



Conference Paper

Compresión de Audio Basada en Codificación Subbanda

Aura Polo Llanos^{1,2}, Carlos Robles Algarín¹, and Jordan Guillot Fula^{1,2}¹Universidad del Magdalena, Santa Marta, Colombia²Universidad del Norte, Barranquilla, Colombia

Abstract

An eight channels subband audio codec is implemented for signals with 44.1 KHz and 16 bits per sample using Matlab. To achieve perfect reconstruction, a two channels QMF filter bank with cutoff frequency $\omega = \pi/2$ is designed, based on an equiripple filter of 99 order. Seven stages of this bank are used to split the input into eight signals with sample rates from 2.76 to 11 KHz, which are coded from 1 to 16 bits depending on the band energy. To evaluate performance for three tracks in terms of similarity of input and output signals, a Mean Opinion Score (MOS) experiment with fifteen subjects was performed. The Euclidean Distance between spectrums was also measured. Results showed a fair similitude for two tracks and excellent for one. Compression factors above 96% were achieved.

Keywords: subband coding, mean opinion score MOS, QMF filter banks, audio compression

Resumen

Se implementa un codificador subbanda de 8 canales para señales de entrada con frecuencia de muestreo de 44.1KHz y 16 bits por muestra, utilizando Matlab. Para reconstrucción perfecta se diseña un banco de filtros QMF de dos canales con frecuencia de corte $\omega = \pi/2$, a partir de un pasabajo equiripple de orden 99. Se utilizan siete etapas del banco diseñado para generar ocho señales con frecuencias de muestreo de 2.76 a 11 KHz, las cuales son codificadas con tamaños de palabra de 1 a 16 bits, dependiendo de la energía de la banda. Para hacer una evaluación subjetiva del desempeño, se mide la similitud de la señal original con la codificada para tres pistas musicales, realizando una prueba MOS a quince sujetos. Como medición objetiva se calcula distancia euclidiana entre espectros de la señal original y la reconstruida. Los resultados mostraron similitud aceptable para dos pistas y excelente para una. Se logran factores de compresión superiores al 96%.

Corresponding Author:

Aura Polo Llanos

aura.polollanos@gmail.com

Received: 15 November 2017

Accepted: 5 January 2018

Published: 4 February 2018

Publishing services provided
by Knowledge E

© Aura Polo Llanos et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Selection and Peer-review under the responsibility of the ESTEC Conference Committee.

OPEN ACCESS

Palabras claves: codificación subbanda, prueba MOS, bancos de filtros QMF, compresión de audio

1. Introducción

Los sistemas multitasa en Procesamiento Digital de Señales (PDS) responden a la necesidad que se presenta en muchas implementaciones de utilizar más de una tasa de muestreo. La codificación subbanda (CSB) es un claro ejemplo de aplicación de sistemas multitasa, donde se busca hacer un uso más eficiente de recursos como el ancho de banda de un canal, comprimir datos sin tener efectos trascendentales sobre la calidad de las señales o simplemente descomponer una señal en componentes de frecuencia para procesarlas separadamente (Proakis and Manolakis, 1996; Vaidyanathan, 1990). Dicha técnica encuentra aplicación en muchos campos: en la bioingeniería, se utilizan técnicas de descomposición para hacer clasificación de señales electrofisiológicas como las EEG (Kevric and Subasi, 2017), para hacer compresión (Oosugi et al., 2017) y también eliminación de ruido (Poornachandra, 2008); en procesamiento de imágenes se trabaja el mejoramiento de imágenes y su compresión (Anbarjafari and Demirel, 2010) y en procesamiento de audio se ha convertido en una herramienta poderosa en la implementación de codecs de audio para reducir la tasa de bits (Winkler and Faller, 2005; April, 2015).

El objetivo del presente trabajo es implementar la codificación en ocho bandas de señales de audio monofónicas de calidad CD, con una frecuencia de muestreo de 44100Hz y 16 bits por muestra. Se explican a continuación las etapas de análisis, codificación, decodificación y síntesis utilizadas para lograr dicho objetivo.

2. Reconstrucción perfecta con banco de filtros qmf

Los tres componentes básicos de un sistema CSB son filtros, diezmadores e interpoladores. Donde el diseño de los filtros es crucial para evitar problemas de distorsión del espectro y aliasing derivados del submuestreo. La metodología sugerida en el estado del arte para resolver el anterior problema es el uso de bancos de filtros de reconstrucción perfecta de dos canales. Por mencionar algunas soluciones de reconstrucción perfecta, se tienen los bancos de filtros en espejo y cuadratura QMF (Quadrature Mirror Filters), los CQF (Conjugate Quadrature Filters) y los Pseudo-QMF o PQMF utilizados en el algoritmo de codificación MPEG (Bosi and Goldberg, 2003). En el presente trabajo

se utilizan filtros QMF por su baja complejidad de implementación y alto desempeño demostrado en (Vaidyanathan, 1990).

El diseño de un banco QMF consta de cuatro filtros, dos de análisis y dos de síntesis, como se puede ver en la Figura 1.

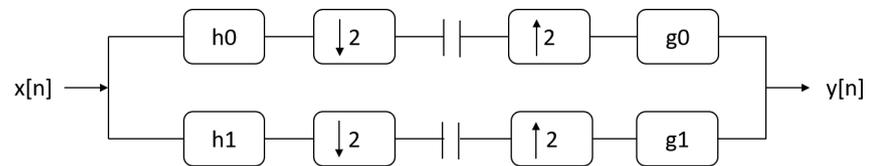


Figura 1: Diagrama de un Banco QMF de dos Canales.

Se parte de la obtención de la respuesta al impulso h_0 del filtro pasobajo de análisis y a partir de ésta se generan las respuestas de los tres filtros restantes. Se utilizaron las ecuaciones (1), (2) y (3), donde h_1 , g_0 y g_1 son respectivamente las respuestas de los filtros pasoalto de análisis, pasobajo de síntesis y pasoalto de síntesis (Proakis and Manolakis, 1996).

$$h_1 [n] = -1^n h_0 [n] \tag{1}$$

$$g_0 [n] = -2h_0 [n] \tag{2}$$

$$g_1 [n] = -2(-1)^n h_0 [n] \tag{3}$$

3. Método

A continuación, se describe la metodología utilizada para el diseño de las etapas de análisis y síntesis implementadas para el CODEC desarrollado e igualmente se presenta el protocolo de prueba seguido.

3.1. Análisis

La señal de audio de entrada es dividida en ocho bandas como muestra la Figura 2. Se diseña un banco QMF de dos canales, el cual es repetido siete veces hasta lograr la distribución del espectro mostrada en la Figura 3. Se diseña, como filtro generador, un filtro FIR pasobajo equiriple de orden 99 con frecuencia de paso de 9600 Hz y frecuencia de rechazo de 12000 Hz, que satisface las condiciones de fase lineal, baja atenuación en la banda de rechazo, respuesta plana en banda de paso y corte aproximadamente en $\omega = \pi/2$. A partir de dicho filtro se obtienen las respuestas de los tres

filtros restantes utilizando las ecuaciones (1), (2) y (3). Seguido al filtrado se utiliza el comando *mfilt.firdecim* del DSP System Toolbox de MatLab para implementar los filtros en polifase con un factor de diezmado de 2 y luego el comando *filter* de MatLab para obtener las salidas respectivas.

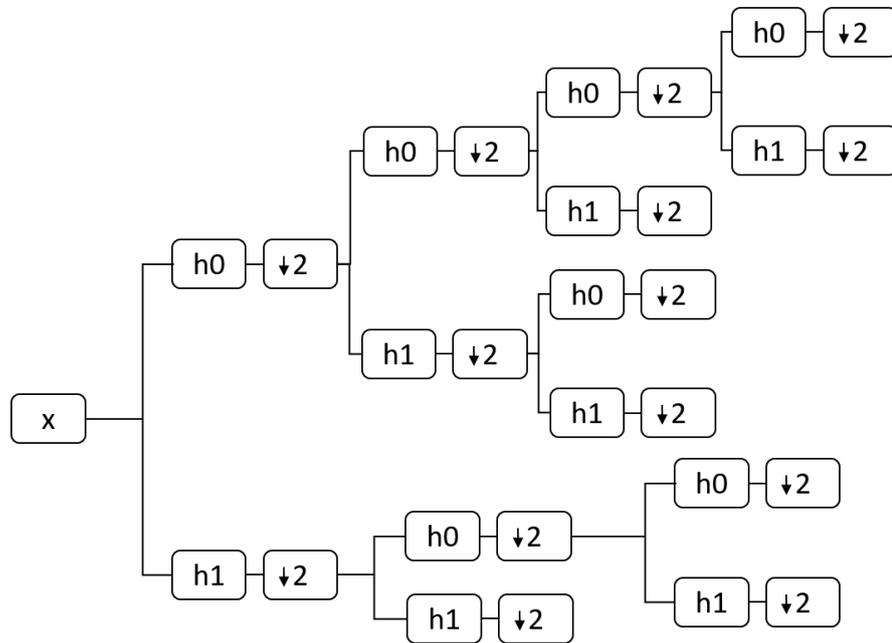


Figura 2: Diagrama General de Etapa de Análisis

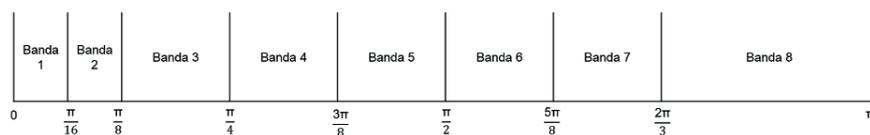


Figura 3: Distribución del Espectro Después del Análisis

Las respuestas de magnitud de los filtros del banco QMF de dos canales diseñado, mostradas en las Figura 4, demuestran que se cumple con la condición de respuestas en espejo y cuadratura para los cuatro filtros.

3.2. Codificación y decodificación

Para la codificación de las señales provenientes de cada una de las subbandas obtenidas en la etapa de análisis, se define un criterio basado en el cálculo de las energías de cada banda, a partir de la ecuación (4). Con respecto a la energía máxima obtenida, se establecen cuatro umbrales para asignar los bits según la Tabla 1.

$$E = \sum_{m=0}^{N-1} |x(n)|^2 \tag{4}$$

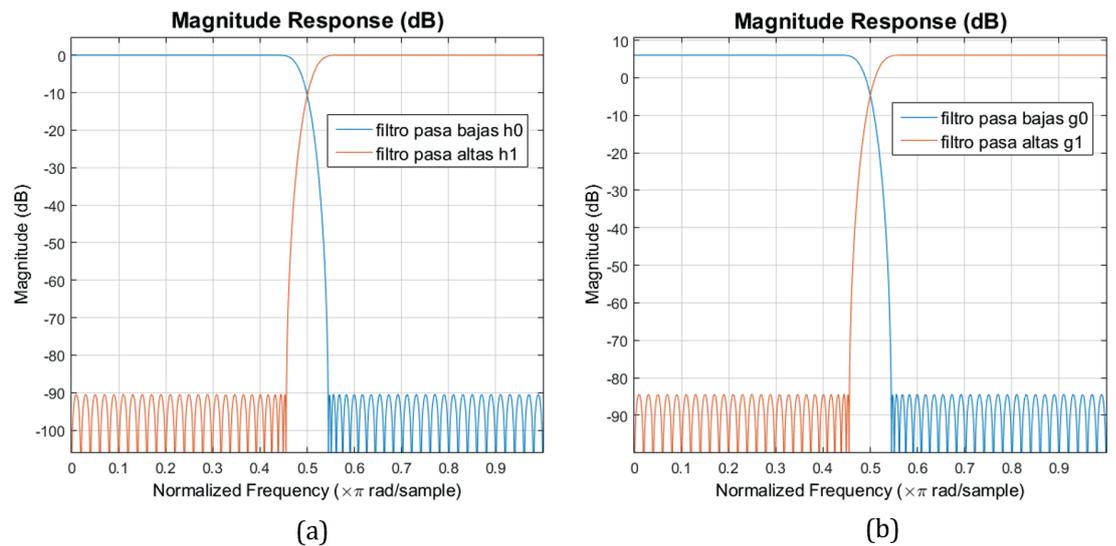


Figura 4: Respuesta de Magnitud: a) Filtros de Análisis. b) Filtros de Síntesis

TABLA 1: Asignación de Número de Bits Según Energía de la Subbanda

Condición	N° Bits
$0.4E_{max} < E \leq E_{max}$	16 Bits
$0.1E_{max} < E \leq 0.4E_{max}$	8 Bits
$0.02E_{max} < E \leq 0.1E_{max}$	4 Bits
$0.01E_{max} < E \leq 0.02E_{max}$	2 Bits
$E < 0.01$	1 Bits

De dicha asignación se crea un vector de ocho posiciones, con el cual se hace la codificación y decodificación de las muestras utilizando el constructor de objetos de punto fijo del DSP System Toolbox de MatLab, fi. Esta herramienta fue utilizada de forma que entregara una cantidad de entrada codificada a 16 bits en una cantidad de salida con signo, del tamaño especificado por la asignación anterior y ajustada a la mejor precisión de longitud de fracción. Para obtener el valor cuantizado a Nb (1) bits de la banda 1 con la mejor precisión se utilizó el código de MatLab: $c = double(fi(x, 1, Nb(1)))$.

Cabe anotar que los umbrales fueron ajustados empíricamente con sucesivas reproducciones del audio de salida y aplicando una prueba MOS preliminar, hasta obtener una combinación de límites apropiada.

3.3. Síntesis

Las ocho señales que salen de la decodificación, son interpoladas por un factor de dos, filtradas utilizando los filtros diseñados y sumadas en siete etapas para obtener la versión reconstruida de la señal original de audio. En la Figura 5 se muestra el esquema de la operación de síntesis.

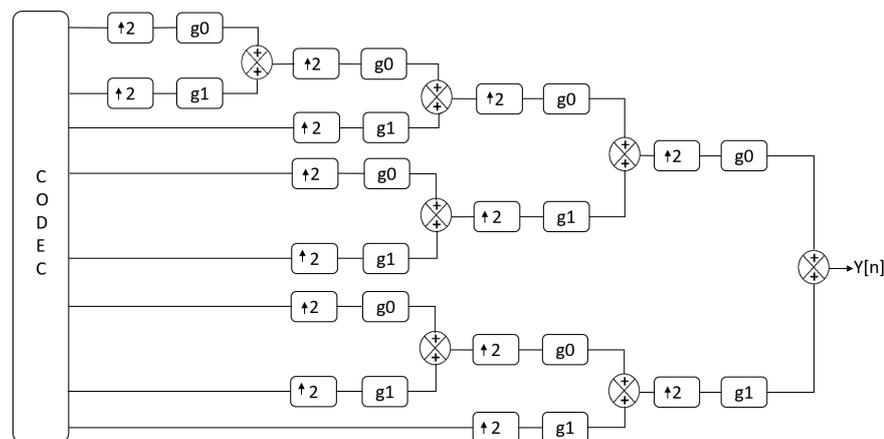


Figura 5: Diagrama General de Etapa de Síntesis

De forma similar que en el análisis se implementan los filtros, pero en este caso utilizando el comando *mfilt.firinterp* del DSP System Toolbox de MatLab para construir filtros polifase con un factor de interpolación de 2.

3.4. Medición de percepción de calidad (prueba mos)

Para medir el desempeño del codificador diseñado, en términos de la similitud de la señal original y la codificada, se utilizan dos tipos de pruebas, una subjetiva y una objetiva. Una prueba MOS (Mean Opinion Score), es realizada adaptando las recomendaciones presentadas en la serie de estándares P.800 de la ITU (International Communication Union) (ITU-T, 2016). El protocolo seguido fue el siguiente:

- A cada sujeto se le pide escuchar atentamente en un ambiente silencioso una pista musical con duración de 8 segundos en dos versiones, una la original y otra la codificada. Al sujeto le son presentadas las dos versiones en orden aleatorio.
- Luego se le pide calificar en una escala continua de 1 a 5 la imparidad de las versiones. Dando una puntuación máxima en caso de no percibirse diferencia y mínima al caso en que se perciben totalmente diferentes y alguna de ellas es perturbante al oído.

- Además, se le pide calificar la percepción de calidad de cada una de las versiones en una escala continua de 1 a 5, siendo 5 calidad excelente y 1 mala calidad.

Se aplica la prueba a cada uno de 15 sujetos, para 3 pistas musicales seleccionadas: Furelise, Timbalero y BigBoy. La primera es una pista instrumental de música clásica, la segunda una salsa instrumental y la última un rap con bastante contenido de voz. Los datos son recolectados en una hoja de cálculo de MS EXCEL, donde se calculan los siguientes parámetros:

- MOS₁: percepción de imparidad. Promedio de puntajes emitidos por sujetos con respecto a la percepción de imparidad de las versiones escuchadas.
- MOS₂: original. Promedio de puntajes emitidos por sujetos sobre la percepción de calidad de la versión original.
- MOS₃: códec. Promedio de puntajes emitidos por sujetos sobre la percepción de calidad de la versión procesada por el codificador subbanda.
- Diferencia. Promedio de las diferencias entre los puntajes emitidos por cada sujeto sobre la calidad de la versión original y de la versión codificada.

3.5. Medición objetiva de desempeño (distancia euclidiana)

Como prueba objetiva de la similitud entre las señales original y codificada, se escogió el cálculo de la distancia euclidiana entre los espectros resultantes de las dos señales. De acuerdo con la ecuación (5) se obtienen los valores DE para cada punto de la FFT y luego se promedian.

$$DE = |fft(x[n])| - |fft(y[n])| \quad (5)$$

4. Resultados y discusión

Se presentan los resultados obtenidos de la realización del experimento MOS, la medición de Distancia Euclidiana y la medición de compresión.

4.1. Resultados prueba mos

En la Tabla 2 se muestran los resultados obtenidos para los parámetros MOS calculados a partir de los datos recolectados. La pista con mayor puntuación es Timbalero, con un resultado en la diferencia entre versiones que muestra que se percibió más calidad

en la versión codificada que en la versión original. Los resultados de percepción de imparidad son muy buenos para la pista Timbalero, no se percibe diferencia; y más que aceptables para las dos pistas restantes, pues se obtienen promedios superiores a 3, cercanos a 4, implicando que si hay percepción de diferencia entre las versiones pero que no es molesta.

En la Figura 6 se presentan los gráficos de los puntajes obtenidos en las tres pistas para cada parámetro contra los sujetos de prueba. Se evidencia la característica subjetiva de esta prueba ya que presenta gran dependencia de los sujetos de prueba, pues en varios casos presentan grandes variaciones. Este es el caso de la pista Furelise, en la que los puntajes oscilan entre 1 y 5, incluso para la versión original. Lo anterior puede incluso implicar que no hay afinidad de algunos sujetos por ese tipo de música. En el caso de Timbalero, los datos muestran menos variación, especialmente para la versión codificada, con puntajes que oscilan entre 4 y 5.

TABLA 2: Resultados MOS Obtenidos Para Cada Pista

Parámetro	Pista		
	Furelise	Timbalero	BigBoy
MOS ₁	3.65	4.37	3.65
MOS ₂	4.36	4.21	4.54
MOS ₃	3.32	4.46	3.47
Prom Diferencia	1.04	-0.25	1.07

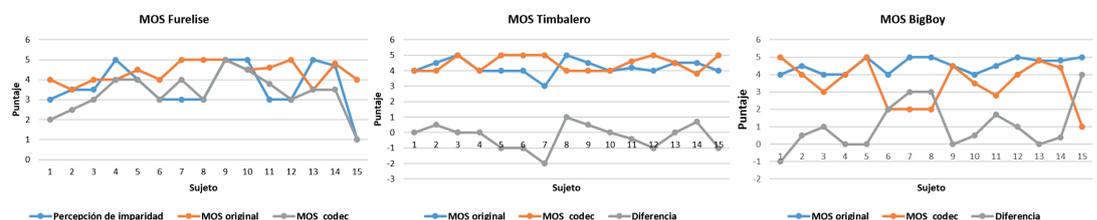


Figura 6: Gráficas de Puntajes Vs. Sujetos Para las Pruebas MOS Realizadas con las Tres Pistas

4.2. Resultados prueba distancia euclidiana

Los resultados para los promedios DE muestran una medida de mayor similitud para la pista Furelise con una desviación relativa de los espectros del 0.45%, le sigue BigBoy con un 4.56% y Timbalero con 5.78%. La aparente contradicción entre los resultados de las pruebas subjetiva y objetiva, demuestra en este caso que la percepción en audio

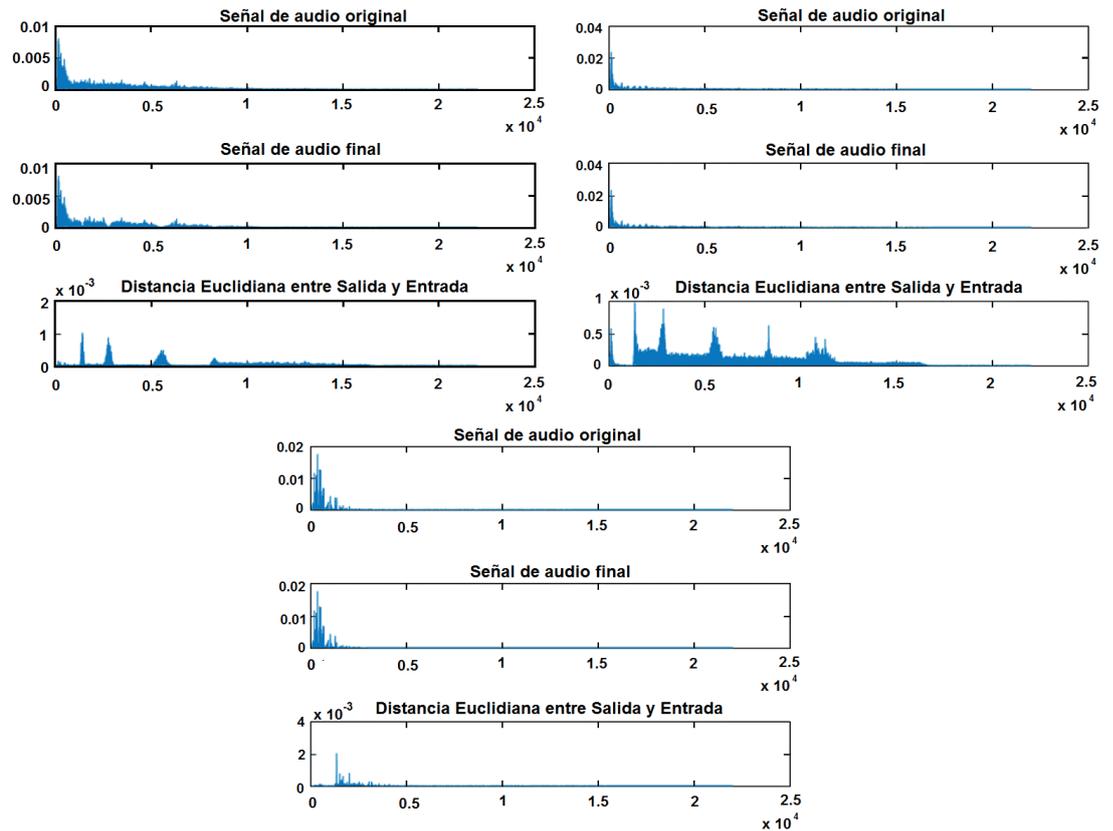


Figura 7: Espectros de Entrada y Salida Para: a) Timbalero. b) BigBoy. c) Furelise

es definitivamente muy dependiente de factores como la afinidad hacia el tipo de muestra escuchada y el público que escucha.

En la Figura 7 se muestran los espectros de entrada y salida para las tres pistas, junto con las DE Vs frecuencia. Analizando los espectros, se evidencia gran similitud entre entrada y salida de Timbalero en la mayor parte del espectro, pero con unas discontinuidades abruptas en 3 zonas del espectro, que se ven claramente en la gráfica respectiva de DE. Lo anterior explica entonces el DE promedio más alto obtenido para esta pista, es decir que dichas discontinuidades parecen no tener impacto negativo sobre el audio resultante, tanto que fue la pista mejor evaluada en la prueba subjetiva. El caso de BigBoy muestra que en la mayor parte del espectro se presentan diferencias entre las dos señales y concordaría con la evaluación aceptable obtenida en la prueba subjetiva. Finalmente, con Furelise evidentemente los espectros presentan menor diferencia que en las otras dos pistas.

TABLA 3: Resultados de Compresión

Banda	D	Fs(Hz)	Furelise			Timbalero			BigBoy		
			N (bits)	Tb (Kbps)	FC	N (bits)	Tb (Kbps)	FC	N (bits)	Tb (Kbps)	FC
0	16	2756.25	16	44.10	0.94	16	44.10	0.94	16	44.10	0.94
1	16	2756.25	2	5.51	0.99	8	22.05	0.97	4	11.03	0.98
2	8	5512.5	1	5.51	0.99	8	44.10	0.94	4	22.05	0.97
3	8	5512.5	1	5.51	0.99	4	22.05	0.97	4	22.05	0.97
4	8	5512.5	1	5.51	0.99	8	44.10	0.94	4	22.05	0.97
5	8	5512.5	1	5.51	0.99	1	5.51	0.99	1	5.51	0.99
6	8	5512.5	1	5.51	0.99	1	5.51	0.99	1	5.51	0.99
7	4	11025	1	11.03	0.98	2	22.05	0.97	4	44.10	0.94
FC Promedio					0.98			0.96			0.97

4.3. Resultados de compresión

Un factor importante para medir el desempeño de un codificador es el factor de compresión. En este caso se halló la relación entre la tasa de bits de la señal de entrada, que es de 705.6 Kbps con las tasas de bits logradas en cada una de las 8 bandas para las tres pistas evaluadas. La Tabla 3 muestra los resultados obtenidos, enfatizando el promedio de factor de compresión logrado para cada pista. Se evidencia que el codificador diseñado logra altos factores de compresión, desde el 96% para Timbalero que fue la pista mejor evaluada.

5. Conclusiones

Se logró implementar un codificador subbanda de señales de audio monofónicas con tasas de compresión altas superiores al 94%, para tres pistas de prueba con diferentes características espectrales. La medición perceptual de desempeño arrojó buenos resultados para el codificador, a pesar que se pudo evidenciar que la percepción varía mucho dependiendo del tipo de audio o melodía, los gustos musicales y las condiciones del entorno. En el diseño del filtro pasabajo de análisis es importante lograr un rizado bajo en la banda de rechazo y una banda de transición angosta. Con los resultados presentados se confirma la importancia de implementar codificadores basados en características perceptuales, para lograr altas tasas de compresión sin afectar significativamente la calidad del audio.

Referencias

- [1] Anbarjafari, G., and Demirel, H. (2010). "Image super resolution based on interpolation of wavelet domain high frequency subbands and the spatial domain input image". *ETRI J.*, Vol. 32, No. 3, pp 390-394.
- [2] April, M. (2015). "Sub-band Coding of Speech Signals using Multirate Signal Processing and comparing the various parameter of different speech signals". *IJETCS*, Vol. 4, No. 2, pp 217-221.
- [3] Bosi, M., and Goldberg, R.E. (2003). *Introduction to Digital Audio Coding and Standards*, 1st edition, Springer, USA.
- [4] ITU-T. (2016). Recommendation ITU-T P.800.1, <https://goo.gl/sGc7iA>, 06/15/2017. (date accessed)
- [5] Kevric, J., and Subasi, A. (2017). "Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system". *Biomed. Signal Process. Control*, Vol. 31, pp 398-406.
- [6] Oosugi, N., Kitajo, K., Hasegawa, N., Nagasaka, Y., Okanoya, K., and Fujii, N. (2017). "A new method for quantifying the performance of EEG blind source separation algorithms by referencing a simultaneously recorded ECoG signal". *Neural Networks*, Vol. 93, pp 1-6.
- [7] Poornachandra, S. (2008). "Wavelet-based denoising using subband dependent threshold for ECG signals". *Digit. Signal Process. A Rev. J.*, Vol. 18, No. 1, pp 49-55.
- [8] Proakis, J.G., and Manolakis, D.G. (1996). *Digital Signal Processing*, 3rd edition, Prentice Hall, USA.
- [9] Vaidyanathan, P.P. (1990). "Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial". *Proceedings of the IEEE*, Vol. 78, No. 1, pp 56-93.
- [10] Winkler, S., and Faller, C. (2005). "Maximizing Audiovisual Quality at Low Bitrates". *Electron. Imaging 2005*, Vol. 2005, pp 139-148.