

CAPÍTULO 8:

SONIDO

El sonido es un medio percibido de un modo completamente distinto a como se perciben el resto de medios presentes en una producción Multimedia. No solamente se percibe por un órgano distinto, el oído, si no que en su percepción entran a jugar parte fenómenos psicológicos.

Por otro lado, el sonido no siempre está presente en las producciones Multimedia, y de estar presente, siempre debe ser posible desactivarlo, de algún modo, por parte del usuario. También hay que tener en cuenta que ciertos usuarios no pueden escuchar el sonido porque, por ejemplo, sus equipos no cuentan con altavoces.

Finalmente hay que distinguir dos tipos de sonidos en las producciones Multimedia, la voz y el resto de sonidos tales como música.

En este Capítulo aprenderás:

Cual es la naturaleza y principales características del sonido.

Cual es el proceso de digitalización del sonido y de qué factores depende su calidad.

Las técnicas de compresión de ficheros de sonido.

Los principales formatos de ficheros de sonido existentes.

El modo en que sonido, imágenes y animación se deben combinar en una producción Multimedia.

ESQUEMA DEL CAPITULO 8: SONIDO.

8.1 LA NATURALEZA DEL SONIDO.

- 8.1.1 Características de los sonidos.
- 8.1.2 Representación gráfica de los sonidos.

8.2 DIGITALIZACIÓN DEL SONIDO.

- 8.2.1 Muestreo.
- 8.2.2 Cuantización.
- 8.2.3 Calidad del sonido. Tamaño de los ficheros.

8.3 Procesamiento del sonido.

- 8.3.1 Grabación de sonido.
- 8.3.2 Edición del sonido.

8.4 Compresión de ficheros de sonido.

- 8.4.1 Compresión basada en un modelo psicológico perceptual.

8.5 Formatos.

- 8.5.1 Formatos de ficheros.
- 8.5.2 Difusión de sonido.
- 8.5.3 MIDI.

8.6 Combinación de sonido e imagen.

Resumen.

8.1 LA NATURALEZA DEL SONIDO.

El sonido, las ondas sonoras, necesitan de un medio material para propagarse. Este medio puede ser el aire, el agua, los metales, etc. La generación de un sonido provoca cambios de presión en el medio. Una onda sonora es una sucesión de compresiones y rarefacciones del medio. Las compresiones se corresponden con los máximos de la onda y las rarefacciones con los mínimos.

Hasta aquí en cuanto a la naturaleza física del sonido, pero el sonido finalmente percibo, captado por el cerebro humano, es un proceso donde entran en juego no sólo procesos físicos sino también procesos fisiológicos, los derivados de la estimulación de los órganos de la audición, y de procesos psicológicos, los derivados del acto consciente de escuchar un sonido.

De este modo, y como veremos más adelante, el proceso de la audición se convierte al algo más complicado de modelar que el modelo simple de una onda.

8.1.1 Características de los sonidos.

Un sonido rara vez está compuesto por una única frecuencia. Los sonidos de la naturaleza están compuestos por una suma de una onda principal, de frecuencia base o principal, más una serie mas o menos numerosa de armónicos, que son ondas con frecuencias múltiplos enteros de la frecuencia principal.

Tres son las características que nos permiten diferenciar unos sonidos de otros: el tono, el timbre y la intensidad.

El tono se relaciona con la frecuencia base. El tono es la característica que nos permite distinguir, por ejemplo, las notas musicales entre sí; es la característica que nos permite distinguir un Do de un Re o un Re de un Re#.

El timbre se relaciona con los armónicos asociados al tono o frecuencia principal. Son los que dotan de "calidad" al sonido. A través del timbre somos capaces de diferenciar, por ejemplo, la misma nota producida por dos instrumentos musicales diferentes; es la característica que nos permite reconocer y diferenciar la voz de dos tenores, etc.

La intensidad de un sonido se relaciona con la amplitud de la onda, cuanto mayor es la amplitud de la onda mayor es su intensidad. Como comentamos en la introducción del capítulo, el proceso de percepción de sonidos por el cerebro humano es bastante más complicado que el registro de una onda. En particular el oído humano no percibe la intensidad del sonido de forma lineal, es decir, un sonido de amplitud doble que otro no es percibido con el doble de intensidad. La percepción de la intensidad de un sonido por el oído humano sigue un comportamiento logarítmico. Por ello, la escala de medida de intensidad de un sonido se define de forma logarítmica, los llamados decibelios:

$$dB = 10 \log_{10} \frac{a^2}{a_{ref}^2}$$

donde a_{ref} es el valor de la amplitud de referencia con la que se compara. Si se toma como valor de referencia el umbral de audición la Tabla 8.1 muestra algunos ejemplos de intensidades sonoras:

Descripción	Nivel (dB)	Intensidad
Umbral del dolor	130	10^{13}
Concierto heavy metal	120	10^{12}
Martillazos sobre metal	110	10^{11}
Tráfico de vehículos	70	10^7
Conversación normal	60	10^6
Restaurante concurrido	50	10^5
Casa en la ciudad	40	10^4
Iglesia vacía	30	10^3
Estudio de grabación	20	10^2
Umbral de audición	0	1

Tabla 8.1: Distintos niveles de intensidad referidos a la intensidad umbral de audición.

La intensidad de un sonido percibido por el oído humano no depende exclusivamente de la amplitud del sonido si no también de la frecuencia de este, es decir, para que dos sonidos de distintas frecuencias se perciban con la misma intensidad por el oído, la intensidad 'absoluta' medida en decibelios no es la misma. En la Figura 8.1 se muestra el nivel umbral de audición dependiendo de la frecuencia del sonido.

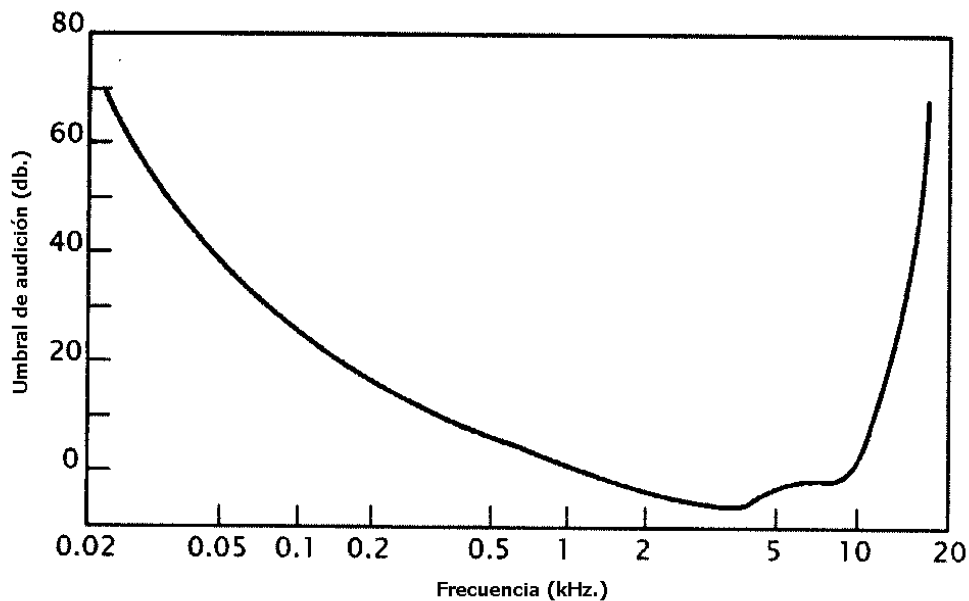


Figura 8.1: Nivel de audición en función de la frecuencia del sonido.

8.1.2 Representación gráfica de los sonidos.

El modo usual de representación de una onda es representando en una grafica el valor de la amplitud para cada instante del tiempo. A esta representación se le llama forma de onda.

En la Figura 8.2 se muestran varios ejemplos de ondas sonoras. En ellos se puede observar que la frecuencia base es siempre de mayor amplitud que el resto de armónicos. Sabiendo cual es la contribución en amplitud de cada uno de los armónicos es sencillo obtener la forma de onda final. Resulta bastante más complicado obtener, a partir de una forma de onda, la contribución de cada uno de los armónicos. El modo de conocer la contribución de cada uno de los armónicos es mediante una transformación de Fourier de la onda. Aquí, no entraremos en los detalles de la transformada de Fourier.

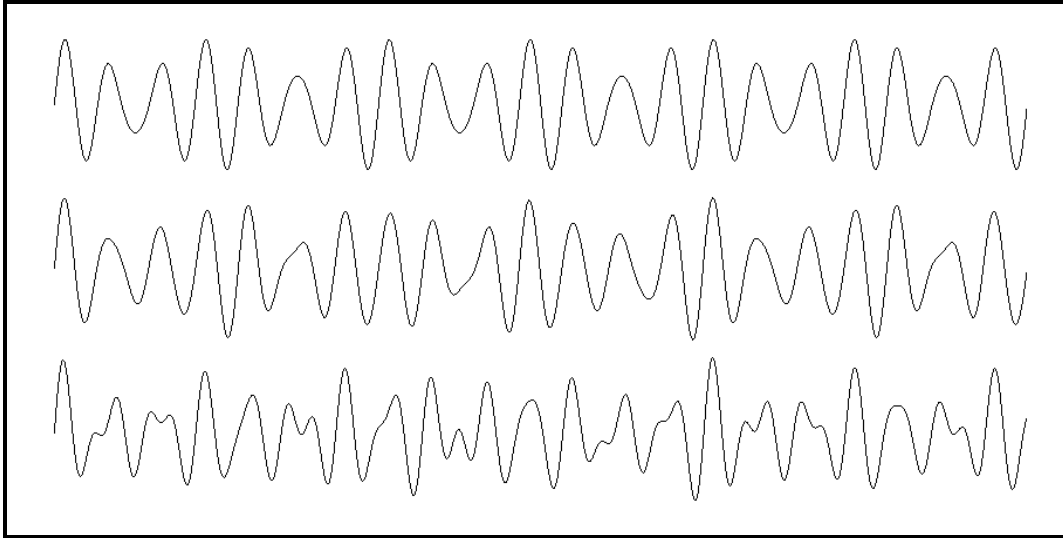


Figura 8.2: Algunas formas de onda. Cada una de ellas está formada por la frecuencia principal y uno, dos o tres armónicos respectivamente.

A partir de la Figura 8.2 resulta evidente que cuantos más armónicos se añada a la frecuencia fundamental más compleja será la forma de onda final, obtenida como la suma de la contribución de todos los armónicos

8.2 DIGITALIZACIÓN DEL SONIDO.

Como ya sabemos, la Naturaleza es eminentemente analógica. Para poder manipular, almacenar y transmitir esta información por medios digitales, necesitamos transformar la información sonora a un formato digital. Como ya vimos en el capítulo 2, al proceso de conversión de una señal analógica a señal digital se le llama digitalización. La digitalización consta de dos partes diferenciadas, el muestreo y la cuantización.

8.2.1 Muestreo.

El muestreo es la acción de tomar muestras. Si de una señal analógica tomamos su valor cada cierto intervalo de tiempo estamos muestreando la señal. En la práctica el intervalo de tiempo utilizado es siempre constante. En la Figura 8.3 vemos un caso de una señal analógica y el muestreo a intervalos regulares de tiempo de esa misma señal.

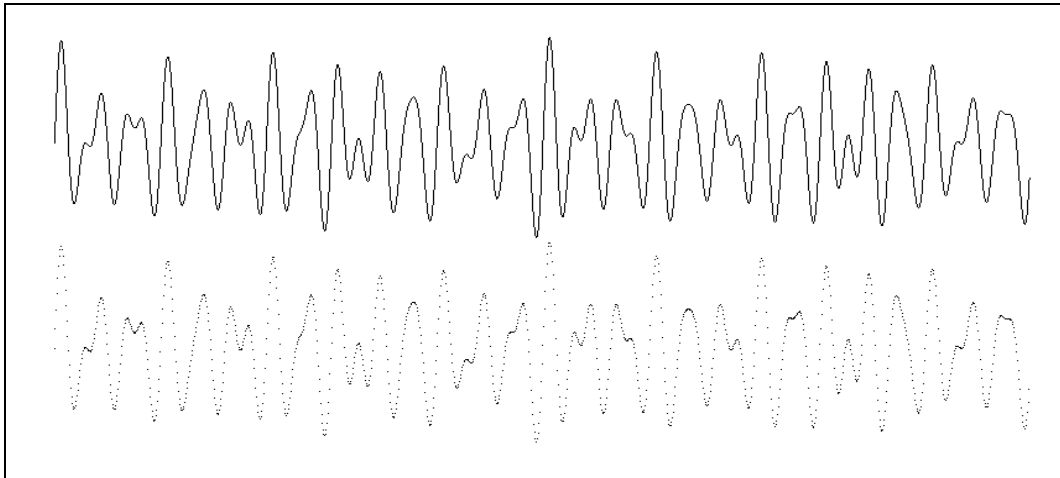


Figura 8.3: Señal analógica y su correspondiente muestreada.

En el caso del sonido la magnitud que se muestrea es la amplitud de la onda en un cierto instante de tiempo. El instante de tiempo ha de ser lo suficientemente 'estrecho' como para que la amplitud no se promedie, en la Figura 8.4 se muestra un ejemplo de una señal muestreada con dos ventanas temporales de distinto tamaño. El caso de los CDs, la ventana temporal es de 200 picosegundos ($200 \cdot 10^{-12}$ s.).

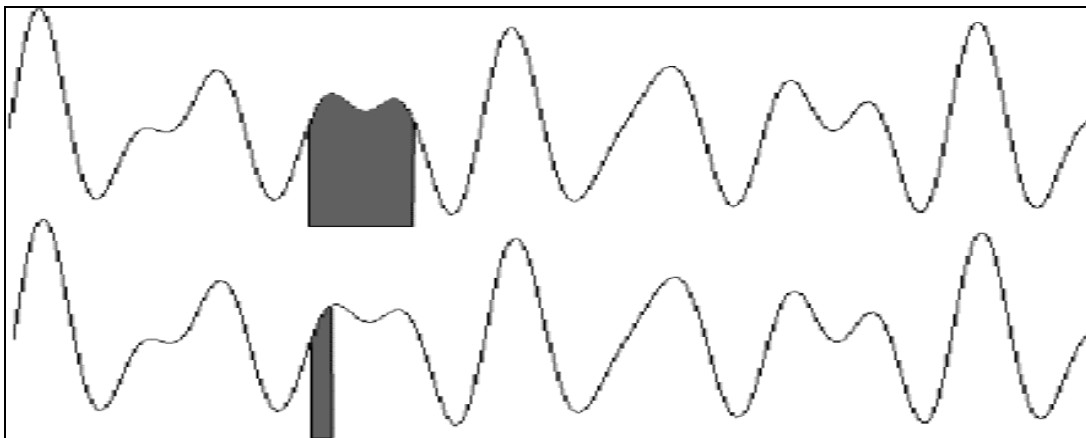


Figura 8.4: Señal muestreada con dos ventanas temporales de distinta anchura.

Como también comentamos en el capítulo 2, no siempre es válida cualquier frecuencia de muestreo. El teorema de muestreo indica que si una señal analógica se quiere reconstruir, a partir de su muestreo, sin pérdida de información, es necesario que la frecuencia de muestreo sea superior al doble de la máxima frecuencia presente en la señal.

El oído humano es capaz de percibir sonidos con frecuencias dentro del rango de los 20 a los 20.000 Hz. Así pues, si queremos digitalizar sonido y que este no presente pérdida de calidad para el oído humano cuando se reconstruya la señal, la frecuencia de muestreo debe ser superior a los 40kHz. En el caso de los CDs de audio, la frecuencia utilizada es de 44'1 kHz. Para las DATs (Digital Audio Tape), la frecuencia de muestreo es de 48 kHz.

Por otro lado, cuando se muestrea una señal analógica el instante en que se capta la amplitud de la señal

Cuando se quiere reconstruir un sonido a partir de la señal digitalizada, se requiere que el muestreo se haya llevado a cabo con un frecuencia fija. No podemos utilizar, por ejemplo, dos señales digitales no homogéneas con frecuencias de muestreo distintas, como por ejemplo de un CD y de una DAT. En este caso la técnica consiste en reconstruir cada una de las señales por separado, sumarlas y digitalizar de nuevo la señal suma. El teorema del muestreo nos garantiza que si la frecuencia de muestreo utilizada en la digitalización final es adecuada no se perderá calidad en la señal. En la Figura 8.5 se muestra un esquema del proceso.

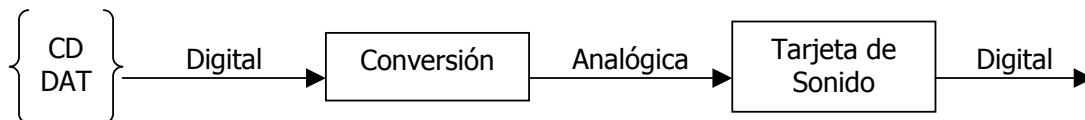


Figura 8.5: Combinación de dos señales digitales sonoras con frecuencias de muestreo distintas.

En la transmisión de un sonido a través de Internet es tan importante que este se transmita con rapidez como su calidad, por ello, la frecuencia típica de muestreo utilizada para sonido a través de Internet es de 22.05 kHz. En el caso de las líneas telefónicas, de nuevo es importante disminuir el flujo de señales y sólo es necesario que se identifique las palabras durante una conversación, la frecuencia de muestreo en las líneas telefónicas es de 11.025 kHz.

No es casualidad que las dos frecuencias anteriores sean divisores enteros de la frecuencia de muestreo utilizada en los CDs. ¿Puedes dar una explicación a este hecho?.

8.2.2 Cuantización.

Cuantizar significa restringir los posibles valores de una variable. Un ejemplo sencillo es restringir los valores de una variable a enteros; la variable no podrá tomar el valor 6'5 ó 3'2, únicamente los valores 6 ó 3. Además, en el caso de información que manejará un ordenador, existen un cierto intervalo, o número de niveles, fuera del cual la variable no puede tomar ningún valor. Por ejemplo, en el caso de las señales de audio para CD sólo existen 65.536 (2^{16}) niveles enteros distintos. La Figura 2.1 muestra una señal digitalizada con 16 niveles distintos. El mínimo número de niveles para que la digitalización sea aceptable es de 256.

Antes de digitalizar un sonido se ha de tener cuidado de que el máximo valor de la señal no cae fuera del intervalo de niveles que se está utilizando, de lo contrario se puede producir el fenómeno del *recortado* (*clipping*) de la señal tal y como se muestra en la Figura 8.6.

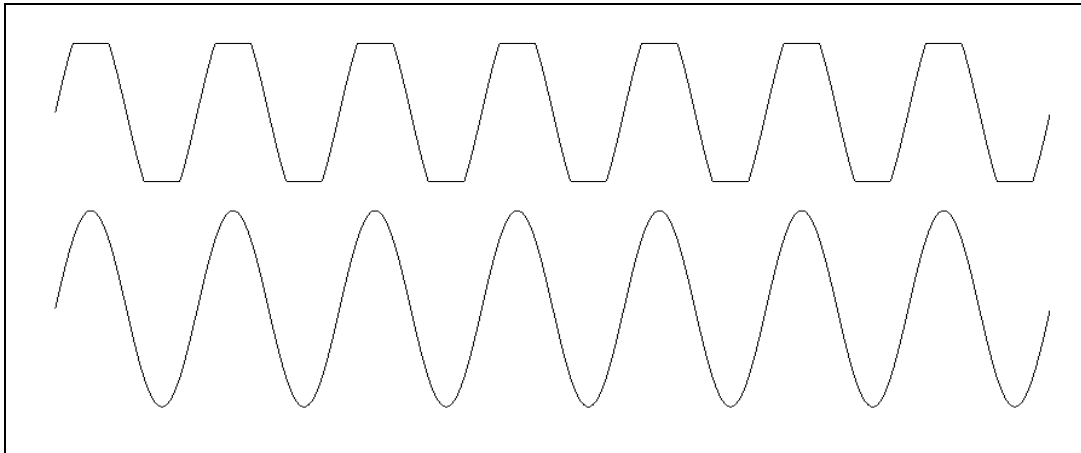


Figura 8.6: Fenómeno del recortado producido al utilizar un intervalo de digitalización demasiado pequeño.

Un fenómeno que puede aparecer al contar con pocos niveles de cuantización es la aparición de dientes de sierra. El resultado final es la aparición de chasquidos durante la reconstrucción del sonido. Este fenómeno se puede atenuar introduciendo un ruido aleatorio antes del proceso de cuantización. En la Figura 8.7 se muestra una señal y su correspondiente cuantizada utilizando pocos niveles; claramente se aprecia la aparición de dientes de sierra. En la parte inferior se introduce un ruido en la señal original antes de digitalizar.

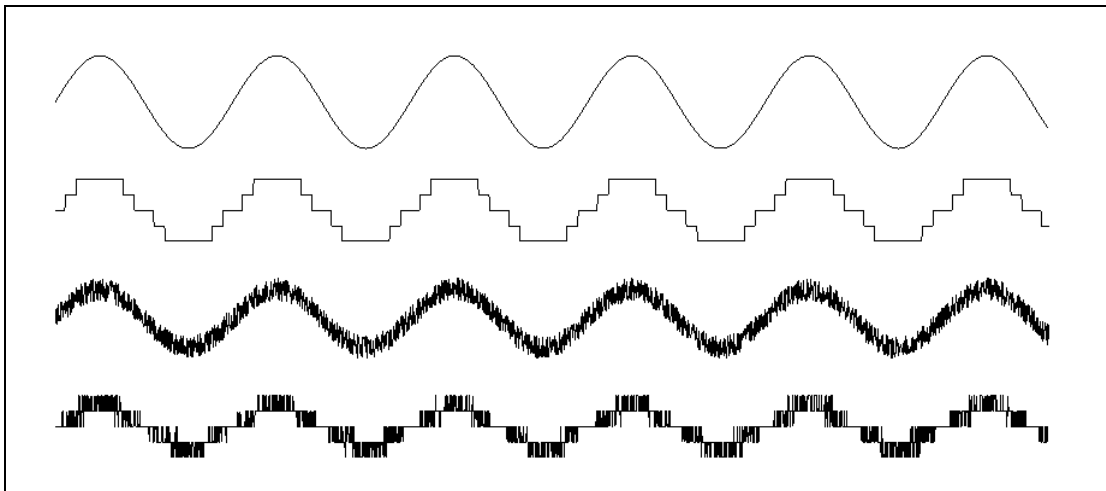


Figura 8.7: Introducción de un ruido aleatorio para evitar el efecto de diente de sierra.

8.2.3 Calidad del sonido. Tamaño de los ficheros.

Como sabemos, la calidad del sonido digital depende de la frecuencia de muestreo que se utilizó durante la fase de muestreo y del número de niveles utilizados en la fase de cuantización. La frecuencia de muestreo utilizada en la digitalización de sonido para CD es de 44'1 kHz con 16 bits para la cuantización. Por lo tanto, un segundo de música en un CD ocupa 86 kb y un minuto aproximadamente 8 Mb. Estas cantidades se duplican si se trabaja con sonido estéreo. Si la señal así obtenida se desea almacenar y reproducir desde un disco duro, nos encontramos con que posiblemente el tamaño de fichero generado excede del tamaño que puede manejar el sistema operativo, o velocidad de lectura no sea la suficientemente rápida.

8.3 Procesamiento del sonido.

Dos son las tareas básicas en el procesamiento del sonido para aplicaciones Multimedia. La primera de ellas es la grabación del sonido. En esta fase se debe actuar lo más cuidadosamente posible, ya que un sonido digital de calidad evitará mucho trabajo durante la fase de edición que es la segunda tarea básica.

8.3.1 Grabación de sonido.

Para evitar obtener una señal digital lo más perecida posible al sonido analógico lo más recomendable es utilizar el material de captura adecuado. Por ejemplo, micrófonos de calidad al igual que las cintas DAT.

El modo más sencillo de capturar sonido digital es utilizar el que ya existe en CDs, MP3, etc. pero esto es ilegal la mayoría de las veces.

8.3.2 Edición del sonido.

Existen numerosas aplicaciones para la edición de sonido. Todas ellas cuentan con una serie de operaciones comunes tales como selección de un intervalo dentro de la forma de onda, borrado de un intervalo, cortar un intervalo, copiar un intervalo, duplicar un intervalo, etc.

Sobre el sonido se pueden realizar tareas encaminadas a aumentar la calidad del sonido o bien a enriquecer la señal con algún tipo de efecto. Al primer grupo pertenecen operaciones como eliminar un sonido de fondo, eliminar chasquidos, filtrado de un determinada frecuencia, etc. En el segundo grupo se encuentran los efectos de eco, repeticiones, *fader in/fader out*, etc.

8.4 Compresión de ficheros de sonido.

En §8.2.3 se dijo que un minuto de sonido estéreo en calidad de CD ocupa alrededor de 10 Mb, una canción de entre tres y cuatro minutos puede llegar ocupar 30 ó 40 Mb. Manejar, y almacenar ficheros de tal tamaño se hace impracticable. Este problema se acrecienta cuando pensamos en transmitir estos ficheros a través de Internet. Se hace absolutamente necesario comprimir estos ficheros.

Un primer intento para comprimir ficheros de sonido consiste en sustituir los silencios, o niveles muy bajos de sonido, por un valor que indique la duración de este sonido. Este método es parecido a lo que se hace en la compresión RLE, no obstante en este caso sí que hay pérdidas porque un silencio casi nunca es una ausencia total de sonido.

Otro método es utilizar una cuantización logarítmica siguiendo una función como la mostrada en la Figura 8.8, de este modo, los valores de sonido poco intenso tienen más niveles en su descripción que los valores altos. Este método está basado en la evidencia de como funciona el oído humano.

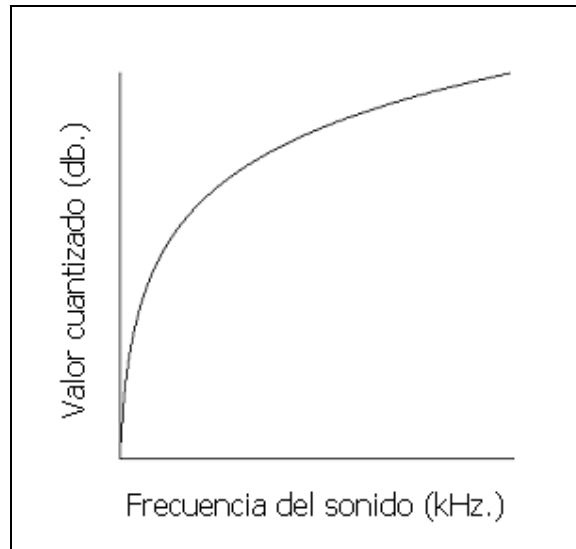


Figura 8.8: Función de cuantización logarítmica.

El proceso inverso es necesario cuando se reconstruye la señal.

Otras funciones utilizadas en la cuantización logarítmica son:

$$\frac{\log(1 + \mu x)}{\log(1 + \mu)}, x \geq 0 \quad \frac{Ax}{1 + \log(A)}, 0 \leq |x| \leq \frac{1}{A} \quad \frac{1 + \log(Ax)}{1 + \log(A)}, \frac{1}{A} \leq |x| \leq 1$$

Finalmente, otro método consiste en almacenar las diferencias entre dos niveles consecutivos, no los valores absolutos de intensidad del sonido.

8.4.1 Compresión basada en un modelo psicológico perceptual.

La audición es el efecto de una serie de procesos fisiológicos y psicológicos que implican tanto al oído como al cerebro, que es donde se produce finalmente la audición.

Una característica del proceso de la audición es que el umbral audible depende de la frecuencia del sonido. Es decir, los sonidos demasiado graves o demasiado agudos necesitan ser más intensos que aquellos que se encuentran en un rango de entre dos y cinco kilo hertzios, para ser escuchados.

La Figura 8.1 muestra cual es el mínimo nivel de intensidad de un sonido para que sea audible en función de su frecuencia. La Figura 8.10 muestra el efecto del enmascaramiento para las frecuencias adyacentes a una dominante.

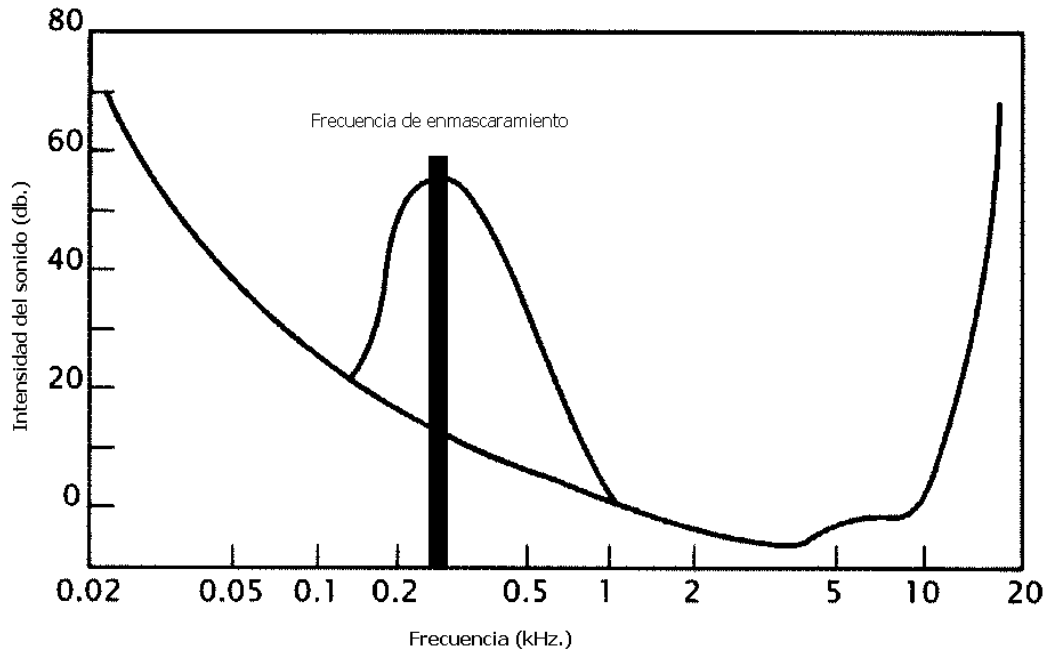


Figura 8.10: Nivel mínimo de audición al superponer una frecuencia de enmascaramiento.

Un fenómeno interesante en la audición es lo que se conoce como enmascaramiento; cuando, en un sonido existe una frecuencia dominante, las frecuencias alrededor de esta se perciben con más dificultad, necesitan ser más intensas para ser percibidas.

Este fenómeno es utilizado en la compresión de audio en el estándar MPEG. La capa III de MPEG-2 es la correspondiente al audio en este estándar y lo que popularmente se conoce como MP3.

La aproximación que usualmente se adopta es dividir la señal en treinta y dos bandas de frecuencia. Para cada banda se calcula una frecuencia de enmascaramiento a partir de la señal promediada en cada banda y un modelo psico-acústico. De este modo se supone que la curva de enmascaramiento se puede sustituir por el valor calculado y, si la señal en una banda está por debajo del nivel de enmascaramiento, la banda se desecha. En otro caso la señal se cuantiza utilizando el menor número de bits para que el nivel de ruido desaparezca.

Con la compresión MP3 se consiguen tasas de compresión 10:1.

8.5 Formatos.

8.5.1 Formatos de ficheros.

Se han desarrollado distintos formatos de ficheros para cada una de las plataformas existentes en el mercado. Así, por ejemplo, en MacOS encontramos el formato AIFF, en Windows el formato Wave y en Unix el formato AU.

También se puede almacenar audio en ficheros multimedia como QuickTime, AVI o los swf de Flash.

8.5.2 Difusión de sonido.

El sonido se puede difundir a través de una línea de comunicación de datos y en particular a través de Internet.

La información se transmite para ser escuchada inmediatamente como en el caso de la radio, no para ser almacenada. Ejemplos de formatos de difusión de audio por Internet son RealAudio o QuickTime.

8.5.3 MIDI.

MIDI son las iniciales de Musical Instrument Digital Interface. Este es un protocolo estándar de comunicación entre instrumentos electrónicos.

Una composición en protocolo MIDI está formada por una secuencia de mensajes. Cada mensaje controla alguna característica tal como *Note On*, mensaje que tiene dos parámetros, el primero identifica la nota y el segundo es la velocidad de ataque de la nota.

Otros mensajes son *Note Off*, *Key Pressure* y *Program Change*. Este último controla el tipo de instrumento que ejecuta la nota.

QuickTime incorpora toda la funcionalidad del protocolo MIDI por lo que cualquier PC con el software QuickTime puede ser utilizado como un sintetizador.

8.6 Combinación de sonido e imagen.

La sincronización entre audio y vídeo es una tarea importante para evitar, por ejemplo, el extraño efecto que se produce cuando un personaje habla y su voz suena un instante antes o un instante después de que se produzca el movimiento de sus labios.

En particular Flash de Macromedia cuenta con facilidades de sincronización.

Resumen

Aunque la producción de sonido sea un proceso físico bien definido, en la percepción del sonido entran a tomar parte otros procesos como los fisiológicos, procesos derivados del propio órgano de audición, y psicológicos, derivados de los procesos de reconocimiento que se llevan a cabo en el cerebro.

Los sonidos percibidos por el oído humano se diferencian entre sí por tres características esenciales: el tono, el timbre y la intensidad.

La digitalización del sonido consta de dos partes: el muestreo y la cuantización. Durante la fase de muestreo se toma el valor de la amplitud del sonido a intervalos regulares de tiempo. Durante la fase de cuantización se restringe el valor de la intensidad a ciertos valores enteros.

La calidad del sonido depende tanto de la frecuencia de muestreo como del número de niveles utilizados en la fase de cuantización.

Para evitar tener que limpiar la señal digital obtenida lo más sensato es realizar la digitalización del mejor modo posible, esto es, utilizando las mejores herramientas a nuestro alcance.

El tamaño de los ficheros de audio depende de la calidad exigida al sonido digitalizado. Por ejemplo, un minuto de sonido estéreo digitalizado a con una frecuencia de

muestreo de 44.1 kHz u 16 bits para la cuantización ocupa alrededor de 10 Mb. una canción de cuatro minutos puede ocupar alrededor de 40 Mb.

Almacenar, manipular y transmitir ficheros tan grandes es impracticable, se hace necesaria su compresión. Los métodos de compresión de ficheros de audio que obtienen mayores tasas de compresión sin mermar la calidad de la grabación son los basados en modelos perceptuales, como la compresión MPEG-2 de audio, lo que popularmente se conoce como MP3. Con este tipo de compresión se consiguen tasas de hasta 10:1.

Cada plataforma tiene su propio formato de fichero de sonido. El formato MIDI sí que es un formato estándar de comunicación entre instrumentos musicales electrónicos. Está incluido en QuickTime.

Los ficheros de sonido se pueden difundir a través de Internet igual que las emisiones de radio. El objeto es que sean escuchados inmediatamente sin necesidad de ser almacenados.