Capítulo 6

Audio

hat
ideweb
ork
email

Capítulo 6 - Archivos de audio

Un formato de archivo audio es un contenedor multimedia que guarda una grabación de audio (música, discurso, etc.). Lo que hace a un archivo distinto del otro son sus propiedades; cómo se almacenan los datos, sus capacidades de reproducción, y cómo puede utilizarse el archivo en un sistema de administración de archivos (etiquetado).

Existen diferentes tipos de formato según la compresión del audio.

Por un lado hay formatos de audio sin compresión como es el caso de WAV, y por otro hay formatos de audio con pérdida y formatos de audio sin pérdida.

Codec de audio

Un códec de audio es un tipo de códec específicamente diseñado para la compresión y descompresión de señales de sonido audible para el ser humano. Por ejemplo, música o conversaciones.
Los códec de audio cumplen fundamentalmente la función de reducir la cantidad de datos digitales necesarios para reproducir una señal auditiva. Lo que comúnmente se denomina "compresión de datos", pero aplicado a un fin muy concreto.

Por ello, existen fundamentalmente dos aplicaciones de los códec de audio:

- Almacenamiento: útil para reproductores multimedia que pueden reproducir sonido almacenado, por ejemplo, en un disco duro, CD-ROM o tarjeta de memoria.

- Transmisión: útil para implementar redes de videoconferencia y Telefonía IP.

Los códecs de audio se caracterizan por los siguientes parámetros:

1. Número de canales: un flujo de datos codificado puede contener una o más señales de audio simultáneamente. De manera que puede tratarse de audiciones "mono" (un canal), "estéreo" (dos canales, lo más habitual) o multicanal. Los códec de audio multicanal se suelen utilizar en sistemas de entretenimiento "cine en casa" ofreciendo seis (5.1) u ocho (7.1) canales.

2. Frecuencia de muestreo: de acuerdo con el teorema de Nyquist, determina la calidad percibida a través de la máxima frecuencia que es capaz de codificar, que es precisamente la mitad de la frecuencia de muestreo. Por tanto, cuanto mayor sea la frecuencia de muestreo, mayor será la fidelidad del sonido obtenido respecto a la señal de audio original. Por ejemplo, para codificar sonido con calidad CD nunca se usan frecuencias de muestreo superiores a 44,1 kHz, ya que el oído humano no es capaz de escuchar frecuencias superiores a 22 kHz.

3. Número de bits por muestra. Determina la precisión con la que se reproduce la señal original y el rango dinámico de la misma. Se suelen utilizar 8 (para un rango dinámico de hasta 45 dB), 16 (para un rango dinámico de hasta 90 dB como el formato CD) o 24 bits por muestra (para 109 a 120 dB de rango dinámico). El más común es 16 bits.

4. Pérdida. Algunos códecs pueden eliminar frecuencias de la señal original que, teóricamente, son inaudibles para el ser humano. De esta manera se puede reducir la frecuencia de muestreo. En este caso se dice que es un códec con pérdida o lossy codec (en inglés). En caso contrario se dice que es un códec sin pérdida o lossless codec (en inglés).

El parámetro tasa de bits o bit-rate es el número de bits de información que se procesan por unidad de tiempo, teniendo en cuenta la frecuencia de muestreo resultante, la profundidad de la muestra en bits y el número de canales. A causa de la posibilidad de utilizar compresión (con o sin pérdidas), la tasa de bits no puede deducirse directamente de los parámetros anteriores

Lista de códecs de audio

Sin pérdida: Wav

WAV (o WAVE), apócope de WAVEform audio format, es un formato de audio digital normalmente sin compresión de datos desarrollado y propiedad de Microsoft y de IBM que se utiliza para almacenar sonidos en el PC, admite archivos mono y estéreo a diversas resoluciones y velocidades de muestreo, su extensión es .wav. Es una variante del formato RIFF (Resource Interchange File Format, formato de fichero para intercambio de recursos), método para almacenamiento en "paquetes", y relativamente parecido al IFF y al formato AIFF usado por Macintosh. El formato toma en cuenta algunas peculiaridades de la CPU Intel, y es el formato principal usado por Windows.

A pesar de que el formato WAV puede soportar casi cualquier códec de audio, se utiliza principalmente con el formato PCM (no comprimido) y al no tener pérdida de calidad puede ser usado por profesionales. Para tener calidad CD de audio se necesita que el sonido se grabe a 44100 Hz y a 16 bits. Por cada minuto de grabación de sonido se consumen unos 10 megabytes de espacio en disco. Una de sus grandes limitaciones es que solo se puede grabar un archivo de hasta 4 gigabytes, que equivale aproximadamente a 6,6 horas en calidad de CD de audio. Es una limitación propia del formato, independientemente de que el sistema operativo donde se utilice sea MS Windows u otro distinto, y se debe a que en la cabecera del fichero se indica la longitud del mismo con un número entero de 32 bit, lo que limita el tamaño del fichero a 4 GB.

En Internet no es popular, fundamentalmente porque los archivos sin compresión son muy grandes. Son más frecuentes los formatos comprimidos con pérdida, como el MP3 o el Ogg Vorbis. Como éstos son más pequeños la transferencia a través de Internet es mucho más rápida. Además existen códecs de compresión sin pérdida más eficaces como Apple Lossless o FLAC.

Con pérdida

MP3

MPEG-1 Audio Layer 3, más conocido como MP3, es un formato de audio digital comprimido con pérdida desarrollado por el Moving Picture Experts Group (MPEG) para formar parte de la versión 1 (y posteriormente ampliado en la versión 2) del formato de vídeo MPEG. El mp3 estándar es de 44 kHz y un bitrate de 128 kbps por la relación de calidad/tamaño. Su nombre es el acrónimo de MPEG-1 Audio Layer 3 y el término no se debe confundir con el de reproductor MP3.

Este formato fue desarrollado principalmente por Karlheinz Brandenberg, director de tecnologías de medios electrónicos del Instituto Fraunhofer IIS, perteneciente al Fraunhofer-Gesellschaft - red de centros de investigación alemanes - que junto con Thomson Multimedia controla el grueso de las patentes relacionadas con el MP3. La primera de ellas fue registrada en 1986 y varias más en 1991. Pero no fue hasta julio de 1995 cuando Brandenburg usó por primera vez la extensión .mp3 para los archivos relacionados con el MP3 que guardaba en su ordenador. Un año después su instituto ingresaba en concepto de patentes 1,2 millones de euros. Diez años más tarde esta cantidad ha alcanzado los 26,1 millones.

El formato MP3 se convirtió en el estándar utilizado para streaming de audio y compresión de audio de alta calidad (con pérdida en equipos de alta fidelidad) gracias a la posibilidad de ajustar la calidad de la compresión, proporcional al tamaño por segundo (bitrate), y por tanto el tamaño final del archivo, que podía llegar a ocupar 12 e incluso 15 veces menos que el archivo original sin comprimir.

Fue el primer formato de compresión de audio popularizado gracias a Internet, ya que hizo posible el intercambio de ficheros musicales. Los procesos judiciales contra empresas como Napster y AudioGalaxy son resultado de la facilidad con que se comparten este tipo de ficheros.

Tras el desarrollo de reproductores autónomos, portátiles o integrados en cadenas musicales (estéreos), el formato MP3 llega más allá del mundo de la informática.

A principios de 2002 otros formatos de audio comprimido como Windows Media Audio y Ogg Vorbis empiezan a ser masivamente incluidos en programas, sistemas operativos y reproductores autónomos, lo que hizo prever que el MP3 fuera paulatinamente cayendo en desuso, en favor de otros formatos, como los mencionados, de mucha mejor calidad. Uno de los factores que influye en el declive del MP3 es que tiene patente. Técnicamente no significa que su calidad sea inferior ni superior, pero impide que la comunidad pueda seguir mejorándolo y puede obligar a pagar por la utilización de algún códec, esto es lo que ocurre con los reproductores de MP3. Aun así, a finales de 2009, el formato mp3 continua siendo el más usado y el que goza de más éxito.
Detalles técnicos

En esta capa existen varias diferencias respecto a los estándares MPEG-1 y MPEG-2, entre las que se encuentra el llamado banco de filtros híbrido que hace que su diseño tenga mayor complejidad. Esta mejora de la resolución frecuencial empeora la resolución temporal introduciendo problemas de pre-eco que son predichos y corregidos. Además, permite calidad de audio en tasas tan bajas como 64 kbps.

Banco de filtros

El banco de filtros utilizado en esta capa es el llamado banco de filtros híbrido polifase/MDCT. Se encarga de realizar el mapeado del dominio del tiempo al de la frecuencia tanto para el codificador como para los filtros de reconstrucción del decodificador. Las muestras de salida del banco están cuantizadas y proporcionan una resolución en frecuencia variable, 6x32 o 18x32 subbandas, ajustándose mucho mejor a las bandas críticas de las diferentes frecuencias. Usando 18 puntos, el número máximo de componentes frecuenciales es: 32 x 18 = 576. Dando lugar a una resolución frecuencial de: 24000/576 = 41,67 Hz (si fs = 48 kHz.). Si se usan 6 líneas de frecuencia la resolución frecuencial es menor, pero la temporal es mayor, y se aplica en aquellas zonas en las que se espera efectos de pre-eco (transiciones bruscas de silencio a altos niveles energéticos).

La Capa III tiene tres modos de bloque de funcionamiento: dos modos donde las 32 salidas del banco de filtros pueden pasar a través de las ventanas y las transformadas MDCT y un modo de bloque mixto donde las dos bandas de frecuencia más baja usan bloques largos y las 30 bandas superiores usan bloques cortos. Para el caso concreto del MPEG-1 Audio Layer 3 (que concretamente significa la tercera capa de audio para el estandar MPEG-1) especifica cuatro tipos de ventanas: (a) NORMAL, (b) transición de ventana larga a corta (START), (c) 3 ventanas cortas (SHORT), y (d) transición de ventana corta a larga (STOP).

El modelo psicoacústico

La compresión se basa en la reducción del margen dinámico irrelevante, es decir, en la incapacidad del sistema auditivo para detectar los errores de cuantificación en condiciones de enmascaramiento. Este estándar divide la señal en bandas de frecuencia que se aproximan a las bandas críticas, y luego cuantifica cada subbanda en función del umbral de detección del ruido dentro de esa banda. El modelo psicoacústico es una modificación del empleado en el esquema II, y utiliza un método denominado predicción polinómica. Analiza la señal de audio y calcula la cantidad de ruido que se puede introducir en función de la frecuencia, es decir, calcula la “cantidad de enmascaramiento” o umbral de enmascaramiento en función de la frecuencia.

El codificador usa esta información para decidir la mejor manera de gastar los bits disponibles. Este estándar provee dos modelos psicoacústicos de diferente complejidad: el modelo I es menos complejo que el modelo psicoacústico II y simplifica mucho los cálculos. Estudios demuestran que la distorsión generada es imperceptible para el oído experimentado en un ambiente óptimo desde los 256 kbps y en condiciones normales. Para el oído no experimentado, o común, con 128 kbps o hasta 96 kbps basta para que se oiga "bien" (a menos que se posea un equipo de audio de alta calidad donde se nota excesivamente la falta de graves y se destaca el sonido de "fritura" en los agudos). En personas que escuchan mucha música o que tienen experiencia en la parte auditiva, desde 192 o 256 kbps basta para oír bien. La música que circula por Internet, en su mayoría, está codificada entre 128 y 192 kbps.

Codificación y cuantificación

La solución que propone este estándar en cuanto a la repartición de bits o ruido, se hace en un ciclo de iteración que consiste de un ciclo interno y uno externo. Examina tanto las muestras de salida del banco de filtros como el SMR (signal-to-mask ratio) proporcionado por el modelo psicoacústico, y ajusta la asignación de bits o ruido, según el esquema utilizado, para satisfacer simultáneamente los requisitos de tasa de bits y de enmascaramiento. Dichos ciclos consisten en:

Ciclo interno

El ciclo interno realiza la cuantización no-uniforme de acuerdo con el sistema de punto flotante (cada valor espectral MDCT se eleva a la potencia 3/4). El ciclo escoge un determinado intervalo de cuantización y, a los datos cuantizados, se les aplica codificación de Huffman en el siguiente bloque. 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.

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 con los datos previamente calculados en el análisis psicoacústico. El ciclo externo termina cuando una de las siguientes condiciones se cumple:

- Ninguna de las bandas del factor de escala tiene mucho ruido.
- Si la siguiente iteración amplifica una de las bandas más de lo permitido.
- Todas las bandas han sido amplificadas al menos una vez.

Empaquetado o formateador de bitstream

Este bloque toma las muestras cuantificadas del banco de filtros, junto a los datos de asignación de bits/ruido y almacena el audio codificado y algunos datos adicionales en las tramas. Cada trama contiene información de 1152 muestras de audio y consiste de un encabezado, de los datos de audio junto con el chequeo de errores mediante CRC y de los datos auxiliares (estos dos últimos opcionales). El encabezado nos describe cuál capa, tasa de bits y frecuencia de muestreo se están usando para el audio codificado. Las tramas empiezan con la misma cabecera de sincronización y diferenciación y su longitud puede variar. Además de tratar con esta información, también incluye la codificación Huffman de longitud variable, un método de codificación entrópica que sin pérdida de información elimina redundancia. Actúa al final de la compresión para codificar la información. 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.

Estructura de un fichero MP3

Un fichero Mp3 se constituye de diferentes frames MP3 que a su vez se componen de una cabecera Mp3 y los datos MP3. Esta secuencia de datos es la denominada "stream elemental". Cada uno de los Frames son independientes, es decir, una persona puede cortar los frames de un fichero MP3 y después reproducirlos en cualquier reproductor MP3 del Mercado. La cabecera consta de una palabra de sincronismo que es utilizada para indicar el principio de un frame válido. A continuación siguen una serie de bits que indican que el fichero analizado es un fichero Standard MPEG y si usa o no la capa 3. Después de todo esto, los valores difieren dependiendo del tipo de archivo MP3. Los rangos de valores quedan definidos en la ISO/IEC 11172-3.

Otros Codecs con perdida.

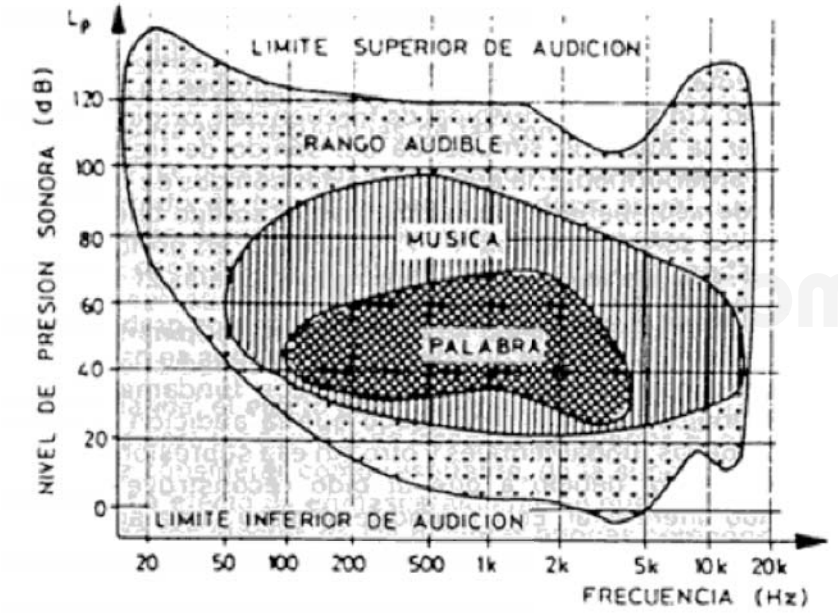
- Advanced Audio Coding (AAC).
- Ogg Vorbis
- WMA (Windows Media Audio).
- Musepack
- AC3 (Dolby Digital A/52).
- DTS (Digital Theather Systems).
- ADPCM.
- ADX (usado en videojuegos).
- ATRAC (Adaptive TRansform Acoustic Coding).
- Perceptual Audio Coding
- TwinVQ

El Sonido

Introducción

El sonido es una vibración de partículas en un medio, gaseoso, líquido o sólido.

Si consideramos sonido, como las vibraciones que el oído humano es capaz de percibir, estas vibraciones se limitan a frecuencias comprendidas entre 20 y 20.000 Hz.



Parámetros básicos del Sonido

Existen tres parámetros básicos que identifican el sonido:

- la amplitud,
- la frecuencia, y
- el timbre

El conocimiento de estos parámetros nos permitirán obtener un producto homogéneo con la máxima calidad según el soporte final que queramos obtener.

Amplitud

La cantidad de energía que tiene una onda sonora representa su amplitud

La amplitud o intensidad del sonido se mide en dB, que es una medida relativa.

La medida de la amplitud, tiene ciertos problemas debidos a que el oído no se comporta de una forma lineal, sino que lo hace de forma logarítmica, es decir no interpreta las diferencias, sino la relación.
En base a este criterio, se eligen unos niveles fijos de referencia, así tenemos que:

dB = 20 log V1/V2, (siendo V2 el nivel de referencia)

Puede ser interesante recordar las siguientes relaciones

0 dB	V1 = V2
+6 dB	V1 = V2 x 2
-6 dB	V1 = V2 / 2
-20 dB	V1 = V2 / 10

El oído humano es capaz de adaptarse y compensar las diferencias de nivel y seguir apreciando la diferencia que existe entre diversos sonidos.

Esta característica del oído humano es difícil de trasladar a los equipos electrónicos.

Así tenemos que si estamos digitalizado una señal con una resolución de 8 bit esto nos indica que el máximo número de niveles que podemos discriminar es 256. Si ajustamos la entrada de nuestro equipo a la intensidad de sonido de un concierto de rock, nos encontraremos que cuando alguien hable con un nivel de conversación no se le entenderá, en caso contrario si ajustamos a un bajo nivel nos encontraremos que una gran cantidad de sonidos estarán saturados.

Frecuencia

La frecuencia como ya conocemos, se mide en Hz, el ser humano generalmente tiene un rango de audición que va de 20 Hz a 20khz

Habla humana: De acuerdo con estudios realizados, se puede asegurar que las bajas frecuencias (inferior a 500 Hz) afectan poco a la compresión de la palabra, así como las frecuencias superiores a los 4000Hz. Las frecuencias más importantes para la inteligibilidad están entre 500 y 3000Hz.

Aunque las frecuencias superiores a 4000Hz, no afectan a la inteligibilidad, si son necesarias para la correcta reproducción del Timbre, ya que el hombre llega a emitir sonidos hasta los 8000Hz y la mujer hasta los 9000Hz.

Según las componentes de frecuencia del sonido que vayamos a digitalizar, nos indica que frecuencia de muestreo es la que vamos a requerir, si nuestro deseo es recomponer la señal digitalizada con la mayor calidad posible.

Como vimos anteriormente la frecuencia de muestreo debe ser de 2 veces la frecuencia mayor que se desee digitalizar.

Timbre

El sonido en contadas ocasiones se presenta como una frecuencia pura, esto sólo se da en generadores, lo normal es que estas frecuencia vayan acompañadas por sus armónicos,

Estos armónicos y sus características de nivel, etc., son las que nos permiten identificar cada uno de los sonidos. Así podemos diferenciar un instrumento de otro aun y cuando la nota básica sea la misma, etc.
Lo que ocurre cuando muestreamos con una frecuencia baja, que perdemos armónicos y por tanto brillantez.

Estéreo

Existe otra característica del sonido que es el estéreo, nosotros escuchamos en estéreo y esta es una característica muy importante a la hora de realizar la toma del sonido, la edición y la reproducción, ya que afecta al volumen de la información, esta es exactamente el doble que el de una señal monoaural.