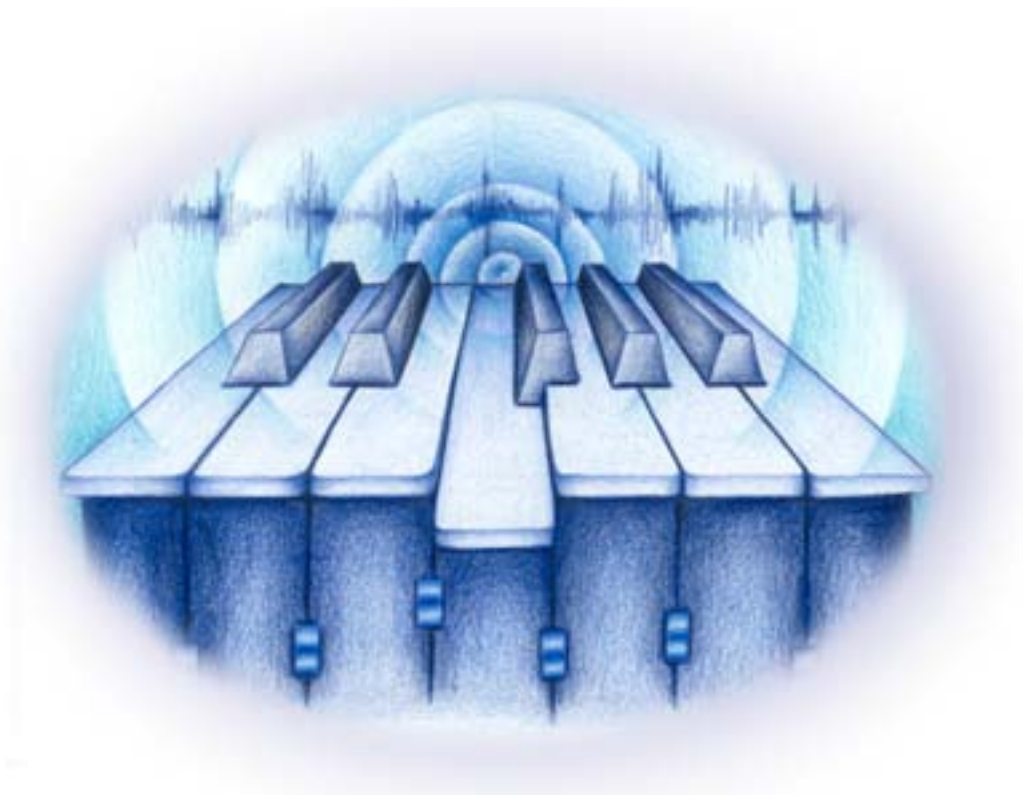


# Principios del sonido digital





## Índice

---

<b>Etapa 1: Introducción al sonido digital</b> .....	5
<b>El sonido analógico. Transductores</b> .....	5
<b>Grabación analógica del sonido</b> .....	6
La calidad del sonido analógico.....	7
<b>El sonido digital. Introducción</b> .....	7
<b>Comparación del sonido con la imagen digital</b> .....	8
<b>Etapa 2: Muestreo en frecuencia</b> .....	10
<b>Frecuencia de muestreo</b> .....	10
<b>Teorema de Nyquist</b> .....	11
<b>Muestreo con frecuencias inferiores a 44.100 Hz</b> .....	12
<b>Visualización de los componentes frecuenciales.</b>	
<b>Espectrograma de un sonido</b> .....	13
¿Qué es un espectrograma sonoro? .....	14
¿Cómo interpretar un espectrograma sonoro? .....	15
¿Cómo obtener representaciones espectrales en Sound Forge? .....	15
<b>El Aliasing</b> .....	16
<b>Filtrado Anti Aliasing</b> .....	17
<b>Filtrado paso-bajo en la grabación digital</b> .....	18
<b>Muestreo con frecuencias superiores a 44.100 Hz</b> .....	19
<b>Etapa 3: Cuantificación</b> .....	20
<b>Número de bits y resolución</b> .....	20
<b>Error de cuantificación</b> .....	21
<b>Bits y rango dinámico</b> .....	21
Algunas consideraciones adicionales sobre los 16 bits .....	23
<b>Los conversores A/D y D/A</b> .....	24
<b>Saturación y distorsión</b> .....	26
Saturación en la grabación .....	26
Saturación en la reproducción.....	26
<b>Sobremuestreo</b> .....	27
<b>Etapa 4: Formatos</b> .....	28
<b>Tamaños en el audio digital</b> .....	28
¿Cuánto ocupa un segundo de sonido digital estéreo de 16 bits y 44.100 Hz?.....	28
<b>¿Qué calidad seleccionar?</b> .....	29
<b>Formatos de sonido digital</b> .....	30
<b>Compresión de ficheros de audio</b> .....	31
<b>La compresión MP3</b> .....	32

Compresión y descompresión .....	33
Calidad y bit ratio .....	33
<b>Etapa 5: Historia del audio digital .....</b>	<b>35</b>
<b>Grabación y generación digital de sonido .....</b>	<b>35</b>
<b>Los ordenadores y el audio digital .....</b>	<b>35</b>
Audio digital para las masas .....	36
<b>Autoevaluación .....</b>	<b>38</b>

## Etapa 1: Introducción al sonido digital

### El sonido analógico. Transductores

Antes de proceder a estudiar en profundidad el sonido digital, comentaremos brevemente algunas características de su “antecesor tecnológico”, el sonido analógico, y es que, hasta la aparición de los ordenadores, el sonido se grababa siempre de forma analógica. Sin embargo, ¿qué significa exactamente este término?

*El término analógico se utiliza actualmente en contraposición a digital, especialmente en el campo del sonido, pero también en el de la imagen. Las señales analógicas son funciones de variables continuas (frecuentemente del voltaje) y se denominan así porque sus evoluciones temporales imitan (son una analogía de) las señales originales.*

La representación gráfica de un sonido grabado analógicamente tendrá la misma forma que el sonido original, pero la curva indicará variaciones de voltaje, en lugar de variaciones de presión de aire. Estas nuevas señales analógicas se obtienen por medio de transductores.

*Un transductor es un dispositivo capaz de convertir una magnitud física en otra.*

#### Ejemplos de transductores

El fotómetro de una cámara fotográfica automática	Convierte la intensidad luminosa en voltaje. Para ello, utiliza una fotorresistencia, es decir, una resistencia que varía en función de la luz que incide sobre ella.
El ratón de un ordenador	Convierte los desplazamientos en x e y, en dos señales de voltaje. Existen varias técnicas posibles, pero la más extendida utiliza dos pequeñas ruedas dentadas que se mueven junto con la “bola” del ratón, y que dejan pasar una señal luminosa a intervalos regulares. De esta forma, el número de pasos o impulsos luminosos registrados en uno u otro sentido de giro es proporcional a la distancia recorrida.

## Grabación analógica del sonido

El micrófono convierte la variación de la presión de aire ejercida sobre su membrana en una señal de voltaje variable en el tiempo.

La variación de este voltaje se puede grabar analógicamente utilizando diferentes tecnologías, sobre una cinta magnética o en los surcos de un disco de vinilo.

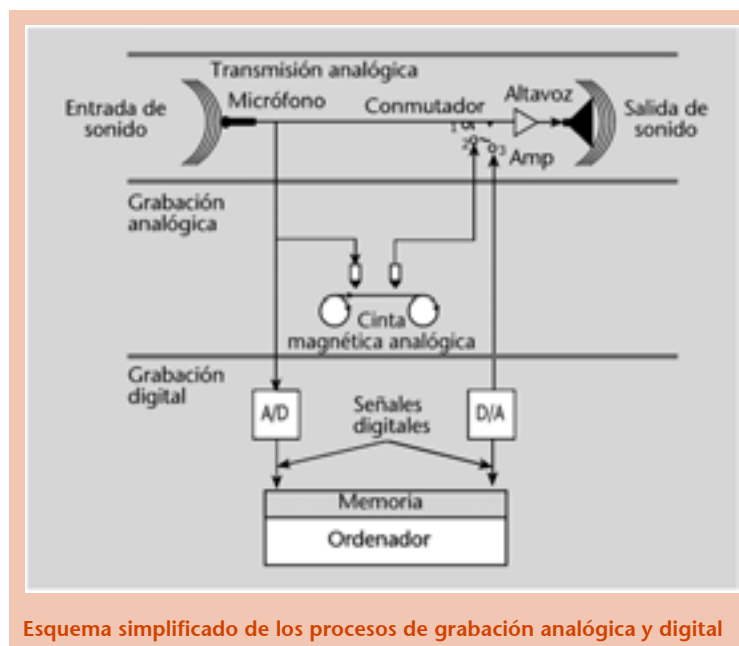
En el caso del disco de vinilo, por ejemplo, los surcos dibujan, sobre la espiral del disco, trayectorias que reproducen las formas (son una analogía) de la señal original.

Cuando deseamos reproducir el sonido, la señal eléctrica generada por el cabezal de la pletina o la aguja del tocadiscos se amplifica y envía a los altavoces, donde un nuevo transductor la convierte en un campo magnético capaz de desplazar y de hacer que oscilen (con las frecuencias originales) los conos de papel de los altavoces.

El micrófono y los altavoces son los dos transductores básicos utilizados en la grabación y reproducción del sonido

Y no son los únicos, pues también lo son los cabezales de un magnetófono o de una pletina de casetes, o la aguja y la cápsula de un plato tocadiscos, pero conviene resaltar que, aunque la grabación se realice digitalmente, tal como se estudiará a continuación, las transducciones analógicas en el micrófono y en los altavoces se seguirán llevando a cabo inevitablemente.

En la figura se esquematizan los dos procesos de grabación alternativos (analógico y digital).



## La calidad del sonido analógico

Al contrario de lo que opina la mayoría, una grabación analógica no tiene por qué sonar peor que una grabación digital (en condiciones óptimas suena, de hecho, mejor). Sus inconvenientes radican en que:

- la señal analógica se degrada mucho más rápidamente (las cintas magnéticas se desmagnetizan, y tanto las agujas como los surcos en el vinilo se desgastan);
- en cada nueva generación se produce una pequeña pero inevitable pérdida, de forma que, a cada nueva copia, la señal se parece cada vez menos a la original.

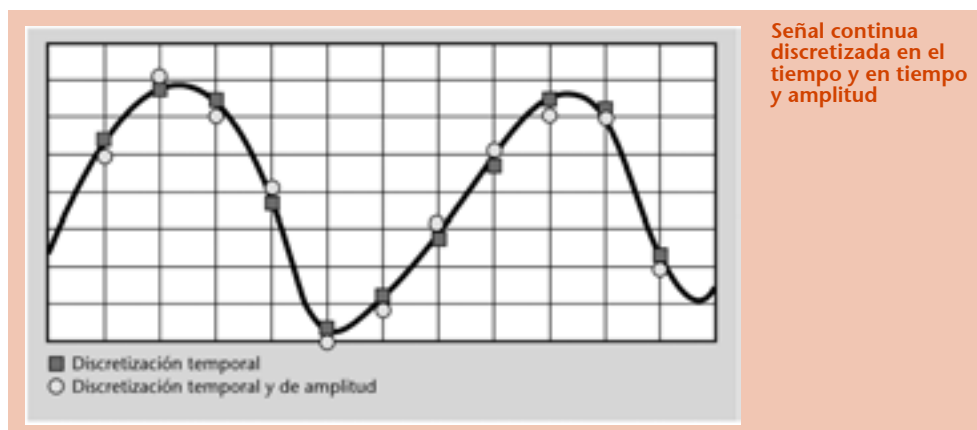
Por estas razones, el sonido digital ha tomado definitivamente el relevo del analógico. En éste y los siguientes módulos se estudiará por qué y cómo el sonido digital aporta un sinnúmero de nuevas y excitantes posibilidades en la producción, manipulación, creación y experimentación sonoras.

## El sonido digital. Introducción

El principio fundamental del audio digital consiste en discretizar las señales sonoras continuas (como las captadas por un micrófono) para convertirlas en secuencias de números.

La discretización de estas señales se lleva a cabo en dos escalas diferentes, la temporal y la de la amplitud.

En la siguiente figura se muestra una señal continua, discretizada sólo en el tiempo (cuadros oscuros) y conjuntamente en el tiempo y en la amplitud (puntos claros), de forma que sólo pueda tomar valores situados sobre las líneas. Se intuye que cuanto menor sea el tamaño de la cuadrícula, mayor similitud existirá entre la señal original y la señal digitalizada.



## Comparación del sonido con la imagen digital

Para aclarar algunos conceptos, haremos una analogía con el cine y el vídeo digital.

En ambos medios, existe una discretización temporal (incluso en el caso del cine, que no es un medio digital). En el cine, la unidad de discretización temporal es el fotograma (24 fotogramas/segundo), mientras que en el vídeo esta unidad se suele denominar con el término inglés *frame* (dependiendo del sistema, este valor suele ser de 25 ó 30 *frames*/segundo). Esto correspondería, en el audio, a la discretización temporal.

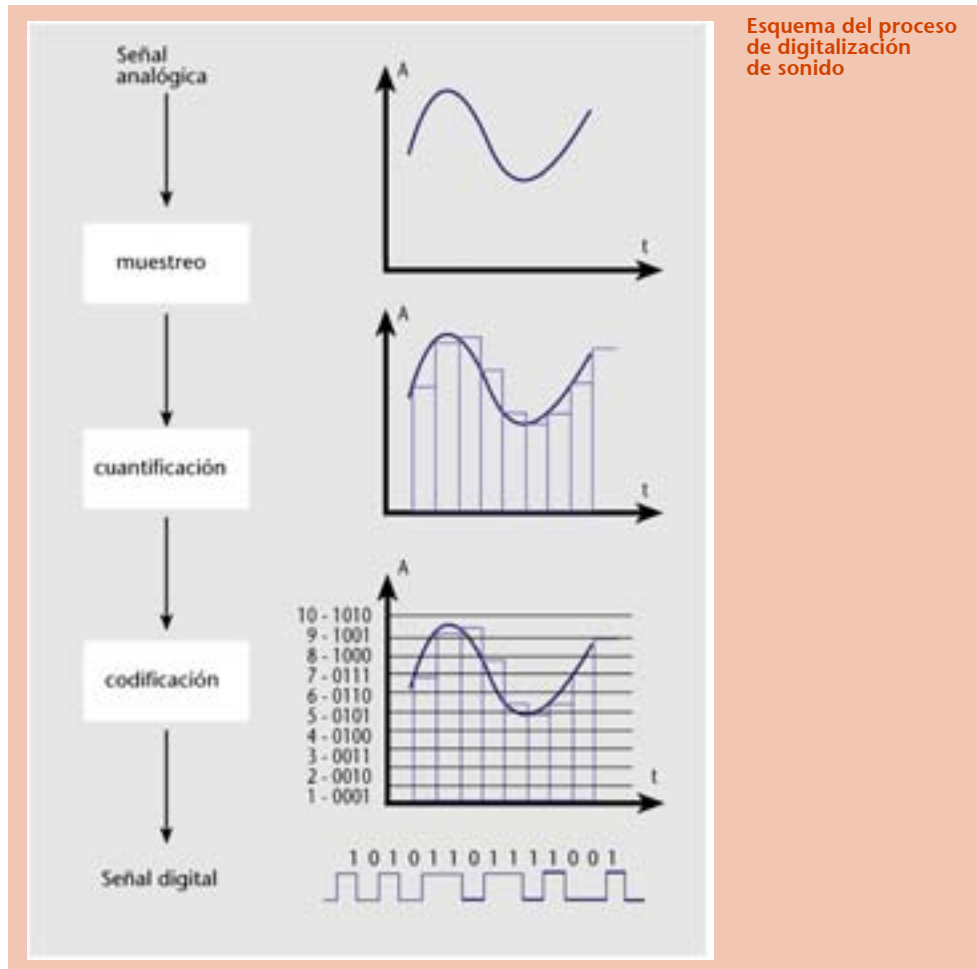
Sin embargo, en el caso del vídeo digital se producen dos discretizaciones adicionales. La segunda convierte cada *frame* en una matriz de puntos (por ejemplo,  $800 \times 600$ ), y la tercera asigna un número (de entre un conjunto finito de valores) a cada punto, de forma que cada uno de estos números corresponde a un color y un brillo determinado.

Es obvio que cuantos más puntos apliquemos, y cuantos más números utilicemos para cada punto, mayor será la similitud entre la señal analógica original y la señal digitalizada. Si disponemos de pocos puntos, la imagen aparecerá cuadrículada (pixelada), y si el número de valores posibles para cada punto es pequeño, perderemos matices en los colores (o en los niveles de grises) y en los brillos. En la siguiente figura se muestran estos casos.





A continuación veremos que algo parecido sucede con el sonido. En primer lugar, será necesario realizar una discretización temporal denominada **muestreo**, y, a continuación, discretizar de nuevo estos valores muestreados en un proceso denominado **cuantificación**. En la siguiente figura se representa el proceso completo, que nosotros estudiaremos con detalle en las dos etapas “Muestreo en frecuencia” y “Cuantificación”.



## Etapa 2: Muestreo en frecuencia

### Frecuencia de muestreo

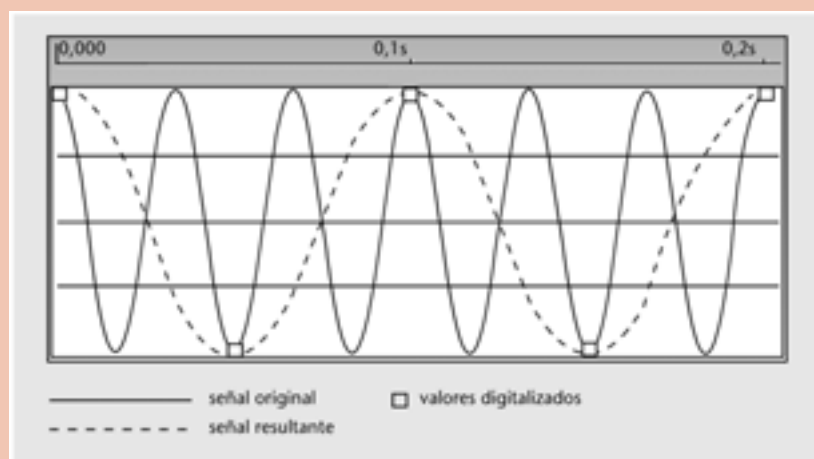
Hasta ahora hemos utilizado el verbo *discretizar* para referirnos a la acción de tomar valores discretos de una señal continua.

De forma más precisa, se utiliza el verbo *muestrear* (y el sustantivo *muestreo*, equivalentes de los términos ingleses *to sample / sampling*) para indicar la acción de tomar muestras a intervalos de tiempo regulares. Para digitalizar un sonido es, pues, necesario muestrearlo, pero ¿con qué frecuencia o regularidad?

*Muestrear significa tomar muestras o valores a intervalos de tiempo regulares.*

En la etapa “Introducción al sonido digital” se comenta que para la imagen animada, la frecuencia de muestreo suele estar entre los 24 y los 30 fotogramas por segundo. Con esta frecuencia se consigue engañar a la retina y al cerebro, haciéndoles creer que lo que reciben no son imágenes discontinuas, sino un flujo continuo de luz, pero, como veremos, estos valores son totalmente insuficientes en el caso del sonido.

Para comprender este fenómeno se puede estudiar la siguiente figura que representa con trazo continuo una señal sinusoidal de 30 Hz, y con trazo discontinuo, el resultado de muestrearla con una frecuencia insuficiente.



Señal continua muestreada con una frecuencia insuficiente

Muestreo con frecuencia insuficiente	
¿Cómo se puede comprobar visualmente que la frecuencia es efectivamente de 30 Hz?	Si se observa con detalle la escala temporal de la figura, se puede comprobar que en 0,1 segundos se producen tres ciclos completos de la onda. Mediante una sencilla regla de tres se puede, por tanto, deducir que en 1 segundo se producirán 30 ciclos completos.
¿Cuál es la frecuencia de muestreo de la figura (cuadrados blancos)?	Se están tomando 2 valores en cada décima de segundo, por lo que la frecuencia de muestreo es de 20 Hz.
¿Qué ocurre cuando muestreamos una señal de 30 Hz con una frecuencia de muestreo de 20 Hz? ¿Cuál es la frecuencia de la señal resultante?	En este caso, estaremos tomando 20 valores de la onda original por segundo, es decir, un valor cada 0,05 segundos, lo que corresponde a los seis cuadros blancos de la figura. La nueva señal obtenida juntando estos puntos muestreados, que se representa en la figura mediante un trazo discontinuo, tiene un periodo aparente de 0,1 segundos, es decir, ¡una frecuencia de 10 Hz!, lo cual no se corresponde en absoluto con la señal original.
¿Podrías describir un fenómeno comparable, debido al submuestreo, en el caso de la imagen en movimiento?	Un ejemplo típico es el del movimiento aparente de los radios de las ruedas de los carros o los coches, en las películas. A veces se observa que los radios avanzan lentamente o incluso que retroceden. Esto es debido a que la frecuencia de muestreo del cine (24 fotogramas/segundo) no es suficiente para mostrar la frecuencia de rotación de la rueda.
¿Cuál debería ser, pues, la mínima frecuencia de muestreo correcta, en el ejemplo anterior?	En el caso de la rueda, intuitivamente se puede entender que sólo se mostrará el movimiento de forma correcta si, entre un fotograma y el siguiente, el desplazamiento angular de cada radio es menor al ángulo entre radios. En caso contrario, el movimiento grabado aparente no coincidirá con el real. Algo muy similar, aunque tal vez menos intuitivo, sucede con el sonido.

## Teorema de Nyquist

El **Teorema de Nyquist** o teorema del muestreo afirma que:

para muestrear correctamente una señal periódica de cualquier frecuencia, se requiere como mínimo una frecuencia de muestreo doble.

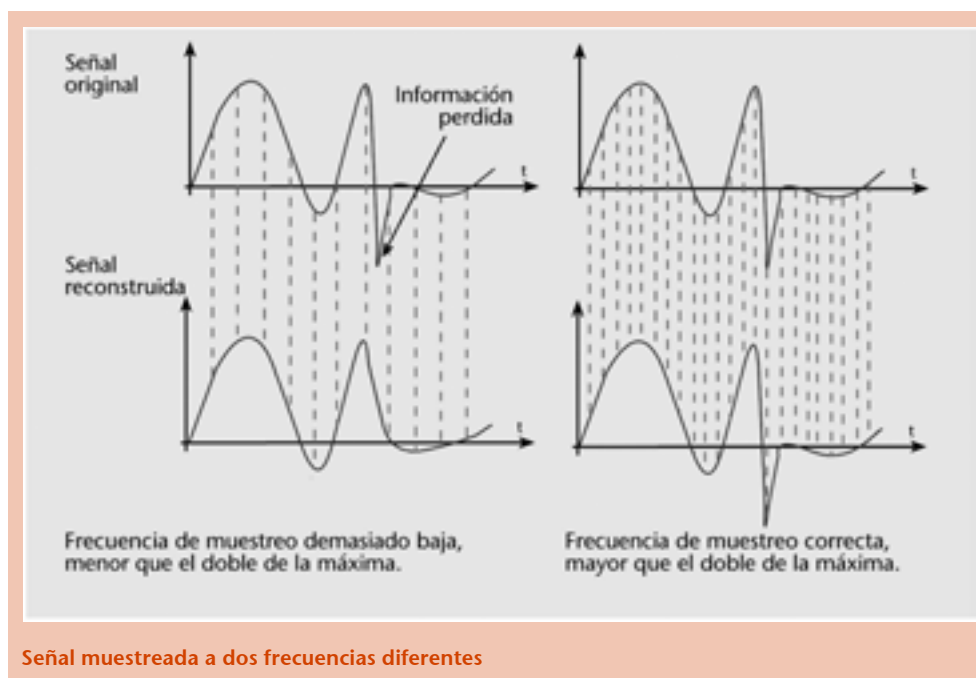
Dicho de otra forma,

cualquier señal digitalizada sólo puede representar correctamente frecuencias inferiores a la mitad de la frecuencia de muestreo. Esta frecuencia mitad se denomina *frecuencia de Nyquist*.

En el ejemplo del apartado “Frecuencia de muestreo” hubiésemos necesitado, por lo tanto, una frecuencia mínima de 60 Hz para muestrear correctamente la señal original, ya que ésta tenía 30 Hz.

En el siguiente gráfico se puede comprobar el efecto de muestrear una señal con dos diferentes frecuencias de muestreo. En el primer caso, la frecuencia es claramente

insuficiente, ya que se pierde la información y no se consigue reconstruir correctamente la señal original.



### Muestreo con frecuencias inferiores a 44.100 Hz

Muestrear a 44.100 Hz significa tomar 44.100 muestras del sonido cada segundo, lo que para sonidos largos puede llegar a significar una considerable cantidad de memoria o de espacio en el disco duro. Por ello, en aplicaciones no estrictamente musicales en las que la calidad ya no es lo primordial, es frecuente almacenar los sonidos con frecuencias inferiores, para ahorrar espacio.

En un sonido con una frecuencia de muestreo de, por ejemplo, 20.000 Hz, sólo se podrán representar correctamente los componentes frecuenciales situados por debajo de los 10.000 Hz (la mitad de la frecuencia de muestreo).

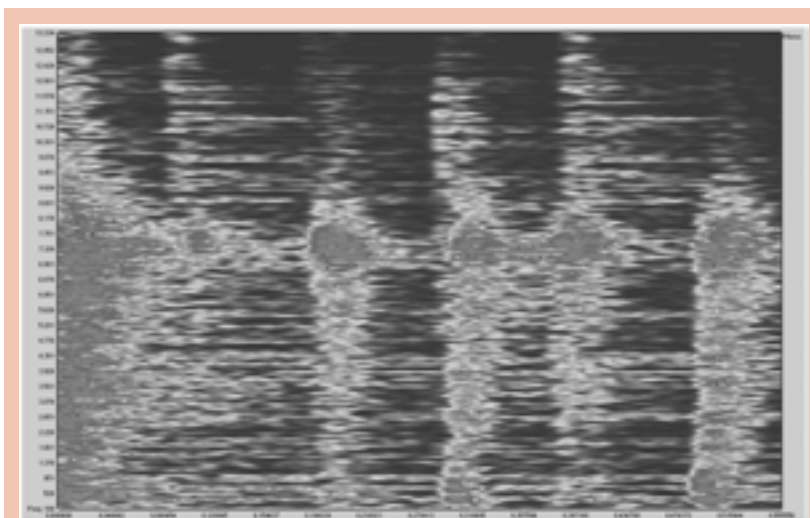
Para las aplicaciones que no precisen la máxima fidelidad, esto es más que suficiente. El sonido resultante será un poco menos brillante, pero se parecerá todavía bastante al original, pues la zona frecuencial con mayor energía suele estar aproximadamente entre los 1.000 y los 3.000 Hz.

En las versiones CD y web pueden apreciarse ejemplos ilustrativos.

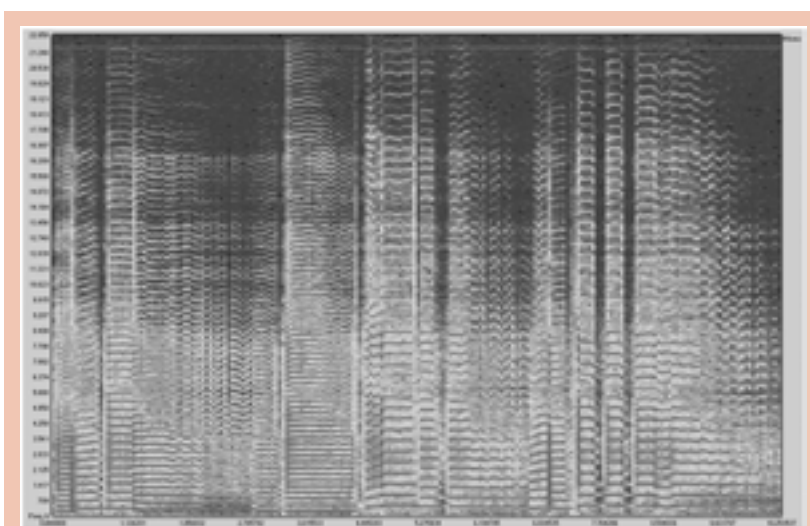
## Visualización de los componentes frecuenciales. Espectrograma de un sonido

---

Las variaciones producidas en un sonido aplicando diferentes frecuencias de muestreo pueden comprobarse también visualmente, estudiando sus análisis espectrales. Los siguientes espectros, que representan respectivamente los dos sonidos (batería y saxo) estudiados en el apartado “Muestreo con frecuencias inferiores a 44.000 Hz”, se han obtenido mediante la herramienta **Spectrum Analysis**, disponible en el menú **Tools** del programa Sound Forge.



Espectro del sonido de batería escuchado en el anterior apartado (a 44.100 Hz)



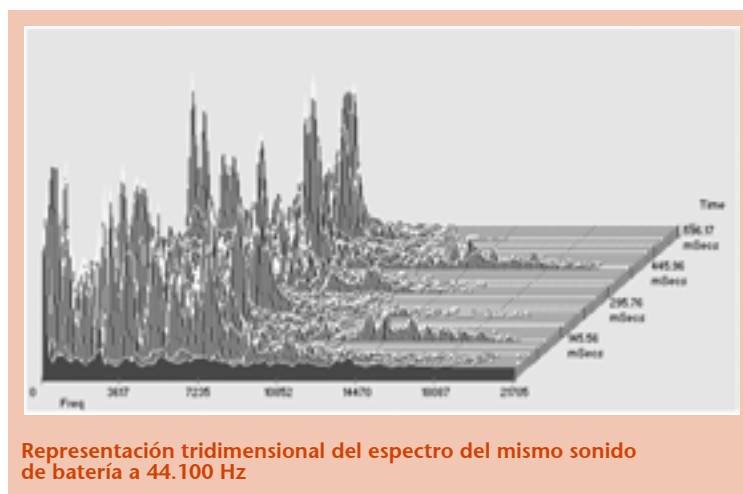
Espectro del sonido de saxo escuchado en el anterior apartado (a 44.100 Hz)

- El eje horizontal representa el tiempo (desde el inicio hasta el final del sonido).
- El eje vertical la frecuencia (de 0 a 22.050 Hz, que es la frecuencia de Nyquist).

- Los colores representan el nivel de amplitud en función de la frecuencia y del tiempo.

### ¿Qué es un espectrograma sonoro?

- Cuando visualizamos un sonido cualquiera desde un programa editor gráfico de audio, como Sound Forge, lo que aparece es la variación de la presión o intensidad sonora en el tiempo. Es lo que se denomina *representación en el dominio temporal*.
- Sin embargo, cuando visualizamos el espectrograma de un sonido, obtenido a partir de la transformación de Fourier, lo que se muestra es la evolución temporal de las diversas frecuencias que integran este sonido. Dado que se trata de una representación tridimensional (en lugar de la bidimensional del caso anterior) en la que se deben representar simultáneamente tiempo, frecuencia y amplitud, esta visualización puede llevarse a cabo de varias formas. Las dos más frecuentes son:
- Utilizando colores (o valores de grises) diferentes para mostrar las diferentes intensidades de frecuencia. Ésta es la estrategia que utiliza Sound Forge.
- Recreando una visualización tridimensional, tal como se muestra en la siguiente figura.



- Sobre el eje horizontal se representan las frecuencias (desde 0 a la frecuencia de Nyquist, que en este caso vale 22.050 Hz).
- Sobre el eje diagonal se representa el tiempo (en milisegundos).
- Sobre el eje vertical se representa la amplitud.

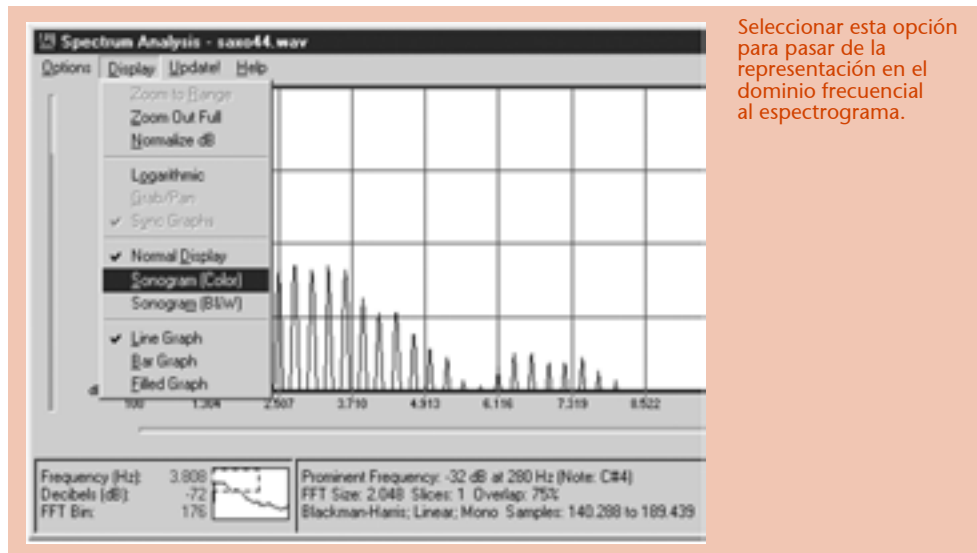
### ¿Cómo interpretar un espectrograma sonoro?

Aunque sus aspectos sean muy diferentes, los dos tipos de espectrogramas muestran la misma información, es decir, **la evolución de la intensidad para todo el rango de frecuencia, a lo largo de toda la duración del sonido**. Estudiando cualquiera de los dos espectrogramas (aunque la información se ve más claramente en el espectrograma “de colores”) del sonido de batería, se puede detectar por ejemplo:

El sonido de batería	
El sonido se compone de 6 “golpes” separados, con breves silencios entre ellos.	Tenemos 6 “columnas” verdes o rojas separadas por zonas azules.
El primer golpe es el más intenso y también el más grave.	Es el que tiene más rojo.
El segundo es el menos intenso y tan sólo tiene componentes muy agudos.	Tan sólo tiene un pequeño punto rojo, y el verde se sitúa en la zona de altas frecuencias, por encima de otras “columnas” o golpes.
El quinto golpe también presenta componentes agudos intensos.	Tiene zonas verdes para las mismas frecuencias que el caso anterior.

### ¿Cómo obtener representaciones espectrales en Sound Forge?

1. Abrir el fichero de sonido que se va a estudiar (y activar su ventana en el caso de que tuviésemos varios ficheros abiertos).
2. Si sólo quisiéramos analizar un fragmento del sonido, podemos seleccionar el fragmento con el ratón.
3. Abrir el menú **Tools/Spectrum Analysis**.
4. La visualización que aparece por defecto representa el promedio de la transformada de Fourier (FFT) para todo el sonido. En esta representación no aparece el tiempo, sino tan sólo los valores de la amplitud en función de la frecuencia, promediados para todo el sonido. Es lo que se denomina la *representación en el dominio frecuencial* (ya que el eje de abscisas que suele corresponder al tiempo corresponde aquí a las frecuencias).
5. Abrir el menú **Display/Sonogram** (nos permite elegir entre la representación con colores o con escala de grises).
6. **Display/Zoom to Range** selecciona automáticamente la zona de frecuencias en la que se concentra la energía, mientras que **Display/Zoom Out Full** muestra todo el rango de frecuencias.
7. Con la barra deslizante horizontal situada debajo del gráfico, se puede ajustar la intensidad del color (o de la escala de grises) para visualizar mejor una zona determinada.

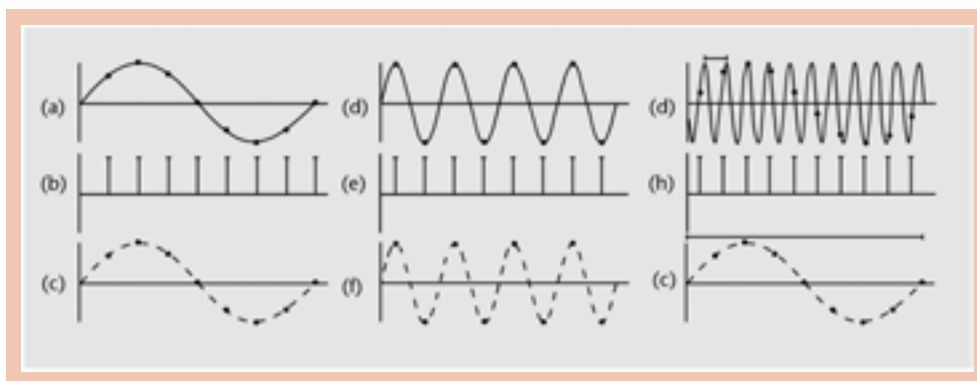


## El Aliasing

Hemos podido comprobar que en algunos casos una reducción de la frecuencia de muestreo puede ser perfectamente tolerable, aunque en sonidos muy agudos o brillantes, como por ejemplo los platillos de una batería, se producirá una inevitable pérdida de brillo.

En cualquier caso, para aplicar correctamente esta reducción de la frecuencia de muestreo y obtener los mejores resultados posibles, se deberán tener en cuenta unas consideraciones adicionales que se detallan a continuación.

En este gráfico (interactivo en las versiones web y CD) se muestra el efecto que puede producirse al utilizar una frecuencia de muestreo insuficiente. En los tres casos, la frecuencia de muestreo se mantiene, pero varía la frecuencia de la señal que se debe muestrear.



En el primer caso la frecuencia de muestreo es más que suficiente para la frecuencia de la señal que hay que muestrear, por lo que la señal se reconstruye sin ningún problema.



El segundo caso es crítico, ya que la frecuencia de la señal es igual a la mitad de la frecuencia de muestreo (e igual, por tanto, a la frecuencia límite de Nyquist). Se debería utilizar una frecuencia de muestreo superior.

En el tercer caso la frecuencia de muestreo es claramente insuficiente, ya que es aproximadamente igual a la frecuencia de la señal que hay que muestrear. En este caso, la señal reconstruida muestra una frecuencia que nada tiene que ver con la señal original.

Tal como se demuestra gráficamente en el tercer caso, al muestrear con frecuencias inferiores surgen frecuencias “fantasmas” que realmente nada tienen que ver con la frecuencia original.

En este ejemplo hemos utilizado ondas sinusoidales para simplificar el entendimiento de este principio. Cuando muestreamos sonidos reales (compuestos a partir de la superposición de muchas frecuencias diferentes) con una frecuencia de muestreo insuficiente, este fenómeno se produce con todos los armónicos más altos de este sonido (los que tengan una frecuencia superior a la de Nyquist).

El resultado es que aparecen varias frecuencias “fantasmas” más graves que se superponen y distorsionan el sonido muestreado. Este fenómeno recibe el nombre de *aliasing*.

---

*El aliasing se produce cuando la frecuencia de muestreo es insuficiente.*

*El aliasing “inventa” frecuencias que no se encuentran en el sonido original, con valores aproximados a la diferencia entre la frecuencia original y la frecuencia de muestreo.*

---

## **Filtrado Anti Aliasing**

---

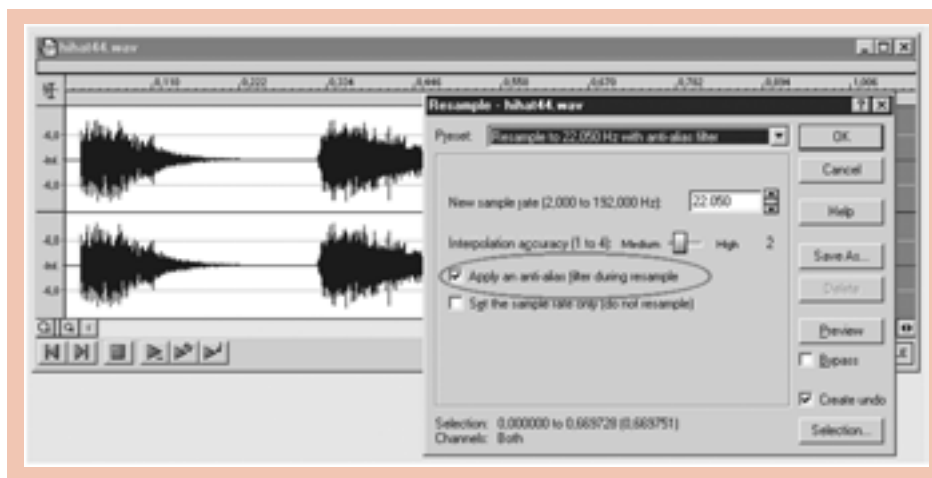
Para evitar el *aliasing*, cuando se desee muestrear a frecuencias inferiores a 44.100 Hz, se debería filtrar previamente la señal entrante, eliminando todos sus valores por encima de la mitad de la frecuencia de muestreo o frecuencia de Nyquist. Es decir, para muestrear correctamente a 20.000 Hz, se debería filtrar previamente la señal original, eliminando todos sus componentes situados por encima de 10.000 Hz.

Esto no siempre es posible si no se dispone del *hardware* adecuado, ya que la mayoría de tarjetas de sonido no profesionales no permite modificar la frecuencia de corte del filtro de entrada.

Existe una solución para realizar todo este proceso por *software* consistente en:

- muestrear a 44.100 Hz;
- filtrar por *software* el sonido obtenido, a la mitad de la frecuencia deseada (p.ej. 10.000 Hz);
- reconvertir por *software* el sonido a la frecuencia deseada (p.ej. 20.000 Hz).

Aunque este proceso pueda parecer algo complicado y engorroso, es la forma de obtener los mejores resultados. De hecho, los dos últimos pasos se pueden realizar de forma automatizada en programas como Sound Forge, al aplicar la opción **Resample** (modificación de la frecuencia de muestreo), accesible desde el menú **Tools**.



La herramienta **Resample**, que permite modificar la frecuencia de muestreo de un sonido almacenado, presenta la opción de filtrar el sonido durante el proceso. Para ello basta con activar la opción **Apply an anti-alias filter during resample** (aplicar filtro anti-alias durante el *remuestreo*).

Conviene indicar también que esta herramienta ofrece distintas posibles calidades para llevar a cabo la conversión (**Interpolation accuracy**). El valor por defecto es 2, más que suficiente para la mayoría de las necesidades. Si se deseara probar otras calidades, se deberá tener en cuenta que a mayor calidad, más lento será el proceso de conversión.

## Filtrado paso-bajo en la grabación digital

Que el oído humano no pueda apreciar frecuencias por encima de los 20.000 Hz no significa necesariamente que éstas no existan. Por ello, este proceso de filtrado que en el apartado “Filtrado Anti Aliasing” hemos realizado por *software* también debe llevarse a cabo aun cuando muestreemos a 44.100 Hz. Por esta razón, los dispositivos de digitalización de sonido incorporan un filtro paso-bajo que recorta todas las frecuencias entrantes situadas por encima de los 22.050 Hz, antes de realizar la digitalización.



Un filtro paso-bajo es un filtro que sólo deja pasar las frecuencias por debajo de determinado valor, denominado *frecuencia de corte*. Los filtros se estudiarán con detalle en el módulo “Síntesis y generación digital del sonido”.

## Muestreo con frecuencias superiores a 44.100 Hz

Además de la frecuencia característica de 44.100 Hz utilizada en los discos compactos y disponible en todas las tarjetas de sonido actuales, algunos dispositivos pueden trabajar también con frecuencias de muestreo superiores.



Los sistemas DAT (Digital Audio Tape) de grabación digital en cinta permiten trabajar, por ejemplo, con frecuencias de 48.000 Hz, aunque esta prestación no estaba originalmente destinada a ofrecer una mayor calidad (la diferencia entre 44,1 KHz y 48 KHz

es muy pequeña para ser realmente apreciable), sino más bien a dificultar las copias digitales de discos compactos, en una época en que las prestaciones de los ordenadores todavía no ponían las cosas tan fáciles para la “piratería casera”.

- Algunos dispositivos profesionales permiten grabar y reproducir hasta frecuencias de 96 KHz o incluso superiores.
- La mayoría de los programas editores de audio, como Sound Forge, permiten trabajar también a estas frecuencias (la última versión, Sound Forge 5.0 admite frecuencias de hasta 192 KHz). Sin embargo, hay que tener cuidado: una cosa es que el *software* lo permita y otra es que lo acepte el conversor de nuestra tarjeta de sonido, que, con suerte, llegará hasta los 48 KHz.

## Etapa 3: Cuantificación

### Número de bits y resolución

Hasta aquí, hemos resuelto la primera parte teórica de la digitalización, consistente en obtener una serie de muestras a intervalos regulares. Falta todavía la segunda parte, que consiste en asignar a cada una de estas muestras un valor numérico que pueda manejar un ordenador.

Para ello, volveremos momentáneamente al ejemplo del vídeo digital, que tal vez resulte más familiar e intuitivo. La resolución de color de una imagen (y de las tarjetas de vídeo) se mide en bits e indica el número de bits asignado a cada píxel de la imagen para almacenar su color. Así, una imagen de 8 bits, podrá incluir 256 ( $2^8$ ) colores diferentes, mientras que una de 24 bits podrá representar más de 16 millones ( $2^{24}$ ) de colores.

*Un bit sólo puede tener dos valores: 1 ó 0. La combinación de dos bits nos da cuatro posibles valores: 00, 01, 10 y 11. Conforme aumenta el número de bits aumenta también el rango de valores representables, que se multiplican por dos con cada nuevo bit.*

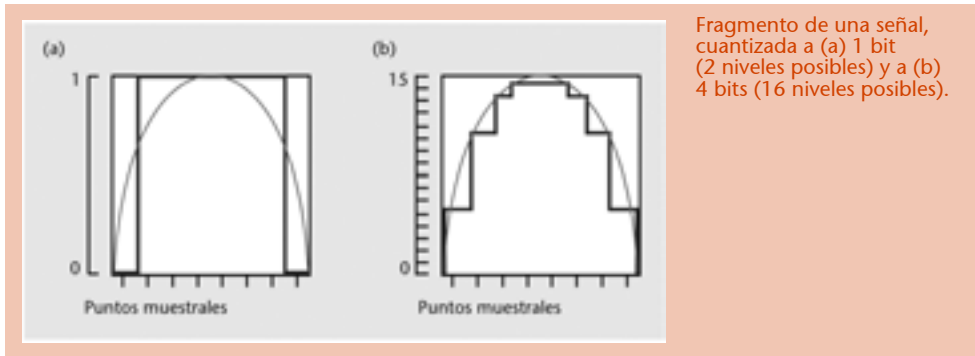
*Por ello, el número de posibles valores viene dado por la fórmula  $2^{\text{número de bits}}$ .*

En el caso del sonido digital, cada uno de los valores muestreados se guardará con un número determinado de bits, y cuantos más bits le asignemos, más niveles o escalones posibles tendrá este sonido digitalizado y más parecido será, en consecuencia, al sonido analógico original (que, al ser continuo, poseía un número infinito de niveles).

*El término resolución de un sonido digital indica el número de bits que se han utilizado para almacenar cada muestra.*

*La resolución determina el número de posibles valores diferentes, o rango, que cada muestra de sonido puede tomar.*

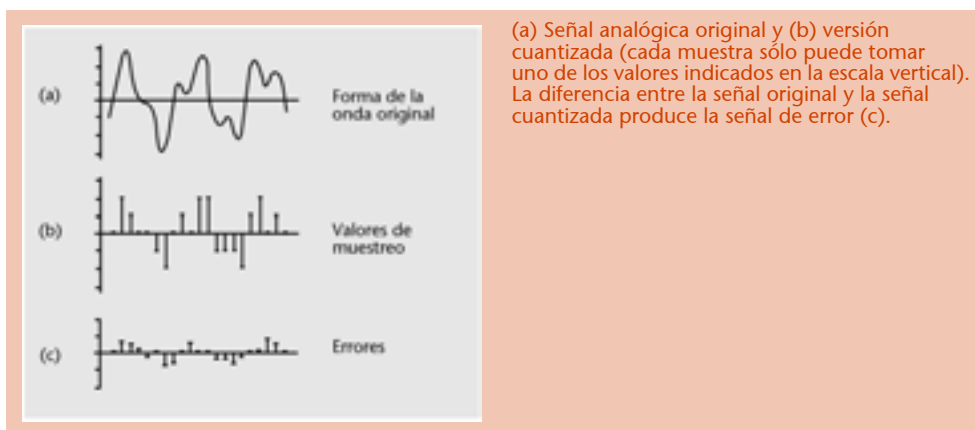
*Así, un sonido digitalizado a 8 bits posee 256 niveles posibles, mientras que un sonido a 16 bits presenta 65.536 niveles posibles.*



## Error de cuantificación

En el gráfico anterior se observa que la señal cuantizada a 4 bits se asemeja algo más a la original que la de 1 bit. Obviamente, cuantos más bits utilizemos, más semejanza obtendremos, pero la cuantificación siempre conllevará un error, por pequeño que éste pueda ser.

Si utilizamos, por ejemplo, 8 bits para digitalizar una señal analógica cuyos valores de voltaje oscilen entre -10 y +10 voltios, cada valor de los 256 posibles no representará un valor exacto, sino un valor dentro de un margen de 0,078125 V (20/256). Esto produce un margen de error, o error de cuantificación, igual a la mitad de este margen. Este error de cuantificación es inevitable e inherente a todo sistema digital, aunque puede reducirse aumentando el número de bits de resolución. En el gráfico siguiente, la tercera figura indica este error.



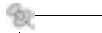
## Bits y rango dinámico

El rango dinámico de una sistema de sonido, expresado en decibelios, viene dado por la fórmula:

$$\text{rango dinámico en dB} = 10 \times \log_{10} (\text{amplitud máxima}^2 / \text{amplitud mínima}^2)$$

y depende, por lo tanto, del cociente entre los cuadrados de la máxima y la mínima amplitud que el sistema puede producir. Conviene remarcar que esta relación no es en absoluto exclusiva de los sistemas digitales, y también es aplicable a los sistemas analógicos.

En todos los dispositivos de sonido electrónicos (ya sean digitales o analógicos), un concepto muy vinculado al rango dinámico es el de la relación señal/ruido (para simplificar, nosotros confundiremos ambos términos).



En un aparato electrónico, la relación señal/ruido indica la diferencia entre el nivel máximo que el dispositivo puede emitir, y el nivel de ruido existente cuando la señal es silencio (el ruido de fondo). Cuanto mayor sea esta diferencia, más limpio será el sonido del dispositivo. En una cadena de varios dispositivos de audio (emisor, amplificador, altavoces, etc.), el valor real, es decir, el

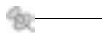
máximo rango dinámico que podremos llegar a apreciar, corresponderá al del dispositivo con peores características.

En la mayoría de los dispositivos de audio (amplificadores, pletinas de casete, reproductores de compactos, altavoces, etc.), esta relación señal/ruido figura entre las especificaciones técnicas.

Volviendo a nuestra fórmula original, en el caso de un sistema digital, podemos reinterpretar esta expresión a partir de las consideraciones estudiadas en el apartado “Error de cuantificación”:

- la amplitud máxima será igual a  $2^{\text{bits}}$ ;
- la amplitud mínima o nivel de ruido será igual al error de cuantificación, que se puede tomar como igual a 1.

Si realizamos los cálculos necesarios, tendremos que **un sistema de conversión de 8 bits posee un rango dinámico de aproximadamente 48 dB, mientras que en uno de 16 bits, el rango dinámico es de 96 dB.**



Rango dinámico de un sistema de 8 bits (demostración):

$$10\log_{10}\left(\frac{(2^8)^2}{1^2}\right) = 10\log_{10}65,536 = 48,16\text{dB}$$

*Una forma aproximada y rápida de calcular el rango dinámico de un sistema digital es mediante la fórmula:*

$$\text{número de bits} \times 6$$

*La resolución en bits de un sonido digital incide directamente en el rango dinámico y en el ruido de fondo.*

*Cuanto más bits utilizemos, más nítido y con menos ruido se percibirá el sonido.*

Basta con escuchar un sonido cuantizado a 16 bits y el mismo a 8 bits, para comprobar que el segundo posee mucho más ruido de fondo.

En las versiones CD y web pueden apreciarse ejemplos ilustrativos.

*Para cambiar los bits de resolución en Sound Forge, se utiliza el comando Process/Bith Depth Converter.*

*Conviene tener muy en cuenta que, si bien al pasar de 16 a 8 bits se pierde calidad, al pasar de 8 a 16 bits, esta calidad no se recupera.*

### **Algunas consideraciones adicionales sobre los 16 bits**

Todos sabemos que el estándar que se ha adoptado en la mayoría de los sistemas comerciales (discos compactos, tarjetas de sonido multimedia, etc.) es el de 16 bits.

Si bien el disco compacto se precia de no añadir casi ningún ruido de fondo (al contrario que los discos de vinilo o las grabaciones en casete), lo cierto es que los 96 dB de rango dinámico (máximo teórico) que ofrecen estos sistemas digitales comerciales podrían desde luego ser mejores, ya que el oído humano es capaz de percibir confortablemente rangos dinámicos superiores.

Dado que el umbral de dolor del oído humano no se sitúa hasta alrededor de los 130 dB, sería de hecho deseable que un sistema de alta fidelidad pudiese alcanzar este rango dinámico, para lo cual serían necesarios unos 22 bits de resolución (i.e.  $130/6$ ).

Algunos sistemas de grabación y reproducción digitales profesionales utilizan, de hecho, conversores con resoluciones de 24 o incluso de 32 bits, ¡que permiten rangos dinámicos superiores a los 144 dB y 192 dB respectivamente!

24 bits es también la resolución empleada por el DVD.

En la práctica, todos estos valores máximos teóricos se ven siempre disminuidos por factores de circuitería analógica, que añaden algo más de ruido.

### ¿Por qué la calidad CD no es siempre calidad CD?

Existe la creencia popular de que en el sonido digital no hay mejores ni peores, “como todo son ceros y unos, mientras no se confundan unos con otros...”. Sin embargo, nada más lejos de la realidad. Los fabricantes de equipos multimedia se escudan detrás de los 16 bits y 44.100 Hz, para colocar la indiscriminada etiqueta “calidad CD”. Sin embargo, ¿os habéis preguntado alguna vez por qué existen actualmente lectores de CD-ROM por 5.000 pts., mientras que se pueden adquirir lectores de CD Audio por 100.000 pts.? La respuesta está en los conversores D/A (que comentaremos en el apartado “Los conversores A/D y D/A”) y en otros componentes electrónicos analógicos, que son los responsables finales del sonido. Es cierto que el láser no suele confundir los ceros con los unos,

pero para que estos enteros binarios lleguen a sonar, tienen que pasar por un complicado proceso de conversión, que se puede realizar con criterios de calidad muy diferentes.

Los valores “reales”, en lo que a rango dinámico se refiere, para tarjetas de sonido y reproductores de discos compactos convencionales se sitúan aproximadamente entre los 65 dB de los equipos más económicos y los 90 dB para los equipos de mayor calidad; valores, en cualquier caso, siempre bastante por debajo de los 96 dB, máximo teórico alcanzable con 16 bits de resolución. Como referencia, indicaremos que la relación señal-ruido de una pletina de cassetes convencional es de unos 60 dB.

Algunos valores típicos de rango dinámico o relación señal/ruido	
Tarjeta de sonido multimedia de 8 bits	40 dB
Pletina de cassette (sin sistema de reducción de ruido)	60-70 dB
Tarjeta de sonido multimedia de 16 bits convencional	70-80 dB
Tarjeta de sonido multimedia de 16 bits “profesional”	80-90 dB
Reproductor de discos compactos	70-90 dB
Sistemas de grabación profesionales de 20, 24 o 32 bits	> 120 dB

## Los conversores A/D y D/A

Hasta aquí la teoría. Veamos ahora cómo funciona un sistema digitalizador de sonido.

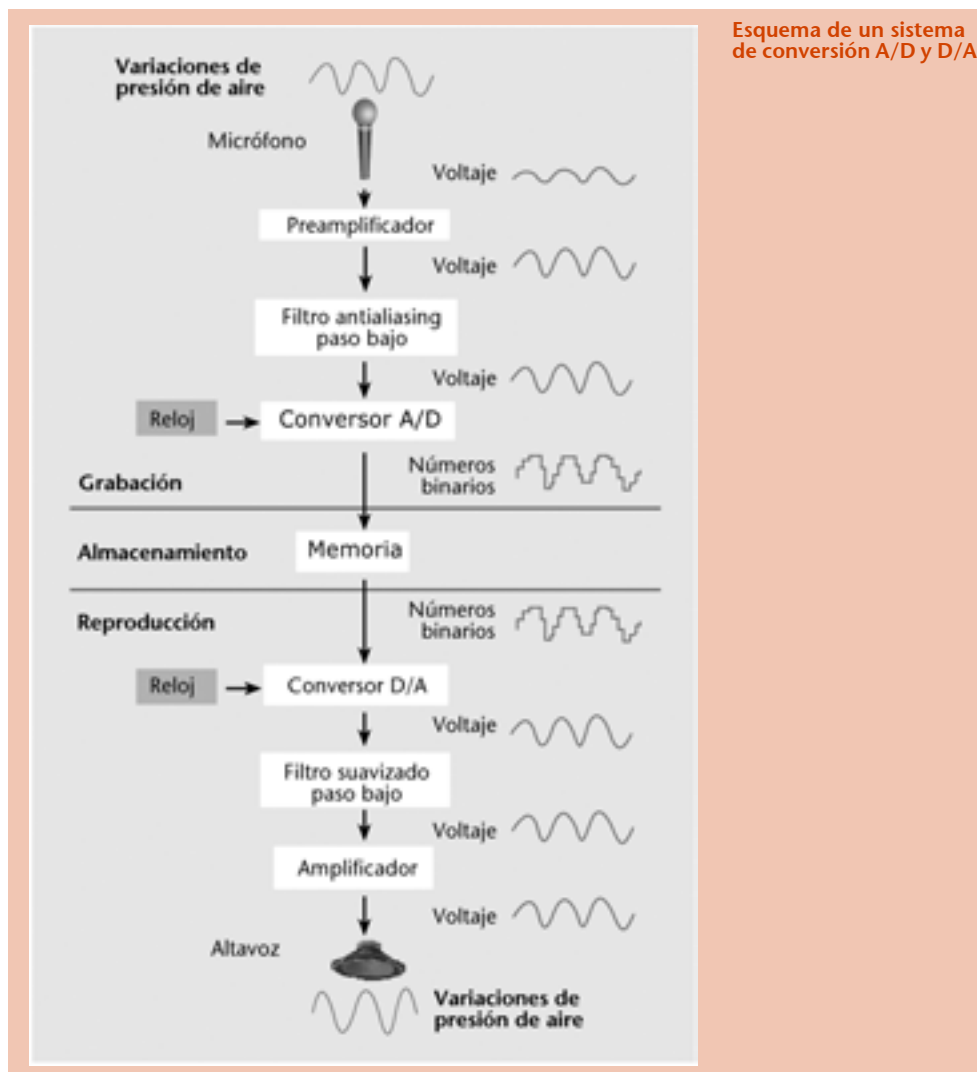
El sistema se compone de dos conversores: un conversor analógico/digital (a partir de ahora, A/D) en la entrada, y un conversor digital/analógico (a partir de ahora, D/A) en la salida. Ambos están controlados por un reloj digital que determina la frecuencia de muestreo.



Para evitar el *aliasing* se coloca en la entrada un filtro pasa-bajo, que elimina de la señal analógica todos los componentes con frecuencias superiores a la frecuencia de Nyquist (normalmente 22.050 Hz). En el proceso de digitalización, el conversor A/D genera un número binario (normalmente de 8 ó 16 bits) a cada pulso del reloj. Este número se almacena en la memoria o se graba en un disco duro.

Para la audición de esta señal digital, es necesaria la reconversión inversa, de forma que los números almacenados son enviados a un conversor D/A que los convierte en voltajes, a la misma frecuencia de reloj. Esta señal analógica es suavizada mediante un nuevo filtro, amplificada y enviada a unos altavoces, que con su vibración, convierten finalmente los voltajes en variaciones de presión de aire.

La siguiente figura esquematiza el proceso completo.



Este proceso de grabación digital se denomina a veces PCM (Pulse Code Modulation) por la tecnología que utilizan los conversores A/D y D/A.

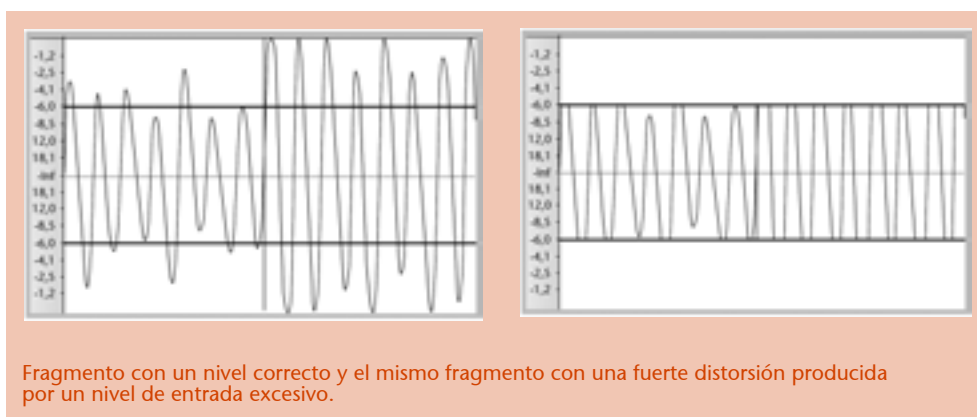
## Saturación y distorsión

### Saturación en la grabación

En cualquier sistema de grabación, incluidos los sistemas analógicos, cuando el nivel de entrada supera un cierto valor umbral se produce una saturación que tiene como consecuencia una distorsión en la reproducción del sonido.

En el caso de la grabación analógica, una ligera saturación no resulta demasiado molesta para el oído, y, de hecho, en ciertas grabaciones analógicas se tiende con frecuencia a saturar ligeramente para obtener un sonido más “lleno”.

No obstante, en el caso de la grabación digital, la saturación puede ser muy molesta y es algo que se debe evitar totalmente. En la siguiente figura se observa una señal grabada con un nivel correcto y, a continuación, la misma señal saturada a causa de una grabación con un nivel de entrada excesivo. Se observa que los picos aparecen totalmente recortados como consecuencia de haber superado los máximos valores almacenables en un sistema de 16 bits.



En las versiones CD y web pueden apreciarse ejemplos ilustrativos.

En el módulo “Grabación y técnicas de estudio” se estudiará con mayor detalle la forma de realizar grabaciones con niveles correctos.

### Saturación en la reproducción

La distorsión digital se puede producir también en la conversión D/A de salida, incluso a partir de señales con niveles correctos. Esto puede suceder, por ejemplo, cuando se reproducen simultáneamente varios sonidos (i.e. varias pistas), ya que, aunque el nivel de cada uno de ellos sea correcto, puede ocurrir que en determinados

instantes, la suma de todos ellos (que es lo que, en definitiva, se manda por el conversor D/A) supere el nivel permitido.

## Sobremuestreo

---

Hasta ahora hemos asumido implícitamente que el número de bits utilizado en la cuantificación (p.ej. 16) es igual al número de bits utilizado en el posterior almacenamiento de estos datos. Existe la posibilidad de utilizar conversores con mayor resolución, aunque finalmente la señal digitalizada se almacene con la misma resolución de 16 bits.

El conjunto de estas distintas técnicas que permiten obtener mejores resultados sin por ello necesitar más espacio de almacenamiento, se denominan *sobremuestreo* o *oversampling*. Actualmente son frecuentes los reproductores de CD Audio que utilizan alguna de estas técnicas en la conversión D/A de salida.

La teoría que subyace detrás del sobremuestreo es demasiado complicada para ser expuesta aquí, por lo cual daremos tan sólo algunas breves indicaciones.

Un sistema que utilice, por ejemplo, un sobremuestreo cuádruple realiza, antes de mandar la señal digital al conversor D/A de salida, una interpolación de los datos digitales colocando tres muestras intermedias entre cada dos muestras originales. De esta forma se obtiene una señal con una frecuencia de muestreo cuatro veces superior (i.e. 176.400 Hz en lugar de 44.100 Hz) que se pasa a través de un conversor D/A que trabaje a esta misma frecuencia y, posteriormente, de un filtro suavizante paso bajo que recorta las frecuencias por encima de los 30 KHz. Con estas técnicas se consigue un incremento del rango dinámico de aproximadamente 6 dB para el sobremuestreo cuádruple y de unos 12 dB para el sobremuestreo óctuplo.

Existe otra familia de técnicas más recientes, denominadas genéricamente *sobremuestreo de 1 bit* o *modulación sigma-delta*, que operan de forma diferente y con las que se obtienen mejores resultados.

## Etapa 4: Formatos

### Tamaños en el audio digital

Cuando los ordenadores sólo manejaban texto, un *megabyte* (aproximadamente un millón de caracteres) parecía una cifra monstruosa. Hoy día, con la llegada del multimedia, los sonidos y especialmente las imágenes digitales han disparado estas cifras.

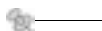
#### ¿Cuánto ocupa un segundo de sonido digital estéreo de 16 bits y 44.100 Hz?

El cálculo es sencillo: en un sonido de 16 bits, cada muestra ocupa dos *bytes* (un *byte* son ocho bits), y si la frecuencia de muestreo es de 44.100 Hz, significa que cada segundo requiere 44.100 muestras. Si el sonido es estéreo, utiliza dos canales, por lo que estas necesidades se ven duplicadas.

$2 \text{ bytes/muestra} \times 44.100 \text{ muestras/segundo} \times 2 \text{ (canales)} = 176.400 \text{ bytes/segundo}$  ó 172,2 Kb/s.

Realizando una multiplicación más, se observa que **un minuto de sonido digital estéreo de calidad ocupa un valor muy cercano a los 10 Mb**. Sabemos por experiencia que los CD Audio raramente superan los 70 minutos. Esta duración límite aproximada viene dada por su capacidad, que suele ser de 720 Mb.

*Un minuto de sonido digital estéreo a calidad CD (16 bits y 44.100 Hz) ocupa aproximadamente 10 Mbytes.*



Comprobar con la versión web la tasa de transferencia a la que trabajan los lectores de CD Audio, y es la que los fabricantes de lectores CD-ROM designaron en su día como *velocidad simple*. A partir de aquí, se tomó la

costumbre de medir la velocidad de los lectores de CD-ROM en múltiplos de este valor (igual que la velocidad de los aviones reactores se mide a veces en múltiplos de la velocidad del sonido).

Cuando el sonido no es estéreo, estos tamaños descienden a la mitad, y lo mismo sucede si la resolución es de 8 en lugar de 16 bits, o la frecuencia de muestreo es de 22.050 Hz. Por ello, limitando la calidad al mínimo, el tamaño necesario para un

minuto de sonido mono de 8 bits y 11.025 Hz se reduce aproximadamente a 646 Kb (estas últimas condiciones suelen aparecer en la configuración del sonido en Windows, como *calidad telefónica*).

La siguiente tabla resume los tamaños (en *Kbytes/segundo*) que ocupan diferentes calidades de sonido digital. Observad que se han redondeado los valores para facilitar su retención mnemotécnica.

Frecuencia de muestreo	Profundidad de cuantización	Kbytes/segundo	Kbytes/segundo
11 KHz	8 bits	11 Kb/s	22 Kb/s
11 KHz	16 bits	22 Kb/s	44 Kb/s
22 KHz	8 bits	22 Kb/s	44 Kb/s
22 KHz	16 bits	44 Kb/s	88 Kb/s
44 KHz	16 bits	88 Kb/s	176 Kb/s
Ancho de banda del oído humano: de 20 Hz a 1g KHz			

## ¿Qué calidad seleccionar?

No existe una respuesta unívoca a esta pregunta, ya que todo depende del uso que queramos dar a cada sonido.

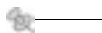
Cuando la intención sea crear (o trabajar con) música o sonido cuyo soporte final vaya a ser un CD Audio, no existe otra posibilidad que la máxima calidad: 44.100 Hz, 16 bits y en formato estéreo.

En cualquier otro caso, no existe una solución unívoca, aunque sí que podemos indicar, a modo de guía, el orden de los pasos que se deben seguir si necesitamos ahorrar memoria o espacio de disco duro.

- La reducción de calidad menos perceptible es utilizar 22.050 Hz en lugar de 44.100 Hz.
- Si el efecto estéreo no es importante, se puede también limitar el sonido a un único canal, pero manteniendo los 16 bits de resolución. Con esto nos situamos a 42,05 Kb/segundo, que es el *mínimo* aconsejable para un sonido “musical”.
- Si lo que se desea es grabar mensajes de voz, se puede seguir reduciendo la frecuencia, o pasar directamente a los 8 bits.

- El que se indiquen valores de frecuencia de 44.100, 22.050 ó 11.025 no es casual, ya que éstos son los valores estándar que soporta la mayoría de las tarjetas (aunque algunas soporten también frecuencias intermedias).
- Hace pocos años, la mayoría de las tarjetas de sonido era de 8 bits, por lo que algunos CD-ROM interactivos utilizan todavía esta resolución, más por criterios de compatibilidad que de economía a ultranza.
- Sin embargo, esta economía sí que es importante cuando el destino de nuestros ficheros de sonido está en la Web, ya que a mayor volumen de los ficheros, mayores tiempos de transferencia.
- Mediante un cálculo sencillo, si consideramos una conexión de 56 kbps (Kbits/segundo) tendremos que un minuto de audio en calidad CD (44.100 Hz, 16 bits y estéreo), ¡tardaría en bajar cerca de 24 minutos! Esto explica el éxito que ha obtenido en Internet el formato de compresión MP3, que consigue, mediante sofisticados algoritmos, reducir el volumen de los ficheros en un factor 1:10 con pérdidas de calidad poco importantes. De éste y otros formatos de compresión hablaremos más adelante.

## Formatos de sonido digital



La información incluida en un fichero de sonido digital no es más que un *array* de *bytes* o de enteros de 16 bits (dependiendo de la resolución aplicada, que, como hemos visto, puede ser de 8 ó 16 bits). Sin embargo, a pesar de la sencillez de su contenido, la lista de formatos existentes es inmensa, pues cada plataforma, y a veces cada programa, dispone de una serie de formatos preferidos con ligeras variaciones entre uno y otro. Las diferencias pueden radicar en el tipo aplicado a las muestras (por ejemplo, enteros con

o sin signo), el tipo de compresión utilizado (si es que lo hubiera) y la forma en que estas muestras se ordenan. Para un sonido estéreo algunos formatos optan por guardar cada canal de forma independiente, mientras que otros eligen guardar muestras alternas de cada una de las pistas. Todos los formatos incluyen una cabecera (en la que se indica la resolución, la frecuencia de muestreo, el número de canales, etc.) cuyo tamaño, contenido y ordenación varía también de un formato a otro.

*En Windows, el formato estándar es el WAV.*

- Esta extensión es una abreviación de la palabra inglesa *wave*, que significa ola u onda; por ello, este tipo de ficheros se conoce también como *ficheros de onda*. Si sólo trabajáis con PC compatibles, probablemente nunca tendréis que preocuparos de otros formatos, salvo si deseáis importar ficheros procedentes de otras plataformas.

- En el entorno Macintosh, uno de los formatos más extendidos es el AIF, algo que deberán tener muy en cuenta los programadores de aplicaciones multimedia multiplataforma. En Internet se utiliza también mucho este formato, junto con el AU, propio de los sistemas Unix.
- También es posible trabajar con ficheros sin cabecera (suelen venir con las extensiones PCM o RAW). Éstos constituyen un caso particular, ya que, a la hora de abrir un fichero de estas características, deberemos indicar nosotros la resolución, la frecuencia de muestreo y el número de canales. Si el fichero suena bien, habremos acertado. Si suena mal, habrá que volver a probar con parámetros diferentes.
- Los programas de edición de audio como Sound Forge ofrecen normalmente opciones para cargar y salvar formatos diferentes. En cualquier caso, conviene tener en cuenta que dos ficheros no comprimidos, y con la misma resolución y frecuencia de muestreo, contendrán la misma información (aunque organizada de formas distintas) y sonarán exactamente igual, independientemente del formato. En el cuadro siguiente se enumeran algunos de estos formatos más frecuentes.

Extensión	Plataforma – aplicaciones
AU	Next/Sun - Internet
AIF	Macintosh - Multimedia ( <i>Director</i> , etc.) e Internet
IFF	Amiga
PCM	Cualquiera
RAW	Cualquiera
WAV	PC

## Compresión de ficheros de audio

Dado el tamaño que pueden adquirir los ficheros de audio digital, es lógico que se hayan buscado formas de compresión que permitan reducir esta cantidad de información.

Al igual que ocurre con la imagen, existen **técnicas sin pérdida** y **técnicas con pérdida**.

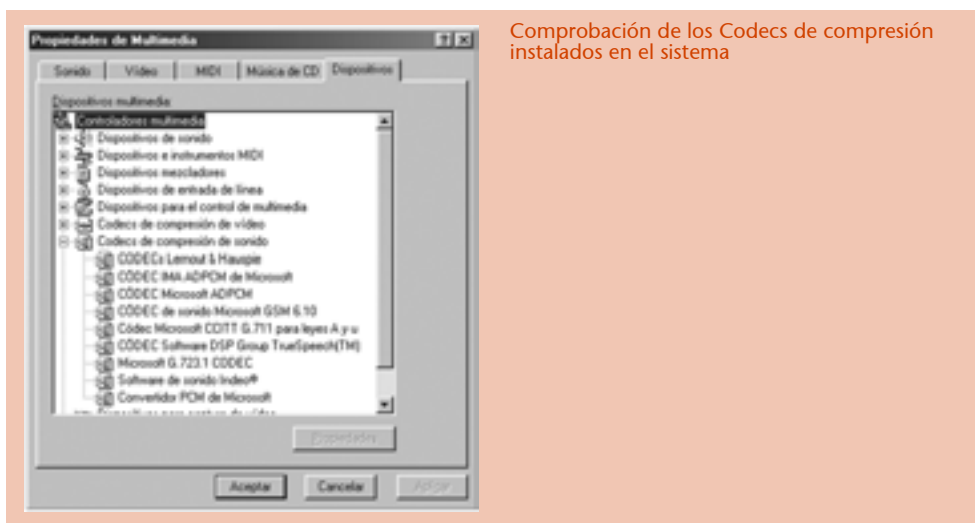
Un factor importante en los sistemas de compresión de audio es que interesa que sean capaces de comprimir y descomprimir en tiempo real (para comprimir y descomprimir en diferido existen ya multitud de sistemas de uso general como el ZIP o el ARJ, que obviamente también pueden ser aplicados a ficheros de audio).

Las matemáticas involucradas en cualquier sistema de compresión son demasiado complejas para ser tratadas con rigor en esta obra, por lo que nos limitaremos a esbozar ciertas ideas. Muchos de estos sistemas son independientes de los formatos

citados en el apartado “Formatos de sonido digital”, de modo que un mismo formato de fichero puede soportar varios métodos de compresión.

- Las siglas ADPCM corresponden a *Adaptative Delta Pulse Code Modulation*, una forma de compresión de la que existen múltiples variantes, en la cual se reduce el número de bits de la señal (por ejemplo a 4 bits) tratando únicamente las diferencias entre una muestra y la siguiente. Es fácil intuir que esto provoca una inevitable degradación de la señal, ya que el sistema es incapaz de representar saltos grandes de amplitud entre dos muestras.
- Las compresiones  $\mu$ -law y A-law utilizan un sistema de compresión no lineal que permite utilizar una resolución de 8 bits, pero ofreciendo una calidad sonora y un rango dinámico próximos a los obtenidos con 14 bits. Son, por lo general, más rápidos que los métodos basados en ADPCM.

Algunos de estos sistemas se ven favorecidos por la presencia de *hardware* especializado, como puede ser la inclusión de un DSP (procesador digital de señal) en la tarjeta de sonido, mientras que otros funcionan perfectamente por *software*. Si abris el icono de multimedia en el panel de control de Windows 95, y seleccionáis la forma de visualización **Avanzado**, podréis ver, en el apartado de **Codecs de compresión de audio**, todos los compresores instalados en su sistema, tal como se observa en la figura.



## La compresión MP3

El famoso MP3 surgió con posterioridad a los métodos previamente descritos.

A diferencia de estos últimos, basados todos ellos en principios matemáticos más o menos simples, el formato MP3, abreviación con que se conoce popularmente el MPEG-1 Layer III, utiliza algoritmos mucho más sofisticados inspirados en la



psicoacústica, y que tienen en cuenta, por lo tanto, la forma en que los humanos percibimos (o mejor dicho ¡no percibimos!) determinados sonidos.

Con este método, descubierto a mediados de los años ochenta en el Instituto Fraunhofer, en Alemania, se logran espectaculares compresiones del orden del 90%, y se consigue que un fragmento musical comprimido ocupe una décima parte de su tamaño original, con una pérdida de calidad casi imperceptible.

Para conseguir esta reducción, el MP3 se basa en el fenómeno del enmascaramiento descrito en el módulo “**Fundamentos de psicoacústica musical**”, que explica cómo, en determinadas condiciones, ciertos sonidos dejan de percibirse totalmente, ocultados o enmascarados por otros. El MPEG-1 Layer III aprovecha estas imperfecciones auditivas humanas para eliminar la información que no podemos percibir, manteniendo la máxima calidad en las zonas de mayor sensibilidad (de forma similar al caso de que, en una imagen 3D, eliminásemos toda la información de los objetos tapados por planos más próximos al observador).

Antes del auge que este formato ha obtenido en los últimos años gracias a Internet, con las consecuentes batallas legales por los derechos de autor que el fenómeno ha acarreado, variantes de este método de compresión ya se empezaron a utilizar en los dos sistemas de grabación digital domésticos surgidos a principios de los noventa, el DCC y MiniDisc.

## **Compresión y descompresión**

---

Las peculiares características del algoritmo hacen que mientras que la descompresión (y por consiguiente la reproducción) se puede realizar en tiempo real en un ordenador corriente, la compresión sólo pueda realizarse a tiempo real en ordenadores equipados con procesadores de última generación.

*Las siglas MP3 corresponden al formato de compresión MPEG-1 Layer III.*

*Sus principios están basados en la psicoacústica.*

*Con este método se consiguen reducciones del 90% con muy poca pérdida de calidad.*

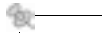
*La fase de descompresión se realiza en tiempo real durante la reproducción del fichero, pero la compresión es mucho más lenta.*

## Calidad y bit ratio

---

Este formato ofrece varios niveles de calidad posibles, seleccionables por el usuario a la hora de comprimir y determinados por la tasa de transferencia o bit ratio elegido.

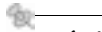
**La tasa que se utiliza en la mayoría de los ficheros disponibles en Internet es la de 128 Kbits/s**, es decir un valor cercano al 9% del que ofrece el audio digital convencional a 44.100 Hz y 16 bits.



Recordemos que un segundo de audio convencional ocupa: 44.100 muestras/segundo  $\times$  2 *bytes*/muestra  $\times$  2 canales, es decir, aproximadamente 176.400 *bytes*/segundo, y

este valor debe multiplicarse por 8 si queremos hablar de bps (i.e. bits/segundo), ya que cada *byte* está constituido por 8 bits, lo cual nos da un valor de 1.411.200 bps.

Con este ratio de 128 Kbps, la pérdida de calidad es casi inapreciable, y se podrá comprimir mucho más (hasta 8 Kbps) cuando se requieran tamaños muy inferiores (aunque en estos casos de compresión tan extrema la degradación sonora sí que resulta muy evidente).



### Más información

Miles de páginas en Internet ofrecen no sólo música en formato MP3, sino también todos los recursos necesarios para reproducir estos ficheros e incluso para comprimir nuestra propia música. Una buena dirección de partida la encontramos en: <http://www.mp3.com/>.

## Etapa 5: Historia del audio digital

### Grabación y generación digital de sonido

---

- Los primeros experimentos en audio digital datan de finales de los cincuenta, cuando Max Mathews obtuvo en los laboratorios Bell los primeros sonidos generados por ordenador. En esta época, la escasa potencia de las computadoras no permitía una velocidad suficiente para digitalizar el sonido. Los experimentos consistieron en obtener sonidos sintéticos generando los números en el ordenador. En módulos posteriores trataremos la síntesis y el procesado del sonido digital.
- Tras muchos experimentos que ayudaron a comprender mejor la naturaleza del sonido digital, el primer grabador comercial, el Sony PCM-1, vio la luz en 1977. El conversor era de 13 bits y utilizaba como soporte cintas de vídeo Betacam.
- Al año siguiente (1978), el modelo PCM-1600 ya utilizaba conversión de 16 bits.
- En 1982 apareció el disco compacto, desarrollado conjuntamente por Sony y Philips, con un éxito fulminante.
- A principios de los ochenta aparecieron también los primeros sintetizadores digitales y los primeros *samplers*. De ellos hablaremos profusamente en módulos posteriores.
- Los primeros sistemas domésticos de grabación digital no llegaron hasta finales de los ochenta, inicialmente con la cinta DAT (Digital Audio Tape), y, posteriormente, con el DCC (Digital Compact Cassette), el MiniDisc, la grabación a disco duro, los CD-R (grabadores de discos compactos) y varios formatos de multipistas digitales (Alesis, Tascam, etc.).
- Las tarjetas de sonido para ordenadores, dotadas de conversores A/D y D/A, aparecidas en los últimos años, han supuesto la definitiva democratización del sonido digital. A ellas dedicamos el apartado “Los ordenadores y el audio digital”.

### Los ordenadores y el audio digital

---

Hoy día, a todos nos parece normal disponer en nuestro ordenador de un sistema de audio digital con el que hubiese soñado “el más profesional de los músicos” hace poco más de 10 años.

Para concluir este módulo haremos un breve resumen histórico de lo que ha sido el binomio **ordenador/audio digital** a lo largo de las últimas décadas.

Tal como se comenta en el apartado “grabación y generación digital de sonido”, Max Mathews consiguió sintetizar sonido digital en un ordenador ya finales de la década de los cincuenta! Aunque se tardó unos años más en disponer de la velocidad suficiente para muestrear sonido analógico, las posibilidades abiertas por Mathews iniciaron dos largas décadas de investigaciones en el terreno de la síntesis digital, llevadas a cabo inevitablemente en ordenadores *mainframes* y estaciones de trabajo Unix, en centros universitarios o laboratorios de importantes empresas informáticas.

### **Audio digital para las masas**

- El Commodore Amiga fue el primer ordenador doméstico que incorporó, en 1985, ciertas posibilidades de audio digital. El Amiga disponía de cuatro conversores D/A de 8 bits. Carecía de conversores A/D, por lo que, si no se compraba un dispositivo muestreador adicional conectable al puerto paralelo, sólo funcionaba como reproductor. Estos primeros conversores D/A adicionales tenían una limitación importante, ya que por motivos de velocidad sólo permitían muestrear en memoria RAM, y no directamente al disco duro, por lo que la duración de los sonidos almacenables era muy limitada. Aun con todas estas limitaciones, que lo hacían inviable para un trabajo profesional, el Amiga supuso para muchos la entrada al sonido digital.
- Las primeras aplicaciones serias se dieron a finales de los ochenta en dos plataformas que, ya de por sí, no eran nada baratas (Apple Macintosh y NeXT), mientras que el coste del *hardware* adicional superaba fácilmente el medio millón de pesetas. No eran productos “para las masas”, pero los estudios de grabación y algunos músicos profesionales se los pudieron ya permitir. Especialmente en el caso del Mac, el *hardware* y el *software* de Digidesign, que ofrecía cuatro canales de 16 bits, se convirtió en un estándar (y lo sigue siendo hoy) de la grabación y edición de audio digital en disco duro.
- El bautizo sonoro del PC fue realmente tímido y tardío. En 1988, una compañía canadiense, Ad Lib, fabricó las primeras tarjetas de sonido. Soportaban tan sólo cuatro voces con sonidos sintéticos y, aunque muchos juegos comenzaron a tener en cuenta el nuevo *hardware*, los tipos de sonidos ofrecidos por el pequeño sintetizador permitían poco más que generar una cierta musiquilla de fondo, con una calidad sonora bastante infame. Al no disponer de sonido digitalizado, la Ad Lib no favorecía la reproducción de efectos convincentes (golpes, motores, etc.) o voces (o gritos), que el mercado de los juegos estaba esperando.
- Un año más tarde, Creative Labs, una empresa de Singapur que había trabajado en síntesis de voz, lanzó al mercado la primera Sound Blaster, que añadía a la compatibilidad con la Ad Lib unas pequeñas posibilidades de audio digital. Pese a que estas posibilidades eran limitadas (dos canales de 8 bits), el éxito fue

tremendo y propulsó a la compañía al primer puesto del mercado multimedia internacional, lugar que sigue ocupando en la actualidad.

- Hoy día todas las tarjetas ofrecen como mínimo dos canales de audio a 16 bits y hasta 44.100 Hz de frecuencia de muestreo, con posibilidad de grabación y reproducción a disco duro. Algunas, especialmente orientadas al músico semiprofesional, incorporan además un mayor número de salidas, así como entradas y salidas digitales. Cualquiera de ellas permite, sin embargo, con la ayuda del *software* necesario, descubrir las inmensas posibilidades del audio digital.

## Autoevaluación

---

1. ¿Qué es un transductor?
  - Cualquier dispositivo que convierta una magnitud física en otra.
  - Un conversor de sonido analógico a digital.
  - Un dispositivo integrado dentro del ordenador.
  - Un micrófono.
  
2. La digitalización de sonido...
  - presupone conectar un micrófono al ordenador.
  - puede producirse en una única fase.
  - se produce en dos fases: cuantificación y muestreo, pero el orden no es fundamental.
  - se produce en dos fases: muestreo y cuantificación, y siempre en este orden.
  - supone un incremento en la calidad sonora.
  
3. Muestrear...
  - implica aplicar frecuencias a una señal analógica.
  - implica modificar la frecuencia de una señal.
  - significa convertir el sonido en números que el ordenador pueda gestionar.
  - significa lo mismo que discretizar.
  - significa tomar valores de cualquier magnitud a intervalos regulares.
  
4. El teorema de Nyquist...
  - determina cuál es la máxima frecuencia de muestreo aplicable para una señal dada.
  - determina cuál es la máxima frecuencia que puede tener una señal.
  - determina cuál es la máxima frecuencia que se puede muestrear correctamente para una frecuencia de muestreo determinada.
  - determina que la frecuencia de muestreo debe ser como máximo el doble de la frecuencia de la señal.
  
5. La frecuencia de Nyquist...
  - es igual a 22.050 Hz.
  - es igual a la frecuencia de muestreo.
  - es igual al doble de la frecuencia de muestreo.
  - es la máxima frecuencia que puede tener un sonido.
  - es la máxima frecuencia que se puede muestrear correctamente.
  - es la mínima frecuencia que puede tener un sonido.
  
6. ¿Por qué es frecuente muestrear a 44.100 Hz?
  - No es excesivamente frecuente.
  - Porque así no se consume demasiada memoria.
  - Porque el oído humano no detecta frecuencias aproximadamente por encima de los 20.000 Hz.
  - Porque es el doble de la frecuencia de Nyquist.
  - Porque se obtiene más calidad que si muestreásemos a 48.000 Hz.

7. ¿Cuál es la frecuencia de Nyquist de un disco compacto?
  - 11.025 Hz.
  - 20.000 Hz.
  - 22.050 Hz.
  - 44.100 Hz.
  - 48.000 Hz.
  - 96.000 Hz.
  - Depende de la grabación.
  
8. Muestrear a frecuencias menores...
  - ahorra espacio pero pierde calidad.
  - consume más espacio y pierde calidad.
  - es más rápido.
  - gana calidad a costa de consumir más espacio.
  - nunca es recomendable.
  - produce sonidos más brillantes.
  - puede producir frecuencias “fantasmas” si no se filtra la señal convenientemente.
  
9. Si queremos ahorrar espacio o memoria, ¿qué es más conveniente?
  - Muestrear a 22.050 Hz y filtrar con un paso bajo a 44.100 Hz.
  - Muestrear a 44.100 Hz, filtrar con paso alto a 22.050 Hz y convertir a 22.050 Hz.
  - Muestrear a 44.100 Hz, y filtrar con un paso bajo a la mitad de la frecuencia a la que queramos convertir.
  - Muestrear a cualquier frecuencia, siempre que sea inferior a 22.050 Hz.
  - Muestrear directamente a 22.050 Hz.
  - Muestrear directamente a 48.000 Hz.
  
10. El espectrograma...
  - es la representación de un sonido en el dominio frecuencial.
  - es un gráfico con cuatro dimensiones.
  - es un gráfico con escala de grises.
  - permite representar sonidos mediante colores.
  - representa la evolución temporal de todas las frecuencias de un sonido.
  - representa la frecuencia de Nyquist en función del tiempo.
  
11. El *aliasing*...
  - añade frecuencias “fantasmas” a un sonido digitalizado.
  - casi nunca es perceptible.
  - es más importante cuanto más agudo sea un sonido.
  - no es deseable, pero tampoco evitable.
  - se produce cuando muestreemos una señal con una frecuencia de muestreo demasiado alta.
  
12. La frecuencia de muestreo de una tarjeta de sonido...
  - depende de la resolución en bits de la tarjeta.
  - determina la calidad de la tarjeta.
  - es fija para cada tarjeta.

- es igual a 44.100 Hz.
- es variable y configurable por *software* de forma automática.
- no puede ser superior a 48.000 Hz.

13. El rango dinámico de un sistema de sonido...

- depende de las amplitudes máxima y mínima que el sistema puede reproducir.
- es igual al número de bits de resolución por 8.
- es superior en los sistemas digitales.
- se expresa en unidades de presión o de voltaje.

14. Cuanto mayor sea la resolución de un sistema digital, ...

- más débil sonará.
- más fuerte sonará.
- más niveles de amplitud y menos ruido de fondo presentará el sonido.
- más ruido de fondo presentará el sonido.
- mayor será el error de cuantificación.
- menor será el rango dinámico.
- menos bits se emplearán para codificar el sonido.

15. Una resolución de 16 bits...

- da un rango dinámico máximo de 96 dB.
- da un rango dinámico mínimo de 96 dB.
- es la máxima posible.
- es la que se utiliza en los discos compactos y DVD.
- es mejor que una de 24 bits.
- presenta 32.767 posibles niveles.

16. El sobremuestreo...

- aumenta en 12 dB la relación señal/ruido.
- es una técnica que permite mejorar la relación señal/ruido a costa de aumentar la resolución real.
- es una técnica que permite mejorar la relación señal/ruido sin aumentar la resolución real.
- hace que los ficheros de sonido ocupen el doble.
- hace que los ficheros de sonido ocupen la mitad.

17. Un minuto de sonido monofónico a 16 bits y 44.100 Hz...

- cabe perfectamente en un disquete.
- ocupa aproximadamente 10 Mbytes.
- ocupa aproximadamente 172 Kbytes.
- ocupa aproximadamente 5 Mbytes.
- ocupa un tamaño que depende de la frecuencia de salida.
- ocupa un tamaño que depende de la resolución.

18. El MP3...

- es la abreviación de MPEG-II Layer 3.
- es la abreviación de MPEG-III.



- es un formato de audio mejor que el WAV.
- es un sistema de compresión/descompresión de sonido.
- es un sistema de intercambio de ficheros por Internet.
- se inventó en torno a 1995.

