

Sistema de compresión de audio con transformadas integrales de soporte compacto

Marcelo Herrera Martínez*

Recibido: noviembre de 2013

Aceptado: mayo de 2014

72 ▶ 85

RESUMEN

El principio de incertidumbre de Gabor, determina que no es posible tener óptima resolución en tiempo-frecuencia, al realizar un procesamiento digital de la señal con herramientas convencionales como la FFT. Este procesado es esencial en diversos sistemas de audio, como reverbadores digitales, filtros y también en sistemas de compresión. Los sistemas de compresión perceptual de audio, minimizan el volumen informático de la señal, hasta una tasa de 1:10. Esto proporciona la posibilidad de ahorrar espacio en los sistemas de almacenamiento y economizar anchos de banda de canales de comunicación, para transmisiones de Radio y Video. El presente proyecto implementa un compresor perceptual, con una técnica no-convencional. Esta es basada en la Transformada Wavelet- Daubechies Tipo IV, en cascada. Los coeficientes de los filtros paso-alto y paso-bajo son obtenidos en MATLAB, y la implementación completa es realizada en Pure Data. Esto permite su posterior implementación en un sistema embebido. La señal resultante queda reducida en una tasa de 1:2, y conserva el formato original (wav).

Palabras clave: Compresor perceptual, MP3, Transformadas Integrales, Wavelets, Modelo Psico acústico, Cuantificación de señales

* Universidad de San Buenaventura, Bogotá.

ABSTRACT

The uncertainty principle by Gabor, determines that it is not possible to have optimum resolution in time-frequency when a digital signal processing with conventional tools such as FFT is performed. This processing is essential in different audio systems such as digital reverberators, filters and in compression systems too. The audio perceptual compression systems minimize the computer volume of the signal at a rate of 1:10. This provides the ability to save space in storage systems and to economize bandwidth of communication channels for broadcast Radio and Video. This project implements a perceptual compressor with a non-conventional technique. This is based on The Daubechies D4 Wavelet Transform in cascade. The coefficients of the high-pass and low-pass filters are obtained in MATLAB, and complete implementation is done in Pure Data. This allows subsequent implementation in an embedded system. The resulting signal is reduced at the rate of 1: 2 and retains the original format (wav).

Keywords: Perceptual Compressor, MP3, Integral transforms, Wavelates, Psychoacoustic model, Quantization.

Introducción

Los sistemas de compresión perceptual de audio son esquemas de DSP (Procesamiento Digital de Señales), que combinan elementos claves de teoría de la información y la psicoacústica.

Están basados en la reducción del volumen informático de la señal, específicamente en fenómenos como irrelevancia y redundancia. La irrelevancia se define como el descarte de componentes de la señal debido a fenómenos psicoacústicos como enmascaramiento (temporal o frecuencial), descritos ampliamente en (Fastl, Zwicker, 2006). Este fenómeno tiene su explicación fisiológica, debido a las capacidades limitadas de las ciliadas dentro del órgano interno auditivo (Kranz, 1952).

La eliminación de componentes redundantes es lograda con procedimientos de teoría de la información (David, 2008), como la codificación de Huffman, la cual reduce el volumen de datos, tomando en cuenta la probabilidad de eventos de símbolos, asignando palabras de datos mayores a eventos con menor probabilidad, y lo opuesto.

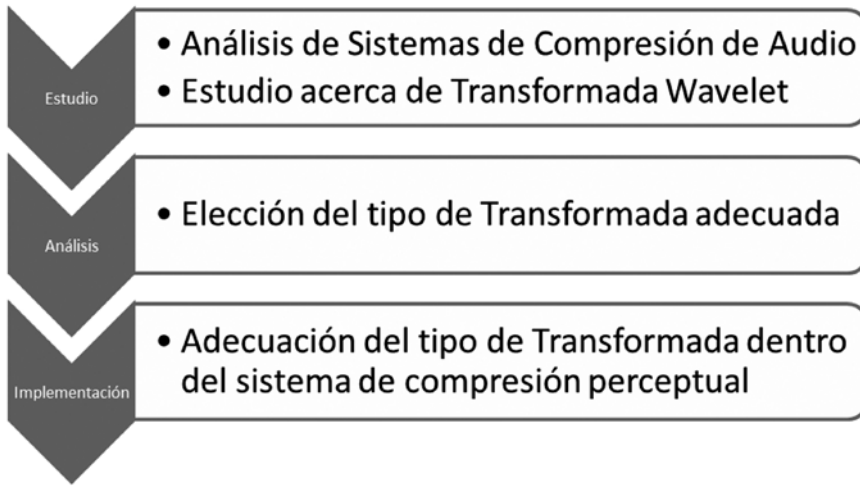
Por lo anterior, el grupo de investigación **de Acústica Aplicada** ha desarrollado una técnica, la cual procesa la señal de audio, con una Transformada Integral, de soporte compacto, como es la Transformada de Wavelet-Daubechies Tipo IV con dos niveles de descomposición, en cascada. Esta Transfor-

mada posibilita solucionar el conocido “Principio de Incertidumbre de Heisenberg”, dentro de la dualidad tiempo-frecuencia, en el procesamiento digital de audio (Proakis, Manolakis, 2007).

Esto posibilita la localización exacta de eventos en Tiempo-Frecuencia, con un alto grado de precisión. La ventaja de la combinación de DWT dentro del modelo psicoacústico es la posibilidad de comprimir hasta la mitad de datos del volumen original, manteniendo la misma estructura de datos, el formato wav. Este hecho posibilita un almacenamiento/transmisión eficiente con equipamiento que no requiere conversión de formato.

Metodología de diseño

El presente proyecto está motivado por la necesidad de crear un dispositivo eficiente y de bajo costo para la compresión de señales de audio, para su posterior transmisión en Radio HD. Los sistemas de compresión actuales basan su procesamiento en la denominada Transformada de Fourier, la cual realiza un mapeo entre los valores psicoacústicos y su respectiva descripción en frecuencia. Sin embargo esta transformada no se traduce en un ahorro significativo del contenido de la señal. Este es el objetivo principal en el desarrollo de un compresor con otras transformadas, como la transformada Wavelet, y la transformada Karhunen Loeve.

FIGURA 1. Procesos involucrados en el desarrollo del codificador perceptual

El proyecto inicia con el modelo matemático del compresor de audio con los nuevos tipos de transformada discreta a utilizar, así como el modelo de los distintos tipos de cuantificación y codificación entrópica. Posteriormente se procederá a simulaciones de los compresores modelados. Finalmente se procederá a la implementación del códec en un procesador digital de señales, donde se implementará además el modelo multicanal, la aplicación para propósitos de Radio HD.

El proyecto tiene en su desarrollo tres fases: una de diseño y simulación, la segunda de programación, construc-

ción y pruebas y la tercera de construcción del prototipo final.

Basado en los principios de bancos de filtros Wavelet fue diseñado un nuevo codificador perceptual de audio. El diagrama de bloques del codificador puede ser visto en la Figura 3.

El sistema tiene diseño modular, de manera que cada bloque puede ser mejorado sin la necesidad de modificar todos los bloques. En la primera fase de la realización del codificador los bloques son implementados de manera sencilla y se espera que mejoren posteriormente después de la evaluación del sistema.

FIGURA 2. Fases de desarrollo del codificador perceptual

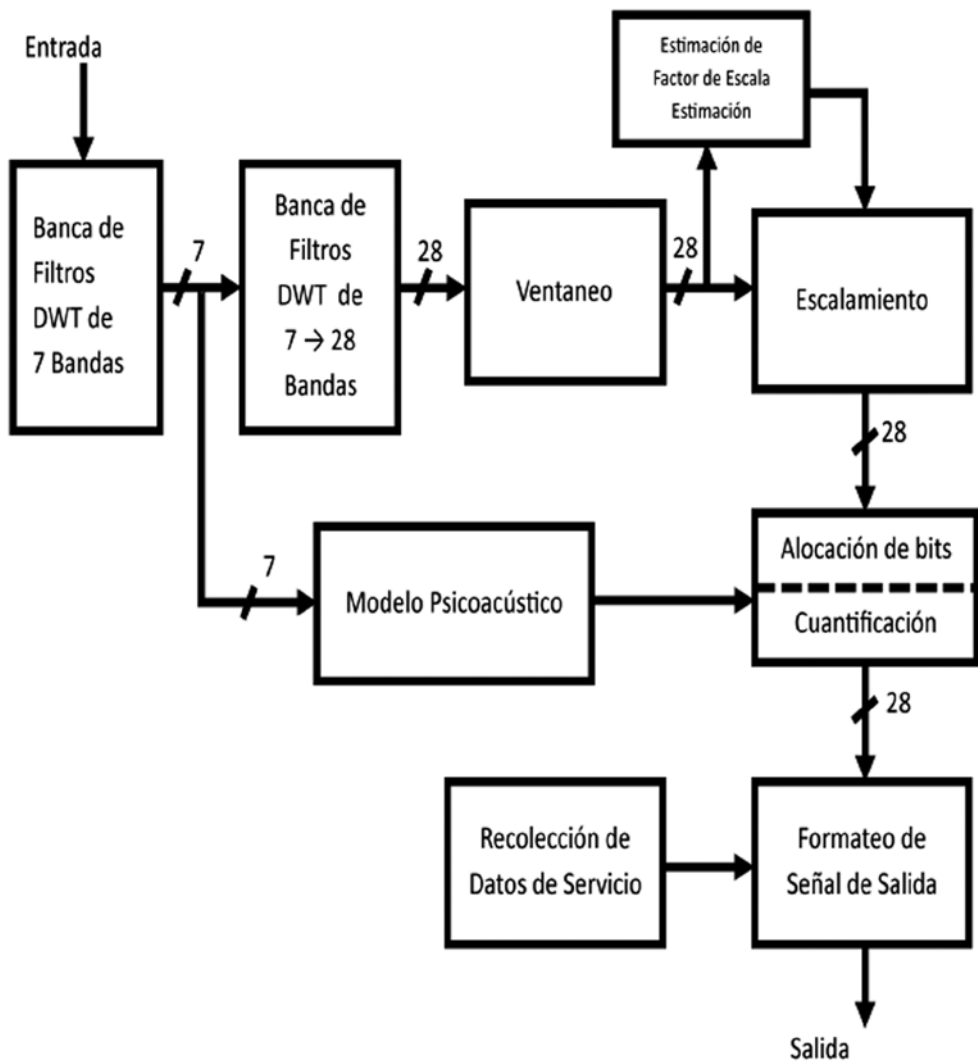
Modelo del compresor perceptual

El núcleo del codificador perceptual es una Banca de Filtros Wavelet. En comparación con la banca de filtros PQMF, usada en ISO/IEC MPEG-1 (1993), la ban-

ca de filtros Wavelet no tiene divisiones en sub-banda igualmente espaciadas, lo cual está más acorde con la teoría de bandas críticas (Zwicker, Fastl, 1990).

Inspirado por Jhang-Liang L., et. al. (1998), un árbol de descomposición Wavelet fue diseñado. Las bandas de frecuencia de las salidas de las sub-

FIGURA 3. Sistema de compresión de audio basado en la transformada Wavelet



bandas del segundo filtro de Wavelet corresponden con las bandas críticas, como se muestra en la figura 1.

Modelo Psicoacústico

El modelo psicoacústico analiza la señal de entrada, desde un punto de vista perceptual y calcula un umbral de enmascaramiento global. El umbral de enmascaramiento global, varía con el tiempo, por lo tanto es calculado para cada ventana de manera separada. El modelo Psicoacústico en el presente proyecto, es una banca de filtros de Wavelet, basado en ISO/IEC MPEG-1. El análisis del Modelo Psicoacústico está descrito en Bosi, Goldberg, (2003) y Mallat, (1998).

La señal entrante al modelo Psicoacústico tiene de siete sub-bandas críticamente muestreadas con diferentes frecuencias de muestreo. Estas siete señales son analizadas por el modelo psicoacústico. La salida del modelo es un umbral global dependiente de la frecuencia, en otros términos, la relación Señal-Enmascaramiento (SMR).

Ventaneo y FFT

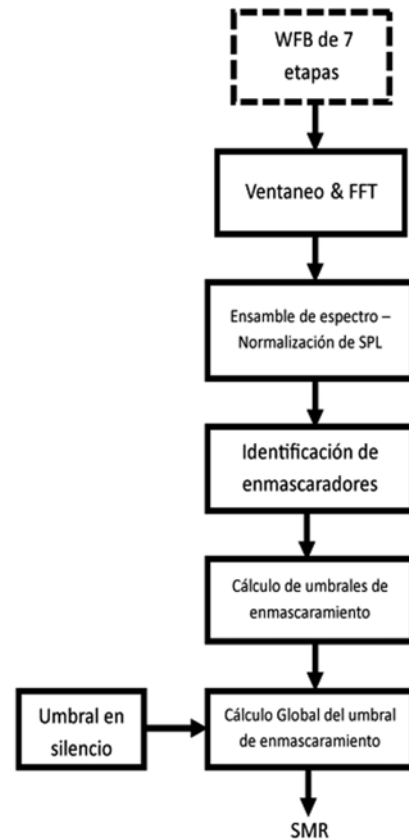
La Ventana de Hamming es usada en el modelo, variando entre ventanas de 32 o 64 muestras.

La señal ventaneada está descrita por, donde s_n representa la señal en la banda $-n$, W

$$(1) \quad s_{wn}(i,j) = s_n \left(i \cdot \frac{N}{2} + j \right) \cdot W(j) \quad j = 0 \dots N$$

Representa la función de ventana usada (La Ventana de Hamming), i es la posición temporal y N es la longitud de la ventana.

FIGURA 2. Modelo Psicoacústico de la compresión de Audio.



La señal ventaneada es transformada por la Transformada de Fourier, como puede observarse,

$$(2) \quad S_n(i) = \text{FFT}(s_{wn}(i))$$

La salida de este bloque es una representación parcial de las señales ventaneadas de las siete sub-bandas bajo testeo.

Conformación del espectro y Normalización de SPL

La representación del espectro global de la señal bajo test está conformada por las contribuciones parciales del espectro calculadas en cada banda de frecuencia.

El vector de los componentes espectrales S , es, posteriormente transformado a niveles de presión sonora por medio de la ecuación 3 (Bosi, Goldberg, 2003)

(3)

$$S_{SPL}(i) = 90.302 + 20 \cdot \log_{10}(|S(i)|)$$

Identificación de enmascaradores tonales y enmascaradores por ruido

Las curvas de enmascaramiento tonal y las curvas de enmascaramiento por ruido tienen diferentes formas (ISO/IEC 11172-3, 1993). Por lo tanto, existe la necesidad de tratarlas de manera separada. De acuerdo con ISO/IEC 11172-3 (1993), los componentes espectrales que exceden a los vecinos con una diferencia de dB, como mínimo, son tonales. Cuando se encuentran los componentes tonales, es necesario encontrar el máxi-

mo local primero, y después compararlo con los componentes del vecindario, tal como se describe en la ecuación 4.

(4)

$$S_{SPL}(i) - S_{SPL}(i \pm \Delta_i) \geq 7$$

donde representa el vecindario examinado. Cada componente espectral que satisface la ecuación 4 es tonal. es un parámetro del modelo, y es usualmente.

De acuerdo con el análisis del modelo psicoacústico del estándar ISO/IEC MPEG-1, el nivel de presión sonora de los enmascaradores tonales es calculado por la ecuación 5.

(5)

$$X_{TM}(i) = 10 \cdot \log_{10} \sum_{j=-1}^1 10^{\frac{S_{SPL}(i+j)}{10}} \quad [\text{dB}]$$

El nivel de presión sonora de los enmascaradores por ruido es calculado, de acuerdo con la ecuación 6, como la suma de los niveles de presión sonora de cada componente espectral en la banda crítica correspondiente.

(6)

$$X_{NM}(b) = 10 \cdot \log_{10} \sum_i 10^{\frac{S_{SPL}(i+j)}{10}} \quad [\text{dB}]$$

Donde b representa la banda crítica, i , el índice de componente espectral, localizado en la correspondiente banda crítica.

Cálculo de los umbrales de enmascaramiento

Cuando el enmascaramiento por ruido y tonal están identificados, el umbral de enmascaramiento es determinado. Como está definido en el Modelo Psicoacústico en el estándar de audio ISO/IEC MPEG 1, la curva de enmascaramiento puede ser calculada por la ecuación 7.

(7)

$$M_{TM}(i+j) = X_{TM}(i) + MF(i,j) - 0.275z(j) - .025[\text{dB}]$$

Donde es el SPL del tono de enmascaramiento, $z(j)$ es la posición de la curva de enmascaramiento en el eje de barks, $MF(i,j)$ es una función de enmascaramiento definida por la ecuación 8 (de ISO/IEC 11172-3, 1993) y la constante 6.025 representa la distancia en SPL entre el enmascarador y la cima de la curva de enmascaramiento.

(8)

$$MF(i,j) = \begin{cases} 17\Delta_z - 0.4X_{TM}(i) + 11 & \Delta_z \in <-3, -1> \\ (0.4X_{TM}(i) + 6)\Delta_z & \Delta_z \in <-1, 0> \\ -17\Delta_z & \Delta_z \in <0, 1> \\ -(\Delta_z - 1) \cdot (17 - 0.15X_{TM}(i)) - 17 & \Delta_z \in <1, 3> \end{cases}$$

Donde representa la distancia en barks desde el enmascarador.

Fuera del intervalo <-3,3>, tenemos $MF = (-,)$.

Las curvas de enmascaramiento para los enmascaradores por ruido están definidos por el Modelo Psicoacústico de ISO/IEC MPEG-1, definidos de manera similar como los enmascaradores tonales (ecuación 9).

(9)

$$M_{NM}(i+j) = X_{NM} + MF(i,j) - 0.175z(j) - 2.025 [\text{dB}]$$

Donde es el SPL del enmascarador de ruido, $z(j)$ es la posición de la curva de enmascaramiento en el eje de barks, $MF(i,j)$ es una función de enmascaramiento definido por la ecuación 8 con cambiado a y la constante 2.025 representa la distancia SPL entre el enmascarador y la cima de la curva de enmascaramiento. Fuera del intervalo <-3,3>, $MF = (-)$.

Cálculo del umbral de enmascaramiento global

Cuando las curvas de enmascaramiento tonales y de ruido son determinadas, el umbral de enmascaramiento global es calculado. De acuerdo con el Modelo Psicoacústico de ISO/IEC MPEG-1, el umbral de enmascaramiento global puede ser calculado con la Eq. 10. Esta expresión toma incluso en cuenta el umbral en silencio.

(10)

$$G_{th}(i) = 10 \cdot \log_{10} \sum_j (10^{0.1 \cdot M_{tm}(i,j)} + 10^{0.1 \cdot M_{nm}(i+j)} + 10^{0.1 \cdot T(i)}) [\text{dB}]$$

donde representa el umbral en silencio a una frecuencia particular.

Cálculo de SMR

La relación señal-enmascaramiento (SMR) es calculada con la ecuación 11 como una substracción entre el Nivel de Presión Sonora y el umbral de enmascaramiento de un componente espectral dado,

(11)

$$SMR(i) = S_{SPL}(i) - G_{th}(i) \text{ [dB]}$$

Un valor SMR positivo indica que la señal está por encima del umbral de enmascaramiento, mientras que un valor SMR negativa indica que la señal está por debajo de ese nivel y puede ser excluida de futuro procesamiento.

Implementación en un sistema embebido

El esquema general de un codificador perceptual de audio está descrito en la figura 1.

El modelo fue realizado en PureData, y posteriormente fue implementado en un Sistema Embebido (ver Figura 3).

El sistema consiste en un bloque de adquisición de la señal, un módulo para la Transformada Wavelet, un bloque de cuantificación y un bloque donde la señal es reproducida (ver figura 4 y 5). En este módulo, la descomposición en anchos-de-banda variables es calcu-

lado (Wavelets), realizando la descomposición de la señal en aproximaciones y detalles, caracterizando de manera eficiente las transientes de la señal y sus discontinuidades.

FIGURA 3. Prototipo del ensamble Electrónico



Una descomposición en dos (2) niveles es realizada con la Transformada Wavelet Daubechies Tipo IV. Esta descomposición es realizada con una cascada de filtros FIR (Finite Impulse Response), cuyos coeficientes son determinados con la ayuda del programa de MATLAB, con el comando *wfilters*.

Módulo de Filtros de Wavelet en conjunto con el Modelo Psicoacústico

El modelo Psicoacústico, proporciona el número de bits con los cuales se puede re-cuantificar la señal. Esto se calcula con la relación señal-enmascaramiento (SMR). La descomposición con la Transformada de Wavelet, consiste, en cada nivel, de una descomposición en filtro paso-bajo y paso-alto, como ha sido mencionado.

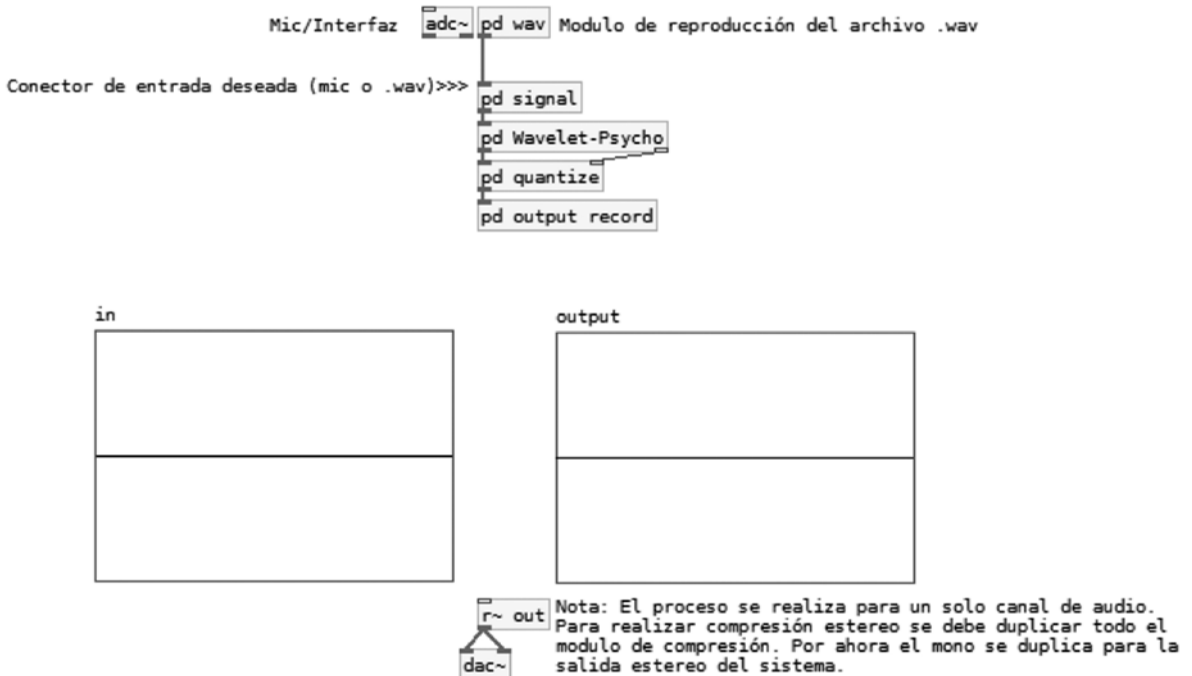
Después de estos filtros, una decimación es realizada por un factor de dos (2). Ésta es realizada, debido a que si la señal original contiene muestras, después del filtrado a través de la sección paso-alta y paso-baja, la señal a la salida de estos filtros, tendrá el mismo número de muestras que la señal original. De esta manera, puede observarse que el número de muestras se duplica. Por esta razón, el bloque de decimación (por un factor de dos (2)), es implementado a la salida de cada uno de los filtros de descomposición. Esta descomposición es conocida como Descomposición Diádica o Análisis de Multirresolución.

Este proceso de filtrado da como resultado los detalles de la señal. Si un nivel más alto de descomposición es realizado, un conjunto de detalles más exacto puede ser obtenido.

Modelo Psicoacústico

El modulo analiza los campos de 1024-muestras, renueva el espectro y después de un análisis por sub-bandas, origina el número de bits necesarios para recuantificación. El campo de la señal es analizado, a través de cada sub-banda, y los umbrales tonales y de ruido son calculados, para colocar dentro de estos modelos, el ruido de cuantificación.

FIGURA 4. Codificador Audio Perceptual en Beagle-Board XM



Análisis de resultados

Comparación entre el espectro de la señal original y la comprimida

Una comparación entre el espectro de la señal original y la comprimida es presentada en la figura 6.

FIGURA 5. Modelo Psicoacústico con Banco de Filtros Wavelet

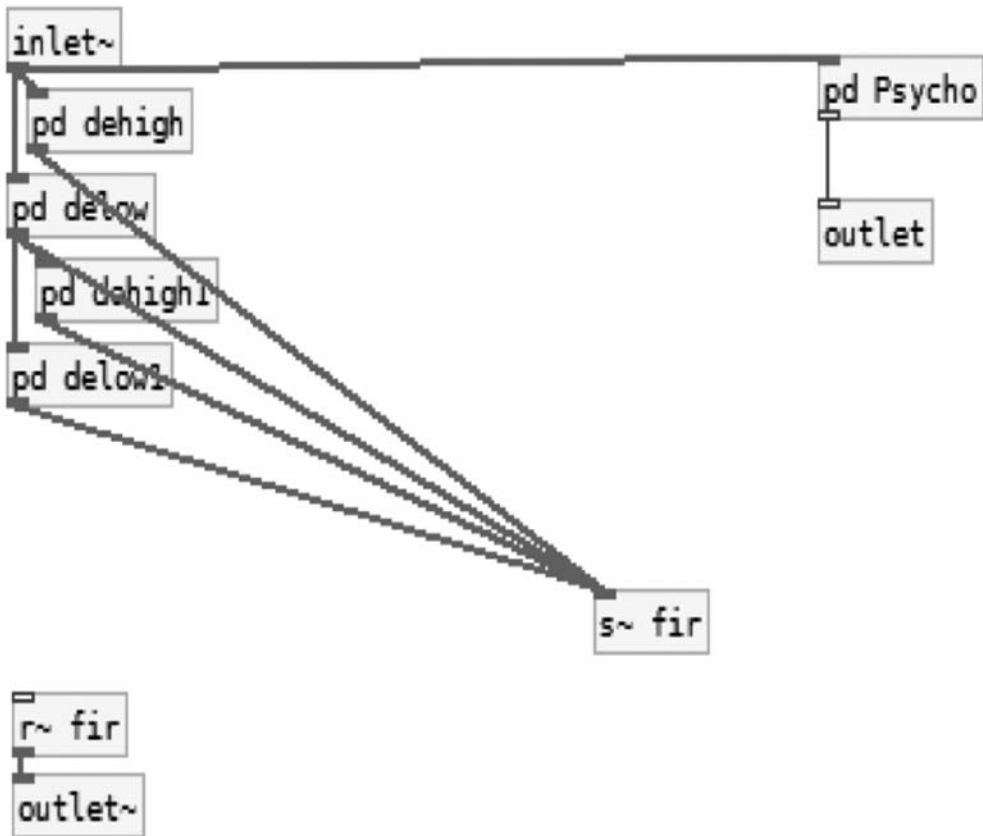
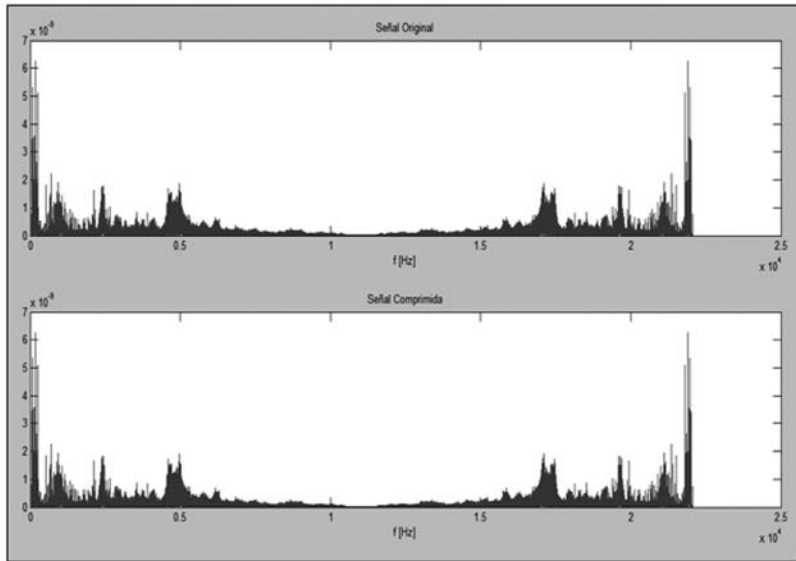


FIGURA 6. Comparación entre el espectro de la señal original y la señal comprimida.

La compresión es implementada con la Transformada de Wavelet-Daubechies Tipo 4. Una señal de “Castañuelas”, es la señal de prueba, con cambios abruptos en el tiempo, con dispersión espectral a través de varias bandas de frecuencia.

Puede observarse, sin embargo, que en el dominio de la frecuencia, la región de frecuencias altas no fue afectada significativamente. Los codificadores de compresión perceptual convencionales introducen distorsiones y degeneraciones en el espectro de este tipo de señales transientes, introduciendo perceptualmente una disminución en el brillo de la señal, apreciando a la señal de una forma más “opaca”; en este caso, con el codificador basado en la Transformada Wavelet-Daubechies Tipo II, no observamos una degradación importante. Verificando este fenó-

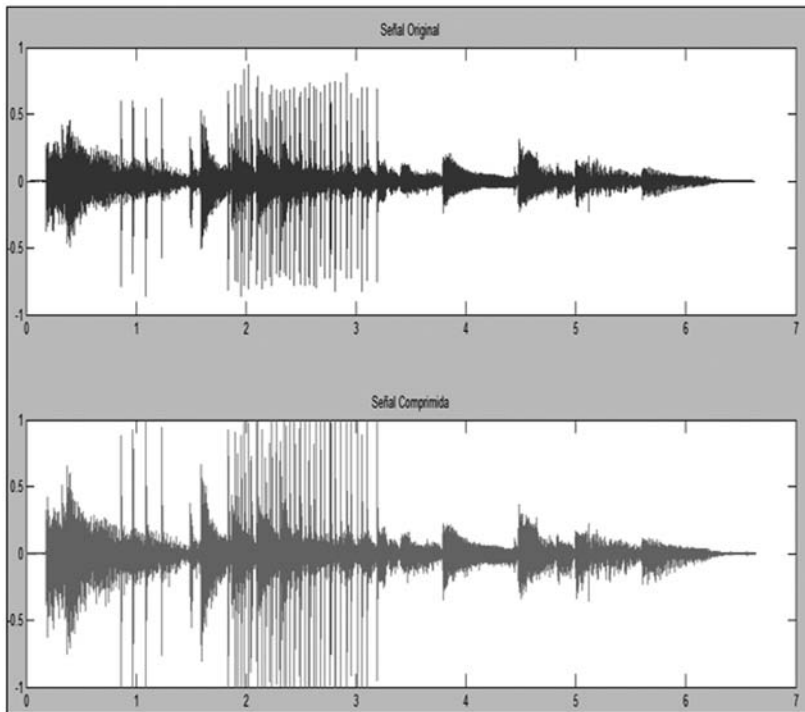
meno auditivamente, percibimos que efectivamente las diferencias entre la señal original y la señal comprimida son realmente insignificantes, y que el compresor, incluso con señales de este tipo, se comporta realmente de manera adecuada.

Cabe resaltar que si este comportamiento es observado en este tipo de señales, las diferencias entre señal-original y señal-comprimida en señales de naturaleza tonal, estacionaria, homogénea, serán inclusive mejores.

Diferencias perceptuales entre la señal original y la comprimida, son mínimas, pero la verificación debe realizarse con test subjetivos que apoyen la afirmación.

Una segunda comparación es realizada en el dominio del tiempo (Figura 7).

FIGURA 7. Comparación entre la representación temporal de la señal original y la señal comprimida.



Se puede observar que en el dominio del tiempo los cambios de la señal son abruptos en ambos casos; es una señal transiente. Las señales transientes son ideales para identificar el efecto de los codificadores de compresión perceptual, debido a que es precisamente con este tipo de formas-de-onda que estos sistemas encuentran problemas.

El modelo matemático implementado (Transformada Wavelet-Daubechies Tipo IV con dos niveles de descomposición), es más adecuado que los modelos tradicionales (FFT, DCT). La explicación a esto es sencilla. El modelo tradicional divide a la señal a través de una banca de filtros, con anchos de

banda iguales, y no permite por lo tanto una resolución significativa en ninguna de ellas. Esto, unido con el principio de incertidumbre de Gabor, el cual determina que no es posible, con transformadas convencionales (FFT, DCT) lograr resolución adecuada en frecuencia y paralelamente resolución adecuada en tiempo.

Conclusiones

Un codificador perceptual es implementado basado en la Transformada Wavelets-Daubechies Tipo 4. Este tipo de transformada posibilita simular el

modelo Psicoacústico sin la introducción de la representación clásica de Fourier.

La Transformada de Wavelet puede ser aprehendida como una descomposición de la señal, donde a medida que se avanza en nivel de descomposición, obtenemos versiones más refinadas de la señal.

Esta Transformada es similar a lo que conocemos como descomposición en fractales.

BIBLIOGRAFÍA

- Fastl, H., Zwicker, E.** (2006). *Psychoacoustics: Facts and Models* (3ed). Berlin: Springer-Verlag
- Kranz, F. W.**(1952). *Hearing aids*. Sonotone Corp.
- David, S. A.** (2008). *Concise Introduction to Data Compression*. Springer-Verlag.
- Proakis, J. Manolakis, D.** (2007). *Tratamiento Digital de Señales (4 ed.)*. Upper Saddle River, N.J. : Pearson- Prentice Hall.
- ISO/IEC 11172-3.** (1993). *Information Technology - Coding of moving pictures and associate audio for digital storage media at up to about 1,5 Mbit/s - Part 3: Audio*.
- Zwicker, E. & Fastl, H.** (1990). *Psychoacoustics*. Berlin Springer-Verlag, Berlin.
- Jhang-Liang L., Yan-Chen L. & Hsueh-Ming H.** (1998). *Scalable Audio Coding Using Wavelet Packet Transform*, Taipei: Proc. ISCE.
- Bosi, M., Goldberg, R. E.** (2003). *Introduction to digital audio coding and standards*, Boston: Kluwer Academic Publishers.
- Painter T., Spanias A.** (2000). *Perceptual Coding of Digital Audio*. Proceedings of the IEEE. 88(4), 451 – 515.
- Mallat S.** (1998). *A Wavelet tour of signal processing*. San Diego: Academic Press.