UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE INGENIERIA

“MP3 Formato de compresión de audio digital: ánalisis y proyección a futuro”

Tesis

Que para obtener el título de:

Ingeniero Eléctrico Electrónico

Presentan

Canales Lizaola, Francisco Adrián
Martínez Araiza, Omar

Director de Tesis: Ing. Francisco Rodríguez Ramírez

México, D.F. 2008

AGRADECIMIENTOS:

A la Universidad Nacional Autónoma de México y a la Facultad de Ingeniería que me dio la oportunidad de utilizar sus instalaciones para desarrollarme como ser humano.

A mis profesores que me enseñaron con dedicación y con diferentes puntos de vista los conocimientos necesarios para comprender todo lo que nos rodea y así percibir el mundo desde una perspectiva más amplia.

A mis padres, que siempre han estado a mi lado dándome su apoyo y su amor incondicional; y que con sus enseñanzas y consejos me guiaron por el camino de la honestidad y de la superación el cual me ha llevado a conseguir las metas que me he propuesto.

A mis hermanas, Adriana, Monica y Laura, con las cuales he vivido experiencias que me han enseñado la importancia de la familia.

A mi primo Daniel que me enseño que el trabajo duro y la perseverancia siempre son bien recompensados.

A mis amigos Angel, Luis Ciro y Mauricio, con los cuales he compartido infinidad de experiencias y se que siempre estarán ahí para reír o para apoyarme en momentos difíciles.

A mi compañero de tesis, Omar Martínez, que emprendió este viaje conmigo y que con este trabajo se ve culminado, que compartió conmigo momentos difíciles y momentos de alegría durante la realización de este trabajo y que en todo momento fue un gran amigo.

A mi asesor, Francisco Ramirez por su apoyo y su infinita paciencia que nos ayudo a terminar este trabajo.

A todos mis amigos y personas que me acompañaron durante la carrera que me hicieron disfrutar cada clase y cada momento durante esta, que sin ellos la vida en la facultad hubiera sido insoportable.

A todos y cada uno muchas gracias.

« Là où est votre trésor, là aussi sera votre cœur »

San Lucas

Francisco Adrian Canales Lizaola.

Agradecimientos

Este trabajo se lo dedico a mi gran inspiración, a la persona que siempre ha estado conmigo en los momentos más difíciles y ha sido mi gran apoyo, mi guía, mi amigo, mi héroe, mi hermano. Gracias Coco por todo, gracias por darme la oportunidad de poder ser tu hermano y me doy cuenta que la vida sin un ejemplo como tú hubiera sido muy aburrida, gracias por los consejos que siempre han sido en el momento exacto y justo, gracias pues por existir, te quiero mucho mi querido hermano.

Sin sus consejos y apoyo no hubiera llegado a ningún lado, gracias por dejarme sar tan independiente y dejarme ser yo mismo, gracias por amarme como lo hacen y por tener fe ciega en mi, espero que sea un orgullo para ustedes como lo son para mí, gracias padres por todo. Gracias Madre por todo tu apoyo y consejo; es cierto que no hablamos mucho de mi vida personal pero créeme que un solo abrazo tuyo arregla todos los problemas y me hace sentir fuerte otra vez, eres el bastión de la casa y de nuestras vidas, gracias por la educación que me diste y por amarme. Gracias Padre, por enseñarnos el significado de la responsabilidad y por guiarnos por el camino a ser buenos hombres, espero que algún día pueda ser tan buen padre como tu lo has sido con mi hermano y conmigo, gracias por atenderme tan bien, te estas ganando una propina!!! Los amos a los dos!!!!

Gracias a Sylvana por ser tan maravillosa mujer, por amar a mi hermano, por apoyarlo y por que en ti encontró esa persona que me dijo que existe con la cual TODO hace clic y TODO tiene sentido. Gracias por tu apoyo y tu cariño, gracias cuñada y espero que pronto me conviertan en tío.

A mi adorada Keisha quien me mostró que un ser tan pequeño puede querer incondicionalmente y que siempre estará ahí para darme muchos besos cuando la cargue. Gracias querida hija por que tu siempre has estado presente en estos maravillosos 10 años y espero que nos dures otros 10 más. A Vocho, que a pesar de que acabas de llegar a nuestras vidas eres todo un relajo y se ve que nos quieres mucho como nosotros a ti.

A Francisco Miranda y familia por darme la oportunidad de entrar a este maravilloso mundo del audio y además darme muchas lecciones de vida. Gracias por confiar en un niño de 16 años y darle su apoyo, todo lo que me han enseñado es verdaderamente maravilloso. Gracias a ustedes aprendí mucho de todo, mis mejores deseos hoy y siempre.

A Humberto Terán por confiar en la juventud y darme la oportunidad de trabajar contigo. Gracias por todos esos buenos consejos que me has dado, son bastante filosóficos y algunos todavía no los entiendo, pero que compartas un poco de tus conocimientos es invaluable, gracias amigo.

A Francisco “Sabu” por que en este camino de la vida nos volvimos a encontrar en la facultad y además tuvimos las ganas de hacer un proyecto juntos, a pesar de todas las dificultades que se nos presentaron y nuestros malditos genios combinados terminamos la tesis, gracias además por tu amistad y sí lo sé te sigo debiendo una gringa.

A mi adorada UNAM, que tanto me negue a ella, pero al final ella es quien me ha dado toda la educación y una gran vida, que orgullo poder decir que estudié en tus aulas, infinitamente gracias mi querida UNAM, y gracias a los PUMAS por el bicampeonato.

A Francisco Rodríguez por su infinita paciencia y ayudarnos a concluir de manera correcta con nuestra educación profesional de la mejor manera.

A mis tíos Esther, Ricardo, Cristina, Gilberto, Beatriz, Jaime y Ernesto (q.d.e.p.) por todo su cariño y apoyo. A mis primos Alejandro, Alejandra, Claudia, Sergio, Gilberto por que el tiempo que hemos pasado juntos ha sido muy divertido y ya después les enseño mi última locura, los quiero a todos.

A Oscar y Francisco "Ro" Mendoza por que me han apoyado siempre, están en los mejores momentos y también en los momentos de dolor, y gracias a esto ahora puedo decir con orgullo que tengo dos hermanos más, que buenas fiestas, que buen relajo. Gracias por su amistad mis queridos hermanitos.

A la banda, sin un orden en particular, Gabo por que gracias a ese consejo que nunca olvidaré el túnel sigue oscuro pero se que me están echando las luces para salir adelante y para no sentirme tan solito; Eli, por que eres a todo dar y una gran amiga; a Tony "Santo" por tu ayuda para poder terminar la tesis y por tu forma de ser que sólo alguien como tú podría tener; Tala, talita, Talayero gracias por esas pláticas tan interesantes y todos esperamos que nos operes la nariz cuando termines la especialidad, y recuerda que todas nuestras operaciones corren por tu cuenta; Val aunque no nos tratamos mucho eres a todo dar y además siempre con una sonrisa; Monica que manera tan extraña de comenzar una amistad, pero que bien que nos la hemos pasado, gracias por todo; David y Magali "La Maga" miren que conocerlos en Acapulco siendo yo el colado y poder decir que los considero mis amigos es un placer, gracias por todo. Que nuestra amistad perdure para siempre, que nuestros sueños se cumplan sobre todo que nunca se terminen nuestras preguntas indiscretas!!!!!. Gracias amig@s por todo, son lo máximo.

A Mintel por que tus consejos y tus regaños siempre han sido en el momento adecuado, esperemos que tengamos más trabajo en el futuro. Gracias a todos los vividores.

Al final, pero no por eso menos importante, a ti que me enseñaste tantas cosas, que me diste tu amor sin condición, que me entregaste una parte de tu vida como lo hice yo contigo, gracias por ese tiempo juntos, gracias por empujarme a terminar la tesis, gracias por estar en los momentos difíciles, gracias por un "bebé" de relación maravilloso, nunca te olvidaré, por esto y muchas cosas más que te diré en persona, un millón de gracias!!!!!! Gracias Jesse.

Y gracias a todas las personas que he conocido en este camino, que me han dejado enseñanzas, unas buenas, otras malas, pero siempre uno aprende de las personas que lo rodean y por eso siempre les estaré agradecido.

Omar Martínez Araiza.

INDICE

Capítulo I

INTRODUCCIÓN

Durante la última década, la calidad del audio digital de los discos compactos ha reemplazado al audio analógico. De esta manera, emergieron nuevas aplicaciones para el audio digital tales como su uso en redes y en multimedia, sin embargo estos sistemas se enfrentan con una serie de limitaciones tales como anchos de banda reducidos, capacidad limitada para almacenamiento y altos costos.

Estas nuevas aplicaciones han creado una demanda de obtener audio digital de alta calidad con tasas de bits bajas. En respuesta a esta necesidad, diversos investigadores se han dado a la tarea de desarrollar un algoritmo de compresión de audio digital que cumpla con dichas necesidades. Como resultado, muchos algoritmos han sido propuestos, y en muchos casos se han convertido en productos comerciales o estándares internacionales.

En el capítulo II se abordarán los conceptos básicos del sonido; empezando con los diferentes tipos de ondas y las principales características de estas. A continuación se hablará de las propiedades y fenómenos que caracterizan al sonido. Quedando establecido el concepto de sonido, se mostrará la estructura y funcionamiento del aparato auditivo para comprender con mayor claridad la última parte en que se divide este capítulo donde se abordan los conceptos básicos de la psicoacústica, este último tópico es de vital importancia dado que a partir de dichos principios se realiza la compresión de audio digital.

En el capítulo III tendremos un estudio sobre el audio digital; como se digitaliza el sonido [6], la teoría matemática necesaria para comprender la digitalización. Así mismo se hablará de los diferentes tipos de archivos de compresión de audio.

En el capítulo IV se explicará el algoritmo de codificación y decodificación de la norma ISO/IEC 11172-3 conocida comúnmente como MP3, dentro de este se explican los diferentes procesos matemáticos a los que es sometida una señal digital de audio para su compresión así como el proceso inverso para su recuperación y reproducción.

El análisis matemático se especificará en el capítulo V y las conclusiones y resultados que demuestre nuestro estudio serán expuestos en el capítulo VI.

Capítulo II

CONCEPTOS DE SONIDO

Introducción

Antes de entrar de lleno a lo que es el audio digital y de cómo se realiza la compresión de audio en el formato MPEG-1 Layer 3, mejor conocido como MP3; debemos de saber en que consiste el sonido, así como sus principales características y los fenómenos que lo afectan, también debemos conocer a grandes rasgos como percibimos nosotros el sonido y las unidades en la que se mide este; ya que los esquemas de compresión de audio se basan fundamentalmente en modelos psicoacústicos (modelos realizados en base a la forma en la que los humanos percibimos el sonido). Por lo tanto en el siguiente capítulo se da una explicación de las principales características del sonido, así como del aparato auditivo humano y de cómo éste percibe los diferentes sonidos.

2.1 Origen y formación del sonido

2.1.1 Movimiento ondulatorio y ondas.

El **movimiento ondulatorio** se define como el proceso por el que se propaga energía de un lugar a otro sin transferencia de materia, mediante ondas mecánicas, como es el caso del sonido, o electromagnéticas. En cualquier punto de la trayectoria de propagación se produce un desplazamiento periódico, u oscilación, alrededor de una posición de equilibrio. Puede ser una oscilación de moléculas de aire, de moléculas de agua o de porciones de una cuerda o un resorte. En todos estos casos, las partículas oscilan en torno a su posición de equilibrio y sólo la energía avanza de forma continua.

Una **onda** es una perturbación de alguna propiedad de un medio la cual se propaga a través del espacio transportando energía. El medio perturbado puede ser de naturaleza diversa, como el aire, agua, un trozo de metal, en incluso el vacío; y las propiedades que sufren la perturbación pueden ser también variadas, por ejemplo, densidad, presión, campo eléctrico, campo magnético.

Las ondas se clasifican atendiendo a diferentes aspectos:

1. En función de su **naturaleza**.
    a) **Ondas mecánicas**: las ondas mecánicas necesitan un medio elástico (sólido, líquido o gaseoso) para propagarse. Las partículas del medio oscilan alrededor de un punto fijo sin desplazarse, por lo que no existe transporte neto de materia a través del medio. Como en el caso de una alfombra o un látigo cuyo extremo se sacude, la alfombra no se desplaza, sin embargo una onda se propaga a su través.
    b) **Electromagnéticas**: las ondas electromagnéticas se propagan por el espacio sin necesidad de un medio pudiendo, por tanto, propagarse en el vacío. Esto es debido a que las ondas electromagnéticas son producidas por las oscilaciones de un campo eléctrico en relación con un campo magnético asociado.
2. En función de su **propagación**.
    a) **Ondas monodimensionales**: las ondas monodimensionales son aquellas que se propagan a lo largo de una sola dirección del espacio, como las ondas en los muelles o en las cuerdas. Si la onda se propaga en una dirección única, sus frentes de onda son planos y paralelos.
    b) **Ondas bidimensionales o superficiales**: son ondas que se propagan en dos direcciones. Pueden propagarse, en cualquiera de las direcciones de

una superficie, por ello, se denominan también ondas superficiales. Un ejemplo serían las ondas que se producen en la superficie de un lago cuando se deja caer una piedra sobre él.

c) **Ondas tridimensionales o esféricas**: son ondas que se propagan en tres direcciones. Las ondas tridimensionales se conocen también como ondas esféricas, porque sus frentes de ondas son esferas concéntricas que salen de la fuente de perturbación expandiéndose en todas direcciones. El sonido es una onda tridimensional. Son ondas tridimensionales las ondas sonoras (mecánicas) y las ondas electromagnéticas.

3. En función de la **dirección de la perturbación**
   a) **Ondas longitudinales**: el movimiento de las partículas que transportan la onda es paralelo a la dirección de propagación de la onda. Por ejemplo, un muelle que se comprime da lugar a una onda longitudinal.
   b) **Ondas transversales**: las partículas se mueven perpendicularmente a la dirección de propagación de la onda.
4. En función de su **periodicidad**
   a) **Ondas periódicas**: la perturbación local que las origina se produce en ciclos repetitivos por ejemplo una onda senoidal.
   b) **Ondas no periódicas**: la perturbación que las origina se da aisladamente o, en el caso de que se repita, las perturbaciones sucesivas tienen características diferentes. Las ondas aisladas se denominan también **pulsos**.

Un movimiento de onda continuo, u onda periódica, requiere de una perturbación constante de una fuente de oscilación, tal movimiento de onda tendrá forma senoidal tanto en el tiempo como el espacio. Las ondas senoidales se describen mediante características específicas.

Al igual que para una partícula en movimiento armónico simple, la **Amplitud** (**A**) de una onda es la magnitud del desplazamiento máximo o la distancia máxima de la posición de equilibrio de la partícula, véase la figura 2.1. Esto corresponde a la altura de una cresta o a la profundidad de un valle. La distancia entre dos crestas sucesivas (o valles) se llama **Longitud de Onda** (**λ**). En realidad, es la distancia entre dos partículas sucesivas que están en fase (en puntos idénticos de la forma de onda). Por conveniencia, se utilizan por lo general las posiciones de las cresta y los valles. Se puede observar que el espacio que ocupa una longitud de onda corresponde a un ciclo.

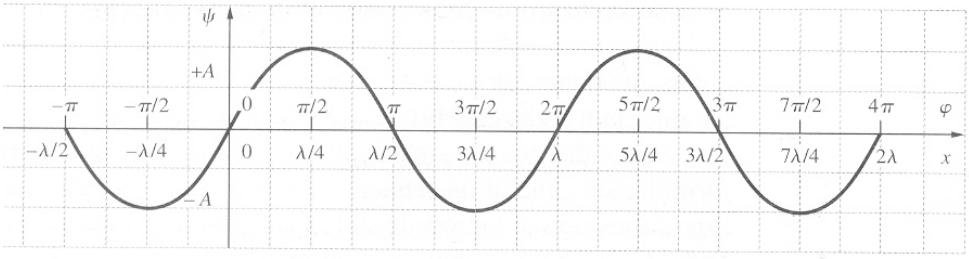


Figura 2.1

La **Frecuencia** (**f**) de una onda es el número de ciclos completos por unidad de tiempo, en este caso por segundo, esto es, el número de formas de onda completas, o longitudes de onda, que pasan por un punto dado durante un segundo. La frecuencia se mide en

ciclos por segundo, o Hertz (Hz). El **Periodo** (T) es el tiempo que toma a una forma de onda completa (longitud de onda) pasar por un punto dado, este se mide en segundos. La frecuencia de la onda está relacionada con el periodo de la onda por la siguiente ecuación:

$$f = \frac{1}{T} \qquad (2.1.1)$$

Donde f es la frecuencia y T es el período. Si el período de una onda es de 10 segundos (por ejemplo, le toma 10 segundos a la onda completar un ciclo), entonces la frecuencia es de 0.1 Hz. En otras palabras, la onda completa 0.1 ciclos cada segundo, o 0.1 Hz.

Una onda es una alteración o disturbio que viaja o se mueve. La velocidad de la onda es una descripción de cuán rápido viaja una onda. La velocidad de la onda está relacionada con la frecuencia, el período y la longitud de onda a través de las simples ecuaciones:

$$v = \frac{\lambda}{T} \quad (2.1.2) \quad \text{o bien} \quad v = \lambda \cdot f \quad (2.1.3)$$

La velocidad de la onda se mide en unidades de longitud por unidad de tiempo, en este caso metros por segundo (m/s). Tomemos, por ejemplo, a la nota musical “LA” que es un sonido con una frecuencia de 440 Hz, su longitud de onda es de 78.4 cm. Para determinar la velocidad de una onda, podemos usar la ecuación 2.1.3 y sustituir los valores dados por longitud de onda y frecuencia, asegurándonos que estamos usando unidades de un mismo sistema de medición, en este caso el Sistema Internacional.

$$f = 440Hz$$
$$\lambda = 78.4cm = 0.784m$$
$$v = \lambda f = (440Hz) \cdot (0.784m) = 345m/s$$

El valor (345 m/s) es el valor aproximado de la velocidad del sonido en el aire.

2.1.2 Ondas sonoras

Las **ondas sonoras** son ondas mecánicas (ondas de compresión), pues precisan de un medio (sólido, líquido o gaseoso), que trasmita la perturbación. Es, el propio medio, el que produce y propicia la propagación de estas ondas, con su compresión y expansión. Para que este medio, pueda comprimirse y expandirse es un requisito fundamental que se trate de un medio elástico. Sin medio elástico, no habría sonido, pues las ondas sonoras no se propagan en el vacío. El aire es el medio transmisor más común del sonido

Un cuerpo en oscilación pone en movimiento a las moléculas del medio que lo rodean. Estas, a su vez, transmiten ese movimiento a las moléculas vecinas y así sucesivamente. Cada molécula del medio entra en oscilación en torno a su punto de reposo. Es decir, el desplazamiento que sufre cada molécula es pequeño. Pero el movimiento se propaga a través del medio.

Entre la fuente sonora y el receptor tenemos entonces una transmisión de energía pero no un traslado de materia. El (pequeño) desplazamiento (oscilatorio) que sufren las distintas

moléculas de aire genera zonas en las que hay una mayor concentración de moléculas (mayor densidad), zonas de condensación, y zonas en las que hay una menor concentración de moléculas (menor densidad), zonas de rarefacción, como se ilustra en la figura 2.2. Esas zonas de mayor o menor densidad generan una variación alterna en la presión estática del aire (la presión del aire en ausencia de sonido). Es lo que se conoce como presión sonora.

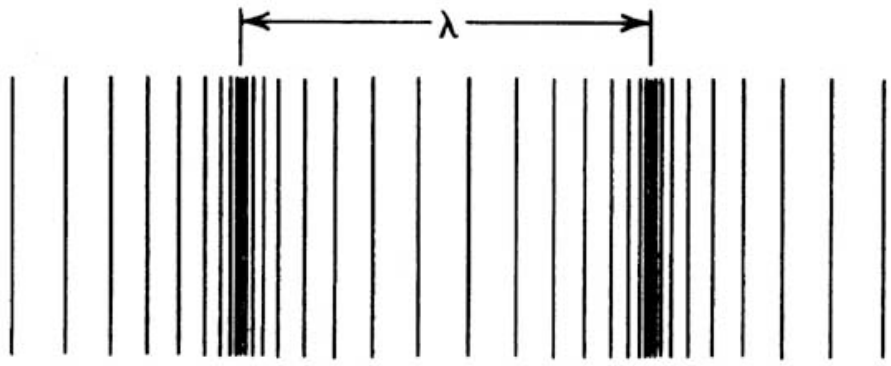


Figura 2.2

Por otro lado, la presión de las partículas que transportan la onda se produce en la misma dirección de propagación de la onda. Por tanto, las ondas sonoras son ondas longitudinales. Además, las ondas sonoras se desplazan en tres dimensiones y sus frentes de onda son esferas concéntricas que salen desde el foco de la perturbación en todas las direcciones. Por esto, son ondas esféricas o tridimensionales.

El oído humano puede percibir ondas sonoras de frecuencias entre los 20 y los 20,000 Hz. Este intervalo de frecuencias se denomina región audible del espectro de frecuencias del sonido. Las frecuencias menores a los 20 Hz están en la región infrasónica. Las ondas longitudinales generadas por los sismos tienen frecuencias infrasónicas. Arriba de los 20 kHz está la región ultrasónica. Las ondas ultrasónicas pueden generarse por vibraciones de alta frecuencia en cristales.

La velocidad de propagación de la onda sonora (velocidad del sonido) depende de las características del medio en el que se realiza dicha propagación y no de las características de la onda o de la fuerza que la genera. En el caso de un gas (como el aire) es directamente proporcional a su temperatura específica y a su presión estática e inversamente proporcional a su densidad. Dado que si varía la presión, varía también la densidad del gas, la velocidad de propagación permanece constante ante los cambios de presión o densidad del medio.

Pero la velocidad del sonido sí varía ante los cambios de temperatura del aire (medio). Cuanto mayor es la temperatura del aire mayor es la velocidad de propagación. La velocidad del sonido en el aire aumenta 0,6 m/s por cada grado centígrado de aumento en la temperatura.

La velocidad del sonido en el aire es de aproximadamente 344 m/s a 20º C de temperatura, lo que equivale a unos 1,200 km/h (1,238.4 km/h, para ser precisos). Es

decir que necesita unos 3 s para recorrer 1 km. (Como posible referencia recordemos que la velocidad de la luz es de 300,000 km/s.)

El sonido se propaga a diferentes velocidades en medios de distinta densidad. En general, se propaga a mayor velocidad en líquidos y sólidos que en gases (como el aire). La velocidad de propagación del sonido es, por ejemplo, de unos 1,440 m/s en el agua y de unos 5,000 m/s en el acero.

2.1.3 Características del sonido

Las ondas sonoras que percibe el oído humano se distinguen por tres características: **intensidad**, **tono** y **timbre**.

La **intensidad** del sonido percibido, o propiedad que hace que este se capte como fuerte o como débil, está relacionada con la intensidad de la onda sonora correspondiente, también llamada intensidad acústica. La intensidad acústica es una magnitud que da idea de la cantidad de energía que está fluyendo por el medio como consecuencia de la propagación de la onda.

Se define como la energía que atraviesa por segundo una superficie dispuesta perpendicularmente a la dirección de propagación. Equivale a una potencia por unidad de superficie y se expresa en W/m^2. Por ejemplo, consideremos un punto de origen que envía fuera ondas de sonido esféricas; si no hay pérdidas la intensidad del sonido a una distancia R del origen es

$$I = \frac{P}{A} = \frac{P}{4\pi R^2} \quad (2.1.4)$$

En donde P es la potencia de la fuente y $4\pi R^2$ es el área de una esfera con un radio R, a través de la cual la energía sonora cruza perpendicularmente. La intensidad para un punto de origen es, por consiguiente, inversamente proporcional al cuadrado de la distancia al origen. Debido a que la intensidad decrece en un factor de $1/R^2$, al duplicar la distancia la intensidad decrece a la cuarta parte de su valor original.

En promedio, el oído humano puede detectar ondas sonoras con una intensidad tan baja como de 10^{-12} W/m^2. A esta intensidad se le conoce como **umbral de audición**. A una intensidad de 1.0 W/m^2, el sonido es demasiado fuerte, y puede ser incluso doloroso para el oído. A esta intensidad se le conoce como **umbral del dolor**. El intervalo de intensidades acústicas que va desde el umbral de audición hasta el umbral del dolor es muy amplio, estando ambos valores límite en una relación del orden de 10^{12}.

Debido a la extensión de este intervalo de audibilidad, para expresar intensidades sonoras se emplea una escala cuyas divisiones son potencias de diez y cuya unidad de medida es el decibel (dB). Ello significa que una intensidad acústica de 10 decibeles corresponde a una energía diez veces mayor que una intensidad de cero decibeles; una intensidad de 20 dB representa una energía 100 veces mayor que la que corresponde a 0 decibeles y así sucesivamente. Para obtener el nivel de intensidad sonora en decibeles utilizamos la siguiente ecuación.

$$I_{dB} = 10\log\left(\frac{I}{I_0}\right) \quad (2.1.5)$$

Donde $I_0 = 10^{-12}$ W/m^2 que es la mínima intensidad sonora que el oído humano puede escuchar, todas las demás intensidades acústicas son referidas a este valor.

En la tabla 2.1.1 se muestran algunos niveles de sonido representativos y sus intensidades correspondientes.

SONIDO	INTENSIDAD (W/m^2)	INTENSIDAD RELATIVA (I/I_0)	NIVEL DE SONIDO (dB)
Umbral de audición	1×10^{-12}	10^0	**0**
El murmullo de las hojas	1×10^{-11}	10^1	**10**
Un murmullo (a 1m)	1×10^{-10}	10^2	**20**
Calle de la ciudad, sin transito	1×10^{-9}	10^3	**30**
Oficina, aula	1×10^{-7}	10^5	**50**
Conversación normal (a 1 m)	1×10^{-6}	10^6	**60**
Martillo perforador (a 1 m)	1×10^{-3}	10^9	**90**
Grupo de rock	1×10^{-1}	10^{11}	**110**
Umbral del dolor	1	10^{12}	**120**
Motor de propulsión a chorro (a 50 m)	**10**	**10^{13}**	**130**

Tabla 2.1 Algunas intensidades y niveles de sonido

El **tono** o altura de un sonido depende únicamente de su frecuencia, es decir, del número de oscilaciones por segundo. La altura de un sonido corresponde a nuestra percepción del mismo como más grave o más agudo. Los sonidos percibidos como graves corresponden a frecuencias bajas, mientras que los agudos son debidos a frecuencias altas. Así, por ejemplo, el sonido más grave de una guitarra corresponde a una frecuencia de 82,4 Hz y el más agudo a 698,5 Hz. Junto con la frecuencia, en la percepción sonora del tono intervienen otros factores de carácter psicológico. Así sucede por lo general que al elevar la intensidad se eleva el tono percibido para frecuencias altas y se baja para las frecuencias bajas. Entre frecuencias comprendidas entre 1 000 y 3 000 Hz el tono es relativamente independiente de la intensidad.

El **timbre** o calidad es la cualidad del sonido que permite distinguir sonidos procedentes de diferentes instrumentos, aun cuando posean igual tono e intensidad. La calidad del tono depende de la forma de onda, específicamente del número de armónicas presentes; en la figura 2.3 podemos observar una forma de onda derivada de la suma de una frecuencia fundamental y sus armónicos. Debido a esta misma cualidad es posible reconocer a una persona por su voz, que resulta característica de cada individuo.

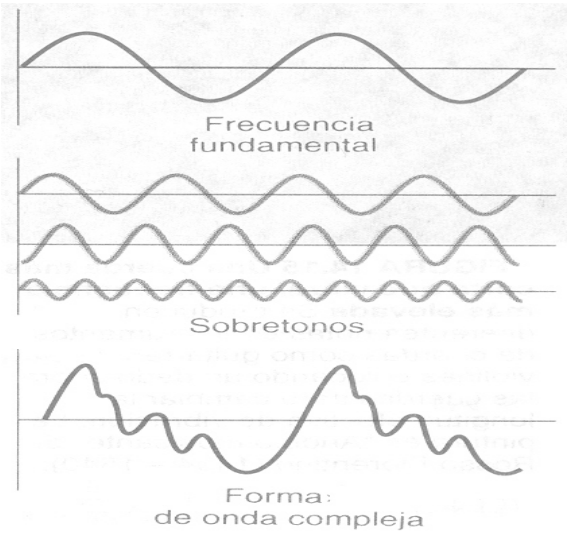


Figura 2.3

Pocas veces las ondas sonoras corresponden a sonidos puros, sólo los diapasones generan este tipo de sonidos, que son debidos a una sola frecuencia y representados por una onda armónica. Los instrumentos musicales, por el contrario, dan lugar a un sonido más rico que resulta de vibraciones complejas. Cada vibración compleja puede considerarse compuesta por una serie de vibraciones armónico simples de una frecuencia y de una amplitud determinadas, cada una de las cuales, si se considerara separadamente, daría lugar a un sonido puro. Esta mezcla de tonos parciales es característica de cada instrumento y define su timbre.

2.1.4 Fenómenos del sonido

Una onda sonora se refleja (rebota al medio del cual proviene) cuando topa con un obstáculo que no puede traspasar ni rodear. El tamaño del obstáculo y la longitud de onda determinan si una onda rodea el obstáculo o rebota hacia la dirección de la que provenía. Si el obstáculo es pequeño en relación con la longitud de onda, el sonido lo rodeara (difracción), en cambio, si sucede lo contrario, el sonido rebota (**reflexión**).

Si la onda rebota, el ángulo de la onda reflejada es igual al ángulo de la onda incidente, de modo que si una onda sonora incide perpendicularmente sobre la superficie reflejante, vuelve sobre sí misma. La reflexión no actúa igual sobre las altas frecuencias que sobre las bajas. Lo que se debe a que la longitud de onda de las bajas frecuencias es muy grande (pueden alcanzar los 18 metros), por lo que son capaces de rodear la mayoría de obstáculos.

En acústica esta propiedad de las ondas es sobradamente conocida y aprovechada. No sólo para aislar, sino también para dirigir el sonido hacia el auditorio mediante placas reflectoras (reflectores y tornavoces). El eco es un ejemplo familiar de reflexión del sonido

La **refracción** del sonido, por ejemplo, se puede experimentar en el verano. En esos días, es posible oír voces a distancia u otros sonidos que por lo general no son audibles. Este efecto se debe a la refracción, o desviación, de las ondas sonoras cuando pasan por una

región en la que la densidad del aire es diferente. El efecto es similar al que sucedería si el sonido pasara a otro medio.

La **difracción** es un fenómeno que afecta a la propagación del sonido. Hablamos de difracción cuando el sonido en lugar de seguir en la dirección normal, se dispersa. La explicación la encontramos en el Principio de Huygens que establece que cualquier punto de un frente de ondas es susceptible de convertirse en un nuevo foco emisor de ondas idénticas a la que lo originó. De acuerdo con este principio, cuando la onda incide sobre una abertura o un obstáculo que impide su propagación, todos los puntos de su plano se convierten en fuentes secundarias de ondas, emitiendo nuevas ondas, denominadas ondas difractadas. La difracción se puede producir por dos motivos diferentes:

a) Porque una onda sonora encuentra a su paso un pequeño obstáculo y lo rodea. Las bajas frecuencias son más capaces de rodear los obstáculos que las altas. Esto es posible porque las longitudes de onda en el espectro audible están entre 3 cm y 12 m, por lo que son lo suficientemente grandes para superar la mayor parte de los obstáculos que encuentran.
b) O porque una onda sonora topa con un pequeño agujero y lo atraviesa.

La cantidad de difracción estará en función del tamaño de la propia abertura y de la longitud de onda. Si una abertura es grande en comparación con la longitud de onda, el efecto de la difracción es pequeño. La onda se propaga en líneas rectas o rayos, como la luz. Cuando el tamaño de la abertura es considerable en comparación con la longitud de onda, los efectos de la difracción son grandes y el sonido se comporta como si fuese una luz que procede de una fuente puntual localizada en la abertura.

Al igual que las ondas de cualquier clase, las ondas sonoras se interfieren cuando se encuentran. Suponiendo que dos oradores separados por cierta distancia emiten ondas sonoras en fase y a la misma frecuencia; en regiones determinadas del espacio, habrá interferencias constructivas o destructivas. Si las ondas se encuentran en una región en la que quedan exactamente en fase, es decir, en la que coinciden dos crestas o dos valles, habrá una interferencia constructiva total. En estos puntos habrá un incremento en la intensidad del sonido. Inversamente, si las ondas se juntan de modo que la cresta de una coincide con el valle de la otra, habrá una interferencia destructiva total, por lo que no se escuchará ningún sonido en estos puntos.

El **efecto Doppler** se origina cuando hay un movimiento relativo entre la fuente sonora y el escucha cuando cualquiera de los dos se mueve con respecto al medio en el que las ondas se propagan. El resultado es la aparente variación de la altura del sonido. Existe una variación en la frecuencia que percibimos con la frecuencia que la fuente origina.

Si el movimiento del emisor va de izquierda a derecha (velocidades positivas), la longitud de onda medida por el observador situado a la derecha es más pequeña, y la longitud de onda medida por el observador situado a la izquierda del emisor es mayor.

Si el emisor emite ondas sonoras, el sonido escuchado por el observador situado a la derecha del emisor, será más agudo y el sonido escuchado por el observador situado a la izquierda será más grave. En otras palabras, cuando el emisor se acerca al observador, éste escucha un sonido más agudo, cuando el emisor se aleja del observador, éste escucha un sonido más grave. Por ejemplo, el tono del silbato de una locomotora o de la

sirena de un carro de bomberos es más alto cuando la fuente se aproxima al escucha que cuando ha pasado y se aleja.

2.2 La Audición

2.2.1 El oído Humano

La función de nuestro sistema auditivo es, esencialmente, transformar las variaciones de presión originadas por la propagación de las ondas sonoras en el aire en impulsos eléctricos (variaciones de potencial), información que los nervios acústicos transmiten a nuestro cerebro para la asignación de significados.
El sistema auditivo periférico (el oído) está compuesto por el **oído externo, el oído medio y el oído interno**, esto se muestra en la figura 2.4**.**

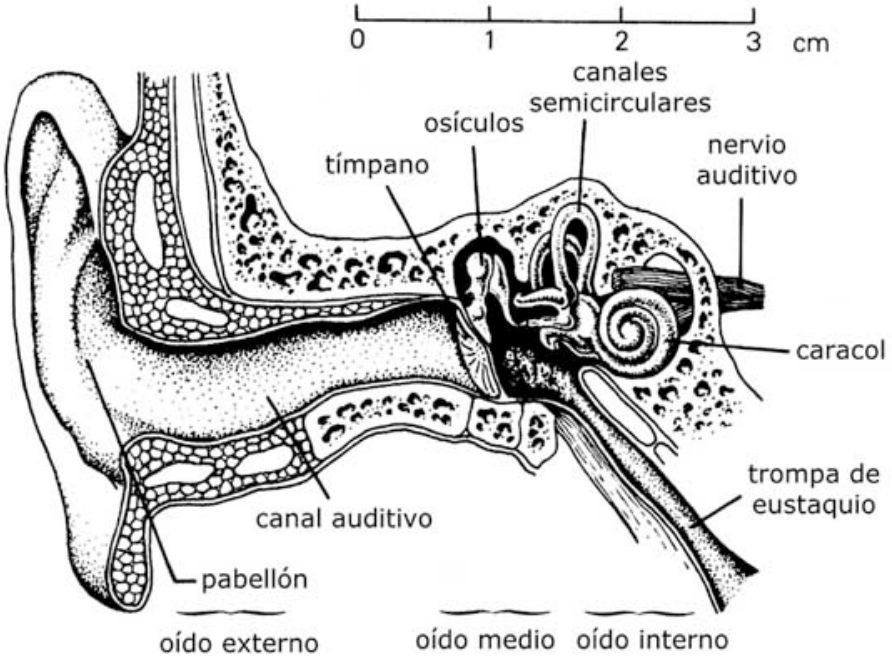


Figura 2.4. El oído humano.

El **oído externo** está compuesto por el pabellón, que concentra las ondas sonoras en el conducto, y el conducto auditivo externo que desemboca en el tímpano. El canal auditivo externo tiene unos 2.7 cm. de longitud y un diámetro promedio de 0.7 cm. Al comportarse como un tubo cerrado en el que oscila una columna de aire, la frecuencia de resonancia del canal es de alrededor de los 3,200 Hz.

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

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

La parte central del tímpano oscila como un cono asimétrico, al menos para frecuencias inferiores a los 2,400 Hz. Para frecuencias superiores a la indicada las vibraciones del tímpano ya no son tan simples, por lo que la transmisión al martillo es menos efectiva.

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

El **oído interno** resuelve este desajuste de impedancias por dos vías complementarias. En primer lugar la disminución de la superficie en la que se concentra el movimiento. Por otra parte el martillo y el yunque funcionan como un mecanismo de palanca y la relación entre ambos brazos de la palanca es de 1.31: 1. En general entre el oído externo y el tímpano se produce una amplificación de entre 5 dB y 10 dB en las frecuencias comprendidas entre los 2,000 Hz y los 5,000 Hz, lo que contribuye de manera fundamental para la gama de frecuencias a la que nuestro sistema auditivo es más sensible.

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

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

Si en el oído externo se canaliza la energía acústica y en el oído medio se la transforma en energía mecánica transmitiéndola (y amplificándola) hasta el oído interno, es en éste en donde se realiza la definitiva transformación en impulsos eléctricos.

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

El oído interno está inmerso en un fluido viscoso llamado endolinfa cuando se encuentra en el laberinto membranoso y perilinfa cuando separa los laberintos óseo y membranoso.
La cóclea (o caracol) es un conducto casi circular enrollado en espiral (de ahí su nombre) unas 2.75 veces sobre sí mismo, de unos 35 mm de largo y unos 1.5 mm de diámetro como promedio. El ducto coclear divide a la cóclea en dos secciones, la rampa vestibular y la rampa timpánica. Esto se puede observar en la figura 2.5.

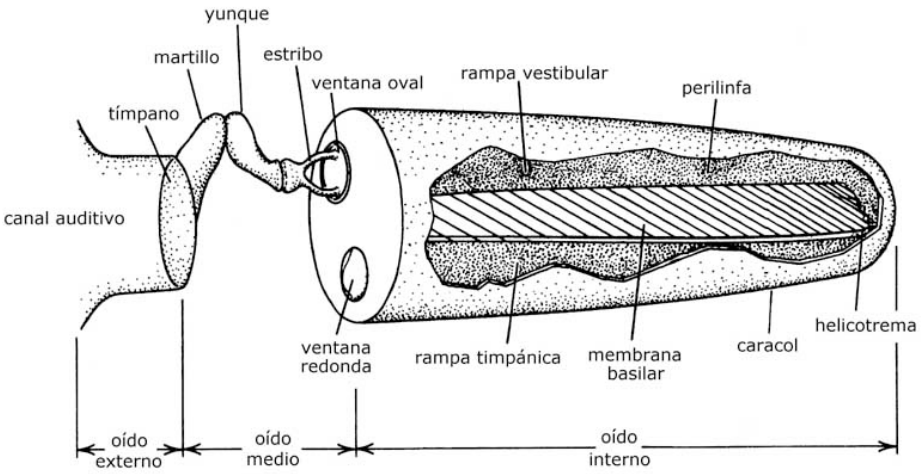


Figura 2.5 Esquema del sistema auditivo periférico con la cóclea desenrollada.

La cóclea está dividida a lo largo por la membrana basilar y la membrana de Reissner, como lo muestra la figura 2.6.

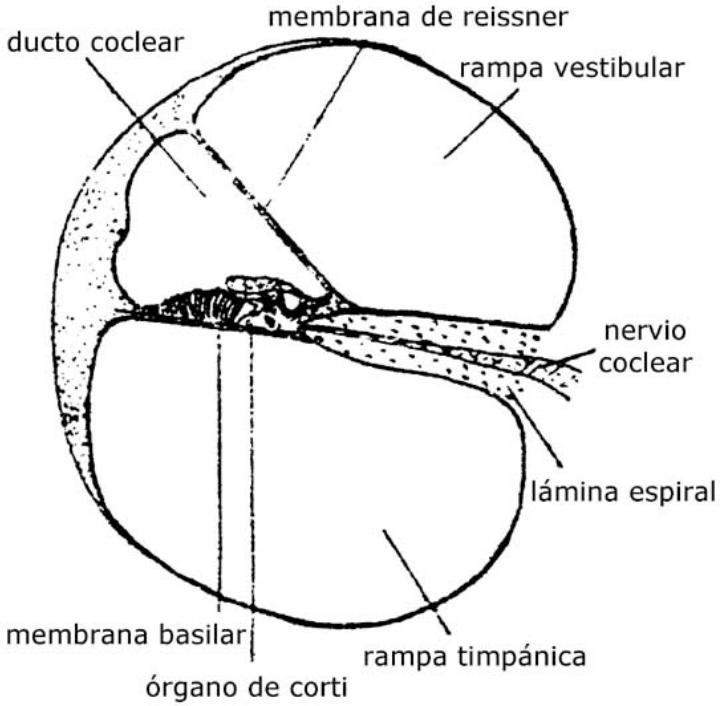


Figura 2.6 Corte de la cóclea.

El movimiento de la membrana basilar afecta las células ciliares (también llamadas capilares o pilosas) del órgano de Corti que al ser estimuladas (deformadas) generan los

impulsos eléctricos que las fibras nerviosas (nervios acústicos) transmiten al cerebro. Puede haber hasta cinco filas de células ciliares en el órgano de Corti, constando las más largas de unas 12,000 células en fila. En la figura 2.7 podemos observar el órgano de Corti.

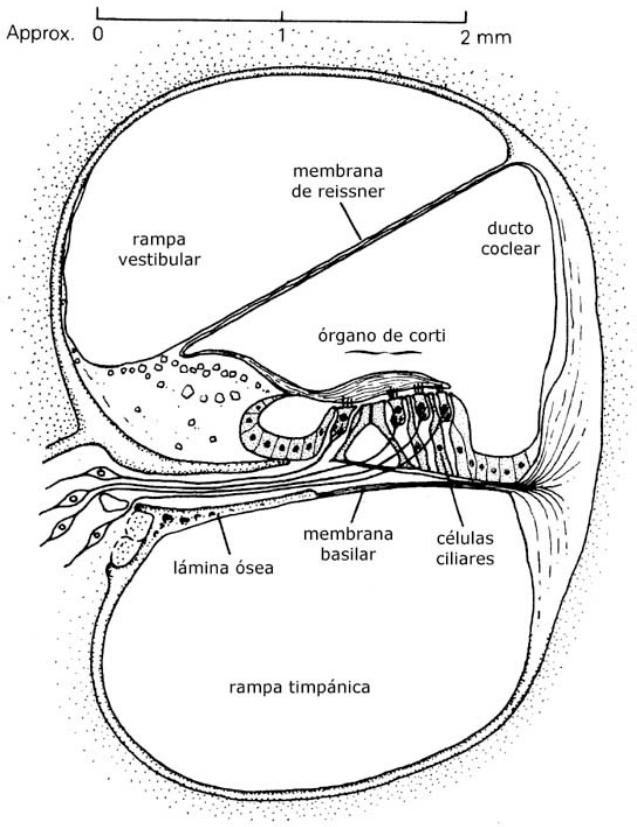


Figura 2.7 El órgano de Corti.

La membrana basilar no llega hasta el final de la cóclea dejando un espacio para la intercomunicación del fluido entre la rampa vestibular y la timpánica, llamado helicotrema que tiene aproximadamente, una superficie de unos 0.25 mm^2 como se muestra en la figura 2.8.

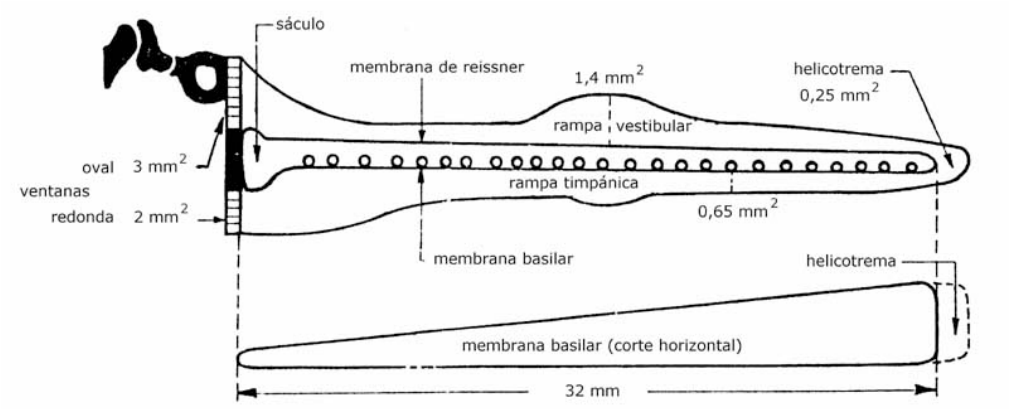


Figura 2.8 La membrana basilar.

La membrana basilar se deforma como producto del movimiento del fluido linfático dentro de la cóclea. El punto de mayor amplitud de oscilación de la membrana basilar varía en función de la frecuencia del sonido que genera su movimiento, produciendo así la información necesaria para nuestra percepción de la altura del sonido. Las frecuencias más altas son procesadas en el sector de la membrana basilar más cercana al oído medio y las más bajas en su sector más lejano (cerca del helicotrema), como se muestra en la figura 2.9. La cantidad de células ciliares estimuladas (deformadas) y la magnitud de dicha deformación determinarían la información acerca de la intensidad de ese sonido.

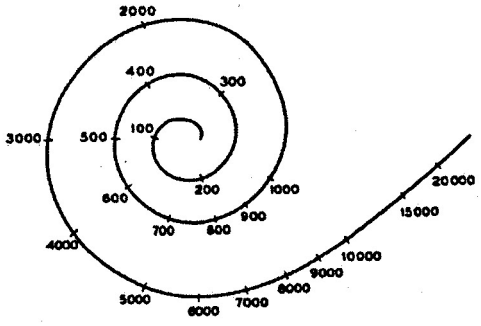


Figura 2.9 Ubicación de la zona de respuesta de frecuencias sobre la membrana basilar.

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

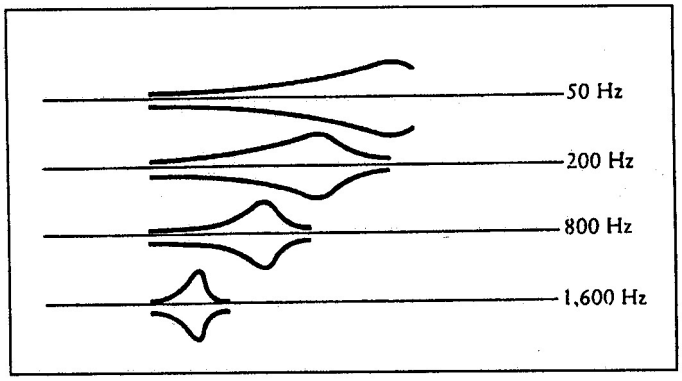


Figura 2.10 Esquema vibratorio de la membrana basilar.
El punto de mayor oscilación depende de la frecuencia

Además de a través del oído medio (el tímpano, los osículos), las ondas sonoras llegan al oído interno directamente por medio de la oscilación de los huesos del cráneo. Dado que el oído interno se encuentra inserto en una cavidad del hueso temporal, las oscilaciones del cráneo hacen entrar en oscilación directamente al fluido linfático, de una manera que no está totalmente clara aún. Lo que sí resulta evidente es que cualquiera de las dos formas de transmisión de las ondas es igualmente efectiva, sirviendo la transmisión ósea como medio alternativo cuando hay enfermedades en el oído medio. La transmisión ósea es también la responsable de que escuchemos nuestra voz con un timbre distinto al que lo escucha el resto de las personas.

2.3 Fundamentos de Psicoacústica

2.3.1 Nivel de Presión Sonora

El nivel de presión sonora (SPL por sus siglas en ingles) es una medición de la intensidad de un estímulo acústico. El SPL nos da el nivel de la presión sonora en decibeles en relación a un nivel de referencia, por ejemplo:

$$L_{SPL} = 20\log_{10}\left(\frac{p}{p_0}\right)dB$$

Donde L_{SPL} es el nivel de presión sonora de un estímulo, p es la presión sonora del estímulo en Pascales y p_0 es la referencia estándar de 20 µP. Este nivel de referencia se toma por ser la mínima presión sonora que percibe nuestro oído, además corresponde a la intensidad de sonido I_0 que se menciono anteriormente.

2.3.2 Sonoridad

El concepto de sonoridad se utiliza para describir la forma en la que percibimos los sonidos, es decir, que tan “fuerte” escuchamos un sonido, ya que esto dependerá de la frecuencia y la intensidad de este.

Con objeto de dar una medida lo más fiel posible de la "sonoridad" de un sonido, Fletcher y Munson establecieron una medida para la sonoridad y una serie de curvas de **igual sonoridad** o curvas isofónicas para varios niveles de presión acústica, desde el umbral de audición (0 dB), hasta niveles dañinos para la salud (120 dB), en intervalos de 10 dB. La unidad de sonoridad es el Fon el cual está definido arbitrariamente como la sonoridad de un sonido senoidal de 1 kHz con un nivel de presión sonora (intensidad) de 0 dB_{SPL}. Así, 0 dB es igual a 0 fon y 120 dB es igual a 120 fon. Estas curvas se pueden observar en la siguiente figura (figura 2.11)

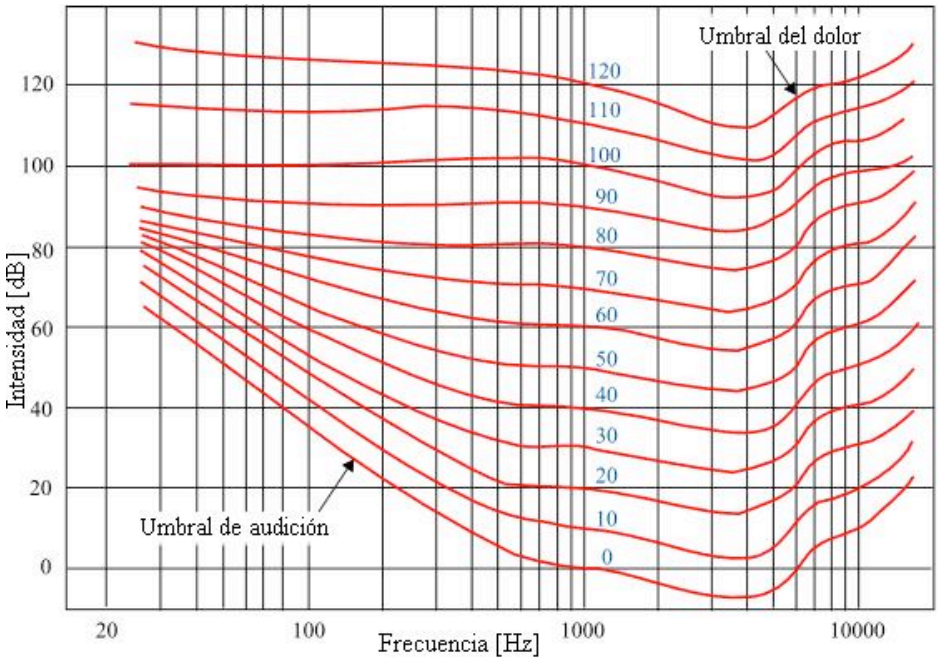


Figura 2.11 Curvas de sonoridad.

Así un sonido de 100 Hz a 52 dB y otro de 10 kHz a 52 dB tienen la misma sonoridad que una nota de 1 kHz a 40 dB, lo que significa que los tres tienen una sonoridad de 40 Fonios, esto quiere decir que estos tres sonidos se perciben igual de fuertes aun cuando no tengan la misma intensidad sonora.

2.3.3 Intervalo auditivo

El ser humano es capaz de detectar únicamente aquellos sonidos que se encuentren dentro de un determinado intervalo de amplitudes y frecuencias. Se puede definir al rango dinámico[1] del oído como la relación entre la máxima potencia sonora que este puede manejar y la mínima potencia necesaria para detectar un sonido. Asimismo, el rango de frecuencias asignado convencionalmente al sistema auditivo va desde los 20 Hz hasta los 20 kHz, aun cuando este rango puede variar de un sujeto a otro o disminuir en función de la edad del sujeto, de trastornos auditivos o de una pérdida de sensibilidad (temporal o permanente) debida a la exposición a sonidos de elevada intensidad. En la figura 2.12 podemos observar las diferentes curvas de SPL en función de la frecuencia.

[1] No es propiamente un rango sino un intervalo pero en el medio del audio se referiere a este intervalo como Rango Dinámico.

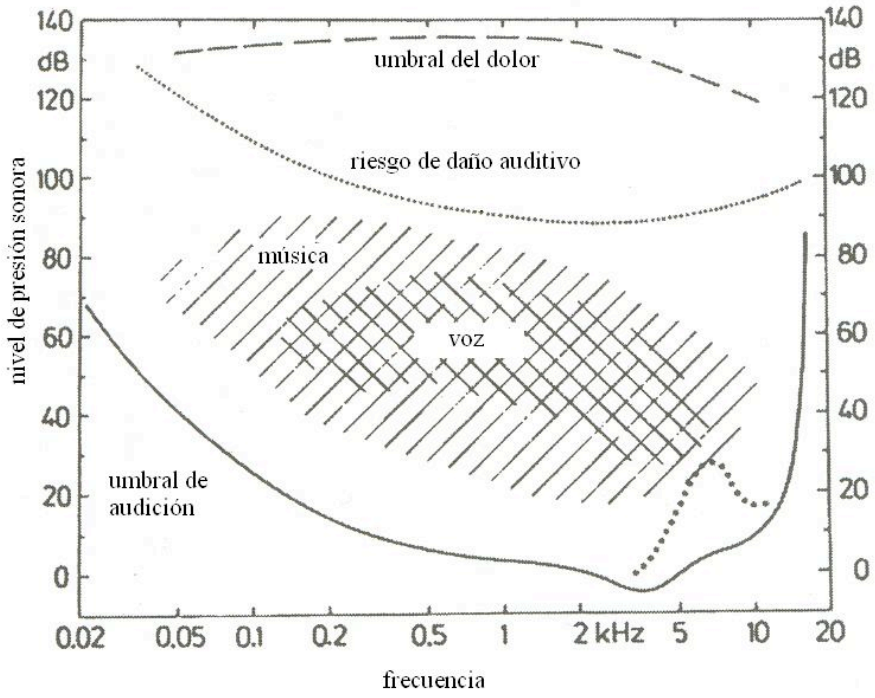


Figura 2.12 Intervalo auditivo

En la parte mas baja de la gráfica podemos observar la curva que representa el umbral de audición, el cual es el nivel de audibilidad para tonos puros en condiciones de tranquilidad; en el extremo superior se puede observar el umbral de dolor, el cual define las máximas presiones sonoras que puede soportar el oído humano, el área entre el umbral del dolor y el umbral auditivo representa el rango auditivo del humano.

La voz humana se encuentra en el intervalo de frecuencias que va desde los 100 Hz a los 8 kHz, aunque en términos prácticos se considera solamente entre 400 Hz y 4 kHz, mientras el rango en niveles de presión sonora va desde 30 dB hasta los 70 dB, donde una conversación típicamente toma valores entre los 50 y los 60 dB. Mientras tanto la música tiene un intervalo mas grande en frecuencia así como para niveles de presión sonora tiene un rango que va aproximadamente desde los 20 Hz (aunque algunos instrumentos alcanzan frecuencias menores) hasta los 15 kHz, además su rango dinámico varia típicamente entre los 20 dB y los 95 dB. También podemos observar que cerca de los 100 dB se encuentra el riesgo de daño para el oído humano.

2.3.4 Umbral de Audición.

El umbral de audición representa el nivel mas bajo de un sonido que puede ser escuchado a cierta frecuencia, incluso en condiciones de extrema quietud y silencio, el ser humano no puede percibir los SPL debajo de este umbral.

Esta curva es muy importante para la codificación de audio ya que todas las componentes de frecuencia que estén por debajo de esta curva son irrelevantes para la percepción del sonido por lo cual estas se pueden eliminar sin afectar la percepción, e incluso el ruido

que se pueda obtener de la cuantización que se encuentre debajo de esta curva no será percibido por el oído humano.
El umbral de audición se puede medir tomando los valores de presión sonora a los cuales la persona que escucha perciba el sonido a diferentes frecuencias, estos valores varían dependiendo de cómo se tomen los datos, ya sea aumentando las frecuencias o disminuyéndolas, lo cual provoca una gráfica en forma de zigzag la cual tiene un rango de 6 dB entre el punto en donde el sonido definitivamente es audible y el punto donde es completamente inaudible; el promedio entre las dos curvas, es decir los puntos más altos y los puntos más bajos de la gráfica, es lo que se usa para determinar el umbral de audición, de acuerdo con esto el umbral de audición de una persona tiene un rango de ± 3 dB, esto lo podemos observar en la figura 2.13.

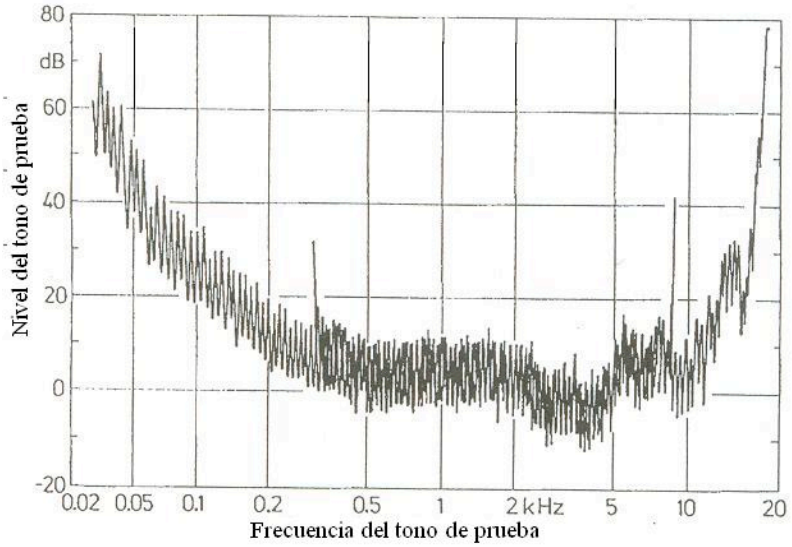


Figura 2.13 Resultados de un experimento para obtener el umbral de audición

La dependencia de la frecuencia del umbral de audición esta muy bien establecido, el umbral es relativamente alto a bajas frecuencias, y se puede observar que de 40 dB a 50Hz cae hasta 0 dB a 500 Hz, de 500 Hz hasta los 2 kHz se mantiene constante, muy cercano a los 0 dB; de los 2 kHz a los 5 kHz puede obtenerse valores debajo de los 0 dB para personas con una buena audición, arriba de los 5 kHz el umbral empieza a crecer con pequeños valles y crestas dependiendo la persona, alrededor de los 16 kHz el umbral empieza a crecer rápidamente. El umbral de audición también varía dependiendo la edad, aunque para frecuencias menores a 2 kHz prácticamente es independiente de la edad, para frecuencias mayores a 2 kHz se puede observar que sucede lo contrario ya que por ejemplo una persona de 60 años tendrá un umbral 30 dB mayor a 10 kHz que una persona de 20 años, esto lo podemos ver en la figura 2.14.

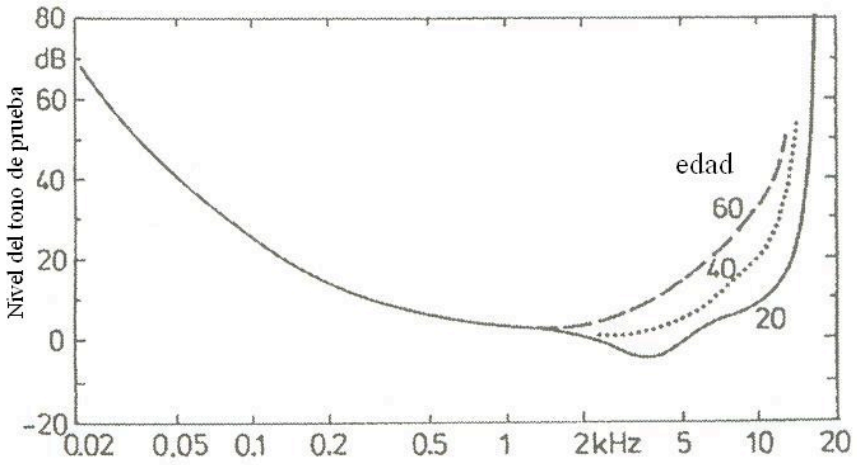


Figura 2.14 umbral de audición para personas de diferentes edades.

Se puede obtener una buena aproximación del umbral por medio de la siguiente función, la cual depende de la frecuencia:

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

Donde esta ecuación se obtiene tomando en cuenta los efectos que se dan dentro del oído externo y medio así como los efectos de cancelación de ruido que se da en el oído interno. Una gráfica de la función la podemos ver en la figura 2.15 donde podemos observar su cercanía a los valores tomados experimentalmente. Además tenemos que considerar que para comparar una señal con el umbral de audición es necesario conocer el nivel de la señal de audio, lo cual no se puede conocer a priori, por lo que se considera el nivel de la señal de audio lo mas pequeño posible dentro del sistema de codificación, esto es cercano a los 0 dB, esto equivale a hacer corresponder la parte más plana del umbral de audición, que es entre los 500 Hz y 2 kHz, con el bit menos significativo de la amplitud espectral de la señal dentro del sistema de codificación.

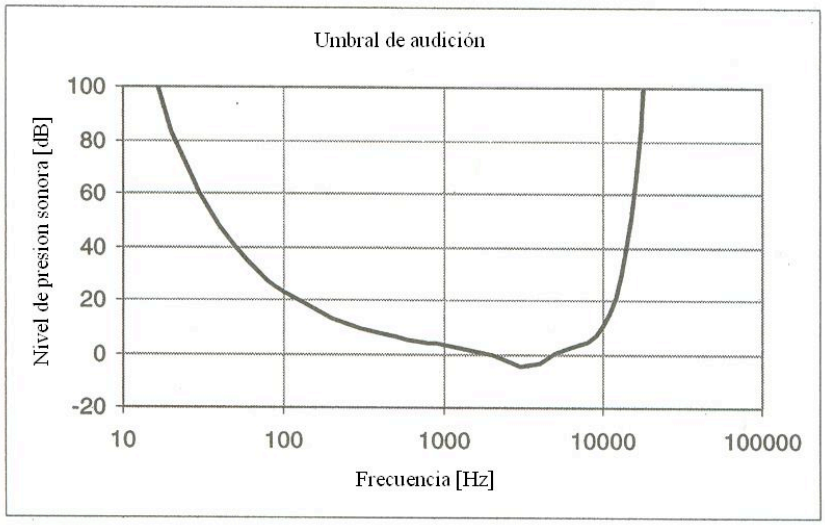


Figura 2.15 Aproximación de la ecuación para el umbral de audición.

2.3.5 Efecto de Enmascaramiento.

La mayor parte del tiempo el mundo se presenta con gran variedad de sonidos simultáneos; el ser humano automáticamente lleva a cabo la tarea de distinguir cada uno de ellos y atender a los de mayor importancia. A menos de que realmente se preste atención a algún sonido que se desee escuchar, pero que sea muy difícil; el ser humano no se percata de todos los sonidos que no escucha a lo largo del día, pero que si existen.
Es muy difícil percibir un sonido cuando existe otro de mayor intensidad presente al mismo tiempo. Este proceso, al parecer intuitivo, a niveles psicoacústicos y cognoscitivos es muy complejo. A este fenómeno se le denomina como enmascaramiento (*masking*).

El enmascaramiento esta definido por la Asociación Americana de Estándares (ASA) como el proceso por el cual el umbral de audición para un sonido (enmascarado) es elevado en la presencia de otro sonido (enmascarante). Por ejemplo si nos encontramos en la calle teniendo una conversación con otra persona, al aparecer el ruido de un camión muy fuerte automáticamente paramos la conversación puesto que esta se hace inaudible, es decir, la conversación fue enmascarada por el ruido del camión.

2.3.6 Enmascaramiento frecuencial

El enmascaramiento frecuencial o simultáneo es un fenómeno que tiene lugar en el dominio de la frecuencia donde las señales de bajo nivel pueden volverse inaudibles para el oído humano si simultáneamente una señal más fuerte está lo suficientemente cerca de las señales de bajo nivel. En la figura 2.16 podemos observar el tono enmascarante y dos tonos de bajo nivel muy cercanos al tono enmascarante, además se muestra el umbral de audición y adicionalmente se observa una nueva curva enmascarante, la cual se llama umbral de enmascaramiento la cual representa el nuevo umbral de audibilidad para señales en presencia de una señal enmascarante.

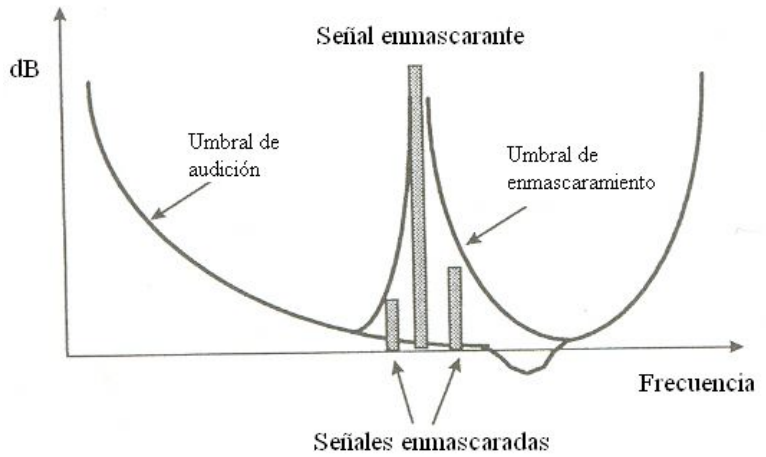


Figura 2.16 Enmascaramiento frecuencial

Cualquier señal o componente frecuencial debajo del umbral de enmascaramiento no podrá ser escuchada, esto lo podemos observar en la figura, ya que aunque las señales que se encuentran debajo de esta curva están por encima del umbral de audición, estas no podrán ser escuchadas debido al tono o señal enmascarante, pues este ha cambiado

los umbrales para estas frecuencias. Por lo tanto todos los puntos que se encuentran entre las dos curvas (umbral de audición y umbral de enmascaramiento) corresponden a sonidos enmascarados por el tono de mayor magnitud.

Dentro del campo de la codificación de audio esto nos permite identificar las componentes de la señal que no necesitan ser transmitidas y también para determinar la cantidad de ruido debido a la cuantización que es permitido para las componentes de señal que son transmitidas.

2.3.7 Enmascaramiento Temporal.

Dependiendo de la ubicación temporal de la señal enmascarada con respecto a la señal enmascarante se pueden distinguir tres situaciones posibles:

a) Enmascaramiento simultaneo. La señal enmascarada y la señal enmascarante se presentan durante toda la duración de la señal enmascarada (enmascaramiento frecuencial).
b) Pre-enmascaramiento. Cuando la señal enmascarante se presenta después de la señal enmascarada.
c) Post-enmascaramiento. La señal enmascarante se presenta antes que la señal enmascarada.

En la figura 2.17 se pueden observar las regiones temporales en las cuales ocurren los fenómenos de pre-enmascaramiento, post-enmascaramiento y enmascaramiento simultáneo, así como la evolución en el tiempo de los mismos.

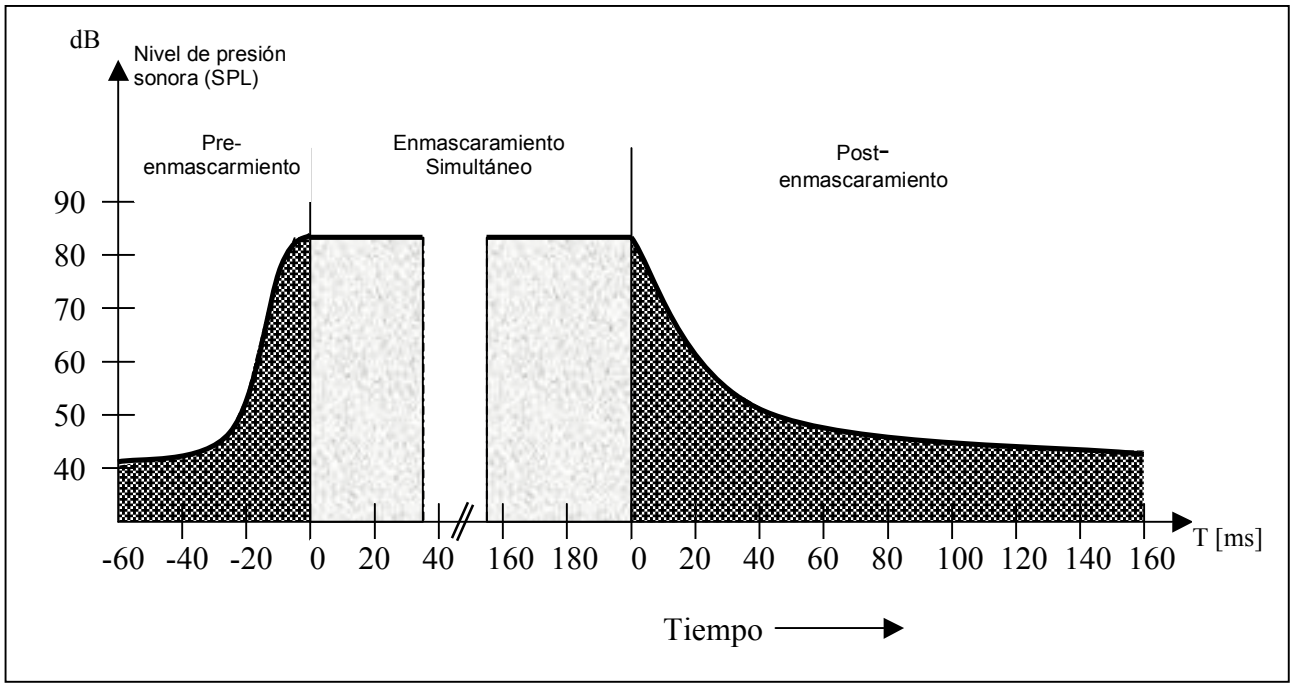


Figura 2.17 Efecto del pre-enmascaramiento y del post-enmascaramiento

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; de hecho, un estímulo sonoro debe tener una duración mínima para que se generen impulsos en las terminaciones nerviosas del órgano de Corti.

Las señales de gran intensidad requieren de un tiempo de formación de la sensación menor que el de las señales de baja intensidad; así, si una señal grande se presenta unos pocos milisegundos después que una señal pequeña, la sensación asociada a ésta puede no llegar a producirse, quedando efectivamente enmascarada. El fenómeno se extiende hasta unos 20 ms antes de la aparición de la señal enmascarante, independientemente del nivel de esta. Aun cuando el pre-enmascaramiento no tenga un efecto tan dramático como el enmascaramiento simultáneo o el post-enmascaramiento, es un importante punto que tomar para el diseño de codificadores perceptúales de audio ya que esta relacionado con el fenómeno de pre-eco o pre-ruido que es causado cuando se codifican bloques de muestras de la señal de entrada.

El efecto de post-enmascaramiento existe durante un intervalo máximo de unos 200 ms después de la desaparición de la señal enmascarante. Para explicar mejor el fenómeno supóngase el siguiente experimento:

Se ejecuta un tono enmascarante de 1 kHz a 60 dB junto con un tono de prueba de 1,1 kHz a 40 dB, el tono de prueba no puede oírse, está enmascarado. Se detiene el tono enmascarante y, luego de un pequeño retardo, se detiene el tono de prueba. Se ajusta el retardo al mínimo tal que el tono de prueba todavía pueda ser oído (por ejemplo 5 ms) y se registra dicho valor de tiempo. Si se repite la prueba para distintas intensidades del tono de prueba y se registran los diferentes tiempos se obtiene una curva como la de la Figura 2.18.

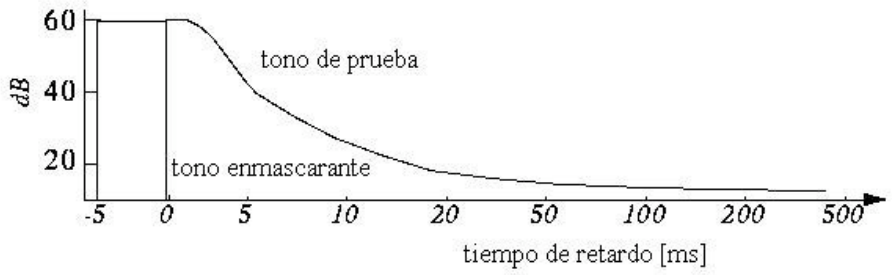


Figura 2.18 Variación del post-enmascaramiento en el tiempo.

El post-enmascaramiento depende del contenido frecuencial de las señales enmascarante y enmascarada. Diversos experimentos descritos en la literatura permiten concluir que la cantidad de post-enmascaramiento es mayor en las bajas frecuencias que en las altas.

Por último, los dos tipos de enmascaramiento interactúan produciendo una curva tridimensional como la que se muestra en la figura 2.19, donde todos los sonidos que se encuentren bajo la curva son inaudibles.

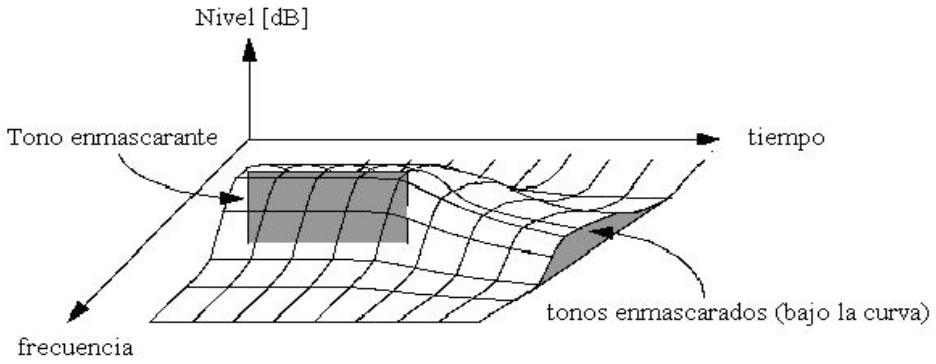


Figura 2.19 Efecto total del enmascaramiento frecuencial y temporal.

2.3.8 Relación Señal a Máscara

Para obtener la medición de las curvas de enmascaramiento se realizan pruebas similares a las utilizadas para obtener la curva característica del umbral de audición, pero existe una gran diferencia, la cual es que las curvas de enmascaramiento depende del tipo de señal con el cual se realice la prueba, es decir si se utilizan tonos o bandas estrechas de ruido ya sea como señal enmascarante o como señal enmascarada.

El primer tipo de prueba que mencionaremos será en la cual se utiliza una banda estrecha de ruido como señal enmascarante y un tono como señal enmascarada, en donde, la banda de ruido es menor o igual a una banda critica (lo cual se explicará posteriormente), en la figura 2.20 podemos observar las curvas de enmascaramiento características de 250 Hz, 1 kHz y 4 kHz, es decir, cuando la banda de ruido se encuentra centrada en dichas frecuencias y tienen un ancho de banda de 100 Hz, 160 Hz y 700 Hz respectivamente, además que las pendientes laterales de las bandas de ruido son de más de 200 dB por octava por lo cual se pueden considerar rectangulares y estas tienen un nivel de 60 dB, también podemos observar dentro de la figura con línea punteada el umbral de audición.

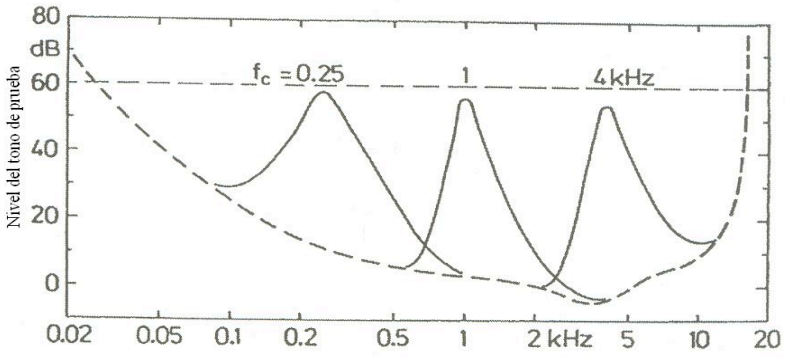


Figura 2.20 Umbrales de Enmascaramiento para banda angosta de ruido como señal enmascarante y tono como señal enmascarada.

Lo que podemos observar en la gráfica es que la curva de enmascaramiento depende de la frecuencia a la que esté centrada la señal enmascarante, y podemos notar que se extiende en un rango más amplio hacia las altas frecuencias que hacia las bajas, también se pude observar que las curvas correspondientes a 1 kHz y 4 kHz son muy parecidas pero aproximadamente debajo de los 500 Hz ya no se parecen, todas las que están debajo de este nivel tienden a ser mas anchas hacia las bajas frecuencias, esto se pude observar cuando se gráfica de forma logarítmica.

La diferencia entre el nivel de la señal enmascarante y el umbral de enmascaramiento se le conoce como Relación Señal a Máscara (SMR, por sus siglas en ingles); entre más grande sea este valor menor enmascaramiento existe, por lo tanto es muy importante conocer el SMR mínimo que genera una señal enmascarante para el diseño de un codificador de audio, por ejemplo, en la figura 2.20 podemos observar un valor mínimo de SMR de 2 dB para la señal de ruido centrada a 250 Hz y de 3 dB para la de 1 kHz.

En la figura 2.21 podemos observar los distintos umbrales de enmascaramiento para una banda de ruido centrada a 1 kHz a diferentes niveles de intensidad sonora, en esta se puede ver que el SMR mínimo no varía con la intensidad de la señal enmascarante, se mantiene alrededor de los 3 dB, y que para las frecuencias que se encuentran debajo de la frecuencia central de ruido, se conserva la pendiente del umbral de enmascaramiento en los diferentes niveles de intensidad, sin embargo esto no ocurre para las altas frecuencias ya que se presenta una deformación cuando aumenta el nivel de la señal enmascarante, esto es debido a que el oído humano tiene un comportamiento no lineal.

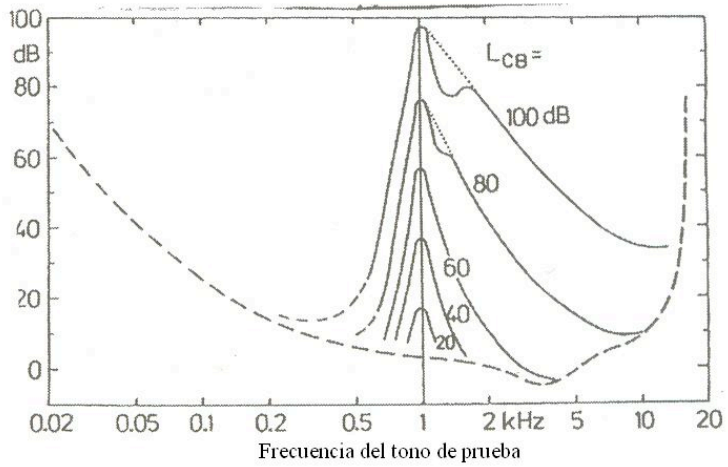


Figura 2.21 Umbrales de enmascaramiento para una banda de ruido estrecha centrada a 1 kHz para diferentes niveles de intensidad sonora

Otro tipo de forma de medir las curvas de enmascaramiento es utilizando como señal enmascarante un tono y como señal enmascarada igualmente un tono, pero este tipo de experimentos presentan muchas dificultades ya que los sujetos llegan a percibir tonos adicionales al utilizado en la prueba; pero se obtienen prácticamente los mismos resultados gráficos, pero existen diferencias en lo que respecta a la relación señal a

máscara, pues esta es mayor cuando se utiliza un tono como señal enmascarante, por lo que se puede concluir que una banda de ruido estrecha es mejor enmascarante que un tono.

De las cuatro combinaciones posibles, las correspondientes a banda estrecha de ruido y tono como señales enmascarantes y banda estrecha de ruido como señal enmascarada, no se ocupan para la realización de los modelos psicoacústicos utilizados en la codificación de audio por su poca practicidad, en lo que respecta a la relación señal a máscara, puesto que su SMR mínimo esta por arriba de los 20 dB.

2.3.9 Bandas Críticas

Para la obtención de las curvas que representan los umbrales de enmascaramiento se habló del concepto de banda crítica, esto es un rango de frecuencia en donde el umbral de enmascaramiento tiene una respuesta plana, hasta que cierta frecuencia decae súbitamente, esto lo podemos observar en la figura 2.22, en donde para obtener esta gráfica se enmascara un tono de 2kHz por medio de dos bandas de ruido estrechas a la misma distancia del tono con una intensidad de 50 dB , la gráfica se obtiene graficando la separación entre las bandas de ruido contra el nivel de intensidad del tono.

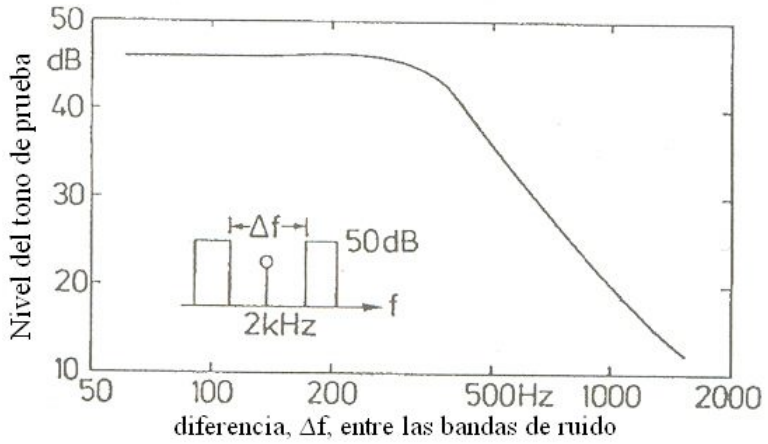


Figura 2.22

Se puede ver claramente como el nivel del tono se mantiene constante aproximadamente en 46 dB hasta que la distancia entre las dos bandas llega a ser de, aproximadamente, 300 Hz, es decir a 150Hz de el tono, en ese punto el nivel de intensidad del tono empieza a disminuir.

El sistema auditivo del ser humano tiene una respuesta en frecuencia limitada en cuanto a resolución, es decir, que existen bandas de frecuencia que el oído humano percibe como una sola, siendo incapaz de identificar diferencias entre estas dos frecuencias distintas dentro de una misma banda, esto se debe a que al vibrar la membrana basilar ocupa cierta longitud y cierta cantidad de células ciliares para identificar una frecuencia, por lo que si existen dos frecuencias muy cercanas provocando vibraciones en la misma zona de la membrana, esta no puede distinguir entre las dos. Por lo que puede entenderse la banda crítica como la mínima banda de frecuencias alrededor de una frecuencia determinada que activan la misma zona de la membrana basilar.

Los anchos de banda de las bandas críticas dependen de la frecuencia, estos se mantienen constantes, de unos 100 Hz, de 0 Hz hasta los 500 Hz; mientras que por arriba de los 500 Hz crecen en proporción de la frecuencia y son aproximadamente el 20% de la frecuencia central de la banda crítica. En la figura 2.23 podemos observar una aproximación del ancho de banda de las bandas críticas en función de la frecuencia.

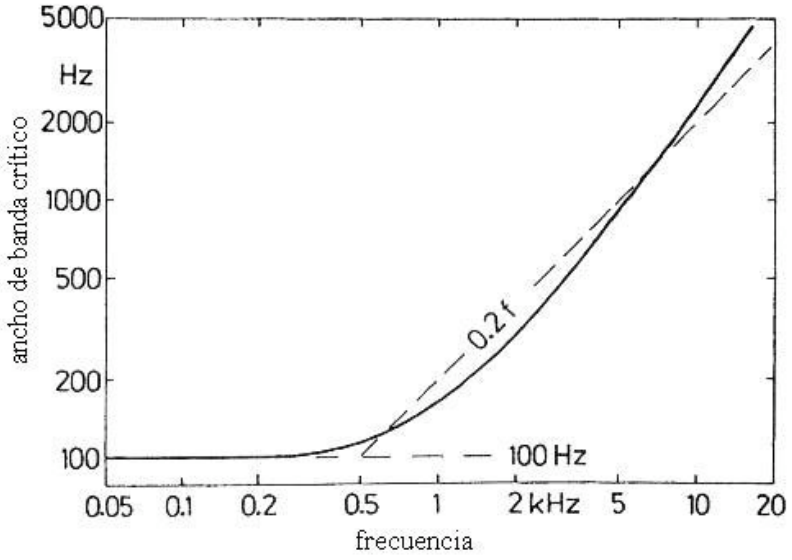


Figura 2.23 Ancho de banda de las bandas críticas en función de la frecuencia.

Una expresión analítica que puede describir la variación del ancho de banda de las bandas críticas con respecto a las frecuencias es:

$$\Delta f_{BC} = 25 + 75\left[1 + 1.4\left(\frac{f_c}{1000}\right)^2\right]^{0.69}$$

Esta ecuación fue dada a conocer por el investigador E. Zwicker, la cual es aceptada como la descripción estándar de la variación de los anchos de banda críticos.

Sin embargo existe una gran cantidad de publicaciones en los que se menciona un desacuerdo con esta ecuación, en especial en la estimación de los anchos de banda por debajo de los 500 Hz. En especial Moore y Glasberg definieron lo que se conoce como “ancho de banda rectangular equivalente” (ERB por sus siglas en ingles), la cual es una función que se acerca con más exactitud a los anchos de banda críticos reales, esta función se define como:

$$ERB = 24.7\left(\frac{4.7 f_c}{1000} + 1\right)$$

En la siguiente figura (2.24) se puede observar las curvas de cada ecuación así como algunos de los valores experimentales de los anchos de banda críticos encontrados por diversos autores.

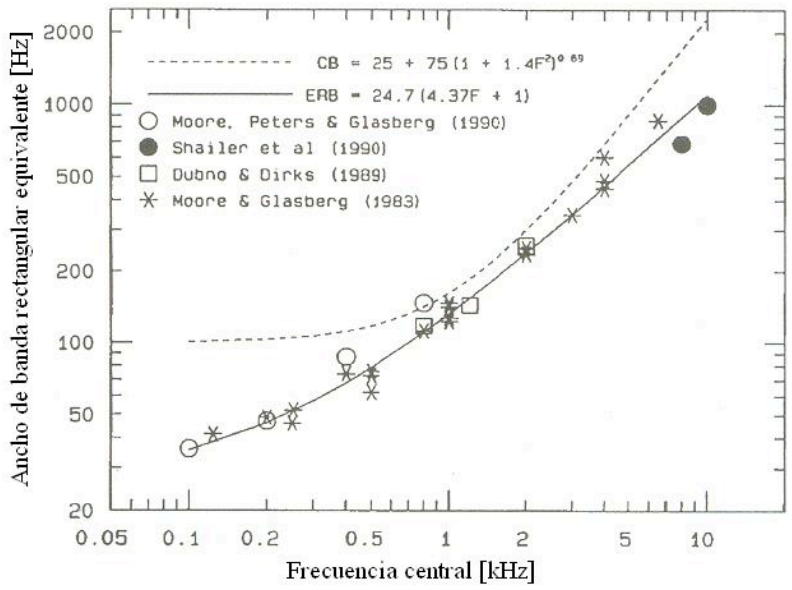


Figura 2.24 Comparativa entre la ecuación de Zwicker y el ERB.

Es posible subdividir el rango de frecuencias audibles en intervalos adyacentes de una banda crítica de ancho, que no se traslapan entre sí, y que representan una primera aproximación al problema de modelar la selectividad en frecuencia del oído interno. En la siguiente figura (2.25) podemos observar el número de bandas críticas adyacentes dentro del rango audible, las cuales dividen al rango audible en 24 bandas y estas mismas se enumeran en la tabla 2.2 con su frecuencia central y frecuencia de corte superior.

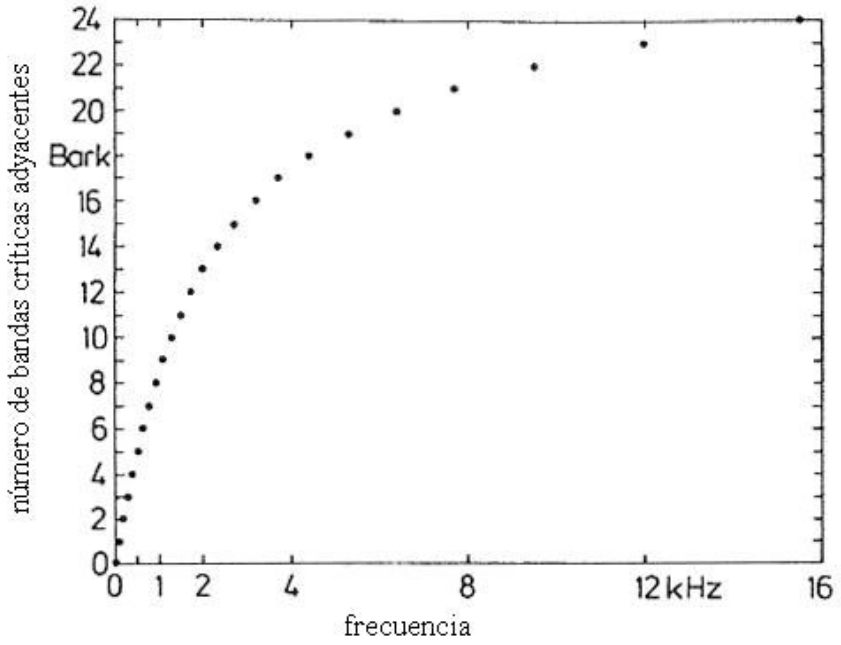


Figura 2.25 Número de bandas críticas en función de la frecuencia.

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

Tabla 2.2 Distribución de las bandas críticas en función de la frecuencia

2.3.10 Escala Bark.

Aparentemente cada banda crítica correspondería a una distancia fija a lo largo de la membrana basilar, de aproximadamente 1.3 mm de longitud (independientemente de la frecuencia central), abarcando unas 150 células receptoras en el órgano de Corti, de un total de 3,600 células capilares que hay en línea entre el helicotrema y la ventana oval bajo la membrana basilar.

Asumiendo que a cada banda critica le corresponde una longitud dentro de la membrana basilar, se puede definir como unidad de longitud z (f) a cada banda critica. Esta unidad es conocida como Bark, un intervalo de frecuencia de 1 bark es, por definición, un intervalo de una BC de ancho en cualquier punto del rango de frecuencias audibles.

Como la ecuación de banda critica representa la cantidad df/dz, lo cual nos dice el cambio de frecuencia a lo largo de la membrana basilar, podemos invertir e integrar esta ecuación para así obtener z (f) que se le conoce como la tasa de bandas criticas. La relación entre

la tasa de bandas críticas y la frecuencia puede ser expresada mediante la siguiente ecuación, la cual permite calcular la tasa de bandas críticas (en barks), z, correspondiente a la frecuencia en Hz, f, con un error inferior a ±0.2 barks:

$$z(f) = 13\arctan\left(\frac{0.76f}{1kHz}\right) + 3.5\arctan\left(\left(\frac{f}{7.5kHz}\right)^2\right)$$

Por lo tanto podemos observar que en la tabla 2.2 el número de banda crítica es también el valor de la banda crítica en barks.

Capítulo III

AUDIO DIGITAL

Introducción

En esta sección abordaremos la teoría necesaria para entender como se digitaliza una señal de audio, la historia de este revolucionario concepto, desde sus inicios hasta fechas recientes así como los tipos de compresión de audio digital que existen en el mercado.

3.1 Historia

El audio profesional ha sido manejado, predominantemente, en equipos analógicos por más de un siglo, pero en la actualidad se ha dado una transición de manera vertiginosa a la tecnología de audio digital. La explicación de dicha transición es que con los elementos digitales se pude procesar y almacenar la información de manera más segura y económica.

La introducción del audio digital era una idea un poco revolucionaria, sin embargo, las cadenas de televisión tales como la BBC de Londres y la NHK de Japón, fueron precursoras de dicha introducción a principios de 1967.

Algunos acontecimientos importantes de la historia del audio y del audio digital se encuentran en la tabla 3.1

Tabla 3.1

1857	Leon Scott muestra una maquina parlante
1877	Edison archiva la patente “La mejora en el fonógrafo o máquina parlante”
1878	Primera grabación de música del concertista Jules Levy
1887	Emile Berliner patenta el gramófono
1895	Marconi lleva a cabo la primer transmisión radiofónica de Italia a América del Norte
1906	Lee DeForest inventa el bulbo
1910	Primer transmisión radiofónica de Caruso desde Nueva York
1916	Fundación de la Sociedad de Ingenieros de Películas (SMPE, por sus siglas en ingles)
1919	Fundación de la Corporación de Radio de América (RCA, por sus siglas en ingles)
1925	Grabación y reproducción eléctrica introducida por Laboratorios Bell
1927	Primer película con sonido (The Jazz Singer)
1929	Harry Nyquist publica su fundamento matemático para el teorema de muestro
1937	Código binario PCM inventado por A.H. Reeves
1941	Primer transmisión comercial en FM en Estados Unidos de Norteamérica
1948	Columbia introduce la grabadora de LP (long playing)
1961	Reverberación digital simulada por computadora
1967	NHK (Japón) muestra la primer grabadora de cinta en formato de audio digital
1969	El Dr. Thomas Stockham comienza a experimentar con cintas digitales para grabación
1971	Introducción del retardo digital (digital delay) para efectos La BBC muestra su grabadora digital estéreo
1972	Primeras grabaciones masterizadas digitalmente por Nippon Columbia,

	Japón.
	El código PCM es utilizado en el Reino Unido para transmisiones de alta calidad de sonido en radio y TV
1975	Sistema de reverberación digital en tiempo real disponible
1976	Lanzamiento de una grabación de Caruso restaurada digitalmente
	La BBC muestra una grabadora digital de 10 canales
1977	Lanzamiento masivo de grabadoras profesionales de audio digital
1980	Se anuncia la idea del sistema de disco compacto (compact-disk)
1982	Lanzamiento de los primeros reproductores de disco compacto
1983	El cable de fibra óptica es utilizado para transmisitir audio digital desde Washington a Nueva York
1984	Lanzamiento de la primer consola de audio digital.
1987	Digidesign produce la primer estación de trabajo de audio digital llamada Sound Tools
1990	Dolby propone su sistema de Teatro en Casa que consiste en 5 canales y un subwoofer
1991	El formato quicktime de Apple aparece en el mercado
1992	Sony lanza a nivel comercial sus primeras grabadoras/reproductores de audio digital
1996	Grabaciones experimentales en formato digital a 24 bits 96 kHz
1997	Presentación del DVD
1998	El primer reproductor de MP3 aparece
1999	Shawn Fanning crea el NAPSTER que consiste en un intercambio de archivos musicales por internet
2001	Aparece el reproductor iPod de Apple y el codificador MP3 Pro
2004	El Instituto Fraunhofer termina el programa de MP3 surround

3.2 Muestreo

El uso de métodos digitales para la grabación, reproducción y almacenamiento de las señales de audio digital encierran conceptos fuera del alcance de los métodos utilizados en el audio analógico. El audio tiene un origen netamente analógico, por lo cual los sistemas digitales utilizan como pilares dos procesos: el muestreo y la cuantización, los cuales abordaremos a continuación.

Antes de entrar de lleno al muestreo y a la cuantización, es necesario aclarar que algunos conceptos matemáticos se darán por entendidos, tomando en cuenta que el lector tiene conocimientos previos de dichas materias.

La unión entre las formas de onda analógicas, en nuestro caso el audio, y su versión digital es a lo que se le conoce como el proceso de muestreo [24]. Es posible llevar a cabo este proceso de diferentes maneras, uno de los procedimientos mas utilizados es haciendo una operación de *muestreo y retención* (sample-and-hold); en la cual interactúan un switch y un mecanismo de almacenamiento (tal como un transistor y un capacitor), de los cuales obtenemos una secuencia de muestras de la señal de entrada. A la salida del proceso de muestreo se le conoce como *modulación por amplitud de pulso*[25] *o PAM* (por sus siglas en ingles) ya que los intervalos sucesivos pueden ser considerados como una secuencia de pulsos con amplitudes derivadas de las muestras de la señal de entrada. La forma de onda original puede ser obtenida a partir de su forma PAM a través de un filtro paso-bajas.

3.2.1 Teorema del muestreo

El teorema del muestreo establece que una señal en el tiempo $s(t)$, cuya transformada de Fourier sea igual a cero para frecuencias mayores de f_m [Hz], es decir, limitada en banda[25], y de la que se conozcan todos los valores de la integral de n para $t = nT_s$, donde:

$$T_s \leq \frac{1}{2f_m} \quad (3.1)$$

entonces $s(t)$ puede ser determinada por medio de los valores de una secuencia de puntos equidistantes o muestras. A esto se le conoce como el *teorema del muestreo uniforme*.

El limite superior de T_s $2f_m$, es conocido como *tasa de muestreo de Nyquist - Shannon* y puede ser expresado a través de su reciproco

$$f_s = \frac{1}{T_s} \quad (3.2)$$

Con lo cual la ecuación (3.1) se puede expresar como,

$$f_s \geq 2f_m \quad (3.3)$$

A f_s se le conoce como frecuencia de muestreo de Nyquist [24][25] y debe ser por lo menos igual al doble de la frecuencia máxima de la señal a muestrear. A continuación se demostrara el teorema con tres diferentes aproximaciones

3.2.2 Muestreo y retención, sample and hold

En la figura 3.1 se muestra un tren de pulsos multiplicando a una señal $s(t)$. Como se puede observar la señal $p(t)$ es un pulso rectangular unitario, por lo que en la figura 3.2 veremos el resultado de esta multiplicación que podría considerarse como una forma de onda muestreada de la señal $s(t)$.

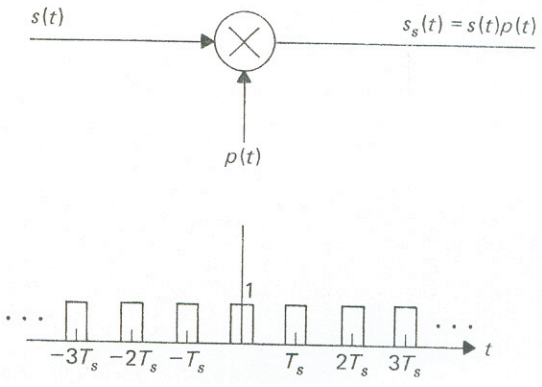


Figura 3.1

Al multiplicar $s(t)$ por $p(t)$ no estamos muestreando, en el estricto sentido de la palabra, sino que estamos realizando una selección de aquellas porciones de una onda que existe durante ciertos intervalos o la que tiene ciertas magnitudes, conocido como *gating* en el tiempo, por que se puede interpretar como el abrir y cerrar de una compuerta.

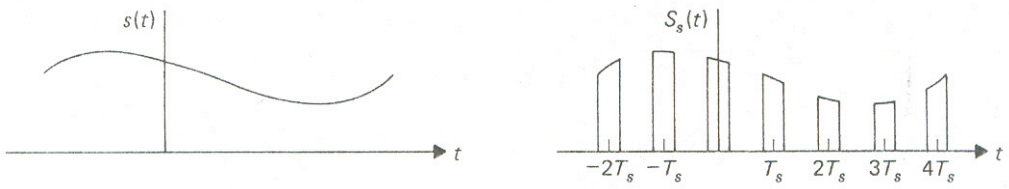


Figura 3.2

Para demostrar el teorema del muestreo es necesario obtener la transformada de Fourier de la forma de onda muestreada y, a partir de ahí, recuperar la señal $s(t)$, para ello asumiremos la forma de $S(f)$, que es la transformada de $s(t)$.

Bajo las condiciones del teorema de muestreo, se delimita una frecuencia máxima f_m para $S(f)$, como se muestra en la figura 3.3.

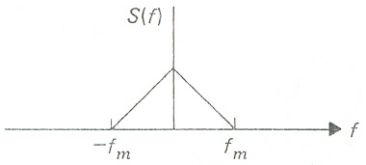


Figura 3.3

Dado que el tren de pulsos se considera que es una función periódica, se puede expandir en una serie de Fourier. Para simplificar el análisis se asume que $p(t)$ es impar. Si este no fuera el caso, la serie de Fourier contendrá términos senoidales en lugar de cosenoidales. Esto puede solucionarse añadiendo un desplazamiento en fase y no se afectarían las magnitudes de los términos, por lo cual el siguiente análisis es valido en cualquiera de las dos opciones.

La salida del multiplicador se puede expresar como,

$$s_s(t) = s(t)p(t)$$
$$= s(t)\left[a_0 + \sum_{n=1}^{\infty} a_n \cos 2\pi n f_s t\right]$$

$$= a_0 s(t) + \sum_{n=1}^{\infty} a_n s(t) \cos 2\pi n f_s t \quad (3.4)$$

en donde $f_s = \frac{1}{T_s}$.

La transformada de Fourier de $s_s(t)$ se puede encontrar por la aplicación sucesiva del teorema de modulación[24]. El primer término de la ecuación (3.4) es una versión a escala de la función original y su transformada de Fourier es la primera sección de $S_s(f)$ centrada alrededor de la frecuencia cero. El resto de los términos representan a la señal $s(t)$ multiplicada por una función coseno, cuya transformada puede ser encontrada, nuevamente, por el teorema de modulación. La transformada de la suma es, por lo tanto, una suma de versiones distintas de $S(f)$, dichas versiones no son mas que el desplazamiento de la frecuencia central del espectro debido a la frecuencia de la función coseno y cuya amplitud se ve afectada por el término a_n. Una representación de $S_s(f)$ se muestra en la figura 3.4. Si se cumple con la condición de $f_s \geq 2f_m$ los lóbulos no deben de traslaparse.

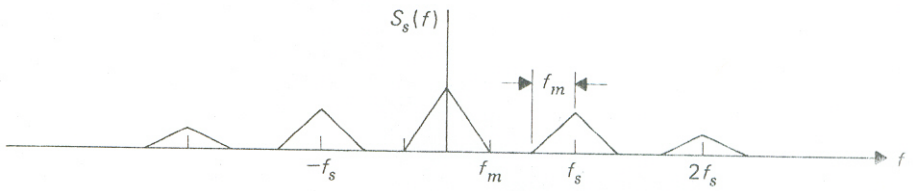


Figura 3.4

Los sistemas y señales poseen diversas propiedades básicas, una de las cuales nos dice que es trivial separar señales si una o ambas condiciones (dualidad) existen:
-Las señales no deben estar traslapadas en el dominio del tiempo o de la frecuencia.
-Las señales que no estén traslapadas en frecuencia pueden ser separadas completamente con el uso de filtros paso-bajas.
Al cumplirse ambas condiciones, es posible filtrar el espectro de $S_s(f)$ y así obtendremos el espectro de $s(t)$, por lo que el teorema del muestreo queda demostrado.

3.2.3 Muestreo natural

Asumiremos ahora que $S(f)$ del muestreo anterior, es diferente a cero a lo largo de una porción limitada del eje de la frecuencia por lo que se puede expandir en una serie de Fourier en el intervalo

$$-f_m < f \leq f_m$$

En la serie de Fourier expandida obtendremos lo siguiente:

$$S(f) = \sum_{-\infty}^{\infty} c_n e^{jnt_0 f}$$

donde

$$t_0 = \frac{\pi}{f_m}$$

El término c_n en la ecuación anterior esta dado por,

$$c_n = \frac{1}{2f_m} \int_{-\infty}^{\infty} S(f) e^{-jnt_0 f} df \quad (3.5)$$

La integral inversa de Fourier nos dice que,

$$s(t) = \int_{-\infty}^{\infty} S(f) e^{j2\pi ft} df = \int_{-f_m}^{f_m} S(f) e^{j2\pi ft} df \quad (3.6)$$

En la ecuación (3.6) los limites se reducen ya que $S(f)$ es igual a cero fuera del intervalo de $-f_m$ a f_m, comparando la ecuación (3.6) con la ecuación (3.5), tenemos que

$$c_n = \frac{1}{2f_m} s\left(-\frac{nt_0}{2\pi}\right) = \frac{1}{2f_m} s\left(-\frac{n}{2f_m}\right)$$

Se conocerá c_n una vez que se conozca la señal $s(t)$ en los puntos $t = \frac{n}{2f_m}$. Sustituyendo los valores de c_n en la serie expandida de $S(f)$,

$$S(f) = \sum_{n=-\infty}^{\infty} c_n e^{jnt_0 f} = \frac{1}{2f_m} \sum_{n=-\infty}^{\infty} s\left(-\frac{n}{2f_m}\right) e^{jn\pi \frac{f}{f_m}} \quad (3.7)$$

La ecuación (3.7) es una declaración matemática del teorema del muestreo, la cual establece que $S(f)$ es completamente conocida y limitada por el valor de las muestras $s(-n/2f_m)$. Estos son los valores de $s(t)$ de puntos equidistantes entre si en el dominio del tiempo. Podemos sustituir la ecuación (3.7) en la integral inversa de Fourier para encontrar $s(t)$en términos de sus muestras.

$$s(t) = \sum_{n=-\infty}^{\infty} \frac{1}{2f_m} \int_{-f_m}^{f_m} s\left(-\frac{n}{2f_m}\right) e^{jn\pi f / f_m} e^{j2\pi ft} dt$$

$$= \sum_{n=-\infty}^{\infty} s\left(-\frac{n}{2f_m}\right)\left[\frac{\sin(2\pi f_m t + n\pi)}{2\pi f_m t + n\pi}\right] \quad (3.8)$$

La ecuación (3.8) nos dice como encontrar $s(t)$ para cualquier tiempo t para los valores de $s(t)$ en los puntos $t = n/2f_m$. Este tipo de muestreo emplea la tasa de muestreo mínima que es $T_s = 1/2f_m$.

3.2.4 Muestreo por pulso o impulse sampling

Dado que $S(f)$ es la transformada de $s(t)$ y es igual a cero fuera del intervalo anteriormente establecido, $-f_m$ a f_m, la figura 3.3 nos servirá de base para representar a $S(f)$ como una señal de banda limitada[4],[14],[19], Ahora consideramos el producto de $s(t)$ con un tren de pulsos unitario, que se define de la siguiente manera,

$$s_\delta(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \quad (3.9)$$

La propiedad de desplazamiento *shifting* de una función impulso nos dice que,

$$s(t)\delta(t - t_0) = s(t_0)\delta(t - t_0) \quad (3.10)$$

Utilizando dicha propiedad, observamos que la versión muestreada de $s(t)$, llamada, $s_s(t)$ esta definida por,

$$s_s(t) = s(t)s_\delta(t) = \sum_{n=-\infty}^{\infty} s(t)\delta(t - nT_s)$$

$$= \sum_{n=-\infty}^{\infty} s(nT_s)\delta(t - nT_s) \quad (3.11)$$

Utilizando la propiedad de la transformada de Fourier para la convolucion en el dominio de la frecuencia, el producto $s(t)s_\delta(t)$ de la ecuación (3.11) queda representado en el dominio de la frecuencia como la convolucion de $S(f) * S_\delta(f)$, donde $S_s(f)$ es la transformada de Fourier del tren de impulsos,

$$S_\delta(f) = \frac{1}{T_s}\sum_{n=-\infty}^{\infty} \delta(f - nf_s) \quad (3.12)$$

La convolucion con una función impulso simplemente desplaza la función original, de la siguiente manera,

$$S(f) * \delta(f - nf_s) = S(n - f_s) \quad (3.13)$$

por lo que obtenemos,

$$S_s(f) = S(f) * \frac{1}{T_s} \sum_{n=-\infty}^{\infty} \delta(f - nf_s)$$

$$= \frac{1}{T_s} \sum_{n=-\infty}^{\infty} \delta(f - nf_s) \quad (3.14)$$

la figura 3.5 nos muestra la gráfica de la función $S_s(f)$

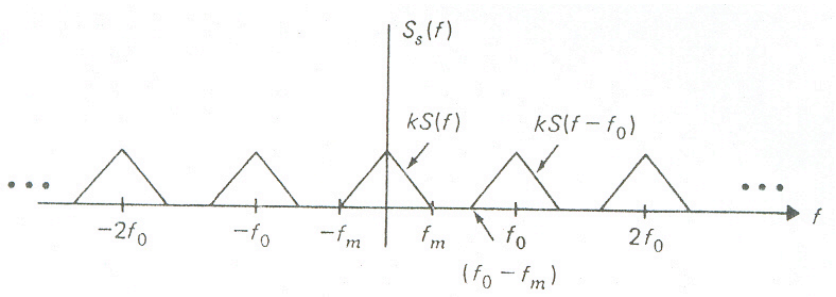


Figura 3.5

Por lo que concluimos que el espectro de la función $S_s(f)$ que es la señal muestreada $s_s(t)$ es, con un factor constante de $1/T_s$, es el mismo que el de la función original, $s(t)$.

Cuando se escoge la tasa de muestreo, como se hizo en este caso, tal como $f_s = 2f_m$, cada réplica del espectro está separada a una banda de frecuencia igual a f_s [Hz], y la señal analógica original puede ser recuperada por medio de sus muestras utilizando un filtro. Como se observa en la figura 3.5, al definir una tasa de muestreo correcta o mayor a la frecuencia máxima las replicas se encontrarán con una mayor distancia entre ellas, esto facilita la operación de filtrado. Un filtro paso-bajas puede ser utilizado para separar la señal original. Sin embargo, cuando la tasa de muestreo es $f_s \langle 2f_m$, las replicas se traslaparán, como se muestra en la figura 3.6, por lo que parte de la información se perderá. Este fenómeno es el resultado del submuestreo[24] y es conocido como *aliasing* o traslape.

3.3 Traslape (Aliasing)

Uno de los principales retos cuando se digitalizan las señales es el efecto conocido como *traslape*, lo cual es un tipo de confusión a la hora del muestreo y se puede presentar a la hora de grabar la señal. El traslape puede crear componentes falsos de la señal que se este muestreando. Estos errores pueden aparecer dentro del ancho de banda de la señal y son imposibles de distinguir de la original.

3.3.1Traslape debido a frecuencias superiores a la frecuencia de muestreo

Hemos visto que una de las condiciones primordiales para el muestreo de una señal es que sea de banda limitada, esto se puede lograr con un filtro paso-bajas. Para nuestro

caso particular el filtro debe de atenuar todas las frecuencias arriba de 20 [kHz]. En caso de que la señal no tuviera dicho filtro se pueden presentar efectos indeseados, como por ejemplo la distorsión por *traslape*. Los cambios de amplitud de las frecuencias más altas no serán codificados correctamente por lo que dicha información no es correcta. El submuestreo de dichas frecuencias puede crear una serie de frecuencias erroneas, lo cual puede presentar otra forma de distorsión.

El efecto de *traslape* es consecuencia de muestrear una señal que no sea de banda limitada. Cuando las frecuencias de la señal a muestrear son cada vez mayores que la frecuencia de muestreo, el numero de muestras por ciclo es menor; podemos observar este efecto en la figura 3.6

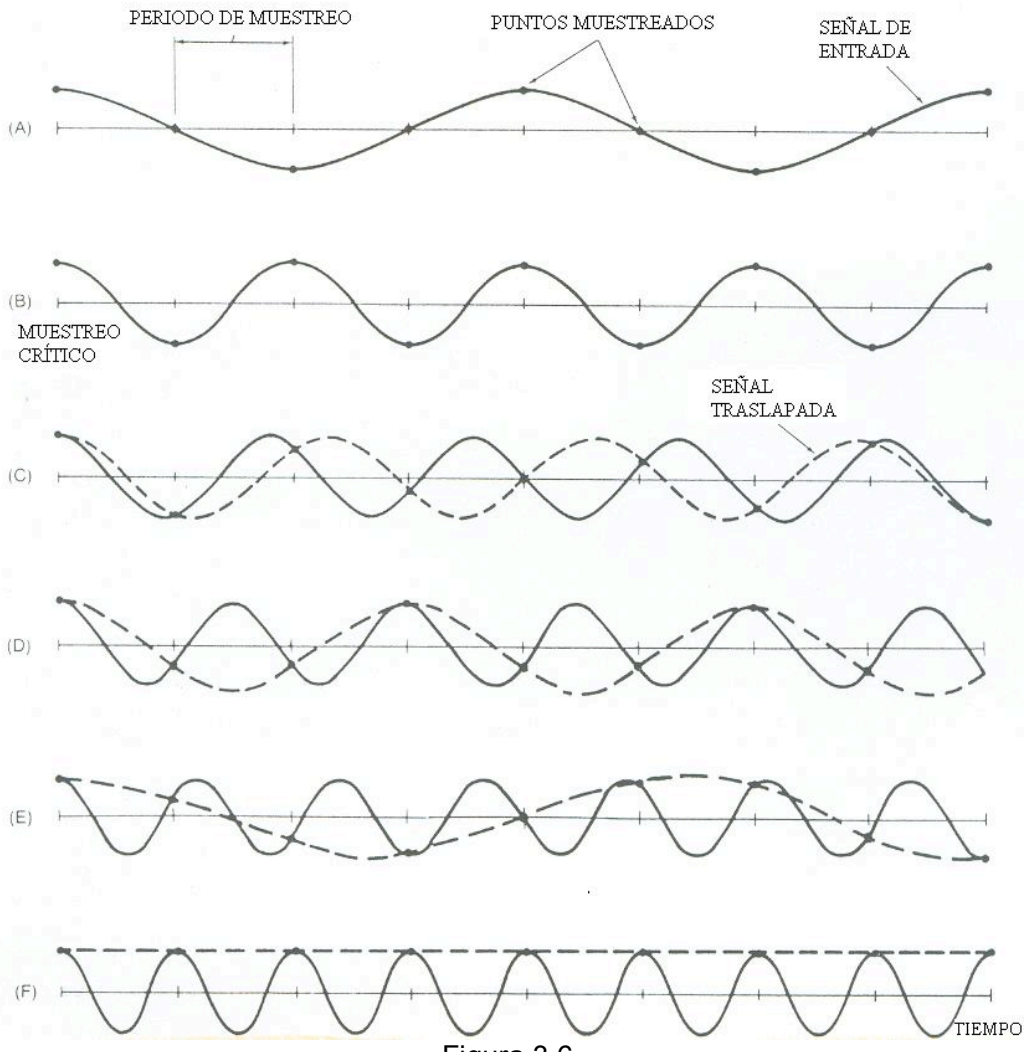


Figura 3.6

Si tratamos de muestrear frecuencias aun mayores, el dispositivo producirá muestras "corregidas" a su tasa de muestreo, pero la variación en amplitud de las muestras de esas frecuencias superiores crea información errónea y esta afecta a otras frecuencias. En medida que esas frecuencias sean mayores, se crearán nuevas que caerán dentro del ancho de banda de nuestra señal.

Si la tasa de muestreo es *S* y *F* es una frecuencia mayor que la mitad de la tasa de muestreo, entonces una frecuencia nueva F_a es creada a:

$$F_a = S - F \quad (3.15)$$

Por ejemplo, si intentamos muestrear una señal a 40 [kHz] con una tasa de muestreo de 44 [kHz], utilizada en los discos compactos, se nos presentara una señal a 4 [kHz]. En otras palabras una nueva frecuencia de 4 [khz] es producida.

3.3.2 Efecto de traslape en el espectro

Las componentes de *traslape* no ocurren solo dentro de nuestra tasa de muestreo, si observamos el espectro producido por el muestreo, notaremos que existen frecuencias tipo *alias* arriba de nuestra frecuencia f_s como se observa en la figura 3.7

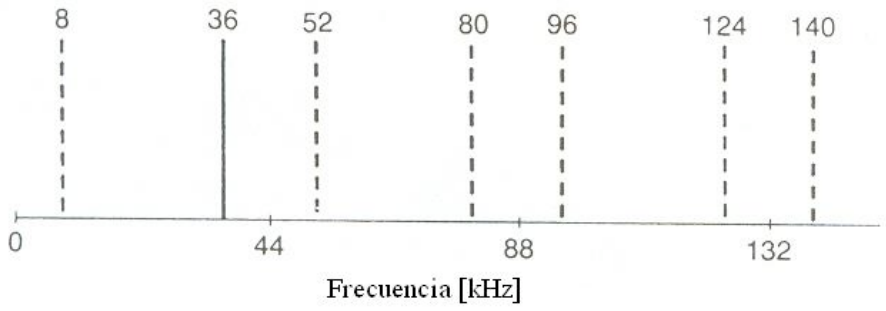


Figura 3.7

Todas estas nuevas componentes son creados por el efecto de imagen del *alias* con la característica de que dichas componentes se encontraran a lo largo del espectro de la siguiente manera:± *S* ± *F,* ± *2S* ± *F,* ± *3S* ± *F,* etc.

Las únicas componentes que nos preocupan son aquellas que se encuentran dentro de nuestro ancho de banda, de cualquier manera es importante conocer que existe este tipo de errores.

3.3.3 Prefiltrado para evitar el efecto de traslape

En la mayoría de los casos es deseable minimizar la tasa de muestreo de un sistema discreto en el tiempo para procesar sistemas analógicos. Esto se debe a la cantidad de procesos aritméticos que se requieren para implementar el sistema, ya que este es proporcional al número de muestras a procesar. En caso de que la señal de entrada no esté limitada en banda o que la señal tenga una frecuencia mayor a la frecuencia de muestreo, el prefiltrado es usado regularmente. Aun si la señal es naturalmente limitada en banda es recomendable colocar la etapa de prefiltrado.

A este filtro paso bajas que precede al dispositivo que lleva a cabo el muestreo es llamado *filtro anti-traslape.*

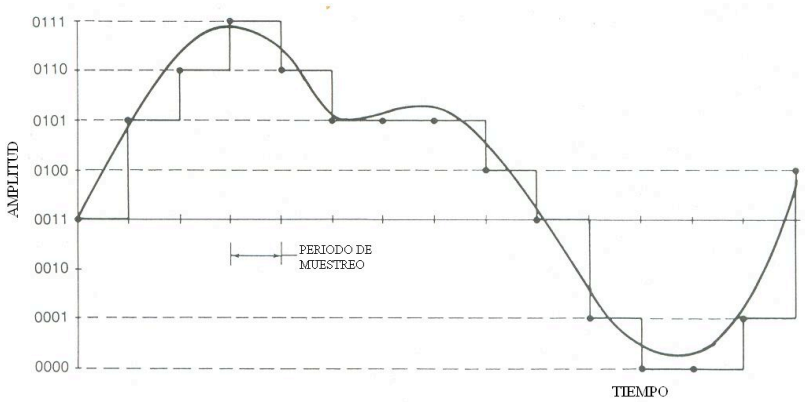
3.4 Cuantización

Para grabar una señal de audio en un medio digital, se necesitan tener dos parámetros de información, o tener la información en dos dimensiones. El muestreo guarda la información en el tiempo de la señal y la cuantización guarda la información referente a la amplitud. Por lo tanto, la cuantización es el valor medido de la señal analógica al tiempo que se toma una muestra de dicha señal.

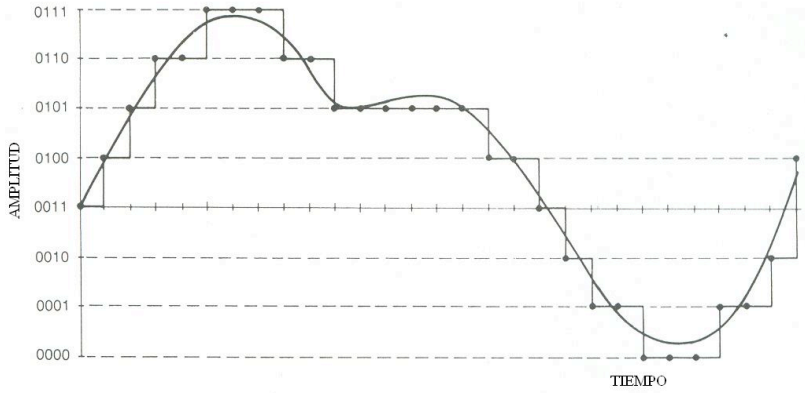
3.4.1 Aproximación en las mediciones

El muestreo representa el tiempo de la medición y la cuantización representa el valor de dicha medición, o en el caso concreto del audio, la amplitud de la forma de onda de la señal al tiempo del muestreo. El muestreo y la cuantización son la base fundamental para la digitalización.

La interacción entre la tasa de muestreo y la cuantización se observa claramente en la figura 3.8. El muestreo correcto de una señal limitada en banda es un proceso sin perdidas, *loosless*[13][14] en ingles, pero el escoger el valor de la amplitud en el tiempo de muestreo no lo es. Cualquiera de las escalas o códigos, mostrados en la figura 3.9, que se escoja para la cuantización nos arroja el mismo resultado; aun cuando aumentemos el numero de bits para llevar a cabo la cuantización nunca podremos codificar la señal de audio en su totalidad, esto se debe a que una señal analógica tiene un valor infinito de amplitud entre dos puntos cualesquiera y nosotros solo tenemos un numero finito de incrementos por lo cual, nuestro valor solamente será una aproximación del valor real.

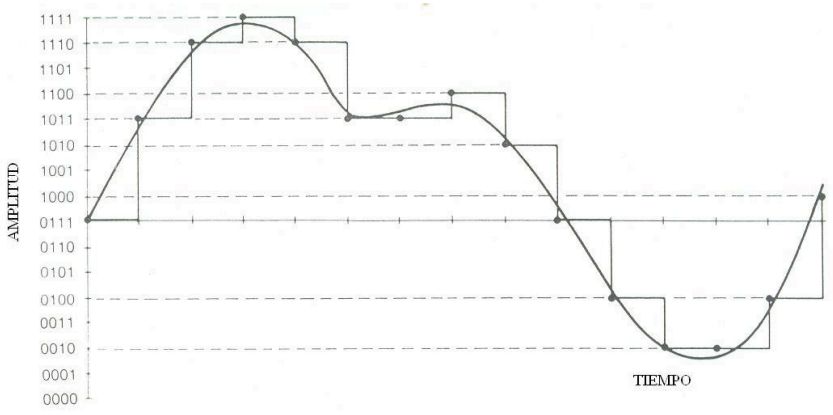


(A) El número de intervalos es bajo, resultando en una mala aproximación

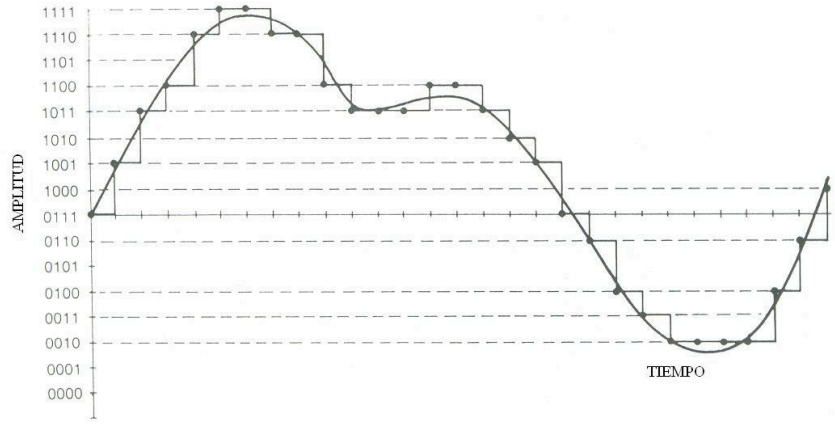


(B) La frecuencia de muestreo se duplica, pero la aproximación todavía es mala

Figura 3.8



(C) Con la frecuencia de muestreo baja y el número de intervalos de cuantización duplicados, la aproximación mejora



(D) Frecuencia de muestreo y número de intervalos de cuantización duplicados

Figura 3.9

3.4.2 Error de cuantización

El error de cuantización es la diferencia entre el valor actual analógico al momento de la muestra y el valor que determina el cuantizador en ese instante (figura 3.10). En el momento que se realiza la muestra, el valor de la amplitud debe ser aquel que se encuentre más cercano a uno de los valores de cuantización. En el mejor de los casos (muestras 11 y 12) la forma de onda coincide con el intervalo de cuantización, en el peor (muestra 1) la forma de onda se encuentra exactamente entre dos valores posibles de cuantización. Por lo tanto el error de cuantización esta limitado a ± ½ intervalo.

$$e = g_a - g_d \quad (3.16)$$

Donde e se define como el valor del error, g_a es el valor actual de la amplitud de la señal analógica y g_d es el valor del cuantizador.

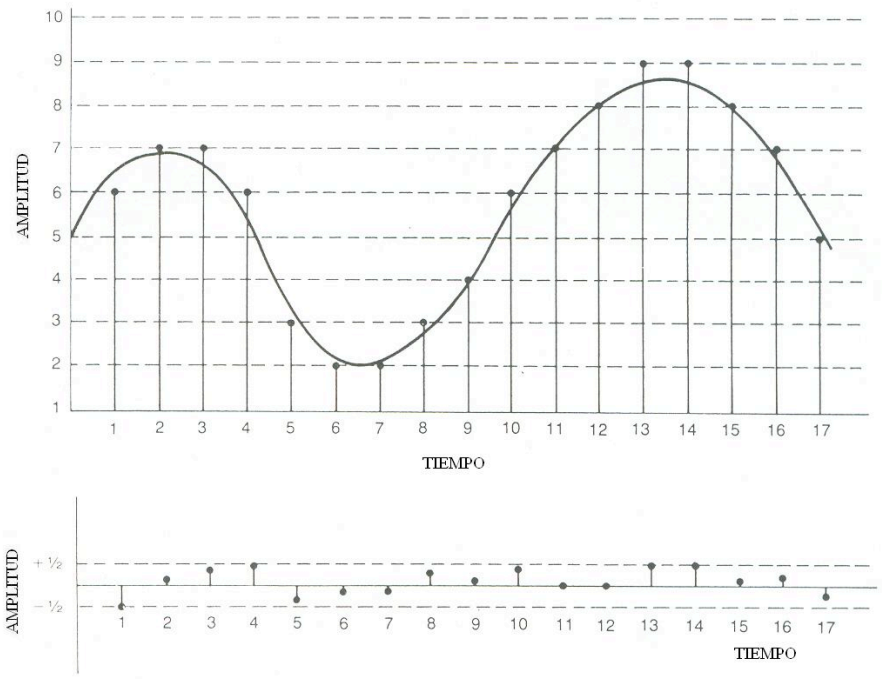


Figura 3.10

3.4.3 Relación señal a ruido (SNR, por sus siglas en ingles)

Con un sistema de números binarios, la longitud de la palabra determina el número de incrementos disponibles para hacer la cuantización; se puede obtener el valor elevando el número 2 a la longitud de la palabra. Así, una palabra de 8 bits nos permitirá 256 incrementos, $2^8 = 256$, una palabra de 16 bits tendrá $2^{16} = 65{,}536$ incrementos.

El error de cuantización puede ser inaudible en cierto punto, es por ello que los discos compactos de audio tienen una resolución de 16 bits.

La aproximación de la relación señal a ruido esta dada por el número máximo de incrementos entre el máximo error que se obtiene. En el caso que la longitud de palabra sea de 16 bits,

$$S/N = \frac{65{,}535}{.5} = 131{,}070$$

$$S/N_{dB} = 98dB$$

El valor de 0.5 se obtiene del error máximo de cuantización, que se presenta en la ecuación 3.16

Esto determina el valor máximo de la relación señal a ruido del sistema.

3.4.4 Otros métodos de cuantización

La cuantización es más que la longitud de una palabra de cómputo, es también una cuestión de diseño del hardware y formatos. Existen muchas técnicas para complementar la cuantización. Por ejemplo, podemos utilizar una distribución lineal o una no lineal en los intervalos de cuantización para la escala de la amplitud, o bien, se puede utilizar un sistema con modulación delta en el que solamente se utilice un bit de cuantización para codificar la amplitud, utilizando dicho bit solo como signo. Este tipo de decisiones influyen en la eficiencia del sistema.

3.5 Digitalización del sonido. Modulación por codificación de pulsos (PCM)

Por definición encontramos que la modulación por código de pulso es la representación digital de una señal análoga, donde la magnitud de la señal es muestreada y cuantizada a una serie de símbolos en código digital, generalmente código binario.

PCM es la representación por medio de dígitos binarios de los valores muestreados de una señal analógica. Cuando el audio PCM, por nombrarlo de alguna manera, es transmitido cada “1” es representado por un pulso de tensión positiva y un “0” por la ausencia del pulso tal como se muestra en la figura 3.11

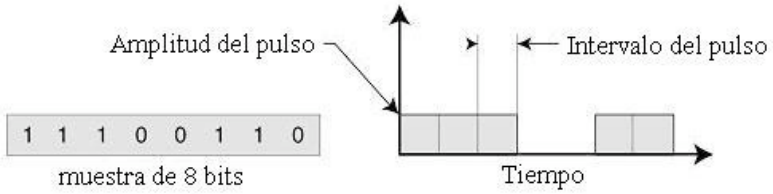


Figura 3.11

Visto lo anterior podemos concluir que el audio digital es una representación en bits de la señal analógica original, es importante recalcar que existen diferentes formas de representar dichas señales, esto es, la cantidad de muestras por segundo y el número de bits que realicen la digitalización de la señal, es ahí en donde se diferencian dos tipos de digitalización que podemos identificar como formas de compresión de audio digital.

3.6 Formatos de compresión con pérdidas. (Tipo loosy).

La forma de compresión del tipo *loosy* es aquella que a la hora de llevar a cabo el muestreo de la señal remueve ciertos rangos del sonido que un escucha promedio no puede oír. Para llevar a cabo dicha tarea, se utilizan algoritmos muy complejos basados en diferentes tipos de codificación y de modelos psicoacústicos. Con ellos se pueden eliminar dichos sonidos y por ende reducir el tamaño de archivo.

El problema de este formato de compresión es que, al remover sonidos y reducir el tamaño del archivo de audio digital, la frecuencia de muestreo es menor por lo cual se puede ver afectada la señal al grado tal que al compararlo con la original los cambios sufridos sean muy audibles y resulte en un audio comprimido pero de mala calidad.

El formato de compresión *lossy* más utilizado en la actualidad es el conocido como MP3 (capítulo IV), sin embargo existen diferentes formatos de este tipo de compresión, a continuación se enlistan algunos de ellos:

- Advanced Audio Coding **(AAC)** —También conocido como MPEG-4 AAC, este formato de audio es el que se utiliza en el software de Apple, iTunes, y en los reproductores de audio iPod. AAC ofrece una mejor calidad de audio que el MP3 con servicios extras tal como administración de derechos digitales de autor, con el fin de evitar las copias ilegales.
- ATRAC3 **(OMA, OMG)** —Sus iniciales OMG que significan *Open Magic Gate*, se deben a que es el administrador de derechos que Sony utiliza para su formato de audio digital ATRAC3 (*Adaptive Transform Acoustic Coding 3*, Codificador acústico de transformada adaptable) La desventaja de este formato es que solamente es compatible con los productos de Sony.
- Liquid Audio **(LAT, LQT, LSL)** —Se suponía fuese la competencia directa de MP3, se popularizo al final de los 90's, en la actualidad es poco utilizado.
- MP3 **(MP3)** —Ver capítulo IV.
- MP3PRO **(MP3)** —Es una versión mejorada del MP3 y salió al Mercado en el 2001.
- OGG Vorbis **(OGG)** —Con tecnología de codificación de fuente abierta conocida como *"Squish",* OGG Vorbis fue diseñado como sustituto del MP3 y del WMA. Utiliza una tasa de compresión de bits variable la cual codifica diferentes partes de una canción con mayor o menor compresión, esto con el fin de tener una mejor calidad.
- QuickTime Audio **(MOV)** —En esencia es la misma tecnología que utiliza el MPEG-4.
- RealAudio Media **(RA, RM, RMA)** —Formato que utiliza Real Networks, diseñado específicamente para uso en redes.
- Windows Media Audio **(WMA)** —Formato nativo de Microsoft cuya línea es ser una alternativa para los archivos MP3 con calidad de audio similar pero con una longitud de archivo menor, casi la mitad.

3.7 Formatos de compresión sin pérdidas. (Tipo loossless)

Dentro de este formato de compresión se garantiza que la calidad de audio sea igual a la que se reproduce de un CD. Para aplicaciones de alta fidelidad es el formato que se debe de utilizar ya que en el formato *loosy* no importa que tan alto sea el *bitrate,* siempre habrá perdidas en la compresión.

El tamaño de archivo que se obtiene utilizando este tipo de formatos es equivalente a 5 veces el tamaño de un archivo comprimido con un formato *loosy.*

Los formatos de compresión sin perdidas son:

- Apple Lossless Audio Codec **(ALAC, M4A)** —Una opción de compresión disponible para los usuarios de iPod.

- Free Lossless Codec **(FLAC)** —Formato *lossless* de fuente abierta, apoyado principalmente por consumidores de equipos electrónicos y que se puede utilizar con la mayoría de los sistemas operativos, incluyendo Windows y Linux.
- Monkey's Audio **(APE)** —Formato sin perdidas gratuito, casi no se utiliza.
- Windows Media Audio Lossless **(WMA)**, Formato de compresión *lossless* de Microsoft, disponible a partir de la versión 9 de su software Windows Media Player.
- WavPack **(WV, WVC)** — Formato lossless de fuente abierta, similar al FLAC.

3.8 Formatos sin compresión.

El tipo de archivo que se encuentra dentro de un CD de música es un formato sin compresión. Dentro de estos formatos los diferentes tipos son:

- **AU** —Formato de audio, abreviación de audio, que se utilizaba en los sistemas de las computadoras Sun y NeXT.
- **Audio Interchange Format (AIF, AIFF)** —Formato de archivo para sistemas Macintosh, similar al formato WAV de Windows.
- **Compact Disc Digital Audio (CDA)** —Este formato es utilizado para la codificación de música en todos los discos compactos. Este tipo de archivo es exclusivo de los discos compactos, las computadoras no pueden salvar archivos con ese tipo de extensión.
- **SND** —Formato de archivo, abreviación de *sound* o sonido, similar al AU y utilizado por el sistema operativo de las computadoras Macintosh.
- **Waveform Sound Files (WAV)** —Este formato (pronunciado como wave) produce una copia exacta de la señal original, sin compresión alguna. El resultado es una fidelidad perfecta pero con archivos muy grandes, de la misma longitud que el original. Es el formato preferido para archivar sin compresión.

Capítulo IV

ISO/IEC 11172-3

Introducción

En este capítulo abordaremos de manera clara la norma ISO/IEC 11172-3 para saber la forma en la cual se lleva a cabo la codificación de archivos de audio en el formato MP3. Cabe señalar que este capítulo es una adaptación de la norma publicada así como de los artículos que se especifican en la bibliografía.

4.1 Historia

En 1985, en Europa, se creó una fundación para la investigación y desarrollo con fines lucrativos llamada *EUREKA* [26][27] la cual formó parte de un equipo de trabajo conocido como EU-147. Dicho equipo desarrolló el sistema Eureka 147 DAB que, básicamente, funciona para transmisiones digitales de radio y fue diseñado específicamente para equipos con una recepción bastante robusta utilizando antenas simples no-direccionables.

A su vez un equipo formado por CCETT (Francia), IRT (Alemania) y Phillips (Holanda) desarrollaron el *MUSICAM* [19][25][26][27] *(Masking pattern adapted Universal Sub-band Integrated Coding and Multiplexing)*, por sus siglas en ingles, que es un algoritmo que fue utilizado en el sistema Eureka 147. El director de este proyecto fue Egon Meier-Engelen y fue terminado para 1987.

Casi al paralelo otro grupo formado por los AT&T Bell Labs, Thomson, Sociedad Fraunhofer y CNET propusieron un algoritmo para transmisión de audio a través de la Internet conocido como *ASPEC [26] (Adaptive Spectral Perceptual Entropy Coding).*

Ambos sistemas fueron objeto de pruebas de audición tras las cuales se encontró que el algoritmo *MUSICAM* presentaba un retardo en la codificación a diferencia del *ASPEC*. Pero el codificador *ASPEC* funcionaba mejor a tasas de bits menores. Entonces se creo el grupo llamado *MPEG.*

El *MPEG (Movie Pictures Expert Group),* formalmente conocido como ISO/IEC JTC1/SC29/WG11, fue creado en 1988 y trabaja bajo la dirección de la *ISO (International Standards Organization)* y la *IEC (International Electrotechnical Comisión).* El objetivo principal de este grupo es desarrollar estándares generales para la representación codificada de videos, audio y la interacción entre ambos. Desde entonces MPEG se ha hecho cargo de la estandarización de las técnicas de compresión para audio y video.

4.1.1 MPEG-1

MPEG-1 es el nombre de la primera fase del trabajo de *MPEG*. Esta fase terminó con la creación de la norma *ISO/IEC 11172* en 1992. La sección de esta norma (11172-3) dedicada a la codificación de audio describe un sistema de codificación genérico diseñado para abarcar la demanda de diversas aplicaciones. MPEG-1 Audio consiste de tres modos operativos llamados *Layers* (capas), las cuales incrementan la complejidad y desempeño, llamados *Layer-1, Layer-2 y Layer-3.* El Layer-3, de mayor complejidad, fue diseñado para proveer de la mayor calidad de audio a tasas de bits menores (alrededor de 128 kbit/s para una señal estéreo).

4.1.2 MPEG-2

MPEG-2 denota la segunda fase de MPEG. Introduce nuevos conceptos a la codificación de video, incluyendo soporte de señales de video entrelazadas. El área de mayor aplicación del MPEG-2 es la televisión digital. La norma MPEG-2 Audio (ISO/IEC 13818-3) fue terminada en 1994 y consiste en dos extensiones de MPEG-1 Audio:

*Codificación multicanal, incluyendo configuraciones para audio 5.1.
*Codificación a frecuencias de muestreo menores; 16, 22.05 y 24 [kHz]

4.1.3 MPEG-2 AAC

A principios de 1994, pruebas de verificación mostraron que nuevos algoritmos de codificación, sin compatibilidad con MPEG-1, presentaban mejorías significativas en la eficiencia de la codificación. Como resultado, se formó una nueva norma de codificación de audio conocida como *MPEG-2 AAC (Advanced Audio Coding)*. La norma fue terminada en 1997 (ISO/IEC 13818-7). La codificación avanzada de audio, AAC, es un esquema de codificación de segunda generación para codificación genérica de señales estéreo y multicanal, hasta 48 canales, con capacidad de soportar frecuencias de muestreo de 8[kHz] a 96[kHz].

4.1.4 MPEG-3

Originalmente *MPEG-3* estaba destinado para desarrollar las normas a utilizarse en televisión de alta definición o HDTV, por sus siglas en ingles. Sin embargo, se encontró que con las herramientas del MPEG-2 se pueden cubrir los requerimientos de dicha aplicación, por lo cual MPEG desarrolla otra norma para esta fase.

4.1.5 MPEG-4

La intención de *MPEG-4* es de ser el estándar mundial de multimedia. La primera versión fue terminada a finales de 1998 (ISO/IEC 14496-3), la segunda versión a finales de 1999. El énfasis de MPEG-4, a diferencia de MPEG-1 y MPEG-2, es añadir nuevas funcionalidades. Terminales fijas y móviles, acceso a bases de datos, comunicaciones y servicios interactivos son las aplicaciones a las que se enfoca MPEG-4.

MPEG-4 Audio consiste en una familia de algoritmos de codificación de audio que permiten tasas de bits muy bajas (menores a 2kbit/s) utilizadas en voz así como audio de alta calidad con codificaciones de hasta 64 kbit/s por canal.

4.1.6 MPEG-7

MPEG-7, llamado formalmente *Multimedia Content Description Interface*, es una representación estándar de la información audiovisual. Permite la descripción de contenidos por palabras clave y por significado semántico (quién, qué, cuándo, dónde) y estructural (formas, colores, texturas, movimientos, sonidos). El formato MPEG-7 se asocia de forma natural a los contenidos audiovisuales comprimidos por los codificadores MPEG-1, MPEG-2 y MPEG-4, pero se ha diseñado para que sea independiente del formato del contenido.

El nuevo estándar ayuda a las herramientas de indexación a crear grandes bases de material audiovisual (imágenes fijas, gráficos, modelos tridimensionales, audio, discursos, vídeo e información sobre la interacción entre ellos en una presentación multimedia) y buscar dentro de ellas manual o automáticamente.

4.2 Codificación.

4.2.1 Introducción

Para entrar de lleno al proceso de codificación del MP3 es necesario entender algunas funciones básicas, las cuales abordaremos a continuación.

4.2.1.1 Modos de operación

La sección MPEG-1 Audio trabaja tanto para señales monoaurales como señales estereo. Existe una técnica de codificación llamada conjunto estereo (*joint stereo*), que puede ser utilizada para lograr una mayor eficiencia en el proceso de codificación de un canal estereo ya que realiza el enmascaramiento simultáneamente en ambos canales, izquierdo y derecho. En general, los modos de operación del codificador son los siguientes:

*Canal mono
*Doble canal (dos canales independientes, por ejemplo cada uno conteniendo una versión diferente del lenguaje)
*Estereo
*Conjunto estereo

4.2.1.2 Frecuencia de muestreo

La compresión de audio de MPEG trabaja en diferentes frecuencias de muestreo. La fase 1 delimita la compresión de audio a 32, 44.1 y 48 [kHz]. MPEG-2 extiende este rango de frecuencias de 16, 22.05 y 24 [kHz], además de tener la capacidad de codificar la información de una señal surround 5.1.

4.2.1.3 Tasa de bits

MPEG Audio no solamente trabaja con una relación de compresión fija. La elección de la tasa de bits para la compresión de audio es, con algunas limitantes, dependiendo de quien implemente u opere un codificador MPEG. Para la capa 3, la norma establece un rango de tasa de bits desde 8 kbit/s hasta 320 kbit/s.

4.2.2 Análisis del codificador

La figura 4.1 muestra un diagrama de bloques del codificador MPEG-1. El algoritmo opera con bloques de datos. El bloque de audio a ser codificado, audio digital PCM, pasa a través de un banco de filtros (FB, por sus siglas en ingles) el cual divide la señal de entrada en múltiples sub-bandas. De forma paralela, el audio es sometido a un modelo psicoacústico (capítulo II) el cual determina la relación de la energía contenida en la señal con el umbral de enmascaramiento para cada sub –banda, en otras palabras el modelo psicoacústico determina la relación que existe entre la señal a enmascarar y la señal que llevará a cabo el enmascaramiento. Basado en el resultado del análisis psicoacústico y los bits disponibles, el bloque de cuantización asigna bits iterativamente a las diferentes sub-

bandas para minimizar el ruido de cuantización. Estas muestras cuantizadas de las sub-bandas son “empaquetadas” en una cadena de bits por medio de la codificación entropica. Cada bloque de datos, es representado como un frame en la cadena de bits de codificación.

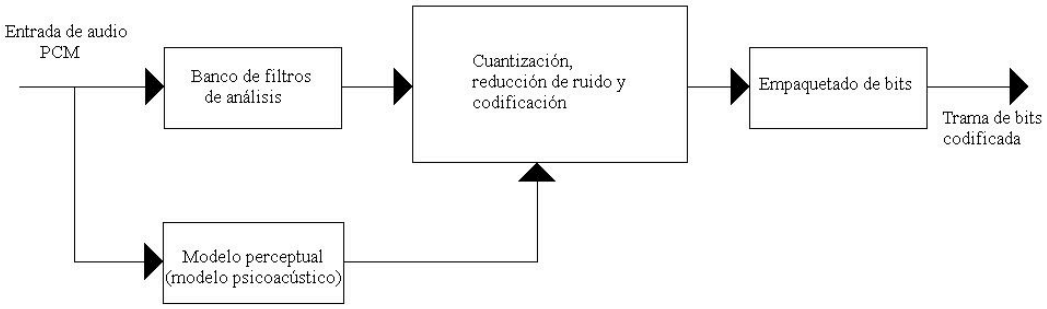


Figura 4.1

4.2.2.1 Banco de filtros de análisis

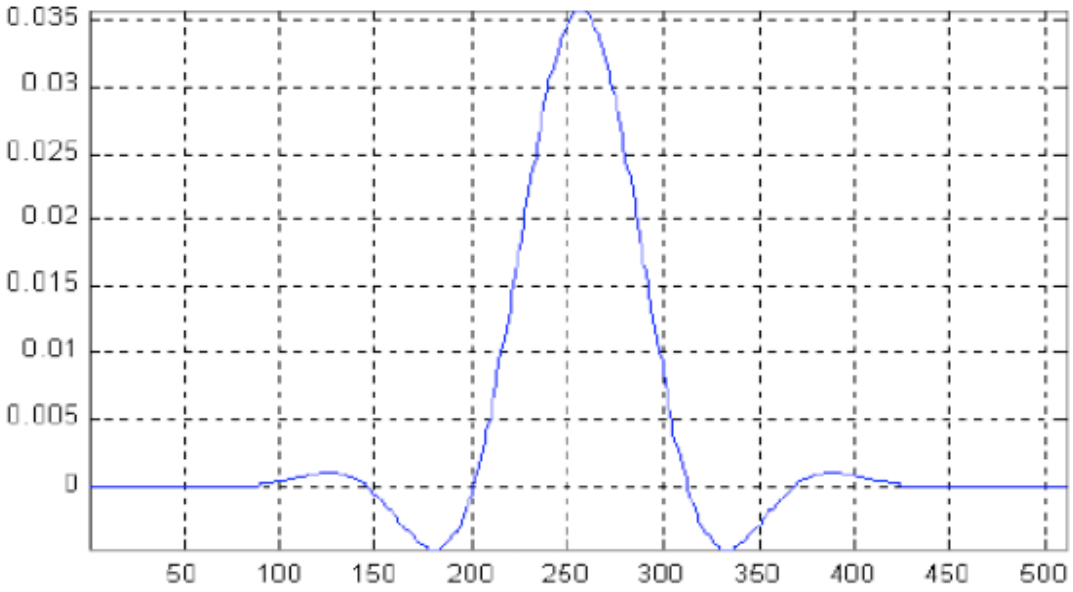


Figura 4.2

El banco de filtros de análisis es común a las tres capas del algoritmo. Este banco de filtros divide el bloque de audio en 32 bandas, cada una de ellas con ancho de banda nominal de $\pi/(32T)$, en donde *T* es el intervalo de muestreo. Los 512 coeficientes del prototipo del filtro paso-bajas se muestran en la figura 4.2. La respuesta al impulso del banco de filtros, figura 4.3, atenúa los lóbulos adyacentes en más de 96 [dB]. El filtro paso-bajas es transformado por medio de un coseno modulado para obtener un banco de

filtros, cuyas frecuencias centrales se encuentran localizadas en múltiplos impares de $\pi/(64T)$, figura 4.4

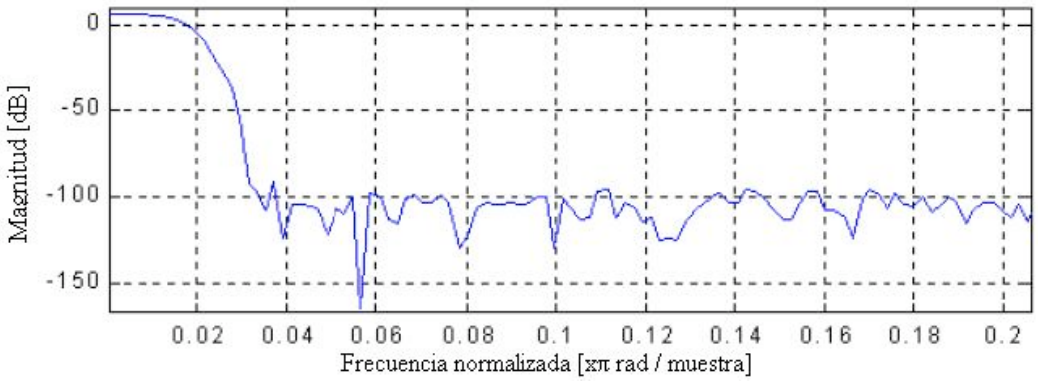


Figura 4.3

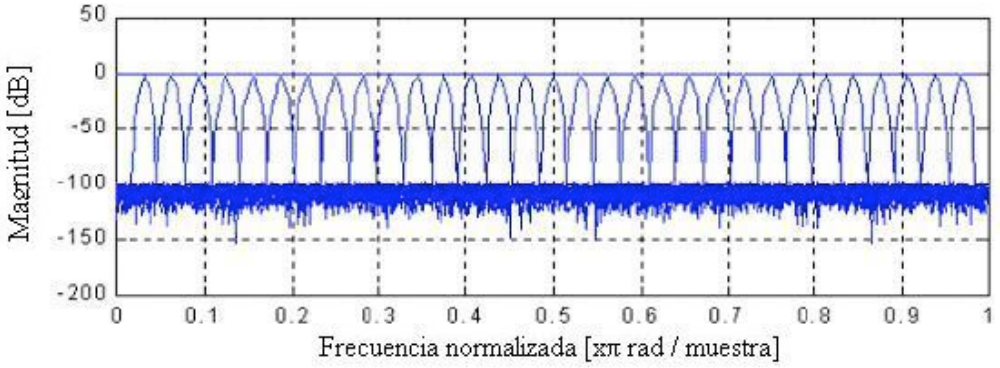


Figura 4.4

En resumen, el banco de filtros puede pensarse como un conjunto de 32 filtros paso-banda trabajando en paralelo, todos reciben la misma señal de entrada y todos entregan una porción del espectro a la salida.

Como es de esperarse todo este proceso nos arrojara un retraso a la señal, aproximadamente de 256 muestras, dicho retraso será compensado en la sección del modelo psicoacústico.

A pesar del retraso que presenta, la respuesta del banco de filtros es admirable, sin embargo, tiene 3 grandes defectos.

Primero, al ser un banco de filtros real, el corte de frecuencias no es lo suficientemente rápido para evitar que se vean afectadas las sub-bandas adyacentes, por lo cual nos puede llegar a generar un tono, consideremos el siguiente ejemplo, tenemos una señal de audio sintetizada hecha por medio de una combinación de dos señales senoidales a 675 y 11.100 [Hz]. A una frecuencia de muestreo de 44,1 [kHz], el ancho de banda nominal del prototipo del banco de filtros es, aproximadamente, de 689 [Hz]. Si la respuesta del banco de filtros fuera ideal, solamente la sub-banda 0 y la sub-banda 15 tendrían señal de salida. Debido a que existe superposición entre canales adyacentes, como se observa en

la figura 4.4, una porción de energía se difunde a las bandas cercanas. La respuesta se ilustra en la figura 4.5

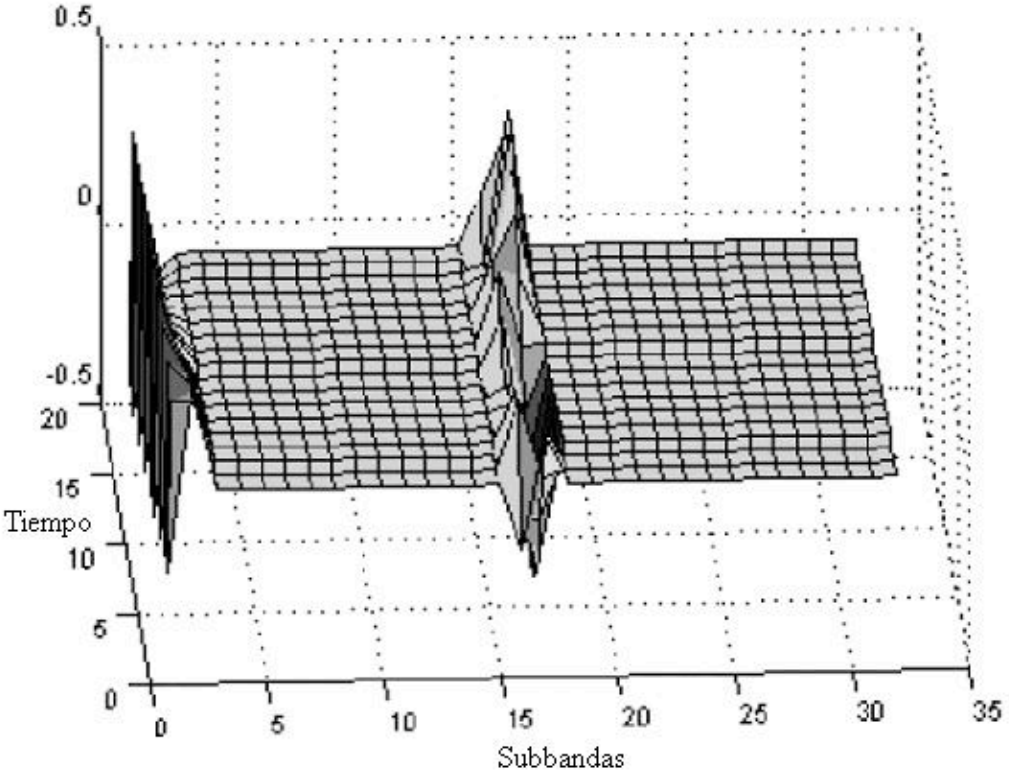


Figura 4.5

El segundo problema que se nos presenta en esta etapa se debe a la división del espectro frecuencial en sub-bandas con el mismo ancho de banda, como vimos en el capítulo 2 la membrana basilar no tiene una división frecuencial constante. Como resultado de esta diferencia tendremos que, a frecuencias bajas, una sola sub-banda contendrá muchas bandas criticas. Esto nos conlleva a que el cálculo de los umbrales de enmascaramiento se vuelve inexacto.

Tercero, el banco de filtros y su inverso (el banco de filtros sintetizado) son transformadas con perdidas por lo tanto, si ambos bancos de filtros no cuentan con un proceso de cuantización entre ellos la señal no será reconstruida exactamente. Sin embargo, el banco de filtros ha sido diseñado de tal manera que los errores sean imperceptibles al oído.

4.2.2.2 MDCT y banco de filtros híbrido

La salida del banco de filtros es procesada utilizando una Transformada Discreta del Coseno Modificada *MDCT,* por sus siglas en ingles, debido a las buenas propiedades que presenta en cuanto a energía se refiere. A diferencia del banco de filtros, la MDCT es una transformada sin perdidas.

La primera observación que podemos resaltar es que por cada muestra de entrada se tendrán 32 muestras a la salida del banco de filtro, por lo cual las operaciones a realizar se incrementan por 32 y la memoria necesaria para llevar a cabo estos cálculos se incrementa en el mismo número. Ahora bien, si pensamos en la salida de los filtros como un vector, en cada sub-banda solamente necesitaremos utilizar una muestra de cada uno de ellos ya que cada vector no es más que una ventana de las últimas 32 muestras de la señal original. A este tipo de muestras de la salida de los filtros se le conoce como sub-muestra.

Por especificaciones de la norma ISO/IEC 11172-3 el paquete de información a utilizarse debe de ser igual a 1152 muestras en el dominio del tiempo, ya que la MDCT transforma la salida del banco de filtros al dominio de la frecuencia y para cumplir con lo establecido por la norma agrupa en bloques de 36 sub-muestras de las sub-bandas. En el dominio de la frecuencia dichos bloques equivalen 18 muestras.

Existen dos tamaños de bloque definidos por la capa III. Para las porciones de señal que contengan transientes, es decir ataques, como por ejemplo una tarola se utilizara un conjunto de bloques cortos de 6 muestras para evitar problemas de *pre-echo* ya que los bloques cortos tienen una mejor resolución en el tiempo. En caso contrario, señales constantes como por ejemplo una guitarra rítmica, dado que la señal es constante, estadísticamente hablando, se puede utilizar un bloque largo de 18 muestras con lo cual tendremos una mejor resolución en el dominio de la frecuencia. La decisión del modo de bloque a utilizar recae sobre el modelo psicoacústico.

Cabe hacer la aclaración que al utilizar tres bloques cortos tendremos la misma longitud que un bloque largo, con lo cual aseguramos un cambio entre bloques suave.

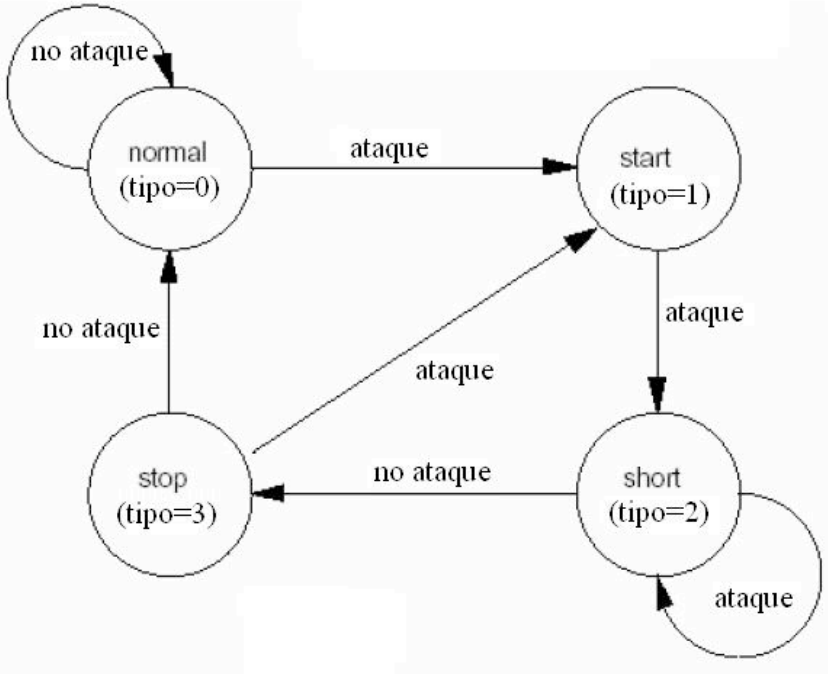


Figura 4.6

En la figura 4.7 encontramos la estructura del banco de filtros híbrido así como la MDCT.

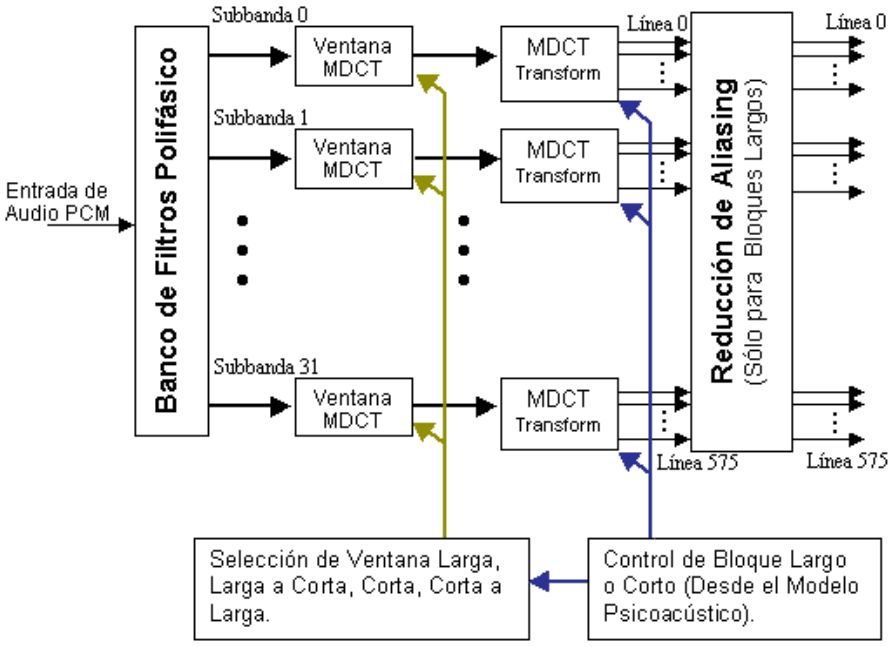


Figura 4.7

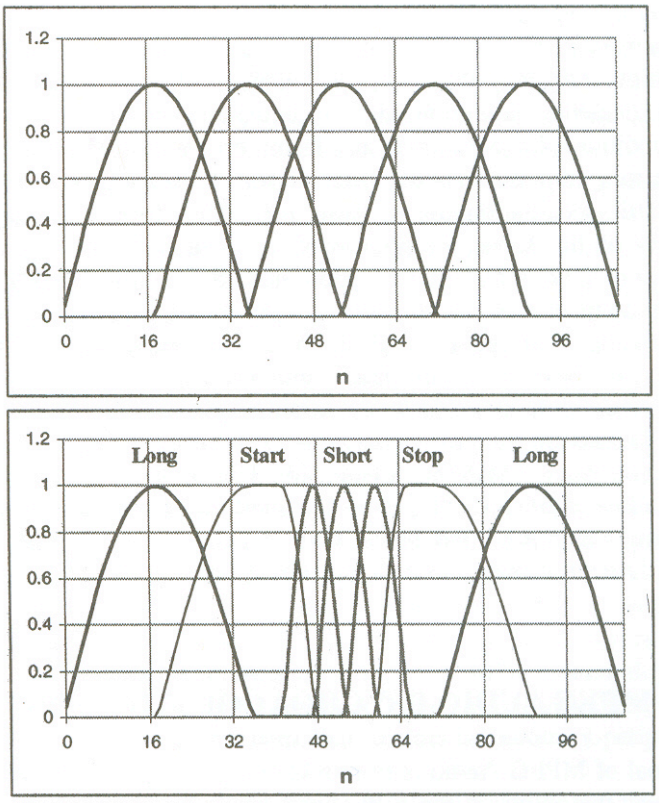


Figura 4.8

En la figura 4.8 podemos encontrar una secuencia de cambio de ventanas de larga a corta y de corta a larga así como la cantidad de traslape que existe entre cada una de ellas. La ventana básica, w(n), utilizada en la capa III es una ventana senoidal. Los tipos de ventanas quedan definidos de la siguiente manera.

$$w(n) = \sin\left(\frac{\pi}{36}\left(n+\frac{1}{2}\right)\right) \quad n = 0,\ldots, 35 \text{ (ventana larga)}$$

$$w(n) = \sin\left(\frac{\pi}{12}\left(n+\frac{1}{2}\right)\right) \quad n = 0,\ldots, 11 \text{(ventana corta)}$$

$$w(n) = \begin{cases} \sin\left(\frac{\pi}{36}\left(n+\frac{1}{2}\right)\right) \\ 1 \\ \sin\left(\frac{\pi}{12}\left(n-18+\frac{1}{2}\right)\right) \\ 0 \end{cases} \text{(ventana de inicio)}$$

$$w(n) = \begin{cases} 0 \\ \sin\left(\frac{\pi}{12}\left(n-6+\frac{1}{2}\right)\right) \\ 1 \\ \sin\left(\frac{\pi}{36}\left(n+\frac{1}{2}\right)\right) \end{cases} \text{(ventana de fin)}$$

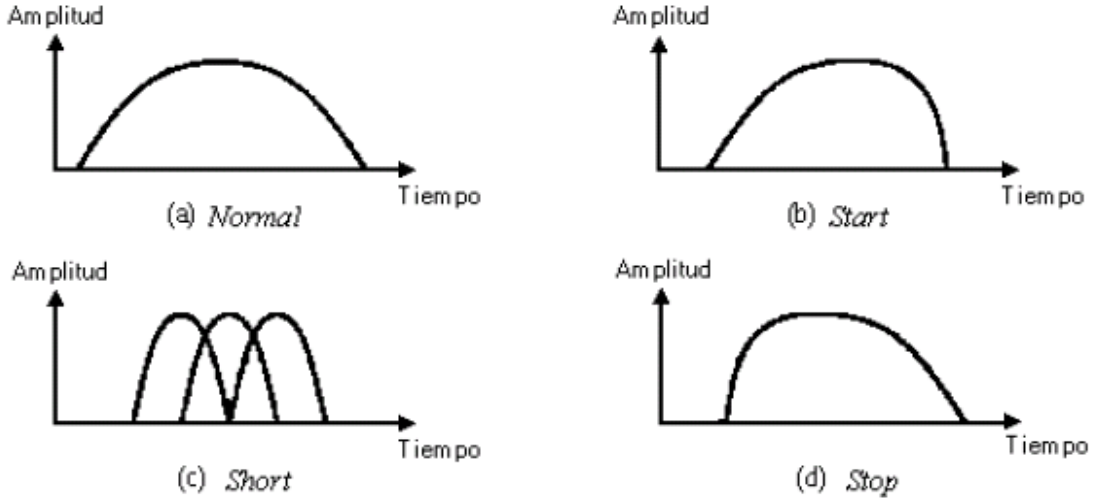


Figura 4.9
Tipo de ventanas

En la capa III se permite un modo de bloques mixto, en el cual las dos sub-bandas que se encuentren en el rango de las frecuencias mas bajas se procesan siempre con bloques largos, mientras que el resto de las 30 sub-bandas son procesadas con bloques pequeños. Con ello aseguramos una gran resolución frecuencial en los graves y mayor resolución en el tiempo para las frecuencias altas

4.2.2.3 Modelo psicoacústico

El modelo psicoacústico de un codificador MPEG-1 es un modelo matemático del comportamiento del oído humano y, como su nombre lo indica, tiene en cuenta los principios psicoacústicos de enmascaramiento frecuencial y temporal (capítulo 2).

Durante el proceso de codificación, la señal de entrada es analizada paquete (*frame)* por paquete y se determina que porción de la señal puede ser enmascarada. Por cada paquete, basados en los umbrales de enmascaramiento, se distribuyen los bits disponibles para lograr la mejor representación de la señal original. Existen dos tipos de modelos psicoacústicos el modelo I y el modelo II, para la norma 11172 capa 3 se utiliza el modelo II. El modelo psicoacústico lleva a cabo esta tarea de la siguiente manera.

Primero el modelo convierte el audio al dominio de la frecuencia, usando una FFT de 1024 puntos para conseguir una buena resolución 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. El modelo calcula el mínimo umbral de enmascaramiento para cada subbanda; dichos valores se usan luego para calcular la distorsión permitida, también conocida como *JND Just Noticeable Distorsion.* A continuación se muestran los pasos generales para el cálculo psicoacústico de la señal.

1) Alineación en el 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. Es importante puntualizar que los datos sufren un retraso después de ser procesados por el banco de filtros.

2) Representación espectral. El modelo lleva a cabo una conversión del tiempo a la frecuencia totalmente independiente del mapeo del banco de filtros ya que necesita una mejor resolución en frecuencia para calcular los umbrales de enmascaramiento. Esto se lleva a cabo por medio de una transformada de Fourier. El modelo II usa una FFT de 1024 puntos y se ejecutan dos cálculos. El primero se encarga de las 576 muestras iniciales y el segundo cálculo se realiza sobre las últimas 576 muestras. A esta división de muestras se le conoce como granulo. El modelo combina los resultados de ambos cálculos, de tal manera que el resultado total implique la selección del umbral de enmascaramiento (*Noise Masking Threshold, NMT*) más bajo en cada subbanda. Para simplificar los cálculos los valores espectrales se procesan en barks.

3) Componentes tonales y no tonales. El modelo identifica y separa 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.
   Aunque el modelo II nunca separa las componentes tonales ni las no tonales, éste calcula un índice de tonalidad en función de la frecuencia, el cual mide el comportamiento que presenta cada tipo de componente. Este valor es utilizado

para interpolar entre valores puros TMN (*Tone Masking Noise*) y valores puros NMT (*Noise Masking Tone*). El índice de tonalidad es en realidad una medida anticipada, también conocida como "predictability measure". El modelo utiliza datos de los dos cálculos anteriores para predecir, a través de una extrapolación lineal, 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.

4) Función de dispersión. La capacidad enmascarante de una componente determinada se distribuye por toda la banda crítica que la rodea. El modelo determina el umbral de enmascaramiento de ruido aplicando una función de dispersión.

$$B(dz) = 15.8111389 + 7.5 * (1.05 * dz + 0.474) - 17.5 * \sqrt{1.0 + (1.05 * dz + 0.474)^2}$$
$$+8 * MIN\left(0, (1.05 * dz - 0.5)^2 - 2 * (1.05 * dz - 0.5)\right)$$

en donde dz es la distancia medida en barks entre la señal a enmascarar y la máscara.

5) Umbral de enmascaramiento global. El umbral de enmascaramiento absoluto es un valor determinado empíricamente: el mínimo umbral auditivo en un ambiente silencioso, lo cual se puede definir como la intensidad del sonido más débil que se puede escuchar cuando no hay mas sonidos presentes. En el modelo II este enmascaramiento global no es calculado, sino que se trabaja todos los datos dentro de cada subbanda, de acuerdo con el índice de tonalidad que tenga cada componente enmascarante en esa subbanda.

6) Umbral de enmascaramiento mínimo. Se selecciona el mínimo de todos los umbrales de enmascaramiento en cada subbanda sólo para regiones de frecuencia donde el ancho de banda 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 de esa subbanda.

7) Relaciones señal a máscara. El modelo calcula la relación señal a máscara, SMR, como la relación entre la energía de la señal en la subbanda o grupo de subbandas y el mínimo umbral de enmascaramiento para esa subbanda. El valor que se entrega a la etapa de cuantización es el JND, el cual determina cuál es la cantidad máxima de ruido de cuantización que se permite.

4.2.2.4 Cuantización (noise/bit allocation)

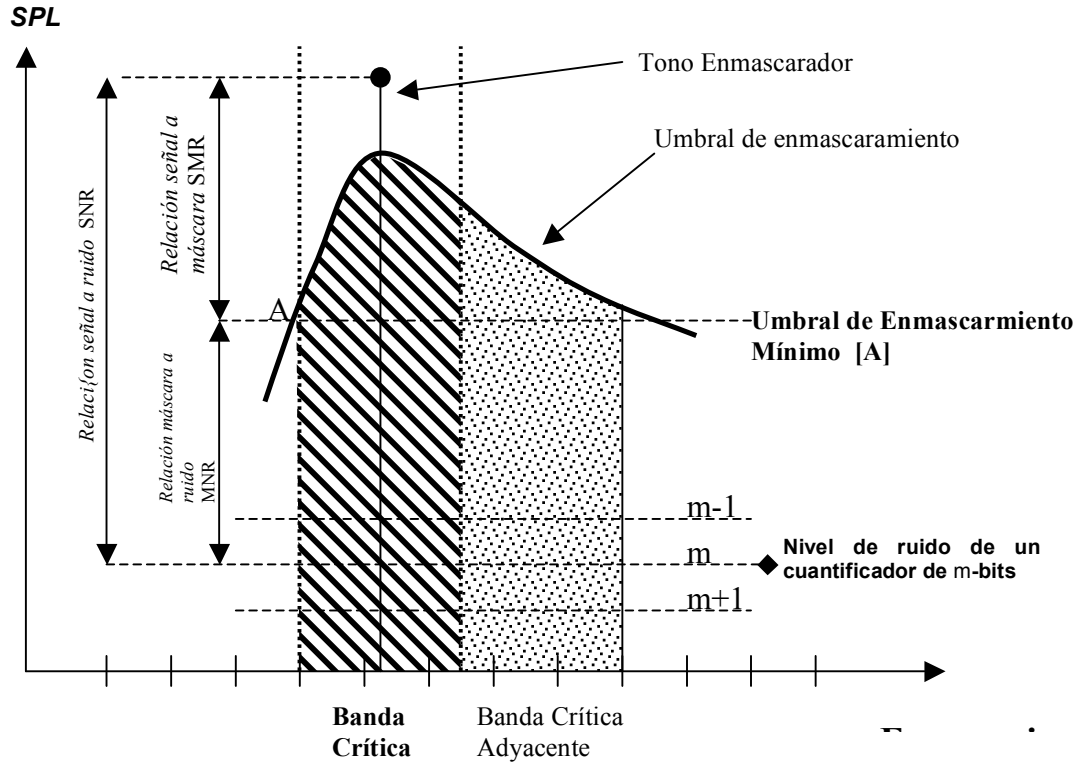


Figura 4.10

Considere el tono de la figura 4.10 El enmascaramiento producido por este tono se determina simulando el patrón de excitación de la membrana bacilar, ignorando la desviación y comparando con el umbral en cuestión. Asumiendo que la *máscara* se cuantiza utilizando una escala de *m*-bits, el ruido se considerara que tiene un nivel *m*.

La relación señal-máscara *(SMR)* y máscara-ruido *(MNR)* representan las distancias entre el umbral de enmascaramiento mínimo y la máscara y el nivel de ruido, respectivamente, en el dominio logarítmico.

La salida definitiva del bloque (la cantidad de bits de cuantización por banda) se expresa en términos de una relación denominada *Mask to Noise Ratio* (relación enmascaramiento/ruido) calculada de la siguiente forma:

$$MNR_{dB} = SNR_{dB} - SMR_{dB}$$

donde

MNR_{dB} : *Mask to Noise Ratio*

SNR_{dB} : *Signal to Noise Ratio*

SMR_{dB} : *Signal to Mask Ratio* (del modelo psicoacústico)

Este bloque cuantifica las salidas de los filtros de acuerdo a la cantidad de bits disponibles para cada una de las bandas; información suministrada por el modelo psicoacústico.

La Capa III cuantifica los valores espectrales asignando los bits correctos en cada sub-banda para mantener una transparencia perceptual para la tasa de bit dada. Esta capa controla la forma del espectro de ruido de cuantización para que quede por debajo de los niveles audibles. Este esquema es llamado asignación o reparto de ruido, *noise allocation*.

La repartición de bits únicamente aproxima la cantidad de ruido causado por la cuantización, mientras la repartición de ruido verdaderamente calcula el ruido. La repartición se hace en un ciclo de iteración que consiste de un ciclo interno y uno externo.

1) Ciclo interno. El ciclo interno realiza la cuantización no-uniforme de acuerdo con el sistema de punto flotante verdadero (cada valor espectral MDCT se eleva a la potencia 3/4). El ciclo escoge un determinado intervalo de cuantización (quantization step size), y a los datos cuantizados se les aplica codificación de Huffman. Si al realizar este proceso se encuentra que el número de bits requerido para codificar los valores excede la cantidad de bits disponibles, de acuerdo con la tasa de bits escogida, 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.

2) Ciclo externo. Ahora el ciclo externo se encarga de verificar si el factor de escala para cada subbanda tiene más distorsión de 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 del factor de escala 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 mucho ruido.
- La próxima iteración amplificaría una de las bandas más de lo 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.

3) Codificación de Huffman. El MP3 también emplea la clásica técnica del algoritmo de Huffman. Actúa al final de la compresión para codificar la información; por lo tanto, no es un algoritmo de compresión, sino más bien un método de codificación. Esta técnica crea códigos de longitud variable sobre un número total de bits, donde los símbolos con más alta probabilidad tienen códigos más cortos. Los códigos de Huffman tienen la propiedad de poseer un único prefijo y por lo tanto, pueden ser decodificados correctamente a pesar de su longitud variable; el proceso de la decodificación es muy rápido, a través de una tabla de correspondencias. Este tipo de codificación permite ahorrar, en promedio, aproximadamente un 20% en espacio de almacenamiento.

Las técnicas que se han mostrado son el complemento ideal para la codificación psicoacústica: durante gran polifonía, muchos sonidos están enmascarados o disminuidos, logrando que la codificación psicoacústica sea muy eficiente; y debido a que hay poca información idéntica, entonces el algoritmo de Huffman presenta poca eficiencia. Pero durante los sonidos "puros" hay muy pocos efectos de enmascaramiento, y es aquí donde la codificación de Huffman se vuelve muy eficiente debido a que los sonidos puros, cuando se digitalizan, contienen gran cantidad de bytes redundantes, que entonces serán reemplazados por códigos más cortos.

4.3 Trama

El último paso en el proceso de codificación es producir un flujo de bits MP3 válido. Este bloque es el encargado de almacenar el audio codificado y algunos datos adicionales en tramas, cada trama en la capa III contiene información de 1152 muestras. Una trama consiste de encabezado y datos de audio junto con la revisión de errores y los datos auxiliares, estos dos últimos 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. En el caso de la Capa III, las tramas no son totalmente independientes siempre: debido al posible uso del bit reservoir, que es una especie de búfer, las tramas son a menudo dependientes unas de otras.

La estructura de la trama del mp3 se puede observar en la figura 4.11

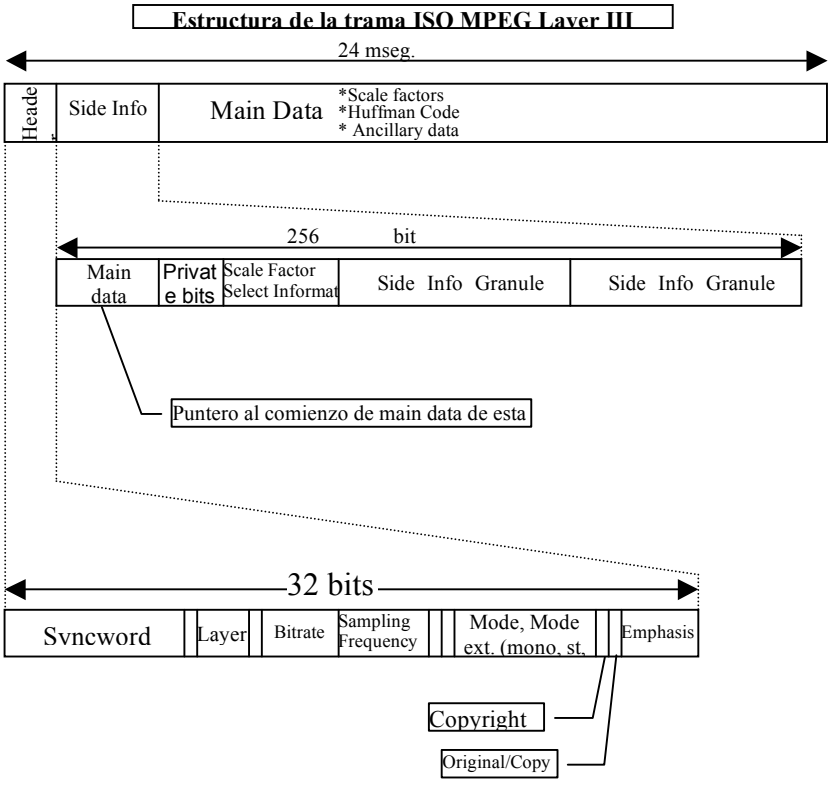


Figura 4.11

La trama esta compuesta por una cabecera (*Header*) que es común para las 3 capas y que contiene 32 bits de información la cual se subdivide en:

-**Syncword**. Palabra de sincronía que se compone de 12 bits.
-**ID**: Un bit usado para identificación del audio. Siempre en '1', para indicar que se trata de audio MPEG-1.
-**Layer**. De acuerdo al contenido de esta información se determina el tipo de capa (2 bits) que fue usado durante la codificación de audio.

00	Reservado
01	Capa III
10	Capa II
11	Capa I

-**Protection Bit**. Es un bit de protección que indica si se ha introducido redundancia en el flujo de información para facilitar la detección y/o cancelación de errores.
-**Bitrate**. Cuatro bits para proporcionar el índice de la tasa de bits, de acuerdo con la siguiente tabla.

Código	Tasa de bits MPEG-1 (Kbps)		
	Capa I	Capa II	Capa III
0000	Formato libre	Formato libre	Formato libre
0001	32	32	32
0010	64	48	40
0011	96	56	48
0100	128	64	56
0101	160	80	64
0110	192	96	80
0111	224	112	96
1000	256	128	112
1001	288	160	128
1010	320	192	160
1011	352	224	192
1100	384	256	224
1101	416	320	256
1110	448	384	320
1111	No permitido	No permitido	No permitido
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 (320 Kbps para la Capa III).			

-**sampling_frequency**. Dos bits que indican la tasa de muestreo.

00	44.1 KHz
01	48 KHz
10	32 KHz
11	Reservado

-**padding_bit**. Un bit usado para relleno. Si está en '1' la trama se rellena con una ranura extra. Únicamente se usa para frecuencias de 44.1 [kHz]. Por ejemplo, un sonido 128 kbps 44.1 [kHz] Capa II usa muchas tramas de 418 bytes de largo y unas pocas de 417 bytes para cumplir exactamente la tasa de transferencia de 128 kbps. La ranura consume 8 bits (1 byte) para las Capas II y III.
-**private_bit**. Un bit para uso privado. No se usa generalmente.
-**mode**. Dos bits que indican el modo de canal, tal y como se muestra a continuación.

00	*Stereo*
01	*Joint Stereo*
10	*Dual Channel* (2 canales monofónicos independientes)
11	*Single Channel* (1 canal monofónico)

En 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 mono 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. Single Channel consiste en un único canal de audio.
-**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 III, estos dos bits indican qué tipo de codificación Joint Stereo se está usando, Intensidad estéreo o Estéreo M/S. 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 fórmula:

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

Mi = Middle; Si = Side; Li = Izquierdo; Ri = Derecho
Los valores Mi se transmiten por el canal izquierdo y los valores Si se transmiten por el canal derecho.
En el lado del decodificador los canales izquierdo y derecho se reconstruyen así:

$$L_i = \frac{M_i + S_i}{\sqrt{2}} \quad y \quad R_i = \frac{M_i - S_i}{\sqrt{2}}$$

Intensidad estéreo 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 de la siguiente manera:

Código	Intensidad estéreo	M/S estereo
00	No	No
01	Sí	No
10	No	Sí
11	Sí	Sî

-**copyrigth**. Un bit usado para derechos de autor. Tiene el mismo significado que el bit de copyright en CD y cintas DAT, indica que es ilegal copiar el contenido del archivo si el bit está en '1'.
-**original/copy**. Un bit usado para indicar si se trata de un medio original '1' o si es una copia '0'
-**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. Se usa raramente.

00	Ninguna
01	50/15 ms
10	Reservado
11	CCITT J.17

Posterior a la cabecera se encuentra la información lateral (*side info*) la cual 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 (scfsi y private_bits), información secundaria para el gránulo 0, e información secundaria del gránulo 1.

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

-**main_data_begin**. El campo main_data 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**. La variable 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 '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 '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:

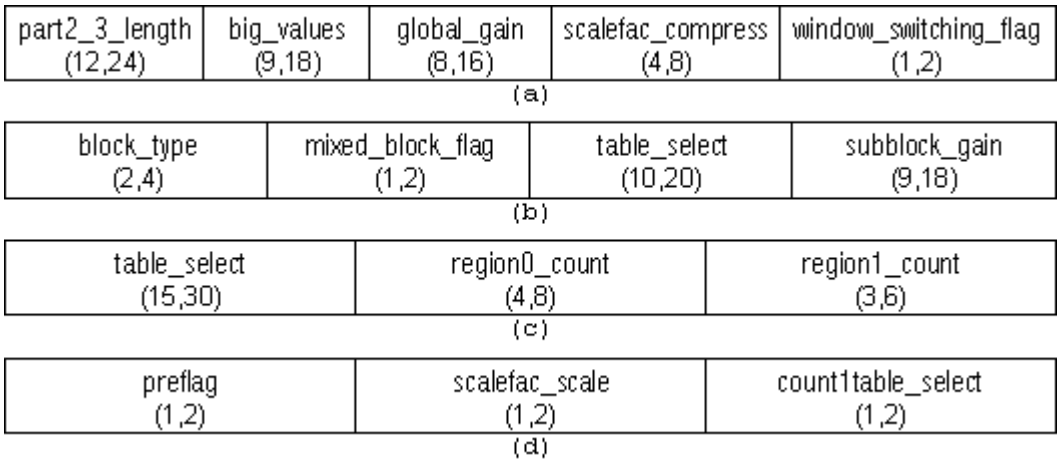


Figura 4.12
Información secundaria para cada gránulo.

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 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 partición de la mitad, 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 directa (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 (9) bits se usan para big_values en modo mono y 18 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 cuatro bits (en modo mono), 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 11172-3, 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 '1', se añaden los bits de la figura 4.12 (b); si está en '0', se añaden los bits de la figura 4.12 (c).
-**table_select**. Habilita el uso de 32 diferentes tablas para el código de Huffman, 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 region0 y region1. Esta variable de 4 bits (en modo mono) especifica la cantidad de bandas del factor de escala incluidas en esta región, pero disminuidas en 1.

region0_count = bandas del factor de escala en region0 – 1

-**region1_count**. Especifica el límite entre 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 = bandas del factor de escala en region1 – 1

-**preflag**. Un bit por canal, indicando que se usó preénfasis (o sea, amplificación adicional en las altas frecuencias). Este valor apunta a una tabla en el estándar ISO 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 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 4.12 (c), las cuales son reemplazadas por aquellas de la figura 4.12 (b). Las otras variables mostradas en la figura 4.12 no cambian.

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

-**mixed_block_flag**. Esta variable, que consume 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 '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 subbloque particular. Esta variable usa 3 bits por canal.

Por último encontramos los datos principales o *main data* los cuales se dividen en las siguientes secciones.

Factores de escala	Código de Huffman	Datos auxiliares
Variable length	*Variable length*	*Variable length*

-**Scale factors**. Factores de escala que se usan para colorear el ruido 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.

-**Huffman Code**. 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 (0). La partición count1 contiene cuádruplos de valores iguales a -1, 0 ó 1, que se codifican usando una de 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 ('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, como se muestra en la siguiente fórmula:

$$linbits = \log_2(máximo\ valor\ cuantizado - 14) \Rightarrow \texttt{Se redondea por exceso}$$

De nuevo, para cada valor diferente de cero, se agrega un bit de signo ('0' si es positivo). Por ejemplo: Asúmase, primero que la tabla de Huffman ya ha sido seleccionada, y también:

Par de valores cuantizados (x,y) = (0,15)
Máximo valor cuantizado de la región = 1039
Código de Huffman para el par (0,15) = '01101'
Valor adicional para 'y' = linbitsy = 15-15 = 0

$$linbits = \log_2(1039 - 14) \cong 10{,}0014 \Rightarrow linbits = 11$$

$$linbitsy = 15 - 15 = 0 = '00000000000'$$

$$Codificación\ del\ par\ (0{,}15) = Codificación\ del\ par\ (0{,}15) + linbitsy$$

$$Codificación\ del\ par\ (0{,}15) = '01101'\ '00000000000'$$

$$bits\ necesarios\ para\ codificar\ el\ par\ (0{,}15) = 16\ bits$$

(x,y)	linbitsx	linbitsy	signx	signy
5 bits	0 bits	11 bits	0 bits	1 bit
Flujo de bits '01101' '00000000000' '0'				

En el caso de que 'x' también sea mayor que 14, se debe buscar el código de Huffman para el par (15,15), y además también se debe codificar un valor adicional llamado "linbitsx", que indica la diferencia entre 15 (máximo valor de las tablas) y el valor verdadero de 'x'. Adicionalmente, por cada valor diferente de cero se debe agregar un bit de signo ('0' si es positivo, '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 (0).

-**Datos auxiliares**. Éstos son opcionales, y la cantidad de bits repartidos para este campo, se define por el usuario.

4.4 Decodificación

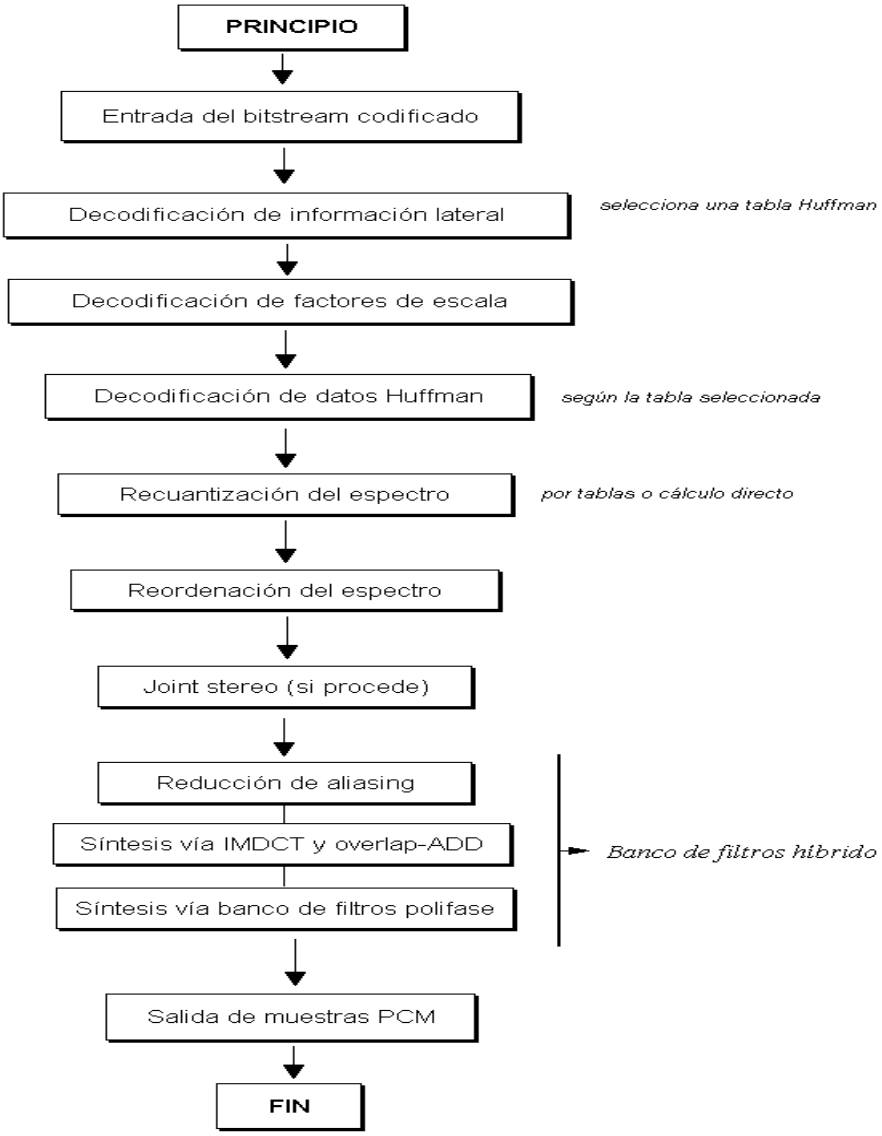


Figura 4.13

Como se observa en la figura 4.13 la primer tarea que lleva a cabo el decodificador es la de localizar la palabra de sincronización que marca el inicio de una trama de audio valido, esta palabra de sincronización está dictaminada por el cambio de ventana en el codificador y viene dentro del cabezal del flujo de bit, que además proporciona la información referente a la frecuencia de muestreo y la configuración de canales, esto es, si la canción viene en formato estereo, monoaural o estereo enlazado (*joint stereo*). Encontraremos en la cabecera la información referente a la *tasa de bits* que fue utilizada, con lo que el decodificador

identificará la longitud de la trama, cuando inicia y cuando termina. Cada trama contiene 27 [ms] de información musical.

La información requerida por el decodificador que no pertenece al audio codificado se le conoce como información lateral. Existe un bloque de dicha información por cada canal. Este bloque contiene información de los parámetros referentes a la reconstrucción de la señal. Selecciona la tabla de Huffman correspondiente y divide el espectro en el dominio de la frecuencia en bandas, que son determinadas por la frecuencia de muestreo y su correspondencia mutua con las bandas criticas del oído humano. Para cada banda existe un factor que se utilizara en el control de ganancia durante la decuantización.

La decodificación de los factores de escala nos proporcionan los valores necesarios para ingresar a una tabla en la norma ISO/IEC 11172-3 que a su vez nos indica el número de bits utilizados por los factores.

En la decodificación de los datos de Huffman, se escoge la tabla correspondiente para decodificar los *big_values*. La decodificación se realiza hasta que todos los códigos Huffman han sido decodificados o hasta que los valores cuantificados que representen 576 líneas (granulo) han sido decodificados.

En el siguiente paso, las muestras de la cadena de bits son decuantizadas y escaladas a los valores apropiados utilizando el control de ganancia.

Como vimos en el proceso de codificación tendremos bloques largos y cortos. Los bloques cortos deben ser reordenados para poder continuar con el proceso. En el caso de los bloques largos se aplica una cancelación del traslape para compensar errores que se hayan presentado durante el banco de filtros.

Posteriormente cada sub-banda se transforma al dominio del tiempo, esto se lleva a cabo gracias a la *IMDCT* (Transformada Inversa Discreta del Coseno Modificado). Para los bloques largos una *imdct* de 36 puntos calcula las 36 muestras de salida de manera directa. En el caso de los bloques cortos, la salida de 3 *imdcts* de 12 puntos se combina para tener 36 muestras a la salida.

Finalmente las 32 sub-bandas entran a un filtro de síntesis en el dominio del tiempo para cubrir todo el espectro frecuencial. Se obtienen 1152 muestras de audio por trama. Las muestras de audio PCM son calculadas por un arreglo de ventanas.

Capítulo V

ANÁLISIS Y EVOLUCIÓN DEL MP3

Introducción

Observamos en los capítulos anteriores que el objetivo de la compresión de audio es que se obtenga la misma calidad auditiva del CD pero con un archivo de información mas pequeño y que esto se logra por medio de los codificadores que ha desarrollado el grupo *MPEG*, sin embargo, hasta ahora no hemos visto de que manera se dictamina si un codificador es de buena calidad o no y en que basan dichos resultados.

Para certificar la calidad de los codificadores diseñados por el grupo *MPEG* se pueden llevar a cabo dos tipos diferentes de pruebas: objetivas y subjetivas. Mientras que el primer tipo de pruebas pueden medir la calidad del audio tal como la relación señal a ruido o el total de distorsión de la misma, estas no miden el efecto de la psicoacústica por lo que el resultado de la codificación puede distar mucho de la señal original; en cambio las pruebas del tipo subjetivo se basan en el criterio del escucha, que finalmente es quien decide si un archivo codificado es de buena o mala calidad. En nuestro caso utilizaremos las pruebas de tipo subjetivo.

Existen diferentes pruebas para medir la calidad del codificador tales como: *Doble ciego triple estímulo con referencia escondida* o *el método MUSHRA (Multiple Stimulus with Hidden Reference and Anchors)*, introducidas por la división de radiocomunicación de la *ITU (International Telecommunications Union)*. Para efectos prácticos utilizaremos el método doble ciego triple estímulo con referencia escondida, que es la prueba a la que se someten los codificadores que se utilizan en las normas del grupo *MPEG* y que se encuentra abalada por la *ITU* en su recomendación *ITU-R BS.1116-1* [32].

En este capítulo haremos una descripción completa del método y las condiciones en las cuales se debe llevar a cabo la prueba para evaluar la codificación MP3 en sus diferentes tasas de compresión, posteriormente obtendremos los resultados de la misma.

5.1 Mediciones de calidad de los codecs de audio

5.1.1 Escala de los 5 grados de diferencia

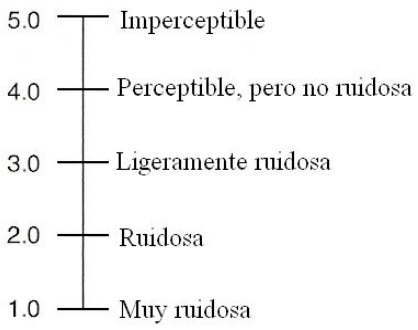


Figura 5.1

La escala de 5 grados a utilizar en las pruebas, tabla 1, esta basada en la escala de diferencias que se muestra en la figura 5.1. De acuerdo a la *ITU,* cualquier diferencia

percibida entre la señal de referencia y la señal a prueba debe ser interpretada como una disparidad y con la ayuda de la escala se puede medir dicha diferencia. Dicha escala está dividida por factores de una unidad, siendo 5.0 el equivalente a una codificación transparente y 1.0 una codificación bastante ruidosa.

Calidad	Diferencias
5 Excelente	5 Imperceptible
4 Buena	4 Perceptible, pero no ruidosa
3 Media	3 Ligeramente ruidosa
2 Pobre	2 Ruidosa
1 Mala	1 Muy ruidosa

Tabla 1. Relación entre la calidad y la escala de diferencias.

Para facilitar el manejo de la información el grado de diferencia que percibe el escucha entre la referencia y la señal codificada es cuantificado. El valor que se obtiene, llamado grado de diferencia subjetiva, *SDG* por sus siglas en ingles, esta definido por:

$$SDG = Grado_{\text{señal codificada}} - Grado_{\text{señal de referencia}}$$

EL *SDG* presenta un valor negativo cuando el escucha distingue satisfactoriamente la señal de referencia de la señal codificada, y un valor positivo cuando el escucha selecciona erróneamente la señal codificada como la de referencia. Un valor de cero indica que la codificación se encuentra en la zona de transparencia por lo cual las pequeñas diferencias que pudiera mostrar la señal codificada con la original son imperceptibles al oído humano; por otro lado un valor equivalente a –4 indica que la codificación es muy ruidosa. En la tabla 2 podemos encontrar la relación entre el grado de diferencia y el valor del *SDG*.

Descripción de la diferencia	Grado	SDG
Imperceptible	5	0.0
Perceptible, pero no ruidosa	4	-1.0
Ligeramente ruidosa	3	-2.0
Ruidosa	2	-3.0
Muy ruidosa	1	-4.0

Tabla 2.

5.1.2 Método de prueba doble ciego triple estímulo referencia escondida

Este método es el más utilizado para probar sistemas con diferencias pequeñas. En este método se le proporcionan al escucha tres señales (estímulos): la señal de referencia, R, y las señales a prueba identificadas como A y B. Una de las dos señales de prueba será idéntica a la señal de referencia y en la otra tendremos la señal codificada. La prueba se lleva a cabo a “doble ciego” ya que el escucha y el administrador de la prueba no deben de saber cual de las señales corresponde a que letra. La asignación de las señales A y B debe llevarse a cabo por alguna persona o entidad ajena tanto al administrador de la prueba como al escucha.

Se le pide al escucha que valore las diferencias que encuentra comparando la señal A con la R, y las diferencias de B con R basándose en la escala de la figura 1. Dado que una de las dos señales de prueba es la referencia, una de ellas debiera de recibir un grado igual a cinco, mientras que el otro estímulo debe recibir un grado que describa la valoración de diferencias que percibe el escucha. Si el sistema a prueba nos arroja una

calidad de codificación que se encuentre dentro de la zona de transparencia, quiere decir que el escucha no nota ninguna diferencia entre la señal de referencia y la codificada. En este caso, debe considerarse la posibilidad de variar el factor de compresión del sistema para estimar el margen de codificación del mismo.

El método anterior ha sido implementado de diferentes maneras. Por ejemplo, el sistema a prueba puede ser una implementación en tiempo real de un hardware o una simulación por software. Se recomienda que solamente se someta a un sujeto a la prueba a la vez. El escucha debe tener la facilidad de poder cambiar entre la señal R, A y B así como repetir la sección de prueba o canción, tantas veces como lo crea necesario.

El incluir la referencia escondida dentro de la prueba es una manera sencilla de revisar que el escucha no cometa errores constantemente y provee, de alguna manera, un control de calidad.

5.1.3 Selección de panel de escuchas

Antes de llevar a cabo la sesión de entrenamiento se necesita hacer una selección de personas las cuales serán sometidas a la prueba. Dicha selección es necesaria ya que al ser una prueba que evalúa pequeñas diferencias auditivas existirán sujetos los cuales no podrán distinguir entre una señal y otra cuando la señal a prueba tenga una mejor calidad de compresión.

La recomendación *ITU-R BS.1116-1* indica que para cada prueba en donde exista un cambio en la señal a evaluar se necesita que por lo menos 1.5 expertos en el tema realicen dicha prueba, con un mínimo de 5 expertos.

5.1.4 Sesiones de entrenamiento para el escucha

Una prueba de escucha generalmente consiste de dos partes: una fase de entrenamiento y una fase de graduación. La fase de entrenamiento consiste en exponer a los escuchas todo lo relacionado con la prueba, es decir, que se familiaricen con el lugar en donde se llevará a cabo la prueba, el proceso para calificar la prueba y las diferencias entre los rangos de compresión que se utilizarán. Es imperativo que el grupo de escuchas se familiarice con el equipo a utilizarse durante las pruebas así como poseer de antemano un ejemplo de lo que se espera, este ejemplo puede ser reproduciendo una señal de un CD y otra señal con mucha compresión para que ellos hagan hincapié en las diferencias que presentan ambas señales y en base a ellas, busquen esas diferencias cuando se este realizando la prueba.

Para la fase de graduación también es importante llevar a cabo un ejemplo de puntuación de las diferentes señales, dejando en claro las diferencias que se estén evaluando.

5.1.5 Condiciones de escucha

En base a las condiciones establecidas por la *ITU* se deben describir claramente las condiciones de escucha y el equipo a utilizar, en las condiciones de escucha deben incluirse las características acústicas del cuarto, así como las características de las bocinas y posición de las mismas con respecto al punto de escucha. Para pruebas con señales monoaurales y estereofónicas se sugiere utilizar audífonos.

De manera adicional se recomienda tener un nivel de audición fijo, ya que las variaciones de nivel arrojan errores en las pruebas debido a que se producen desviaciones en los umbrales de enmascaramiento.

5.1.6 Selección del material a escuchar

El material que se utilizará para las pruebas se escogerá en base al contenido de instrumentos de las canciones. Conforme avance la prueba el material ira incrementando su complejidad en cuanto a instrumentación ya que esto forzará al sistema y a su vez al individuo al dificultar la apreciación entre la señal de referencia y la señal codificada.

5.1.7 Análisis de datos

Cabe destacar que la escala de graduación antes mostrada no representa una medida física, sino la interpretación de la misma.

Para dar un factor con el cual se pueda medir la prueba se recomienda utilizar un modelo de análisis de la varianza tal como el *ANOVA*. La base para un análisis estadístico apropiado es el grado de diferencia o *SDG*, definido con anterioridad, y no así el grado absoluto. Los resultados del análisis deben ser tales que nos muestren el desempeño promedio de nuestro sistema así como las diferencias entre varios sistemas, si fuese el caso.

La resolución de la prueba se logra por el intervalo de confianza que se utilice. Este intervalo contiene los valores del *SDG* con un grado de confianza, 1-α, donde α representa la probabilidad de que las diferencias inaudibles sean marcadas como audibles. En la práctica se recomienda utilizar un valor α = 0.05, que corresponde a un intervalo de confianza del 95%.

5.2 Evaluación del codificador MPEG-1 Layer III

De acuerdo con el grupo *MPEG* la norma 11172-3 posee las cualidades necesarias para que, a tasas de compresión mayores a 96 kbit/s, la diferencia entre una señal original y una comprimida sea imperceptible al oído humano. Esta aseveración depende del tipo de señal que se quiere comprimir, ya que el contenido en frecuencias de una guitarra con voz dista mucho del contenido de una canción de rock o de una orquesta sinfónica.

Por tal motivo hemos decidido comprobar esta afirmación mediante la prueba *ITU-R BS.1116-1*, explicada con anterioridad, utilizando compresiones desde 32 hasta 160 kbps con diferentes géneros musicales.

Las canciones a utilizar son:

Género	**Intérprete**	**Título**
Trova	Fernando Delgadillo	Hoy ten miedo de mí
Instrumental	Orquesta Sinfónica Nacional de México	Huapango
Banda	Banda “El Recodo”	El club de las feas
Ranchero	Vicente Fernández	Mujeres Divinas
Merengue	Garibaldi	Banana
Rock	Zoé	Vía Láctea

Electrónico	Belanova	Tus ojos

Es importante señalar que las pruebas a realizarse deben de ser en el idioma del escucha, ya que la pronunciación de las vocales y consonantes difiere en cada idioma y esto nos conlleva a que el contenido frecuencial de las palabras no es el mismo.

La prueba se realizó con un grupo compuesto por 27 personas y 8 expertos en procesamiento de audio. La edad del grupo de personas osciló entre los 20 y 30 años de edad, mientras que el grupo de expertos osciló entre los 40 y 55 años.

Las bocinas a utilizar para la prueba son:

Marca: Genelec
Modelo: 8040 A

Especificaciones Técnicas	
Tipo de entrada	Analógica
Intensidad sonora máxima por un par de bocinas con material musical	>115 dB SPL a 1m
Amplificador: Graves Agudos	 90 W 90 W
Dimensiones	365 x 237 x 223 mm
Peso	8.6 kgs
Respuesta en frecuencia en campo libre (±2.0 dB)	48 Hz – 20 kHz
Nivel de ruido autogenerado a 1 m	< -10 dB
Distorsión armónica a 90 dB SPL a 1 m Frecuencias de 50 a 100 Hz >100 Hz	 < 2% < 0.05%
Relación señal a ruido de los amplificadores: Graves Agudos	 > 100 dB > 100 dB

Las características acústicas del recinto en donde se llevaron a cabo las pruebas son las siguientes:

Parámetro	**Especificaciones Estudio 19**
Dimensiones	6.32 x 5.24 m
Tiempo de reverberación	0 ms @ 63 Hz 228 ms @ 125 Hz 253 ms @ 200 a 4000 Hz 189 ms @ > 8000 Hz
Reflexiones primarias	-7.5 dB para 10 ms
Respuesta en frecuencia del cuarto	-4.1 dB @ 50 Hz -3.3 dB @ 125 Hz 0 dB @ 250 – 2000 Hz -1 dB @ 4000 Hz -3.2 dB @ 8000 Hz -5.3 dB @ 16000 Hz

Ruido de fondo	NR 10

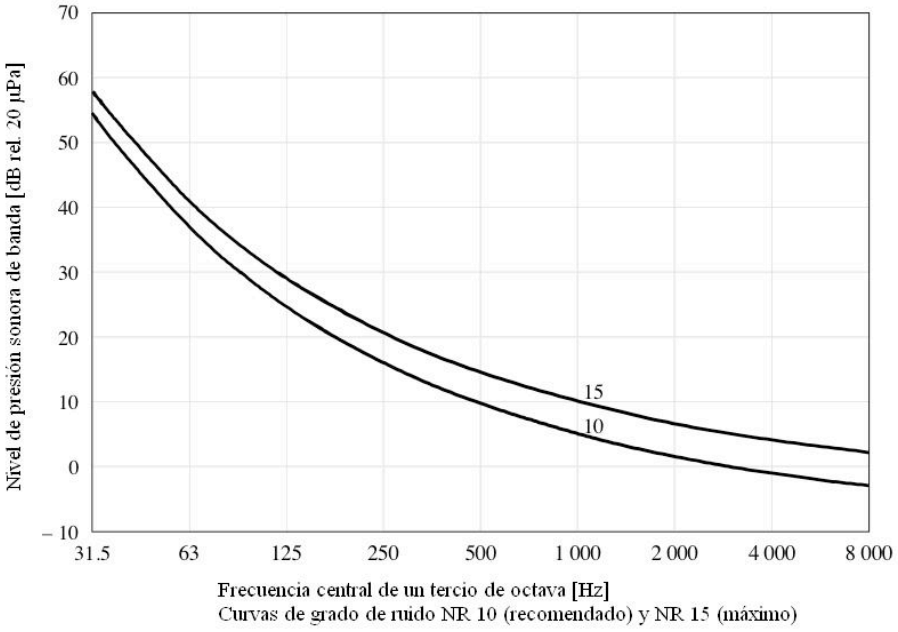


Figura 5.2

Configuración de bocinas:

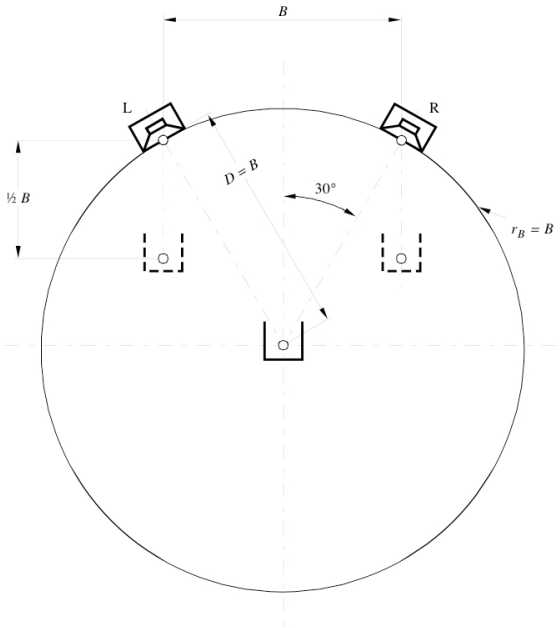
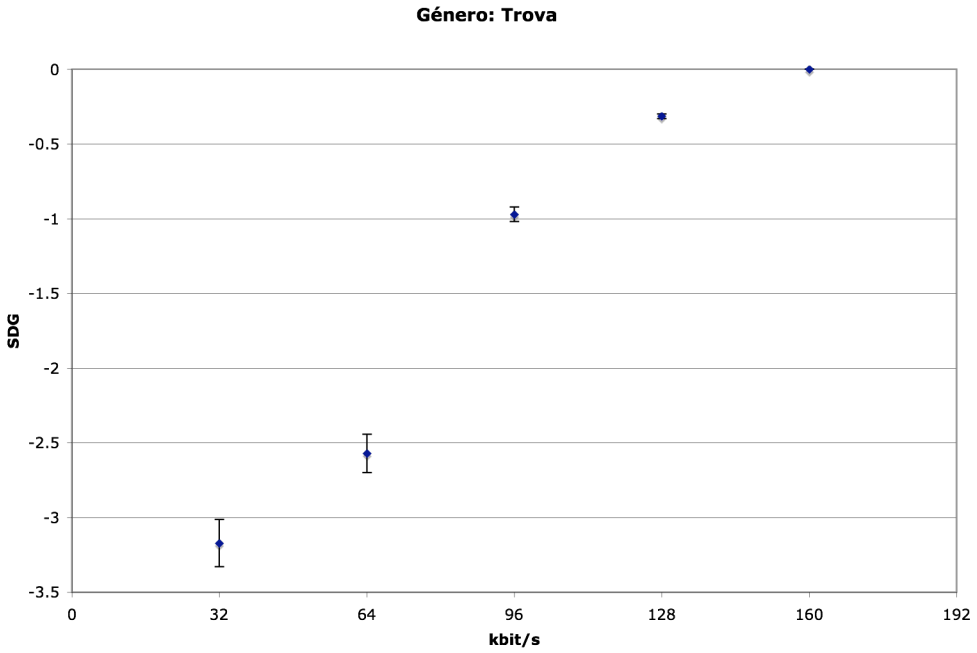


Figura 5.3
D = B = 2m

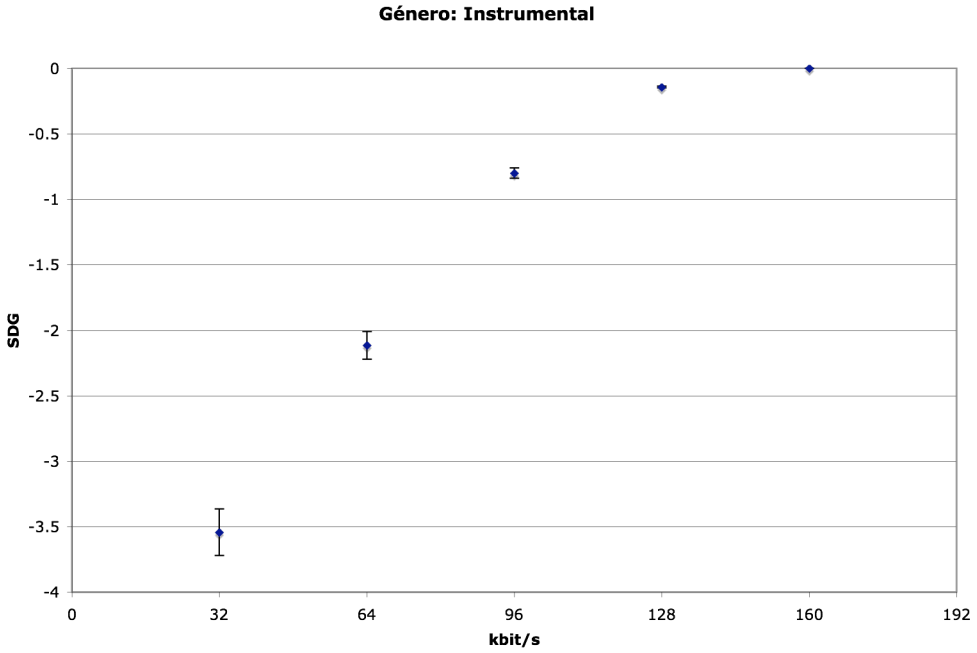
La altura del piso a las bocinas de escucha es de 1.10 m.

5.3 Datos

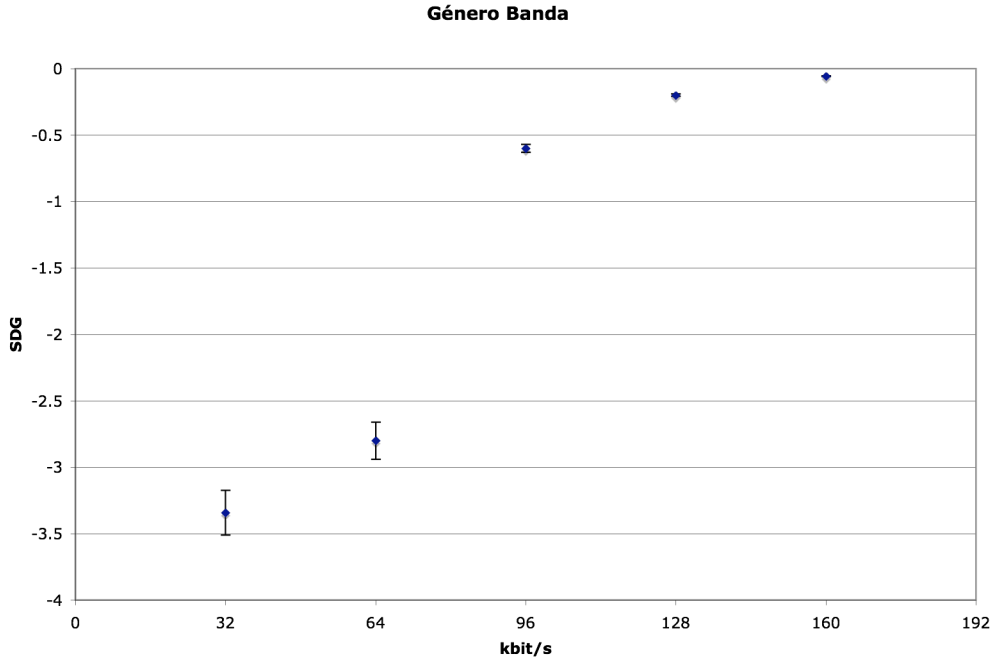
A continuación se presentan las gráficas donde se muestra el SDG de cada bitrate correspondiente a cada género.



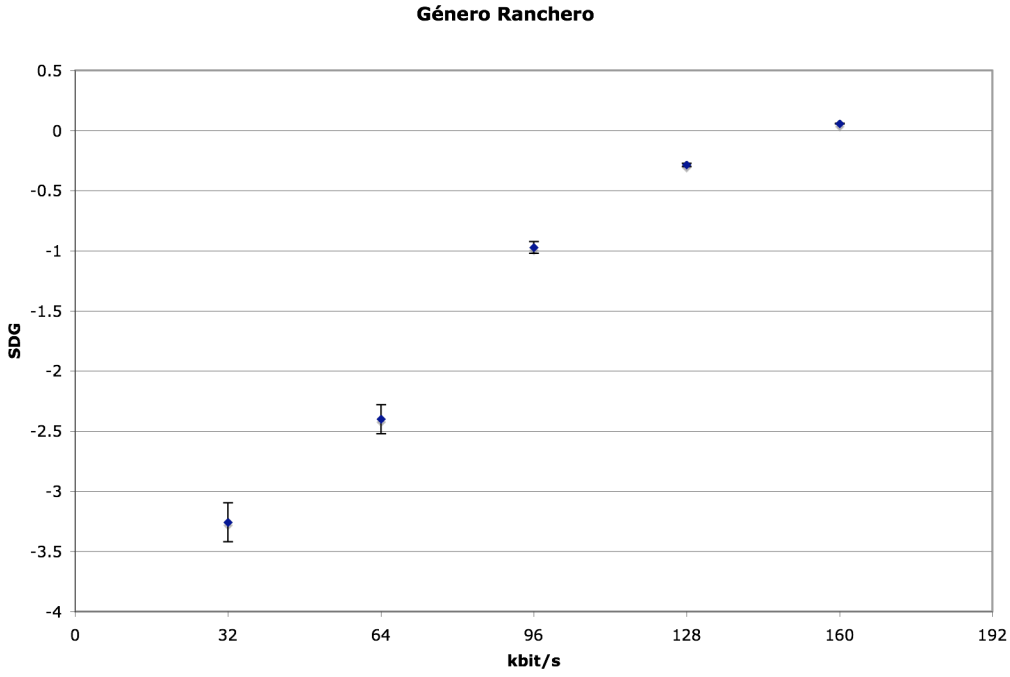
Gráfica 5.1



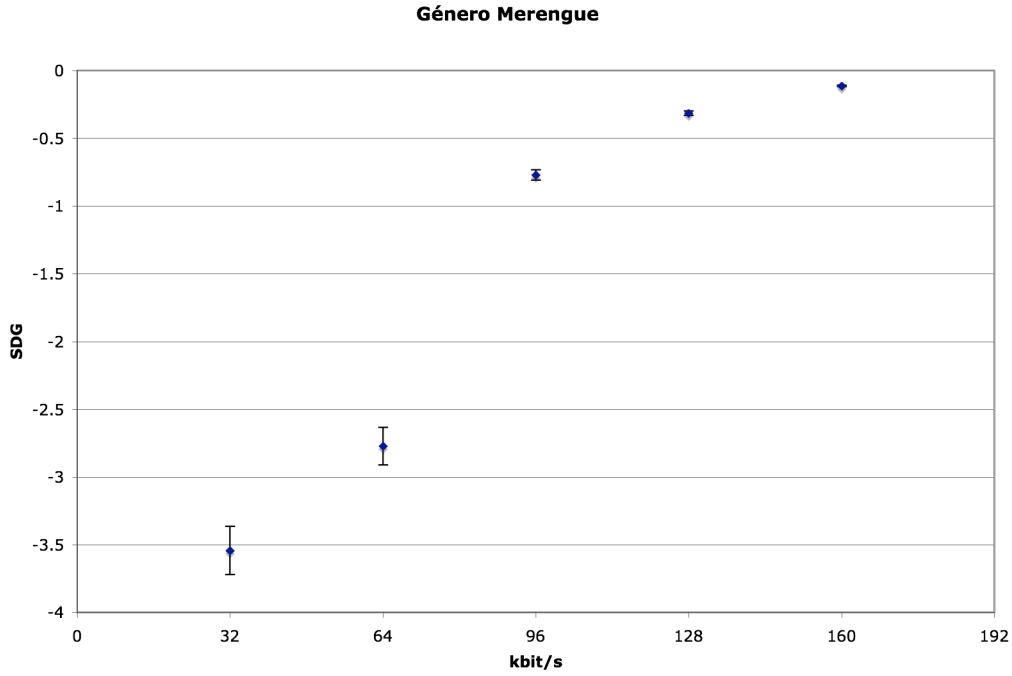
Gráfica 5.2



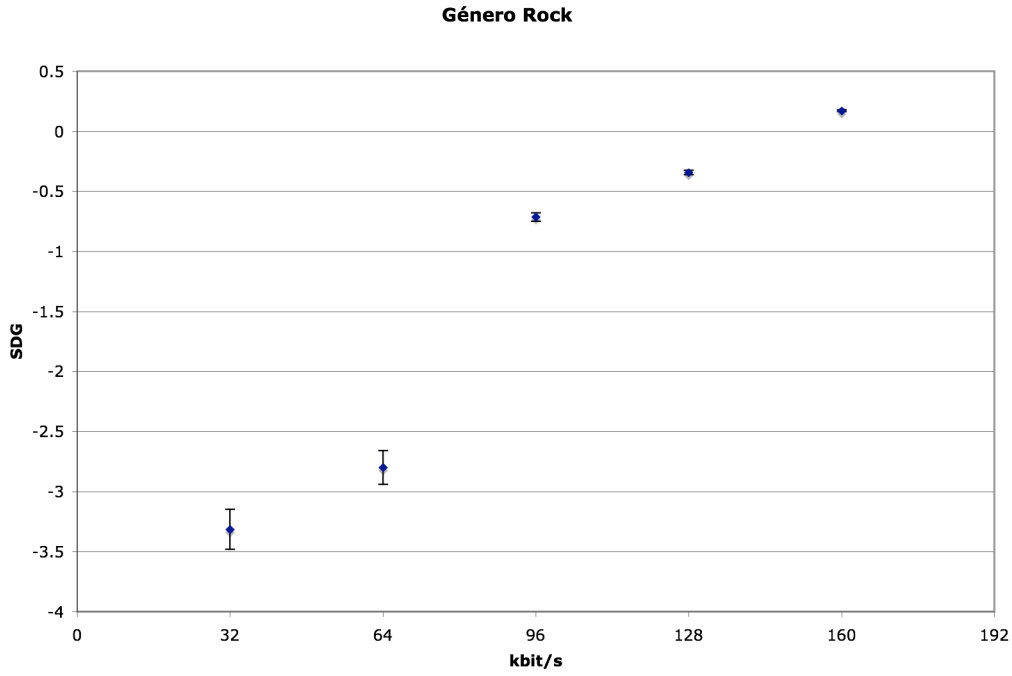
Gráfica 5.3



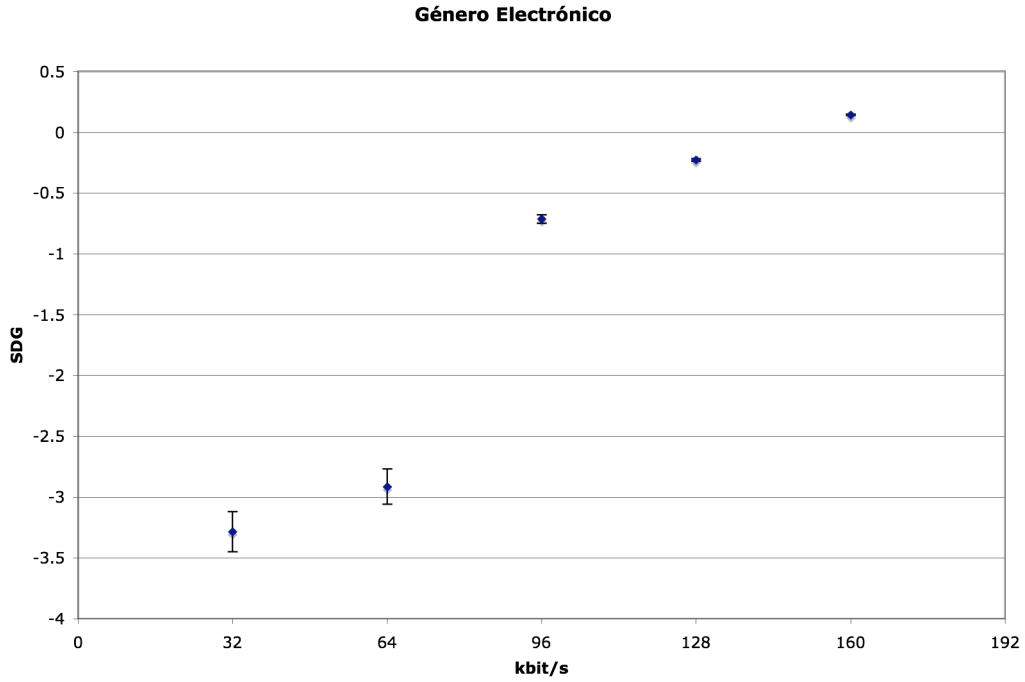
Gráfica 5.4



Gráfica 5.5



Gráfica 5.6



Gráfica 5.7

Como observamos en las gráficas el grupo de escuchas no tuvo dificultad en encontrar la señal codificada en los rangos de compresión de 32 y 64 kbit/s, sin embargo a partir de 128 kbit/s y conforme se hacia más complejo el contenido instrumental de cada canción no les fue tan sencillo distinguir entre la referencia y la señal codificada.

Por lo que encontramos de manera general, que a una tasa de 128 kbit/s las personas entre 20 y 30 años de edad no encontrarán diferencias entre una señal codificada en MP3 y la señal original.

5.4 Evolución del MP3

Como observamos al principio del capítulo IV, la fase 1 del grupo MPEG concluyó en 1992 con la creación de la norma, sin embargo el desarrollo de este codec continua en la actualidad. Esto lo podemos observar en los formatos de compresión mp3PRO y el más novedoso MP3 Surround, así como en la creación de circuitos integrados dedicados, exclusivamente, a la decodificación de este formato.

5.4.1 mp3PRO

Como observamos en los resultados de nuestra prueba, el audio a 128 kbit/s no tiene grandes diferencias perceptibles para el escucha promedio, sin embargo para tasas de compresión menores se distinguen ruidos causados por la codificación en las frecuencias altas. Esto es debido a que a tasas de compresión menores el codificador mp3 utiliza casi todos los bits disponibles en las frecuencias bajas, por lo que no le es posible abarcar el ancho de banda del audio.

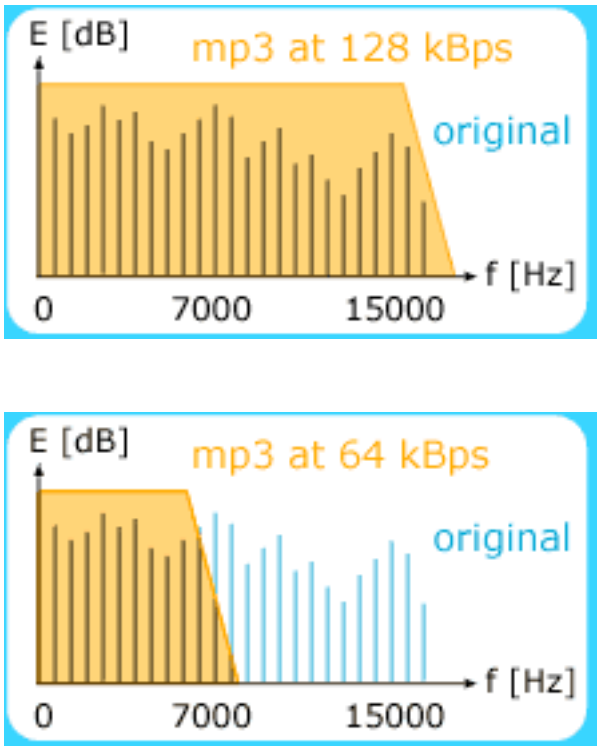


Figura 5.4

Para incrementar la calidad del sonido del mp3 a estas tasas de compresión (menores a 96 kbit/s) la compañía Coding Technologies desarrollo una tecnología conocida como *Spectral Band Replication (SBR norma ISO/IEC 14496-3)* [36] la cual consiste en generar las altas frecuencias de una señal de audio, a través de un generador de ruido, basándose en la información proporcionada por el códec, y con información estadística, como el nivel, la distribución, frecuencias armónicas o el rango.

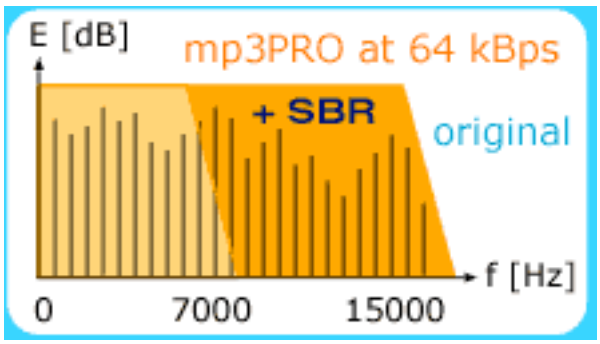


Figura 5.5

Combinando el mp3 con la tecnología SBR se genera una señal de audio con ancho de banda mayor utilizando tasas de compresión bajas. El formato de audio resultante de esta unión, mp3PRO, esta compuesto por dos secciones, la parte mp3 que contiene las bajas frecuentas y la parte SBR o “PRO” que contiene las altas frecuencias. Dado que la sección denominado “PRO” requiere de algunos kbit/s, el formato es compatible con el mp3 original. Esto permite que los reproductores de mp3 puedan reproducir archivos codificados en mp3PRO. Simplemente los reproductores ignoran la parte PRO. El único

requerimiento es que las frecuencias de muestreo sean coherentes con las del mp3 (16, 22.5, 24, 32, 44.1, 48 kHz). Todos los reproductores de software cumplen con estas características, no así todos los reproductores portátiles.

5.4.2 MP3 Surround

Establecido como una norma (ISO/IEC 23003-1) el mp3 Surround es desarrollado debido a la popularización de los sistemas de cine en casa (Home theater) y la demanda que se ha creado a partir del sonido envolvente 5.1.

El modo en el que opera este codec no dista mucho del mp3 que se mencionó en el capítulo 4, siendo el proceso conocido como *downmix* [17][18] y el uso del *Binaural Cue Coding (BCC)* [37] las implementaciones que se agregaron para obtener el mp3 surround.

Partiendo de un archivo multi-canal, por ejemplo el audio de una película o de un DVD, el cual contiene, por lo general, información de audio en 5 canales (L, C, R, Ls y Rs) previo a la codificación mp3 se llevan a cabo dos procesos en paralelo, como se muestra en la figura 5.6

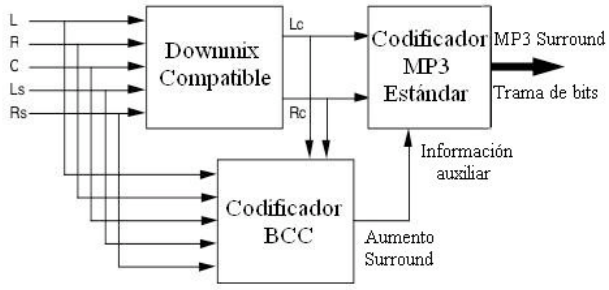


Figura 5.6

El proceso de *downmix* consiste en reducir estos 5 canales en un canal estereo (2 canales) los cuales contengan toda la información del archivo original. Para llevar a cabo esta reducción se utilizan ecuaciones normalizadas por la ITU-R, una de ellas la encontramos en la ecuación 5.1

$$L' = L + \frac{1}{\sqrt{2}} * C + \frac{1}{\sqrt{2}} * Ls$$
$$R' = R + \frac{1}{\sqrt{2}} * C + \frac{1}{\sqrt{2}} * Rs \qquad 5.1$$

De manera paralela se procesa en el *BCC* el archivo con los 5 canales y el *downmix* con lo cual obtenemos información lateral que contiene la diferencia de tiempo (ICTD), la diferencia de nivel (ICLD) y la coherencia (ICC) entre los canales. Esta información junto con la salida del downmix entran a un codificador de mp3 comun y asi obtenemos un archivo mp3 estereo.
En la figura 5.7 se muestra la estructura general de un decodificador de mp3 surround, la cual consiste de un decodificador de mp3 normal y de un decodificador BCC. Dado que el

archivo esta codificado en formato mp3, cualquier decodificador puede producir una salida estéreo; los decodificadores mp3 surround son mas avanzados dado que detectan la presencia de la información surround embebida y producen una señal multicanal.

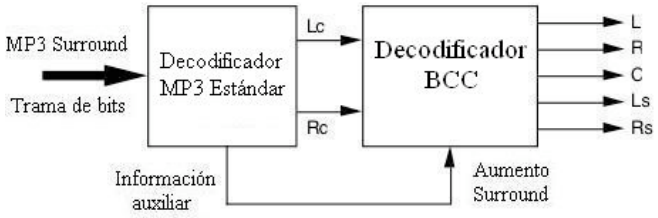


Figura 5.7

Las aplicaciones de este tipo de codificación abarcan desde la radio por Internet hasta el audio para videojuegos.

Capítulo VI

CONCLUSIONES

En la última década hemos observado un rápido crecimiento en la demanda y oferta de los archivos MP3 a lo largo del mundo. Inclusive, algunos artistas de talla internacional han optado por que este formato sea su forma de distribución final hacia el usuario de su música por lo cual, en algunos casos, éstos últimos han dejado atrás la distribución y contrato con disqueras transnacionales.

Ha sido tal la aceptación de estos formatos de audio, que en la actualidad algunas páginas de internet son líderes en la distribución de nueva música y así han surgido nuevos artistas que sin este tipo de herramientas no hubieran tenido una carrera. A esto podemo agregarle un sinfín de programas para descargar música de forma gratuita.

Con la nueva tecnología se han desarrollado equipos capaces de hacer la codificación de archivos de audio en tiempo real, es decir, el audio pasa de su naturaleza puramente analógica a un archivo MP3. Evidentemente, todos estos archivos tienen que ser reproducidos en equipos dedicados y ha sido tal su demanda que en la actualidad existen reproductores cuyo costo es bastante accesible para la mayor parte de la población.

En nuestro país la demanda por archivos de este tipo va en aumento de tal manera que han surgido empresas dedicadas a la comercialización de archivos MP3 de los artistas mexicanos y de habla hispana más importantes en la actualidad.

El presente trabajo, demuestra, apoyados por una prueba de la Internacional Telecomunications Union, como afecta la compresión de archivos de audio a la calidad de la misma.

La compresión en archivos MP3 es buena siempre y cuando se tengan en consideración sus limitantes. Es decir, estar conscientes de que la calidad de una compresión de 56 kbit/s no podrá ser nunca parecida a la que nos ofrece un disco compacto.

En base a los resultados podemos asegurar que los archivos de audio que tengan una compresión mayor o igual a 128 kbit/s pueden ser considerados de buena calidad.

Bibliografía.

[1] NUÑEZ, ADOLFO.
Informática y Electrónica Musical. Segunda edición. Editorial Paraninfo S. A., 1996 España
[2] PEREZ MIÑANA, JOSE.
Compendio practico de acústica. Editorial Labor S.A., 1969 España.
[3] K. BLAIR BENSON
Audio Engineering Handbook. McGraw-Hill, E.UA.
[4] COULTHER DOUG
Digital Audio Processing. CMP Books, 2000, E.U.A
[5] STREICHER, RON & EVEREST, F. ALTON
The New Stereo Soundbook. Segunda edición. Audio engineering associates, 1998. E.U.A.
[6] WORAM, JOHN M.
Soun recording handbook. Primera edición. SAMS, 1992, E.U.A.
[7] KARHS, MARC & BRANDENBURG, KARLHEINZ
Applications of digital signal processingto audio and acoustics. Kluwer Academic Publishers, 1998, E.U.A.
[8] POHLMANN, KEN C.
Principles of digital audio. Editorial Howard W. Sams & Co., 1987 E.U.A.
[9] PIERCE, ALLAN D.
Acoustics. An Introduction to its physical principles and applications. Acoustical society of America, 1991 E.U.A.
[10] BOURLOT, CARLOS JOSÉ,VOLLMANN, ARIEL ROBERTO
Reproductor Portátil De Audio Digital Comprimido, Tesis de Licenciatura, 2000 Argentina
[11] LATHI, B.P.
Introducción a la teoría y sistemas de comunicación, Limusa, 1974. E.U.A.
[12] KREYSZIG, ERWIN
Matemáticas avanzadas para ingeniería Vol. II. Limusa Wiley, 2000,EUA
[13] COUCH II, LEON W.
Modern Communication Systems, Principles and Applications, Prentice Hall, 1994, EUA
[14] OPPENHEIM, ALAN V. & SCHAFER, RONALD W. SCHAFER
Discrete-Time Signal Processing, Prentice Hall 1989, EUA
[15] BALLOU, GLEN M.
Handbook for Sound Engineers. Tercera edición. Focal Prees, 2002, E.U.A.
[16] DAVIS, CAROLYN & DAVIS, DON
Sound system engineering. Segunda edición. Focal Press, 1997, E.U.A.
[17] HOLMAN TOMLINSON
5.1 Surround Sound up and running. Primera edición. Focal Press, 2000, E.U.A.
[18] RUMSEY FRANCIS
Spatial audio. Focal Press, 2001, Reino Unido.
[19] PAINTER, TED
Perceptual Coding of Digital Audio. Articulo publicado en el Vol. 88 de la IEEE, abril 2000, E.U.A.

[20] ZOLSER, UDO
Digital audio signal processing. Primera reimpresión. John Wiley & Sons, 1998. Alemania
[21] MANGIATERRA, SERGIO
Compresión de audio MPEG, Proyecto final, 2003. Argentina
[22] ROMAN, LOPEZ DAVID, et ad
Análisis del algoritmo MPEG capa III. Tesina, mayo 2005. Universidad de Valladolid, España.
[23] RODEN, MARTÍN S.
Analog and Digital Communication System. Segunda edición. Prentice Hall, 1985, E.U.A.
[24] SKLAR, BERNARD
Digital Communications: Fundamentals and Applications. Segunda edición. Prenctice Hall, 1988, E.U.A
[25] BRANDENBURG K., POPP H.
An introduction to MPEG Layer-3. Reseña técnica, Junio 2000 .
[26] RANGACHAR, RAMAPRIYA
Análysis and improvement of the MPEG-1 Audio Layer-III algorithm at low bit-rates*. Tesis, diciembre 2001, Arizona State University*
[27] JACABA, JOEBERT S.
Audio compression using modified discrete cosine transform: the mp3 coding standard. Tesis, octubre 2001,Universidad de Filipinas.
[28] KIM, SEONJOO
Real time MPEG1 audio encoder and decoder implemented on a 16-bit fixed point DSP. Documento tecnico, Samsung Electronics Co. Corea
[29] FÄLTMAN, IRINA; HARST, MARCUS
A hardware implementation of an MP3 decoder. Digital IC-Project, mayo 2003,Suecia.
[30] EHLIAR, ANDREAS & EILERT, JOHAN
A hardware MP3 decoderwith low precision floating point intermediate storage. Documento tecnico, 2003 Linköping University Suecia. http://www.ep.liu.se/
[31] BOSI, MARINA & GOLDBERG, RICHARD E.
Introduction to digital audio coding and standards. Primera edición. Springer Science+Bussiness Media, Inc. 2003, E.U.A.
[32] INTERNATIONAL TELECOMMUNICATIONS UNION
Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. Recommendation ITU-R BS.1116-1. 1994-1997
[34] INTERNATIONAL STANDARS ORGANIZATION, INTERNATIONAL ELECTROTECHNICAL COMISION
ISO / IEC 11172-3.
[35] INTERNATIONAL STANDARS ORGANIZATION, INTERNATIONAL ELECTROTECHNICAL COMISION
ISO / IEC 14496-3.
[36] INTERNATIONAL STANDARS ORGANIZATION, INTERNATIONAL ELECTROTECHNICAL COMISION
ISO / IEC JTC 1/SC 29/WG 11Description of Spectral Band Replication. 2005

[37] FALLER, CHRISTOF
Binaural Cue Coding Applied to Stereo and Multi-Channel Audio Compresion. convention paper, 2002. Alemania. AES, 112th Convention
[38] RESNICK, ROBERT; HALLIDAY, DAVID; KRANE, KENNETH S.
Fisica Vol. I. Cuarta edición. Editorial continental S.A. de C.V. 1993 México.
[39] WILSON JERRY D.
Física. Prentice Hall Hispanoamericana. 1996

http://www.audioworld.com.ar/Sonido_profesional.htm#PRO1
http://members.tripod.com/~MoisesRBB/adc.html
http://www.ii.uam.es/~taao1/teoria/tema1/pdf/Procesadores_dig.pdf
http://www.tiscali.co.uk/reference/dictionaries/computers/data/m0034303.htmlç
http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MonogSO/GESTES02.htm
http://www.teamcombooks.com/mp3handbook/12.htm
http://es.wikipedia.org/wiki/Onda_%28f%C3%ADsica%29
http://www.eumus.edu.uy/docentes/maggiolo/acuapu/sap.html
http://www.peruprom.com/orientacion/oido.html