

Capítulo 15

Audio Digital

15.1. Introducción

Las técnicas digitales han cobrado en las últimas décadas una importancia fundamental en el desarrollo de nuevas tecnologías para la generación, el procesamiento, el almacenamiento y el análisis del sonido. Ello ha sido posible gracias al avance vertiginoso de la microelectrónica y su aplicación a la producción de dispositivos poderosos y complejos capaces de manejar y transformar cada vez más con mayor precisión y rapidez la enorme cantidad de información contenida en el sonido.

Una de las primeras consecuencias de la aplicación de la tecnología digital al audio fue el desarrollo de sistemas de almacenamiento del sonido de gran confiabilidad, inalterabilidad y fidelidad. Otra fue el gran impulso al desarrollo de instrumentos musicales electrónicos de gran complejidad y versatilidad. La tercera consecuencia fue el desarrollo y aplicación de técnicas para el procesamiento de la señal sonora, que permitieron no sólo el mejoramiento de procesos que antes se llevaban a cabo analógicamente sino también la introducción de nuevos procesos, entre los cuales se encuentran una gran cantidad de efectos tales como retardos, modulaciones, reverberaciones y espacializaciones de gran realismo y naturalidad, cuya implementación analógica sería mucho más costosa y por lo tanto destinada a un mercado mucho más restringido.

La idea básica detrás del audio digital es la de *representar el sonido por medio de números* ("digital" viene de *dígito*, es decir número). Aún antes de profundizar el análisis, es fácil ver que esto tiene varias ventajas. En primer lugar, se elimina el problema de la alterabilidad de la información. *Es mucho más fácil guardar un número que la magnitud física que ese número representa.* Por ejemplo, si quisiéramos guardar una varilla de 57,235 cm de longitud, tendríamos serias dificultades. Efectivamente, la dilatación a causa de la temperatura, o cualquier partícula de polvo que se adhiriera a sus extremos, o simplemente el desgaste, podrían ocasionar un error. Esto, que es válido para la longitud de una varilla, lo es más para el campo magnético almacenado en una cinta grabada.

En segundo lugar, existen algoritmos (métodos de cálculo) para realizar digitalmente no sólo todos los tipos de procesamiento utilizados en el audio tradicional, como la *amplificación, la mezcla, la modulación, el filtrado, la compresión y expansión, etc.*, sino muchos otros más, entre los cuales se encuentran los *retardos, los sincronismos, los desplazamientos de frecuencia, la generación de sonidos por diversos procedimientos, etc.* Estos algoritmos pueden implementarse en una computadora de propósito general o bien en dispositivos específicos llamados **procesadores digitales de señal (DSP)**.

En tercer lugar, el reemplazo de los procesadores analógicos por sus equivalentes digitales permite evitar la degradación de la señal a causa del ruido analógico, lo cual es conveniente dado que el ruido analógico es muy difícil de eliminar.

15.2. Numeración binaria

Dado que todos los sistemas digitales se basan en la numeración binaria, antes de comenzar a describir los procesos básicos de muestreo y digitalización del sonido nos referiremos brevemente a esa numeración. En la numeración decimal (el sistema que empleamos habitualmente), se utilizan diez símbolos (los dígitos 0, 1, 2, ..., 9) en un sistema *posicional* para representar las sucesivas cantidades. Esto significa que cada nueva cifra que se agrega tiene un peso 10 veces mayor que la que se encuentra a su derecha. Por ejemplo,

$$27 = 2 \times 10 + 7,$$

$$306 = 3 \times 10^2 + 0 \times 10 + 6.$$

En la numeración binaria, se utilizan sólo dos símbolos (los dígitos 0 y 1), también en un sistema *posicional*, sólo que ahora cada nueva cifra tiene un peso sólo 2 veces mayor que la anterior. Por ejemplo,

$$101 = 1 \times 2^2 + 0 \times 2 + 1 = 5,$$

$$11011 = 1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 1 \times 2 + 1 = 27.$$

En esta numeración, el 1 tiene características similares al 9 del sistema decimal, es decir, una vez que llegamos al 1 debemos agregar una nueva cifra 1 y cambiar la primera por 0. En la Tabla 15.1 se muestra la conversión de decimal a binario para los números del 0 al 15.

La razón por la que se utilizan los números binarios es porque eléctricamente es muy fácil codificar los 0's y los 1's. Basta utilizar un nivel de tensión alto (5 V) para un 1 y un nivel de tensión bajo (0 V) para un 0. Esto hace que la representación sea extremadamente insensible al ruido. En efecto, la señal seguiría siendo recuperable aún en presencia de un ruido de 2 V, que corresponde a una relación señal/ruido tan baja como $20 \log 5/2 = 8 \text{ dB}$ (inadmisible si el sistema fuera analógico).

15.3. Muestreo

Pasemos ahora al concepto de muestreo (*sampling*). Las señales acústicas (y por lo tanto las señales eléctricas que las representan) varían en forma continua, lo que significa que en un intervalo de tiempo dado, por pequeño que sea, existen *infinitos* valores diferentes. Sin embargo, a los efectos del mensaje auditivo, no hace falta tanta información. Primero, porque el oído no tiene tanta *discriminación en el tiempo*, y segundo porque tampoco tiene tanta *discriminación en la amplitud* como para distinguir valores que por estar muy próximos en el tiempo difieren muy poco en amplitud. No solamente no



Tabla 15.1. Conversión del sistema binario al decimal.
Los ceros de más a la izquierda son opcionales.

Decimal	Binario
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001
10	1010
11	1011
12	1100
13	1101
14	1110
15	1111

hace falta tanta información sino que desde el punto de vista práctico tampoco es conveniente ni posible manejarla. Entonces surge el concepto de muestreo (sampling). Muestrear una señal significa *reemplazar la señal original por una serie de muestras tomadas a intervalos regulares*. La frecuencia con la que se toman las muestras se denomina *frecuencia de muestreo*, f_M , y el tiempo entre muestras, *periodo de muestreo*, T_M . Se cumple que

$$f_M = \frac{1}{T_M}$$

En la **Figura 15.1** se ilustra el proceso de muestreo. En la grafica superior se grafica la onda original y los instantes de muestreo, y en la figura de abajo se indican las muestras.

15.4. Frecuencia de muestreo

Es intuitivamente evidente que la *frecuencia de muestreo debe ser bastante alta*, ya que entonces se logra un *grado de detalle mucho mayor*, lo cual significa que el sonido será reproducido con mayor fidelidad al original. En realidad existe un criterio que debe cumplirse obligatoriamente en todo proceso de muestreo, y es que *la frecuencia de muestreo debe ser mayor que el doble de la máxima frecuencia presente en la señal*. Es decir

$$f_M > 2f_{\text{máx}}$$

Esto es consecuencia de un teorema llamado **Teorema del muestreo**, que dice que una señal muestreada puede recuperarse totalmente sólo si fue muestreada cumpliendo con el criterio anterior. La frecuencia $f_M/2$ se denomina **frecuencia de Nyquist**.

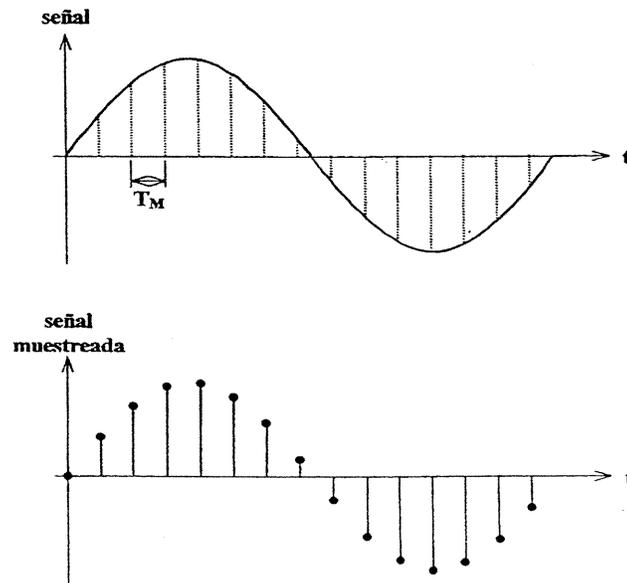


Figura 15.1. Efecto del proceso de muestreo sobre una onda senoidal. La frecuencia de muestreo es en este caso 14,7 veces mayor que la frecuencia de la onda.

Es importante comprender que la frecuencia máxima que aparece en la fórmula anterior no se refiere sólo a la máxima frecuencia *de interés*, sino en realidad a la máxima frecuencia que *efectivamente* aparece en la señal a muestrear, aunque dicha frecuencia provenga de un ruido de alta frecuencia que contamina la señal. En caso de no cumplirse el criterio, al intentar recuperar la señal aparecerán componentes de frecuencia en la banda útil. Para verlo, supongamos que muestreamos con una frecuencia de 40 kHz una señal de audio, y que aparece un ruido (inaudible) de 35 kHz superpuesto a la señal, situación ilustrada en la **Figura 15.2**. Como consecuencia del proceso de muestreo y posterior reconstrucción de la señal, aparece una frecuencia de 5 kHz que no se encontraba presente en la señal original. Esta frecuencia, que sustituye a la original de 35 kHz, se denomina **alias** de aquella. Obsérvese especialmente que la frecuencia original (35 kHz) no producía sensación audible, pero la nueva frecuencia, no sólo es audible sino que está cerca de la región de máxima sensibilidad del oído y por lo tanto se percibirá como un silbido notorio y molesto.

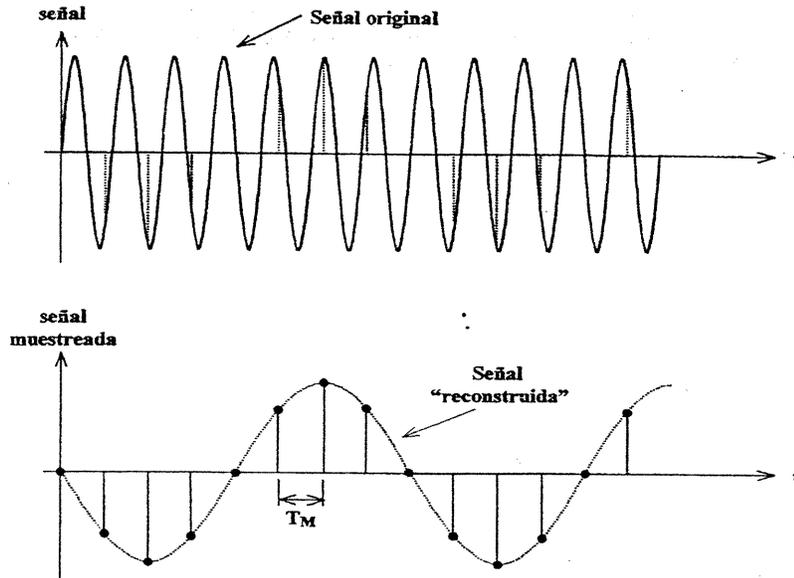


Figura 15.2. Efecto del muestreo con una frecuencia menor que el doble de la máxima frecuencia contenida en la señal. Una señal de 35 kHz se muestrea con una frecuencia de 40 kHz; al intentar reconstruirla, aparece una frecuencia *alias* de 5 kHz.

El ejemplo anterior nos está señalando que si pretendemos que la señal pueda reconstruirse correctamente después del proceso de muestreo es imprescindible eliminar toda frecuencia espuria que caiga más allá del espectro de audio, es decir por encima de 20 kHz. Se utiliza para ello un filtro pasabajos de pendiente muy abrupta en la banda de corte (96 dB/octava ó más), denominado filtro antialias (en inglés *antialiasing filter*).

La elección como frecuencia de muestreo estándar de 44,1 kHz para audio digital obedece precisamente a este problema de las frecuencias alias y la consecuente necesidad de un filtro antialias. Si imponemos una frecuencia máxima de 20 kHz para el audio de alta calidad, el filtro antialias deberá tener su frecuencia de corte en 20 kHz, y como su caída es rápida pero no *infinitamente* rápida, recién después de los 22 kHz se puede considerar que las señales espurias han quedado reducidas a niveles despreciables (Figura 15.3). Por ello se ha adoptado una frecuencia de algo más del doble, es decir 44,1 kHz (el valor *exacto* de 44,1 kHz en lugar de 44 kHz surgió en los comienzos de la grabación digital en cinta de video, para compatibilizar la norma de audio con la de video).

Un inconveniente de los filtros antialias es su gran complejidad y el hecho de que no son del todo inofensivos para la señal dentro de la banda de paso (en este caso la de audio). Aunque el filtro afecte sólo imperceptiblemente la amplitud de la señal en dicha banda, afecta de un modo apreciable la fase, lo cual puede alterar la imagen estéreo.

Además, afecta las señales rápidamente variables agregando pequeños transitorios de frecuencias próximas a la de corte. Por esa razón, otros formatos de audio digital utilizan frecuencias de muestreo más altas, como 48 kHz (el DAT, por ejemplo), que requieren filtros menos complejos. En realidad la frecuencia de 44,1 kHz obedece limitaciones tecnológicas propias de la época en que surgió el compact disc.

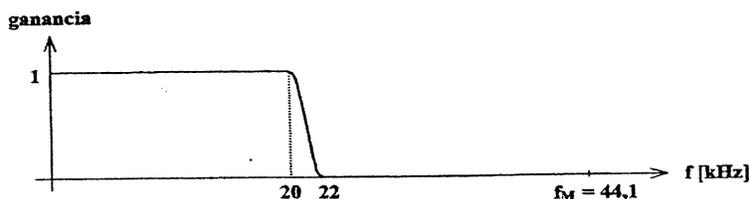


Figura 15.3. Respuesta en frecuencia de un filtro antialias utilizado para audio digital de alta calidad.

15.5. Digitalización

Una vez tomada cada muestra es necesario almacenarla, y para ello debe ser transformada en un número, más específicamente, en un número binario. Esta función la cumple un dispositivo denominado **convertor analógico-digital (A/D)**, que convierte valores de tensión en números binarios.

Consideremos el ejemplo de la Figura 15.4, en el cual utilizamos números binarios de 3 dígitos. Dado que un dígito binario se denomina **bit** (del inglés, **binary digit**), estaremos utilizando, por lo tanto, números de 3 bits. Es fácil ver que existen 8 ($= 2^3$) números de 3 bits: 000, 001, 010, 011, 100, 101, 110 y 111. Para representar los diversos valores de tensión que pueden tomar las muestras, dividimos el rango de variación de la señal en 8 niveles, y aproximamos cada muestra al nivel inmediato inferior.

En la parte central de la Figura 15.4 se comparan las **muestras exactas (puntos vacíos)**, y las **muestras digitalizadas (puntos llenos)**, aproximándolas de la manera comentada. Vemos que el error máximo que se comete es de una división, que a su vez corresponde a 1 bit. La forma de onda reconstruida difiere considerablemente de la original debido a que una resolución de 3 bits es muy pequeña.

En el ejemplo anterior *adoptamos*, en forma arbitraria, una resolución de 3 bits. El resultado fue, como pudo observarse, bastante deficitario, ya que la **onda reconstruida estaba muy distorsionada**. Sería interesante disponer de algún criterio más sistemático para seleccionar la resolución requerida.

El problema es similar al de decidir cuántos dígitos decimales se necesitan para representar una longitud dada con suficiente precisión. Por ejemplo, si quisiéramos representar la longitud de objetos menores de 1 m con una precisión milimétrica, necesitaríamos 3 dígitos decimales, ya que dichos objetos podrían medir entre 0 y 999 mm. Si en cambio quisiéramos una precisión de décimas de milímetro, necesitaríamos 4 dígitos, ya que los objetos podrían medir entre 0 y 9.999 décimas de mm.

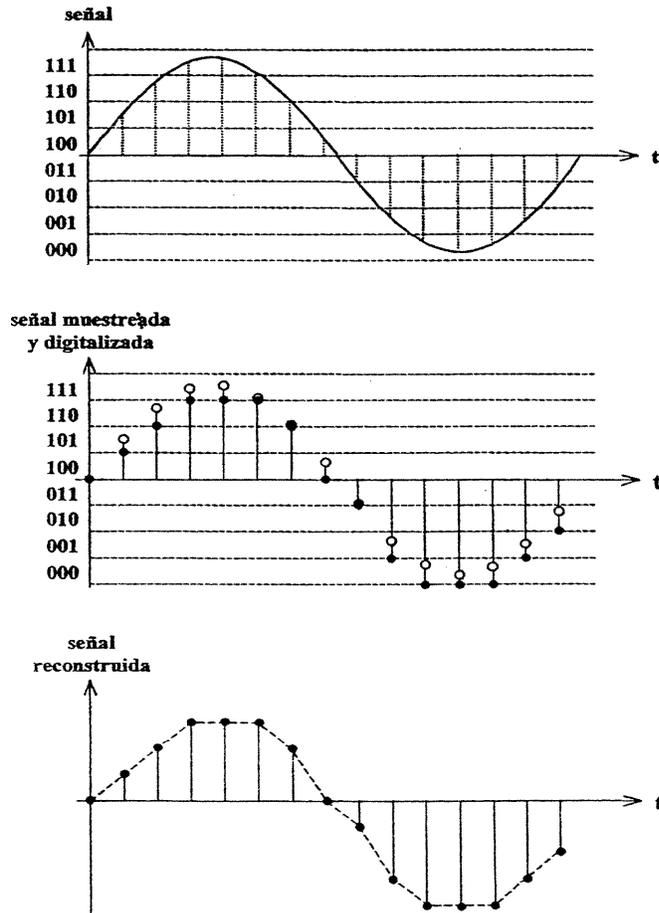


Figura 15.4. Efecto del proceso de muestreo y digitalización sobre una onda senoidal. La resolución es de 3 bits y la frecuencia de muestreo 14,7 veces mayor que la frecuencia de la onda. En la figura central los puntos vacíos representan las muestras exactas y los puntos llenos las muestras digitalizadas. Abajo se muestra la señal reconstruida.

En audio, el criterio para determinar la "precisión" es la relación señal/ruido. Analicemos desde este punto de vista el ejemplo de la Figura 15.4. Dejando de lado el ruido propio que pudiera contener la señal, un efecto colateral de la digitalización es la aparición de un error, el cual puede asimilarse a un ruido. Este ruido se conoce como ruido de digitalización. Bajo esta interpretación, el máximo valor pico a pico de la señal es proporcional a 8, y el máximo valor pico a pico de ruido es proporcional a 1. Entonces, la relación señal a ruido es $8/1 = 8$, que expresada en dB es

$$S/R = 20 \log_{10} \frac{8}{1} = 18 \text{ dB} .$$

Si tenemos en cuenta que en audio de alta fidelidad se manejan hoy en día relaciones señal/ruido mayores de 90 dB, podemos comprender por qué una resolución de 3 bits es por completo insuficiente.

Supongamos ahora que aumentamos la resolución a 4 bits. Dado que ahora hay 16 posibles valores en lugar de 8, la relación señal a ruido en dB será ahora

$$S/R = 20 \log_{10} \frac{16}{1} = 24 \text{ dB} .$$

Vemos que se ha incrementado en 6 dB. Esto puede interpretarse así: si bien la amplitud de la señal no cambió, al duplicar la cantidad de niveles, cada nivel se redujo a la mitad, por lo cual el ruido de digitalización también se redujo a la mitad. Entonces la relación señal a ruido se duplica, y una duplicación equivale a un incremento de 6 dB. Si ahora incrementáramos la resolución nuevamente en 1 bit, llevándola a 5 bits, observaríamos que nuevamente el ruido se reduciría a la mitad, por lo que la relación señal/ruido experimentaría otro incremento de 6 dB.

Podemos obtener una expresión general para la relación señal a ruido. Si adoptamos una resolución de n bits, donde n es cualquier número entero, resulta

$$S/R_{n \text{ bits}} = 6 \cdot n \text{ dB} .$$

Aplicando esta fórmula a la resolución estándar de 16 bits utilizada en los formatos más populares de almacenamiento de sonido digital, resulta una relación señal/ruido de 96 dB. Esta relación señal/ruido es, en condiciones normales, suficiente para crear contrastes dinámicos imponentes. En efecto, tengamos en cuenta que muy rara vez se tiene en la música un nivel sonoro de más de 110 dB (el cual es realmente ensordecedor y para nada recomendable). Si restamos a este valor 96 dB, obtenemos 14 dB, nivel sonoro que probablemente pocas personas tengan el privilegio de haber "escuchado", ya que aún en condiciones de gran silencio durante la noche, en una habitación interior, normalmente es difícil bajar de los 20 dB de nivel de presión sonora.

Es necesario advertir que aunque un sistema funcione con formato de audio digital de 16 bits, su relación señal/ruido *no necesariamente será de 96 dB*. Esto se debe a que en los diversos componentes analógicos que forman parte de todo dispositivo se genera ruido que se agrega al ruido de digitalización. En los equipos de bajo costo, como por ejemplo los discman o los minicomponentes, la electrónica de bajo costo (y baja calidad) utilizada en su fabricación es particularmente ruidosa y la relación señal/ruido real es bastante menor que 96 dB.

15.6. Reconstrucción de la señal

Veamos ahora con más detalle el proceso de reconstrucción de la señal. El procedimiento más simple consiste en obtener, mediante un conversor digital-analógico (D/A), un valor proporcional al número binario de cada muestra y mantenerlo constante hasta

que llegue una nueva muestra, es decir durante un periodo de muestreo T_M . Este procedimiento se denomina **retención simple**. En la Figura 15.5 se muestra este procedimiento para la señal digitalizada de la Figura 15.4.

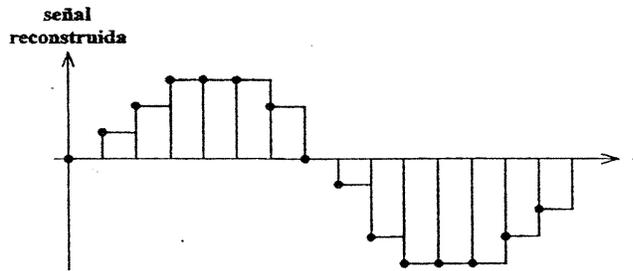


Figura 15.5. Reconstrucción de la señal digitalizada de la Figura 15.4 mediante una retención simple.

Una vez reconstruida la señal, debemos utilizar un **filtro de suavizado**, es decir un **filtro pasabajos** que quite la forma escalonada que resulta de la retención simple. Dicho filtro debe tener características similares a las del filtro antialias que habíamos introducido para la digitalización, es decir, debe ser bastante abrupto para eliminar casi por completo las frecuencias superiores a los 20 kHz, y en cambio permitir pasar las inferiores. Nuevamente, un filtro de ese tipo por un lado es complejo y por el otro produce distorsiones de fase. Para resolver esta situación se ha introducido el concepto de **sobremuestreo (oversampling)**.

El **sobremuestreo** consiste en intercalar, entre las muestras de la señal *realmente* obtenidas o almacenadas, otras "muestras" calculadas por interpolación. Así, por ejemplo, un sobremuestreo por 8 agrega 7 muestras calculadas por cada muestra real. El resultado equivale a una frecuencia de muestreo 8 veces superior a la original. Si $f_M = 44,1$ kHz, entonces la nueva frecuencia de muestreo es de 352,8 kHz, la cual puede eliminarse con filtros pasabajos muchísimo más sencillos y con menos efectos sobre la fase y sobre los transitorios de la señal. El sobremuestreo se utiliza hoy corrientemente en los reproductores de compact disc, lo cual es posible porque la velocidad de la electrónica es mucho mayor que lo que era cuando recién había surgido esta nueva tecnología.

15.7. Memorias electrónicas

Introduzcamos ahora algunos conceptos relativos a las memorias electrónicas. Existen básicamente dos tipos de memorias: las memorias **RAM (random access memory)**, y las memorias **ROM (read only memory)**. Las memorias RAM son memorias de *lectura y escritura*, es decir en las cuales es posible escribir (guardar) números binarios, y posteriormente leerlos. Son memorias *volátiles*, es decir que una vez interrumpido el suministro de energía eléctrica (al desconectar el equipo) la información se pierde. Las memorias ROM son memorias de *lectura solamente*, en las cuales no es posible escribir.

En ellas hay datos fijos de fábrica, que *no se pierden al desenergizar el equipo*. Se utilizan mucho para guardar formas de onda muestreadas, como en la mayoría de los sintetizadores actuales.

En ambos tipos de memoria los datos se almacenan en posiciones sucesivas, cada una de ellas identificada con un número binario denominado **dirección**. En la Figura 15.6 se ilustra con un ejemplo la estructura de una memoria con datos de 4 bits y direcciones de 3 bits. Así, en la dirección 0, es decir 000, se encuentra almacenado el número binario 0110 (en decimal, 6), y en la dirección 3, es decir 011, se encuentra almacenado

000	001	010	011	100	101	110	111
0110	0001	1110	1101	0010	1000	0001	0010

Figura 15.6. Estructura de almacenamiento de una memoria digital. La memoria se organiza en posiciones identificadas con direcciones.

el número binario 1101 (en decimal, 13). En la Figura 15.7 se muestran las entradas y salidas principales de una memoria RAM. Cuando nos referimos a una entrada de datos de 4 bits, por ejemplo, en realidad está formada por 4 líneas o cables, y por eso tanto la entrada como la salida y la dirección se han dibujado como flechas gruesas. Las entradas

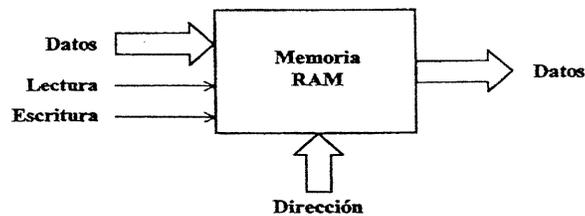


Figura 15.7. Entradas y salidas de una memoria RAM. Las flechas delgadas representan señales binarias de 1 bit de control, y las flechas gruesas son datos digitales de varios bits.

de lectura y escritura son señales binarias de 1 bit, cuya función es de control. Cuando la *entrada de escritura* toma el valor 1, el dato digital X presente en ese instante en la entrada de datos se escribirá en la dirección D indicada por la entrada de direcciones. Para leer dicho valor, bastará volver a aplicar el valor D en la entrada de direcciones y dar valor 1 a la *entrada de lectura*. El valor guardado aparecerá en la salida de datos.

Desde el punto de vista del conexionado, la única diferencia entre una memoria RAM y una ROM es que esta última no tiene entrada de escritura. En el caso de la RAM, después de escribir un dato en una dirección éste quedará almacenado allí hasta que se escriba un nuevo dato en el mismo lugar o hasta que se interrumpa el suministro de energía eléctrica al circuito.

Aunque por razones conceptuales nos hemos referido a la entrada y la salida de datos como si fueran entidades físicamente diferentes, en realidad son el mismo conjunto de líneas cuya función en cada instante depende de cuál de las entradas de control valga 1. Si ambas entradas son 0, dichas líneas se desconectan de modo de no cargar al resto del circuito.

15.8. Dither

Cuando se digitalizan señales de muy bajo nivel (cercano a la resolución del conversor) el ruido de digitalización se convierte en una distorsión, cuyo efecto es más perjudicial que el de un ruido aleatorio. Por ejemplo, si se digitaliza una señal senoidal de 100 Hz y amplitud apenas menor que un escalón (Figura 15.8), se obtiene una señal

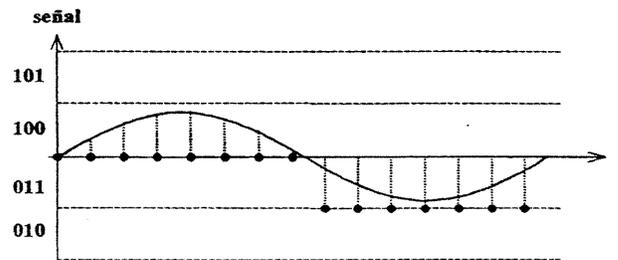


Figura 15.8. Distorsión creada al muestrear señales de bajo nivel.

que al reconstruirse será muy similar a una onda cuadrada, y por lo tanto contendrá armónicos de 300 Hz, 500 Hz, 700 Hz, etc. Si en lugar de una onda senoidal se aplicaran dos o más, aparecería, similarmente, una distorsión por intermodulación sumamente indeseable.

Una manera de evitar estos inconvenientes es aplicar una pequeña cantidad de ruido aleatorio antes del muestreo y posterior digitalización. Este ruido, cuyo valor eficaz es, normalmente, menor que un escalón, se denomina **dither**. Si bien el efecto es empeorar ligeramente la relación señal a ruido, desde el punto de vista auditivo transforma esta distorsión en un ruido aleatorio, que es mucho más aceptable, sobre todo en niveles tan bajos.

También se acostumbra a aplicar **dither** en los procesos de *recuantización*, es decir cuando, por ejemplo, se desea reducir la resolución de una señal grabada en 20 bits a 16 bits a fin de volcarla a un formato comercial como el compact disc. Si solamente se truncaran los datos de 20 bits eliminando los 4 bits menos significativos, se producirían inconvenientes similares al descrito. En ese caso, el ruido se genera digitalmente y se agrega antes de proceder al truncado.