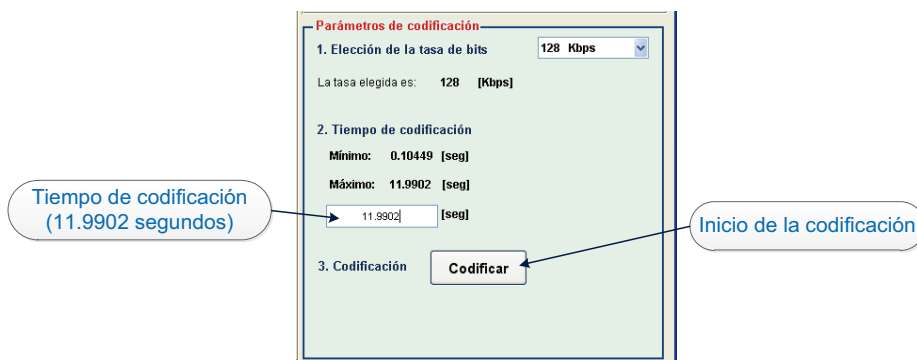


Figura 6. 8: Ingreso del tiempo que se desea codificar del archivo original (no codificado)

La duración total del archivo y el tiempo que se desea codificar son datos que van a ser mostrados en las tablas como se verá más adelante. Una vez que se han ingresado los parámetros para codificar el archivo e iniciar el proceso de la codificación se selecciona el botón “Codificar”.

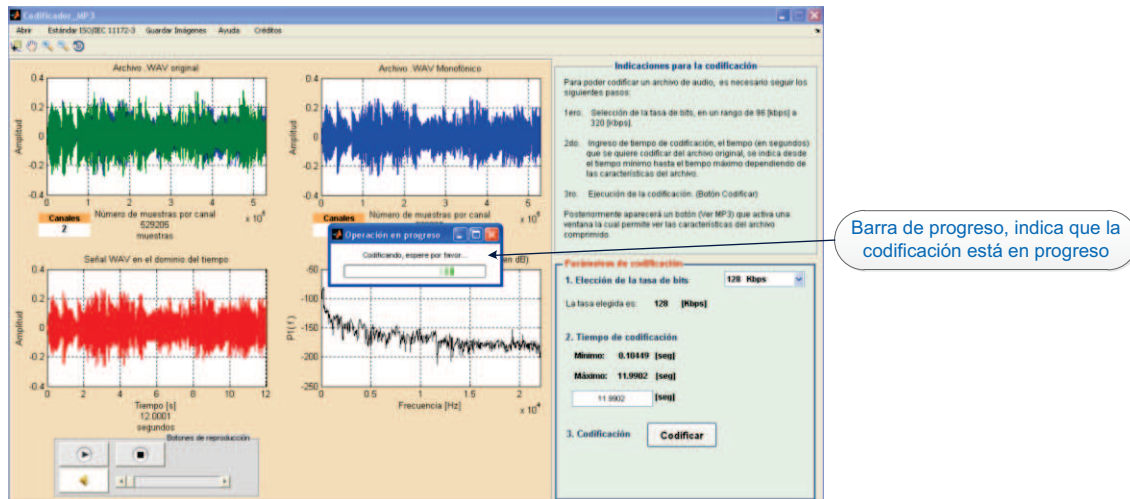


Figura 6. 9: Ejecución de la codificación

Como se indica en la figura 6.9 al iniciarse la codificación se visualiza, el progreso sistemático de la codificación (barra de progreso), pero paralelamente a esto, se inicia un cronómetro el cual permite visualizar el tiempo que se tardó el archivo en ser codificado, este tiempo también será utilizado en la elaboración de la tablas que describen el comportamiento del codificador.

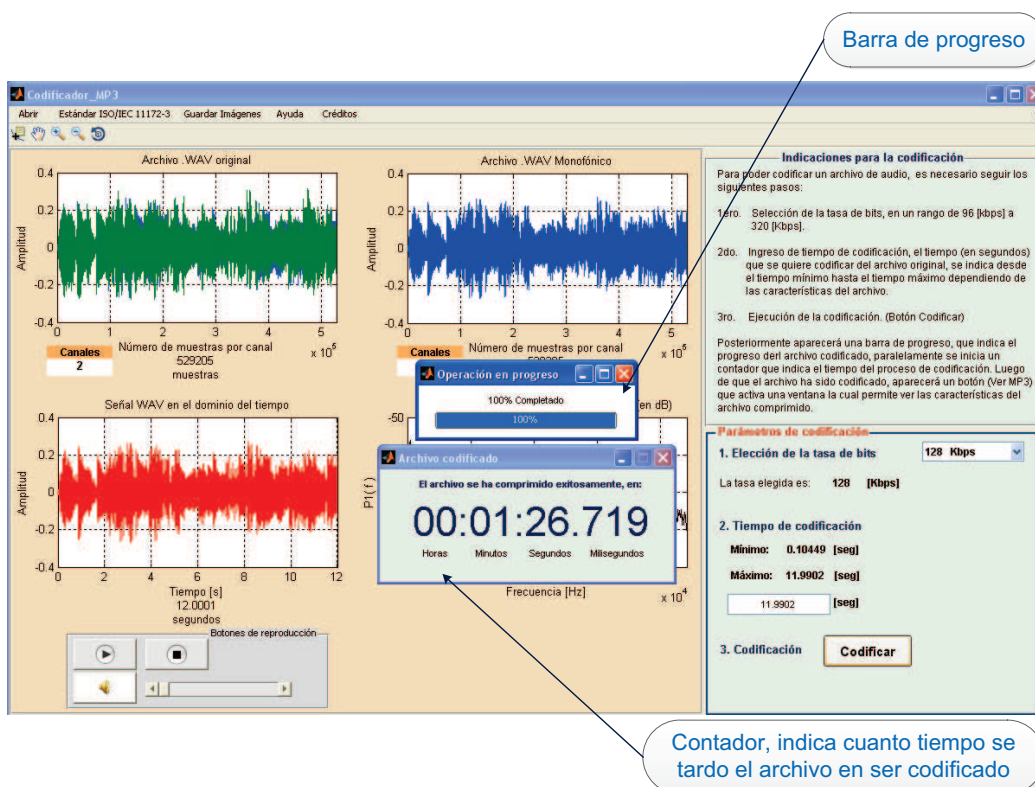


Figura 6. 10: Finalización del proceso de codificación

Cabe mencionar que el cronómetro funciona en modo asíncrono, esto significa que si la carga del procesador es alta, el cronómetro trabaja en segundo plano y no se visualiza el cronómetro en la pantalla, pero si se libera la carga del procesador la ventana del cronómetro se despliega, es por ello que el cronómetro solo se despliega cuando el archivo está casi por terminar su codificación porque en dichos instantes la carga del procesador no es tan alta.

La ventana perteneciente al cronómetro una vez que se ha finalizado la codificación, no solo nos indica el tiempo del proceso de codificación sino que además muestra un mensaje que indica que el archivo de audio ha sido codificado exitosamente, es decir el proceso de codificación ha completado el cien por ciento, sin errores.

Instantes posteriores a la finalización del proceso de codificación, se activa en la parte inferior derecha de la ventana “Codificador_MP3” un botón “Ver MP3” el cual al activarlo da paso a la tercera ventana del codificador, es decir la ventana “Archivo_MP3”.

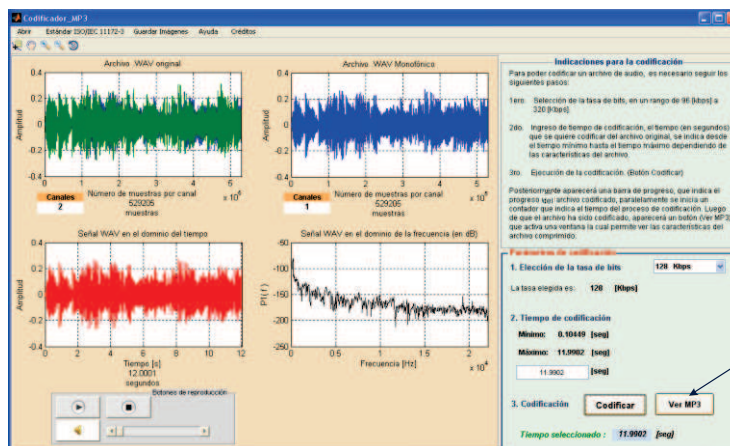


Figura 6. 11: Activación del botón “Ver MP3” para activar la ventana “Archivo_MP3”

Como se puede observar en la figura 6.12, la ventana inicialmente aparece vacía, en esta ventana a diferencia de la ventana “Codificador_MP3” no es necesario ingresar algún parámetro, lo único que se necesita es seleccionar el archivo resultante del proceso de codificación, es decir, el archivo comprimido (*.mp3).

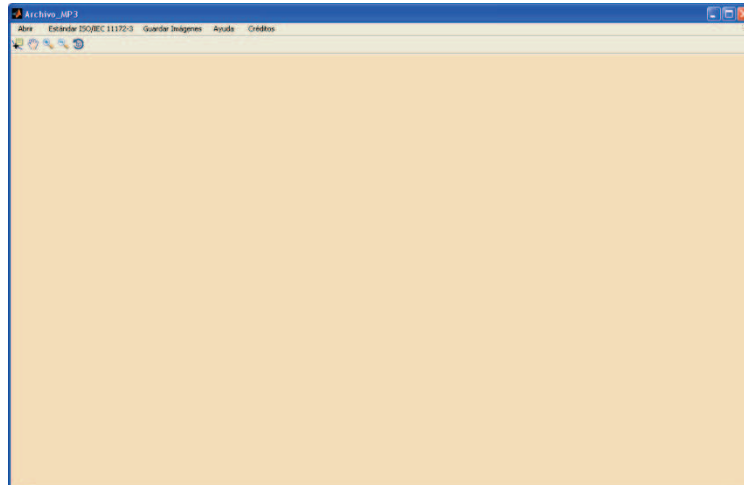


Figura 6. 12: Ventana “Archivo_MP3”

Para seleccionar y abrir el archivo comprimido se escoge la opción “Archivos .MP3” del menú “Abrir”, figura 6.13.

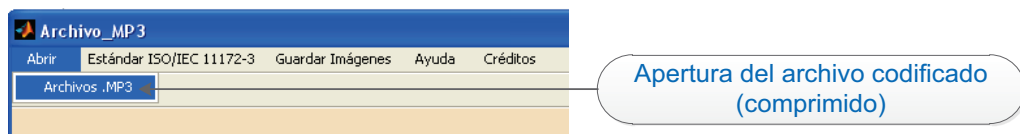


Figura 6. 13: Opción para abrir archivo codificado *.MP3

Al seleccionar la opción “Archivos .MP3” del menú “Abrir” se despliega una ventana, como la de la figura 6.14, la cual permite buscar, seleccionar y abrir el archivo codificado.

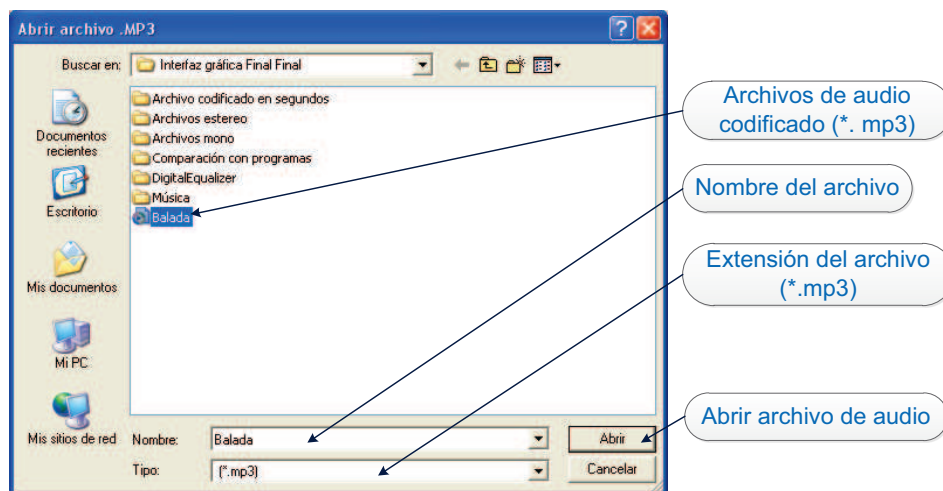


Figura 6. 14: Selección y apertura del archivo del archivo codificado (*.mp3)

Una vez que se abre el archivo comprimido, inicia el procesamiento digital de las señales correspondientes al archivo codificado; dichas señales junto con varias características del archivo codificado y un cuadro de resumen del proceso de codificación se muestran en la ventana "Archivo_MP3".

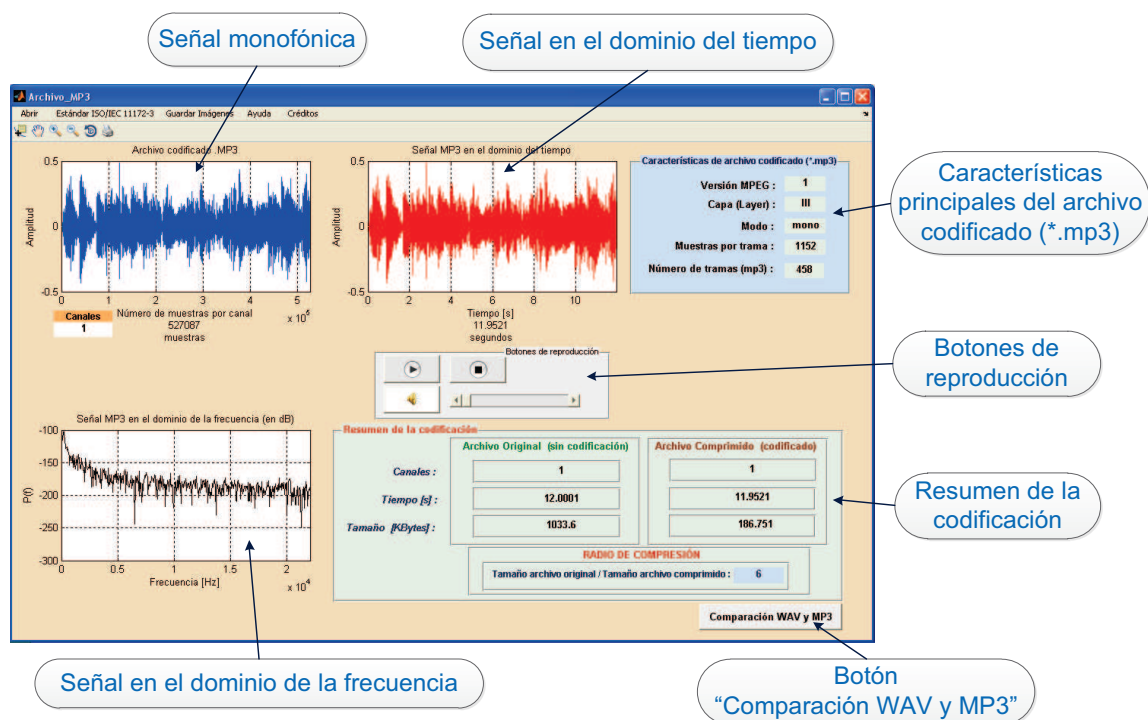


Figura 6. 15: Características del archivo codificado

En el cuadro de resumen como se visualiza en la figura 6.16, se indican características del archivo original (*.wav) y características del archivo comprimido (*.mp3), estas características también son utilizadas para elaborar las tablas de las características del codificador.

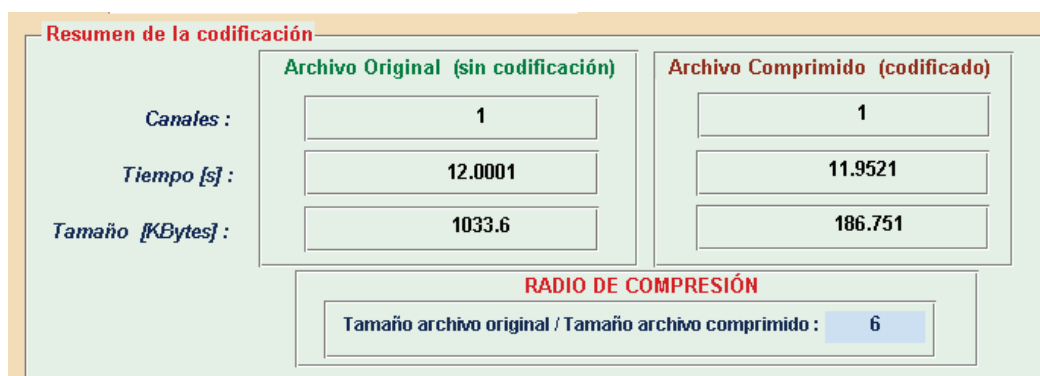


Figura 6. 16: Resumen de la codificación

En la ventana “Archivo_MP3” una vez que se ha procesado todas las señales y se han mostrado todos los cuadros de información aparece, en la parte inferior derecha, un botón “Comparación WAV y MP3”, el cual al ser seleccionado da paso a la ventana “Comparacion_WAV_MP3”.

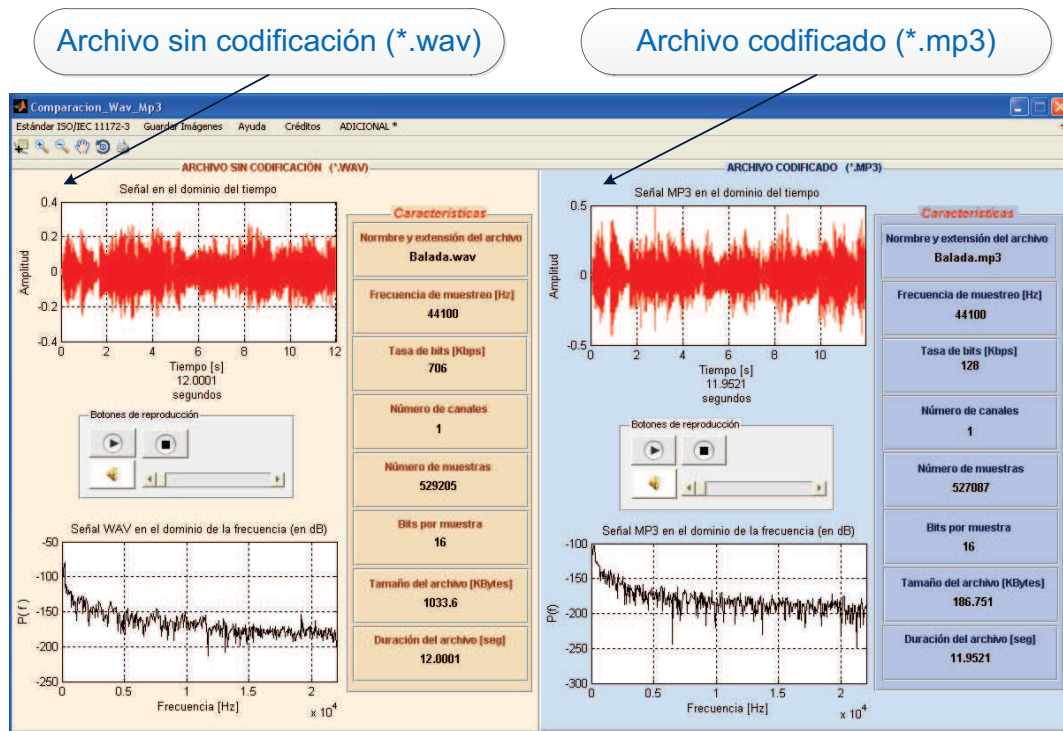


Figura 6. 17: Comparación de las características de los archivos sin codificación (izquierda) y con codificación (derecha)

En esta ventana se muestra información detallada de las características tanto del archivo original (*.wav) como del archivo codificado (*.mp3), además de visualizar las características de los archivos, es posible visualizar los archivos de audio tanto en el dominio del tiempo como en el de la frecuencia. Al igual que en las ventanas anteriores en esta ventana se puede escuchar los archivos de audio (no codificado y codificado) con el uso de los botones de reproducción.

En cuanto a las características de los archivos, las que se muestran en la ventana “Comparación_Wav_Mp3” son: nombre del archivo, extensión del archivo de audio, frecuencia de muestreo, tasa de bits, número de canales, número de muestras, número de bits por muestra, tamaño del archivo y duración del archivo.

6.2.3 CARACTERÍSTICAS DE LOS ARCHIVOS CODIFICADOS

Una vez que se ha codificado el archivo requerido, con todas las tasas de bits disponibles y se ha obtenido información del tiempo que se demora el archivo en ser codificado, tamaños de los archivos original (sin compresión) y codificado (comprimido), todos estos datos son distribuidos y presentados en tablas, las cuales se muestran más adelante.

En cuanto a los archivos que han sido codificados se ha trabajado con archivos monofónicos y se han empleado cinco ritmos los cuales son: folklore, ranchera, rock, balada e instrumental.

Si bien el codificador está diseñado para poder trabajar con archivos en modo estéreo o monofónico, por facilidad en los cálculos y ahorro de recursos y tiempo de procesamiento, el tratamiento del archivo de audio se lo realiza en modo monofónico únicamente.

El incremento de la duración de los archivos monofónicos es cada segundo, empezando con tres segundos en el ritmo folklore y terminando en siete segundos en el ritmo instrumental.

6.2.3.1 Archivos de audio

6.2.3.1.1 Ritmo Folklore

ARCHIVO ORIGINAL .WAV	
Ritmo	Folklore
Modo	Mono
Tamaño archivo original .WAV [Kbytes]	258,473
Duración del archivo original [seg]	3

Tabla 6. 1: Características del archivo monofónico sin codificación, ritmo Folklore

Tiempo seleccionado para codificar: 3 segundos (tamaño total del archivo)			
ARCHIVO CODIFICADO .MP3			
Tasa de bits [Kbps]	Tiempo del proceso de codificación [minutos]	Tamaño del archivo codificado .MP3 [Kbytes]	Radio de compresión
96	0,21	34,76	7
112	0,22	40,55	6
128	0,24	46,34	6
160	0,26	57,93	4
192	0,29	69,51	4
224	0,30	81,10	3
256	0,32	92,69	3
320	0,34	115,86	2

Tabla 6. 2: Características de los archivos codificados, ritmo Folklore

6.2.3.1.2 Ritmo Ranchera

ARCHIVO ORIGINAL .WAV	
Ritmo	Ranchera
Modo	Mono
Tamaño archivo original .WAV [Kbytes]	344,533
Duración del archivo original [seg]	4

Tabla 6. 3: Características del archivo monofónico sin codificación, ritmo Ranchera

Tiempo seleccionado para codificar: 4 segundos (tamaño total del archivo)			
ARCHIVO CODIFICADO .MP3			
Tasa de bits [Kbps]	Tiempo del proceso de codificación [minutos]	Tamaño del archivo codificado .MP3 [Kbytes]	Radio de compresión
96	0,30	46,39	7
112	0,33	54,12	6
128	0,34	61,85	6
160	0,37	77,32	4
192	0,40	92,78	4
224	0,44	108,24	3
256	0,46	123,71	3
320	0,50	154,63	2

Tabla 6. 4: Características de los archivos codificados, ritmo Ranchera

6.2.3.1.3 Ritmo Rock

ARCHIVO ORIGINAL .WAV	
Ritmo	Rock
Modo	Mono
Tamaño archivo original .WAV [Kbytes]	430,707
Duración del archivo original [seg]	5

Tabla 6. 5: Características del archivo monofónico sin codificación, ritmo Rock

Tiempo seleccionado para codificar: 5 segundos (tamaño total del archivo)			
ARCHIVO CODIFICADO .MP3			
Tasa de bits [Kbps]	Tiempo del proceso de codificación [minutos]	Tamaño del archivo codificado .MP3 [Kbytes]	Radio de compresión
96	0,39	58,33	7
112	0,41	68,05	6
128	0,43	77,77	6
160	0,48	97,21	4
192	0,53	116,66	4
224	0,55	136,10	3
256	0,58	155,54	3
320	0,65	194,43	2

Tabla 6. 6: Características de los archivos codificados, ritmo Rock

6.2.3.1.4 Ritmo Balada

ARCHIVO ORIGINAL .WAV	
Ritmo	Balada
Modo	Mono
Tamaño archivo original .WAV [Kbytes]	516,826
Duración del archivo original [seg]	6

Tabla 6. 7: Características del archivo monofónico sin codificación, ritmo Balada

Tiempo seleccionado para codificar: 6 segundos (tamaño total del archivo)			
ARCHIVO CODIFICADO .MP3			
Tasa de bits [Kbps]	Tiempo del proceso de codificación [minutos]	Tamaño del archivo codificado .MP3 [Kbytes]	Radio de compresión
96	0,48	69,66	7
112	0,51	81,26	6
128	0,54	92,87	6
160	0,61	116,09	4
192	0,67	139,31	4
224	0,73	162,53	3
256	0,79	185,75	3
320	0,87	232,18	2

Tabla 6. 8: Características de los archivos codificados, ritmo Balada

6.2.3.1.5 Ritmo Instrumental

ARCHIVO ORIGINAL .WAV	
Ritmo	Instrumental
Modo	Mono
Tamaño archivo original .WAV [Kbytes]	603
Duración del archivo original [seg]	7

Tabla 6. 9: Características del archivo monofónico sin codificación, ritmo Instrumental

Tiempo seleccionado para codificar: 7 segundos (tamaño total del archivo)			
ARCHIVO CODIFICADO .MP3			
Tasa de bits [Kbps]	Tiempo del proceso de codificación [minutos]	Tamaño del archivo codificado .MP3 [Kbytes]	Radio de compresión
96	0,57	81,59	7
112	0,61	95,19	6
128	0,66	108,79	6
160	0,73	135,99	4
192	0,82	163,19	4
224	0,90	190,39	3
256	0,97	217,58	3
320	1,09	271,98	2

Tabla 6. 10: Características de los archivos codificados, ritmo Instrumental

6.2.3.2 Análisis de las características del codificador

A pesar de que parezcan redundantes los datos mostrados en las tablas de la 6.1 a la 6.10, dichos datos muestran que el codificador puede desempeñarse de una manera adecuada con varios ritmos y a diferentes tasas de bits, es decir, el codificador puede desempeñarse en un gran campo de acción.

Al analizar los resultados obtenidos en las tablas, para los archivos de audio codificados se puede mencionar que la tasa de bits es un factor importante en cuanto a la compresión se refiere, ya que codificando el tiempo total de los archivos, a mayor tasa de bits, mayor demora en el proceso de codificación, mayor tamaño del archivo codificado y menor radio de compresión; en tanto con la tasa de bits mínima, el tiempo en codificarse el archivo es el menor, el tamaño del archivo codificado es menor y el radio de compresión es el más alto.

Claramente como se muestra en cada una de las tablas de la 6.1 a la 6.10 el mayor radio de compresión se lo alcanza al seleccionar una tasa de 96 Kbps; el mayor radio de compresión para los archivos de audio fue de siete.

En cuanto a los radios de compresión obtenidos, se puede concluir que la codificación es bastante adecuada. Recordando que la idea original de Karlheinz Brandenburg y varios científicos del Instituto Tecnológico de Fraunhofer, era transmitir por el canal telefónico de 64 Kbps la información de un CD de audio, con una frecuencia de muestreo de 44.100 Hz, 16 bits por muestra y en modo estéreo (2 canales) dando una tasa de bits de 1,411 Mbps como se muestra en la ecuación 6.1.

$$Tasa\ de\ bits\ CD = 44100 \frac{muestras}{segundo} \times 16 \frac{bits}{muestra} \times 2\ canales = 1411200\ bps \quad (Ec.\ 6.1)$$

Tamaño aproximado = 1500 kilobits en 1 segundo

Recordando que el sistema telefónico es un sistema monofónico la tasa que se debería comparar con la tasa de 64 Kbps, debería ser la descrita en la ecuación 6.2.

$$Tasa\ de\ bits\ CD = 44100 \frac{muestras}{segundo} \times 16 \frac{bits}{muestra} \times 1\ canales = 705600\ bps \quad (Ec.\ 6.2)$$

Tamaño aproximado = 705 kilobits en 1 segundo

Al dividir el tamaño del archivo monofónico con calidad CD de 705 kilobits con el tamaño aproximado que se podría transmitir por el canal telefónico de 64 kilobits, el resultado es el que se lo había planteado Karlheinz Brandenburg en el Instituto Tecnológico de Fraunhofer y el cual es expresado cuantitativamente en la ecuación 6.2.

$$Radio\ de\ compresión = \frac{705,6\ kilobits}{64\ kilobits} = 11,025 \quad (Ec.\ 6.3)$$

Radio de compresión aproximadamente = 11

Debido a esta razón al utilizar la menor tasa de bits disponible en el codificador de 96 Kbps (aproximadamente 96 kilobits en 1 segundo) y compararlo con el audio en calidad CD monofónico de 705 Kbps (aproximadamente 705 kilobits en 1 segundo) el radio de compresión que se obtiene, es el que se indica en la ecuación 6.4.

$$Radio\ de\ compresión = \frac{705\ kilobits}{96\ kilobits} = 7,35 \quad (Ec.\ 6.4)$$

Radio de compresión aproximadamente = 7

Con los resultados mostrados en las tablas de la 6.1 a la 6.10 se puede concluir que la relación de compresión del codificador implementado en Matlab y el expuesto en el estándar ISO/IEC 11172-3 es el mismo.

6.2.4 COMPORTAMIENTO DEL CODIFICADOR

Para visualizar el comportamiento del codificador se ha utilizado los datos del archivo original (no codificado) y del archivo codificado (comprimido) correspondientes al ritmo balada, en los gráficos mostrados se vinculan varios parámetros y se observa la dependencia de los diferentes parámetros presentes en el codificador, así mismo con la ayuda de la función “agregar línea de tendencia” de Microsoft Excel se despliega información de la tendencia de los datos incluyendo la ecuación matemática que más se aproxima a la distribución de datos empleados para hacer los diagramas de los parámetros y características del codificador.

ARCHIVO ORIGINAL .WAV	
Ritmo	Balada
Modo	Mono
Tamaño archivo original .WAV [Kbytes]	1033,61
Duración del archivo original [seg]	12

Tabla 6. 11: Características del archivo original (no codificado), ritmo balada

Tiempo seleccionado para codificar: 12 segundos (tamaño total del archivo)			
ARCHIVO CODIFICADO .MP3			
Tasa de bits [Kbps]	Tiempo del proceso de codificación [minutos]	Tamaño del archivo codificado .MP3 [Kbytes]	Radio de compresión
96	1,24	140,06	7
112	1,33	163,41	6
128	1,44	186,75	6
160	1,64	233,44	4
192	1,85	280,13	4
224	2,05	326,82	3
256	2,25	373,50	3
320	2,59	466,88	2

Tabla 6. 12: Características de los archivos codificados (comprimidos), ritmo balada

6.2.4.1 Tiempo del proceso de codificación vs. Tasa de bits

Como se aprecia en la figura 6.18 a mayor tasa de bits hay un incremento del tiempo en que se demora el archivo en ser codificado.

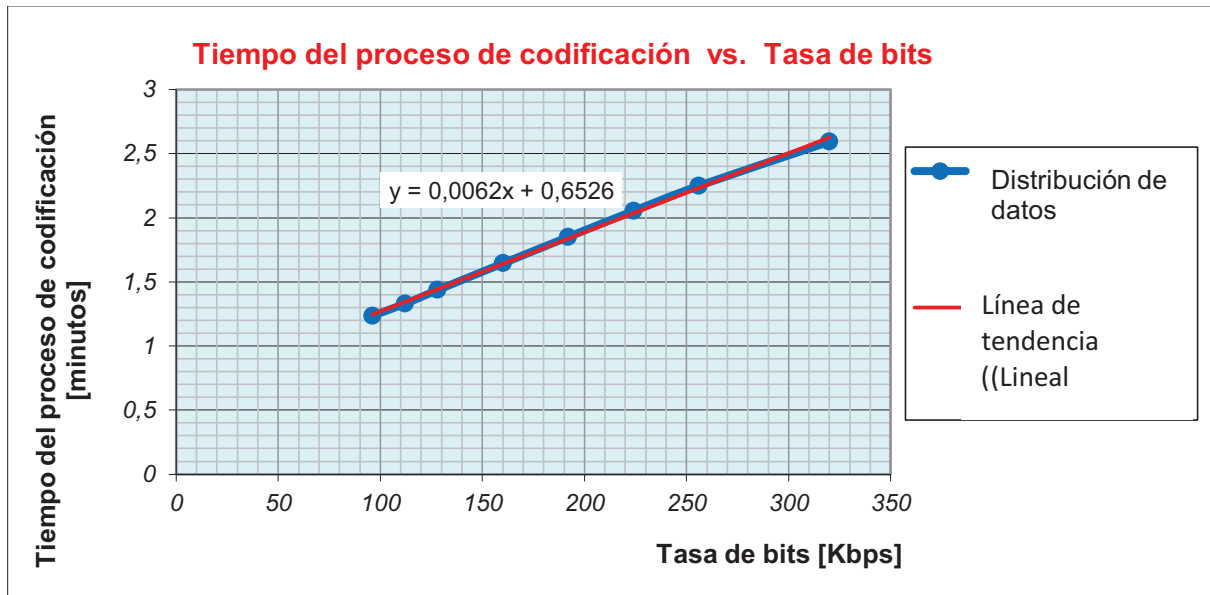


Figura 6. 18: Tiempo del proceso de codificación vs. Tasa de bits

6.2.4.2 Tamaño del archivo codificado vs. Tasa de bits

Si la tasa de bits incrementa como se observa en la figura 6.19, el tamaño del archivo codificado también se incrementa linealmente.

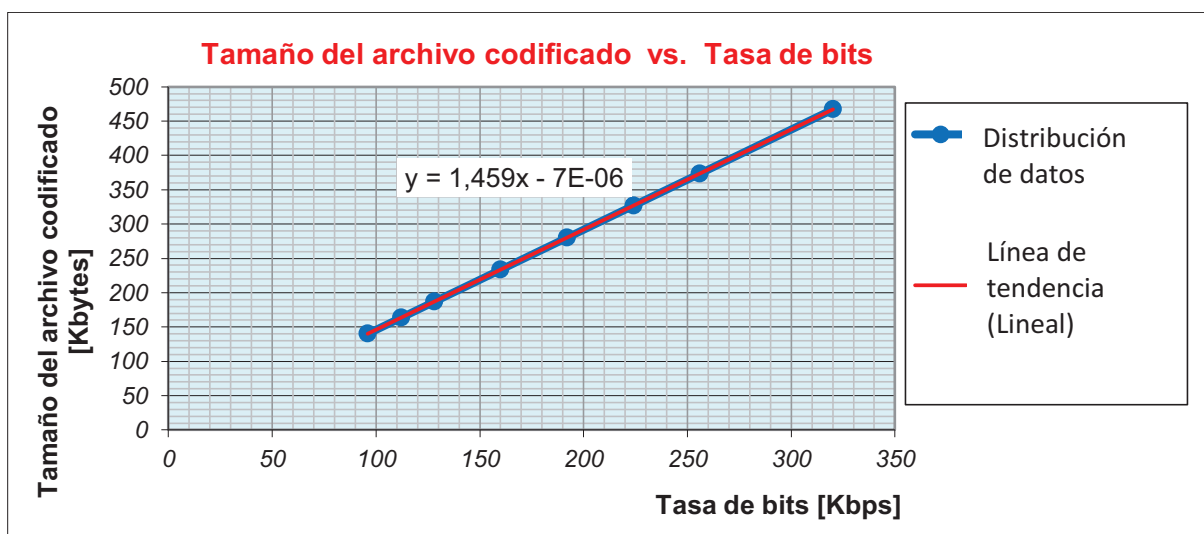


Figura 6. 19: Tiempo del archivo codificado vs. Tasa de bits

6.2.4.3 Radio de compresión vs. Tasa de bits

Al observar la figura 6.20 se distingue claramente que a menor tasa de bits se logra un mayor radio de compresión, si la tasa de bits va aumentando el radio de compresión va disminuyendo exponencialmente.

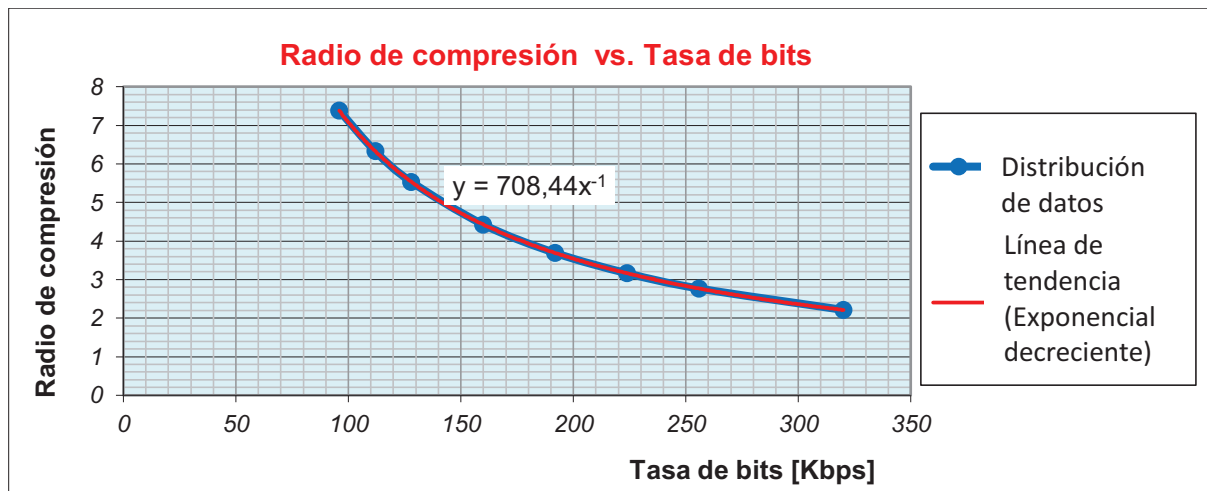


Figura 6. 20: Radio de compresión vs. Tasa de bits

Para observar otras características del codificador se utilizan los datos de las tablas 6.13 y 6.14, en este caso al efectuar el proceso de codificación el parámetro de ingreso correspondiente a la tasa de bits se mantiene fijo en 128 [Kbps], en su defecto el parámetro que varía es el tiempo del archivo original, al igual que en los anteriores casos se codifica todo el archivo de audio. Los archivos originales han sido editados y se van incrementando en un segundo, iniciando en un segundo y llegando a un máximo de doce segundos.

ARCHIVO ORIGINAL .WAV	
Ritmo	Balada
Modo	Mono
Tamaño archivo original .WAV [Kbytes]	1033,61
Duración total del archivo [seg]	12

Tabla 6. 13: Características del archivo original (no codificado).

Duración del archivo original sin comprimir [seg]	Tiempo del proceso de codificación [min]	Tamaño del archivo sin comprimir [Kbytes]	Tamaño del archivo comprimido [Kbytes]	Radio de compresión
1	0,09	86,22	14,91	6
2	0,18	172,31	30,42	6
3	0,25	258,47	46,34	6
4	0,35	344,55	61,85	6
5	0,46	430,73	77,77	6
6	0,56	516,80	92,87	6
7	0,69	602,93	108,79	6
8	0,83	689,07	124,71	6
9	0,97	775,22	139,81	6
10	1,12	861,39	155,73	6
11	1,28	947,54	171,65	6
12	1,45	1033,61	186,75	6

Tabla 6. 14: Características de los archivos codificados a 128 [Kbps]

6.2.4.4 Tamaño del archivo comprimido vs. Duración del archivo sin comprimir

Como se indica en la figura 6.21, lógicamente si se codifica un tiempo mayor el tamaño del archivo codificado será mayor, el tiempo crece linealmente con el tamaño del archivo codificado.

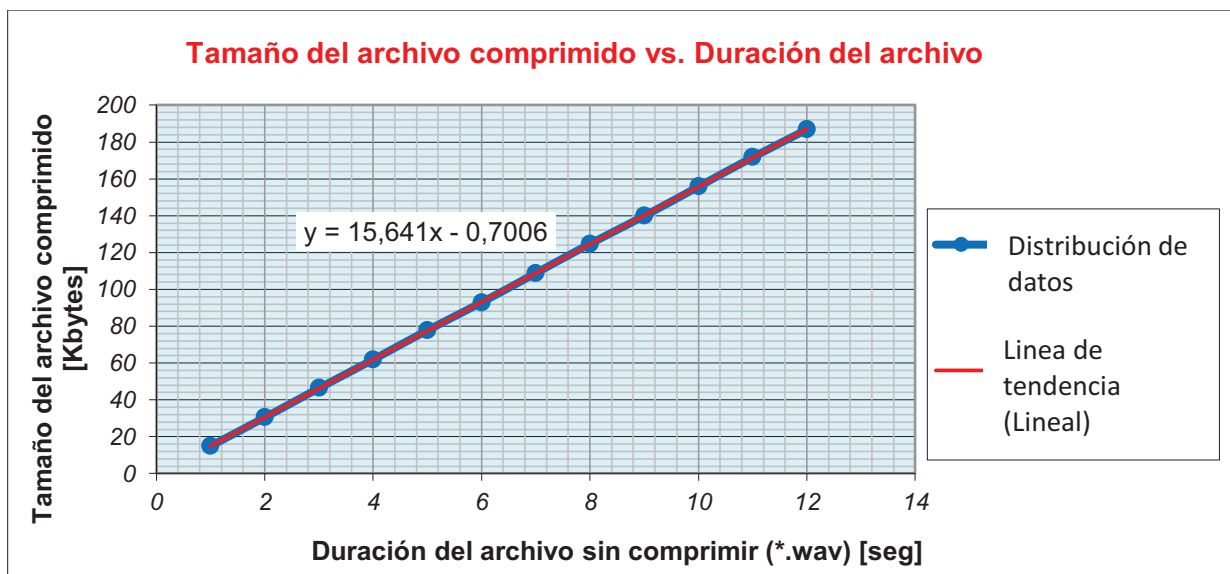


Figura 6. 21: Tamaño del archivo comprimido vs. Duración del archivo sin comprimir

6.2.4.5 Tamaño del archivo comprimido vs. Tamaño del archivo sin comprimir

Tanto el tamaño del archivo sin codificar con el tamaño del archivo codificado como se indica en la figura 6.22, van creciendo linealmente.

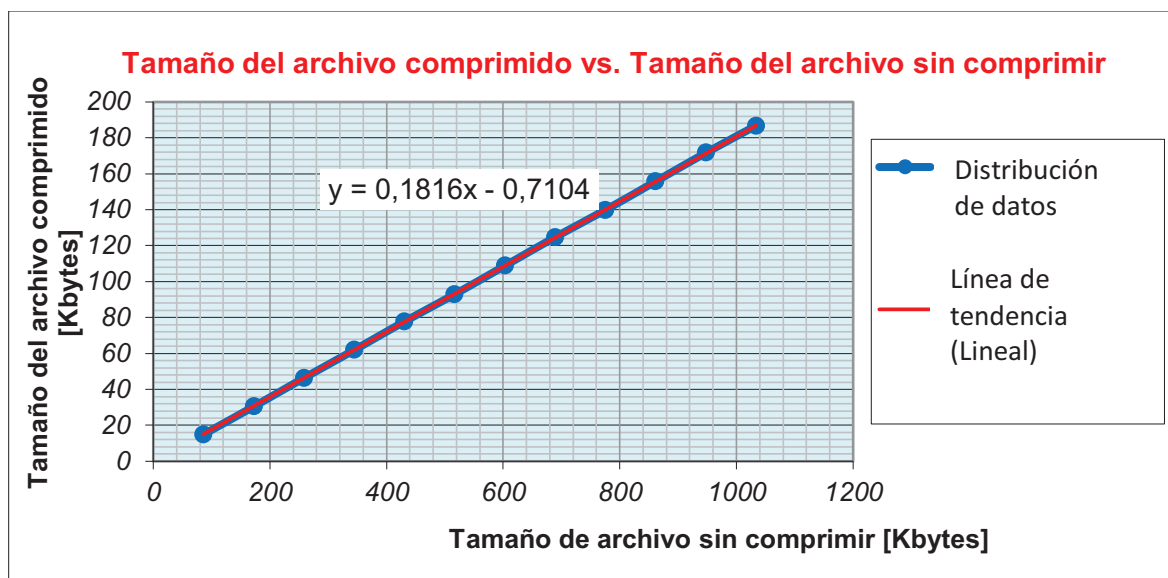


Figura 6. 22: Tamaño del archivo comprimido vs. Tamaño del archivo sin comprimir

6.3 ANÁLISIS DE RESULTADOS DE LOS ARCHIVOS CODIFICADOS

El análisis y verificación de los archivos codificados (archivos *.mp3) van desde comprobar que el archivo pueda ser reproducido por los programas Cool Edit 2.0 y GoldWave 5.2 hasta el análisis y comparación de las características de los archivos codificados en el dominio del tiempo y en el dominio de la frecuencia.

6.3.1 PROGRAMAS UTILIZADOS

Los programas empleados para verificar las características de los archivos codificados fueron básicamente dos, el principal Cool Edit versión 2.0 y el otro GoldWave 5.2, estos programas fueron seleccionados por su robustez y facilidad de utilización.

Los programas mencionados anteriormente fueron seleccionados para verificar la validez de los archivos codificados por que dichos programas son utilizados en ambientes comerciales de producción y edición musical a nivel profesional.

Con la ayuda de los programas se puede verificar las características del archivo tanto en el dominio del tiempo como en el de la frecuencia, así mismo se pueden comprobar características básicas como: el tiempo de duración del archivo, la tasa de bits seleccionada, el modo del archivo (monofónico o estereofónico), entre otras características importantes.

A continuación se hace una breve descripción de los dos programas y se muestra la ventana de trabajo principal de cada uno de ellos.

6.3.1.1 Cool Edit Pro 2.0

Cool Edit Pro 2.0 es un programa de edición de audio muy versátil, como todo programa, cuenta con barras de menús y barras de herramientas, que permite tener acceso a un gran número de efectos de audio y demás utilidades del programa. Cool Edit Pro además posee una interfaz gráfica de usuario bastante amigable, como se muestra en la figura 6.23.



Figura 6. 23: Ventana principal de Cool Edit Pro 2.0

Cool Edit Pro permite editar archivos de audio, se pueden eliminar segmentos de audio con precisiones de milésimas de segundo, añadir otros segmentos que no estaban en la muestra original, añadir efectos (ecos, reverberaciones, amplificar o atenuar la señal, eliminar ruidos, etc.).

Cool Edit Pro permite reproducir y grabar diversos tipos de archivos de audio, soporta más de 20 tipos de archivos diferentes. Incluso al grabar o cambiar el formato de un archivo y dependiendo de las características del formato seleccionado existe la posibilidad de elegir parámetros como: frecuencia de muestreo, número de bits de cuantización, entre otros.

6.3.1.2 GoldWave 5.2

GoldWave es un programa bastante sencillo de usar y más que suficiente para la edición de sonido multimedia, GoldWave permite manejar diversos formatos de audio, lo que lo hace una herramienta útil para un sin número de aplicaciones.



Figura 6. 24: Ventana principal de GoldWave 5.2

En GoldWave los archivos de sonido se presentan en una ventana típica de Microsoft Windows, como se visualiza en la figura 6.24. En la parte inferior de la

ventana de GoldWave existe una barra de estado de gran utilidad ya que muestra información general sobre el archivo activo y el estado del mismo.

6.3.2 PROPIEDADES Y REPRODUCCIÓN DE LOS ARCHIVOS CODIFICADOS

6.3.2.1 Propiedades de los archivos codificados

Conociendo con anterioridad que el programa convierte un archivo de audio de formato Windows PCM (*.wav) de frecuencia de muestreo $F_s = 44.100$ Hz en modo estéreo o monofónico, a un archivo de formato comprimido MPEG-1 Layer III (*.mp3) monofónico; en base a las especificaciones del estándar internacional ISO/IEC 11172-3, los resultados de la verificación de los archivos comprimidos se muestran a continuación.

Con la ayuda de Windows, se puede tener la primera verificación de la validez de los archivos de audio codificados, ya que al ubicar el cursor en el archivo de audio, instantes después de manera automática se despliega un resumen de la información del archivo, como se puede visualizar en la figura 6.25, los datos que comprueban que el archivo se ha codificado exitosamente son el campo tipo, el cual indica “Sonido en formato MP3” y el de la velocidad de transmisión (tasa de bits) que para este caso es de 128 Kbps.

La velocidad de transmisión (128 Kbps) la duración del archivo codificado (11 segundos), la extensión del archivo (*.mp3) y el tamaño del archivo (186 KB) indican que la codificación ha sido exitosa, ya que de lo contrario se tendrían datos erróneos o el archivo, resultaría ser de un formato desconocido o de un formato diferente al de mp3.

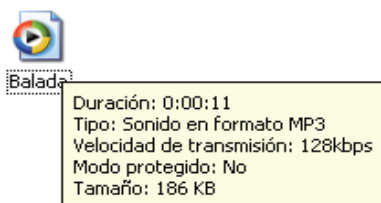


Figura 6. 25: Características de archivo de audio codificado (Explorador de Windows)

Para comprobar las características del archivo codificado empleando Cool Edit Pro 2.0 se procede a abrir el archivo de audio y al dar clic derecho en la zona de edición de la pista de audio, como se muestra en la figura 6.26, inmediatamente aparece una ventana la cual indica en detalle las características del archivo codificado. La información mostrada en esta ventana permite corroborar que el archivo ha sido comprimido exitosamente, y de manera lógica estas características deben ser las mismas que han sido elegidas en el proceso de codificación y además deben cumplir los lineamientos provistos en el estándar ISO/IEC 11172-3.

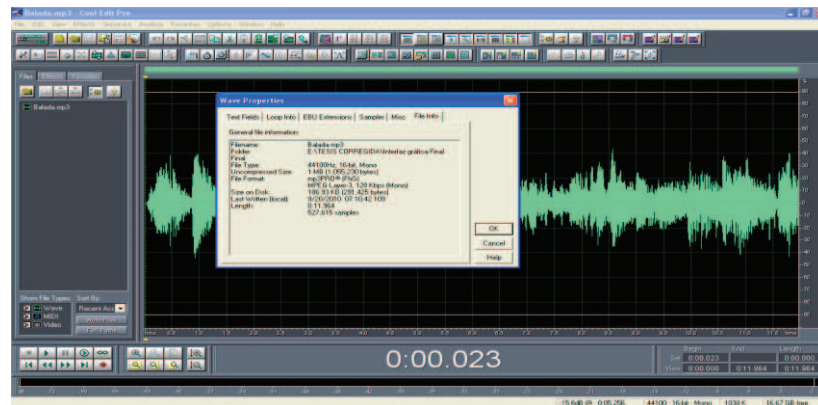


Figura 6. 26: Ventana principal de Cool Edit Pro 2.0 y venta de propiedades del audio

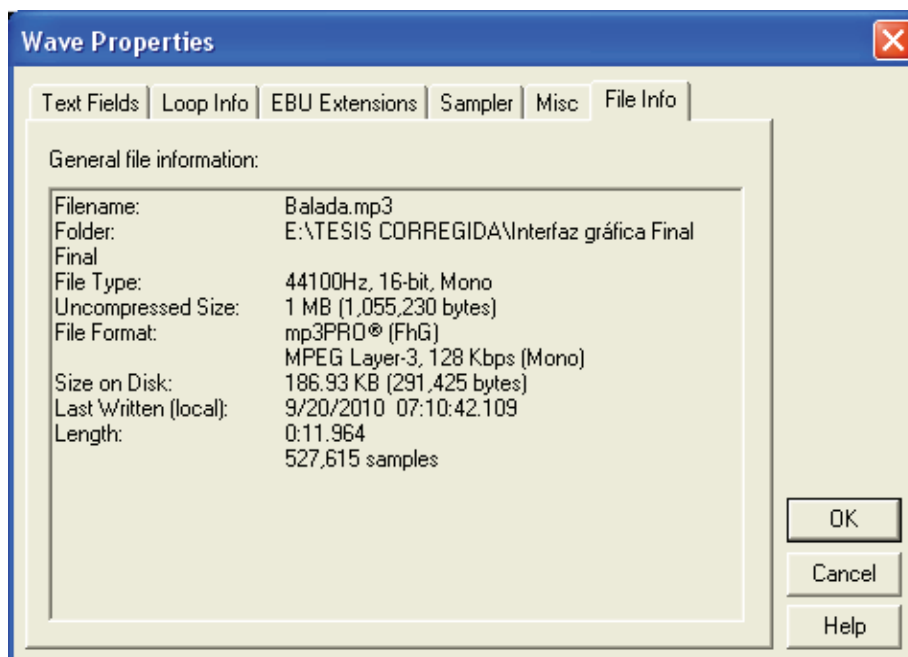


Figura 6. 27: Características del archivo de audio codificado (Programa: Cool Edit Pro 2.0)

Como indica la figura 6.27 los parámetros están de acuerdo a los resultados esperados es decir el archivo se ha codificado de manera exitosa, cumpliendo con requerimientos provistos en el estándar ISO/IEC 11172-3.

6.3.2.2 Reproducción de los archivos codificados

Para saber que el archivo codificado es capaz de ser reproducido en cualquier reproductor digital de audio o en cualquier software compatible con el formato *mp3, se ha procedido a abrir el archivo y a reproducirlo en Cool Edit Pro 2.0 y en GoldWave 5.2.

Como se aprecia en las figuras 6.28 y 6.29, el archivo puede ser reproducido normalmente, se observa que la línea de tiempo y los temporizadores se van incrementando, además en las barras de estado se indica que el archivo está siendo reproducido.

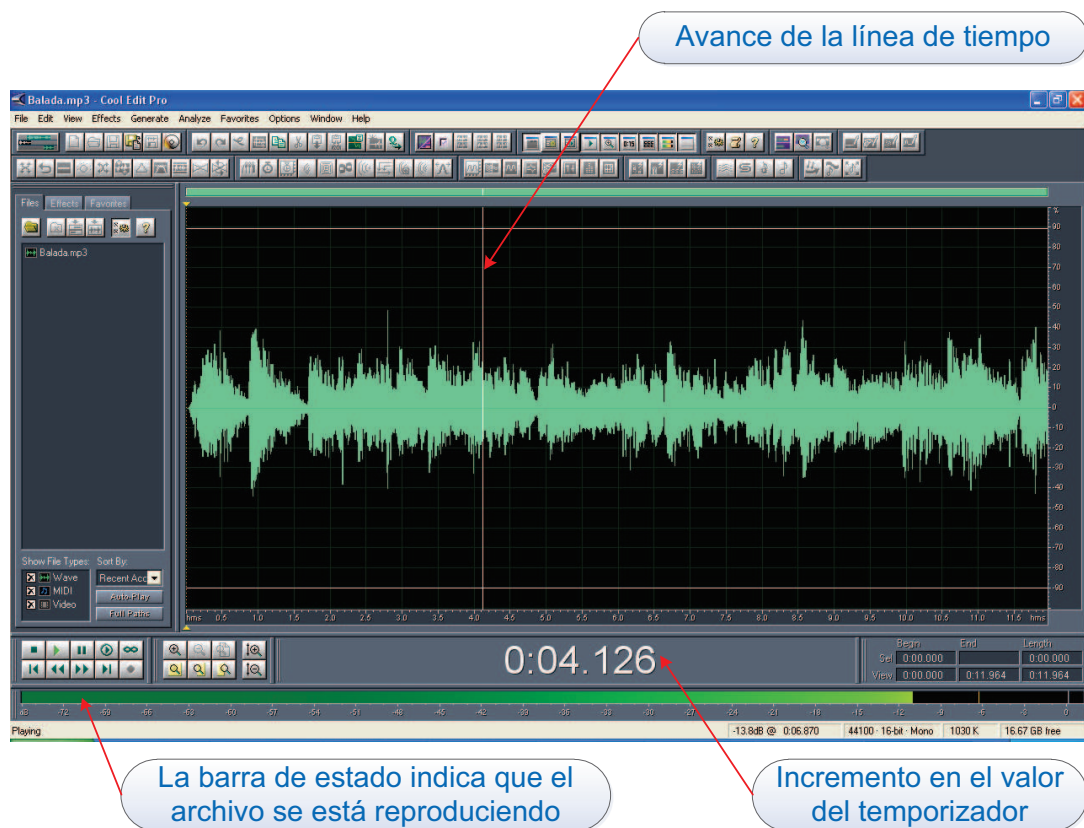


Figura 6. 28: Reproducción del archivo de comprimido (Programa: Cool Edit Pro 2.0)

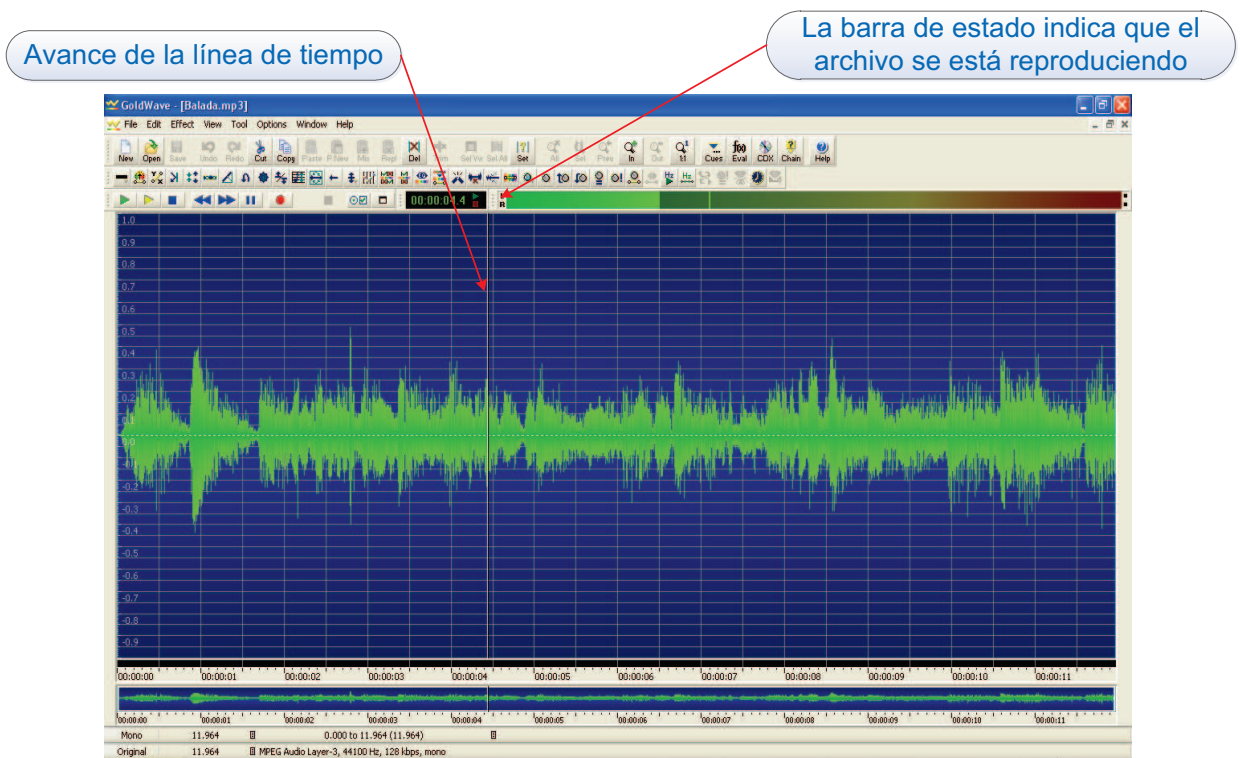


Figura 6. 29: Reproducción del archivo de comprimido (Programa: GoldWave 5.2)

6.3.3 ANÁLISIS DE LAS CARACTERÍSTICAS DE LOS ARCHIVOS CODIFICADOS

Para comprobar que las señales obtenidas en las diferentes ventanas del codificador son correctas se compara las señales en el dominio del tiempo y en el dominio de la frecuencia (espectro) con las señales obtenidas al utilizar Cool Edit Pro 2.0.

Antes de comprobar las señales correspondientes al dominio del tiempo y de la frecuencia del archivo codificado (comprimido) se hará una comparación de las señales del archivo original (no codificado) ya que si estas señales son incorrectas las pertenecientes al archivo codificado también serán incorrectas.

6.3.3.1 Análisis de los archivos sin comprimir

6.3.3.1.1 Análisis del archivo original en el dominio del tiempo

Como se muestra en la figura 6.30 las señales mostradas en el dominio del tiempo son las mismas, esto implica que el procesamiento del archivo de audio original en el dominio del tiempo es correcto.

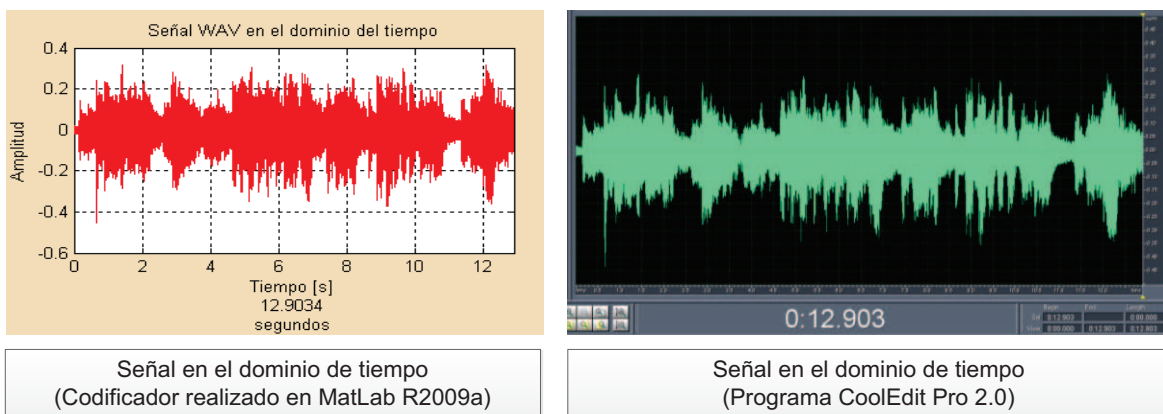


Figura 6. 30: Archivo de audio original en el dominio del tiempo

6.3.3.1.2 Análisis del archivo original en el dominio de la frecuencia

Para efectuar el análisis en el dominio de la frecuencia utilizando Cool Edit Pro 2.0 es necesario utilizar del menú “Analyze” la opción “Show Frequency Analysis”, tal como se indica en la figura 6.31.

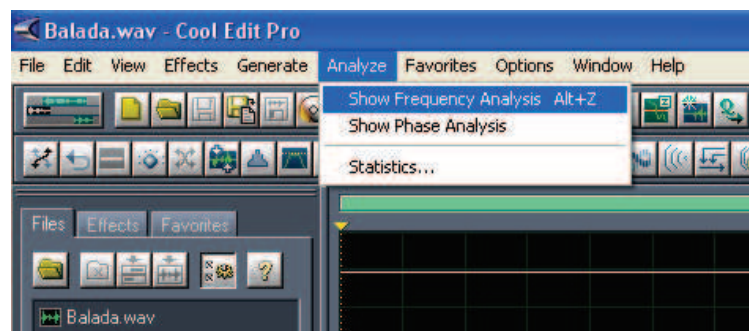


Figura 6. 31: Opción para activar la ventana del análisis de frecuencia

Una vez que se selecciona la opción “Show Frequency Analysis” en la pantalla se visualiza una ventana que indica el análisis en frecuencia correspondiente a un instante de tiempo del archivo de audio. En cada instante de tiempo seleccionado se muestra el análisis de frecuencia correspondiente, en consecuencia el espectro que se obtiene al utilizar Cool Edit Pro es un espectro diferente para cada instante de tiempo, como se aprecia en la figura 6.32.

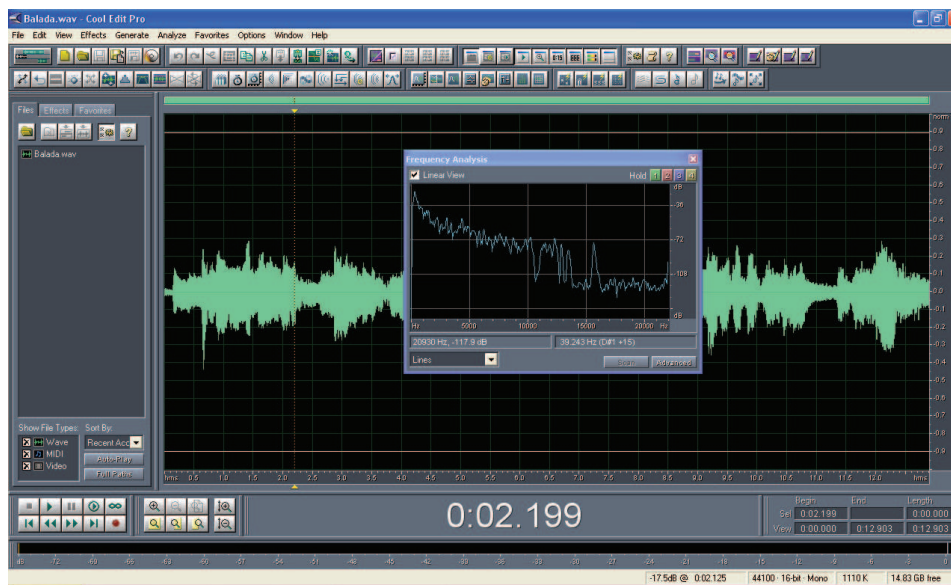


Figura 6. 32: Análisis en frecuencia del archivo audio original (en 2,199 segundos)

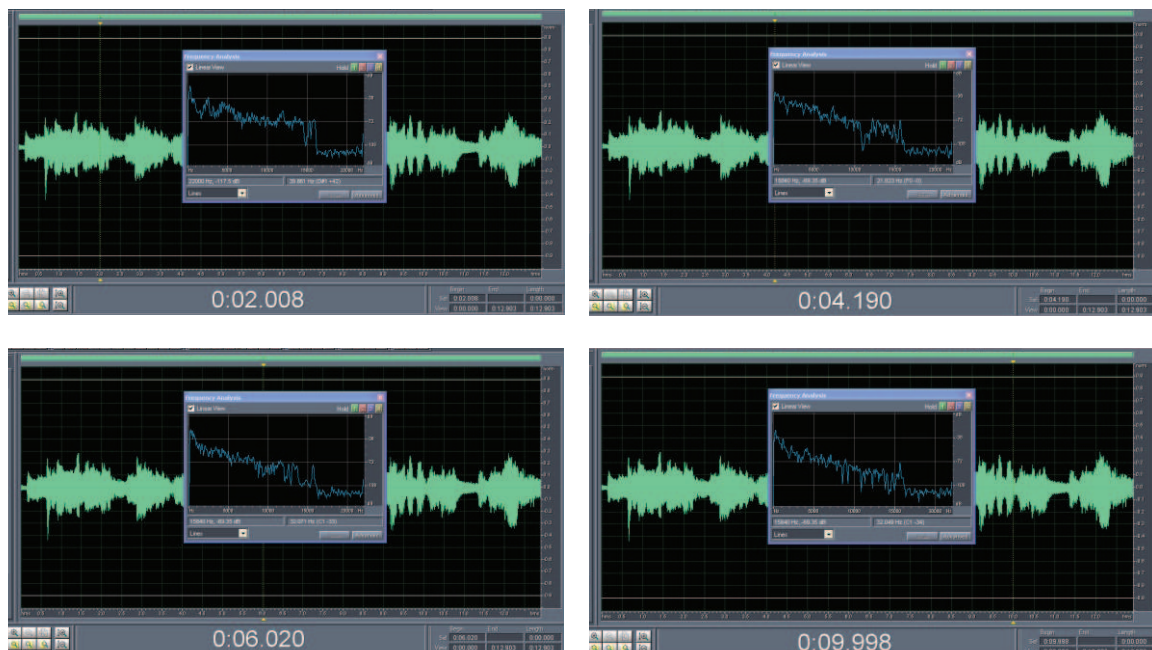


Figura 6. 33: Análisis en frecuencia del archivo audio en varios instantes de tiempo

Al contrario del espectro presentado en Cool Edit Pro 2.0, el cual es diferente en cada instante de tiempo, el espectro presentado en el codificador (diseñado en Matlab R2009a) es un espectro promedio para todo el archivo de audio, la razón es que para el análisis de frecuencia en el codificador diseñado en Matlab se ha empleado la transformada rápida de Fourier, la cual toma todo el archivo de audio es decir todo el tiempo seleccionado y saca un espectro promedio.

Para comparar en frecuencia las señales obtenidas con el codificador y las obtenidas con Cool Edit se verifican los aspectos más relevantes de las señales, se identifica la frecuencia o las frecuencias más sobresalientes en el espectro, como se indica a continuación.

Al observar la figura 6.34, se puede identificar de todo el rango de frecuencias un pico característico que está alrededor de 16.000 Hz, este pico junto con otras características observadas, sirven para comparar, el espectro obtenido con el codificador con el espectro obtenido con el software Cool Edit Pro 2.0. Comparando los espectros y aclarando que en Matlab el espectro que se obtiene es un espectro promedio, se puede concluir que el procesamiento de la señal en el dominio de la frecuencia es correcto.

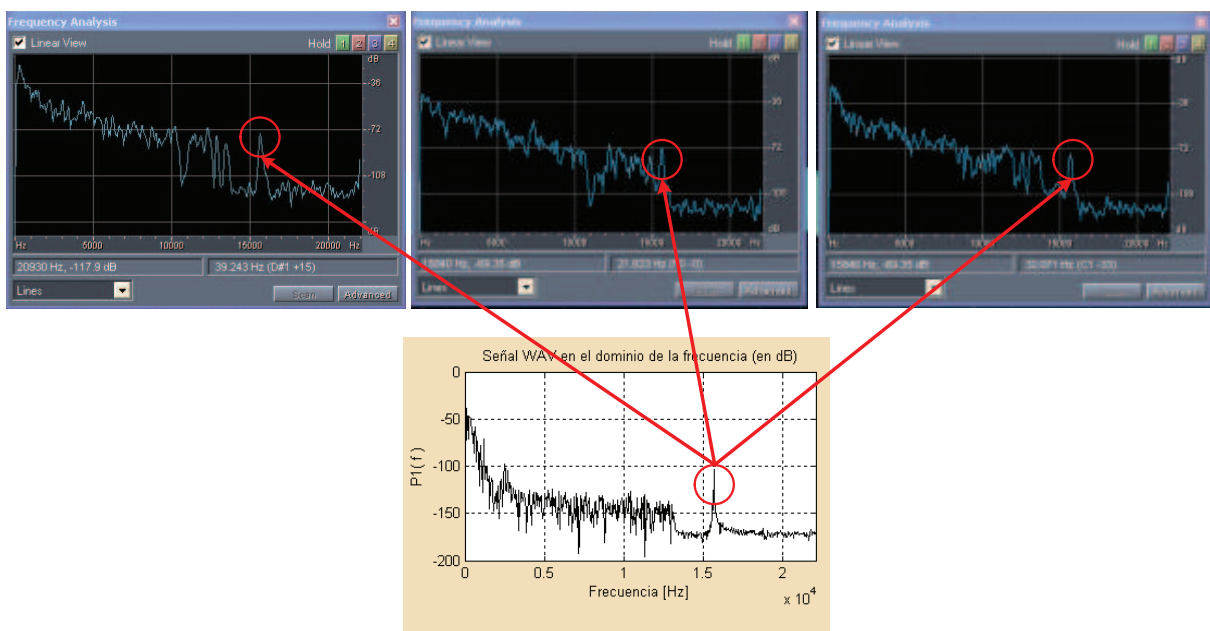


Figura 6. 34: Análisis de frecuencia del archivo audio obtenido con el codificador y comparando con el análisis de frecuencia obtenido con Cool Edit Pro 2.0

6.3.3.2 Análisis de los archivos comprimidos

Verificando anteriormente que tanto las señales en el tiempo y frecuencia están siendo correctamente procesadas y presentadas; a continuación se presentan las señales correspondientes al archivo codificado.

6.3.3.2.1 Análisis del archivo codificado en el dominio del tiempo

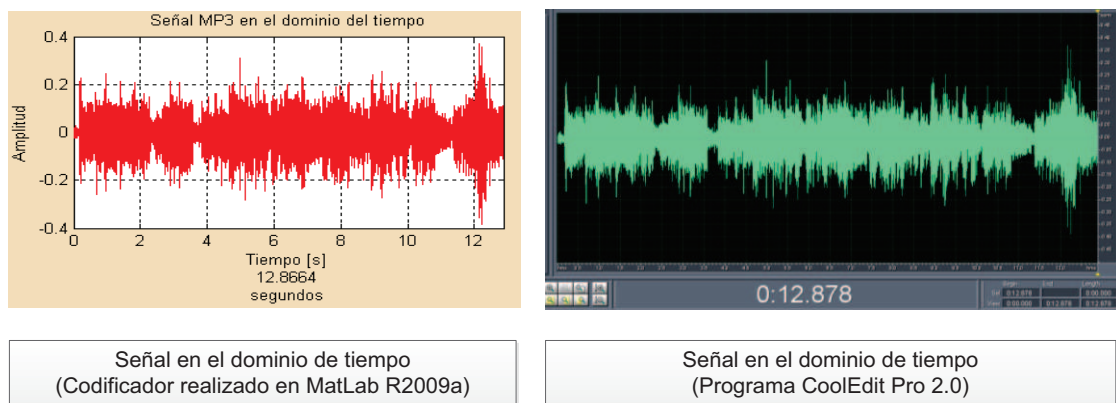


Figura 6. 35: Archivo de audio codificado en el dominio del tiempo

Comparando las señales en el dominio del tiempo en la figura 6.35, se puede apreciar que las señales son idénticas y por ende se puede concluir que las señales correspondientes al archivo codificado en el dominio del tiempo son correctas.

6.3.3.2.2 Análisis del archivo codificado en el dominio de la frecuencia

Como se indicó anteriormente para comparar las señales en el dominio de la frecuencia hay que recordar que el espectro que se obtiene con el codificador es un espectro promedio, un espectro obtenido de todo el archivo; en cambio el espectro obtenido con Cool Edit Pro es un espectro procesado en cada instante de tiempo; por ello para verificar que el espectro obtenido con el codificador es correcto se comparan las características más perceptibles y relevantes.

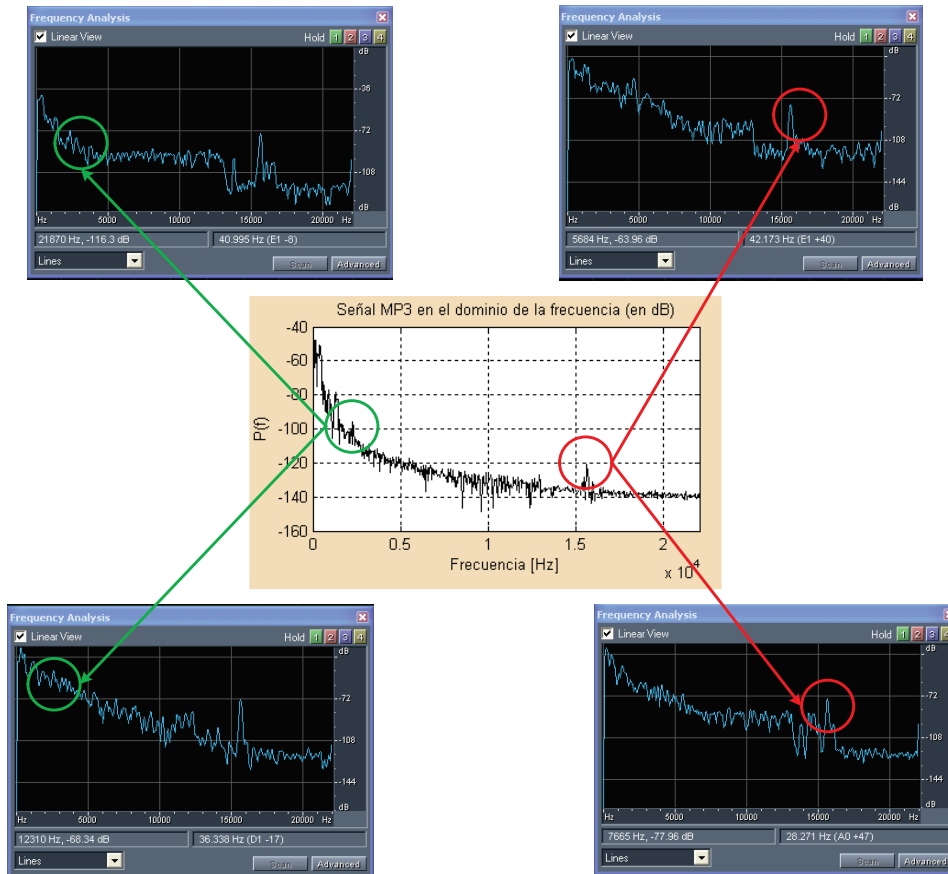


Figura 6. 36: Análisis en frecuencia del archivo de audio codificado obtenido con el codificador y comparando con el análisis de frecuencia obtenido con Cool Edit Pro 2.0

REFERENCIAS BIBLIOGRÁFICAS

CAPITULO 6: PRUEBAS DE FUNCIONAMIENTO Y RESULTADOS

INTERNET:

[49] GoldWave Inc., (2010), Excellence in sound editing software for over 15 years.

<http://www.goldwave.ca/forums>

[50] Adobe Systems, (2010), Introduction Adobe Audition, Adobe Systems.

<http://www.adobe.com/es/products/audition>

[51] The MathWorks, Inc., (2010), Mathworks Homepage, Matworks Inc.,

<http://www.mathworks.com>

CAPÍTULO 7.

CONCLUSIONES Y RECOMENDACIONES

7.1 CONCLUSIONES

- En la actualidad la era de la sociedad de la información y el acceso al internet han permitido el envío, recepción y almacenamiento de información multimedia, a tal punto que diariamente existen millones de archivos de audio que son cargados, descargados, transferidos y almacenados a través del internet. Y si bien es cierto que la tecnología ha ido en mejora y las aplicaciones son más robustas y constan de más utilidades no hay que olvidar que la información multimedia en sí, demanda grandes cantidades de almacenamiento y procesamiento, es por ello, que ha surgido la necesidad de la búsqueda de esquemas y procesos que permitan transferir y almacenar datos de forma eficiente; poder minimizar la cantidad de bits necesarios para representar la misma información ha conducido al desarrollo de algoritmos de compresión. Así por ejemplo, los formatos de audio comprimidos han sido un gran avance, aunque al principio eran criticados por su calidad, en la actualidad es inconcebible almacenar o reproducir archivos de audio en formatos sin compresión.
- En cuanto al audio se refiere los formatos de compresión no solo permiten ahorrar recursos de procesamiento y almacenamiento, además permiten minimizar el ancho de banda, lo que es un punto favorable permitiendo utilizar varias aplicaciones sobre el mismo canal de comunicaciones.
- Una de las principales ventajas para que se pueda dar la compresión del audio sin pérdidas significativas en la calidad, se debe a que el ser humano tiene capacidades limitadas, las cuales pueden ser explotadas en la codificación de audio. Estudios psicoacústicos han demostrado que el ser humano no es

capaz de percibir todas las componentes de frecuencia, por dicho motivo es posible eliminar información que pasa desapercibida por los sentidos y así poder tener relaciones de compresión elevados manteniendo la calidad del audio.

- Debido al crecimiento que ha tenido la utilización del audio digital, y en específico los formatos de audio comprimidos, se ha visto adecuado estudiar en detalle el algoritmo MP3 para compresión de audio. La decisión de hacer énfasis en el formato MP3 se debe a que es uno de los formatos más difundidos, y utilizados y a pesar de coexistir con una gran variedad de formatos que técnicamente podrían ser mejores, tener más utilidades, e incluso estar desarrollados en software libre, el formato MP3 aún actualmente se encuentra en auge y por lo visto se seguirá manteniendo o por lo menos no muestra señales de decaer.
- No se hizo énfasis en los sistemas de alta fidelidad debido a que es una contradicción hablar del audio en alta fidelidad y de formatos de audio comprimidos con pérdidas, como lo es el formato de audio MP3; además la alta fidelidad en los sistemas actuales está asociada a aspectos netamente comerciales más que a conservar una réplica fiel del audio presente en la naturaleza.
- Resultó interesante conocer de una manera muy detallada y extensa el formato mp3; un formato ampliamente difundido y utilizado, del cual la mayoría de personas desconoce las técnicas digitales y demás procesamiento de señales involucrado. Al conocer en detalle los algoritmos pertenecientes a este formato de audio, se evidenció que el proceso es impresionante pero a su vez bastante complejo de lo que se podría imaginar a primera vista, por dicho motivo para poder diseñar e implementar el codificador y cumpliendo con el estándar internacional ISO/IEC 11172-3 se ha procedido a dividir la complejidad del modelo, distribuyéndolo en forma jerárquica en bloques que constituyen subsistemas, de modo que facilitan el estudio y la simulación; ya

que al diseñar el codificador en bloques disminuye la carga computacional y los procesos resultan ser más rápidos y menos complejos.

- El formato MP3 presenta altos niveles de compresión, permite comprimir en un factor aproximado de 12 la información original, la calidad del audio obtenida es buena y existe compatibilidad con una gran cantidad de sistemas de software y de hardware; puede ser reproducido y editado por una gran cantidad de programas en la actualidad. Por las diversas cualidades que este formato presenta, se implementó en software de simulación Matlab 7.8 (R2009a) un programa capaz de leer y reproducir un archivo de audio de Microsoft (*.wav) muestreado a 44.100 Hz en modo estéreo o monofónico y a partir de éste, obtener y reproducir un archivo en formato de audio comprimido MPEG-1 Layer III (*.mp3) monofónico, con frecuencia de muestreo 44.100 Hz, de tasa de bits y tiempo de duración correspondiente a la elección y criterio del usuario del codificador.
- En cuanto a los resultados obtenidos se determinó analítica y gráficamente que los archivos codificados cumplen a cabalidad con los lineamientos provistos por el estándar ISO/IEC 11172-3. Al seguir los procedimientos dados por un estándar de carácter internacional se garantiza la plena compatibilidad de los archivos codificados, verificando que los mismos puedan ser abiertos y reproducidos en cualquier hardware o software compatible con el formato MP3.
- Con la ayuda de los programas comerciales de edición de audio profesional, GoldWave 5.2 y Cool Edit Pro 2.0 se determinó la validez en cuanto a reproducibilidad y la veracidad en cuanto a las características esperadas de los archivos codificados, con el análisis de los parámetros obtenidos de los archivos codificados tanto en el dominio del tiempo como en el dominio de la frecuencia.

- En cuanto a los ratios de compresión esperados, analizando los resultados se puede concluir que la compresión de los archivos es bastante buena, ya que se cumplen con las relaciones de compresión planteadas por Karlheinz Brandenburg y el Instituto Tecnológico de Fraunhofer, dado que para la tasa de bits más baja del codificador de 96 Kbps se presenta la mayor relación de compresión que es de siete.
- Como se mostró con los datos obtenidos al codificar diversos archivos, de diferentes tamaños, duración y diferente ritmo, los parámetros de entrada son importantes e influyen directamente en la compresión, así por ejemplo, codificando todo el archivo el parámetro que determina si se ha codificado en mayor o menor escala en cuanto al tamaño en bytes del archivo, es la tasa de bits; la tasa de bits es un factor que hay que tomar en cuenta si se tiene en mente reducir en mayor o menor cantidad el tamaño de un archivo de audio.

7.2 RECOMENDACIONES

- A pesar de que el programa en su totalidad es muy didáctico y sencillo de utilizar, se recomienda leer el anexo A, el cual constituye un manual de usuario y presenta información general del diseño del programa e información detallada del funcionamiento del codificador.
- Cabe recalcar que el codificador puede ser mejorado, empleando la última versión liberada de Matlab o empleando toolboxes adicionales, o a su vez utilizando una computadora con mejores características a la utilizada durante todo el proyecto.
- Como MatLab, no ha sido diseñado para trabajar en tiempo real, se recomienda codificar tiempos pequeños, ya que de lo contrario, el proceso de codificación puede tardar un tiempo considerable; además el codificador implementado en este proyecto no fue concebido como un software comercial,

sino que fue pensado como un referente técnico para el uso de técnicas digitales en el procesamiento de señales de audio en formato comprimido.

- Como el codificador demanda una gran cantidad de recursos computacionales, principalmente recursos del procesador y memoria, es recomendable ejecutar el programa en una computadora que conste por lo menos de un procesador Pentium IV o AMD Athlon 64 y 512 MB de memoria RAM.

- El programa ha sido implementado para correr sobre un sistema operativo Windows XP pero esto no constituye un limitante en la compatibilidad con otros sistemas operativos tales como GNU Linux o versiones posteriores de Windows. De utilizarse el programa en sistemas operativos como Windows Vista o Windows 7 se recomienda configurar la prioridad del proceso "MATLAB.exe" en "tiempo real" para asignar la mayor cantidad de recursos del procesador al codificador MP3 implementado en Matlab.

ANEXOS

ANEXO A

MANUAL DE USUARIO

A.1 DESCRIPCIÓN DEL PROGRAMA

El programa convierte un archivo de audio de formato Windows PCM (*.wav) de frecuencia de muestreo $F_s = 44.100$ Hz en modo estéreo o monofónico, a un archivo de formato comprimido MPEG-1 Layer III (*.mp3) monofónico; en base a las especificaciones del estándar internacional ISO/IEC 11172-3.

El codificador está diseñado en varios archivos (*.m, *.fig y *.mat), con el objetivo de separar el proceso de codificación en varias etapas, pretendiendo de esta manera entender mejor los lineamientos provistos en el estándar. El programa en su totalidad cuenta con varias interfaces gráficas que permiten una utilización más sencilla del programa, además se emplean varios archivos de imagen (*.jpg y *.png), un archivo de video (*.avi) y varios archivos de documentación (*.pdf), los primeros utilizados en los elementos visuales del programa y los restantes con información del estándar y de la utilización en sí del programa.

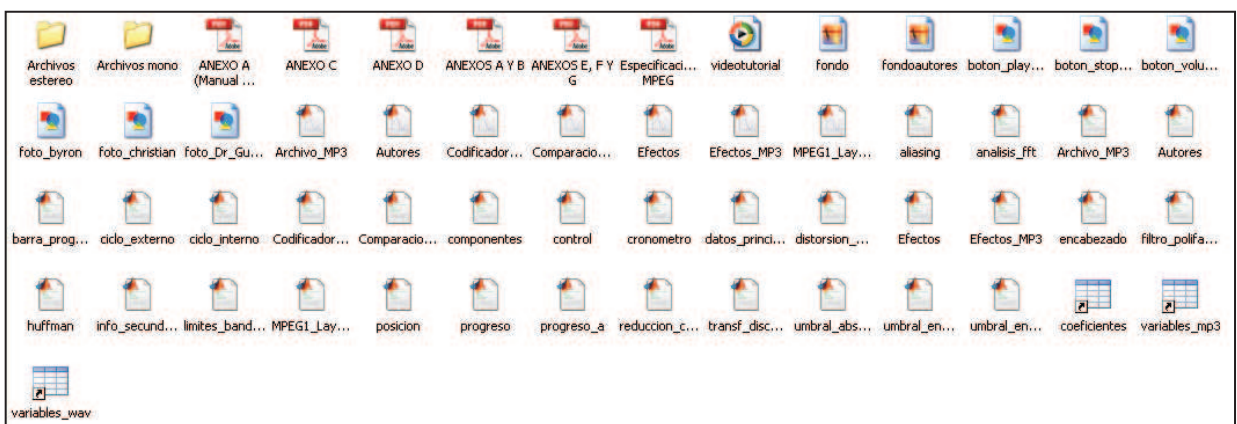


Figura A. 1: Archivos del Codificador MPEG-1 Layer III

El programa en cuanto a la interfaz gráfica se refiere, está constituido por cinco ventanas principales (*.fig), incluida la carátula y biografía de autores, y dos ventanas adicionales (ventanas de efectos de audio).

Cabe mencionar que tanto los elementos de las interfaces gráficas como las interfaces en sí mismo van apareciendo secuencialmente mientras se va procesando la señal.

A.2.1 REQUERIMIENTO DE SOFTWARE Y DE HARWARE

A.1.1.1 Especificaciones de hardware

El programa ha sido elaborado, diseñado y ejecutado en una computadora notebook Compaq Presario C706LA, de las siguientes características:

- Procesador: AMD Turion (tm) 64, Mobile Technology MK38 – 2.21 [GHz].



Figura A. 2: Procesador de la computadora utilizada para correr el programa

- RAM: 1024 MB, DDR2 667 [MHz].
- Disco duro: 120 GB, enhanced IDE 5400 RPM SATA.
- Tarjeta gráfica: Nvidia GeForce Go 1600, 256 MB.

A.1.1.2 Especificaciones de software

El programa ha sido desarrollado en Matlab 7.8 (R2009a), corriendo en un sistema operativo Windows XP SP3 y para que la utilización del programa sea más amigable se ha empleado el entorno de programación visual de Matlab

GUIDE (Graphical User Interface Develop Enviroment). Como el programa ha sido desarrollado en un ambiente de Microsoft Windows, el codificador es compatible con Windows XP, Windows Vista y Windows 7.

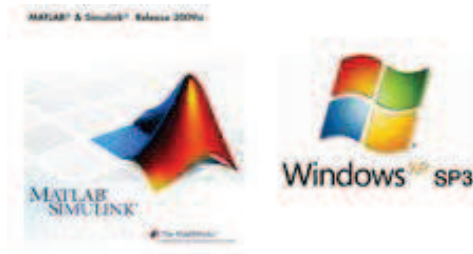


Figura A. 3: Especificaciones de software para el codificador MPEG-1 Layer III

En cuanto a los toolboxes o librerías empleados, el codificador necesita los siguientes toolboxes: Matlab Toolbox™, Signal Processing Toolbox™, Communications Toolbox™, MP3 Toolbox™ y Guide Toolbox™. Al elegir la opción de instalación completa en Matlab, casi todos los toolboxes mencionados anteriormente son instalados automáticamente, excepto el MP3 Toolbox™ el cual es necesario instalarlo (ver anexo B) y se lo puede descargar de la siguiente dirección: <http://www.mathworks.com/matlabcentral/fileexchange/13852-mp3read>.

A.2.2 ACCESO AL PROGRAMA

Como se mencionó anteriormente el programa utiliza el ambiente gráfico de Matlab, y al crear una interfaz gráfica, Matlab de manera automática crea dos archivos uno *.fig y otro *.m, por ello se puede tener acceso al programa mediante dos formas, a través del ambiente gráfico de Matlab o a través de la apertura del archivo *.m correspondiente a la interfaz gráfica.



Figura A. 4: Archivos *.fig y *.m de MatLab

Hay que recordar que el programa está diseñado para que funcione y se presente secuencialmente; por ese motivo la primera ventana a la que se debe tener acceso es la ventana “MPEG1_LayerIII.fig” o al archivo “MPEG1_LayerIII.m”.

A.1.2.1 Acceso al programa a través del ambiente gráfico

Una vez abierto Matlab y listo para ser utilizado; en la ventana de comandos (Command Window) se procede a ingresar el comando *guide*, el cual permite tener acceso a la herramientas del ambiente gráfico.

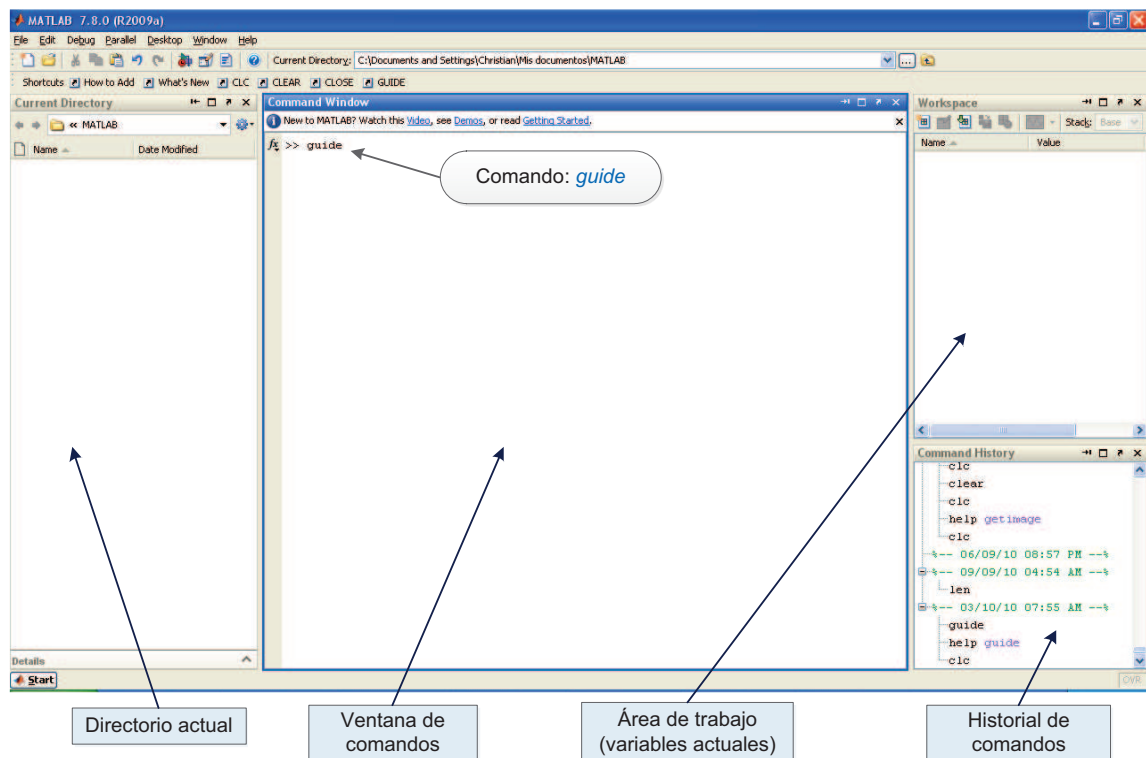


Figura A. 5: Ventana principal de Matlab y comando para acceder a las herramientas de la interfaz gráfica de usuario

Luego de ingresar el comando *guide*, aparece una ventana que permite crear una nueva interfaz gráfica o abrir una interfaz existente creada anteriormente, obviamente se debe elegir abrir una interfaz existente.

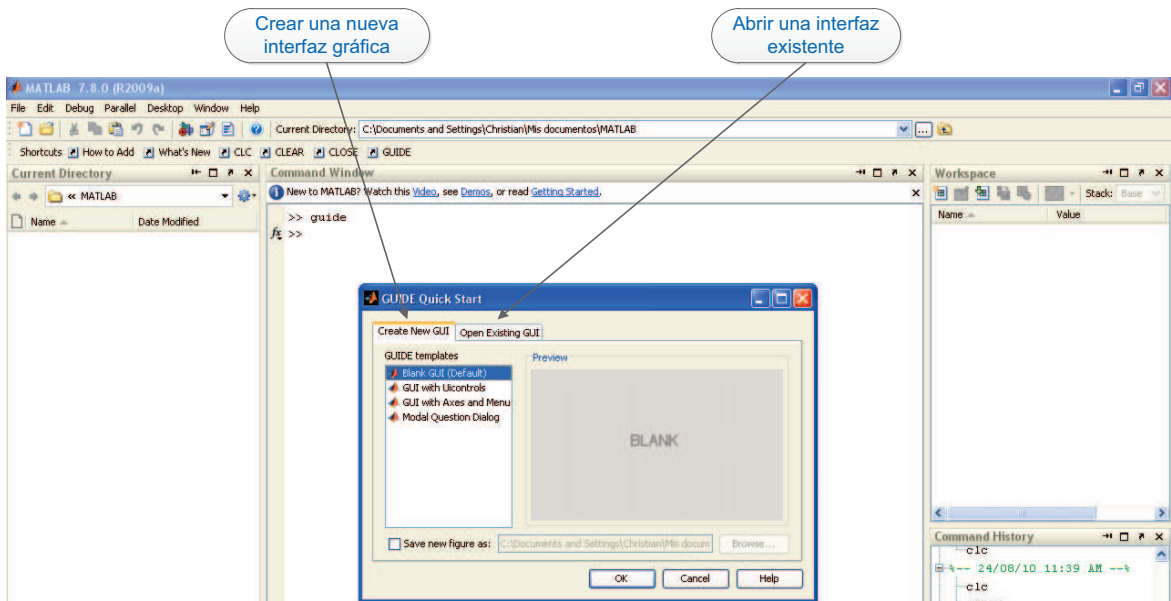


Figura A. 6: Ventana de inicio rápido de la interfaz gráfica de usuario

Al seleccionar la opción abrir una GUI (Graphical User Interface) existente y dependiendo en que directorio de archivos se esté trabajando puede ser necesario buscar la interfaz a la que se quiere tener acceso, esto se lo hace seleccionado el botón de búsqueda (Browse...).

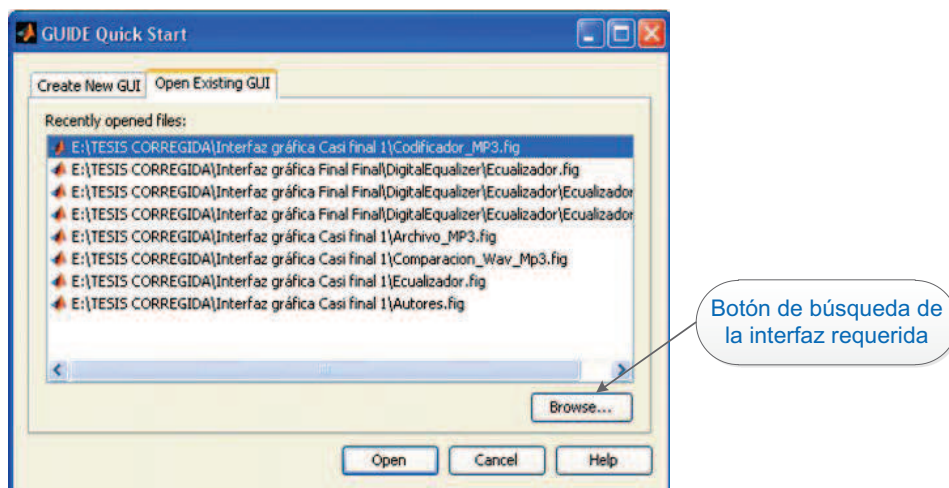


Figura A. 7: Búsqueda de la interfaz gráfica de interés

Una vez que se ha ubicado el directorio en el cual se encuentra almacenada la interfaz a la cual se quiere tener acceso, se selecciona el archivo *.fig de la interfaz correspondiente y se la procede abrir.

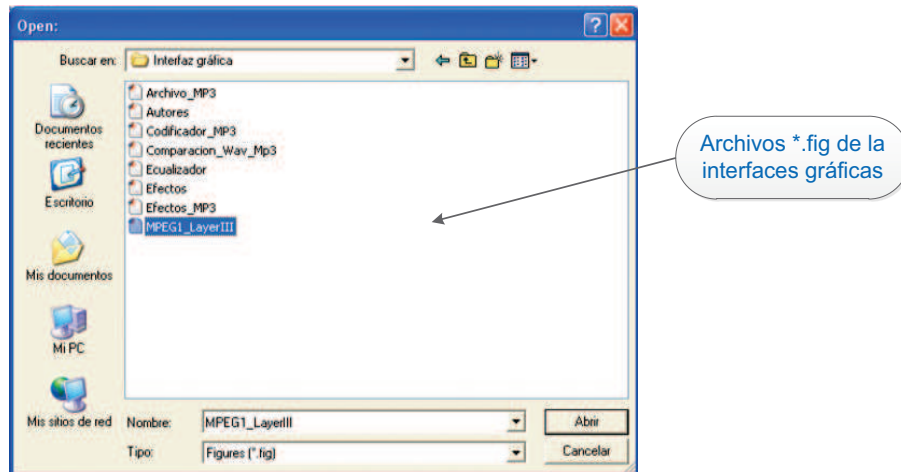


Figura A. 8: Archivos *.fig correspondientes a las interfaces gráficas

Abierto el archivo *.fig correspondiente a la interfaz gráfica, aparece una ventana la cual en su barra de herramientas consta de un botón que permite correr el programa, tal y como se muestra en la figura A.9.

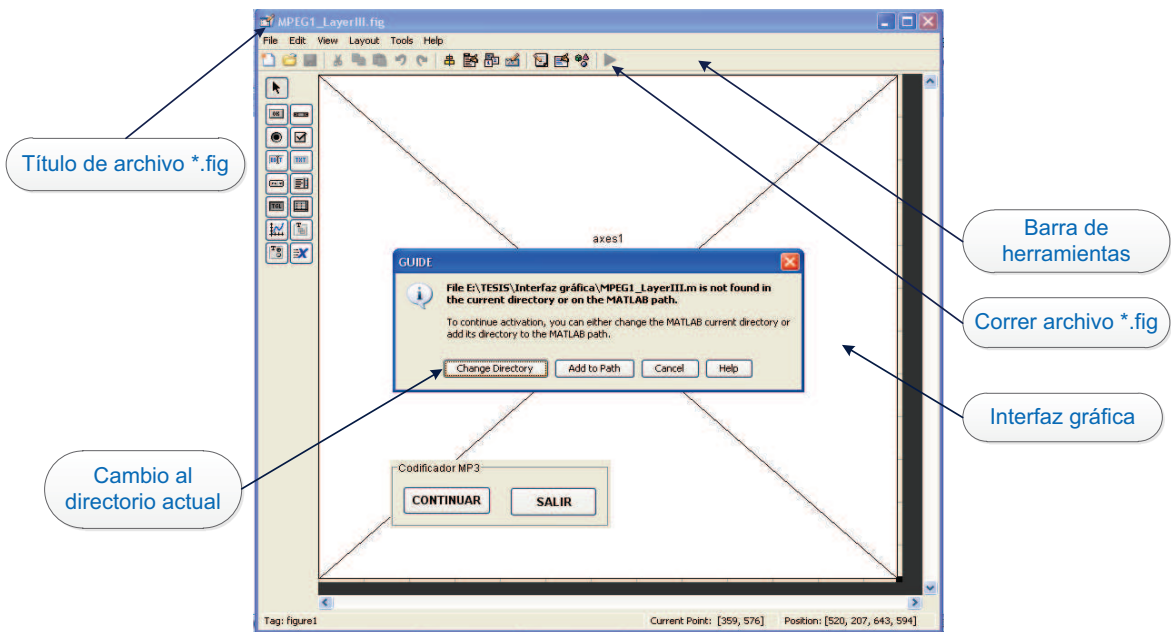


Figura A. 9: Archivo *.fig correspondiente a la interfaz gráfica

Al momento de correr el archivo *.fig, y dependiendo de si el directorio de archivos y la ruta para acceder el archivo están actualizados y son los correctos, puede o no aparecer un mensaje de información, como el que se indica en la figura A.10, en este caso para asegurarse de que se está trabajando correctamente y para

evitar complicaciones futuras es recomendable elegir la opción cambiar directorio (Change Directory).

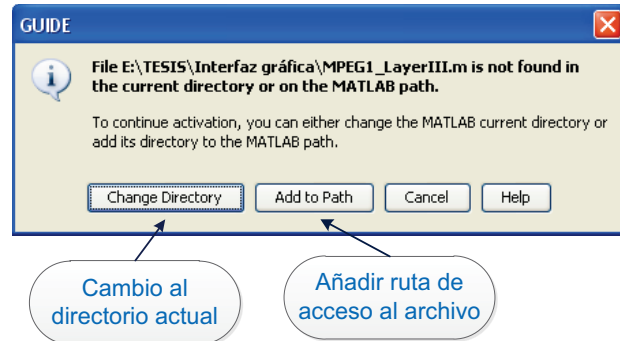


Figura A. 10: Actualización del directorio y ruta de acceso al archivo

En el instante en que tanto el directorio y la ruta de acceso al archivo son los correctos, se tiene acceso a la primera ventana del codificador, es decir a la ventana “MPEG1_LayerIII” (ventana de la carátula).



Figura A. 11: Acceso a la ventana de carátula del codificador mediante la utilización del ambiente gráfico

A.1.2.2 Acceso al programa a través del editor de archivo .m

Una vez abierto Matlab y listo para ser utilizado; se procede a buscar y abrir el archivo *.m de interés, es decir el archivo “MPEG1_LayerIII.m”.

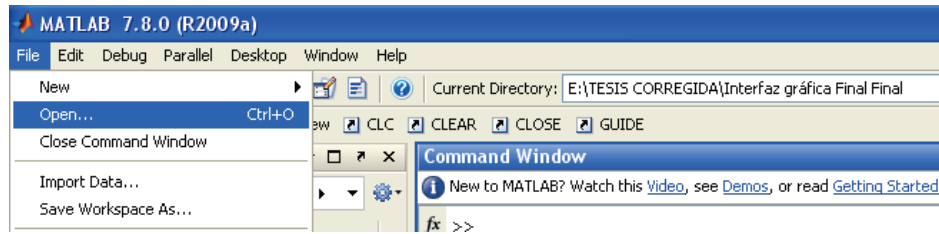


Figura A. 12: Apertura de archivos

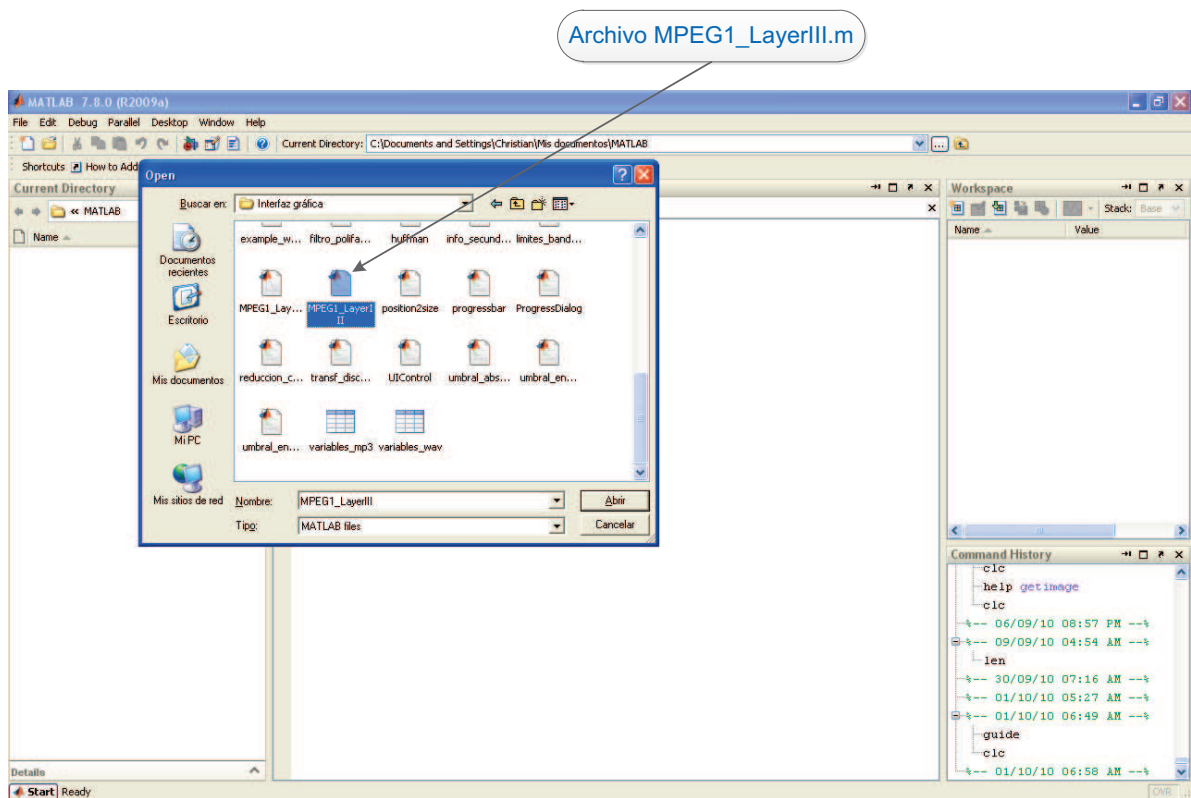


Figura A. 13: Búsqueda y selección del archivo *.m perteneciente al codificador

Cuando se procede a abrir el archivo *.m, el archivo se abre en un típico editor de archivos *.m de MatLab y en la ventana del editor de archivos .m, aparece una barra de herramientas la cual consta de un botón que permite correr el programa, tal y como se muestra en la figura A.14.

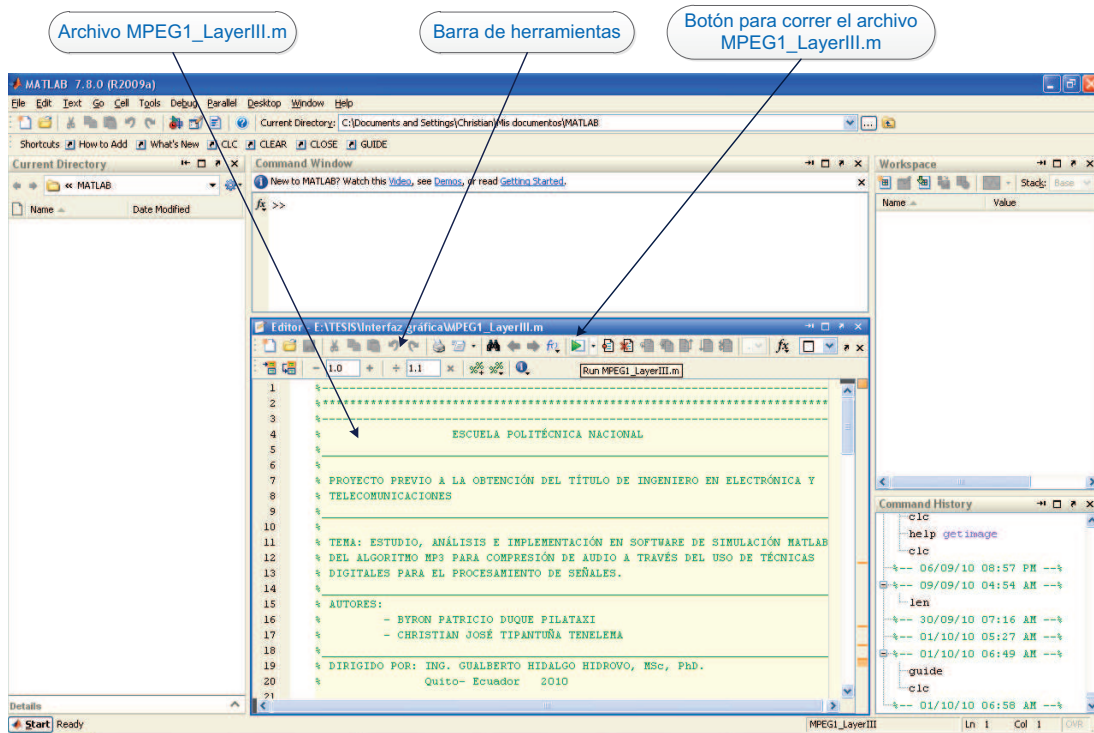


Figura A. 14: Apertura del archivo MPEG1_LayerIII.m

Al correr el programa a través del editor de archivos .m puede o no ser necesario actualizar el directorio de archivos y la ruta para acceder a los archivos del codificador; si es necesario actualizar el directorio y la ruta de acceso se elige la opción cambiar directorio (Change Directory).

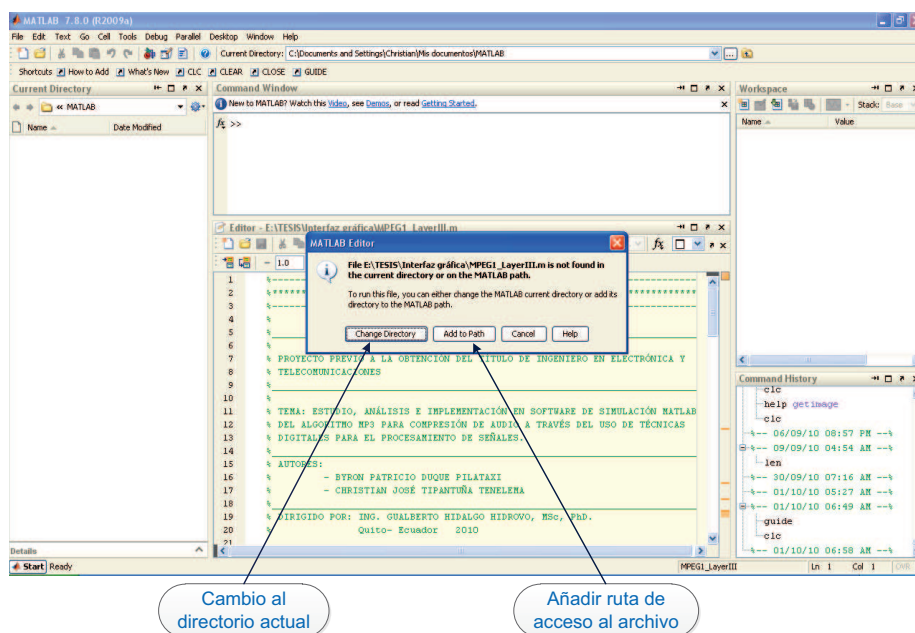


Figura A. 15: Actualización del directorio y ruta de acceso al archivo

Luego de que tanto el directorio como la ruta de acceso están actualizados y al no tener ningún error o advertencia al correr el archivo MPEG1_LayerII.m, se tiene acceso a la primera ventana del codificador, es decir a la ventana “MPEG1_LayerIII” (ventana de la carátula).



Figura A. 16: Acceso a la ventana de carátula del codificador mediante la utilización del editor de archivos .m

Una forma más sencilla de acceder al codificador a través del editor de archivos .m es abrir directamente el archivo “MPEG1_LayerIII.m” posteriormente se corre el archivo, y si no se tiene ningún percance se tiene acceso a la primera ventana del codificador.



Figura A. 17: Archivo MPEG1_LayerIII.m

A.2 EJECUCIÓN DEL PROGRAMA

Como se mencionó anteriormente, el codificador en su totalidad está diseñado para que funcione de una manera secuencial y ordenada, es decir los datos ingresados o calculados en una ventana anterior sirven como datos o resultados en la ventana siguiente o en otra ventana posterior. Como el codificador está dividido en varias ventanas a continuación se presenta una descripción de cada ventana del codificador.

A.2.1 VENTANA MPEG1_LayerIII

La primera ventana del codificador es la ventana de la carátula, la cual consta de dos botones en la parte inferior izquierda, un botón "CONTINUAR" el cual da paso a la ventana siguiente (ventana Codificador_MP3) y un botón "SALIR" el cual al ser seleccionarlo muestra una ventana de consulta, figura A.19, la cual según la opción seleccionada, permite salir del programa o regresar a la ventana MPEG1_LayerIII nuevamente, para posteriormente si así fuese, ir a la ventana siguiente.

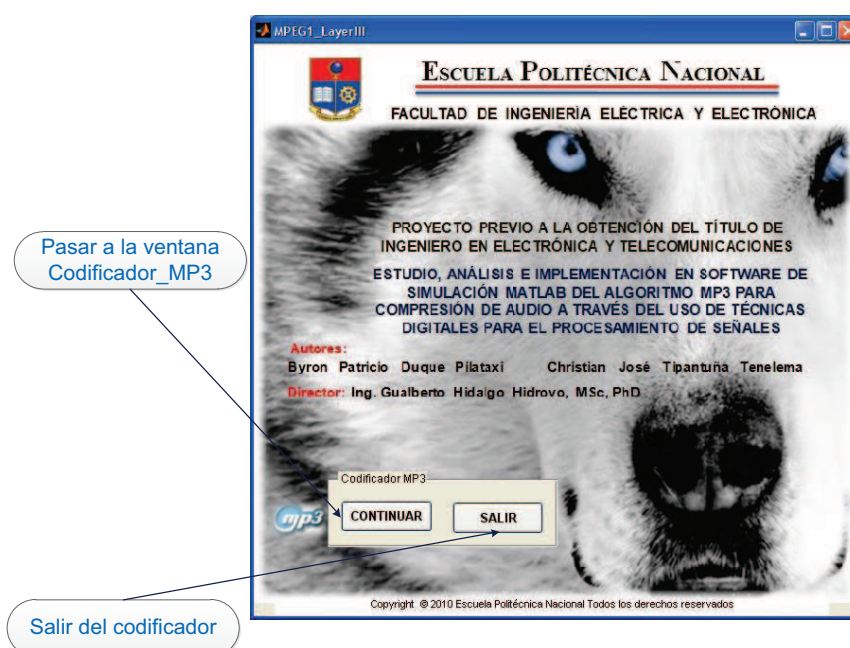


Figura A. 18: Carátula del codificador



Figura A. 19: Mensaje al seleccionar el botón "SALIR"

A.2.2 VENTANA Codificador_MP3

Ventana principal del codificador, en esta ventana se requiere que el usuario interactúe con el programa al elegir de manera secuencial y ordenada los parámetros de codificación, como se puede observar en la figura A.20, la ventana cuenta con una barra de menús, submenús y una barra de herramientas las cuales permiten una sencilla utilización del programa.

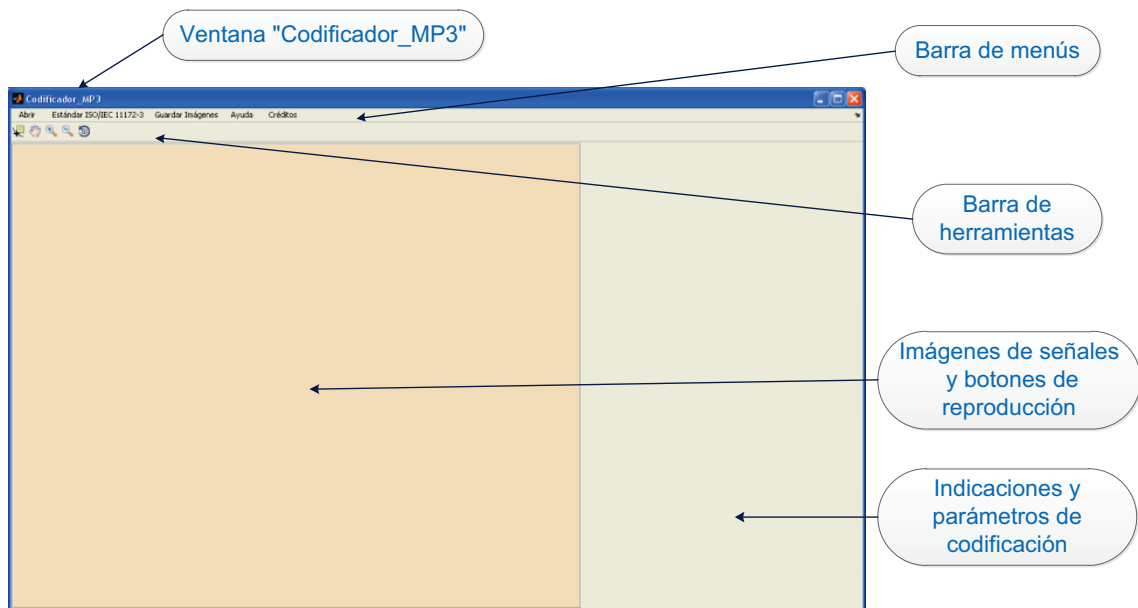


Figura A. 20: Segunda ventana del codificador, ventana "Codificador_MP3"

Esta ventana en primera instancia, aparece vacía, como se visualiza en la figura 20. Una parte del área vacía de la ventana, está destinada para mostrar las imágenes del procesamiento de la señal, y la otra parte de ésta, corresponde a las indicaciones y al ingreso de los requerimientos para la codificación.

Para iniciar el proceso de codificación es necesario e indispensable seleccionar, el archivo de audio que se desea codificar. Para hacerlo, es necesario elegir la opción “Archivos .WAV” del menú “Abrir” de la barra de menús, tal como se indica en la figura A.21. Hay que recordar que el archivo de audio debe estar en formato *.wav, puede ser estereofónico o monofónico y con una frecuencia de muestreo de 44.100 Hz.

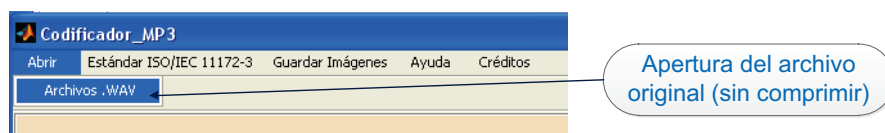


Figura A. 21: Opción para abrir archivo sin comprimir *.wav

Una vez seleccionada la opción “Archivos .WAV” del menú “Abrir” se despliega una ventana de búsqueda, figura A.22, la cual permite ubicar y seleccionar el archivo de audio requerido, la ventana como se muestra a continuación está condicionada para que solo se desplieguen archivos *.wav con ello se garantiza que no se cometan errores al elegir el archivo de audio de entrada.

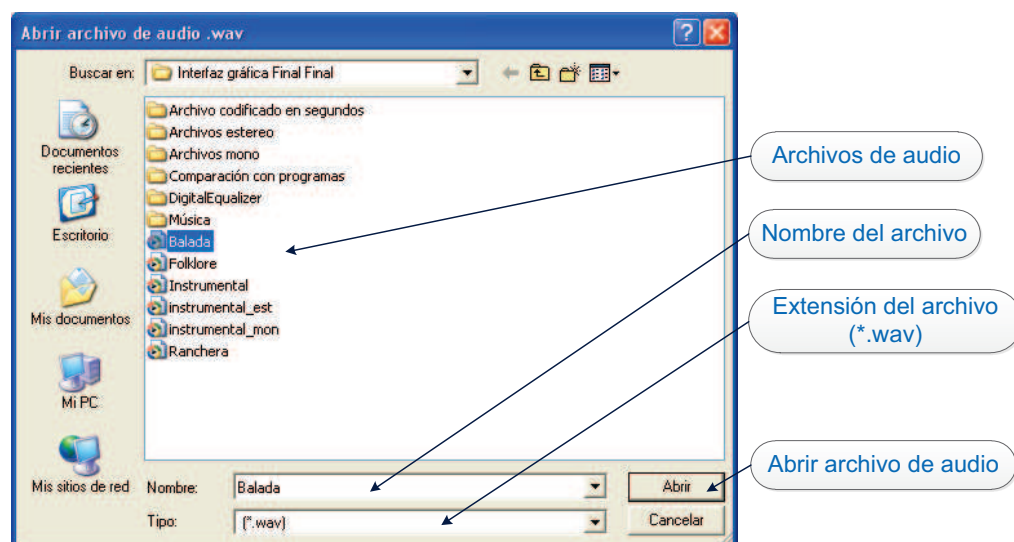


Figura A. 22: Selección y apertura del archivo *.wav

Luego de seleccionar y abrir el archivo de audio, empieza el procesamiento de la señal, y progresivamente van apareciendo las imágenes de las señales procesadas, tal y como se muestra en la figura A.23.

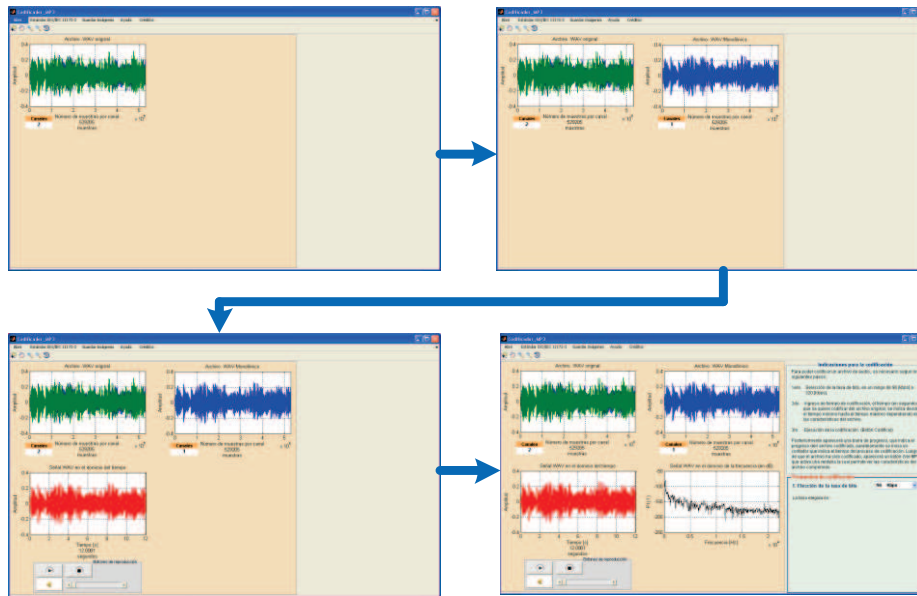


Figura A. 23: Procesamiento de archivo de audio de entrada

Cabe mencionar que las señales que se obtienen del procesamiento del archivo de audio, no sólo se las puede visualizar, sino que también se las puede escuchar, por ello en la parte inferior de la imagen de la señal en el dominio del tiempo se activan botones de reproducción, como se puede ver en la figura A.24; dichos botones de reproducción permiten escuchar, detener y aumentar o disminuir el volumen del archivo de audio; con respecto al volumen, cabe mencionar que por conveniencia el volumen inicialmente se presenta configurado en el nivel mínimo (silencio).

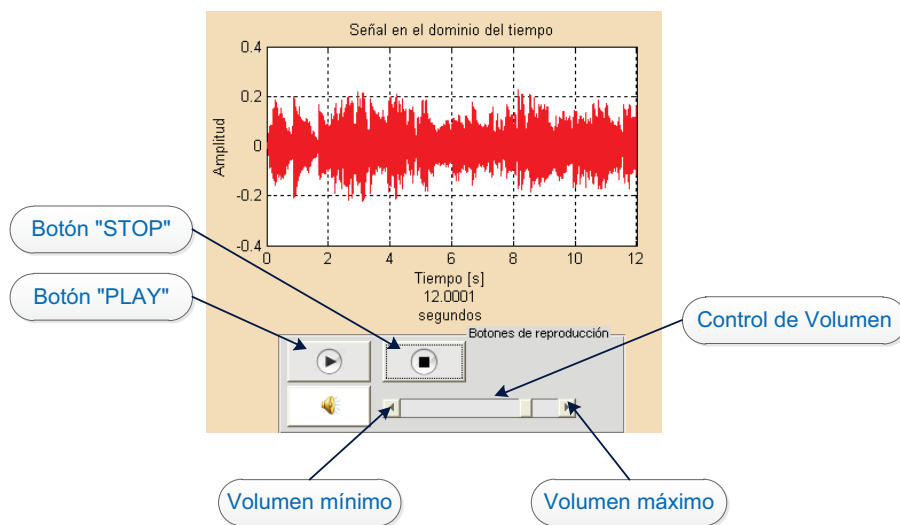


Figura A. 24: Botones de reproducción multimedia

Una vez que se han procesado todas las señales correspondientes al archivo de audio de entrada, en la parte derecha de la ventana se despliega un cuadro que contiene indicaciones de la codificación y más abajo se presenta otro cuadro en el cual se solicitan los parámetros de codificación, figura A.25.

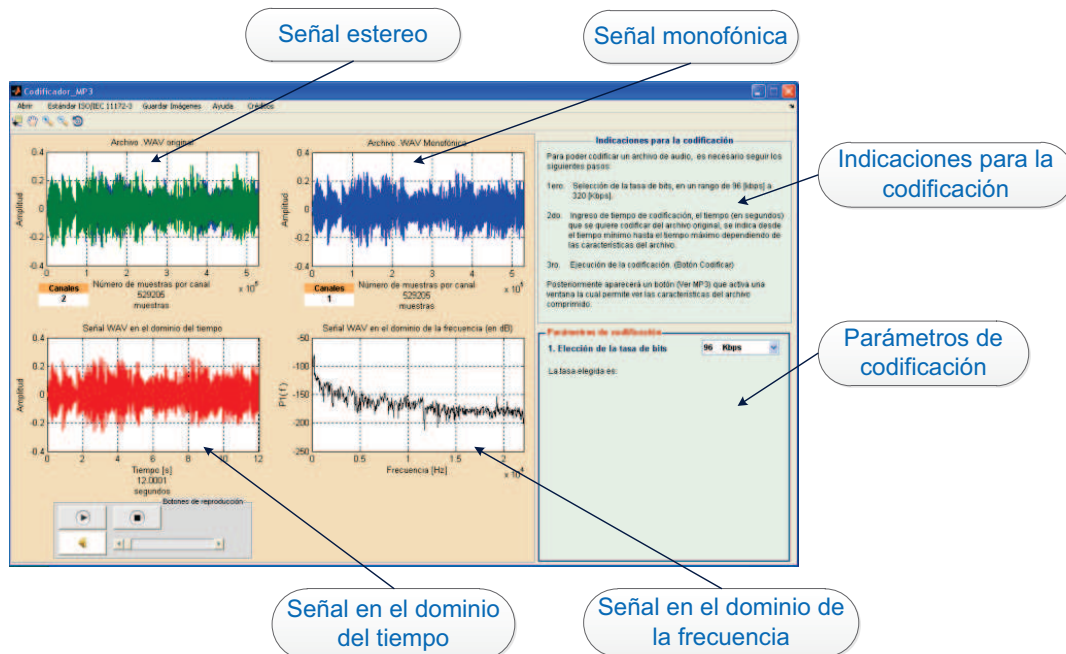


Figura A. 25: Señales de archivo de audio de entrada, indicaciones y parámetros de codificación

Con respecto a las indicaciones de codificación, hay que aclarar que dichas indicaciones se deben cumplir a cabalidad y en el orden establecido, de lo contrario se van a tener errores en la codificación o sencillamente no se activa la siguiente opción que permite el ingreso de los demás parámetros.

El ingreso de dos parámetros, la tasa de bits y el tiempo de codificación, son necesarios para ejecutar el proceso de codificación, no obstante la selección de los parámetros de codificación se la realiza en tres pasos los cuales son descritos a continuación:

- El primer paso es la elección de la tasa de bits, para lo cual se debe seleccionar un valor del menú de la tasa de bits, en un rango de 96 Kbps a

320 Kbps, al seleccionar una tasa de bits en dicho rango se activa el segundo parámetro de codificación, es decir, el tiempo de codificación.

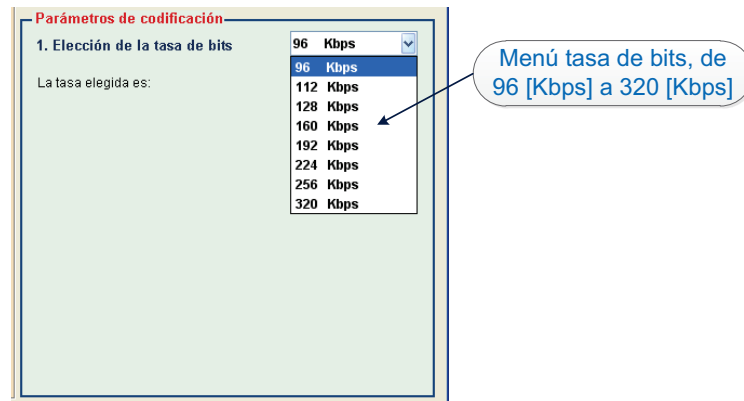


Figura A. 26: Elección de la tasa de bits

Como se observa en la figura A.27, al seleccionar la tasa de bits, se activa una caja de texto de edición la cual permite el ingreso del tiempo requerido, además se despliega información de los tiempos mínimos y máximos que pueden ser codificados.

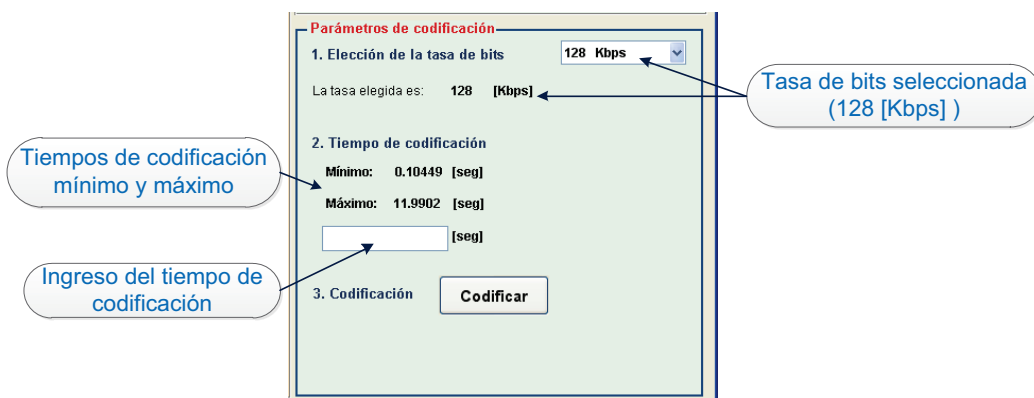


Figura A. 27: Parámetros mostrados después de seleccionar la tasa de bits

- El segundo paso, es el ingreso del tiempo de codificación (cuanto tiempo se quiere codificar del archivo original), el tiempo de codificación está dado en segundos y como se mencionó anteriormente este tiempo viene dado entre un mínimo y un máximo. El tiempo de codificación mínimo para todos los archivos es de 0.10449 segundos, esto es para garantizar el procesamiento mínimo de

4 tramas MPEG-1 Layer III, mientras el tiempo máximo claro está, depende de las características de cada archivo de audio.

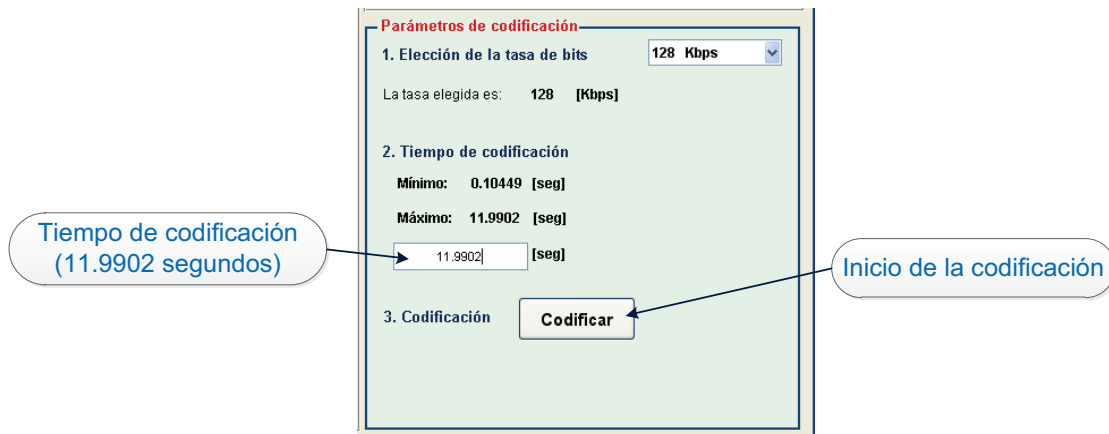


Figura A. 28: Ingreso del tiempo de codificación

- El tercer paso, una vez que se ha ingresado el tiempo de codificación adecuado, es iniciar el proceso de la codificación en sí mismo, esto se lo hace seleccionando el botón "Codificar"; al seleccionar el botón "Codificar" empieza la codificación y se visualiza en la pantalla una barra de progreso, la cual indica el progreso sistemático de la codificación.

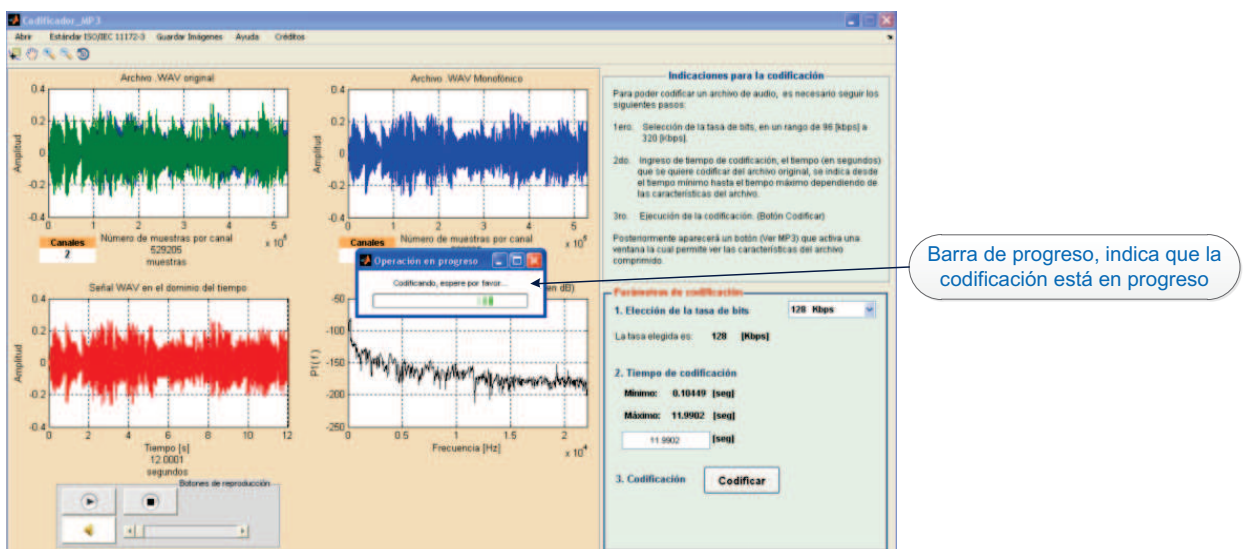


Figura A. 29: Ejecución de la codificación

Al iniciarse la codificación, en la ventana “Codificador_MP3” se visualiza, el progreso de la codificación (barra de progreso), pero paralelamente a esto, se inicia un cronómetro el cual permite visualizar el tiempo que tarda el archivo en ser codificado; el cronómetro cuenta con horas, minutos, segundos y milisegundos, ver figura A.30, con esto se consigue gran exactitud en la lectura del tiempo que se tarda el archivo en la codificación.

Cabe mencionar que el cronómetro funciona en modo asíncrono esto significa que si la carga del procesador es alta no se visualiza el cronómetro en la pantalla, pero si se libera la carga del procesador la ventana del cronómetro se despliega, tal como se muestra en la figura A.30, es por ello que el cronómetro solo se despliega cuando el archivo está casi por terminar su codificación porque en dichos instantes la carga del procesador no es tan alta.

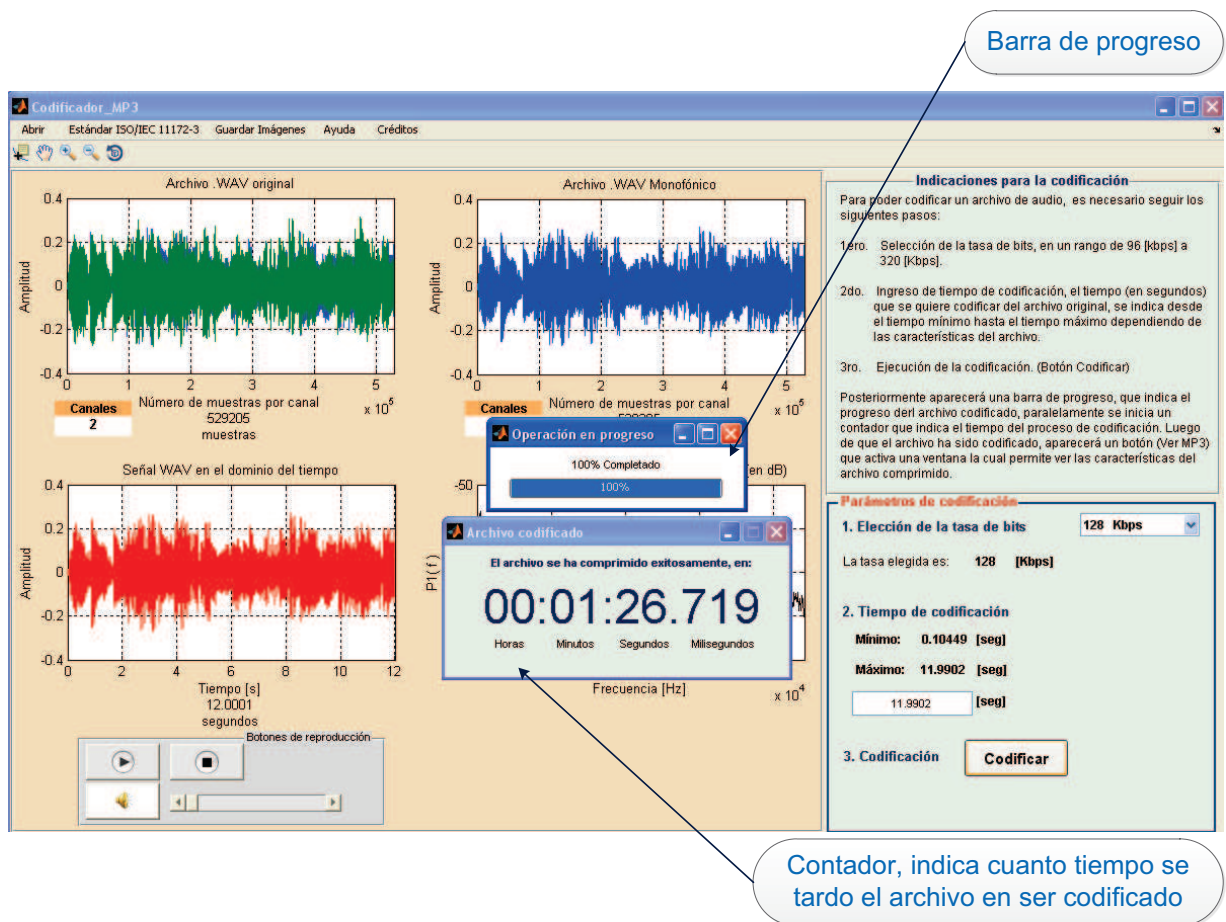


Figura A. 30: Finalización del proceso de codificación

La ventana perteneciente al cronómetro una vez que se ha finalizado la codificación, nos indica el tiempo que se tardó en el proceso de codificación y muestra un mensaje que indica que el archivo de audio ha sido codificado exitosamente, es decir, el proceso de codificación ha completado el cien por ciento, sin errores.

Instantes posteriores a la finalización del proceso de codificación, se activa en la parte inferior derecha de la ventana “Codificador_MP3” un botón “Ver MP3” el cual al ser activado da paso a la tercera ventana del codificador, que es la ventana “Archivo_MP3”.

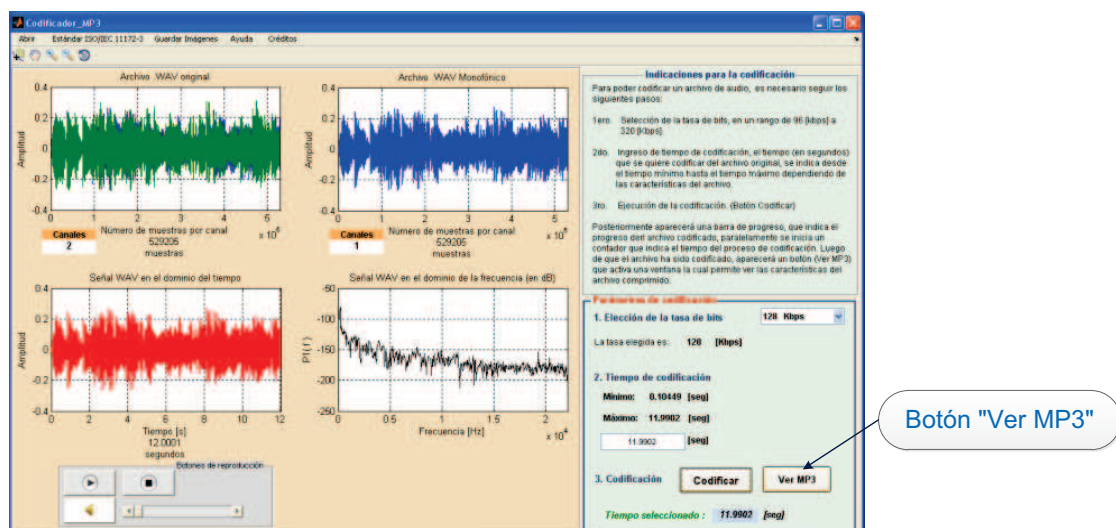


Figura A. 31: Activación del botón “Ver MP3” para activar la ventana “Archivo_MP3”

A.2.3 VENTANA Archivo_MP3

Esta ventana constituye la tercera ventana del codificador, y permite visualizar y escuchar las características del archivo comprimido (*.mp3), cuenta con una barra de menús, submenús y una barra de herramientas, las cuales permiten una sencilla utilización del programa.

Al igual que la ventana “Codificador_MP3”, esta ventana inicialmente aparece vacía como se muestra en la figura A.32, en esta ventana a diferencia de la primera ventana del codificador, el usuario no necesita ingresar parámetros, lo

único que se necesita es seleccionar el archivo resultante del proceso de codificación, es decir, el archivo comprimido (*.mp3).

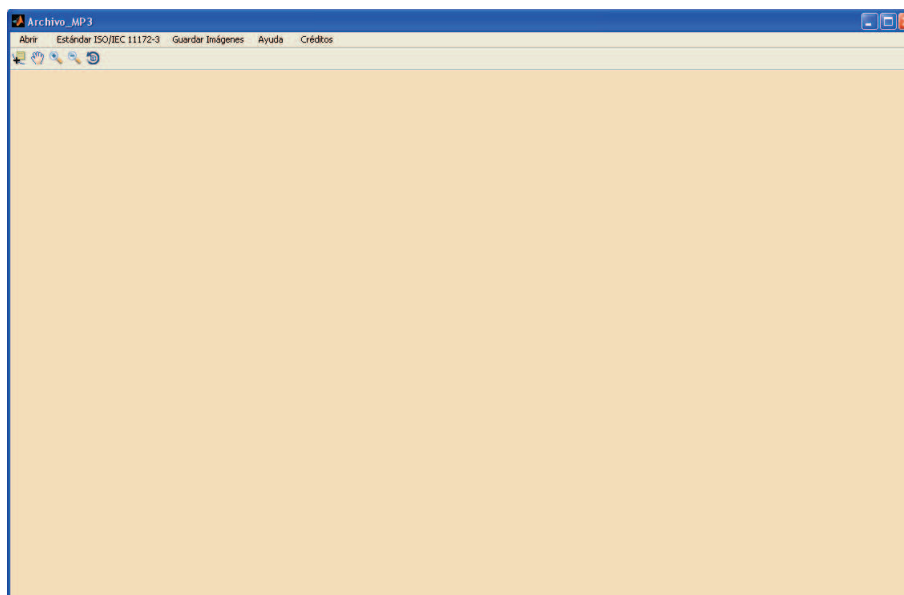


Figura A. 32: Tercera ventana del codificador, ventana "Archivo_MP3"

Para seleccionar y abrir el archivo codificado (*.mp3) es necesario seleccionar el menú "Abrir" la opción "Archivos .MP3", figura A.33.

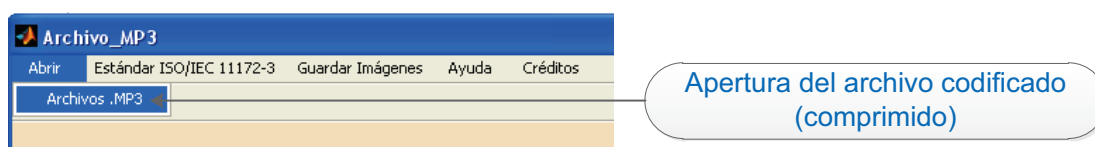


Figura A. 33: Opción para abrir archivo codificado *.MP3

Al seleccionar la opción "Archivos .MP3" del menú "Abrir" se desplegará una ventana, la cual permite buscar, seleccionar y abrir el archivo codificado, la ventana como se muestra a continuación está condicionada para que sólo se desplieguen archivos *.mp3 con ello se garantiza que no se cometan errores al elegir el archivo de audio de codificado; si bien lo lógico es que se abra el archivo codificado, es posible abrir cualquier tipo de archivo cuya extensión sea MPEG-1 Layer III, es decir en formato (*.mp3).

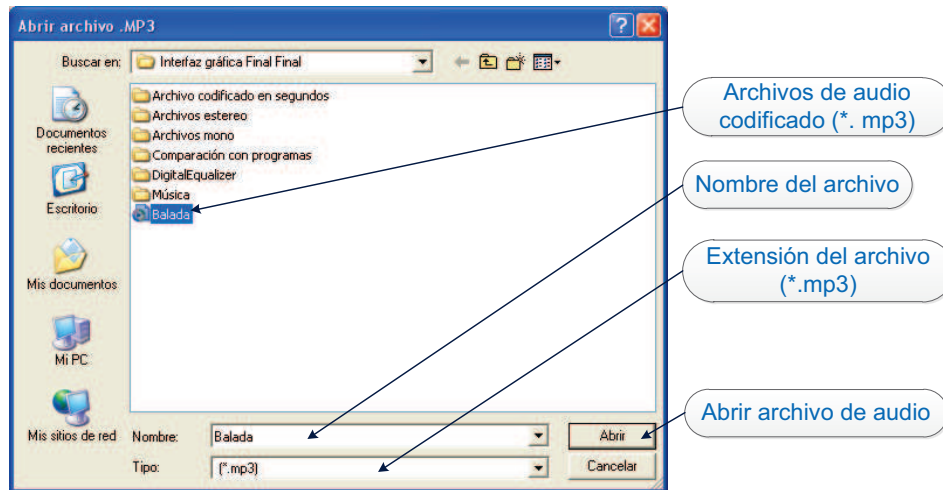
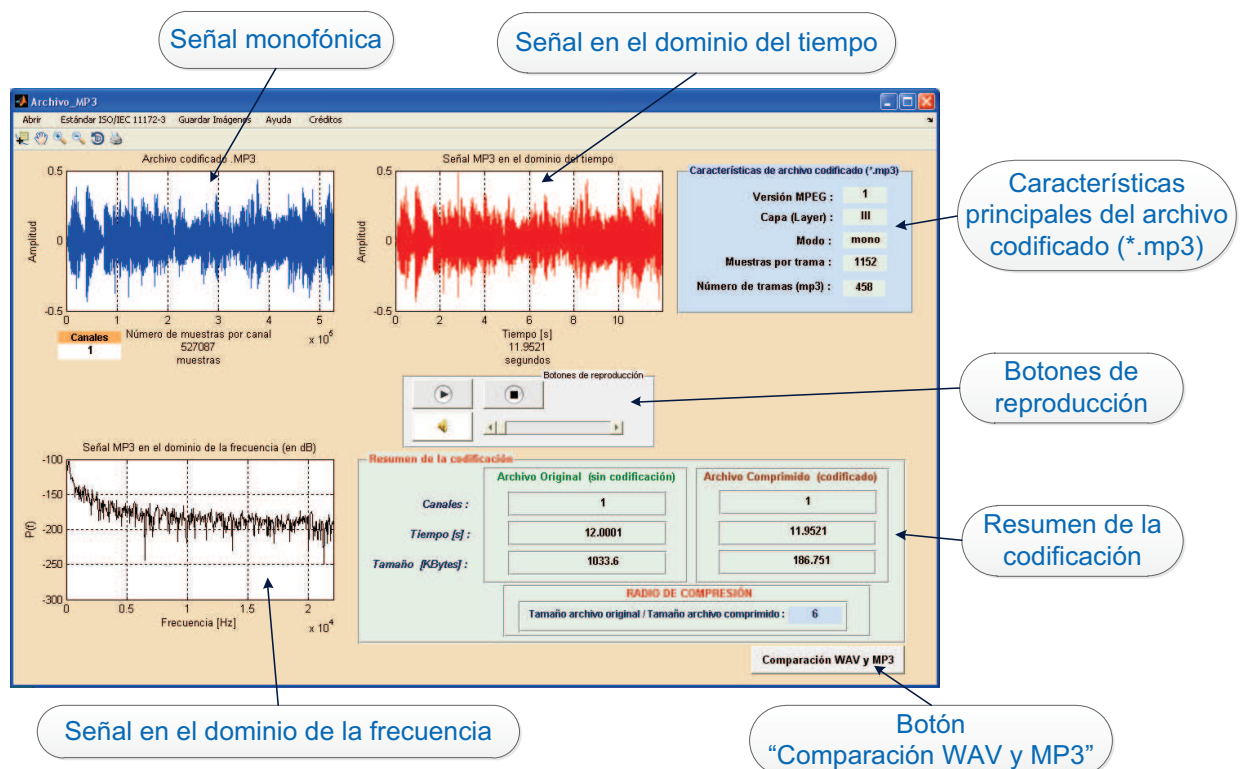


Figura A. 34: Selección y apertura del archivo del archivo codificado (*.mp3)

Una vez que se abre el archivo comprimido, se muestra el procesamiento de la señal de manera secuencial, se muestran varias características del archivo codificado, y además en la parte inferior de la ventana se muestra un cuadro de resumen del proceso de codificación, en el cual se indican características del archivo original (*.wav) y características del archivo comprimido (*.mp3).



Las principales características del archivo codificado, se muestran en la figura A.36, entre dichas características están: la versión de MPEG, la capa correspondiente, el modo del archivo codificado (monofónico), el número de muestras por trama (tamaño de la trama mp3) y el número de tramas del archivo codificado. Estas características son el resultado del proceso de codificación, como se conoce el resultado esperado de la codificación es un archivo MPEG-1 Layer III (*.mp3) con un tamaño por trama de 1152 muestras y que esté en modo monofónico, si hubiese un error en la codificación o simplemente el proceso no ha sido exitoso, los resultados que se mostrarían serían diferentes a los esperados o en su defecto no se mostraría información alguna.

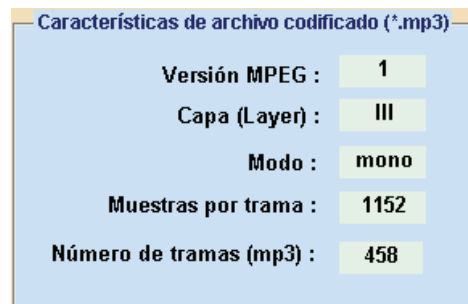


Figura A. 36: Características del archivo codificado

En cuanto al cuadro de resumen, se muestran las características generales de los archivos no codificado y codificado, como se observa en la figura A.37.

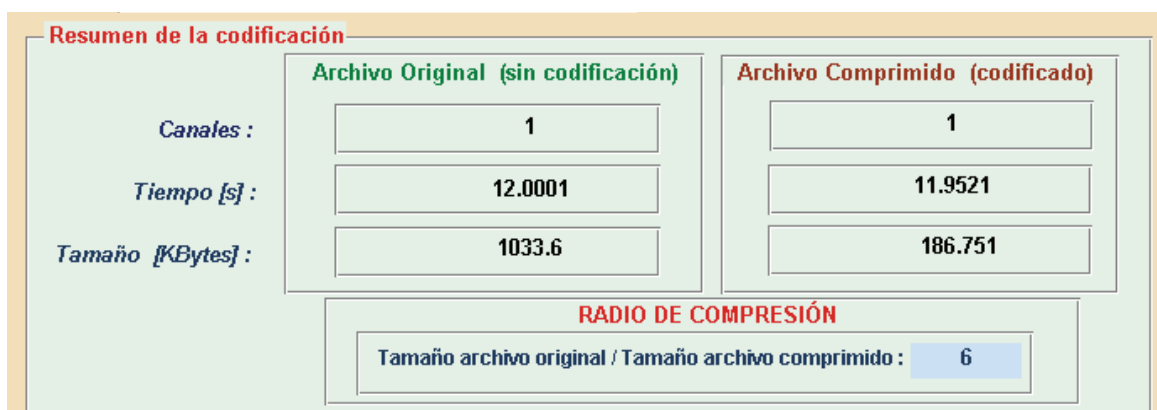


Figura A. 37: Resumen de la codificación

Entre dichas características constan: el número de canales, el tiempo de duración de los archivos (en segundos), el tamaño de los archivos y un parámetro

importante que se muestra es el "radio de compresión", el cual indica cuántas veces se ha reducido el archivo original con respecto al archivo comprimido.

Adicionalmente en la ventana "Archivo_MP3" una vez que se han procesado todas las señales y se han mostrado todos los cuadros de información, aparece en la parte inferior derecha, un botón "Comparación WAV y MP3", el cual al ser seleccionado da paso a la ventana "Comparacion_WAV_MP3".

A.2.4 VENTANA Comparación_Wav_Mp3

En esta ventana se muestra información detallada de las características tanto del archivo original (*.wav) como del archivo codificado (*.mp3), además de visualizar las características de los archivos, es posible visualizar los archivos de audio tanto en el dominio del tiempo como en el de la frecuencia. Al igual que en las ventanas anteriores en esta ventana se puede escuchar los archivos de audio (no codificado y codificado) con el uso de los botones de reproducción.

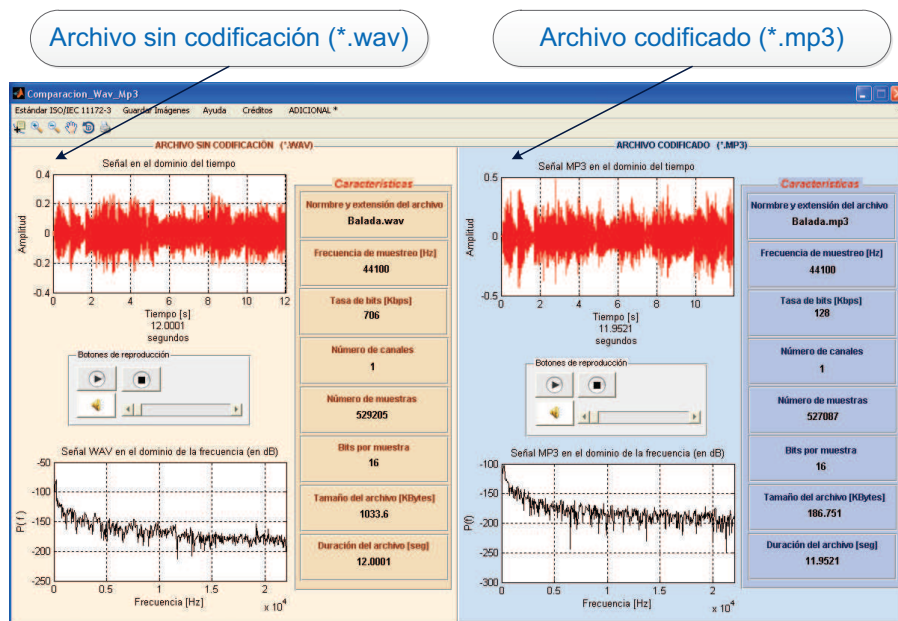


Figura A. 38: Comparación de las características de los archivos sin codificación (izquierda) y con codificación (derecha)

En cuanto a las características de los archivos, las características que se muestran en la ventana "Comparación_Wav_Mp3" son: el nombre del archivo, la

extensión de los archivos de audio, la frecuencia de muestreo, la tasa de bits, el número de canales, el número de muestras, el número de bits por muestra, el tamaño del archivo en Kbytes y la duración del archivo en segundos.

A.3 MENÚ DEL CODIFICADOR

El codificador en general está estructurado en varias ventanas como se indicó anteriormente, y cada ventana a excepción de las ventanas “MPEG1_LayerIII” (ventana de la carátula) y ventana “Autores” constan de una barra de menús y una barra de herramientas, las mismas que difieren en pocas características dependiendo de las ventanas en las cuales se está trabajando. A continuación se describen los menús presentes en las barras de menús de las diferentes ventanas del codificador.

A.3.1 MENÚ ABRIR

Este menú en conjunto con su opción, está disponible en las ventanas “Codificador_MP3” y “Archivo_MP3”, al seleccionar la opción “Archivos .WAV” o “Archivos .MP3” se puede seleccionar y abrir el archivo correspondiente.

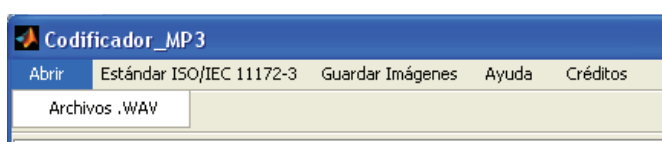


Figura A. 39: Menú “Abrir” y opción “Archivos .WAV”

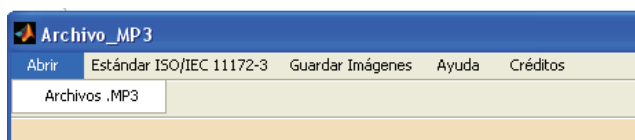


Figura A. 40: Menú “Abrir” y opción “Archivos .MP3”

A.3.2 MENÚ ESTÁNDAR ISO/IEC 11172-3

Dicho menú está disponible en las ventanas “Codificador_MP3”, “Archivo_MP3” y “Comparacion_WAV_MP3”; este menú consta de cinco opciones las cuales son: Especificaciones de audio MPEG (ISO/IEC 11172-3), ANEXOS A y B, ANEXO C, ANEXO D, ANEXOS E, F Y G; al seleccionar cualquiera de estas opciones se abre un archivo *.pdf, correspondiente a la opción seleccionada, como se observa en la figura A.41. Estos archivos *.PDF proporcionan información del Estándar Internacional ISO/IEC 11172-3.

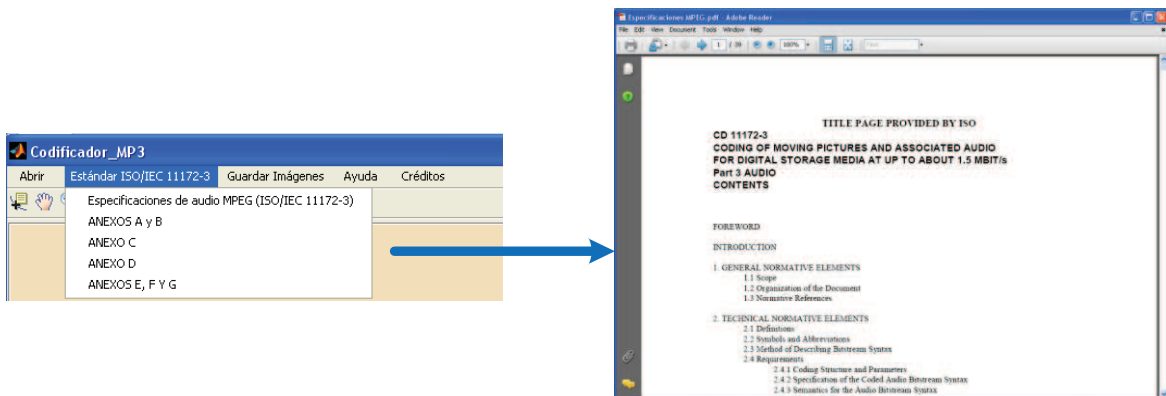


Figura A. 41: Menú “Estándar ISO/IEC 11172-3” y opciones correspondientes

A.3.3 MENÚ GUARDAR IMÁGENES

Las ventanas: “Codificador_MP3”, “Archivo_MP3”, “Comparacion_WAV_MP3” y “Efectos” tienen en su barra de menús, el menú Guardar Imágenes. Las opciones de este menú pueden variar en el nombre, es decir si la imagen corresponde al archivo no codificado o codificado, pero la funcionalidad de las opciones en todas las ventanas es la misma, guardar las imágenes de las señales procesadas.

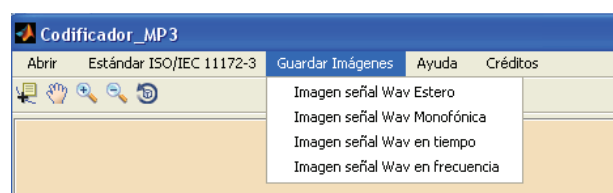


Figura A. 42: Menú “Guardar Imágenes” y opciones correspondientes

Al seleccionar la opción correspondiente de la imagen que se desee guardar, se abre una ventana, figura A.43, en la cual se puede elegir la ubicación y el nombre de la imagen de la señal requerida, por defecto, facilidad y compatibilidad con los sistemas de computación, el formato con el cual se guardan las imágenes es el *.jpg.

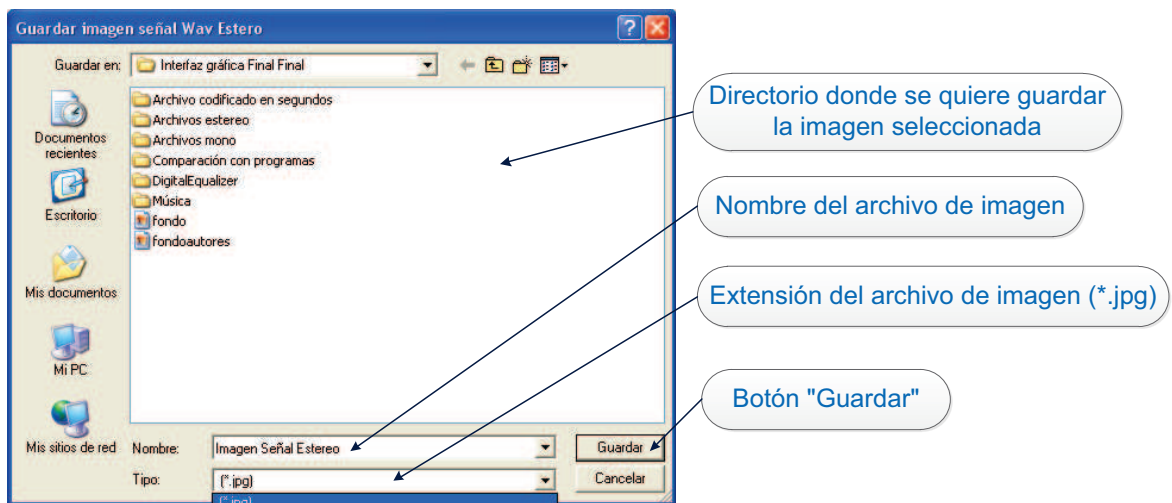


Figura A. 43: Almacenamiento de la imagen de la señal correspondiente

A.3.4 MENÚ AYUDA

Este menú proporciona información concerniente a los aspectos técnicos del programa y a la utilización en sí del programa. La ayuda la constituyen el manual de usuario en formato *.pdf y un video tutorial *.avi.

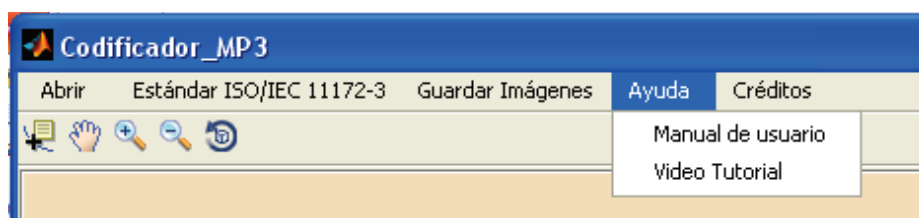


Figura A. 44: Menú "Créditos" y submenú "Autores"

A.3.5 MENÚ CRÉDITOS

El menú “Créditos” está presente en las ventanas: “Codificador_MP3”, “Archivo_MP3”, “Comparacion_WAV_MP3” y “Efectos”. Al seleccionar la opción “Autores” del menú “Créditos”, se despliega una ventana, figura A.46, la cual muestra una breve biografía de los autores y director del proyecto.

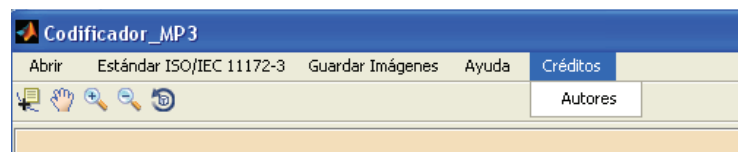


Figura A. 45: Menú “Créditos” y submenú “Autores”

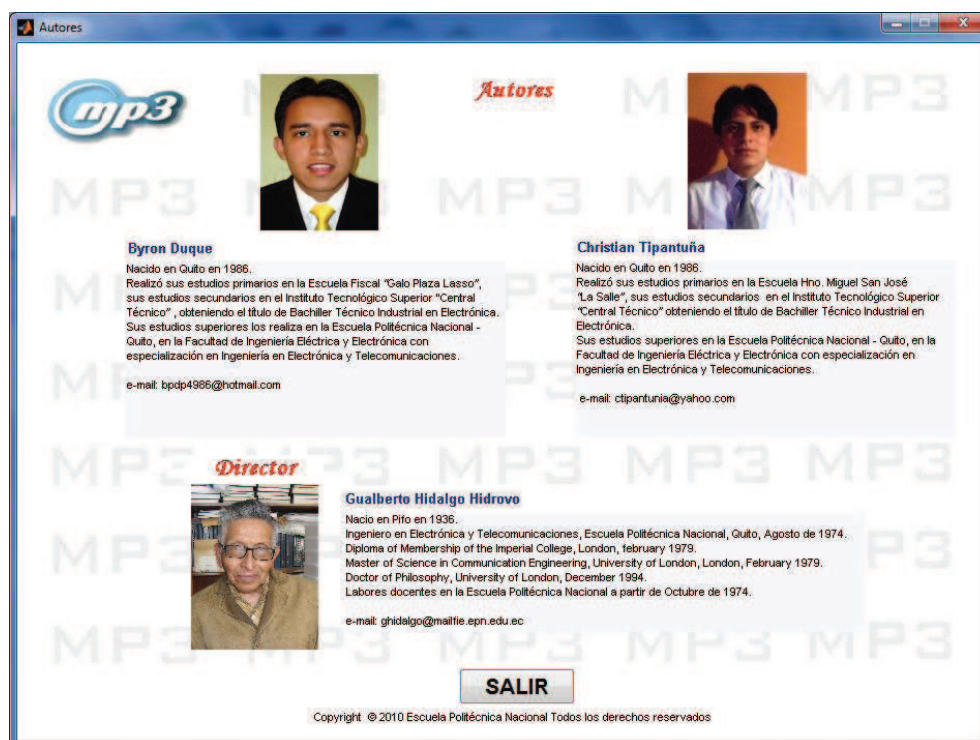


Figura A. 46: Biografía resumida de autores

A.3.6 MENÚ ADICIONAL

Este menú adicional, está disponible sólo en la ventana “Comparacion_Wav_Mp3”, no tiene tanta relación con el codificador en sí, sin embargo como se está trabajando con archivos de audio, se ha considerado conveniente añadir la ventana “Efectos” la cual a través de sus distintas opciones

permite realizar un procesamiento digital de las señales de audio, permitiendo modificar ciertas características de los archivos de audio.

Para tener acceso a la ventana “Efectos” se selecciona del menú “ADICIONAL *” la opción “Efectos de audio”, como se muestra en la figura A.47.

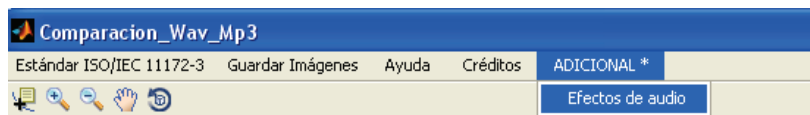


Figura A. 47: Menú “ADICIONAL *” y opción “Efectos de audio”

Una vez seleccionada la opción “Efectos de audio” se visualizará en la pantalla una ventana como la que se muestra a continuación.



Figura A. 48: Ventana principal de los efectos de audio

En su barra de menús la ventana “Efectos” consta del menú “Abrir” el cual a su vez tiene dos opciones, como se observa en la figura A.49, dichas opciones permiten abrir archivos de audio no codificados (*.wav) y codificados (*.mp3) uno a la vez.



Figura A. 49: Apertura de archivos *.WAV o *.MP3 para aplicar efectos de audio

Al seleccionar la opción “Archivos .WAV” o “Archivos .MP3”, se abre una ventana de búsqueda que permite seleccionar el archivo de audio deseado ya sea *.wav o *.mp3.

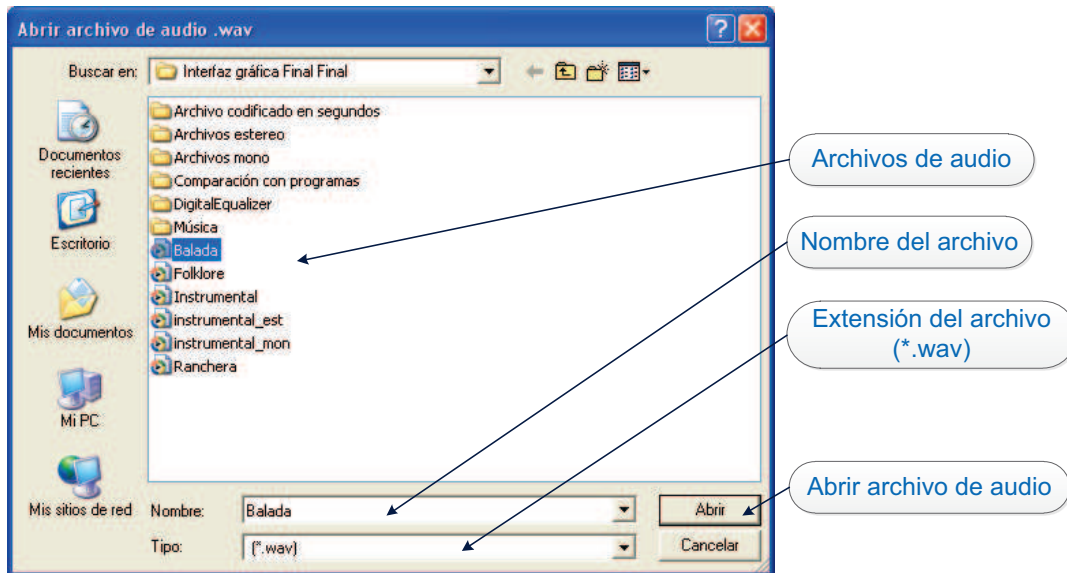


Figura A. 50: Apertura del archivo de audio para aplicar el efecto deseado

Al seleccionar y abrir el archivo requerido, el mismo se despliega en la parte izquierda de la pantalla, tal como se presenta en la figura A.51, tanto en el dominio del tiempo como en el de la frecuencia, además con la ayuda de los botones de reproducción es posible escuchar el archivo seleccionado (archivo sin efectos de audio).

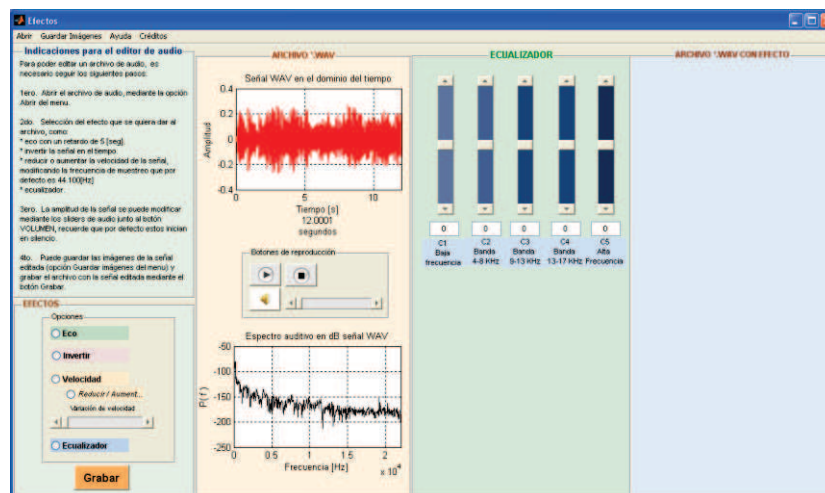


Figura A. 51: Procesamiento del archivo original (archivo sin efectos)

Teniendo el archivo con el que se va a trabajar es posible añadir los efectos que se requiera. Para poder añadir un efecto o varios efectos al archivo de audio original se lo puede hacer seleccionando una de las opciones del menú EFECTOS, figura A.52.

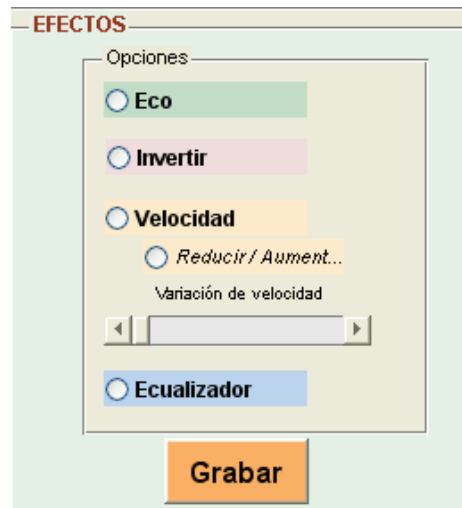


Figura A. 52: Efectos de audio disponibles

Son cuatro las opciones del menú “EFECTOS”, la primera opción es añadir eco con un retardo de 0,5 segundos; la segunda opción es invertir la señal en el tiempo; la tercera opción es reducir o aumentar la velocidad de la señal modificando la frecuencia de muestreo y la última opción es la del ecualizador, un ecualizador que consta de cinco bandas de frecuencia y que permite manipular las bandas de frecuencia desde las bajas hasta las altas frecuencias.

Al seleccionar el efecto (eco, invertir o velocidad) el efecto seleccionado se activa y puede ser utilizado para modificar las características del archivo original, la señal resultante de la adición del efecto es desplegada en la parte derecha de la ventana “Efectos”, como se indica en la figura A.53.

Además con la ayuda de los botones de reproducción se puede escuchar los cambios originados por la aplicación del efecto.

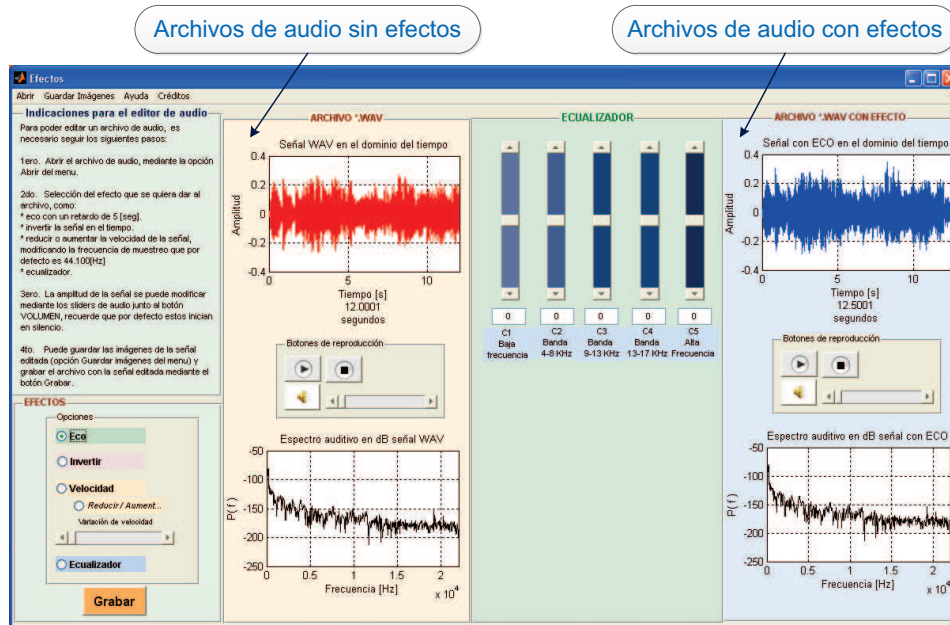


Figura A. 53: Archivos de audio con efectos

Cada efecto para diferenciarse uno de otro es mostrado en la pantalla con un color diferente, como se aprecia en la figura A.54.

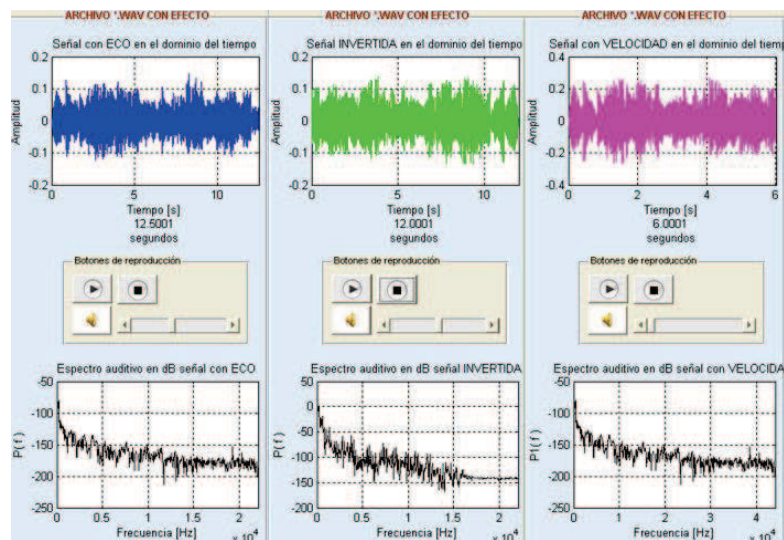


Figura A. 54: Efectos: eco, inversión temporal y variación de la velocidad
(de izquierda a derecha)

Si se utilizase el efecto “Velocidad” es necesario antes de que el efecto actúe, seleccionar la opción “Reducir/Aumentar” y variar la posición del slider como se muestra a continuación, para que el efecto tenga consecuencia.

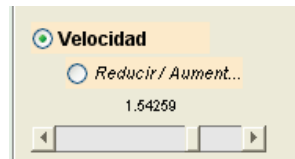


Figura A. 55: Efectos:

Al seleccionar la opción “Ecuador” el ecualizador inicialmente muestra todas sus bandas de frecuencia al mismo nivel; para que tenga efecto la ecualización es necesario desplazar cualquiera de las bandas de frecuencia verticalmente, y automáticamente los cambios realizados pueden ser observados, al seleccionar el botón “Característica” y ser escuchados al utilizar los botones de reproducción.

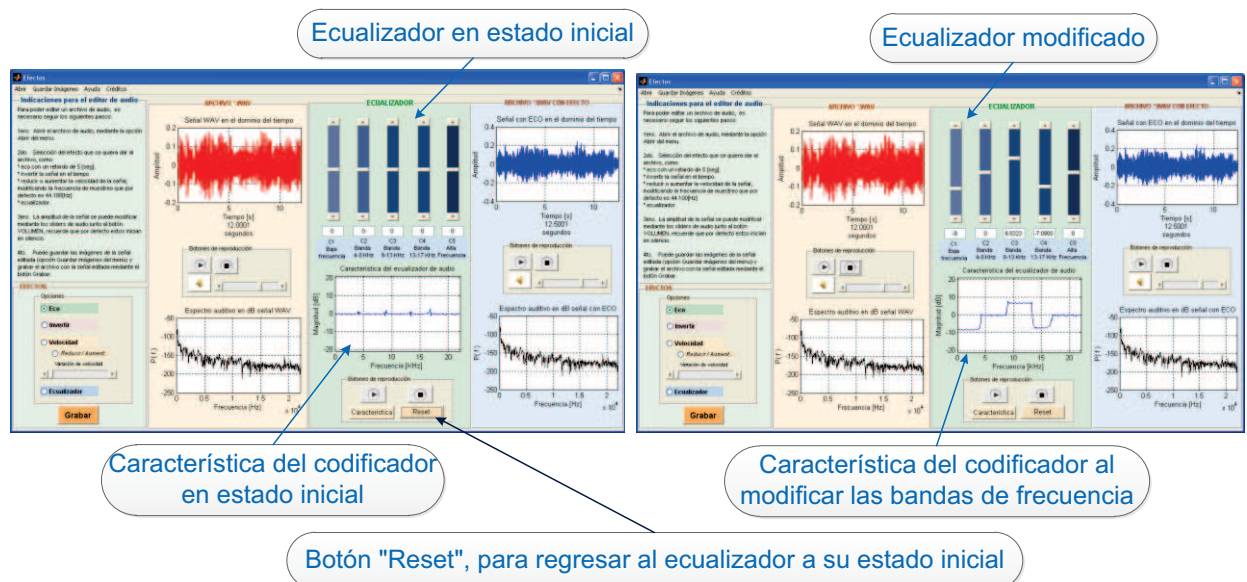


Figura A. 56: Utilización del ecualizador

Una vez que se han realizado los cambios necesarios (adición de efectos de audio) el archivo de audio modificado (archivo de audio con efectos) puede ser almacenado en formato *.wav o *.mp3, según corresponda.

Para poder grabar el archivo de audio con efectos, se selecciona del menú “EFECTOS” el botón “Grabar” (Figura A.52), con ello se despliega una ventana como se indica a continuación, la cual permite seleccionar la ubicación y el nombre del archivo de audio con efectos.

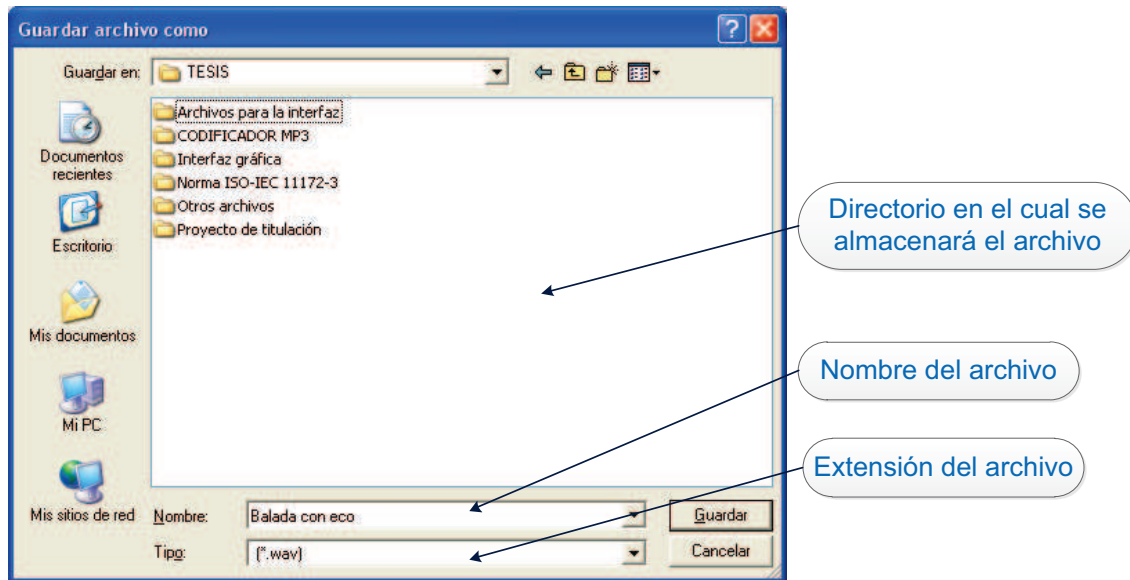


Figura A. 57: Grabación del archivo de audio con efectos

El archivo de audio almacenado en el directorio seleccionado, puede ser reproducido utilizando cualquier software de audio comercial o dispositivo de reproducción.

A.4 RESOLUCIÓN DE PROBLEMAS

El programa cuenta con mensajes de advertencia los cuales son desplegados al ingresar un parámetro incorrecto o estar ejecutando el programa de manera equivocada. Los mensajes son desplegados en su totalidad al estar trabajando en la ventana “Codificador_MP3” ya que en dicha ventana es necesario que el usuario seleccione el archivo de audio de entrada e ingrese los distintos parámetros necesarios para la codificación.

A.4.1 ERROR EN LA FRECUENCIA DE MUESTREO

Si el archivo de audio seleccionado es abierto y tiene una frecuencia de muestreo diferente a 44.100 Hz, aparecerá de inmediato en la pantalla un mensaje de error, figura A.58, indicando que la frecuencia de muestreo es diferente de 44.100 Hz, ya que una de las condiciones del archivo de audio de entrada es que tenga dicha frecuencia.

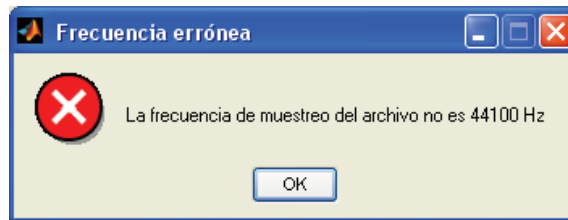


Figura A. 58: Error en la frecuencia de muestreo

A.4.2 ERRORES EN EL INGRESO DEL TIEMPO DE CODIFICACIÓN

A.4.2.1 Ingreso de un tiempo de codificación muy pequeño

Como se indicó anteriormente al seleccionar la tasa de bits, se despliega un cuadro de edición para ingresar el tiempo de codificación requerido, a la par de esto se indican los tiempos mínimos y máximos de codificación (figura A.27) y si el tiempo ingresado es inferior al mínimo se va a tener problemas a futuro en la codificación, por ello para evitar de antemano que suceda lo mencionado, se despliega el mensaje que se observa a continuación, indicando que la cantidad de tiempo es insuficiente para que así se vuelva a ingresar un tiempo adecuado.

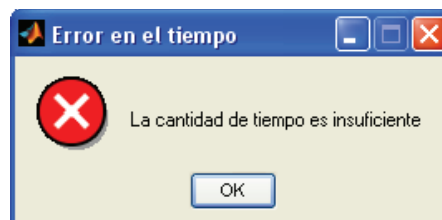


Figura A. 59: Error al ingresar un tiempo muy pequeño

A.4.2.2 Ingreso de un tiempo de codificación muy grande

Si por el contrario el tiempo excede el máximo indicado en la figura A.27, se desplegará un mensaje de error como el que se muestra a continuación para que se corrija este error y en el cuadro de edición se ingrese un tiempo adecuado.

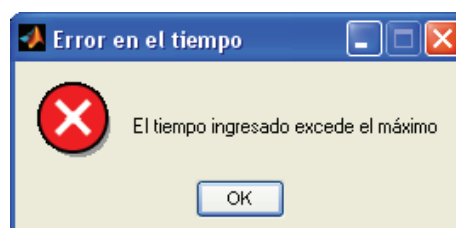


Figura A. 60: Error al ingresar un tiempo muy grande

A.4.2.3 Ingreso de un valor no numérico

Si se ingresa en el cuadro de edición (figura A.28) una letra o carácter no numérico, aparece un mensaje de error como el de la figura A.61, indicando que se elimine dicho carácter y se ingrese un número en un rango válido (entre el mínimo y el máximo permitido).

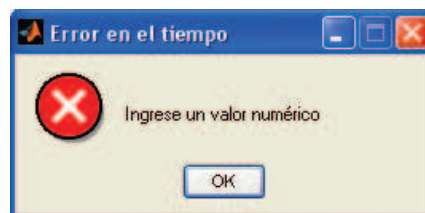


Figura A. 61: Mensaje de error al introducir un valor no numérico (Ejemplo: letra "A")

A.5 RECOMENDACIONES

Como el programa, en el momento de la codificación del archivo de audio demanda de muchos recursos del CPU (procesador y memoria) como se muestra en la figura A.61, es recomendable no correr o ejecutar otra aplicación.

Además para poder tener un rendimiento aceptable es recomendable correr el programa en una computadora que por lo menos conste de un procesador Pentium IV y 512 MB de RAM.

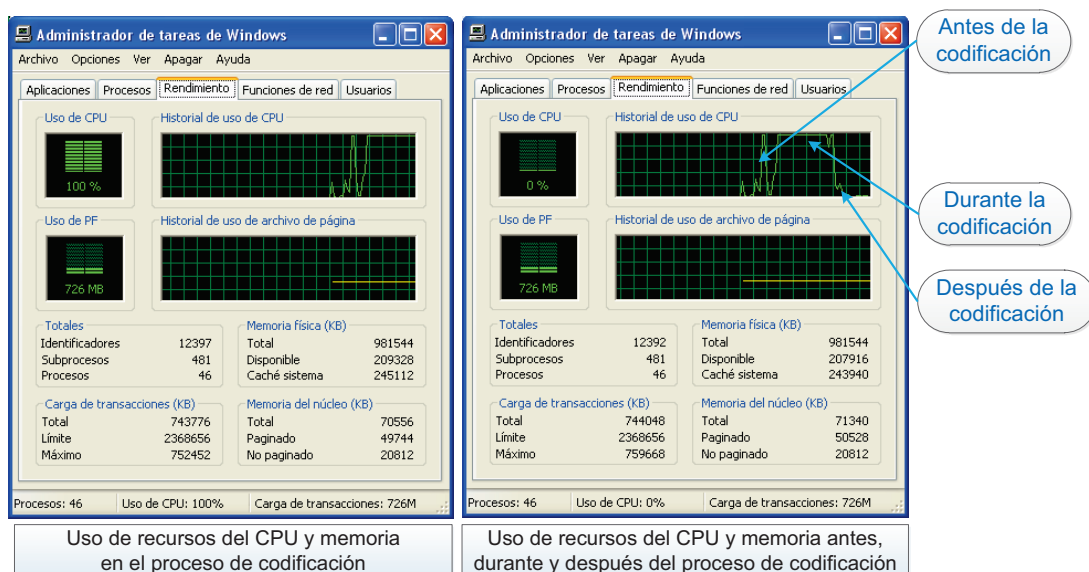


Figura A. 62: Uso de recursos (CPU y memoria RAM)

ANEXO B

INSTALACIÓN DEL TOOLBOX MP3 EN MATLAB

B.1 DESCRIPCIÓN DE LA INSTALACIÓN

En ocasiones al diseñar o ejecutar ciertos programas, resulta indispensable instalar librerías adicionales a las que se instalan por defecto en el momento de la instalación de Matlab, la instalación de las librerías o toolboxes en el caso de Matlab, puede ayudar a que el programa funcione si no lo hace, o mejore su desempeño.

Para el caso del ejemplo que se va describir a continuación, el primer punto es conseguir el toolbox, es decir, el “mp3_toolbox”; dicho toolbox u otros que sean necesarios, pueden ser descargados desde la página oficial de The Mathworks Inc., o de páginas vinculadas al producto.

Asociado a The Mathworks Inc., existe un sitio web denominado Matlab Central, el cual consta de información, toolboxes y programas de propiedad de los ingenieros de soporte de The Mathworks Inc. o de los usuarios de Matlab o Simulink, dicho sitio web cuya dirección es: <http://www.mathworks.com/matlabcentral/>, constituye una ayuda para personas de todos los niveles, desde aquellos que se están iniciando en el mundo de Matlab y Simulink, hasta personas que tienen mediana o mucha experiencia al utilizar Matlab o Simulink.

Específicamente el toolbox requerido para que el codificador cumpla su funcionamiento a cabalidad, puede ser descargado de la siguiente dirección: <http://www.mathworks.com/matlabcentral/fileexchange/13852-mp3read>.

B.2 INSTALACIÓN DEL TOOLBOX MP3

B.2.1 ALMACENAMIENTO DEL TOOLBOX

Una vez que se obtiene el toolbox, y se lo descomprime si así fuera, hay que almacenarlo en la carpeta donde se encuentran los demás toolboxes, la ruta donde se encuentra el directorio puede variar, dependiendo de la configuración del equipo o la selección de los parámetros de instalación; para este caso la dirección en donde se encuentran almacenados todos los toolboxes es: C:\Archivos de programa\MATLAB\R2009a\toolbox.

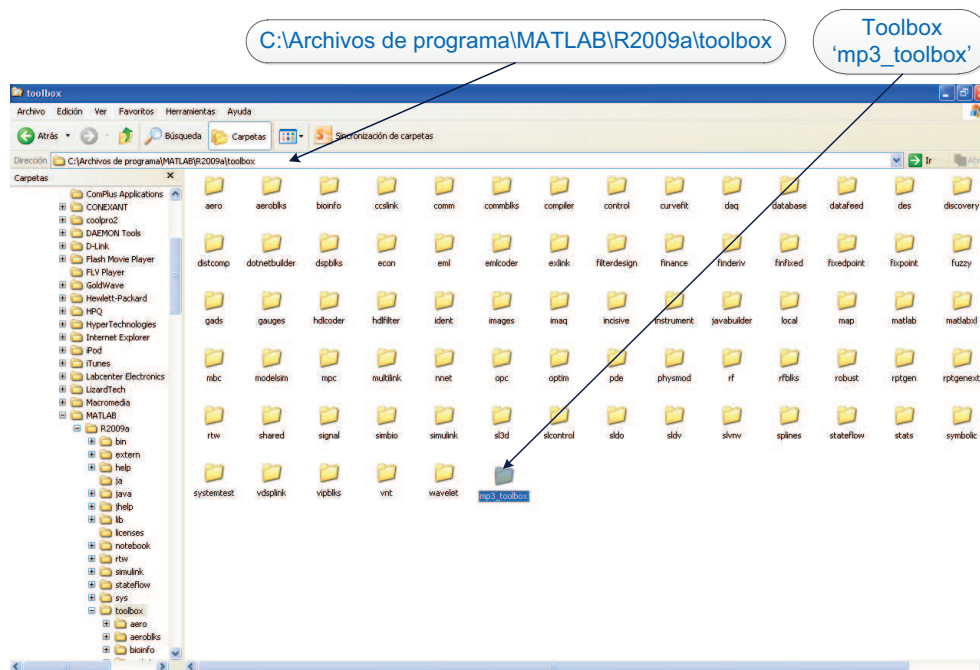


Figura B. 1: Almacenamiento del toolbox en el directorio correspondiente

B.2.2 ADICIÓN DE LA RUTA DE ACCESO AL TOOLBOX

Tener almacenado el toolbox en el directorio correcto no es suficiente para poder utilizarlo, es necesario actualizar la ruta de acceso al toolbox, para ello hay que manualmente guardar la información necesaria de las rutas de acceso. Como primer punto, del menú File se selecciona la opción “Set Path...” o seleccionar ruta, tal y como se observa en la figura B.2.

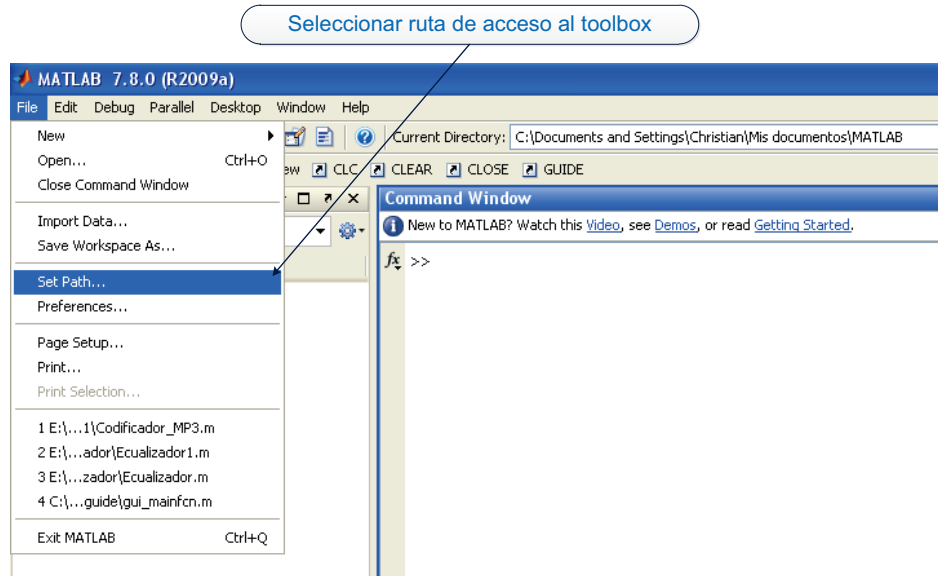


Figura B. 2: Direccionamiento de la ruta donde se encuentra el toolbox

Luego de seleccionar la opción “Set Path...” aparecerá una ventana la cual consta de varias opciones para añadir la carpeta o subcarpetas pertenecientes al toolbox requerido, es recomendable seleccionar la “Add with Subfolders...” como se indica en la figura B.3, seleccionando esta opción se garantiza que todas las carpetas y subcarpetas del toolbox se añadan a la ruta de acceso al toolbox.

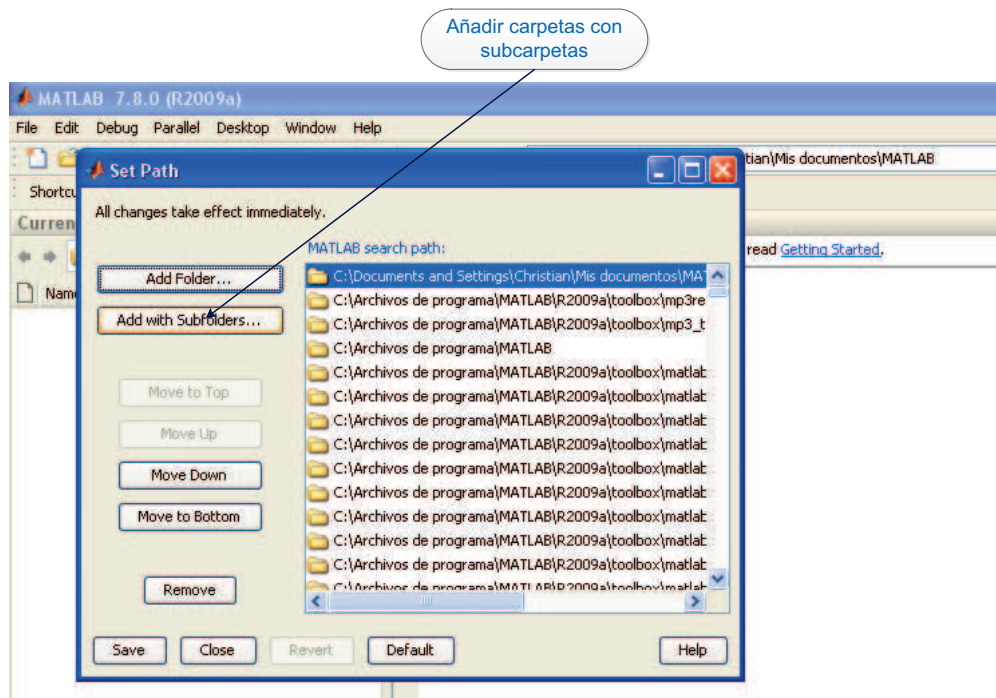


Figura B. 3: Adición de carpetas y subcarpetas pertenecientes al toolbox

A continuación, se despliega una ventana de exploración, en la cual se debe ubicar el toolbox que se desea activar.

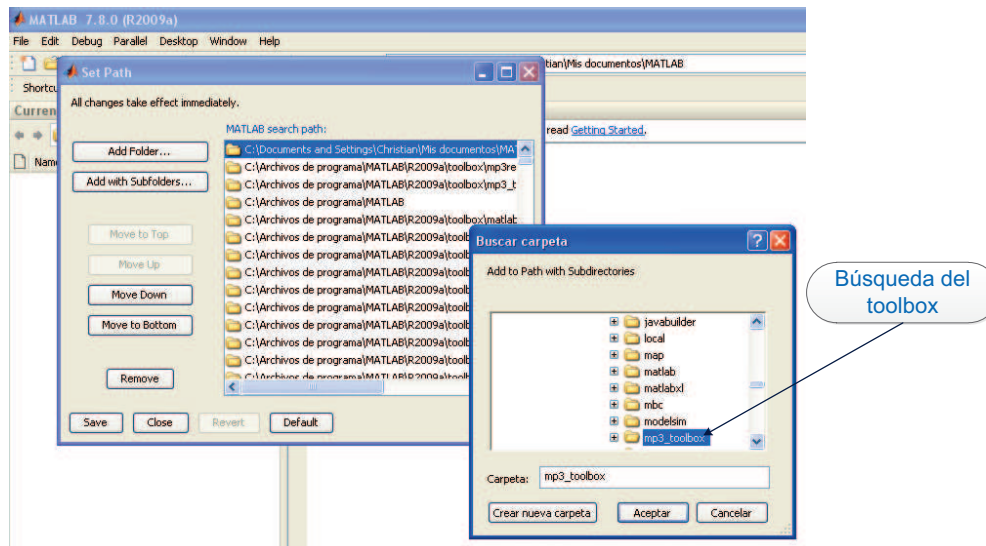


Figura B. 4: Direccionamiento del toolbox requerido

Una vez que se ha localizado y seleccionado el toolbox, es necesario guardar los cambios para que estos tengan efecto, es decir para poder acceder a las utilidades de toolbox; es necesario seleccionar la opción “save” tal y como se muestra en la figura B.5.

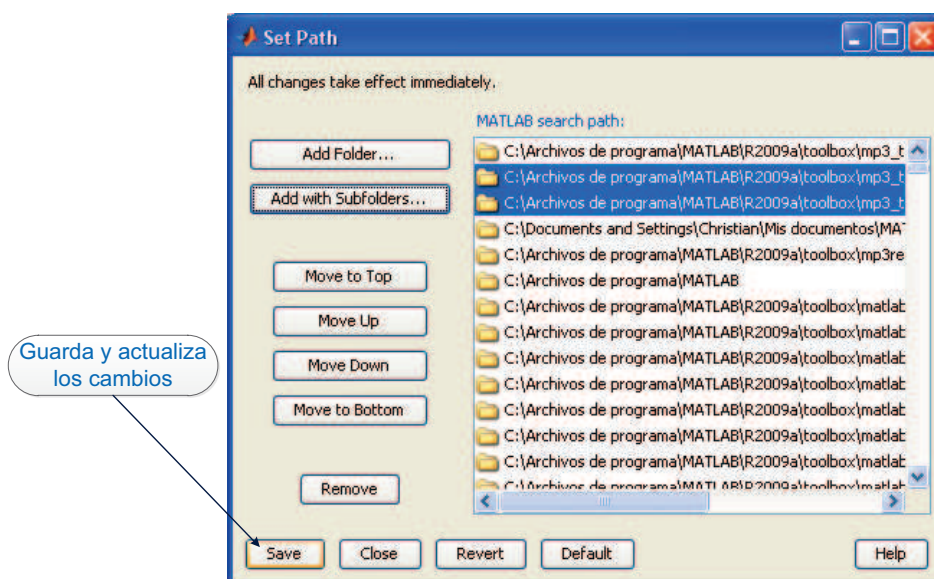


Figura B. 5: Almacenamiento y actualización de cambios

B.2.3 COMPROBACIÓN DE LA INSTALACIÓN DEL TOOLBOX

Cumpliendo a cabalidad todos los pasos mencionados anteriormente, finalmente el toolbox queda almacenado, y activado para ser utilizado. Una de las maneras más sencillas para comprobar que el toolbox está listo y funcionando adecuadamente, es probar la o las funcionalidades para las que está diseñado.

Cabe mencionar que por lo general todo toolbox o programa complementario que se desee instalar, consta de instrucciones tanto de requerimientos como de funcionamiento, por ese motivo es recomendable revisar la guía o manual de usuario de cada toolbox antes de utilizarlo; para conocer sus potencialidades y sus limitaciones.

En el caso del ejemplo, el toolbox instalado (“mp3_toolbox”) consta de varias funciones asociadas al toolbox, una de ellas es “mp3read”, por ello para comprobar si la instalación ha sido exitosa se ha procedido a solicitar la ayuda de dicha función a través del comando “help”, y como se observa en la figura B.6 la ayuda ha sido exitosa, ya que al solicitar la ayuda, automáticamente se despliega el texto de información asociado a la función mp3read, si el toolbox no estaría instalado de una manera correcta en lugar de la información de la función correspondiente aparecería el texto “mp3read not found” en la ventana de comandos.

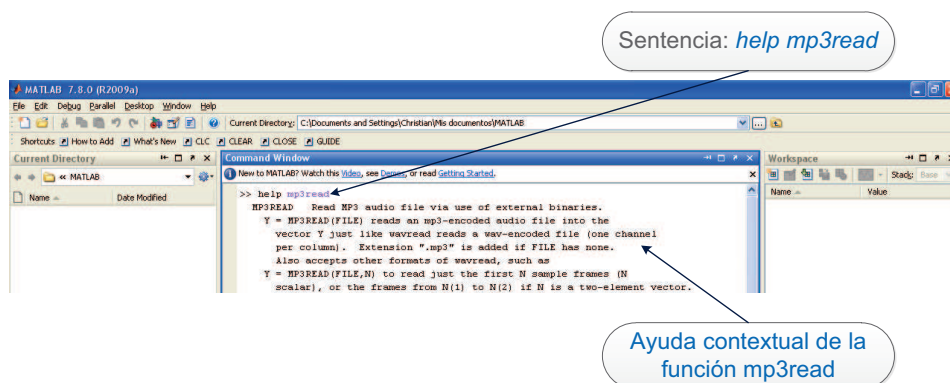


Figura B. 6: Comprobación del funcionamiento del toolbox