

PREGUNTAS FRECUENTES SOBRE AUDIO DIGITAL

Federico Miyara.

Universidad Nacional de Rosario (UNR)

<http://www.eie.fceia.unr.edu.ar/~acustica/audigfaq.htm>

P1. ¿Qué es el *audio digital*?

R: Es la representación de una señal de audio mediante números, en general codificados en forma binaria (es decir con ceros y unos).

P2. ¿Por qué se utiliza el *audio digital*?

R: Porque tiene algunas ventajas sobre el audio analógico que lo hacen insustituible. En primer lugar permite ser almacenado en forma inalterable. Dado que lo que se almacenan son números, es decir símbolos, es mucho más difícil alterar la información guardada que en el caso en que se guarda un campo magnético proporcional a la señal, como en un cassette. Otra ventaja es que permite aprovechar la tecnología de procesamiento digital de señales para introducir efectos, modificaciones o mejoras imposibles o muy difíciles de lograr analógicamente. Por ejemplo, es posible conseguir retardos, efectos de reverberación, supresión de ruido, etc.

P3. ¿Cómo se pasa de una señal eléctrica a una señal digital?

R: Se utiliza un proceso de *muestreo* (discretización en el tiempo) y *digitalización* (discretización en amplitud). El muestreo consiste en tomar valores de la señal a intervalos regulares de tiempo. La digitalización consiste en subdividir el rango útil total de la señal en cierta cantidad de "casilleros" o subintervalos numerados, y asignar a cada muestra el número de subintervalo en el cual se encuentra. Por ejemplo, si el rango de una señal que varía entre 0 y 10 V se subdivide en 16 subintervalos, a una muestra de 7,3 V se le asignará un número igual a la parte entera de $7,3 \cdot 16 / 10 = 11,68$, es decir 11. Este proceso es llevado a cabo por un *convertor analógico/digital*.

P4. ¿Cómo quedan físicamente representados los números asignados después de la digitalización?

R: Pueden representarse de varias formas, pero todas ellas utilizan la *numeración binaria*. En esa numeración sólo se usan los dígitos 0 y 1, denominados *bits* (*b*inary *d*igits, o *d*ígitos *b*inarios). Así como en la numeración decimal al llegar a 9 se pasa a 10, en la numeración binaria al llegar a 1 se pasa a 10. Una vez en versión binaria, cada bit puede estar representado por una tensión eléctrica (por ejemplo "0" = 0 Volt y "1" = 5 Volt), por un campo magnético (por ejemplo "0" = Sur y "1" = Norte), por un haz luminoso (por ejemplo "0" = oscuro y "1" = iluminado), etc. El conjunto de bits que representan a una única muestra se suelen llamar *palabra*.

P5. ¿Cómo se elige la cantidad de subintervalos en que se divide el rango útil de la señal?

R: Normalmente se elige como una potencia de 2, de manera que los valores asignados a las muestras están entre 0 y $2^n - 1$, donde n corresponde a la cantidad de *bits*, es decir de *dígitos binarios*.

P6. ¿Qué es la *resolución* de un sistema de audio digital?

R: Es la cantidad de bits que se utiliza para representar las muestras de audio, es decir la cantidad de bits que conforman cada *palabra* (ver P4). Cuanto mayor sea la resolución, más precisa será la representación. Por ejemplo, con una resolución de 8 bits, el rango de variación de la señal se divide en 256 subintervalos, mientras que con una resolución de 16 bits lo hace en 65536 subintervalos, cuya amplitud será, por consiguiente, mucho menor. El audio digital para el consumo masivo (por ejemplo el CD o las placas de sonido de las computadoras) tiene una resolución de 16 bits. En sistemas de audio profesional se utilizan 20 bits y aún 24 bits

P7. ¿Qué es la *frecuencia de muestreo*?

R: También llamada *tasa de muestreo*, es la cantidad de muestras por unidad de tiempo. Cuanto mayor sea, mayor es la respuesta en frecuencia del sistema. El estándar para los discos compactos (CD) es de 44,1 kHz.

P8. ¿Cómo se elige la frecuencia de muestreo?

R: Debe ser mayor que el doble de la máxima frecuencia $f_{\text{máx}}$ *presente* en la señal. Esta condición se denomina *condición de Nyquist*. Obsérvese que no es suficiente que sea mayor que el doble de la máxima frecuencia *útil*, ya que si hay ruido por encima de ésta, podría producirse un tipo de distorsión denominado *aliasing*

P9. ¿Qué es el *aliasing*?

R: Según el denominado *teorema del muestreo*, si se muestrea con una frecuencia que no cumple la condición de Nyquist (ver pregunta anterior) al intentar reconstruir la señal se generan frecuencias espurias que no estaban presentes originalmente. Supongamos, por ejemplo, que queremos muestrear una señal audible que contiene además un ruido de 35 kHz. Si utilizamos la frecuencia normalizada de 44,1 kHz, a pesar de que ese ruido es originalmente inaudible (por ser mayor que el límite superior de 20 kHz del oído humano), al intentar recuperar la señal aparecerá un ruido de 9,1 kHz (= 44,1 kHz - 35 kHz), que es perfectamente audible. Este tipo de frecuencias que aparecen *dentro* del espectro útil se denominan *frecuencias "alias"*.

P10. ¿Qué sucede cuando la frecuencia de muestreo está fijada por las características del sistema (por ejemplo, porque va a ser utilizada en un CD), y por lo tanto no puede elegirse libremente para cumplir la condición de Nyquist?

R: En ese caso hay que actuar sobre la señal. Se utiliza un *filtro antialias*, que suprime todas las frecuencias por encima de la *frecuencia de Nyquist*, es decir la mitad de la frecuencia de muestreo f_M . En el caso del CD, que utiliza una frecuencia de muestreo de 44,1 kHz, el filtro antialias debe conservar todas las frecuencias por debajo de 20 kHz y eliminar todas las que están por encima de 22,05 kHz (= 44,1 kHz / 2).

P11. La señal almacenada en un CD, por ejemplo, es una señal digital. ¿Cómo se transforma nuevamente en una señal audible?

R: Se utiliza un *conversor digital/analógico*. Este dispositivo recibe las sucesivas muestras digitalizadas y las transforma en valores de tensión eléctrica mediante un factor de escala. Por ejemplo, si el factor de escala es de 10/16 V, una muestra igual a 11 se transformará en un valor de tensión de $11 \cdot 10/16 = 6,875$ V. El valor de tensión que corresponde a cada muestra se mantiene constante hasta que llega la

próxima muestra. Resulta así una onda escalonada formada por tramos constantes.

P12. La señal reconstruida ¿coincide exactamente con la original?

R: No. Si comparamos los ejemplos de las preguntas 3 y 9, vemos que un valor de señal de 7,3 V se "reconstruyó" como 6,875 V, introduciéndose un error de -0,425 V. El error será tanto menor cuanto más pequeños sean los subintervalos en que se divide el rango útil de la señal, es decir, cuanto mayor sea la resolución en bits. La evolución en el tiempo de este error se denomina *ruido de digitalización*.

P13. ¿Qué relación hay entre el ruido de digitalización y la resolución?

R: La mejor manera de evaluar el ruido de cualquier sistema (incluidos los de audio digital) es a través de la relación señal/ruido (S/R) en decibeles. Para el audio digital, la máxima S/R que puede obtenerse es, aproximadamente, igual a $6 \cdot n$, donde n es la resolución en bits. Por ejemplo, un sistema de 16 bits, como el disco compacto (CD), admite una S/R de $6 \cdot 16 = 96$ dB. Nota: Debido a limitaciones en la parte analógica, la S/R suele ser menor que ese valor, por ejemplo 90 dB.

P14. ¿Cómo afecta el hecho de que la señal reconstruida es escalonada en lugar de coincidir con la señal original?

R: Su efecto no debería ser importante, ya que genera frecuencias por encima del espectro audible. Sin embargo, es conveniente agregar un *filtro de suavizado* que limite el contenido de frecuencias a lo estrictamente necesario, para evitar la presencia de frecuencias que podrían interferir con otros procesos, produciendo *batidos* audibles, por ejemplo.

P15. Si la relación señal/ruido aumenta con el número de bits ¿por qué no se aumenta indefinidamente la cantidad de bits?

R: Primero, porque sería impráctico, ya que obligaría a manejar una gran cantidad de información simultáneamente, lo cual implica un elevado costo. De todas maneras, el costo por bit es cada vez menor. Al principio se usaban 8 bits, luego se estableció el estándar de 16 bits para el audio digital comercial. Hoy en día los profesionales trabajan con conversores de hasta 24 bits. Pero hay otra cuestión más. Una resolución de 24 bits implica una relación señal/ruido de 144 dB (ver P13). Ello significa que si la señal es de 4 V (un valor considerado muy alto para una señal de nivel de línea), entonces el ruido de digitalización estará 144 dB por debajo, que, a cálculo hecho, son 0,25 microvolts (esto significa que el salto que se produce entre el escalón correspondiente a un valor digital y el escalón que le sigue es de 0,25 microvolt). Ahora bien, casi todas las salidas de línea tienen una resistencia (impedancia) de salida del orden de 100 ohms. Toda resistencia tiene un ruido eléctrico (de origen térmico) que, calculado para este valor de resistencia, da 0,18 microvolt. Esto significa que con 24 bits estamos prácticamente al límite de lo que puede lograrse con la electrónica analógica. La mayor relación S/R que idealmente podría lograrse con una resolución mayor sería inaprovechable a causa del ruido térmico. Por otra parte, el ruido circuital no es sólo térmico. Los semiconductores en general producen bastante ruido, siendo muy raros (y costosos) los circuitos con relaciones señal/ruido mayores de 120 dB.

P16. A la luz de la respuesta anterior ¿cómo se explica que haya programas de edición de sonido que trabajen con 32 y hasta con 64 bits?

R: Es tratar de reducir los errores acumulativos por truncado en los procesamientos que requieren una gran cantidad de operaciones. Para comprenderlo, y utilizando numeración decimal para simplificar el ejemplo, supongamos que se requiere multiplicar el número 1 veinte veces por 0,800 (un tipo de operación que podría formar parte de una reverberación artificial), y supongamos que se desea una precisión de sólo 3 cifras significativas. Trabajando primero con toda la precisión que permite una calculadora de 10 cifras, tendremos: 0,800 - 0,640 - 0,512 - 0,4096 - 0,32768 - ... - 0,01152929215. Truncando el resultado para conservar sólo las 3 primeras cifras decimales, resulta **0,011**. Si ahora en lugar de truncar al final se trunca al cabo de *cada multiplicación*, tendremos: 0,800 - 0,640 - 0,512 - 0,409 - 0,327 - ... - **0,009**. Vemos que el resultado obtenido difiere del anterior. Esto mismo sucede en el caso del software de edición, sólo que en lugar de veinte operaciones podrían ser decenas de miles. Al trabajar internamente con mayor precisión (por ejemplo, 32 bits) y truncar recién al final para adaptarse a la cantidad de bits del formato utilizado (por ejemplo 16 bits), el error por truncamiento se reduce considerablemente.

P17. ¿Cuál es la relación señal/ruido necesaria para una buena calidad de reproducción?

R: Debería ser comparable con el rango dinámico del oído, que es la diferencia entre el umbral de dolor y el umbral de audición. En el caso más extremo, es decir el de personas jóvenes con excelente audición, estos umbrales están cerca de 120 dB y 0 dB respectivamente, por lo cual una relación señal/ruido de 120 dB debería ser suficiente para las mayores exigencias. Sin embargo, en general las condiciones de escucha normales no permiten llevar a la práctica esta relación señal/ruido, ya que es muy difícil lograr ambientes con ruido de fondo inferior a 20 dB. Por lo tanto una relación señal/ruido de 100 dB debería resultar suficiente en la mayor parte de los casos

P18. Entonces ¿por qué se trabaja con 20 bits y 24 bits, obteniendo relaciones S/R de 120 dB y 144 dB respectivamente?

R: Primero porque permite una mejor calidad en la conversión. Un conversor de 20 bits es mucho más lineal y tiene menor ruido que uno de sólo 16. Segundo, porque da formatos directamente compatibles con las nuevas tecnologías (por ejemplo el DVD). Para mayor información, hay un listado de preguntas y respuestas sobre 24 bits en la dirección <http://www.sonicsense.com/24bitfaq.html>.

P19. ¿Qué es *dither* y para qué se usa?

R: Cuando se está digitalizando una señal de muy poca amplitud, los saltos discretos entre escalones sucesivos adquieren una dimensión comparable con la amplitud de la propia señal. Esto implica que la forma de onda sufre una distorsión que resulta ser perfectamente audible y molesta. Esto es porque además del espectro del sonido propiamente dicho se agregan sus armónicos, que contienen energía concentrada en el espectro en frecuencias discretas. En otras palabras, la energía del *ruido de digitalización* está concentrada. Se ha encontrado que si, antes del muestreo, se agrega una pequeña cantidad de ruido aleatorio (de espectro continuo y no discreto), al cabo del proceso de digitalización la señal resultante también tiene la energía correspondiente al ruido de digitalización distribuida, en lugar de concentrada. Desde el punto de vista de la relación señal/ruido, hubo un ligero empeoramiento, pero desde el punto de vista perceptivo, el ruido se ha

vuelto mucho más tolerable e imperceptible. Hasta se puede trabajar con la forma de su espectro para hacerlo menos notorio. El ruido agragado se denomina *dither* (en inglés, una especie de temblor, como al tiritar).

P20. ¿Por qué a veces se habla de dither *digital*?

R: Esto sucede en los procesos de recuantización para pasar, por ejemplo, de 24 bits a 16 bits. Si simplemente se truncaran los 8 bits menos significativos, estaríamos en presencia de algo equivalente a un muestreo y digitalización sin dither. En este caso, podría agregarse un dither generado analógicamente, pero también es posible agregar uno producido digitalmente en la forma de una sucesión de números pseudoaleatorios (es decir, obtenidos por un algoritmo de cómputo que si bien es determinístico aparenta ser aleatorio). Este dither se genera con la resolución original (más alta), y luego simplemente se redondea.

[E-mail: fmiyara@fceia.unr.edu.ar](mailto:fmiyara@fceia.unr.edu.ar)

[Arriba](#)

[Biblioteca virtual](#)

[Home](#)

[English](#)