El Formato MP3

Introducción

El formato mp3 presenta un sin fin de características que son las que lo convierten, hoy en día, en uno de los tipos de archivos más difundidos. Esto se debe a diversas cualidades que este formato presenta de las cuales, muchas de ellas pasan desapercibidas delante de muchos usuarios, o quizás muchas de ellas no son conocidas debido a que a este formato se lo trata con un muy poco interés cuando se refiere a cómo está compuesto o cómo éste se reproduce y sus diferencias con los demás formatos de audio, como lo son los archivos .wav o .au. Este formato presenta una enorme cantidad de características que a continuación veremos cómo se desarrolla este tema: el MP3.

Autores: Los Tornados

1 Compresión de Audio.

1.1 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 estudiables 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 lo 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 lo 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.

1.2 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 (ver apéndice 2), 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, 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 lo 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.

1.3 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 sub-bandas. 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 un 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 video. Veremos MPEG Audio como ejemplo de un sistema práctico SBC.

2 MPEG-1 en detalle.

Tras haber visto la introducción que figura en los documentos ISO, podemos pasar a analizar en detalle el funcionamiento del sistema. A continuación haremos hincapié en las características y diferencias entre los tres esquemas de MPEG-1.

La codificación:

1. El banco de filtros: realiza el mapeado del dominio del tiempo al de la frecuencia. Existen dos tipos: el polifase y el híbrido polifase/MDCT. Estos bancos proporcionan tanto la separación en frecuencia para el codificador como los filtros de reconstrucción del decodificador. Las muestras de salida del banco están cuantizadas.
2. El modelo psicoacústico: calcula el nivel a partir del cual el ruido comienza a ser perceptible, para cada banda. Este nivel se utiliza en el bloque de asignación de bit/ruido para determinar la cuantización y sus niveles. De nuevo, se utilizan dos diferentes. En ambos, los datos de salida forman el SMR (signal-to-mask ratio) para cada banda o grupo de bandas.
3. Asignación de bit/ruido: examina tanto las muestras de salida del banco de filtros como el SMR proporcionado por el modelo psicoacústico, y ajusta la asignación de bit o ruido, según el esquema utilizado, para satisfacer simultáneamente los requisitos de tasa de bits y de enmascaramiento.
4. El formateador de bitstream: toma las muestras cuantizadas del banco de filtros, junto a los datos de asignación de bit/ruido y otra información lateral para formar la trama.

Los tres esquemas utilizan diferentes algoritmos para cumplir estas especificaciones:

Esquema I:

- El mapeado tiempo-frecuencia se realiza con un banco de filtros polifase con 32 sub-bandas. Los filtros polifase consisten en un conjunto de filtros con el mismo ancho de banda con interrelaciones de fase especiales que ofrecen una implementación eficiente del filtro sub-banda. Se denomina filtro sub-banda al que cubre todo el rango de frecuencias deseado. En general, los filtros polifase combinan una baja complejidad de computación con un diseño flexible y múltiples opciones de implementación.

- El modelo psicoacústico utiliza una FFT (Fast Fourier Transform) de 512 puntos para obtener información espectral detallada de la señal. El resultado de la aplicación de la FFT se utiliza para determinar los enmascaramientos en la señal, cada uno de los cuales produce un nivel de enmascaramiento, según la frecuencia, intensidad y tono. Para cada sub-banda,

los niveles individuales se combinan y forman uno global, que se compara con el máximo nivel de señal en la banda, produciendo el SMR que se introduce en el cuantizador.

- El bloque de cuantización y codificación examina las muestras de cada sub-banda, encuentra el máximo valor absoluto y lo cuantiza con 6 bits. Este valor es el factor de escala de la sub-banda. A continuación se determina la asignación de bits para cada sub-banda minimizando el NMR (noise-to-mask ratio) total. Es posible que algunas sub-bandas con un gran enmascaramiento terminen con cero bits, es decir, no se codificará ninguna muestra. Por último las muestras de sub-banda se cuantizan linealmente según el número de bits asignados a dicha sub-banda concreta.

- El trabajo del empaquetador de trama es sencillo. La trama, según la definición ISO, es la menor parte del bitstream decodificable por sí misma. Cada trama empieza con una cabecera para sincronización y diferenciación, así como 16 bits opcionales de CRC para detección y corrección de errores. Se emplean, para cada sub-banda, 4 bits para describir la asignación de bits y otros 6 para el factor de escala. El resto de bits en la trama se utilizan para la información de samples, 384 en total, y con la opción de añadir cierta información adicional. A 48 Khz., cada trama lleva 8 MS de sonido.

Esquema II:

- El mapeado de tiempo-frecuencia es idéntico al del esquema I.

- El modelo psicoacústico es similar, salvo que utiliza una FFT de 1024 puntos para obtener mayor resolución espectral. En los demás aspectos, es idéntico.

- El bloque de cuantización y codificación también es similar, generando factores de escala de 6 bits para cada sub-banda. Sin embargo, las tramas del esquema II son tres veces más largas que las del esquema I, de forma que se concede a cada sub-banda tres factores de escala, y el codificador utiliza uno, dos o los tres, según la diferencia que haya entre ellos. La asignación de bits es similar a la del esquema I.

- El formateador de trama: la definición ISO de trama es la misma que en el punto anterior. Utiliza la misma cabecera y estructura de CRC que el esquema I. El número de bits que utilizan para describir la asignación de bits varía con las sub-bandas: 4 bits para las inferiores, 3 para las medias y dos para las superiores, adecuándose a las bandas críticas. Los factores de escala se codifican junto a un número de dos bits que indica si se utiliza uno, dos o los tres. Las muestras de sub-banda se cuantizan y a continuación se asocian en grupos de tres, llamados gránulos. Cada uno se codifica con una palabra clave, lo que permite

interceptar mucha más información redundante que en el esquema I. Cada trama contiene, pues, 1152 muestras PCM. A 48 Khz. cada trama lleva 24 MS de sonido.

Esquema III:

El esquema III es sustancialmente más complicado que los dos anteriores e incluye una serie de mejoras cuyo análisis resultaría desbordante, de manera que no entraremos en tantos detalles. Su diagrama de flujos es conceptualmente semejante al visto para los otros dos esquemas, salvo que se realizan múltiples iteraciones para procesar los datos con el mayor nivel de calidad en un cierto tiempo, lo cual complica su diseño hasta el punto de que los diagramas ISO ocupan decenas de páginas.

El mapeado de tiempo-frecuencia añade un nuevo banco de filtros, el DCT (Discrete Cosine Transform), que con el polifase forman el denominado filtro híbrido. Proporciona una resolución en frecuencia variable, 6x32 o 18x32 sub-bandas, ajustándose mucho mejor a las bandas críticas de las diferentes frecuencias.

El modelo psicoacústico es una modificación del empleado en el esquema II, y utiliza un método denominado predicción polinómica. Incluye los efectos del enmascaramiento temporal.

El bloque de cuantización y codificación también emplea algoritmos muy sofisticados que permiten tramas de longitud variable. La gran diferencia con los otros dos esquemas es que la variable controlada es el ruido, a través de bucles iterativos que lo reducen al mínimo posible en cada paso.

El formateador de trama: la definición de trama para este esquema según ISO varía respecto de la de los niveles anteriores: "mínima parte del bitstream decodificable mediante el uso de información principal adquirida previamente". Las tramas contienen información de 1152 muestras y empiezan con la misma cabecera de sincronización y diferenciación, pero la información perteneciente a una misma trama no se encuentra generalmente entre dos cabeceras. La longitud de la trama puede variarse en caso de necesidad. Además de tratar con esta información, el esquema III incluye codificación Huffman de longitud variable, un método de codificación entrópica que sin pérdida de información elimina redundancia. Los métodos de longitud variable se caracterizan, en general, por asignar palabras cortas a los eventos más frecuentes, dejando las largas para los más infrecuentes.

La decodificación:

Es mucho más sencilla que la codificación, de manera que con lo ya comentado en partes anteriores basta para seguir los siguientes diagramas ISO que incluyen algunas notas aclaratorias al margen que no forman parte de las figuras originales de la norma.

3 Aplicaciones del estándar MPEG-1.

Ya tenemos una idea medianamente clara de qué es y cómo funciona, pero ¿para qué sirve emplear tiempo y dinero en comprimir el sonido? Ya hemos visto los diferentes ratios de compresión que alcanzan los tres esquemas:
El esquema-1 obtiene la mayor calidad de sonido a 384 kbps. Las aplicaciones para las que resulta más útil son las relacionadas con la grabación, tanto en cinta como disco duro o discos magneto-ópticos, que aceptan esta tasa de bits sin problemas.
El esquema-2 produce sus mejores resultados de calidad a 256 kbps, pero se mantiene en un nivel aceptable hasta los 64 kbps. Esto hace que se utilice en transmisión de audio, televisión, grabación profesional o doméstica y productos multimedia.
Ciertamente, el mejor miembro de la familia es el esquema-3. Para una determinada calidad de sonido ofrece la menor tasa de bits y viceversa, fijando la tasa de bits ofrece la mejor calidad posible.

Conexiones musicales vía ISDN.

Las redes telefónicas digitales (ISDN = Integrated Services Digital Network) ofrecen servicios seguros de conexión con dos canales de datos de 64 kbps por adaptador; en otras redes los canales son de 56 kbps, pero en ambas los costes de transmisión son similares a las líneas telefónicas tradicionales, analógicas, que permiten un máximo de 33’6 kbps (vía módem). Con el esquema-3 una conexión de banda estrecha ISDN de bajo costo permite transmitir sonido con calidad CD. Los estudios de sonido y estaciones de transmisión se benefician de la posibilidad de la "música por teléfono" de varias maneras. Se ahorra dinero, pues sólo se paga el tiempo de transmisión, a diferencia de la línea telefónica y únicamente se emplea un pequeño conector ISDN para cada canal. Los programas pueden aumentar su atractivo, ofreciendo tomas de alta calidad y noticias en directo sin la pérdida de calidad del sonido telefónico. Aparecen nuevos campos, como el Estudio Virtual, donde artistas en distintas localidades pueden tocar y grabar juntos sin necesidad de viajar hasta el estudio en sí.

Transmisión digital por satélite

Actualmente se encuentra en plena construcción un sistema de transmisión de sonido digital a escala mundial por satélite. El nombre del proyecto es Worldstar y utilizará tres satélites en órbita geoestacionaria, llamados AfriStar1 (21 Este), CaribStar1 (95 Oeste) y AsiaStar1 (105 Este), esperándose el lanzamiento del primero a mediados del año 1998 y partiendo los demás en los siguientes doce meses. Cada uno está equipado con tres canales de conexión que se pueden multiplexar en hasta 96 subcanales de 16 kbps. Estos son combinables dinámicamente, de manera que se pueden agrupar para formar canales de hasta 128 kbps de capacidad, codificados con el esquema-3. Así, se pueden utilizar cuatro subcanales para formar uno de 64 kbps para transmitir un concierto y al finalizar, utilizar cada uno de ellos para enviar las noticias en cuatro idiomas diferentes.
La empresa responsable del proyecto, Worldspace, ofrece canales en sus tres satélites y ha firmado acuerdos con Voice of America, Radio Nederland, Kenya Broadcasting Corporation, National Broadcasting of Ghana, National Broadcasting of Zimbabwe, New Sky Media of Korea y RCN of Columbia, sumando en total un millón de dólares en inversiones. Alcatel Espace, de Francia, se encarga tanto de la contratación del lanzamiento, como del equipo de comunicaciones. Los receptores se han diseñado buscando la máxima simplicidad con los resultados más efectivos. Se han previsto dos millones de estas unidades, que apenas requerirán sintonización y serán totalmente automáticas. El chip principal de estos sistemas ha sido fabricado por ITT Intermetall con tecnología DSP y su nombre es "MAS 3503 C"

Audio en Internet

Como es sabido, Internet es una red mundial de conmutación de paquetes con cientos de miles de máquinas unidas entre sí por medio de varios sistemas de comunicaciones. Los proveedores profesionales normalmente acceden a la red a través de enlaces con un ancho de banda muy elevado (hasta 2 Gbps). El consumidor doméstico, sin embargo, utiliza canales de bajo costo y ancho de banda limitado (ISDN de 64 kbps o conexión telefónica de 28’8 kbps). La tasa de transmisión efectiva varía en función del uso de la parte de la red accedida, situándose en algún punto entre cero y la máxima capacidad del módem.
Sin la codificación de audio, obtener ficheros de audio sin comprimir de un servidor remoto llevaría a unos tiempos de transmisión simplemente inaceptables. Por ejemplo, suponiendo que se alcanza la tasa de 28’8 kbps (lo cual es bastante optimista) una pista de CD de sólo tres minutos tardaría más de dos horas en recibirse. Por lo tanto, el audio en Internet exige un método de codificación que ofrezca la mayor calidad posible, a la vez que permita la decodificación en tiempo real para un amplio número de plataformas sin necesidad de ampliar el hardware, aunque incluya esta

posibilidad como elemento de hipotéticas tarjetas de ampliación. Por supuesto, la elección es el esquema-3. Hay varios reproductores en tiempo real, como el Winplay3, que ofrecen el servicio requerido.

En la práctica, las expectativas se han cumplido con creces, de tal manera que el fenómeno MP3 está en plena expansión en la telaraña mundial. Ya hay innumerables servidores que ofrecen diferentes piezas en el formato esquema-3 (ficheros de extensión .MP3), de los cuales forman parte tanto aficionados como casas de grabación y grupos independientes. Además, se incorporan temas en este formato a las páginas WEB como elemento para incrementar su atractivo, de forma similar a cómo se venía haciendo con el MIDI, salvando la barrera de las muy inferiores posibilidades de éste.

Llegados a este punto, hay que señalar la importancia de respetar los copyrights a la hora de incluir temas de música en un servidor, así como al almacenarlos en disco duro o CD-ROM. La perspectiva de duplicar la capacidad de los CDs tradicionales resulta sumamente interesante a la comunidad informática, y dado el auge de las grabadoras domésticas puede decirse que el mercado pirata de CDs conteniendo las discografías al completo de diversos grupos o compositores es una realidad, sea con ánimo de lucro o no. El más que previsible auge del DVD-ROM como estándar en el futuro cercano no supone sino un agravamiento del problema.

Las aplicaciones legales que se conocen hasta ahora son, por ejemplo, las de Opticom y Cerberus Sound. La primera ofrece soluciones para que las casas ofrezcan a los clientes audio por demanda, enviando los temas seleccionados al ordenador remoto del usuario. Cerberus se dedica a la comercialización directa de estos temas como un sistema más de venta electrónica. Asimismo se avanza en el concepto de Internet Radio, dado que se obtiene calidad superior a la de la onda corta con un ancho de banda tan escaso como 16 kbps. Opticom de nuevo está a la cabeza en este campo, junto a Telos, compañía que asociada con Apple presentó en Septiembre del 96 la tecnología Audioactive. Por último, el gigante Microsoft anunció en Diciembre de ese mismo año su intención de incluir el esquema-3 como parte de la tecnología multimedia Netserver.

Conclusión

La realización de este trabajo nos permitió conocer en su máxima expresión el formato mp3, del cual reconocimos que detrás de un formato mundialmente conocido y distribuido se encontraba un proceso impresionante. El cual incluía una variación de procesos de compresión y/o modificación de información totalmente asombrosa. Conocimos la historia de este formato, cómo y por quién fue creado, y sus diferentes fases. Este formato mundialmente conocido posee infinidades de modificaciones y características propias, la cuales se han ganado un prestigioso lugar dentro de la informática. Utilizados principalmente para la reproducción de música este formato ha llegado a

usarse en todas las partes del mundo, contando con una increíble cantidad de software disponible para realizar diferentes actividades con él, tanto como bajarlos directamente de Internet, así como reproducirlos y modificarlos. Este formato se encuentra en la cima y no muestra señales de decaer. Pero nadie sabe qué sucederá, dado que muy poco pueden saber realmente que pasará.

Bibliografía

Guías de informática del diario Clarín.

- Manual De multimedia
- Curso integral de informática

www.mp3.com

www.hispanmp3.com

www.endads.com

www.mpeg.org

Indice

Pág