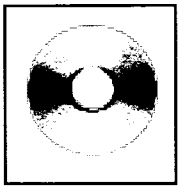


MP3

SONIDO DIGITAL AL ALCANCE DE TODOS

David Rincón Rivera
drincon@mat.upc.es
Departament de Matemàtica Aplicada i Telemàtica
Universitat Politècnica de Catalunya (UPC)

INTRODUCCIÓN

El formato de compresión de sonido *MPEG-1 Layer 3* (también conocido como MP3) está adquiriendo una gran notoriedad, debido al revuelo que está provocando en el campo de la grabación y distribución de audio digital. En los últimos meses se han producido repetidos intentos de las compañías discográficas para limitar sus posibilidades de grabación y reproducción, ya que Internet se ha convertido en una vía de distribución paralela a las habituales (y en muchas ocasiones, ilegal).

En éste artículo se va a describir el formato MP3 desde dos puntos de vista: el del técnico y el del usuario. Desde el punto de vista técnico haremos hincapié en los algoritmos de compresión de sonido utilizados por el estándar, así como algunos comentarios sobre una de las aplicaciones más prometedoras de MP3, que es su transmisión a través de redes de conmutación de paquetes. La segunda parte del artículo estará dedicada a proporcionar información y herramientas para experimentar en nuestro ordenador la calidad de éste estándar.

INTRODUCCIÓN AL AUDIO DIGITAL

Como ya sabemos, para digitalizar una señal lo único que necesitamos es disponer de un convertidor analógico/digital (A/D), que se compone de un módulo de muestreo y un codificador.

El papel del primer bloque, también conocido como *sample & hold*, es discretizar la señal en el tiempo. Para asegurar una correcta reconstrucción de la señal, el teorema de Nyquist nos obliga a que la frecuencia de muestreo sea mayor o igual al doble del ancho de banda de la señal original.

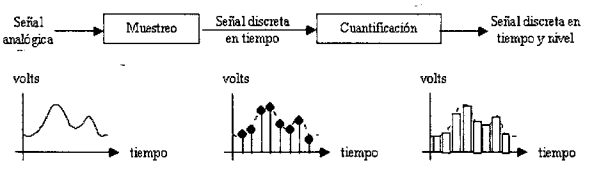


Figura 1. *Proceso de muestreo y cuantificación.*

El segundo bloque cuantifica los valores obtenidos por el primero, con un cierto número de bits por muestra. El sistema de audio digital más sencillo es el PCM (*Pulse Code Modulation*), que se limita a cuantificar de manera uniforme la señal muestreada. La cantidad de bits (n) empleados en la cuantificación determinan la relación señal a ruido (SNR, *signal to noise ratio*) del proceso de digitalización, de manera que cada bit adicional añade 6 dB [1].

$$SNR_q\ (dB) = \text{constante} + 6n\ dB$$

Vamos a comentar dos ejemplos de audio digital PCM: telefonía digital y el Compact Disc. En el primero, digitalizamos una señal de 3.1 Khz a un ritmo de 8000 muestras por segundo, con 8 bits/muestra (SNR » 50 dB), lo que genera una tasa de 64 Kbit/s. El CD muestrea una señal de alta fidelidad (20 Khz) a 44100 muestras/segundo y 16 bits/muestra (SNR » 100 dB), en dos canales (stereo), generando una tasa total de 1.4 Mbit/s. Como vemos, esta tasa es elevadísma, y requiere un gran ancho de banda para ser transmitida. Además, el PCM suele ser una codificación muy ineficiente, ya que cada muestra es muy parecida a la anterior, con lo que tenemos una redundancia muy alta.

Por ello se diseñaron algoritmos de compresión basados en la predicción temporal de las muestras, como el DPCM o la modulación Delta, que se basan en la codificación de la diferencia entre la muestra real y la predicha por el sistema en base a las muestras anteriores. La predicción se hace a partir de unos coeficientes que pueden ser fijos o variables, de manera que se adapten a los cambios de la señal de entrada, haciendo que la señal reconstruida sea más fiel a la inicial (un ejemplo de estos sistemas es el ADPCM).

Sin embargo, aunque estos sistemas explotan con éxito la eliminación de la redundancia, sólo son capaces de reducir la tasa en un factor de entre 2 y 4 (se considera que un ADPCM a 32 Kbit/s ofrece una calidad ligeramente superior al PCM de 64 Kbit/s [1]). Por ello se hizo patente la necesidad de crear nuevos esquemas de compresión que explotaran otro tipo de propiedades. Así en el campo de la telefonía aparecieron los detectores de silencio, que eliminan la transmisión cuando la señal es tan baja que no va a ser captada por el oído humano, o los sistemas llamados *vocoder*, que intentan reproducir las características del tracto vocal humano (cuerdas vocales, boca, lengua...) para analizar y sintetizar digitalmente las for-

mas de onda que salen de nuestra garganta. Pero estos sistemas son óptimos cuando son utilizados para codificar voz humana y no música, que es mucho más rica en matices y que contiene más información.

Por ello se inició otra línea de investigación basada en el otro extremo de la comunicación, en el receptor: el oído humano. A estas técnicas se les llama «psicoacústicas» o de «codificación perceptual», porque se basan en las propiedades de nuestro sistema auditivo para comprimir la información acústica a tasas inimaginables hasta el momento. Hay sonidos que no podemos oír, así que podemos eliminarlos y ahorrar una gran cantidad de información. Vamos a describir con más detalle cuáles son las características del sistema auditivo humano, y cómo podemos aprovecharlas en el proceso de compresión.

¿QUÉ OÍMOS Y QUÉ NO PODEMOS OÍR? EFECTOS PSICOACÚSTICOS

El rango frecuencial en el que el oído humano es capaz de detectar sonido está comprendido entre los 20 Hz y los 20 Khz, con una zona especialmente sensible entre los 2 y los 4 Khz, muy cercana al espectro de la voz, situada entre los 500 hz y los 4 Khz [2]. Por tanto, dos tonos de potencia similar situados en los 3 y los 15 Khz serán percibidos de manera muy diferente (el de 15 Khz pasará mucho más desapercibido, pudiendo ser incluso inaudible). Es decir, nuestro oído no ofrece una respuesta plana con la frecuencia, sino que premia a unas bandas y penaliza otras, llegando al extremo de pasar desapercibidas. Por ello se define el **umbral absoluto de audición** como la frontera entre los sonidos que son perceptibles y los inaudibles. El concepto fue acuñado por Fletcher en 1940, durante una serie de experimentos donde se obtuvieron gráficas como la presentada en la figura 2, que presenta el umbral de audición en función de la frecuencia. El gráfico se obtuvo por métodos empíricos, efectuando un muestreo estadístico entre la población. Los valores se establecen respecto a un tono puro de 1 Khz con una potencia tal que se encuentra en el límite de audición.

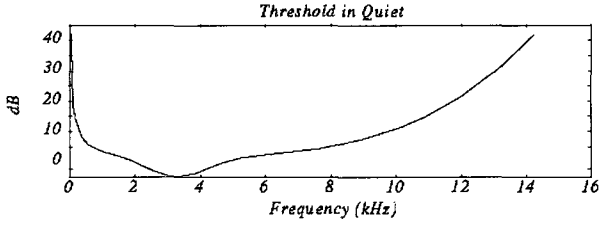


Figura 2. *Umbral de audición del oído humano [3].*

Según esta gráfica, toda señal que quede por debajo de la curva será inaudible para prácticamente todas las personas, así que no tiene sentido codificarla y puede ser eliminada.

Sin embargo, este umbral de audición no es único. La aparición de señales adicionales puede modificar nuestra percepción de un cierto tono, llegando incluso a producir **enmascarados frecuenciales** (*frequency masking*). En la figura 3 se presenta un ejemplo. Supongamos que disponemos de un tono de 1 Khz a un nivel fijo (60 dB por encima del umbral de audición), que llamaremos «tono enmascarador». Generamos otro tono de 1.1 Khz y medimos el nivel de potencia al que se hace indistinguible. Si repetimos el proceso para toda la banda, obtendremos una segunda curva de umbral, esta vez generada por el tono enmascarador. Como en el caso anterior, toda señal que quede por debajo de esta curva será inaudible (se dice que ha sido enmascarada por el tono dominante). Como es lógico, cuanto más cercano esté la señal al tono enmascarador y menor sea su potencia, más posibilidades de que sea enmascarado.

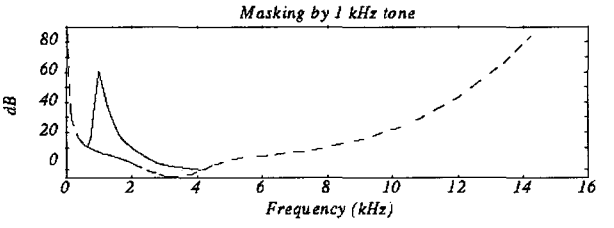


Figura 3. *Enmascarado causado por un tono de 1 Khz [3].*

En la figura 4 se muestra la forma de la campana de enmascarado para diversas frecuencias. La característica más destacable es que a medida que crece la frecuencia, más ancha se hace la campana.

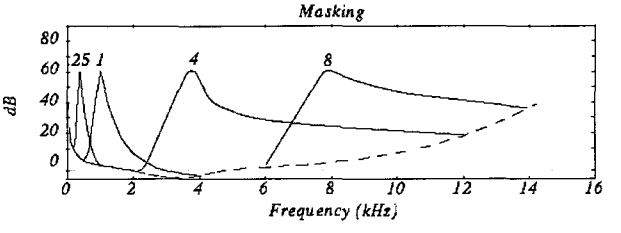


Figura 4. *Enmascarado frecuencial [3].*

Podemos observar que el ancho de la campana de enmascarado sigue una regla aproximadamente logarítmica. Por ello parece natural crear el concepto de **banda crítica** (*critical band*), que se define como cada una de las porciones del espectro en la que el oído percibe una señal uniforme [3]. Para medir estas porciones se crea la unidad llamada *bark* (en honor a Barkhausen) para denominar al ancho de banda correspondiente a una banda crítica. A partir del estudio empírico de las bandas críticas, se llegó a la conclusión de que se podían calcular de la siguiente manera:

- Para frecuencias inferiores a 500 Hz, 1 bark » f/100
- Para frecuencias superiores a 500 Hz, 1 bark » 9 + 4log (f/1000)

Con este convenio se consiguen gráficas como la presentada en la figura 5. Aquí podemos observar cómo la introducción de unidades logarítmicas como los *bark* permiten subdividir el espectro en bandas de tamaño uniforme. Como veremos más adelante, este detalle es importante cuando hay que efectuar un análisis sub-banda.

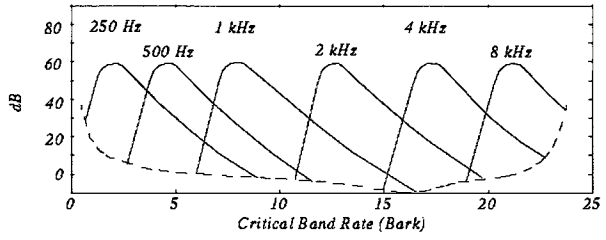


Figura 5. *Bandas críticas, representadas en función de los barks [3].*

Hasta ahora nos hemos limitado a medir los fenómenos de enmascarado en el dominio frecuencial, pero también se producen **enmascarados temporales**. Un tono muy potente enmascarará durante un cierto intervalo temporal cualquier otro tono de frecuencia parecida y que sea temporalmente cercano. En la figura 6 se presenta el caso que exponemos a continuación. Tenemos un tono enmascarador de 1 Khz y 60 dB, y otro tono de 1.1 Khz y 40 dB, que está enmascarado. En t = 0, desactivamos el tono enmascarador y medimos cuanto tarda el oído en percibir el segundo tono (se puede hacer desconectándolo en t = Dt, y disminuir Dt hasta que se deja de percibir). Si repetimos el experimento para diferentes potencias, obtenemos una respuesta como la de la figura 6. Cuanto menos potente sea el tono enmascarado, más tarda el oído humano en recuperarse de la saturación que le ha provocado el tono enmascarador. Este es un efecto que cualquiera de nosotros ha experimentado: después de escuchar un sonido fuerte, como por ejemplo una explosión, nos quedamos momentáneamente sordos y necesitamos un poco de tiempo para recuperar la agudeza auditiva habitual.

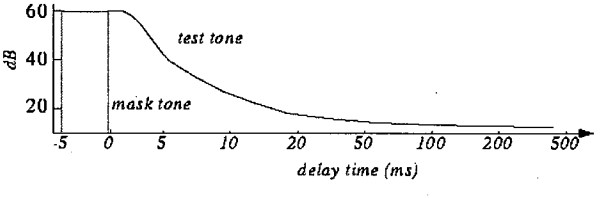


Figura 6. *Enmascarado temporal [3].*

Si repetimos este experimento para diferentes potencias y frecuencias, obtendremos una gráfica como la de la figura 7, donde podemos observar el efecto combinado de los enmascarados frecuenciales y temporales. Curiosamente se puede ver que existe el fenómeno del «enmascarado previo»: hay sonidos que son enmascarados antes de que se genere el tono enmascarador. Esto no es un error; parece ser que nuestro oído necesita un cierto tiempo antes de poder identificar un tono. Si en este tiempo se produce el tono enmascarador, el tono enmascarado no será percibido en absoluto.

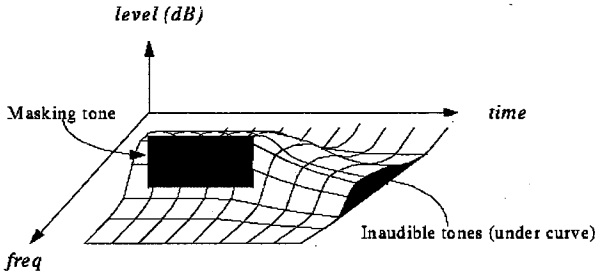


Figura 7. *Efecto combinado del enmascarado [3].*

¿CÓMO PODEMOS APROVECHAR LOS EFECTOS PSICOACÚSTICOS?

Como ya hemos comentado, la gran ventaja de los modelos psicoacústicos es que toda señal que quede por debajo del umbral total de enmascarado (la curva combinada del umbral absoluto, del enmascarado frecuencial y del temporal) es inaudible y, por tanto, no se codifica. También se ha visto que las bandas críticas son las unidades naturales en las que podemos dividir, de manera uniforme, la influencia del enmascarado dentro del espectro. Por tanto parece natural que el proceso de compresión utilice codificación sub-banda (*subband coding*), consistente en separar la señal de entrada en un cierto número de bandas y hacer un análisis independiente de cada una de ellas. Esto se puede conseguir mediante un banco de filtros.

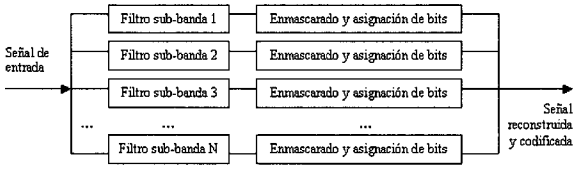


Figura 8. *Esquema del análisis sub-banda.*

A continuación, a cada sub-banda se le aplica un análisis psicoacústico o perceptual que determina las curvas de enmascarado, y determina cuáles son los tonos que se encuentran por encima de dichas curvas (los componentes que están por debajo son simplemente descartados). Así se genera la relación señal a máscara (SMR, *signal-to-mask ratio*). A los tonos supervivientes se les asigna una cierta cantidad de bits en función del ruido que podemos introducir, calculado a partir del enmascarado, la SMR y las necesidades de tasa del sistema global. El criterio para asignar bits a la señal es sencillo: la potencia del ruido debe quedar por debajo del umbral de audición. Veamos un ejemplo: supongamos que en una cierta sub-banda con un nivel de enmascarado de 26 dB existe una señal superviviente con una potencia de 40 dB. Dicha señal debería ser codificada con un mínimo de 7 bits (7 bits

x 6 dB/bit = 42 dB > 40 dB). Sin embargo podemos ahorrar 4 bits, ya que el ruido los enmascara (4 bits x 6dB/bit = 24 dB < 26 dB). Por tanto, nos basta con 3 bits para codificar la señal superviviente.

Mediante este método combinado de eliminación de señales enmascaradas y reducción de la información correspondiente a las supervivientes se pueden conseguir factores de compresión muy elevados, sin comprometer la calidad (aparente) del sonido. En muchos casos se eliminan sub-bandas enteras que han sido enmascaradas por tonos situados en la banda vecina, haciendo que no se utilice ni un solo bit para codificar la sub-banda en cuestión.

LOS ESTÁNDARES MPEG

MP3 es el nombre con el que se conoce a una de las partes del estándar de codificación de vídeo MPEG (Moving Picture Experts Group) [4] de la Organización Internacional de Estándares (ISO) [5]. Concretamente, MP3 hace referencia a la capa 3 del codificador de audio de MPEG-1, así que lo mejor será empezar describiendo el estándar MPEG.

MPEG-1 fue el primer estándar internacional de codificación de vídeo creado por la ISO que aplicó técnicas de compresión basadas en el enmascarado de la información visual y acústica; es decir, lo que el usuario no va a ver ni oír, no se codifica. Esta técnica ya fue aplicada con gran éxito por el comité JPEG (Joint Photograph Experts Group), que dio origen al formato de compresión de imágenes del mismo nombre, y que se ha convertido en el estándar de facto en Internet. Fruto de este éxito, la ISO formó el grupo MPEG a finales de los 80 para crear diversos estándares de vídeo digital de alta calidad. El plan inicial era crear cuatro versiones diferentes, cada una de ellas destinadas a un segmento específico de usuarios y aplicaciones [6]:

- **MPEG-1:** Codificador de vídeo a 1.5 Mbit/s con calidad de videoconferencia mejorada, de resolución 352x288 pixels (CIF) o superior. Apareció como la evolución natural del estándar de videoconferencia ITU-T H.261, con mejoras relacionadas con la compensación de movimiento y la predicción temporal.

- **MPEG-2:** Codificador de vídeo a tasa de 4-10 Mbit/s con calidad de emisión de TV («*broadcast*») comparable a los sistemas analógicos PAL, SECAM y NTSC. Destinado a ser el estándar de emisión de TV digital de consumo masivo.

- **MPEG-3:** Codificador de vídeo a tasas superiores a los 10 Mbit/s, con calidad de TV de alta definición (HDTV), destinado a ser usado en centros de producción y en redes de transmisión.

- **MPEG-4:** Codificador de videoconferencia a tasa muy bajas (64 – 256 Kbit/s) para ser usado sobre redes de banda estrecha, especialmente de telefonía móvil.

De estos cuatro estándares, sólo 3 han visto la luz (MPEG-1, 2 y 4). MPEG-3 se quedó por el camino, ya que los algoritmos desarrollados para MPEG-2 son tan potentes y flexibles que permiten abarcar tanto la calidad *broadcast* como la de producción, simplemente variando la tasa a la que funciona el codificador.

Aunque el esfuerzo más grande de los ingenieros se dedicó a los algoritmos de codificación de vídeo, no se descuidó el sonido que tenía que acompañar a las imágenes. Uno de los handicaps con los que tuvieron que luchar los diseñadores de MPEG fue el requisito de escalabilidad, que consiste en que se debe permitir que equipos de gamas diferentes puedan reproducir el mismo flujo de información, aunque sea a calidades diferentes. Para ello se definió una arquitectura de tres capas, en la que cada «*layer*» se basa en un codificador más sofisticado que el de la capa anterior. Así, tenemos MPEG-1 Layer 1, Layer 2 y Layer 3, siendo esta última la más complicada y la más eficiente desde el punto de vista de compresión. Los reproductores de capa 3 son capaces de reproducir flujos codificados con cualquiera de las tres técnicas, mientras que los de capa 1 sólo pueden reproducir información

	Tasa objetivo	Factor de compresión	Calidad a 64 Kbit/s	Calidad a 128 Kbit/s	Retardo teórico de compresión
Capa 1	192 Kbit/s	4:1			19 ms
Capa 2	128 Kbit/s	6:1	2.1 a 2.6	> 4	35 ms
Capa 3	64 Kbit/s	12:1	3.6 a 3.8	> 4	59 ms

Tabla 1. *Características de las capas de audio de MPEG-1.*

codificada según la capa 1. Las características de cada capa son las siguientes [3]:

La tasa objetivo es el *bitrate* para el que se diseñó cada una de las capas. El factor de compresión nos da la relación entre la tasa generada por el codificador MPEG y la que se necesitaría en PCM para conseguir una calidad equivalente. Vemos que se consiguen factores de hasta 12, lo que nos da una idea de la potencia del algoritmo.

Los otros dos apartados interesantes de la tabla 1 son la valoración de la calidad subjetiva a las tasas de 64 y 128 Kbit/s. Esta medida de calidad se realiza basándose en el criterio MOS (*Mean Opinion Square*), definido por la ITU (Unión Internacional de Telecomunicaciones). Consiste en hacer un análisis estadístico de la calidad percibida por grupos de personas escogidas al azar en diferentes países, que hacen una valoración subjetiva de la calidad de los tests y pruebas presentadas. La escala MOS tiene un rango comprendido entre 1 (ininteligible) y 5

(perfecto). Vemos que incluso a tasas muy bajas, las capas 2 y 3 obtienen valoraciones muy buenas. Una anécdota surgida durante el período de pruebas: parece que uno de los pocos casos en los que el algoritmo no funciona con la calidad adecuada es en la codificación de voz masculina alemana, aunque esto se puede solucionar elevando la tasa del flujo [6].

Existen cuatro modos de funcionamiento para cada una de las capas: canal único (una sola señal de audio en el flujo), canal doble (dos canales separados e independientes), stereo (igual al anterior pero con dos señales pertenecientes a los canales derecho e izquierdo de una señal stereo original), y joint stereo (parecido al anterior pero explota la redundancia entre los dos canales para reducir aún más la tasa).

CODIFICACIÓN Y DECODIFICACIÓN DE AUDIO MPEG

Los estándares MPEG son del tipo denominado **asimétrico**, en el que los codificadores son mucho más complejos que los decodificadores. Esto es así para permitir la comercialización de reproductores baratos, destinados al mercado de electrónica de consumo. Veremos que los codificadores soportan una carga computacional muy superior a la de los decodificadores.

Otro detalle importante desde el punto de vista del implementador es que **no se especifica un estándar de codificación**. Lo que sí existe es una especificación de qué tipo de flujos de bits es capaz de reproducir un cierto «decodificador modelo», y una serie de recomendaciones sobre cómo puede construirse un codificador. Esto permite fomentar el desarrollo de algoritmos de codificación diferentes para cada fabricante (que puede así diferenciarse de sus competidores, promover la investigación y preservar sus patentes), manteniendo al mismo tiempo la compatibilidad (ya que todos los codificadores han de ser compatibles con cualquier reproductor que siga el modelo especificado).

La fuente sobre la cual se aplican los algoritmos debe ser siempre una señal PCM a las frecuencias de muestreo de 32 Khz, 44.1 Khz (propia del Compact Disc) y 48 Khz (propia del sistema DAT), con 16 bits por muestra (unos 100 dB de relación señal a ruido de cuantificación).

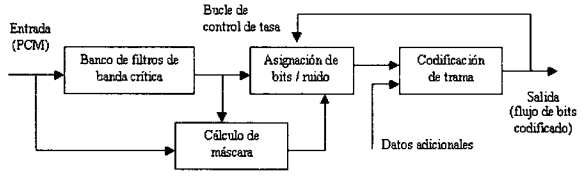


Figura 9. Esquema general del proceso de codificación de audio MPEG-1.

En un apartado anterior hemos comentado cómo se podían aprovechar los efectos psicoacústicos para comprimir (con pérdidas) la información correspondiente a una fuente de sonido. El esquema general utilizado por las tres capas MPEG es el presentado en la figura 9.

El proceso de codificación se realiza sobre el flujo continuo de bits de entrada. Sin embargo, para facilitar la compresión y permitir su segmentación temporal, se define la **trama** (*frame*) como el bloque unitario mínimo que puede ser decodificado completamente. Las tramas se componen de una cierta cantidad de muestras de entrada, que varía según la capa utilizada.

Podemos observar que la funcionalidad básica es la ya mostrada en la figura 8: el banco de filtros que separa la señal en subbandas críticas, el módulo que evalúa los efectos psicoacústicos y elimina las señales que quedan por debajo de la curva, y el bloque que asigna bits a las señales supervivientes en función del nivel de ruido enmascarado y de las necesidades de tasa instantánea (puede observarse la realimentación desde la salida, para el caso en que se exija una tasa constante). Finalmente encontramos un bloque que se dedica a formatear el flujo de bits de salida, con funciones como controlar la tasa (constante o variable), comprimir aún más los datos mediante algoritmos como Huffman o Ziv-Lempel, segmentar el flujo de salida en tramas, añadir marcas temporales para su correcta reproducción en el decodificador, introducir un canal de datos adicionales con información sobre el autor de la música, códigos de acceso, etc.

Como ya comentamos anteriormente, el esquema del decodificador es mucho más sencillo que el codificador. Se limita a extraer la información formateada de las subbandas, reconstruirlas por separado (con las curvas de enmascarado) y finalmente, a unir todas las bandas para formar la señal reconstruida.

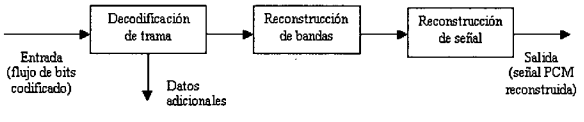


Figura 10. Esquema general del proceso de decodificación de audio MPEG.

Vamos a comentar brevemente las características de cada una de las capas [2, 6, 7, 8]:

MPEG-1 Audio layer 1:

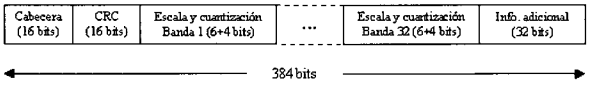
- Segmentación de la señal en 32 subbandas a partir de un filtro polifase de baja complejidad.
- Análisis frecuencial mediante FFT de 512 puntos.
- Cálculo de la SMR a partir del tono dominante en cada subbanda.

- Modelo psicoacústico que sólo tiene en cuenta el enmascarado frecuencial.
- Cuantización fija del máximo valor absoluto de todas las subbandas con 6 bits. Este valor actua como escala para el resto de bandas.

- Formato de trama: 384 bits, correspondientes a 12 muestras de 32 subbandas (a 48 Khz, equivalentes a 8 ms de sonido).

- Sincronización y ordenación: 16 bits

- Corrección de errores (CRC): 16 bits



- Para cada banda: 6 bits de factor de escala + 4 bits de cuantización del valor dominante.

- Información adicional: 32 bits

MPEG-1 Audio layer 2:

Esta capa es muy similar a la anterior, pero consigue tasas de compresión mayores (a costa de incrementar el retardo). Las diferencias principales respecto a la capa 1 son las siguientes:

- FFT de 1024 puntos (mayor resolución espectral).

- 3 valores de escala para cada subbanda en cada trama. Se pueden usar unos, dos o los tres.

- Cuantización variable de cada subbanda, en función del factor de escala.

- El modelo psicoacústico incluye la influencia del enmascarado temporal (no completamente)

- Aumento de la resolución de los cuantizadores a 16 bits.

- Tramas de 1152 bits (equivalentes a 3 de las tramas de capa 1), correspondientes a 36 muestras de 32 subbandas (a 48 Khz, equivalentes a 24 ms de sonido).

MPEG-1 Audio layer 3:

La capa 3 es, con diferencia, la más complicada y sofisticada de toda la familia MPEG. Introduce mejoras sustanciales en casi todos los pasos de la codificación, como:

- Uso de un banco de filtros híbridos conmutados (DCT/polifase) que proporciona resolución variable, ajustándose mejor a las bandas críticas.
- El modelo psicoacústico explota en su totalidad el enmascarado temporal, e incluye una predicción polinómica.

- Aprovecha la redundancia inherente en las señales estereofónicas.

Calidad de sonido	Ancho de banda	Modo	Tasa	Factor de compresión
Sonido telefónico	2.5 Khz	Mono	8 Kbit/s	96:1
≈ Radio AM	4.5 Khz	Mono	16 Kbit/s	48:1
Mejor que AM	7.5 Khz	Mono	32 Kbit/s	24:1
≈ Radio FM	11 Khz	Stereo	56-64 Kbit/s	26-24:1
Cercana a CD	15 Khz	Stereo	96 Kbit/s	16:1
Calidad CD	22.1 Khz	Stereo	112-128 Kbit/s	14-12:1

Tabla 2. *Comparación de las diversas calidades de la capa 3 de MPEG-1.*

- Cuantización variable con control dinámico del ruido.
- Codificación de trama mediante códigos entrópicos de Huffman y RLE (Run-length encoding).

- Formato de trama: 1152 bits. La trama ya no es la unidad de decodificación, sino que necesita de información codificada (almacenada o transmitida) previamente. Orientado a la transmisión en flujo.

La calidad ofrecida por la capa 3 se compara con la de otros sistemas en la Tabla 2 [9].

MP3 Y EL MUNDO DE LA MÚSICA

Hasta aquí hemos descrito el estándar MP3 desde el punto de vista técnico. ¿Cuál ha sido la razón de su éxito? Lo que en principio era solamente una curiosidad técnica se ha ido convirtiendo en una ola que hace que las discográficas empiecen a preocuparse. ¿Por qué?

La razón es que se ha dado la convergencia de tres factores: la popularización de los ordenadores con capacidad multimedia (especialmente las grabadoras de CD-ROM), el fenómeno del software de libre distribución, y el *boom* de Internet.

La gran novedad del MP3 es que es el primer estándar con el que se puede adquirir y, sobretodo, copiar audio de alta calidad, usando un simple PC. ¿Y cómo se puede hacer eso? Con uno de los cientos de programas de software gratuito que se pueden encontrar tanto en Internet como en forma de regalo por parte de revistas de informática. Básicamente hay tres tipos de programas: los *rippers* (programas que leen las pistas del CD y las convierten a ficheros WAV codificados en PCM a 44.1 Khz),

codificadores (compresores que a partir del WAV aplican los algoritmos y generan un fichero MP3) y decodificadores (reproductores de MP3). Se pueden encontrar versiones comerciales mucho mejor acabadas y con gran cantidad de prestaciones, pero los más populares son programas muy sencillos creados por programadores sin afán comercial, que ceden su producto a toda la comunidad de usuarios (*freeware*).

Un usuario con un equipo medio puede comprimir canciones desde el CD sin ningún tipo de problemas: le basta con poner en marcha el *ripper* y a continuación utilizar el codificador. Normalmente el proceso de compresión no se puede hacer a tiempo real; una canción de tres minutos puede tardar entre diez y quince en ser comprimida, con un Pentium de primera generación. Es necesario un equipo bastante más potente (un Pentium II) para poder generar MP3 en tiempo real. Y para reproducir basta con un equipo de gama baja con una tarjeta de sonido. Esto es posible debido a la característica asimétrica del estándar: el decodificador es mucho más sencillo que el codificador. Así, un simple Pentium a 133 Mhz es capaz de reproducir sin problemas un fichero o un flujo MP3 (siempre que sea el único proceso que corre).

Tal como se comentó en un apartado anterior, se puede conseguir una calidad propia del CD con un factor de compresión que se aproxima a un valor de 10-12. Es el caso de la capa 3 a 44.1 Khz, en modo «*joint stereo*» y una tasa de 128 Kbit/s. Con estos parámetros se consigue que una canción de 4 minutos ocupe menos de 4 Mbytes, lo cual posibilita que en un CD-ROM de 600 Mbytes quepan más de 12 horas de música en MP3, o que sea factible enviar canciones por correo electrónico, o bien capturarlas desde Internet.

Y esto es lo que preocupa a las discográficas, la posibilidad de transportar fácilmente la información, ya sea en CD o a través de la red. Por un lado se está produciendo un fenómeno de piratería (ya es posible tener en un solo CD toda la discografía de un artista); por otro lado, están perdiendo el mercado de la distribución (hay artistas noveles que editan sus trabajos en la red, e incluso algún cantante consagrado como David Bowie que distribuye canciones en su servidor web [10]). Están apareciendo multitud de portales y buscadores especializados en música MP3 [11] que son una auténtica mina para los «piratas musicales», ya que la estructura de Internet hace posible establecer servidores en países donde las discográficas no pueden actuar legalmente en su contra. Es por ello que estas compañías están promoviendo diferentes estándares de compresión que incorporan protección contra copias, pero todavía no está claro que consigan imponerlos en el mercado y, sobre todo, en la red.

La batalla continúa. Hay quien cree que es una lucha entre piratas y empresas, y hay quien piensa que es una pugna por la democratización del acceso a la música, acorde con la filosofía original de Internet (acceso universal a la información). Como siempre, depende del lado desde el que se mire...

PARA EMPEZAR A JUGAR

En este apartado vamos a ofrecer información práctica sobre programas y productos que nos permitirán experimentar con el sonido MP3. Comentaremos cuáles son los más populares y dónde encontrarlos.

Figura 11. *Reproductor WinAmp, con la piel original.*

El primer contacto con el mundo MP3 suele ser a través de un reproductor. El más popular es el WinAmp (http://www.winamp.com) de la compañía NullSoft Inc. Es un programa actualmente *freeware* que decodifica tanto ficheros como flujos HTTP y RTP (con un *plug-in*) de los formatos MP3, CD-Audio, WMA, MOD y WAV, entre otros. Una de las características más curiosas es su capacidad de cambiar de piel (*skin*). Existen versiones con motivos de Star Trek, el OS de Apple, o el interfaz X-Window de Unix. Los fans más incondicionales puede diseñar su propia versión y donarla al resto de usuarios.

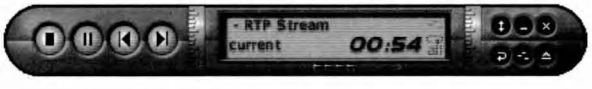


Figura 12. *Reproductor FreeAmp, en su versión Windows.*

Otro reproductor, algo menos popular, es el FreeAmp (http://www.freeamp.org), del cual está disponible el ejecutable y el código, compilable tanto para Windows (Visual C++) como para Linux. Este decodificador permite recibir flujos HTTP/RTP/Multicast sin ningún tipo de plug-in adicional. El código es ideal para estudiar a fondo el funcionamiento del MP3, al menos desde el lado del reproductor.

Si queremos generar nuestros propios ficheros MP3, necesitaremos *rippers* (para fuentes de CD-Audio) y codificadores. Aunque se pueden encontrar por separado, lo habitual es que las dos funciones estén juntas en el mismo programa. Es el caso de AudioGrabber (http://www.audiograbber.com-us.net) y AudioCatalyst (http://audiocatalyst.com). Los dos permiten seleccionar opciones de adquisición (velocidad del CD, protección contra errores, drivers, etc.) y compresión (pre-normalización de la señal, modo stereo/mono, frecuencia de muestreo, tasa

de bits, preénfasis, etc.). De hecho, estos programas suelen ser simples *front-ends*, interfaces que controlan el motor de compresión. Los motores más habituales son el L3encoder, del instituto Fraunhofer [9], LAME y XingMP3 Encoder (http://www.xingtech.com/mp3/encoder/).

Figura 13. Reproductor Rio.

Pero no todo es software: empiezan a aparecer los reproductores basados en hardware. Son portátiles y muy parecidos a los *walkman*. El primero que apareció, y el más popular, es el Rio de Diamond (http://www.diamondmm.com). El último modelo, el Rio 500, dispone de una memoria de 64 Mb, ampliable mediante tarjetas flash. La carga de los ficheros de música se realiza mediante las tarjetas o bien mediante un puerto USB con el que se conecta a un ordenador de sobremesa. El peso es de apenas 100 gramos, lo que hace que sea ideal como reproductor portátil.

TRANSMISIÓN DE AUDIO POR RED: RADIO MP3

Para finalizar este artículo, comentaremos una de las aplicaciones más innovadoras del MP3: la posibilidad de crear emisoras de radio a través de Internet, con calidad suficiente como para ser comerciales y con un ancho de banda suficientemente bajo como para poder ser recibidas a través de una conexión de baja velocidad (un módem de 33/56 Kbit/s o un canal RDSI de 64 Kbit/s).

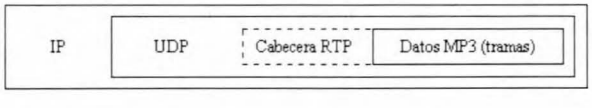


Figura 14. *Protocolos involucrados en la transmisión de MP3 en red.*

Esto se consigue segmentando los ficheros MP3 en trozos pequeños, de manera que en cada uno de ellos exista un número entero (y no muy elevado) de tramas. Como hemos visto en la descripción del estándar, las tramas nos permiten decodificar por entero un cierto número de muestras PCM. Así generamos un flujo continuo (*stream*) de trozos que se encapsulan en paquetes RTP (Real Time Protocol) [12], que inserta marcas temporales que permiten una reconstrucción fiel del flujo original. Estos paquetes, a su vez, son transportados por los protocolos UDP, o TCP (que a su vez descansan sobre el IP).

El uso de UDP, que se define como un protocolo «ligero» (es decir, de funcionamiento sencillo y con muy poca carga de cabeceras) permite que el transporte de tramas MP3 sea muy eficiente. Sin embargo, el TCP proporciona mucha más protección frente a pérdidas, a costa de retardos y saltos en la reproducción.

Estas dos opciones de transporte han hecho aparecer dos tipos de emisoras de MP3. Por un lado tenemos las basadas en TCP, cuyo mejor exponente es ShoutCast (http://www.shoutcast.com). Esta página web, mantenida por la misma empresa del WinAmp, es en realidad un portal que da entrada a varios centenares de emisoras distribuidas por todo el mundo, conectadas a la sede central de Nullsoft a través de Internet. El transporte de la información se realiza sobre HTTP y TCP, y se abre una conexión por cada nuevo usuario, lo que limita la cantidad de oyentes (ya que en caso contrario se desbordaría la capacidad del ordenador emisor y de los enlaces que lo conectan a Internet).

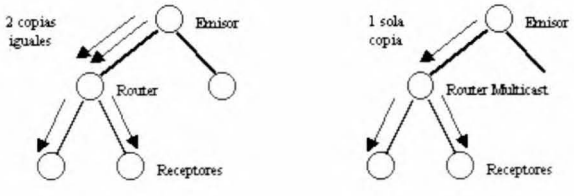


Figura 15. *Comparación entre el modo unicast (izquierda) y el multicast (derecha).*

Este problema de escalabilidad, que es un obstáculo para una emisión masiva, es resuelto por la otra gran familia de emisoras. Éstas se basan en el uso de UDP y del protocolo IP con extensiones Multicast [12], que permite que sólo se emita una copia de la información desde el emisor, independientemente del número de receptores que estén activos. Cuando la información llega al último *router* del árbol de distribución, se efectúa una copia por cada uno de los usuarios que quieren recibir el flujo. Así se minimiza la carga en los tramos superiores del árbol, donde solo circula una copia de los paquetes. Esta técnica requiere

de routers modernos, con capacidad de encaminar los paquetes multicast, por lo que su uso está restringido por ahora a entornos experimentales y universitarios, pero sin duda se extenderá en el futuro hacia los usuarios comerciales de Internet. Una aplicación muy sencilla y potente para la transmisión de flujos MP3 multicast es liveCaster (http://www.livegate.com).

FUTURO

A la vista de la velocidad a la que cambia el mundo de la tecnología, se hace difícil decir cuál va a ser el futuro de los codificadores MPEG y qué impacto van a tener en nuestra vida cotidiana. Lo que sí parece claro es que hay una tendencia hacia la creación de productos de hardware especializado en compresión y reproducción MP3, lo cual podría hacer que se convirtiera en un estándar «de facto» en el mundo de la electrónica de consumo, desplazando a productos como el MiniDisc y compitiendo con el CD Audio.

En el campo de la transmisión de audio con calidad *broadcast*, ya sea asociado a una señal de vídeo o por sí mismo, se continúa el desarrollo de algoritmos basados en efectos psicoacústicos. El último de ellos es el AAC (*Advanced Audio Coding*) de Dolby y NBC, aprobado como estándar para la banda de sonido de MPEG-2. Este sistema se basa en el Dolby Surround de 5+1 canales, y es capaz de multiplexar hasta 48 canales de audio, 15 canales de mejora de baja frecuencia, y 15 canales de datos. Según sus diseñadores, un flujo AAC stereo a 96 Kbit/s ofrece una calidad superior a la de MPEG-1 capa 3 a 128 Kbit/s o MPEG-capa 2 a 192 Kbit/s. Para conseguirlo, suma técnicas predictivas a las psicoacústicas y utiliza un banco de filtros de alta resolución. Pero ya se están anunciando algoritmos de compresión superiores en prestaciones...

Como puede verse, el mundo del audio digital está en plena ebullición. Os recomiendo que os mantengáis al corriente de las últimas novedades visitando los *links* que se proporcionan al final del artículo, y que experimentéis con los programas; es la mejor manera de aprender y disfrutar de la tecnología.

PARA MÁS INFORMACIÓN...

... sobre los estándares MPEG de video y audio:

http://www.mpeg.org

... sobre las cuestiones técnicas relacionadas con el formato de audio MPEG:

http://www.mp3tech.com

... sobre el Instituto Fraunhofer, creador del estándar MP3:

http://www.iis.fhg.de

... sobre música en formato MP3:

http.//mp3.lycos.com

... sobre productos software y hardware MP3:

http://www.mp3.com

... sobre código C de codificadores y decodificadores:

http://mp3tech.free.fr/programmers/programmers.html

REFERENCIAS

[1] B. Sklar, «Digital commmunications fundamentals and applications», Prentice-Hall International, 1988.

[2] Marcos Faúndez Zanuy, «Estándares de codificación de audio MPEG», Mundo Electrónico, Septiembre 1999.

[3] Z. Nian-Li, Audio Compression course notes, http://www.cs.sfu.ca/CC/365/li/material/notes/Chap4/Chap4.4/Chap4.4_prev.html

[4] Official MPEG Website, http://drogo.cselt.stet.it/mpeg

[5] ISO – International Standards Organisation, http://www.iso.ch

[6] Introducción a la compresión de audio: MPEG 1 Layer 3, http.//www.geocities.com/SiliconValley/Vista/5390/index.html,

[7] Davis Pan, «A Tutorial on MPEG Audio compresion», IEEE Multimedia, pp 60-74, 1995

[8] J.L. Mitchell, W.B. Pennebaker, C.E. Fogg, D.J. LeGall, MPEG Video Compression Standard, Chapman and Hall - International Thomson Publishing, 1997.

[9] Fraunhofer IIS. http://www.iis.fhg.de/amm/techinf/layer3/index.html

[10] David Bowie website, http://www.davidbowie.com

[11] Un buscador de ficheros MP3 se puede encontrar en http://mp3.lycos.com

[12] S.A. Thomas, «IPng and the TCP/IP Protocols», Wiley Computer Publishing, 1996.

[13] Kosiur, D.R., «IP multicasting the complete guide to interactive corporate networks», John Wiley & Sons, 1998.