

Evaluación de artefactos de compresión de sonido

Evaluation of audio-coding artifacts

Recibido: septiembre de 2011
Arbitrado: octubre de 2011

Marcelo Herrera Martinez*

Resumen

El presente artículo trata de la evaluación subjetiva de sistemas de codificación de audio. De acuerdo con estos resultados se encuentra que, dependiendo del tipo de la señal y dependiendo del tipo de señal y del algoritmo del sistema de codificación de audio, diferentes tipos de errores audibles se hacen presentes. Estos errores son llamados errores de codificación, sin embargo, mientras pueden ser perceptibles tres clases de artefactos, el autor propone que en el dominio de codificación hay una clase común para la aparición de artefactos: el rastreo no-eficiente de señales transientes estocásticas. Para este propósito el estado del arte de los sistemas de audio codificación usan un gran rango de técnicas de procesamiento de señal, para lo cual la aplicación de la transformada de Wavelet está descrita.

Palabras clave

Codificación de audio, Artefactos, Transformada de Wavelet, Psicoacústica, Transformada Ortonormal.

Abstract

The present paper deals with the subjective evaluation of audio-coding systems. From this evaluation, it is found that depending on the type of the signal and the algorithm of the audio-coding system used, different types of audible errors arise. These errors are called coding artifacts. Nevertheless while in the auditory domain there can be perceivable three kinds of artifacts, the author purposes that in the coding domain there is only one common

cause for the artifact apparition, the non-efficient tracking of transient-stochastic signals. For this purpose, state-of-the art of audio coding systems use a wide range of signal processing techniques, from which the application of the wavelet transform is described here.

Keywords

Audio-coding, Artifacts, Wavelet transform, Psychoacoustics, Orthonormal transforms.

INTRODUCCIÓN

La Tecnología de Información ha tenido grandes avances en el almacenamiento y la transmisión de datos de audio. En 1981 el CD (Compact-Disc) fue desarrollado por la Corporación Philips en Holanda, el cual implementó una solución de almacenamiento basado en el láser óptico y la representación digital. Sin embargo, los límites de transmisión de datos, impuestos por la necesidad de bajar las tasas de transmisión, dieron origen a los algoritmos de compresión para reducir el flujo de información sin distorsionar de manera significativa la señal. En

* Ingeniero Electrónico, Magister en Radioelectrónica y Doctor en Acústica de la Universidad Técnica de Praga. Líder del Grupo de Investigación de Sistemas Interactivos y Videojuegos. E-mail: musicamarcelo@yahoo.es

1987, el Instituto Fraunhofer desarrolló un standard del algoritmo de compresión, basado en modelos perceptuales del sistema auditivo, usando los modelos del fenómeno de enmascaramiento. El ruido de cuantización introducido por estos sistemas de codificación, especialmente cuando se codifica a bajos niveles de compresión da origen a errores de distorsión audible, llamados artefactos. La evaluación subjetiva de estos artefactos lleva a una clasificación de estos artefactos, de los cuales el pre-echo captura atención. El estado del arte para la cancelación de pre-echo es discutida, y al final la técnica de la transformada de wavelet para este propósito es discutida, así como algunas consideraciones matemáticas sobre la transformada. Los codificadores híbridos, que hacen uso de la FFT, o DCT para los componentes quasi-periódicos de la señal, y de la DWT para ataques de la transiente de la señal parecen ser, de acuerdo con el autor, el camino correcto para una futura investigación.

I. DOS MÉTODOS PSICOMÉTRICOS PARA LA EVALUACIÓN DE SISTEMAS DE CODIFICACIÓN

DBTS y SR son dos métodos utilizados para la evaluación subjetiva de codificadores de audio. Los resultados de estos test han sido publicados en [1] y [2], así como la descripción de estos test, señales de audio y resultados. El método DBTS es un método psicométrico que introduce la señal referencia, por lo tanto el escucha realiza una comparación de la señal de codificación con la de referencia, y en el otro la señal de referencia no es introducida.

Source of Variation	Degr of Freed	Sums of Squares	Mean Square	Variance Ratio (F)	Probability
Factor A	5	109,8867	21,9773	59,0312	$p < 0.05$
Factor B	6	12,0919	2,0153	5,4131	$p < 0.05$
Factor A X B	30	47,5625	1,5854	4,2584	$p < 0.05$
Error	840	312,7176	0,3723		
Total	881	482,2587			

Tabla. 1. Resultados de ANOVA para la metodología DBTS

Source of Variation	Degr of Freed	Sums of Squares	Mean Square	Variance Ratio (F)	Probability
Factor A	5	7.1488	1.4298	5.7146	$p < 0.05$
Factor B	6	9.6904	1.6151	6.4552	$p < 0.05$
Factor A X B	30	7.4262	0.2475	1.12	$p < 0.05$
Error	966	241.6617	0.2502		
Total	1007	265.9271			

Tabla. 2. Resultados de ANOVA para la metodología SR

II. ARTEFACTOS DE COMPRESIÓN DE AUDIO

Los test subjetivos realizados en señales de codificación de audio muestran que codificadores individuales varían notoriamente en su rendimiento (esto es validado con el modelo ANOVA), y también tienen diferente rendimiento dependiendo del tipo de señal que está siendo usado para el test. Señales codificadas con un fuerte carácter aperiódico, llamadas "señales de ataque o señales con carácter transiente originan un artefacto llamado pre-echo. Similarmente, la codificación de señales de habla introduce en la señal un artefacto llamado reverberación. Algunas veces, cuando se codifica a bajas tasas de bits, variaciones de los umbrales de enmascaramiento de un campo al otro, llevan a diferentes asignaciones de bits, y como resultado algunos grupos de coeficientes espectrales pueden aparecer o desaparecer [3].

El primer artefacto mencionado, pre-echo, es analizado y algunas técnicas para su cancelación son descritas. Sin embargo, los investigadores analizando estos tres artefactos (aliasing no es discutido aquí, debido a que se está evaluando compresión, no frecuencia de muestreo conveniente), cuando describen generación de artefactos, coinciden en el sentido que el artefacto se origina debido a la asignación incorrecta de bits de cuadro a cuadro debido a la dispersión de la energía de la señal a cuadros vecinos, e incluso sub-bandas. Las relaciones entre las longitudes de dispersión dan origen a artefactos perceptuales diferentes. En el dominio del tiempo, señales con carácter transiente-estocástico, son aquellas que son afectadas, por lo tanto, señales percusivas como las castañuelas, timbales, clicks, palmas y tambores, entre otros, cuando son codificadas dan origen al pre-echo.

Fonemas explosivos, señales de habla estocástica con carácter de ruido, que son originadas por corrientes de aire turbulento dan origen a algunas consonantes. Cuando se codifican estas señales, las cuales están juntas con señales de habla de carácter cuasi-periódico, es decir, las vocales, se percibe reverberación.

A. ORIGEN DE LOS ARTEFACTOS DE COMPRESIÓN

La estructura general de un codificador de audio está dada en el siguiente esquema.

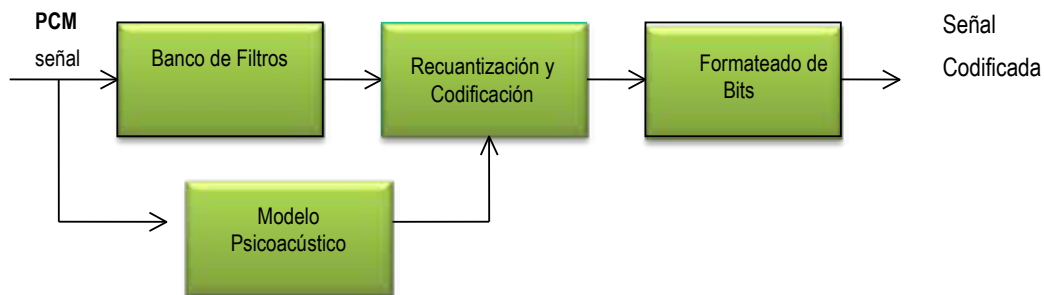


Fig. 2. Diagrama de un compresor de audio

Existen tres tipos de sistemas de codificación de audio que difieren básicamente en la forma como se alimenta la señal de entrada al modelo psicoacústico. El primer tipo, conocido como codificadores de transformación, donde muestras de la señal de entrada son transformadas al dominio de la frecuencia. El segundo tipo codificadores sub-banda, donde en un primer paso la transformación es realizado y luego los umbrales de enmascaramiento son calculados para cada sub-banda. El tercer tipo son

Cuando se codifica una señal que consiste no solo de los componentes explicados antes, pero que tiene una representación frecuencial que da variaciones fuertes del umbral de enmascaramiento de un cuadro al otro, el artefacto de birdies es percibido.

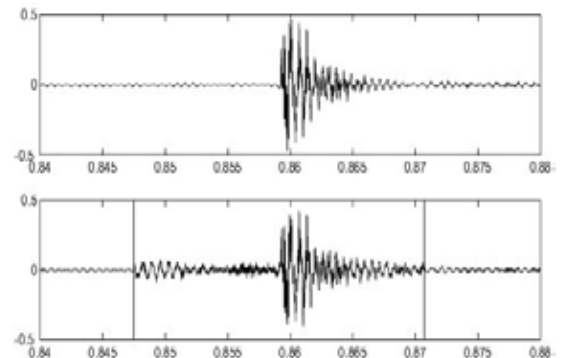


Fig. 1. Artefacto pre-echo generado en la señal de castañuelas [3]

codificadores paramétricos en donde un tipo de parametrización es observada.

Los codificadores sub-banda dan mejores resultados cuando hacen seguimiento de señales de transición, pero las ventanas de longitud finita que ellos aplican no los rastrean de manera exacta. Para este propósito un amplio rango de técnicas son implementadas como se describirá a continuación.

III. SELECCIÓN DEL MATERIAL CRÍTICO DE AUDIO

Durante el presente trabajo, el autor diseñó un programa en MATLAB para describir la energía de la señal en cada una de las sub-bandas que los codificadores sub-banda usan. Señales con carácter transiente mostrarán dispersión de su energía a otras sub-bandas vecinas.

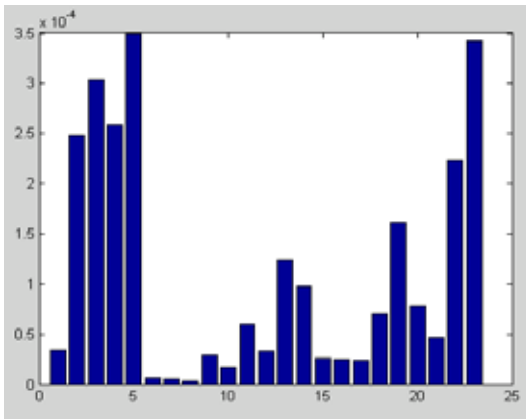


Fig. 2. Densidad espectral de potencia

Por lo tanto para seleccionar el material de audio que es conveniente para el asesoramiento subjetivo de codificadores de audio, el programa realiza una estimación de cuáles señales se comportarán de manera crítica y cuáles no. La investigación futura debe ser realizada para establecer las relaciones entre la representación sub-banda de una señal en particular y el artefacto producido una vez comprimida con un algoritmo definido para el seguimiento de transientes.

Por otra parte, determinar las relaciones entre los niveles de potencia espectral dentro de cada sub-banda debería dar una pista para investigación futura.

La Figura 2 muestra la alocaación de la energía de la densidad espectral de potencia de la señal "Castañuelas".

IV. DETECCIÓN DE SEÑALES TRANSIENTES DE AUDIO

El Procesamiento Digital de Señales evidentemente ofrece algunas posibilidades para la detección de transientes. Algunas de ellas incluyen modificaciones de la Transformada Discreta de Cosenos, DCT, con bloques de longitud variable, rastreando señales transientes de manera más exacta. La Transformada Discreta Wavelet, DWT, también representa una herramienta poderosa para el seguimiento de señales transientes. Algunas implementaciones realizan un DWT/DCT híbrido. Otras aproximaciones combinan codificación con transformadas no-lineales y técnicas de aproximación estructuradas, junto con la modelización híbrida de la señal bajo consideración. Técnicas con transformadas no uniformes con solapamiento (lapped transforms) también son usadas para este propósito. Aquí, un banco de filtros no-uniforme es obtenido, uniendo bancos de filtros modulados usando un filtro de transición. Audio watermarking, en donde una señal watermark modifica las características estadísticas de las señales de audio, en particular su estacionaridad, también es usado.[5].

A. APLICACIÓN DE LA TRANSFORMADA DE WAVELET PARA EL RASTREO DE TRANSIENTES

La representación de la señal en el dominio de la frecuencia en codificadores iniciales, como MPEG-1 Layer III, Ogg Vorbis y otros, está basado en la FFT, o DCT. Actualmente, aplicaciones que buscan un rastreo de transientes, usan DCT híbridos, DWT, entre otros.

La representación de la señal de audio, descargando el componente de ruido, puede ser hecho de la siguiente manera [6],

$$x_{ton} = \sum_{\delta \in \Delta} \beta_{\delta} \omega_{\delta}, \quad x_{tran} = \sum_{\lambda \in \Lambda} \alpha_{\lambda} \psi_{\lambda} \quad (1)$$

Donde $\{\psi_n, n = 0, \dots, N-1\}$ es una base de wavelet, $\{\omega_m, m = 0, \dots, N-1\}$ y es una base de MDCT.

$$\text{La señal resultante es } x = x_{tran} + x_{ton} + r \quad (2)$$

Daudet et all. en su trabajo [6] describieron Λ y Δ como subconjuntos de conjuntos de índice, denominados mapas de significancia. La señal residual r , no está correlacionada con las dos bases consideradas.

La idea principal es que DCT, FFT y los otros algoritmos implementados usualmente en audio compresión son convenientes para el análisis y el rastreo de sinusoides o componentes cuasi-estacionarios de la señal. El rastreo de transientes es más conveniente con la DWT. La transformada DWT, y su habilidad para localizar ataques abruptos en el tiempo se deduce de la transformada Fourier-Plancharel y el principio de incertidumbre.

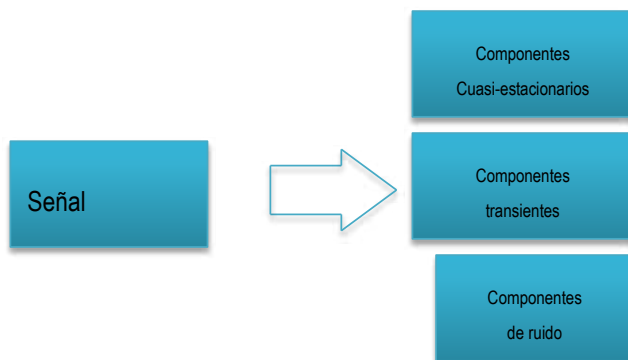


Fig. 3. Descomposición de la señal usada en estado del arte de los códecs

B. DEMOSTRACIÓN DE LA TRANSFORMADA WAVELET CUANDO SE RESUELVE UNA SEÑAL TRANSIENTE

El espectro de “Castañuelas”, uno de los materiales críticos de audio, cuando es procesado por la FFT o DCT con ventana de longitud fija, se dispersa de tal manera que la asignación de bits derivada del modelo psicoacústico no es eficiente y por lo tanto un artefacto audible llamado pre-echo es originado.

La figura 4 muestra la señal “Castañuelas” original,

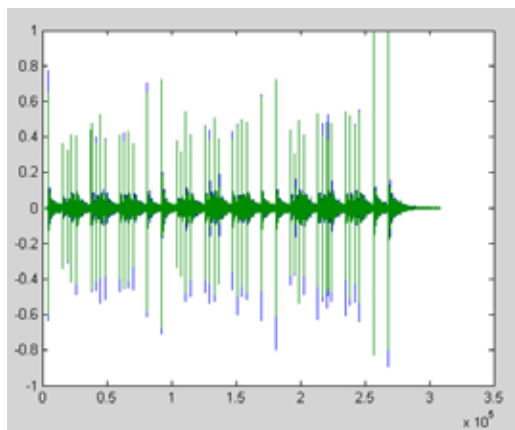


Fig. 4. Señal “Castañuelas” original, material crítico

DWT, FFT, DCT y las otras transformadas ortonormales realizan la descomposición de la señal en la base fundamental. En el caso de la FFT, la base de descomposición ortogonal es el conjunto de todas las funciones:

$$t \rightarrow \frac{1}{\sqrt{N}} e^{j\omega t \frac{2\pi}{N}}, \quad t = \{0, 1, 2, \dots, N-1\} \quad (3)$$

En la base de Fourier, la localización de frecuencia es precisa, pero la localización en tiempo es ineficiente.

La base Euclideana ortonormal, que tiene la forma:

$$(1, 0, 0, \dots, N-1), (0, 1, 0, \dots, N-1), \dots \quad (4)$$

Al contrario de la FFT, realiza una localización precisa en el tiempo, pero ineficiente en frecuencia. STFT soluciona parcialmente el problema con la implementación de la ventana de la señal, y por lo tanto da la posibilidad de separar a la señal en cuadros y obtener la representación frecuencial de estos cuadros de manera separada. Sin embargo, debido a la longitud de ventana fija, las señales de ataque transiente son rastreadas de manera no eficiente.

DWT representa un “compromiso” entre estas dos representaciones límites, y realiza una buena localización en frecuencia o en tiempo. La descomposición de la señal en una base particular puede ser vista como un producto escalar con el correspondiente coeficiente de la base. Matemáticamente,

$$(f, g) = \int_{-\infty}^{\infty} f(x) \overline{g(x)} dx \quad (5)$$

Representando qué tan similar es la función f al correspondiente coeficiente de la base ortonormal g .

La descomposición de la señal, matemáticamente expresada, es un mapeo del con-

junto de los números complejo al conjunto donde la descomposición es descrita,

$$C^n \rightarrow (z(0), z(1), \dots, z(n-1)) \quad (6)$$

Al realizar una descomposición de un paso de la señal de "Castañuelas", con la DWT, después de la descomposición de un paso, se obtienen dos componentes de señales, graficadas en la parte posterior,

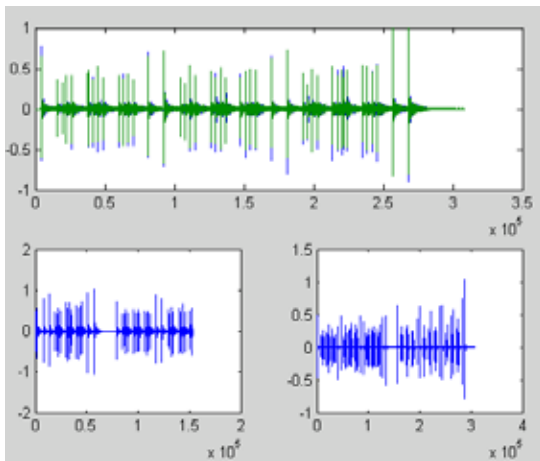


Fig. 5. Descomposición de 1-paso de la señal usando la transformada de wavelet

Reconstruir la señal con los coeficientes que aparecieron después de la descomposición de un paso muestra los siguientes resultados: La figura 6 grafica la señal reconstruida.

La descomposición de la señal a altos niveles, dará representaciones más exactas de la señal de audio, de manera similar como la resolución de altas frecuencias mejora la exactitud de la representación en frecuencia en la FFT.

DWT, entonces, tiene una estructura jerárquica en donde mientras más alto sea el nivel de la descomposición, el árbol de DWT será más largo.

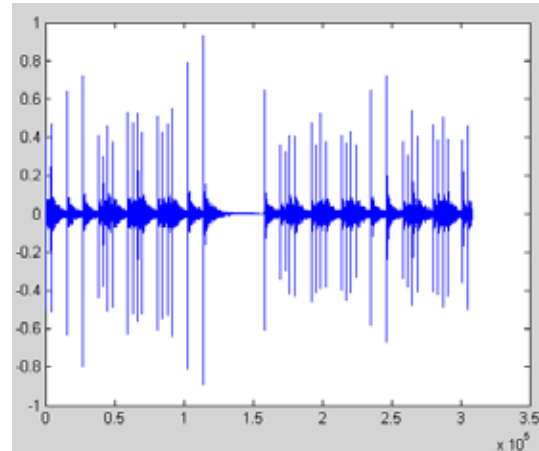


Fig. 6. Descomposición de la señal usando coeficientes.

Comparando las figuras 4 y 6 se observa que la reconstrucción fue realizada satisfactoriamente.

Ahora, se realiza una descomposición en 3-pasos. Un conjunto finito de coeficientes es obtenido. La extracción de coeficientes es realizada, y en las figuras siguientes, presentada,

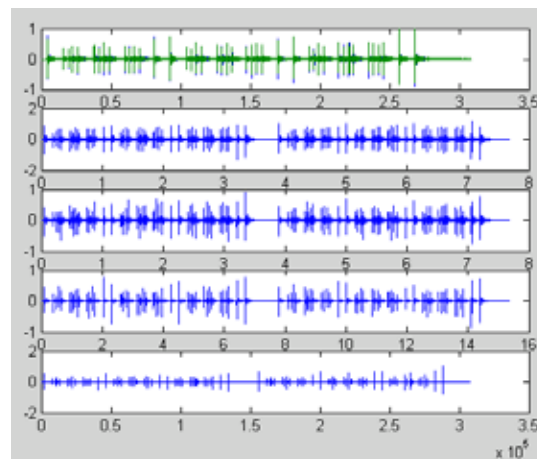


Fig. 7. Detalles de coeficientes en niveles 1, 2 y 3 de la estructura de descomposición de wavelet. Señal original, ca3, cd3, cd2 and cd1

Finalmente se reconstruye la aproximación en el nivel 3 de la estructura de descomposición de wavelet. Se realizan las reconstrucciones de detalles de coeficientes en niveles 1, 2 y 3 de la estructura de descomposición de wavelet.

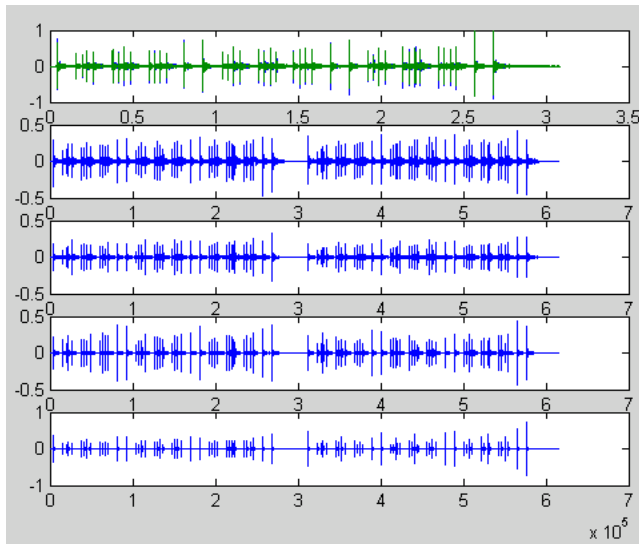


Fig.8. Detalles de coeficientes de reconstrucción en niveles 1, 2 y 3, de la estructura de descomposición wavelet

La figura superior es la señal original, seguida por la señal reconstruida, y al final los coeficientes. El último paso es la reconstrucción de la señal a partir de la estructura de descomposición de wavelet.

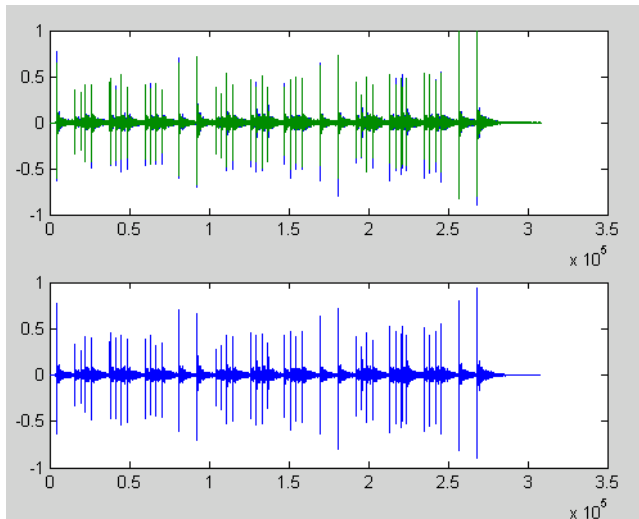


Fig. 9. Señal original y reconstruida

La reconstrucción de la señal transiente muestra que DWT es un método conveniente para descomponer señales transientes, aun realizando una descomposición de 3 niveles. Este resultado muestra que un códec híbrido implementando FFT para la extracción y el procesado de señales transientes es un algoritmo más conveniente para

codificación de sonido que codificadores tradicionales, los cuales rastrean señales con DCT con longitud fija de ventana o transformadas FFT.

CONCLUSIONES

Métodos psicométricos fueron usados para la evaluación de sistemas de codificación de audio. DBTS y SR fueron los métodos seleccionados para realizar la evaluación. A partir de estos test, la validación de resultados de ANOVA muestra que no sólo el rendimiento del codificador, sino además las características de la señal tienen un fuerte impacto en la evaluación.

Las señales con carácter percusivo, como las castañuelas, timbales, palmas, y otros cuando son codificados por algoritmos que implementan DCT y FFT para la representación frecuencial de la señal, muestran precho como un artefacto auditivo producido debido a compresión.

Los otros dos factores, a pesar de tener un carácter en el dominio auditivo diferente del ya mencionado, tienen, de acuerdo con el autor el mismo origen: la incorrecta aloca-ción de bits de coeficientes de enmascara-miento, y esto, porque la señal crítica tiene un espectro de potencia que se dispersa a cuadros vecinos, o a bandas vecinas.

Es posible de verificar la criticidad de la señal con el programa desarrollado. Al final, el estado del arte de algunas técnicas son discutidas para rastrear eficientemente estas señales críticas de audio, dando aten-ción especial a la transformada de wavelet.

REFERENCIAS BIBLIOGRÁFICAS

- [1] HERRERA, M. Summary of the subjective evaluation of audio-coding testing at the CVUT during the period 2003-2005. In *XI International Symposium of audio and video*. Krakov (Poland), 2005.

- [2] HUSNIK, L.-HERRERA, M. Comparison of two methods used for the subjective evaluation of compressed sound signals. In *Forum Acousticum*. Budapest, 2005.
- [3] AES. Tutorial CD-ROM, Perceptual Audio Coders, What to listen for. New York, 2002.
- [4] HERRERA, M.-DOLEJSI, P. Subjective evaluation of audio-coding systems. In *INTERNOISE 2004*. Prague, 2004.
- [5] LARBI, S.-JAIDANE, M. Audio Watermarking: A way to stationnarize audio signals. In *IEEE Transactions of Signal Processing*, Vol 53. No. 2. February 2005.
- [6] DAUDET, L.-MOLLA, S.-TORRESANI, B. Towards a hybrid audio coder. In *Proceedings of the International Conference Wavelet Analysis and Applications*. February 2004.
- [7] <http://www.mathworks.com/access/helpdesk/help/toolbox/wavelet/wavelet.htm>.



UNIVERSIDAD DE
SAN BUENAVENTURA
SEDE BOGOTÁ

Calidad humana y profesional

Ingeniería Mecatrónica



CÓDIGO SNIES	6647
TÍTULO QUE OTORGA	Ingeniero(a) Mecatrónico(a)
METODOLOGÍA	Presencial
DURACIÓN	10 semestres

Objetivo del programa

Formar ingenieros que respondan a las necesidades del entorno, capaces de realizar procesos de la investigación, diseño, desarrollo y aplicación de sistemas mecatrónicos en la industria, en la protección del medio ambiente, a través de la aplicación de los conocimientos adquiridos en las áreas de mecánica, electrónica, teoría del control y de la automatización.

•Competencias del ingeniero mecatrónico, bonaaventuriano

- Planeación, diseño y manejo de proyecto de manufactura, orientados a la industria.
- Adaptación, rediseño e implementación de maquinarias para aplicaciones industriales.
- Asesoramiento en la adquisición, implementación y mantenimiento de equipos de alta tecnología.
- Competencias para la investigación de tecnologías para el control digital y secuencial de procesos de manufactura industrial, aplicables a la agroindustria, la aeronáutica y la medicina.
- Formación interdisciplinaria para el trabajo exitoso en equipo, en cuanto a la investigación y diseños de naturaleza mecatrónica.

UNIVERSIDAD DE SAN BUENAVENTURA, SEDE BOGOTÁ - Carrera 6 - • 172-20 • PBX 667 1000 • Línea gratuita nacional: 01 8000 025 19
 Correo electrónico: informacion@usbog.edu.co • www.usbog.edu.co
 FACULTAD DE INGENIERÍA • Edificio Fray Diego Barrabá, oficina 201 • P.O. Box 1030 extensiones: 275 - 258