

Modelos psicoacústicos en compresión de audio con transformada Wavelet

Recibido: julio de 2012
Arbitrado: agosto de 2012

Marcelo Herrera M.*, Felipe Mesa C.** , Rosas J., Nelson F.***

Resumen

El objetivo del presente trabajo es describir los principios de compresión de señales y principios psicoacústicos para proveer las bases de técnicas digitales usadas en estos campos. La compresión de señales de audio reduce la información de datos para almacenar y/o transmitir una señal de audio. La reducción de memoria es realizada, reduciendo información redundante e irrelevante contenida en la señal.

Palabras clave

Compresión de audio, enmascaramiento, psicoacústica, codificadores, señales de audio.

Abstract

The objective of this work is to describe the principles of signal compression and psychoacoustic principles to provide the basis of the digital techniques used in these fields. The compression of audio signals reduces the amount of data information for storing and/or broadcasting an audio signal. This memory reduction is performed by removing redundant and irrelevant information registered in the signal.

Keywords

Audio compression, masking, psychoacoustics, coders, audio signals.

I. Introducción

La compresión de señales de audio consiste en reducir el tamaño de información de datos usando principios psicoacústicos. Los datos redundantes e irrelevantes de la señal son descargados. Existen dos tipos de codificadores de audio. Los primeros, llamados sin pérdida reducen únicamente la información redundante mientras que los codificadores con pérdida o «perceptuales», reducen tanto la información redundante, como la irrelevante. Mientras el proceso es llevado a cabo, la señal es reconstruida mediante un decodificador, eliminando la información innecesaria, tratando de mantener la calidad de la señal.

El presente texto describe los principios de compresión de señales y describe principios detallados sobre psicoacústica, los cuales dan la base para la compresión, y el desarrollo de las técnicas digitales utilizadas en estos campos.

* Ingeniero Electrónico, Magister en Radioelectrónica y Doctor en Acústica de la Universidad Técnica de Praga. Profesor Titular de la Facultad de Ingeniería de la Universidad de San Buenaventura, Bogotá. Líder del Semillero de Investigación de «Sistemas de Compresión Perceptual de Audio». E-mail: mherrera@usbog.edu.co.

** Estudiante de pregrado de Ingeniería de Sonido de la Universidad de San Buenaventura. Integrante del Semillero de Investigación de «Sistemas de Compresión Perceptual de Audio». E-mail: fmesa@academia.usbbog.edu.co

*** Ingeniero Electrónico, Universidad Nacional de Colombia, sede Bogotá. Magister en Ingeniería - Telecomunicaciones Universidad Nacional de Colombia, sede Bogotá. Profesor Catedrático Facultad de Ingeniería, Universidad de La Salle, sede Bogotá. E-mail: nfrosasj@unal.edu.co

II. Psicoacústica

Las relaciones entre los estímulos físicos y las respuestas psicológicas que estos provocan, son objeto de estudio de la psicoacústica. Las relaciones entre las propiedades físicas del sonido y la interpretación cerebral de ellas por la corteza cerebral son el núcleo de esta disciplina.

Los objetivos generales de la psicoacústica se resumen en determinar:

1. La característica de respuesta del sistema auditivo.
2. El umbral de sensaciones auditivas.
3. La diferencia entre umbrales perceptuales.
4. La interpretación de los estímulos.
5. La variación en el tiempo de la sensación del estímulo.

III. Enmascaramiento auditivo

El enmascaramiento es un fenómeno psicoacústico en el cual el ser humano deja de percibir ciertos estímulos ante la presencia de otros que los opacan en amplitud. Esta información redundante es imperceptible para el sistema auditivo humano y por consiguiente puede ser eliminada.

Existen tres tipos de enmascaramiento auditivo definidos:

3.1 Enmascaramiento temporal

Sucede cuando dos estímulos sonoros llegan al oído de forma cercana en el tiempo, de forma que el estímulo enmascarante hará que el enmascarado sea inaudible, en esta situación el tono más intenso tiende a enmascarar al débil.

Según el instante del tiempo en que se produce el estímulo enmascarante sobre el enmascarado se puede distinguir entre preenmascaramiento y posenmascaramiento.

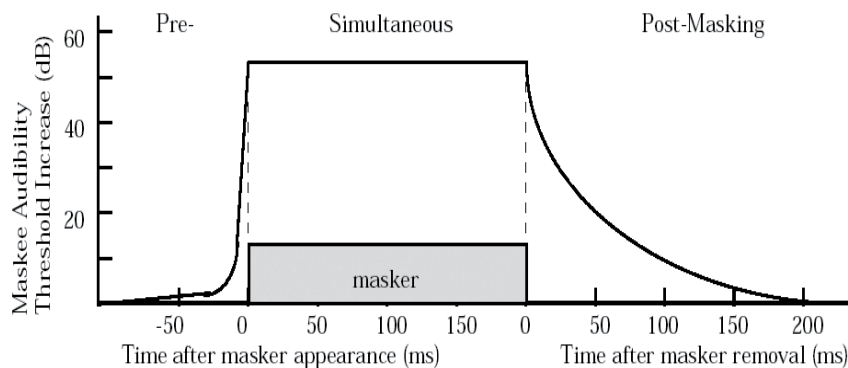


Figura 1. Tipos de enmascaramiento auditivo.

3.2 Posenmascaramiento

Se produce cuando el tono de mayor amplitud aparece primero en el tiempo seguido por un tono de menor amplitud, por lo cual solo es percibido el primer estímulo ya que después de esto el oído necesita un tiempo de adaptación aclarando que este fenómeno ocurre en lapsos de 30 a 60 milisegundos.

3.3 Preenmascaramiento

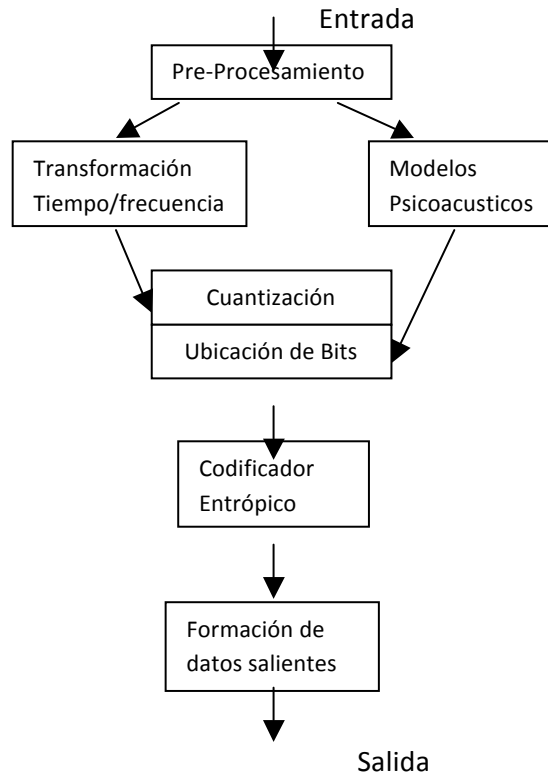
Este se encuentra en el momento en el que se presenta un estímulo suave y posteriormente uno intenso; este último de igual forma logra enmascarar al primero, siempre y cuando la diferencia de tiempo entre estos dos esté definida entre 5 y 10 milisegundos. Esta anticipación puede ser explicada, y se debe a que el cerebro humano procesa la información que

llega, por ráfagas y al ser la diferencia de tiempo tan corta suele ser el tono de mayor amplitud el que es procesado con más énfasis.

IV. Codificadores de audio

La función esencial de los codificadores de audio es reducir la cantidad de bits que ocupa el

archivo, además de comprimir señales o ficheros de audio con un flujo de datos, donde el objetivo es ocupar el menor espacio posible sin quitarle calidad a la señal y descomprimiendo los datos para lograr manejarlos en un formato más adecuado. Se puede clasificar a los codificadores según el tipo de filtros que utilizan y la información que es descartada a partir de la señal de audio original, en 2 tipos.



4.1 Codificadores sin pérdida

Los codificadores sin pérdida procesan una señal de audio en tramos de longitudes variables, de las cuales se obtienen 'residuos' también llamados redundancias, los cuales se dan en tramos de señal consecutivos, posterior a eso la señal es comprimida.

4.2 Codificadores con pérdida

Son llamados también perceptuales, estos aprovechan las limitaciones del sistema auditivo humano para realizar su análisis, usan algoritmos con pérdida de datos para lograr un menor tamaño de archivo, en este se presentan formatos de audio

comunes, que son usados tanto en ordenadores como en reproductores portátiles.

V. Codificación perceptual de señales de audio

Todos los sistemas de codificación de audio utilizan algoritmos muy similares en su estructura, debido a que se basan en los mismos principios psicoacústicos. En general, la reducción de memoria se realiza mediante la eliminación de la información imperceptible, esta tarea se realiza recuantizando la señal y analizando cada uno de los componentes espectrales.

VI. Modelos psicoacústicos

El modelo psicoacústico es un elemento esencial en la codificación de señales de audio, aplicando principios como el enmascaramiento y el umbral auditivo para acomodar la mejor forma de aplicar la compresión.

6.1 Representación frecuencial

Es una manera de mostrar una señal de audio en función del umbral de frecuencias mediante coeficientes, en vez de usarse un filtro polifase ya que para realizar la codificación se requiere una resolución más grande en la representación para poder calcular los umbrales de enmascaramiento presentes en la señal.

6.2 Separación de componentes tonales y no tonales

Se debe hacer una separación en componentes tonales y no tonales debido a que el enmascaramiento se manifiesta de diferente forma en ambos casos.

- Para identificar los componentes tonales, el modelo identifica los picos en el espectro de frecuencias en bandas críticas, para después juntar el resto y marcarlo como un componente no tonal.
- El segundo caso crea índices de tonalidad en función a la frecuencia y después lo usa para hacer un análisis entre ruido enmascarado por tono y viceversa.

VII. La escala 'Bark'

Esta es una escala de medida psicoacústica nombrada con base en Heinrich Barkhausen que fue el primero en proponer esta forma de medida subjetiva. Esta escala define rangos desde 1 a 24, lo cual comprende las 24 bandas críticas de audición en Hz con sus respectivos centros.

Es posible hallar la medida en Bark dependiendo de una determinada frecuencia con la siguiente expresión, la cual es posible implementar en forma de función en Matlab, donde B es el valor y F la frecuencia.

$$B = 13 \arctan(0.00076f) + 3.5 \arctan\left(\left(\frac{f}{7533}\right)^2\right) \quad (1)$$

VIII. La transformada de Fourier

La transformada de Fourier es usada para analizar una señal en el dominio de la frecuencia, para obtener información que no es evidente en el dominio temporal. Es más fácil saber sobre qué ancho de banda se concentra la energía de una señal analizándola en el dominio de la frecuencia.

La transformada de Fourier se emplea con señales aperiódicas a diferencia de la serie de Fourier. Las condiciones para poder obtener la transformada de Fourier son, que la señal sea absolutamente integrable.

La transformada de Fourier se define así:

$$X(\omega) = \int_{-\infty}^{\infty} x(t) \cdot e^{-j\omega t} dt \quad (2)$$

Donde: ω , describe la oscilación de la onda ($\omega = 2\pi f$).

8.1 Transformada Discreta (DFT)

Al trabajar en compresión en un ordenador es más rentable trabajar con muestras tomadas lo más juntas posibles por lo que la transformada discreta es más útil ya que permite evaluar señales no necesariamente periódicas, es decir una frecuencia determinada en un tiempo determinado.

La serie de Fourier discreta es simplemente una modificación de la serie de Fourier tradicional, pero sustituyendo las integrales por sumatorios de las muestras.

IX. Análisis unidimensional de señales mediante la interfaz de Matlab

Matlab posee una interfaz de usuario destinada al análisis de señales a la cual se accede mediante el comando 'wavemenu'.

Para el análisis de una señal monofónica mediante la interfaz de matlab es necesario ingresar a la interfaz mediante el 'wavemenu', en este hay un apartado que abre el cuadro de análisis para señales monofónicas, el cual consta de los siguientes procesos.

1. Cargar la señal
2. Identificar las variables (vector, tamaño del vector).

3. Hacer una descomposición sencilla de la señal (Wavedec).
4. Construir las aproximaciones y detalles de la señal a partir de los coeficientes (Up-coef).
5. Descomponer la señal en niveles más profundos.

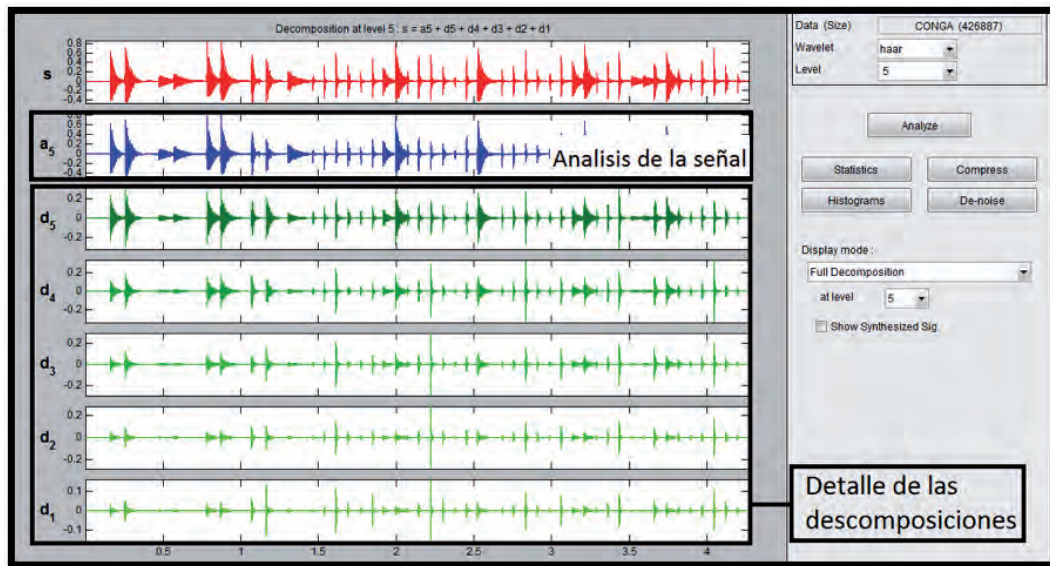


Fig. 3. Interfaz en MATLAB para descomposición Wavelet de una señal monofónica

X. Conclusiones

Mediante la compresión de audio es posible desarrollar sistemas nuevos para compresión y envío de señales más livianas, pero igualmente calificadas en tiempo real, logrando mejorar ampliamente la gama de las comunicaciones ya que sería posible desarrollar un medio de alta definición y rendimiento, tanto para contextos como la radio o televisión e inclusive dispositivos móviles.

XI. Bibliografía

- [1] Zwicker, H. Fastl: «Psychoacoustics», Springer Verlag, 1990.
- [2] Bosi, M. Goldberg: «Introduction to audio coding and standards» Boston, 2003.
- [3] J.G. Proakis and D.G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, Prentice-Hall, NJ, Third Edition, 1996.
- [4] S.G. Mallat. «A Wavelet Tour of Signal Processing». 2nd Edition. Academic Press, 1999. ISBN 0-12-466606-X
- [5] Ken C. Pohlmann «Principles of Digital Audio», McGraw-Hill, Fourth edition, 2000.
- [6] P. Srinivasan and L. H. Jamieson. «High Quality Audio Compression Using an Adaptive Wavelet Packet Decomposition and Psychoacoustic Modeling», *IEEE Transactions on Signal Processing*, Vol 46, N.º 4, April 1998.
- [7] T. von Petersdorff, «How to use the Matlab wavelet toolbox» (web page), Scientific Computing I (MAPL/CMSC 660), University of Maryland, College Park, MD, USA, 1999. URL: www.glue.umd.edu/~tvp/660/matlabwav.html.
- [8] R. A. Devore and B. J. Lucier, «Wavelets», University of South Carolina, Columbia, SC and Purdue University, West Lafayette, IN, 50 pp., 1992.
- [9] ISO/IEC 11172-3, «Information technology - Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mbit/s» - (Part 3), 1992.
- [10] P. Srinivasan, L. H. Jamieson, «High-quality compression using an adaptive wavelet packet decomposition and psychoacoustic modeling», *IEEE Trans. on Signal Processing*, Vol. 46, N.º 4, pp. 1085-1093, April 1998.