

Descripción de la técnica de compresión MP3 para audio



Óscar Machuca Gómez
Manuel Espinosa Candón

**Técnicas de Adquisición y
procesado digital de señales**

Índice

<i>1. Introducción: Qué es el MP3</i>	<i>3</i>
<i>2. Descripción de la técnica de compresión MP3 para audio</i>	<i>4</i>
2.1 Digitalización.	4
2.2 Codificación y Compresión.....	6
2.3 Codificación sub-banda (SBC).	8
<i>3. Bibliografía.....</i>	<i>10</i>

1. Introducción: Qué es el MP3

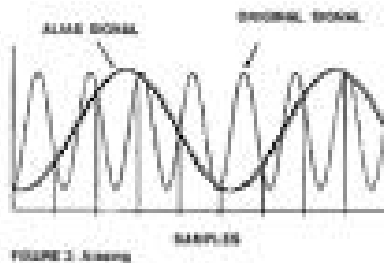
MP3 es un formato de datos que debe su nombre a un algoritmo de codificación llamado **MPEG 1 Layer 3**, el cual, a su vez, es un sistema de compresión de audio que permite almacenar sonido con una calidad similar a la de un CD y con un índice de compresión muy elevado, del orden de 1:11. En la práctica, esto significa que en un CD-Rom se pueden grabar unos 11 CD de audio, es decir, unas 150 canciones aproximadamente.

El sistema de codificación que utiliza MP3 es un **algoritmo de pérdida**. Es decir, el sonido original y el que obtenemos posteriormente no son idénticos. Esto se debe a que MP3 aprovecha las deficiencias del oído humano y elimina toda aquella información que no somos capaces de percibir. Se han realizado multitud de estudios de percepción acústica descubriendo que hay una serie de efectos que pueden ayudar a la codificación del sonido con el objetivo de reducir todo lo posible la cantidad de información inútil o redundante. Los más importantes son:

- ? **Los límites de audición.** Nuestro oído solo trabaja con frecuencias que van entre los 20 Hz y los 20 KHz aproximadamente, con lo que las frecuencias restantes son descartables.



- ? **Efecto de enmascaramiento.** Es aquel que se produce cuando dos señales de frecuencia similar se superponen. Entonces solo podemos percibir aquella que posee más volumen y, por lo tanto, la de volumen menor es susceptible de ser eliminada.



- ? **Redundancia de estéreo.** Existen redundancias entre los componentes tonales y no tonales del sonido en los dos canales estéreo, y, además, por debajo de una cierta frecuencia el oído humano no es capaz de percibir la direccionalidad del sonido, por lo cual por debajo de estas frecuencias es posible incluso codificar un solo canal, junto con información complementaria para restaurar la sensación espacial para el otro canal.

Para realizar ésta acción de "pérdida de información" se utiliza un sistema llamado **Codificación de SubBandas**, proceso por el cual la señal se descompone en subbandas a través de un banco de filtros. Éstas subbandas se comparan a continuación con el original mediante un modelo psicoacústico que es el encargado de determinar que bandas se pueden eliminar y cuales no. Dependiendo de la calidad que deseemos obtener, se eliminarán más o menos bandas. Para finalizar el proceso, se cuantifican y codifican las subbandas resultantes, y el resultado final se comprime mediante un algoritmo estándar, obteniendo así el fichero **MP3** resultante. El proceso de codificación es mucho más complicado que el de decodificación, por ello se tarda mucho más en codificar un archivo MP3 que en reproducirlo.

Este algoritmo de codificación perceptual fue desarrollado por la sociedad MPEG (*Moving Picture Expert Group*) en conjunción con el Instituto Tecnológico de Franunhofer, y se ha estandarizado como una norma ISO.

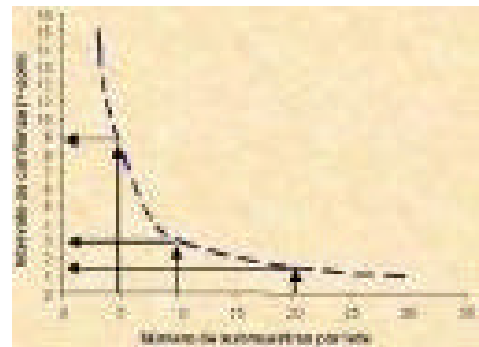
2. Descripción de la técnica de compresión MP3 para audio

2.1 Digitalización.

El sonido es una onda continua que se propaga a través del aire u otros medios, formada por diferencias de presión, de forma que puede detectarse por la medida del nivel de presión en un punto. Las ondas sonoras poseen las características propias y estudiadas de las ondas en general, tales como reflexión, refracción y difracción. Al

tratarse de una onda continua, se requiere un proceso de digitalización para representarla como una serie de números. Actualmente, la mayoría de las operaciones realizadas sobre señales de sonido son digitales, pues tanto el almacenamiento como el procesado y transmisión de la señal en forma digital ofrece ventajas muy significativas sobre los métodos analógicos. La tecnología digital es más avanzada y ofrece mayores posibilidades, menor sensibilidad al ruido en la transmisión y capacidad de incluir códigos de protección frente a errores, así como encriptación. Con los mecanismos de decodificación adecuados, además, se pueden tratar simultáneamente señales de diferentes tipos transmitidas por un mismo canal. La desventaja principal de la señal digital es que requiere un ancho de banda mucho mayor que el de la señal analógica, de ahí que se realice un exhaustivo estudio en lo referente a la compresión de datos, algunas de cuyas técnicas serán el centro de nuestro estudio.

El proceso de digitalización se compone de dos fases: muestreo y cuantización. En el muestreo se divide el eje del tiempo en segmentos discretos: la frecuencia de muestreo será la inversa del tiempo que medie entre una medida y la siguiente. En estos momentos se realiza la cuantización, que, en su forma más sencilla, consiste simplemente en medir el valor de la señal en amplitud y guardarlo. El teorema de Nyquist



garantiza que la frecuencia necesaria para muestrear una señal que tiene sus componentes más altas a una frecuencia dada f es como mínimo $2f$. Por tanto, siendo el rango superior de la audición humana en torno a los 20 KHz, la frecuencia que garantiza un muestreo adecuado para cualquier sonido audible será de unos 40 KHz. Concretamente, para obtener sonido de alta calidad se utilizan frecuencias de 44'1 KHz, en el caso del CD, por ejemplo, y hasta 48 KHz, en el caso del DAT. Otros valores típicos son submúltiplos de la primera, 22 y 11 KHz. Según la naturaleza de la aplicación, por supuesto, las frecuencias adecuadas pueden ser muy inferiores, de tal manera que el proceso de la voz acostumbra a realizarse a una frecuencia de entre 6 y 20 KHz. o incluso menos. En lo referente a la cuantización, es evidente que cuantos más bits se utilicen para la división del eje de la amplitud, más "fina" será la partición y por tanto menor el error al atribuir una amplitud concreta al sonido en cada instante. Por ejemplo, 8 bits ofrecen 256 niveles de cuantización y 16, 65536. El margen

dinámico de la audición humana es de unos 100 dB. La división del eje se puede realizar a intervalos iguales o según una determinada función de densidad, buscando más resolución en ciertos tramos si la señal que se trata tiene más componentes en cierta zona de intensidad, como veremos en las técnicas de codificación.

El proceso completo se denomina habitualmente PCM (Pulse Code Modulation) y así nos referiremos a él en lo sucesivo. Se ha descrito de forma sumamente simplista, principalmente porque está ampliamente tratado y es sobradamente conocido, siendo otro el campo de estudio de este trabajo. Sin embargo, entraremos en detalle en todo momento que sea necesario para el desarrollo de la exposición.

2.2 Codificación y Compresión.

Antes de describir los sistemas de codificación y compresión, debemos detenernos en un breve análisis de la percepción auditiva del ser humano, para comprender por qué una cantidad significativa de la información que proporciona el PCM puede desecharse. El centro de la cuestión, en lo que a nosotros respecta, se basa en un fenómeno conocido como enmascaramiento.

El oído humano percibe un rango de frecuencias entre 20 Hz. y 20 KHz. En primer lugar, la sensibilidad es mayor en la zona alrededor de los 2-4 KHz., de forma que el sonido resulta más difícilmente audible cuanto más cercano a los extremos de la escala. En segundo lugar está el enmascaramiento, cuyas propiedades utilizan exhaustivamente los algoritmos más interesantes: cuando la componente a cierta frecuencia de una señal tiene una energía elevada, el oído no puede percibir componentes de menor energía en frecuencias cercanas, tanto inferiores como superiores. A una cierta distancia de la frecuencia enmascaradora, el efecto se reduce tanto que resulta despreciable; el rango de frecuencias en las que se produce el fenómeno se denomina banda crítica (critical band). Las componentes que pertenecen a la misma banda crítica se influyen mutuamente y no afectan ni se ven afectadas por las que aparecen fuera de ella. La amplitud de la banda crítica es diferente según la frecuencia en la que nos situemos y viene dada por unos determinados datos que demuestran que es mayor con la frecuencia. Hay que señalar que estos datos se obtienen por experimentos psicoacústicos, que se realizan con expertos entrenados en percepción sonora, dando origen con sus impresiones a los modelos psicoacústicos.

Este que hemos descrito es el llamado enmascaramiento simultáneo o en frecuencia. Existe, asimismo, el denominado enmascaramiento asimultáneo o en el tiempo (ver apéndice 2), así como otros fenómenos de la audición que no resultan relevantes en este punto. Por ahora, centrémonos en la idea de que ciertas componentes en frecuencia de la señal admiten un mayor ruido del que generalmente consideraríamos tolerable y, por tanto, requieren menos bits para ser codificadas si se dota al codificador de los algoritmos adecuados para resolver máscaras.

La digitalización de la señal mediante PCM es la forma más simple de codificación de la señal, y es la que utilizan tanto los CD como los sistemas DAT. Como toda digitalización, añade ruido a la señal, generalmente indeseable. Como hemos visto, cuantos menos bits se utilicen en el muestreo y la cuantización, mayor será el error al aceptar valores discretos para la señal continua, esto es, mayor será el ruido. Para evitar que el ruido alcance un nivel excesivo hay que emplear un gran número de bits, de forma que a 44'1 KHz. y utilizando 16 bits para cuantizar la señal, uno de los dos canales de un CD produce más de 700 kilobits por segundo (kbps). Como veremos, gran parte de esta información es innecesaria y ocupa un ancho de banda que podría liberarse, a costa de aumentar la complejidad del sistema decodificador e incurrir en cierta pérdida de calidad. El compromiso entre ancho de banda, complejidad y calidad es el que produce los diferentes estándares del mercado y formará la parte esencial de nuestro estudio.

Calidad	Muestreo	Bits/muestra	Modo	Tasa de bits	Frecuencia
Teléfono	8 KHz	8	Mono	64 kbps	200-3400 Hz
Radio AM	11'025 KHz	8	Mono	88 kbps	
Radio FM	22'050 KHz	16	Estéreo	705'6 kbps	
CD	44'1 KHz	16	Estéreo	1411'2 kbps	20-20000 Hz
DAT	48 KHz	16	Estéreo	1536 kbps	20-20000 Hz

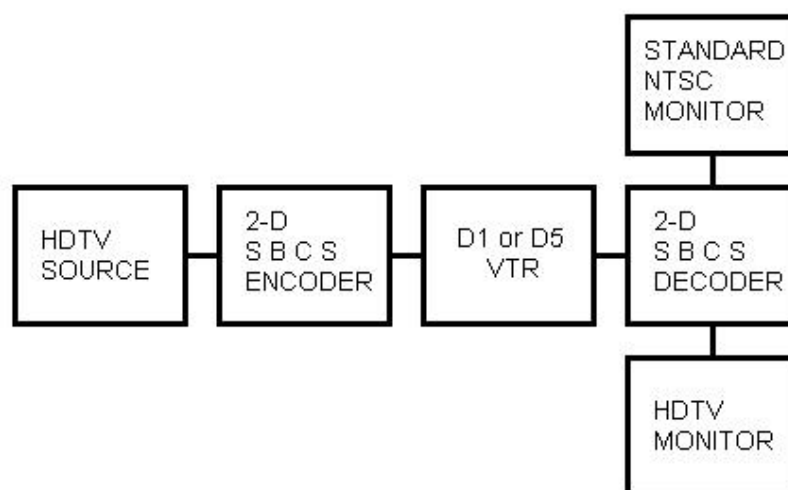
Tabla 1 : comparación de formatos de calidad de audio

Un modo mejor de codificar la señal es mediante PCM no-lineal o cuantización logarítmica, que como ya comentamos consiste en dividir el eje de la amplitud de tal forma que los escalones sean mayores cuanto más energía tiene la señal, con lo que se consigue una relación señal/ruido igual o mejor con menos bits. Con este método se puede reducir el canal de CD audio a 350 kbps, lo cual evidentemente es una mejora

sustancial, aunque puede reducirse mucho más. Otros sistemas similares nos llevan a la cuantización adaptativa (APCM), diferencial (DPCM) y la mezcla de ambas, ADPCM. Así prosigue la reducción del ancho de banda, pero sin llegar a los niveles que proporciona el tener en cuenta los efectos del enmascaramiento.

2.3 Codificación sub-banda (SBC).

La codificación sub-banda o SBC (sub-band coding) es un método potente y flexible para codificar señales de audio eficientemente. A diferencia de los métodos específicos para ciertas fuentes, el SBC puede codificar cualquier señal de audio sin importar su origen, ya sea voz, música o sonido de tipos variados. El estándar MPEG Audio es el ejemplo más popular de SBC, y lo analizaremos posteriormente en detalle.



El principio básico del SBC es la limitación del ancho de banda por descarte de información en frecuencias enmascaradas. El resultado simplemente no es el mismo que el original, pero si el proceso se realiza correctamente, el oído humano no percibe la diferencia. Veamos tanto el codificador como el decodificador que participan en el tratamiento de la señal.

La mayoría de los codificadores SBC utilizan el mismo esquema. Primero, un filtro o un banco de ellos, o algún otro mecanismo descompone la señal de entrada en varias subbandas. A continuación se aplica un modelo psicoacústico que analiza tanto las bandas como la señal y determina los niveles de enmascaramiento utilizando los datos

psicoacústicos de que dispone. Considerando estos niveles de enmascaramiento se cuantizan y codifican las muestras de cada banda: si en una frecuencia dentro de la banda hay una componente por debajo de dicho nivel, se desecha. Si lo supera, se calculan los bits necesarios para cuantizarla y se codifica. Por último se agrupan los datos según el estándar correspondiente que estén utilizando codificador y decodificador, de manera que éste pueda descifrar los bits que le llegan de aquél y recomponer la señal.

La decodificación es mucho más sencilla, ya que no hay que aplicar ningún modelo psicoacústico. Simplemente se analizan los datos y se recomponen las bandas y sus muestras correspondientes.

En los últimos diez años la mayoría de las principales compañías de la industria del audio han desarrollado sistemas SBC. A finales de los años ochenta, un grupo de estandarización del ISO llamado Motion Picture Experts Group (MPEG) comenzó a desarrollar los estándares para la codificación tanto de audio como de vídeo. Veremos MPEG Audio como ejemplo de un sistema práctico SBC.

3. Bibliografía

Libros consultados

? MP3

Juan F. Marcelo, Eva Martín

Madrid Anaya Multimedia, 2000

? Guía esencial MP3

Julio Crespo

Madrid, Prentice Hall, 2001

? PCM y transmisión digital

G.H. Bennett

Barcelona. Ediciones Técnicas Rede, 1988

Páginas Web visitadas

? <http://www.hispamp3.com/tallerm3/como/queesunmp3.shtml>

? <http://www.hispamp3.com/tallerm3/tutoriales/mp3profundidad/2.shtml>

? <http://www-pagines.fib.upc.es/~si/treballs-SI2001/e4022697/tecnicas.html>

? <http://www.cs.ucl.ac.uk/staff/jon/mmbook/book/node120.html>

? http://searchnetworking.techtarget.com/sDefinition/0,,sid7_gci214284,00.htm

|