



Compresión de audio en calidad VOZ

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 identifican los principales parámetros que definen una señal de audio de calidad voz, para posteriormente pasar a describir las principales técnicas de compresión empleadas con este tipo de señal.

Para ello, empezaremos viendo los parámetros que definen típicamente este tipo de audio: principalmente la frecuencia de muestreo y el número de bits por muestra empleados en su digitalización. Posteriormente veremos su representación PCM lineal y logarítmica, la codificación PCM diferencial y la DPCM adaptativa. Finalizaremos viendo técnicas específicas para voz llamadas vo-coding.

2 Introducción

Decimos que una señal de audio es de calidad voz si ha sido capturada según los parámetros clásicos de la telefonía básica.

En la antigua telefonía analógica (*Plain Old Telephone Service*), el sonido transmitido se restringía a las componentes de frecuencia situadas en la banda entre los 300 Hz y los 3300 Hz,

¿por qué crees que se empleaba esta banda de frecuencia?

El motivo era que estos servicios estaban inicialmente diseñados para transmisión de voz en una conversación, y la mayor parte de frecuencias significativas en este tipo de sonidos se encuentran por debajo de los 3,5 KHz.

¿Qué parámetros de digitalización, frecuencia de muestreo y bits/muestra, crees que se deberán emplear en una señal de audio si queremos tener calidad voz?

Según el teorema de muestreo de Shannon y Nyquist, para capturar las frecuencias de hasta 3,5 KHz de un sonido deberemos usar una frecuencia de muestreo de algo más del doble. En calidad voz muestreamos el audio a 8 KHz.

Por otro lado, el número de bits por muestra determinará el error de cuantización. Con 8 bits tendremos 256 niveles de cuantización, con lo que el error será tolerable. Sin embargo, esta no es la única opción, como veremos más adelante.

Esta señal, capturada con estos parámetros y sin compresión, necesitará una tasa de bits de 8×8000 bps, o sea 64 Kbps.

3 Objetivos

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

- Identificar los principales parámetros que definen una señal de audio de calidad voz.
- Describir las principales técnicas de compresión empleadas con este tipo de señal.

4 Desarrollo

4.1 Codificación PCM lineal

Al proceso de digitalización del audio se le llama codificación PCM. En el ejemplo de la Imagen 1, vemos el resultado de digitalizar una señal analógica usando codificación PCM con 3 bits por muestra.

La secuencia de muestras obtenidas como resultado de la digitalización en este ejemplo es:

110,111,111, 110, 100,010, 001, 001,001, 001,
001,001,001, ...

Además, cuando, como en este ejemplo, todos los niveles de cuantización son equidistantes (es decir, están igualmente espaciados), decimos que la codificación es PCM lineal.

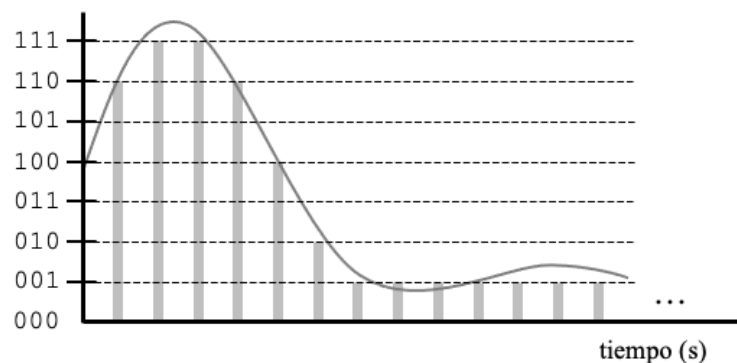


Imagen 1. Señal de audio digitalizada con 3 bits por muestra empleando PCM lineal

4.2 Codificación PCM logarítmica

Dentro del rango auditivo, nuestro oído es más sensible a sonidos débiles que a sonidos más fuertes, por tanto, el error de cuantización introducido en muestras de pequeña intensidad se puede llegar a notar más. Por este motivo, en ocasiones se emplea una codificación PCM no lineal en la que no todas las intensidades se cuantizan igual.

Un ejemplo típico es la cuantización logarítmica, como la que se muestra en la Imagen 2, en la que la distancia entre niveles de cuantización sigue una escala logarítmica. Fíjate como en ella, el error de cuantización es menor en los sonidos más débiles que en los fuertes. Un ejemplo de este tipo de codificación es la norma ITU G.711 empleada habitualmente en telefonía IP.

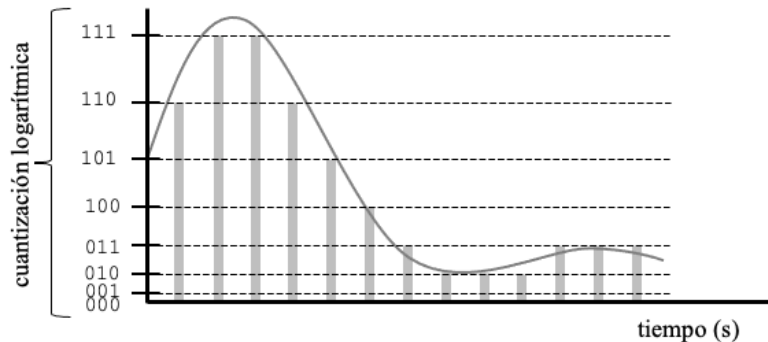


Imagen 2. Señal de audio digitalizada con 3 bits por muestra empleando PCM con cuantización logarítmica

4.3 Codificación PCM diferencial (DPCM)

Al digitalizar una señal de audio con PCM lineal, observa cómo, en muchos casos, las muestras de audio vecinas son muy parecidas. Aprovechando esta similitud, podremos codificar la diferencia entre las muestras consecutivas en lugar de las propias muestras, de manera que, si estas diferencias son pequeñas, podremos emplear menos bits por muestras. A esta técnica se la conoce como codificación PCM diferencial o DPCM.

Por ejemplo, podemos codificar diferencialmente el ejemplo de señal PCM digitalizada con 3 bits por muestra de la Imagen 1. Podemos plantearnos hacer una codificación diferencial usando 2 bits, de forma que en complemento a 2, podamos representar un rango de valores de -2 a 1.

La primera muestra se codificará empleando PCM y para el resto de muestras, codificamos las diferencias. El resultado sería el siguiente:

Si usamos 2 bits (en complemento a 2)

$$\{10, 11, 00, 01\} \equiv \{-2, -1, 0, 1\}$$

110

$$111-110= 01$$

$$111-111= 00$$

$$110-111= 11 (-1)$$

$$100-110= 10 (-2)$$

$$010-100= 10 (-2)$$

$$001-010= 11 (-1)$$

$$001-001= 00$$

$$001-001= 00$$

$$001-001= 00$$

$$001-001= 00$$

Fíjate como en este caso, todas las diferencias se pueden representar con 2 bits en complemento a 2, con lo que el resultado de la codificación sería equivalente a la de PCM con 3 bits. Pero esto no es siempre así. A veces, las diferencias a codificar no se pueden representar con el número de bits empleados, y en las oscilaciones tendríamos una convergencia más lenta a la señal original.

Veamos un ejemplo de esto. Supón que en lugar de utilizar complemento a 2, empleásemos la representación de signo magnitud. Como esta es menos eficiente,

tan sólo podremos representar diferencias entre -1 y +1, ya que tendríamos las siguientes equivalencias, $\{11, 10, 00, 01\} \equiv \{-1, -0, 0, 1\}$

Así, observa en la Imagen 3 que el codificador tarda más tiempo en reaccionar ante las oscilaciones más pronunciadas en la señal original. De hecho, con esta codificación diferencial, sí que se aprecia un mayor error de cuantización entre nuestra señal PCM diferencial, que se decodificaría como muestra la curva azul, y la PCM de 3 bits, cuya decodificación, que mostramos en verde, se aproxima bastante más a la señal original (la gris).

En general, se estima que una codificación PCM de 8 bits la podemos representar bastante fielmente empleando PCM diferencial con 6 bits por muestra.

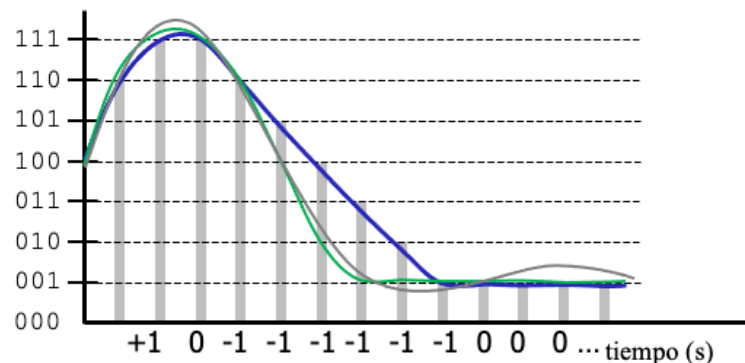


Imagen 3. Señal de audio digitalizada con empleando DPCM con dos bits y representación signo magnitud (en la que las diferencias solo pueden ser -1, 0, +1)

4.4 Codificación DPCM adaptativa (ADPCM)

Como hemos visto, el problema con la codificación diferencial se encuentra cuando queremos representar señales con elevada oscilación, es decir, altas frecuencias. Por ejemplo, en la señal de la Imagen 4, en la zona de altas frecuencias pasamos de una muestra con un valor de 7, a otra con un valor de 0, para pasar a otra con un valor nuevamente de 7. Para representar estas oscilaciones con codificación diferencial necesitaríamos un rango de valores de -7 a 7, es decir, 4 bits, uno más que con la propia codificación PCM, que requería tan solo 3 bits.

Para resolver este problema, la codificación DPCM adaptativa (o ADPCM) modifica dinámicamente la ubicación de los niveles de cuantización que se aplican sobre las diferencias, de forma que cuando se están codificando valores parecidos (o sea en las bajas frecuencias), se emplean niveles de cuantización más próximos, mientras que si las diferencias son mayores, estos se espacian más.

Fíjate que no estamos modificando el número de niveles de cuantización sino la distancia entre ellos, para cubrir así un rango mayor cuando las diferencias entre muestras son más grandes. Esto tiene como efecto lateral el aumento del error de cuantización en las frecuencias más elevadas, pero afortunadamente este error se percibe menos en las altas que en las bajas frecuencias.

Existen varias versiones de ADPCM, como IMA ADPCM (también llamado DVI), o algunos codificadores de la ITU, como el G.721, G.722 y variaciones del G.721 como G.723, G.726 y G.727. Algunos de estos codificadores realizan una codificación por subbandas y una predicción más elaborada de la muestra a codificar.

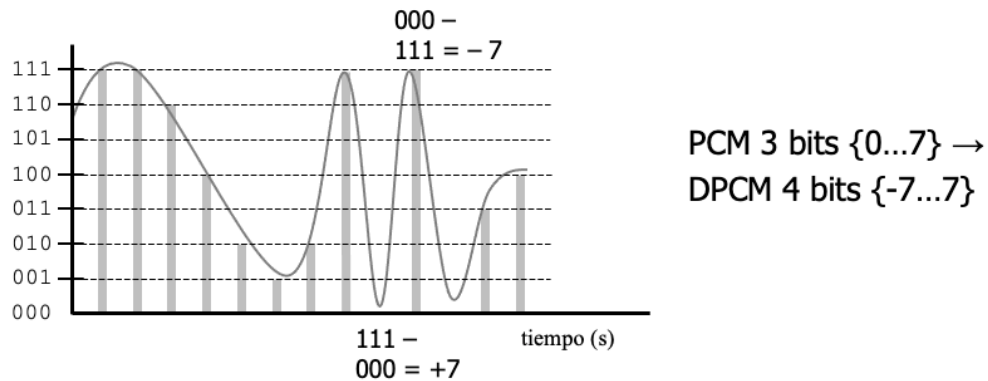


Imagen 4. Señal de audio digitalizada

4.5 Técnicas vo-coding

Las técnicas que acabamos de ver, aunque se ajustan bien a las señales de voz con calidad similar a la telefonía, son de propósito general, es decir, pueden codificar tanto la voz humana como cualquier otro tipo de sonido.

Las técnicas conocidas como codificación de voz o vo-coding definen un modelo del aparato fonador humano e intentan extraer parámetros de este modelo a partir de la señal de audio a codificar, con lo que son técnicas orientadas exclusivamente a codificar voz.

LPC es un vocoder introducido en los años 60. En él se divide la fuente de audio en pequeños segmentos de audio de en torno a 20 ms, y para cada uno de ellos, se buscan valores de los principales parámetros de este modelo. De esta forma, el decodificador recoge estos parámetros para sintetizar la voz artificialmente.

LPC obtiene unas tasas de bit muy reducidas, de 2,4 Kbps, pero la voz suena muy sintética. Una mejora de este codificador es CELP, en el que además de los parámetros de LPC, se transmite el error cometido al sintetizar la voz, empleando para ello un codebook que contiene un conjunto de errores típicos. Con este método se consiguen mejoras en la calidad del audio, con tasas de sólo 4,8 Kbps.

Existen muchas otras variantes de estas técnicas. Entre ellas la empleada en la telefonía móvil digital GSM, y diversos codificadores de la ITU.

5 Cierre

Hemos visto las características de la señal de audio en calidad voz, y los fundamentos de algunos de los principales codificadores de audio para este tipo de señal.

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.