

2. Principios del sonido digital

2.1 Introducción

El principio fundamental del audio digital consiste en discretizar las señales sonoras continuas (como las emitidas por un micrófono) para convertirlas en secuencias de números. La discretización de estas señales se lleva a cabo en dos niveles diferentes, el temporal y el de la amplitud. En la figura 2.1 se muestra una señal continua, discretizada sólo en el tiempo (cuadros blancos) y en el tiempo y la amplitud (puntos negros). Cuanto menor sea la cuadrícula, mayor similitud existirá entre la señal original y la señal digitalizada.

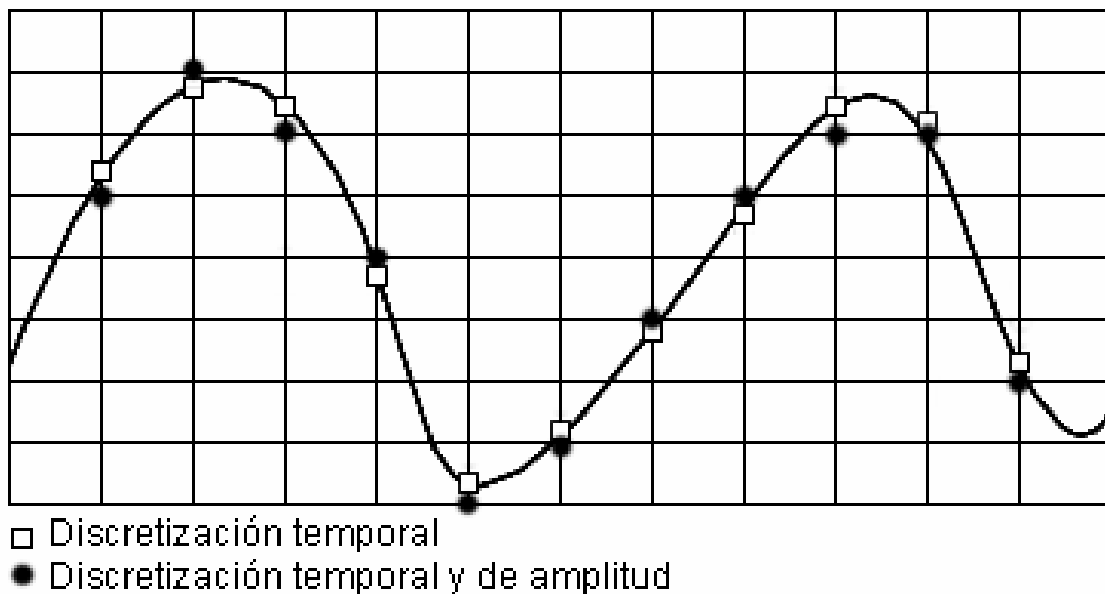


Figura 2.1. Discretización de una señal continua.

2.2. Comparación con el cine y el vídeo 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.

Pero en el caso del vídeo digital se producen dos discretizaciones adicionales. La segunda convierte cada *frame* en una matriz de puntos (por ejemplo 800x600), 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, y si el número de valores posibles para cada punto es pequeño, perderemos matices en los colores y en los brillos. En la figura 2.2 se muestran estos casos. Algo parecido sucede con el sonido.



Imagen original



imagen con baja resolución espacial



imagen con baja resolución de color



imagen con baja resolución espacial y de color

Figura 2.2. Ejemplos de imágenes deficientemente digitalizadas.

2.3. Muestreo del sonido - Frecuencia de muestreo

La palabra muestreo es el equivalente del término inglés *sampling*, y se utiliza para indicar la acción de tomar muestras a intervalos de tiempo regulares. Para digitalizar un sonido es necesario muestrearlo, pero ¿con qué frecuencia? En el apartado anterior comentábamos 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, haciéndole creer que lo que recibe no son imágenes discontinuas, sino un flujo continuo de luz, pero estos valores son *totalmente insuficientes* en el caso del sonido.

Para comprender este fenómeno observemos la figura 2.3 que representa una señal sinusoidal de 30 Hz. (el ciclo se repite 30 veces por segundo) ¿Que ocurre si la muestreamos a una frecuencia de 20 Hz? En este caso, estaremos tomando el valor de la onda original cada 0,05 segundos (corresponden a los seis cuadros de la figura). La nueva señal obtenida juntando los puntos muestreados tiene un período aparente de 0,1 segundos, ¡es decir una frecuencia de 10 Hz, lo cual no se corresponde en absoluto con la señal original!

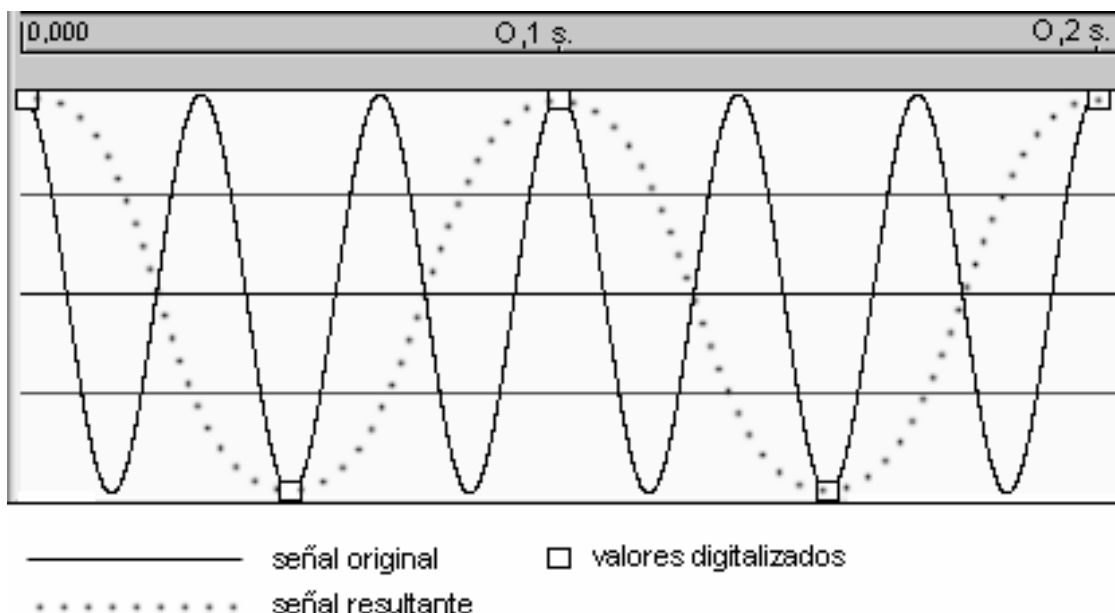


Figura 2.3. Señal muestreada incorrectamente.

¿Cual sería la mínima frecuencia de muestreo correcta?

El teorema del muestreo o teorema de Nyquist, afirma que para muestrear correctamente una señal de X Hz, se requiere como mínimo una frecuencia de muestreo de 2X Hz.

En nuestro ejemplo anterior, una frecuencia de 60 Hz hubiese pues bastado para muestrear correctamente la señal original. Pero tal como indicamos en el capítulo 1, "Principios de acústica", el oído humano es capaz de detectar frecuencias sonoras de hasta aproximadamente 20.000 Hz, por lo que, para muestrear correctamente *cualquier* sonido se necesitará una frecuencia de muestreo superior o igual a 40.000 Hz. ¡De aquí proceden los famosos 44.100 Hz. utilizados en los discos compactos!

2.4. Muestreo con frecuencias inferiores - El aliasing

Si muestreamos a una frecuencia inferior, como por ejemplo a 30.000 Hz, los resultados que obtengamos podrán ser correctos para las frecuencias por debajo de los 15.000 Hz (la mitad de la frecuencia de muestreo). Para las aplicaciones que no precisen de la máxima fidelidad, esto es más que suficiente. El sonido 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.

Existe, sin embargo un problema adicional, algo complicado de entender. Tal como vimos en el ejemplo gráfico anterior, al muestrear con frecuencias inferiores, surgen frecuencias “fantasmas” que realmente no están en el sonido original (como los 10 Hz que aparecían misteriosamente en nuestro ejemplo) y que alteran el sonido muestreado. Este fenómeno recibe el nombre de *aliasing*.

Si lo que sigue le parece excesivamente complicado, no se preocupe: volveremos a tratar el tema con más detalle en el apartado 5.7. “Efectos simples aplicables a la frecuencia”, pero para los más impacientes, he aquí un pequeño adelanto.

El *aliasing* “inventa” frecuencias de valores aproximados a la diferencia entre la frecuencia original y la frecuencia de muestreo. Para evitarlo, 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. Si decidimos muestrear a 20.000 Hz, deberíamos filtrar la señal original, eliminando todos sus componentes 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 permiten modificar el 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 (e.g. 10.000 Hz)
- reconvertir por software el sonido a la frecuencia deseada (e.g. 20.000 Hz)

Parece muy complicado, pero no se apure; como veremos, la mayoría de editores de sonido facilitan bastante esta tarea.

2.5. Número de bits, resolución y rango dinámico

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 ser manejado por un ordenador. Para ello, volveremos momentáneamente al ejemplo del vídeo digital, que para muchos lectores resultará más familiar. La resolución de color de una imagen (y de las tarjetas de vídeo) se mide en bits. 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. Con el sonido sucede algo parecido: cuantos más bits apliquemos, más niveles o escalones tendrá el

sonido digitalizado y más parecido será en consecuencia al sonido analógico original (que al ser continuo, poseía un número infinito de niveles).

El rango dinámico en decibelios viene dado aproximadamente por la fórmula:

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

Si en un sistema digital consideramos la amplitud máxima como 2^{bits} y la amplitud mínima como 1, 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. Dado que el umbral de dolor no se sitúa hasta alrededor de los 130 dB, sería deseable que un sistema de alta fidelidad pudiese alcanzar este rango dinámico. Para ello serían necesarios unos 22 bits de resolución. Todos sabemos que no es éste el estándar que se ha adoptado, que ha quedado establecido en 16 bits. Por ello, *todo sistema digital convencional posee un rango dinámico máximo de 96 dB* (valor teórico que en la práctica se ve disminuido por factores de circuitería).

2.6. La calidad del sonido digital

Estos valores de 44.100 Hz y 16 bits adoptados en el disco compacto y otros dispositivos digitales, son los que nos llevan a afirmar que el sonido digital de alta fidelidad no es todo lo bueno que podría ser, ya que ciertos equipos analógicos superan estas cotas. No se puede negar sin embargo que el sonido digital ofrece muchas ventajas, entre las que hay que destacar el menor desgaste y sobretodo, la posibilidad de realizar copias idénticas, es decir, sin ninguna pérdida. El formato digital permite además, tal como iremos viendo, muchas más posibilidades de manipulación y procesado.

Por otro lado, cuando no se necesite una calidad de alta fidelidad, se pueden utilizar especificaciones inferiores. En este sentido, una frecuencia de muestreo de 11.025 Hz y una resolución de 8 bits, ofrecen una calidad comparable a la de una línea telefónica convencional, por lo que serán suficientes para algunas aplicaciones de voz, donde lo que se busque sea la comprensión de un mensaje, al margen de preocupaciones estéticas.

2.7. Los conversores A/D y D/A

Hasta aquí la teoría. Veamos ahora como funciona un sistema digitalizador de sonido. El sistema se compone de dos conversores: un convertor analógico/digital (a partir de ahora A/D) en la entrada, y un convertor 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 (véase apartado 6.4.3), que elimina de la señal analógica todos los componentes con frecuencias superiores a los 22.050 Hz. En el proceso de digitalización, el convertor A/D genera un número binario (de 8 ó 16 bits) a cada pulso del reloj. Este número es almacenado en la memoria o grabado 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 convertor 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 figura 2.4 esquematiza este proceso.

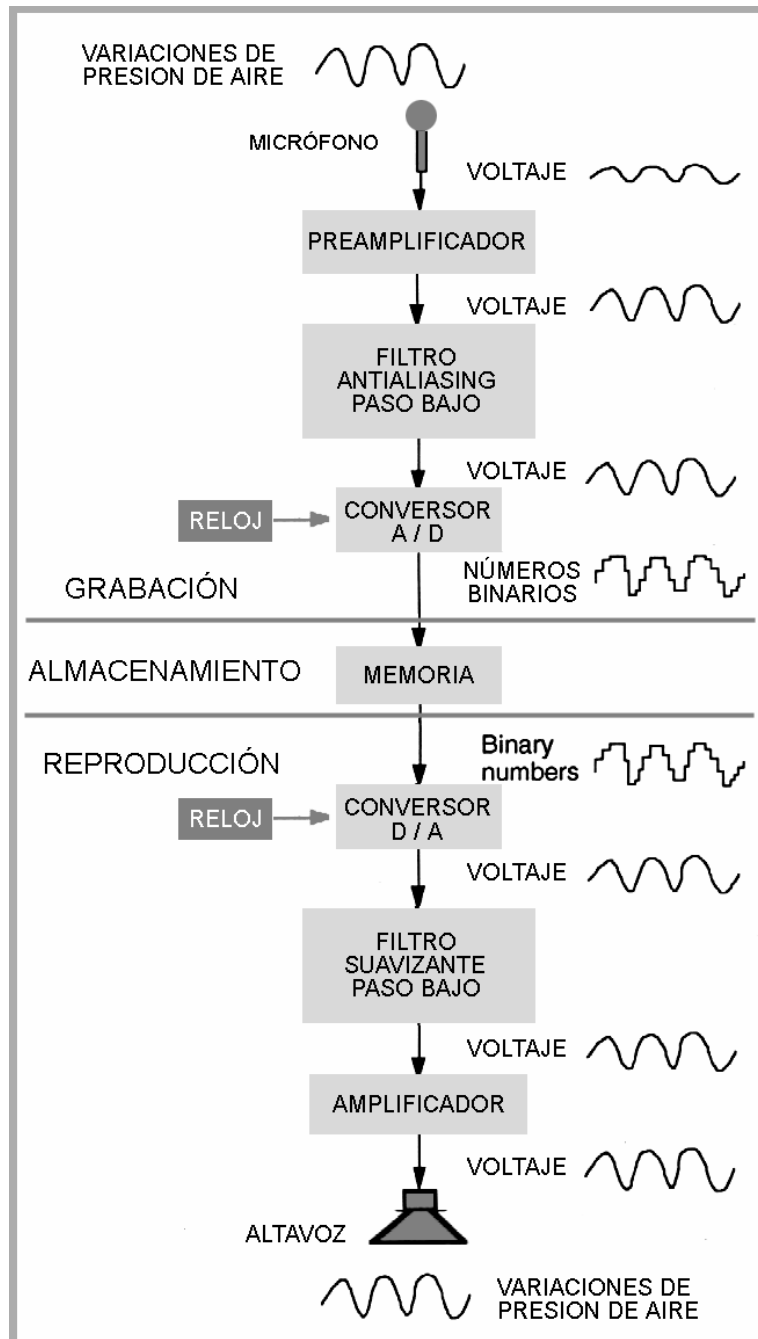


Figura 2.4. Esquema de un sistema de conversión A/D y D/A

2.8. Breve historia del audio digital

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.

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 Beta (al año siguiente, 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, pero de ello hablaremos profusamente en el capítulo 9, “Síntesis y generación digital de sonido”.

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 próximo capítulo.