



Captura de señales de audio y preparación para su compresión

Apellidos, nombre	Oliver Gil, José Salvador (joliver@disca.upv.es)
Departamento	Informática de Sistemas y Computadores
Centro	Escola Tècnica Superior d'Enginyeria Informàtica

1 Resumen de las ideas clave

En este artículo se describe el proceso completo de captura del sonido para poder identificar los elementos clave que posteriormente se emplean a la hora de comprimir una señal de audio. En concreto se definen los parámetros de digitalización fundamentales, como son la frecuencia de muestreo y el número de bits por muestra.

2 Introducción

Previamente al proceso de compresión de una señal de audio, debemos realizar una captura de las ondas acústicas que queremos procesar digitalmente y comprimir. Esta captura se efectúa con un micrófono y dará lugar a una señal analógica, que previamente al proceso de compresión, se digitalizará según los parámetros de muestreo deseados. En concreto, como veremos más adelante, cuanto mayor sea el rango de frecuencias que queramos preservar del audio original, mayor será la frecuencia de muestreo a emplear. Además, veremos que el número de bits que utilicemos para representar cada muestra determinará también el nivel de ruido introducido por la digitalización de esa muestra (que definiremos como error de cuantización).

Otro aspecto que introducimos en este artículo es que no todas las frecuencias de un audio son percibidas de la misma forma por el oyente, de forma que podemos emplear estos efectos psicoacústicos para dar más importancia a la información que mejor se perciba cuando queramos comprimir una señal de audio.

3 Objetivos

Una vez que el alumno lea con detenimiento este documento, será capaz de:

- Identificar los elementos clave del proceso completo de captura de una señal de audio que posteriormente se emplean a la hora de comprimir una imagen.
- Aplicar el teorema de Nyquist-Shannon para saber a qué frecuencia se debe muestrear una señal de audio.

4 Desarrollo

4.1 Definición física

Desde el punto de vista físico, el sonido lo forman ondas acústicas que se transmiten como variaciones de presión propagadas por el aire. El audio es la técnica relacionada con el procesamiento de sonido, incluyendo su captura y transmisión. El resultado de la captura será una señal de audio.

La intensidad del sonido, o dicho de manera más precisa, la presión sonora, se mide físicamente como pascuales o dinas/cm². Pero, ¿te suena encontrar la potencia de un sonido expresada con estas unidades? Seguramente no, es mucho más normal encontrarla como decibelios. Veamos por qué.

La potencia del sonido (o formalmente el nivel de presión sonora) se expresa como la relación entre la "potencia de presión" del sonido que queremos medir y la del mínimo audible, que nos sirve como referencia.

Si queremos medir el nivel de presión de un sonido A, con esta sencilla fórmula:

$$A/B$$

veremos cuántas veces es mayor que la señal de referencia B. Para medir potencias aplicamos cuadrado a estas intensidades.

$$(A/B)^2$$

Además, la sensibilidad del oído humano sigue una escala logarítmica (es más sensible en rangos de baja intensidad sonora), por lo que la forma final de medir el nivel de presión de un sonido queda de esta manera.

$$10\log(A/B)^2 \text{ dB}$$

Para reflejar que a una relación se le ha aplicado una escala logarítmica, al resultado se le añade la palabra *decibelios* (dB).

Algunos valores típicos son: una conversación tranquila registra niveles de presión de en torno a 45 dB, con cierta algarabía supera los 60, mientras que un avión despegando registra 120 dB y una *mascletà* típica de las fallas de Valencia podría llegar a los 130 dB, alcanzando ya el umbral del dolor.

4.2 Psicoacústica

En general, los humanos percibimos sonidos entre los 20 hz y los 20 khz, aunque este límite superior tiende a disminuir con la edad. Además, no todas las frecuencias las percibimos con la misma intensidad. Aquellas que se ubican en el rango de entre 1 y 5 Khz son las que mejor oímos. Este rango tiene bastante sentido si pensamos que hasta los 3,5 Khz se encuentran las componentes de frecuencia más significativas de la voz humana, con lo que nuestro proceso evolutivo nos habría llevado a escuchar mejor el rango en el que se sitúa la voz humana.

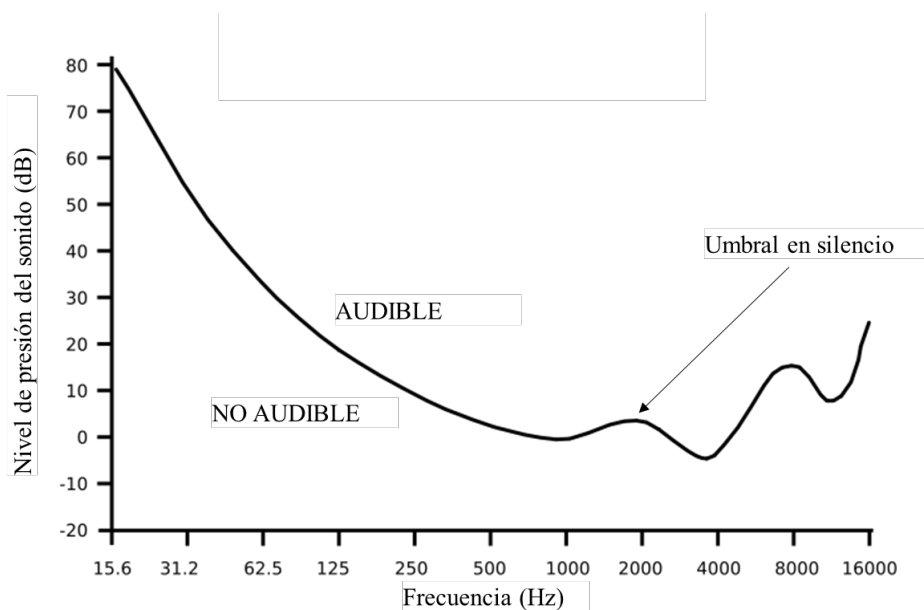


Imagen 1. Rango audible por los humanos

La Imagen 1 nos muestra el umbral de presión sonora a partir del cual empezaríamos a escuchar un tono puro (como un pitido) según su frecuencia en circunstancias óptimas. Fíjate como un pitido de 125 hz con un nivel de presión sonora de 10 dB no llegaríamos a escucharlo, mientras que si ese pitido fuera de 4 khz lo escucharíamos perfectamente.

A este tipo de estudios sobre la percepción del sonido se les llama psicoacústica, y en compresión de audio son muy importantes, ya que la idea es no codificar aquella información que los humanos no percibimos. De esta forma, en el ejemplo anterior, el tono de 125 hz y 10 dB no se codificaría.

4.3 Captura de una señal de audio

La captura del sonido la realiza un micrófono, que es un transductor electroacústico, que convierte ondas acústicas en una señal de audio cuyo nivel de voltaje en cada instante es proporcional a la intensidad de la onda acústica que lo golpea. La señal de audio resultante será por tanto una señal unidimensional.

Si capturamos el audio en una escena con un único micrófono obtendremos una señal de audio monoaural (o simplemente mono). Para integrar mejor al usuario en la escena podremos utilizar dos micrófonos convenientemente espaciados obteniendo así una señal estereofónica o estéreo. Si ubicamos más micrófonos tendremos un sonido envolvente multicanal. En estos casos, para la reproducción usaremos un altavoz especial adicional destinado a emitir sonidos graves denominado subwoofer. Observa que estas señales siguen siendo 1D, pero multicanal.

4.4 Digitalización de una señal de audio

La señal obtenida de un micrófono es analógica, por lo que para su codificación y uso en redes digitales tendremos que realizar un proceso de digitalización mediante un conversor analógico digital. Éste muestrea la señal analógica y la cuantiza a valores digitales.

4.4.1 Frecuencia de muestreo

El proceso de muestreo consiste en tomar valores (o muestras) cada cierto tiempo, tal y como podemos ver en la Imagen 2. Las muestras se toman en un intervalo constante denominado periodo de muestreo (T_m). La frecuencia de muestreo (f_s) (o tasa de muestreo) es la inversa del periodo de muestreo y, en hercios, indica cuantas muestras se capturan cada segundo. Por ejemplo, si tomamos muestras cada 0,125 milisegundos, $T_m=0.000125$ s, estaremos capturando 8000 muestras por segundo, ya que $f_s = 1/0.000125 = 8000$ Hz, con lo que la frecuencia de muestreo usada será de 8 Khz.

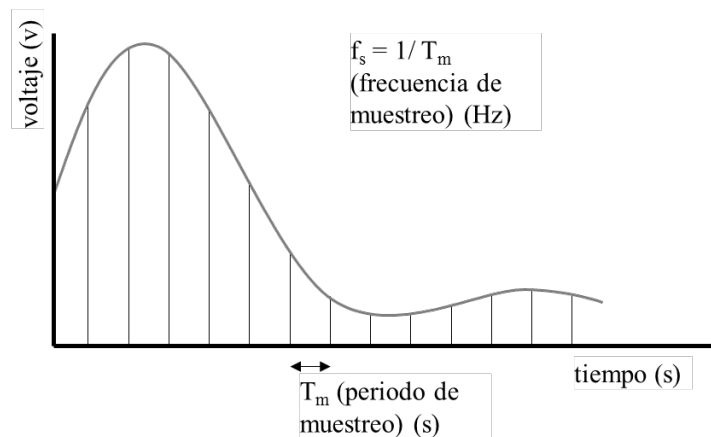


Imagen 2. Digitalización de una señal de audio

Obviamente si tomamos más muestras por segundo, la señal muestreada se acercará más a la original. El teorema de muestreo de Nyquist-Shannon nos dice que si una señal tiene una frecuencia máxima f , podremos reconstruir esta señal original sin pérdidas si la muestreamos con una frecuencia de al menos $2f$. Otra lectura teórica de este teorema es que mediante una frecuencia de muestreo f podemos capturar hasta la frecuencia $f/2$ de una señal (si previamente le aplicamos un filtro paso-bajo para limitar su frecuencia máxima a $f/2$).

Por ejemplo, si quisiéramos muestrear una señal de audio para capturar hasta la frecuencia máxima que podemos percibir los humanos, 20 KHz, ¿cuál debería ser la frecuencia de muestreo teórica a usar? Efectivamente, el doble, 40 KHz.

Sin embargo, para capturar solamente hasta los 20 KHz se debe aplicar un filtro paso bajo. Como no existe el filtro paso-bajo ideal, tendremos que aumentar ligeramente la frecuencia de muestreo a emplear. Un valor típico es incrementarla en un 10%, dándonos como resultado una frecuencia de muestreo de 44 KHz.

4.4.2 Cuantización (bits por muestra)

La señal obtenida en el proceso de muestreo tiene un número discreto de valores en el tiempo, pero cada muestra sigue teniendo un valor continuo. Por lo tanto, el siguiente paso será representar cada muestra con un número de bits concreto. Este proceso, en el que rangos continuos se representan con un valor discreto se denomina cuantización.

Si en nuestro ejemplo de la Imagen 2 empleamos dos bits para cada muestra, podemos definir 2^2 niveles de cuantización, o sea, cuatro niveles de cuantización, tal y como se muestra en la Imagen 3.

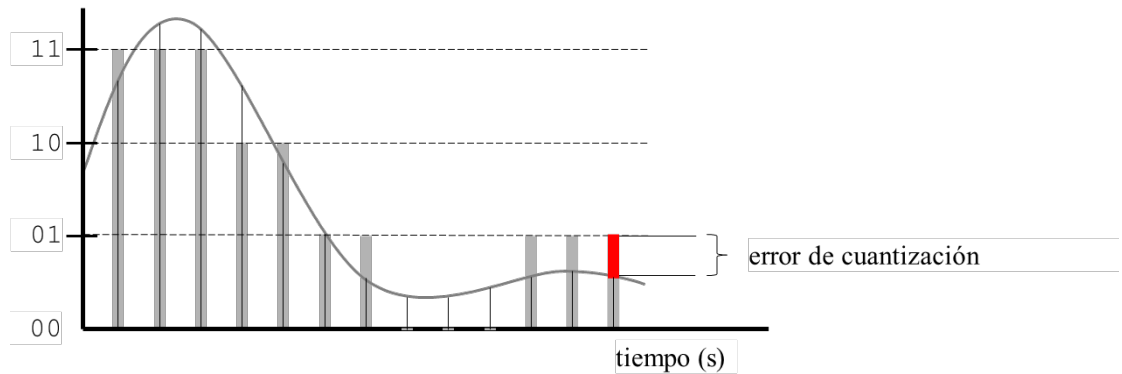


Imagen 3. Cuantización de una señal de audio empleando 2 bits
 (y por tanto 2^2 niveles de cuantización)

Codificaremos cada una de las muestras con el nivel de cuantización más próximo a su valor. En este proceso se introduce un error llamado error de cuantización (o ruido de cuantización), que es la diferencia entre el valor real de la muestra y el valor tal y como se ha digitalizado (la parte en rojo en la Imagen 3).

¿Cómo crees que se puede reducir el error de cuantización? Efectivamente, necesitamos más niveles de cuantización y para ello deberemos emplear más bits por muestra. En general, si empleamos n bits por muestra obtendremos 2^n niveles de cuantización. En la Imagen 4 vemos nuestra señal digitalizada con 3 bits/muestra y por tanto 8 niveles de cuantización. Observa que hemos reducido el error de cuantización respecto al que aparecía en la Imagen 3.

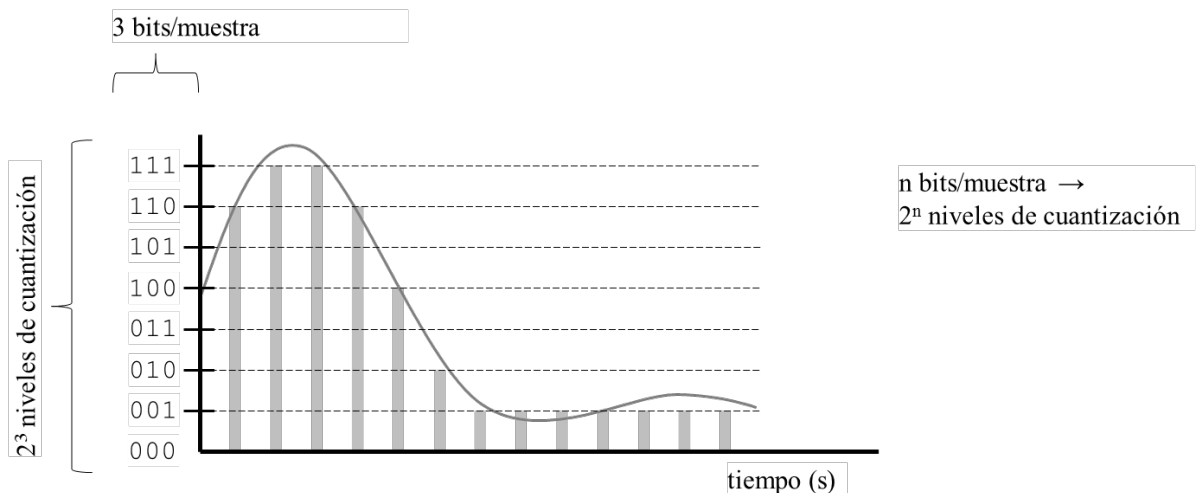


Imagen 4. Cuantización de una señal de audio empleando 3 bits
 (y por tanto 2^3 niveles de cuantización)

De hecho, en la Imagen 5 podemos ver en verde la señal tal y como se reconstruiría para su reproducción usando tres bits. No es muy diferente de la original.

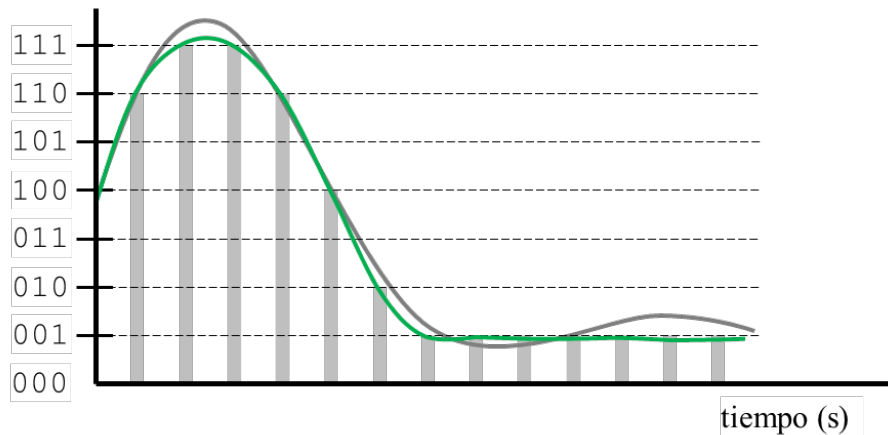


Imagen 5. Señal de audio reconstruida a partir de su codificación empleando 3 bits por muestra

Sin embargo, en la Imagen 6 podemos comparar la señal resultante de emplear dos bits por muestra, en azul, con la señal de la Imagen 5 en la que usamos 3 bits por muestra, en verde. Puedes ver que el error de cuantización introducido es bastante mayor al usar sólo 2 bits.

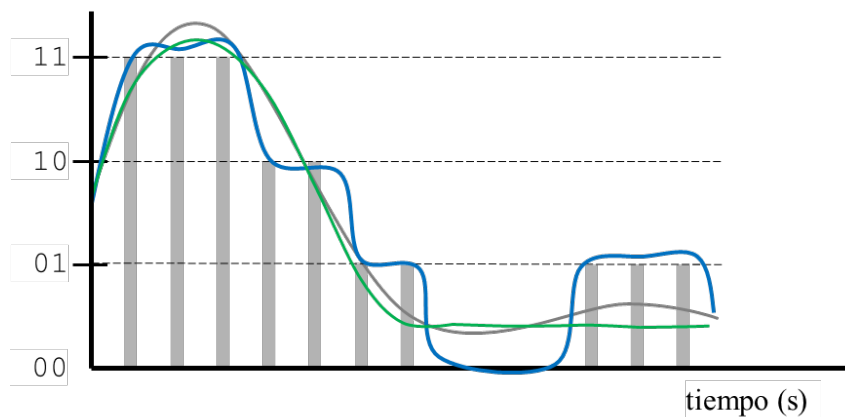


Imagen 6. Señal de audio reconstruida a partir de su codificación empleando 2 bits por muestra (en azul) y comparativa con la señal obtenida usando 3 bits por muestra (en verde)

5 Cierre

Hemos visto el proceso mediante el que un micrófono recoge las ondas acústicas y las transforma en niveles analógicos de voltaje, de forma que un posterior proceso de muestreo y cuantización da lugar a las muestras del audio digitalizado.

6 Bibliografía

Fluckiger, F: "Understanding networked multimedia: applications and technology", Prentice Hall, 1995.



Tanenbaum, A: "Redes de computadoras", Pearson Educación, cuarta edición, 2003, pág. 674-679.