

ANÁLISIS COMPARATIVO DE CODIFICADORES DE AUDIO SIN PÉRDIDAS Y UNA HERRAMIENTA GRÁFICA PARA SU SELECCIÓN Y PREDICCIÓN DE SU DESEMPEÑO

Fernando A. Marengo Rodríguez^a, Eriberto A. Roveri^b, Juan M. Rodríguez Guerrero^c y
Mauro A. Treffiló^d

*Laboratorio de Acústica y Electroacústica, Escuela de Ingeniería Electrónica, Facultad de Ciencias
Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario, Riobamba 245 bis, 2000
Rosario, Argentina, <http://www.fceia.unr.edu.ar/acustica/>*

^afmarengorodriguez@yahoo.com.ar, fmarengo@ie.fceia.unr.edu.ar

^bea_roveri@hotmail.com.ar

^cjmanu1986@gmail.com

^dmaurotreffilo@gmail.com

Palabras Clave: Codificadores de audio, codificación sin pérdidas, rendimiento, predictores, géneros musicales.

Resumen. A diferencia de los codificadores de audio perceptuales como MP3, los codificadores de audio sin pérdidas se caracterizan por no alterar la calidad del archivo original, aunque logran menor compresión. En los últimos tiempos, estos codificadores han ganado gran popularidad principalmente debido al incremento de la capacidad de almacenamiento de los dispositivos digitales y del ancho de banda disponible para la transmisión de datos entre equipos remotos. Actualmente existen varios codificadores de audio sin pérdidas, pero no se ha analizado su rendimiento (en términos de velocidad y compresión) al procesar piezas de diferentes géneros musicales. En este trabajo se realiza dicho análisis con varios codificadores populares y piezas provenientes de CD comerciales. En base a este estudio se diseña una herramienta gráfica que determina los codificadores de mejor rendimiento. Mediante un proceso de normalizado, esta herramienta se utiliza para predecir el rendimiento de cualquiera de los codificadores recién mencionados para una dada colección musical. El estudio se realizó en diferentes computadoras y contemplando las diferentes opciones de optimización de cada codificador. La precisión del método propuesto se corroboró con resultados experimentales.

1 INTRODUCCIÓN

La compresión de datos ha desempeñado un rol fundamental en el procesamiento de señales digitales. Con su aplicación se minimiza el espacio de almacenamiento y el ancho de banda necesario para la transmisión de datos. En cuanto a los datos de audio, se han desarrollado muchos algoritmos tanto para la codificación perceptual (codificadores con pérdidas) como para la codificación sin pérdidas (codificadores sin pérdidas, en adelante CSP). Los primeros son bien conocidos por la alta compresión que alcanzan, gracias a la supresión de componentes psicoacústicamente no perceptibles de la señal de entrada.

Los codificadores de audio sin pérdidas no producen distorsión de la señal. Esto se logra a expensas de obtener una menor compresión en comparación con sus homólogos con pérdidas. En los últimos años, el uso de CSP se ha vuelto más popular por las siguientes razones:

- Tecnológicas. Al aumentar la capacidad de los dispositivos de almacenamiento y la disponibilidad de servicios con mayor ancho de banda para transmisión de datos a través de Internet, se redujo la diferencia entre el uso de codecs (codificadores/decodificadores) con pérdidas frente a los CSP.
- Económicas. Los equipos de audio de alta calidad se han vuelto más accesibles y por lo tanto más populares. En estos dispositivos, la distorsión introducida por los métodos de codificación con pérdida es auralmente perceptible.
- Masterizado. Las grabaciones de estudio pueden ser compartidas entre dispositivos remotos en formato comprimido sin pérdidas, lo cual permite ahorrar tiempo en la transmisión remota de datos así como el espacio de almacenamiento. Los formatos con pérdidas no pueden ser utilizados en estos casos.

En la actualidad existen muchos CSP y se han realizado muchas comparaciones entre ellos ([Wikipedia, 2011](#); [Malvar, 2007](#); [Hans y Schafer, 2001](#)). Sin embargo, se sabe poco sobre el rendimiento de los CSP más populares con respecto a los géneros musicales. Esta información es muy importante, ya que el usuario de este tipo de codecs suele consumir determinados géneros musicales. Además, se sabe empíricamente que la compresión es dependiente del género musical tratado.

Por otra parte, no existe ninguna herramienta para que el audiófilo pueda determinar el mejor CSP en términos de velocidad y compresión. Finalmente, no existe un método para la predicción de la velocidad y compresión de manera que el audiófilo pueda seleccionar el codificador óptimo de acuerdo a su preferencia, sea ésta minimizar espacio de almacenamiento o comprimir una colección musical en el menor tiempo posible.

El propósito de este trabajo es, en primer lugar, caracterizar en términos de velocidad y compresión los resultados de muchos CSP populares con respecto a diferentes géneros musicales. En segundo lugar se propone un método para seleccionar los codificadores óptimos (en términos de velocidad y compresión) de una determinada colección de piezas de audio. En tercer lugar se presenta una herramienta para la predicción de la compresión y del tiempo de procesamiento de una determinada colección musical, utilizando cualquiera de los codificadores seleccionados en el punto anterior. Todos los codecs evaluados en este trabajo han sido probados con piezas musicales extraídas de discos compactos comerciales con una frecuencia de muestreo de 44 100 Hz, resolución de 16 bits y formato PCM estéreo. Estas piezas se corresponden con diferentes géneros musicales muy populares en La Argentina, y el análisis propuesto se puede extender fácilmente a otros géneros.

Este trabajo se encuentra organizado de la siguiente manera. En la sección 2 se describen los aspectos generales de los CSP. En las secciones 3 y 4 se describe el método de análisis comparativo de los CSP y se exponen los resultados, respectivamente. La herramienta gráfica

propuesta para la selección de los codificadores óptimos en el dominio velocidad-compresión se expone en la sección 5, y en la sección 6 se describe el método que permite predecir el desempeño de codificación de una determinada colección de piezas de audio.

2 FUNCIONAMIENTO DE LOS CODIFICADORES DE AUDIO SIN PÉRDIDAS

Como se muestra en el diagrama de bloques de la [Figura 1](#), los CSP procesan la señal de entrada de cada canal en cuadros.



Figura 1: Esquema general de los codificadores de audio sin pérdidas.

En cada cuadro, se reduce la redundancia intracanal (etapa de decorrelación) mediante la estimación eficiente de la señal de entrada $x(n)$. Dicha estimación tiene la finalidad de obtener, mediante pocos parámetros, una aproximación de la entrada. Gracias a esta estimación resulta un error de aproximación o residuo $e(n)$ con menor varianza (es decir menor potencia) que la entrada, resultando por ende un conjunto de datos más fácilmente comprimible ([Salomon, 2007](#)). En muchos CSP, la entrada se estima vía predicción lineal. En cambio, en otros codificadores, dicha estimación se lleva a cabo extrayendo unos pocos coeficientes de la transformada de la entrada en el dominio frecuencial ([Purat et al. 1997](#)).

La tasa de datos del residuo se reduce mediante un codificador de entropía adecuado para su distribución, que generalmente es laplaciana. En estos casos se suele utilizar codificación Rice ([Salomon, 2007](#); [Hans y Schafer, 2001](#); [Robinson, 1994](#)). Según se consultó con varios desarrolladores de CSP, estos datos ocupan más del 90 % del archivo codificado. Cabe aclarar que la información restante de dicho archivo contiene los parámetros del predictor utilizado además de metadatos.

En caso de decorrelacionar la entrada vía predicción lineal, el diagrama de bloques correspondiente es el expuesto en la [Figura 2](#). Cada muestra de la entrada $x(n)$ es aproximada por la combinación lineal de N muestras de entrada anteriores (sección de respuesta finita al impulso o FIR) y de M muestras de salida anteriores (sección de respuesta infinita al impulso o IIR). En general, el modelo es sólo FIR, es decir que $M = 0$. Los N coeficientes del predictor pueden ser fijos ([Robinson, 1994](#); [Coalson, 2009](#)) o dependientes de los datos (codificación lineal predictiva, en adelante LPC). En el primer caso, los coeficientes se derivan de la interpolación de polinomios de Lagrange ([Salomon, 2007](#)), mientras que los coeficientes ajustados por la entrada se calculan por medio de algoritmos de minimización de la varianza del error, como por ej. Levinson-Durbin ([Coalson, 2009](#); [Liebchen, 2011](#)).

Algunos CSP utilizan predictores de tipo IIR ([Liebchen, 2011](#); [Craven y Gerzon, 1996](#); [Craven et al. 1997](#); [Djuric, 2010](#)) que se caracterizan por permitir leves mejoras de compresión a costa de una mayor carga computacional ([Hans y Schafer, 2001](#)). Otros CSP utilizan predictores basados en técnicas diferentes como redes neuronales ([Ashland, 2011](#); [Salomon, 2007](#)).

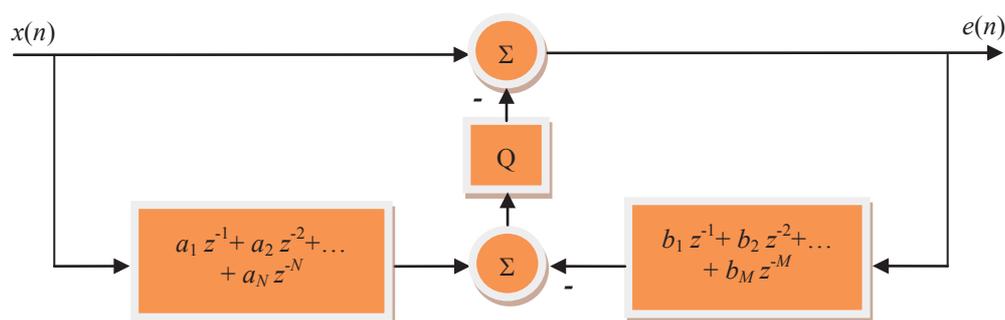


Figura 2: Diagrama de bloques de la sección de decorrelación. Q es el bloque cuantizador.

En el receptor, el decodificador reconstruye la señal original de la siguiente manera. Los datos entrantes se demultiplexan, luego se extrae y decodifica el residuo, y se lo añade a los valores estimados por el predictor. Éste utiliza los mismos coeficientes que en la estructura de codificación (ver Figura 3).

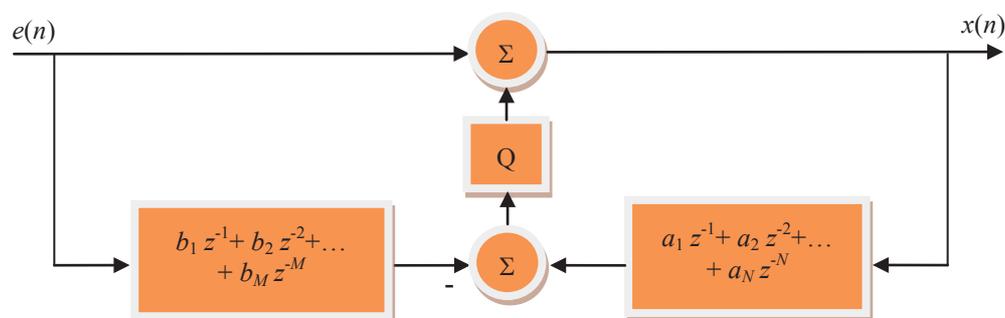


Figura 3: Estructura de la reconstrucción de la señal de entrada en los CSP.

Algunos importantes CSP como Philips (Bruekers et al. 1996) y DVD (Craven y Gerzon, 1996; DVD Forum, 2004) no se evalúan en este documento ya que los archivos ejecutables correspondientes no se encuentran disponibles para nosotros. El primero se basa tanto en un predictor FIR de orden diez como en el codificador Rice para representar el residuo, mientras que el segundo utiliza predicción tipo IIR y codificación Huffman para el residuo.

La Tabla 1 muestra los CSP analizados en este trabajo y los correspondientes comandos DOS para la codificación y decodificación utilizados en nuestras pruebas. Para cada codificador, sólo se consideraron tres modos de codificación: normal (por defecto), rápido y máximo. En estos dos últimos modos se maximiza la velocidad y la compresión, respectivamente. El análisis propuesto se realizó con los archivos ejecutables proporcionados por los desarrolladores originales.

3 ANÁLISIS DEL DESEMPEÑO

Los CSP mencionados en la Tabla 1 se evaluaron con piezas extraídas de discos compactos comerciales (Rodríguez Guerrero et al. 2011). Se seleccionaron más de 15 piezas para cada uno de los siguientes géneros de música: folklore argentino, tango, jazz, música clásica, pop y rock. Además, se incluyó en el análisis una docena de grabaciones de rock en vivo debido a su popularidad entre los amantes de la música. En cada género se eligió una cantidad de piezas suficientemente grande como para contemplar una gran variedad de ritmos y artistas. Se seleccionaron como máximo 20 archivos de audio para cada género, ya que la inclusión de más piezas no mejora significativamente los resultados expuestos en la sección 4 (es decir, la media del cambio de las variables analizadas es menos del 8 %).

Los archivos de audio se codificaron en las computadoras descriptas en la [Tabla 2](#). El desempeño de los codificadores se cuantificó en lo que respecta a los siguientes parámetros:

- *factor de compresión (FC)*, definido como el cociente entre el tamaño del archivo de entrada y del archivo codificado.

- *rate (R)*, definido por el cociente entre el tiempo de reproducción del archivo original y el tiempo de codificación o decodificación según corresponda. Es una medida de la velocidad de procesamiento del codificador o decodificador.

Codec	Modo o Configuración		Comando DOS
FLAC v.1.2.1 (Coalson, 2009)	Decodificación		-d -f <i>infile</i> -o <i>outfile</i>
	Codificación	Rápida	-0 -f --no-padding <i>infile</i> -o <i>outfile</i>
		Normal	-5 -f --no-padding <i>infile</i> -o <i>outfile</i>
		Máxima	-8 -f --no-padding <i>infile</i> -o <i>outfile</i>
LPAC v.3.08 (Liebchen, 2010)	Decodificación		-x -v <i>infile</i> <i>outfile</i>
	Codificación	Rápida	-v -1 <i>infile</i> <i>outfile</i>
		Normal	-v -3 <i>infile</i> <i>outfile</i>
		Máxima	-v -5 <i>infile</i> <i>outfile</i>
Monkey's Audio v.4.06 (Ashland, 2011)	Decodificación		<i>infile</i> <i>outfile</i> -d
	Codificación	Rápida	<i>infile</i> <i>outfile</i> -c1000
		Normal	<i>infile</i> <i>outfile</i> -c2000
		Máxima	<i>infile</i> <i>outfile</i> -c5000
MPEG-4 ALS RM22rev2 (Liebchen, 2011)	Decodificación		-x -v <i>infile</i> <i>outfile</i>
	Codificación	Normal	-v <i>infile</i> <i>outfile</i>
		Máxima	-7 -v <i>infile</i> <i>outfile</i>
OptimFrog v.4.600ex (Ghido, 2011)	Decodificación		--decode --verbose --time --overwrite <i>infile</i> --output <i>outfile</i>
	Codificación	Rápida	--verbose --time --mode fast --overwrite <i>infile</i> --output <i>outfile</i>
		Normal	--verbose --time --mode normal --overwrite <i>infile</i> --output <i>outfile</i>
		Máxima	--verbose --time --maximumcompression --overwrite <i>infile</i> --output <i>outfile</i>
Shorten v.3.6.1 (Robinson, 1994)	Decodificación		-x <i>infile</i> <i>outfile</i>
	Codificación	Normal	<i>infile</i> <i>outfile</i>
		Máxima	-p12 -b1024 <i>infile</i> <i>outfile</i>
TAK v.2.0.0 (Becker, 2011)	Decodificación		-d -overwrite <i>infile</i> <i>outfile</i>
	Codificación	Rápida	-e -p0 -overwrite <i>infile</i> <i>outfile</i>
		Normal	-e -p2 -overwrite <i>infile</i> <i>outfile</i>
		Máxima	-e -p4 -overwrite <i>infile</i> <i>outfile</i>
TTA v.3.4.1 (Djuric, 2010)	Decodificación		-d <i>infile</i> -o <i>outfile</i>
	Codificación		-e <i>infile</i> -o <i>outfile</i>
WavPack ^a v.4.60.1 (Bryant, 2009)	Decodificación		-y <i>infile</i> <i>outfile</i>
	Codificación	Rápida	-f -y <i>infile</i> <i>outfile</i>
		Normal	-y <i>infile</i> <i>outfile</i>
		Máxima	-hx6 -y <i>infile</i> <i>outfile</i>

^aWavPack posee dos archivos ejecutables, uno para la codificación —wavpack.exe— y otro para la decodificación —wvunpack.exe—. Ambos pertenecen a la misma versión (4.60.1).

Tabla 1: Lista de codecs y configuraciones utilizadas. *infile* y *outfile* representan los nombres de los archivos de entrada y salida respectivamente, incluyendo sus rutas correspondientes.

Computadora	Placa madre / μ P	RAM / Disco r�gido	Sistema Operativo
Port�til (PC ₀)	HP Pavilion DV7 -3065DX / 2.4 GHz AMD Turion II Ultra Dual-Core Mobile M600	2 × 2 GB SDRAM DDR2-800 / Samsung HM321HI 5400 rpm SATA (320 GB)	Windows Vista Home Premium
PC de escritorio (PC ₁)	Intel Rock Lake D865PERL / 3 GHz Pentium IV HT	2 × 512 MB SDRAM DDR / Maxtor 6Y080L0, Ultra- ATA133 (80 GB)	Windows XP Service Pack 3
PC de escritorio (PC ₂)	MSI K9N6PGM-F/FI / 2.2 GHz AMD Athlon 64 × 2 Dual Core	2 GB SDRAM DDR2 / WDC WD1600AABS SCSI (149 GB)	Windows XP Service Pack 3

Tabla 2: Computadoras utilizadas para el an lisis. PC₀ se adopta como referencia para los resultados num ricos referidos a una sola computadora.

En este an lisis se promedi  tanto el FC como el R para todos los archivos de audio en cada g nero. Esto permite determinar el rendimiento medio de cada codificador. Adem s, las piezas previamente codificadas fueron decodificadas con el fin de determinar el rate de decodificaci n y tambi n para comprobar si la se al de entrada logra ser reconstruida sin p rdidas.

A pesar de que la codificaci n se puede realizar a trav s de interfaces gr ficas (Pawlowski, 2010), hemos preferido dise ar una interfaz propia que se adapte mejor a nuestras necesidades experimentales (Rodr guez Guerrero, 2011). Entre otras caracter sticas, nuestra interfaz presenta un tiempo de retardo menor que (Pawlowski, 2010) (generalmente 0,38 s en la computadora PC₀) y permite procesar una colecci n de archivos con tan s lo seleccionar la carpeta correspondiente. Esta herramienta tambi n permite al operador pausar el procesamiento con el fin de controlar los errores y muestra los resultados en formato de tabla.

4 RESULTADOS COMPARATIVOS

4.1 Codificaci n en modo normal

Se codific  en modo normal las piezas seleccionadas en la computadora PC₀. El porcentaje de uso de la CPU para cualquiera de los CSP fue entre 24 % y 28 %. El factor de compresi n alcanzado por los g neros analizados se muestra en la Figura 4 (se ordenan los g neros en orden descendente de FC medio). Es evidente que la compresi n es m s sensible al g nero que al codificador, y tambi n que la compresi n m s baja corresponde en todos los casos a Shorten.  ste es un resultado esperado ya que tanto el orden como los coeficientes del predictor de Shorten son fijos en lugar de ser ajustados a la entrada. Adem s, el FC se promedi  para todos los codificadores evaluados en cada g nero, obteni ndose los valores expuestos en la Tabla 3.

El desempe o de cada codificador en el dominio velocidad-compresi n (R - FC) se ilustra en la Figura 5. Cada CSP se identifica con un color diferente (regi n de operaci n), y cada g nero est  representado con un marcador distinto. Se puede observar que la velocidad de codificaci n es m s dependiente del codificador que del g nero, especialmente para velocidades menores (menor rate). Para mayor rate, la velocidad de codificaci n se ve muy afectada por el g nero. En estos casos, la codificaci n es m s r pida para tango y m s lenta para m sica cl sica, rock y grabaciones de rock en vivo y pop. M s importante a n, la Figura

5 muestra que TAK, TTA, WavPack, Shorten y FLAC son los codificadores más rápidos y OptimFrog es el más lento.

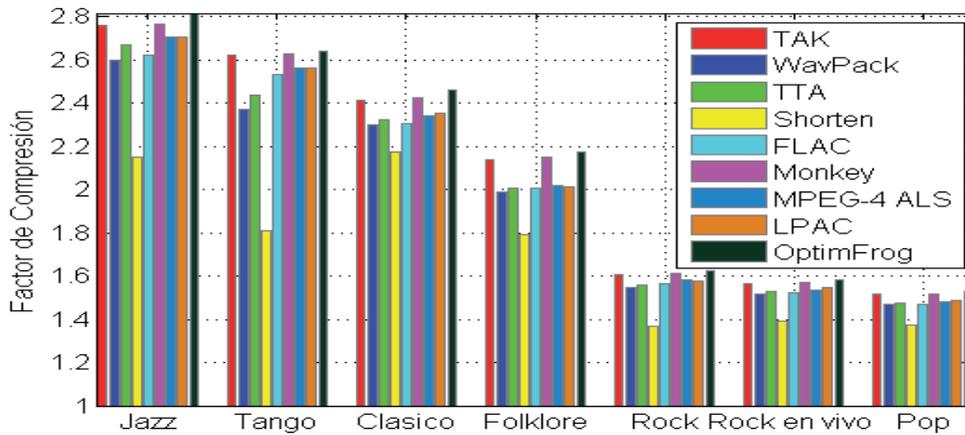


Figura 4: Factor de compresión asociado a cada género musical analizado con los CSP etiquetados en color (codificación en modo normal).

Género	Jazz	Tango	Clásico	Folklore	Rock	Rock en vivo	Pop
FC promedio	2.64	2.46	2.34	2.03	1.57	1.53	1.48

Tabla 3: Factor de compresión promedio de cada género musical.

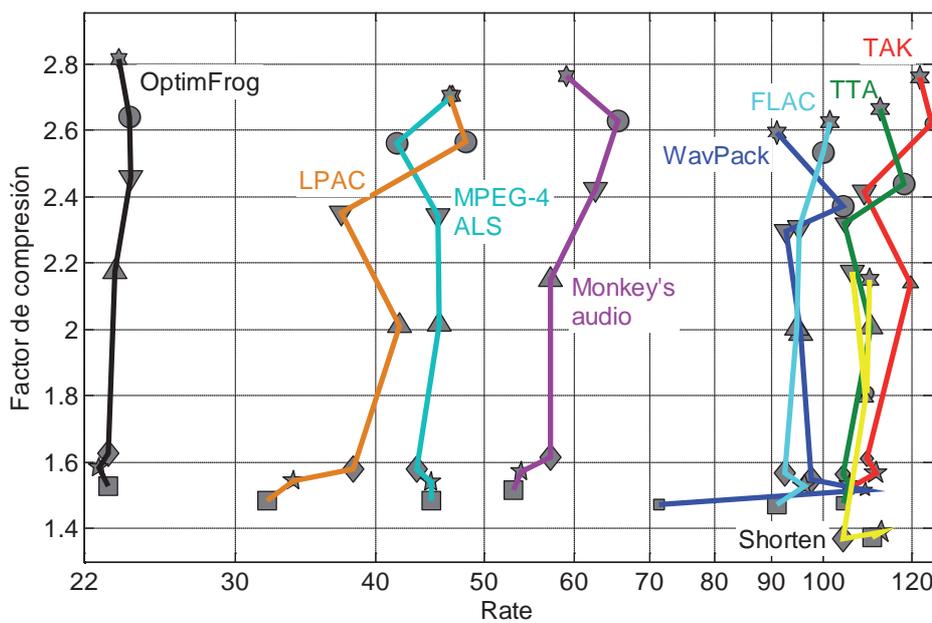


Figura 5: Representación gráfica del rate versus el factor de compresión obtenido para cada género musical con los CSP mencionados en la Tabla 1 en modo normal. Los géneros analizados son jazz (★), tango (●), música clásica (▼), folklore (▲), rock (◆), rock en vivo (★) y pop (■).

4.2 Codificación en modo rápido y modo máximo

En comparación con el modo de codificación normal, el modo rápido de codificación en la computadora portátil aumenta la velocidad de procesamiento en un 45 % (para OptimFrog, Monkey's Audio y WavPack) o incluso más del 45 % (para los codificadores restantes). Sin embargo, el factor de compresión se reduce algo menos de un 10 %. Esto es válido para el conjunto completo de CSP y se representa en la [Figura 6](#). También se puede observar que la velocidad de codificación es más sensible a los codificadores, aunque esto es válido sólo para los CSP más lentos (es decir, para codificadores con rates inferiores a 92). Finalmente, el tiempo de codificación en modo rápido con cualquier codificador (a excepción de OptimFrog) es menor que 1/70 del tiempo de reproducción, por lo tanto muy pequeño.

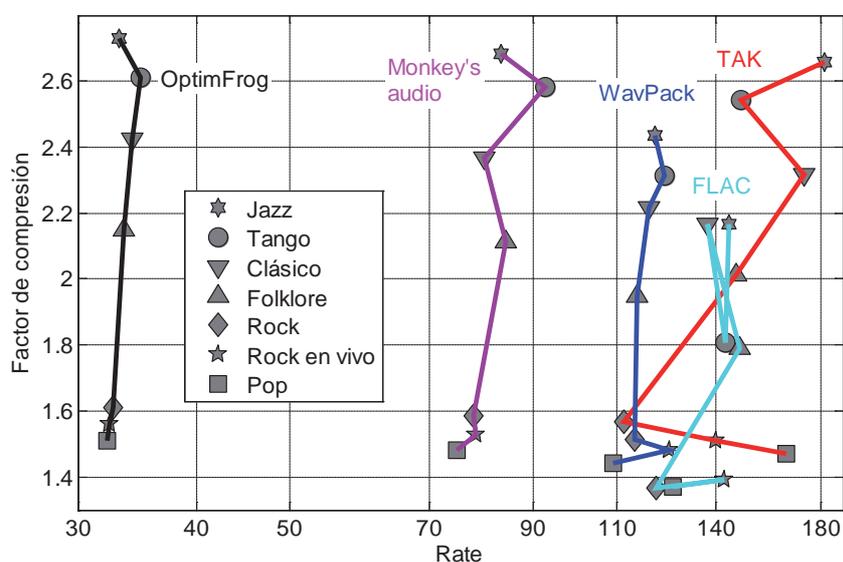


Figura 6: Representación gráfica del rate versus el factor de compresión obtenido para cada género musical con los CSP mencionados en la [Tabla 1](#) en modo rápido.

En comparación con el modo de codificación normal, el modo máximo aumenta la compresión entre un 5 % y 8 % (para todos los codecs), a expensas de incrementar el tiempo de procesamiento, especialmente para los codificadores más lentos (ver [Figura 7](#)). A modo de ejemplo, el tiempo de codificación de OptimFrog puede aumentar de 1/24 (modo normal) al doble del tiempo de reproducción (modo máximo). Además, cada uno de los codificadores analizados puede asociarse con un pequeño intervalo de rates, ya que el tiempo de codificación para el modo de compresión máximo es muy dependiente del codificador, incluso para los CSP más rápidos.

4.3 Decodificación

Se realizó una prueba de precisión en las tres computadoras para garantizar la perfecta reconstrucción de las piezas originales de los archivos codificados, y ninguno de los archivos evaluados sufrió pérdida de información. Más interesante aún, se observaron algunas analogías entre las etapas de codificación y decodificación. Por ejemplo, la velocidad de procesamiento es mucho más sensible al decodificador que al género musical. Dicha velocidad es mayor (menor) si el archivo codificado se generó en el modo de codificación rápido (máximo). En consecuencia, teniendo en cuenta la codificación y decodificación, el

modo de codificación rápido (máximo) es mucho más conveniente si se desea minimizar el tiempo de procesamiento (espacio de almacenamiento).

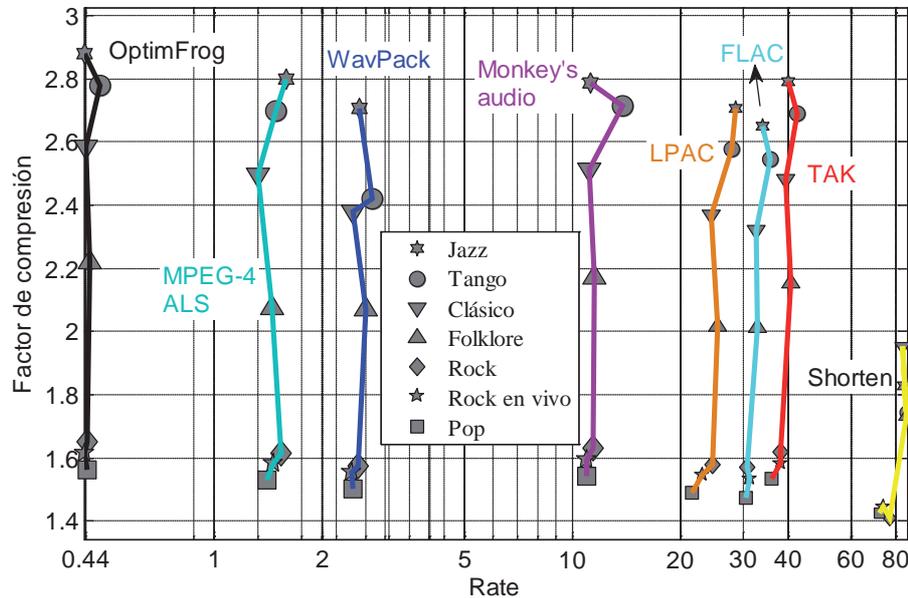


Figura 7: Representación gráfica del rate y factor de compresión obtenidos para cada género musical con los CSP mencionados en la Tabla 1 en modo máximo.

Como era de esperar, la decodificación requiere menos tiempo que la codificación, especialmente para los modos de codificación normal y máximo. Esta asimetría no es válida para Shorten ni Monkey's Audio. La razón de este último caso es el uso de un algoritmo simétrico para la codificación y decodificación. Por lo tanto, la diferencia entre el tiempo de codificación y el tiempo de decodificación viene dada por las diferencias entre los tiempos de transferencia de datos en las instancias de codificación y decodificación (en este último caso, la cantidad de datos escritos en el disco es FC veces la cantidad de datos escritos durante la codificación). En cuanto al modo rápido de codificación, la decodificación no siempre fue más rápida que la codificación, pero el rate fue siempre superior a 30. Para ilustrar estos hechos, la Tabla 4 muestra los tiempos de codificación / decodificación para una sola pieza codificada en los modos rápido y máximo (sólo en los codecs donde fue posible). Como puede observarse, TAK, FLAC y WavPack son los codificadores más rápidos. Esto los hace útiles para la reproducción de archivos codificados en dispositivos de audio de bajo costo o portátiles.

4.4 Codificación en diferentes computadoras

La evaluación de los codecs en las computadoras PC_1 y PC_2 (descritos en la Tabla 2) confirma que la compresión es independiente de la computadora para todos los archivos analizados (siempre que se utilicen las mismas condiciones de compilación). En esas computadoras también se observó que la región de operación de cada codec en los modos rápido, normal y máximo son muy similares a las obtenidas en la computadora portátil. Para que esto sea cierto, se debe escalar el rate y asegurar que éste se mantenga por debajo de un cierto límite.

Codec	Modo máximo	Modo rápido
TAK	40 / 48	181 / 59
WavPack	3 / 51	122 / 58
TTA	119 / 48 (único modo)	
Shorten	82 / 56	115 / 25 (modo normal)
FLAC	32 / 41	124 / 56
Monkey's Audio	11 / 4	79 / 32
MPEG-4 ALS	1 / 9	46 / 35 (modo normal)
LPAC	24 / 32	87 / 35
OptimFrog	0.44 / 1	33 / 30

Tabla 4: Tiempos de codificación y decodificación en PC₀ de la pieza “Si te vas”, interpretada por A. Zitarrosa.

En consecuencia, para la mayoría de los casos estudiados, cada codificador y cada modo se puede asociar a una región operativa, cuyo rate depende de la computadora a través de un factor de escala, siempre y cuando no exceda un cierto límite. Esta propiedad se explota en la siguiente sección.

5 MÉTODO DE FRONTERA

Como se vio previamente, hay demasiados resultados a analizar y es difícil determinar el codec óptimo en términos de velocidad y compresión de una dada colección musical. A fin de superar esta dificultad, se propone un método nuevo y simple orientado al usuario. Este método es gráfico y tiene en cuenta las combinaciones analizadas de codecs y modos, los que llamaremos C-M en adelante.

La idea principal es reducir gráficamente el conjunto completo de los C-M analizados a un pequeño conjunto de C-M óptimos. Esto se realiza descartando los C-M asociados a la menor compresión y menor velocidad. Como consecuencia, la secuencia resultante de C-M óptimos es monótona decreciente en el dominio $R-FC$. Para ilustrar esto, la Figura 8 muestra dos situaciones hipotéticas con dos C-M distintos etiquetados como CM₁ y CM₂. En el panel (a), el CM₁ comprime más que el CM₂, por lo tanto es óptimo en compresión, mientras que el CM₂ es óptimo en cuanto a velocidad. El panel (b) denota otra situación en donde el CM₁ es descartado debido a una menor compresión y una menor velocidad respecto al CM₂. Los puntos que “sobreviven” a este proceso de eliminación son considerados óptimos y constituyen la *frontera* del mejor desempeño que puede obtenerse con los C-M analizados.

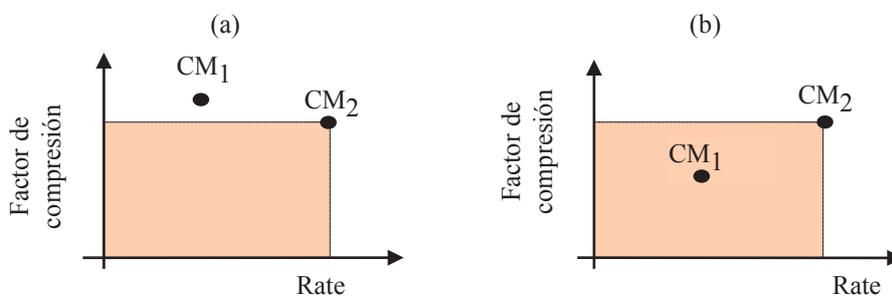


Figura 8: Eliminación de C-M para dos casos hipotéticos: (a) CM₁ y CM₂ son óptimos para compresión y velocidad respectivamente. (b) Se descarta el CM₁ ya que es más lento y comprime menos que CM₂.

La selección de los puntos óptimos $\{R, FC\}$ de un conjunto dado de C-M puede hacerse

como se muestra en el diagrama de flujo de la **Figura 9**. Posteriormente, el conjunto de datos óptimos se representa gráficamente en el dominio R - FC .

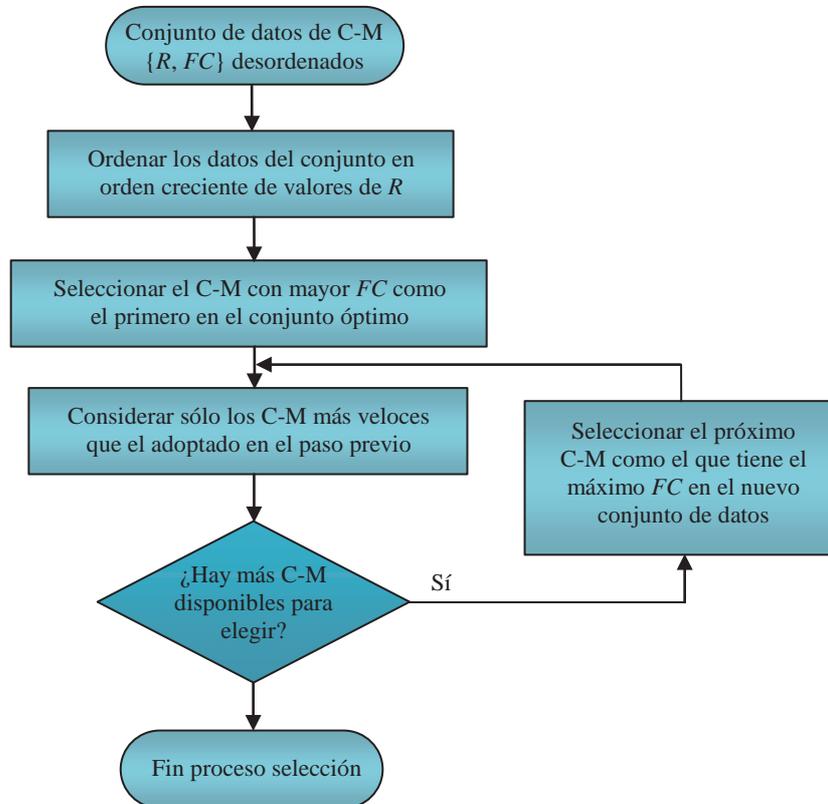


Figura 9: Diagrama de flujo para la selección de C-M óptimos.

5.1 Aplicación para una colección de archivos

Para cada uno de los géneros musicales analizados, la entrada del algoritmo de selección de los C-M óptimos debe consistir en un conjunto de valores $\{R_{col}, FC_{col}\}$ que representen con precisión la velocidad y la compresión de cada C-M en la colección completa del género musical. Según el procedimiento expuesto en la sección 3, estos valores serían el promedio del rate y factor de compresión de todos los archivos en la colección del género. Sin embargo, el usuario puede estar más interesado en el desempeño de la codificación de una colección completa en lugar de la compresión de un archivo promedio por género. Por esta razón, se propone determinar el rate y el factor de compresión de la colección completa para cada C-M como:

$$R_{col} = \frac{\sum_i T_{p_i}}{\sum_j T_{e_j}}, \quad (1)$$

$$FC_{col} = \frac{\sum_i S_{u_i}}{\sum_j S_{e_j}}, \quad (2)$$

donde T_p y T_e representan el tiempo de reproducción y de codificación de cada pieza en la

colección del género musical, y S_u y S_e representan el tamaño original del archivo y el tamaño codificado del mismo, respectivamente.

A continuación se prosigue con un ejemplo. Se determinaron los C-M óptimos para la colección de jazz evaluada en la computadora portátil en la sección 4. Para cada C-M, se computaron las Ecuaciones (1) y (2) y se seleccionaron los C-M óptimos de acuerdo a la Figura 9. La Figura 10 muestra los resultados del proceso, donde los puntos óptimos se representan unidos con línea de trazos intermitentes.

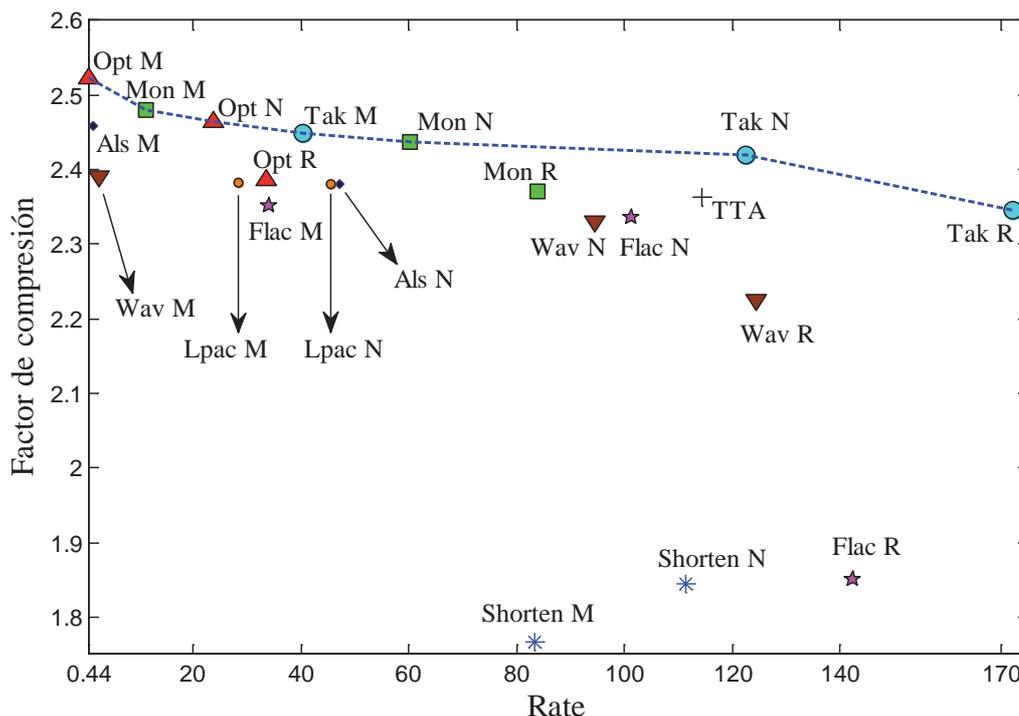


Figura 10: Rate y factor de compresión de todos los C-M para la colección de jazz analizada en la sección 4. Los C-M óptimos que determinan la frontera se unen con línea de trazos intermitentes.

La aplicación de este procedimiento para todas las colecciones de géneros musicales analizadas en la sección 4 en las tres computadoras mencionadas en la Tabla 2 provee los C-M óptimos mencionados en la Tabla 5 y representados gráficamente en la Figura 11.

Como se indica en la Tabla 5, la mayoría de los C-M óptimos son los mismos, independientemente de la computadora. Otro hecho importante es que algunos C-M son óptimos en todas las colecciones como Opt M, TAK M y TAK N o en casi todas como Mon M y Mon N. Esto implica que estos C-M siempre o casi siempre óptimos, más allá del género musical analizado.

En cuanto a la Figura 11, se observó lo siguiente. En primer lugar, la compresión es más sensible al género de la colección que a los C-M óptimos seleccionados (a excepción de los más rápidos). En segundo lugar, la velocidad de procesamiento es mucho más dependiente de los C-M óptimos que del género (por ejemplo, el rate de Opt M varía menos de 3 % entre un género y otro). Estos hechos confirman los resultados presentados en la sección 4. En tercer lugar, todas las fronteras son similares entre sí.

A pesar de poder analizar los C-M óptimos para todos los géneros al mismo tiempo, sólo estamos interesados en los resultados correspondientes a cada género de forma individual, ya

que los usuarios de los CSP suelen consumir música de algunos géneros específicos (como jazz o música clásica).

Género	PC	CM ₁	CM ₂	CM ₃	CM ₄	CM ₅	CM ₆	CM ₇	CM ₈
Jazz	PC ₀ , PC ₁	Opt M	Mon M	Opt N	TAK M	Mon N	TAK N	TAK R	
	PC ₂	Opt M	Mon M	Opt N	TAK M	Mon N	TAK N		
Tango	PC ₀	Opt M	Mon M	TAK M	Mon N	TAK N	TAK R	FLA R	
	PC ₁	Opt M	Mon M	TAK M	Mon N	TAK N	TAK R		
	PC ₂	Opt M	Mon M	TAK M	Mon N	TAK N			
Clásico	Todas	Opt M	Mon M	TAK M	Mon N	TAK N	TAK F		
Folklore	PC ₀	Opt M	Opt N	TAK M	Mon N	TAK N	TAK R	FLA R	
	PC ₁	Opt M	Opt N	TAK M	Mon N	TAK N	TAK R		
	PC ₂	Opt M	Opt N	TAK M	Mon N	TAK N			
Rock	PC ₀	Opt M	Mon M	Opt N	TAK M	Mon N	TAK N	Wav R	
	PC ₁	Opt M	Mon M	Opt N	TAK M	Mon N	TAK N	TAK R	
	PC ₂	Opt M	Mon M	Opt N	TAK M	Mon N	TAK N		
Rock en vivo	PC ₀	Opt M	Mon M	Opt N	TAK M	Mon N	TAK N	TAK R	FLA R
	PC ₁	Opt M	Mon M	Opt N	TAK M	Mon N	TAK N	TAK R	
	PC ₂	Opt M	Mon M	Opt N	TAK M	Mon N	TAK N		
Pop	PC ₀ , PC ₁	Opt M	Mon M	TAK M	TAK N	TAK R			
	PC ₂	Opt M	Mon M	TAK M	TAK N	Sho N			

Tabla 5: C-M óptimos para cada género en las tres computadoras mencionadas en la [Tabla 2](#). Estos C-M se etiquetan en orden ascendente de rate y se grafican en la [Figura 11](#). Los C-M resaltados en negrita son óptimos en las tres computadoras simultáneamente.

5.2 Normalización

La velocidad de procesamiento de cada uno de los C-M está asociada con el hardware en el que estos C-M se ejecutan. Por lo tanto es esperable que la relación de velocidad entre dos C-M diferentes sea aproximadamente la misma, independientemente del equipo en el que se ejecuten. Es por ello que se propone obtener información que sea independiente del hardware a través de un proceso de normalización, el cual consiste en dividir los valores del rate de cada C-M por el rate correspondiente a un solo C-M que se adopte como referencia. Dicha normalización permite medir la velocidad relativa de cada uno de los C-M con respecto a la referencia.

Las fronteras normalizadas correspondientes a las colecciones de géneros estudiadas en la sección 4 se muestran en la [Figura 12](#). Estos resultados corresponden a la computadora PC₀ y se adoptó Opt M como referencia, ya que es óptimo para todos los géneros y su rate se ve menos afectado por el género seleccionado. Se observa que para cada C-M la discrepancia en el rate es generalmente pequeña o moderada (excepto para el C-M más rápido). La [Figura 12](#) también muestra los valores normalizados de compresión con respecto a Opt M. Esta normalización reduce los efectos de la dispersión que se producen al comprimir diferentes conjuntos de archivos dentro de la colección musical de un género dado. Esta propiedad será

utilizada para fines de predicción, como se explicará en la siguiente sección.

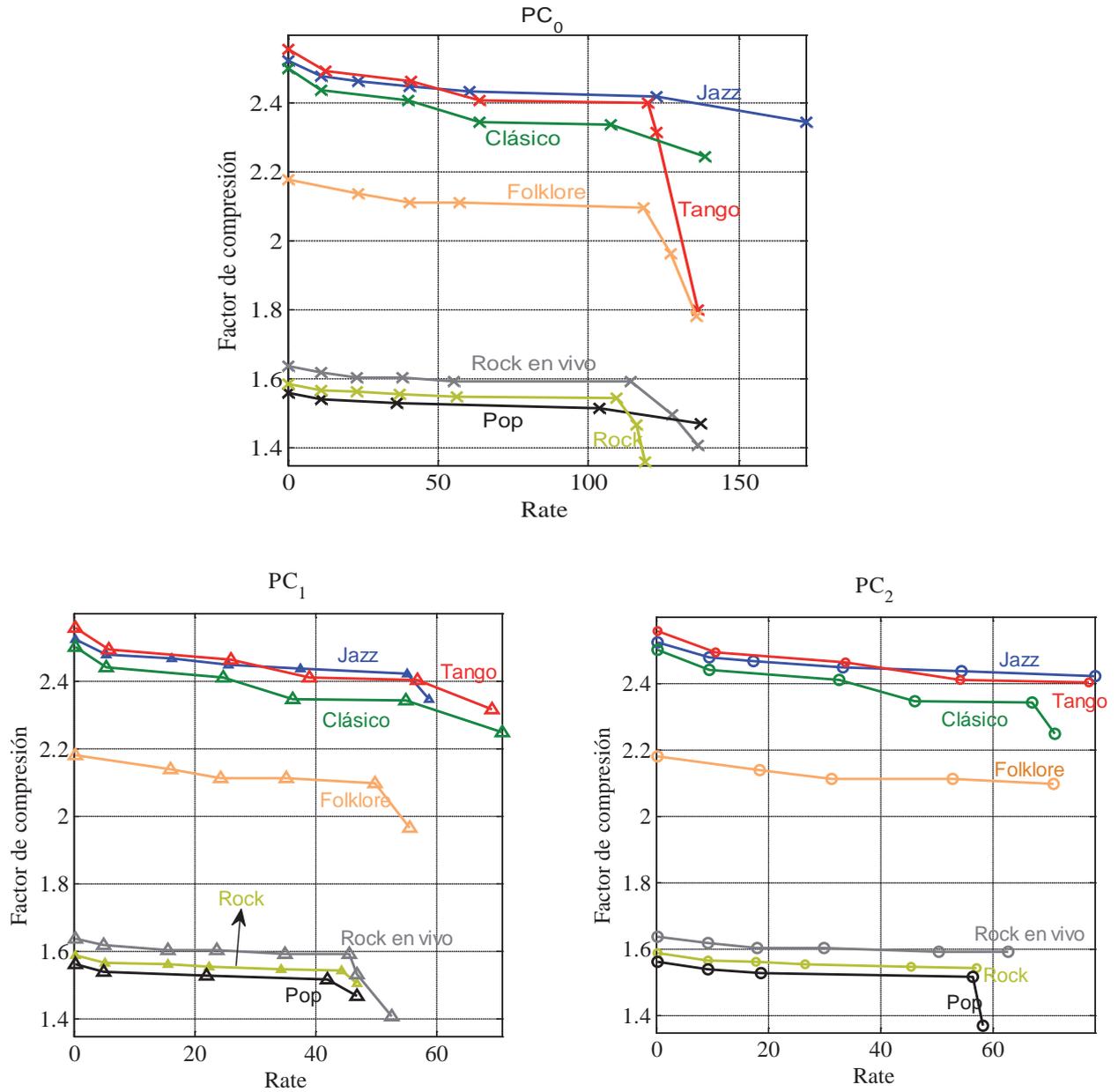


Figura 11: C-M óptimos para cada género musical en las computadoras descritas en la [Tabla 2](#). Los C-M de cada frontera se listan en la [Tabla 5](#).

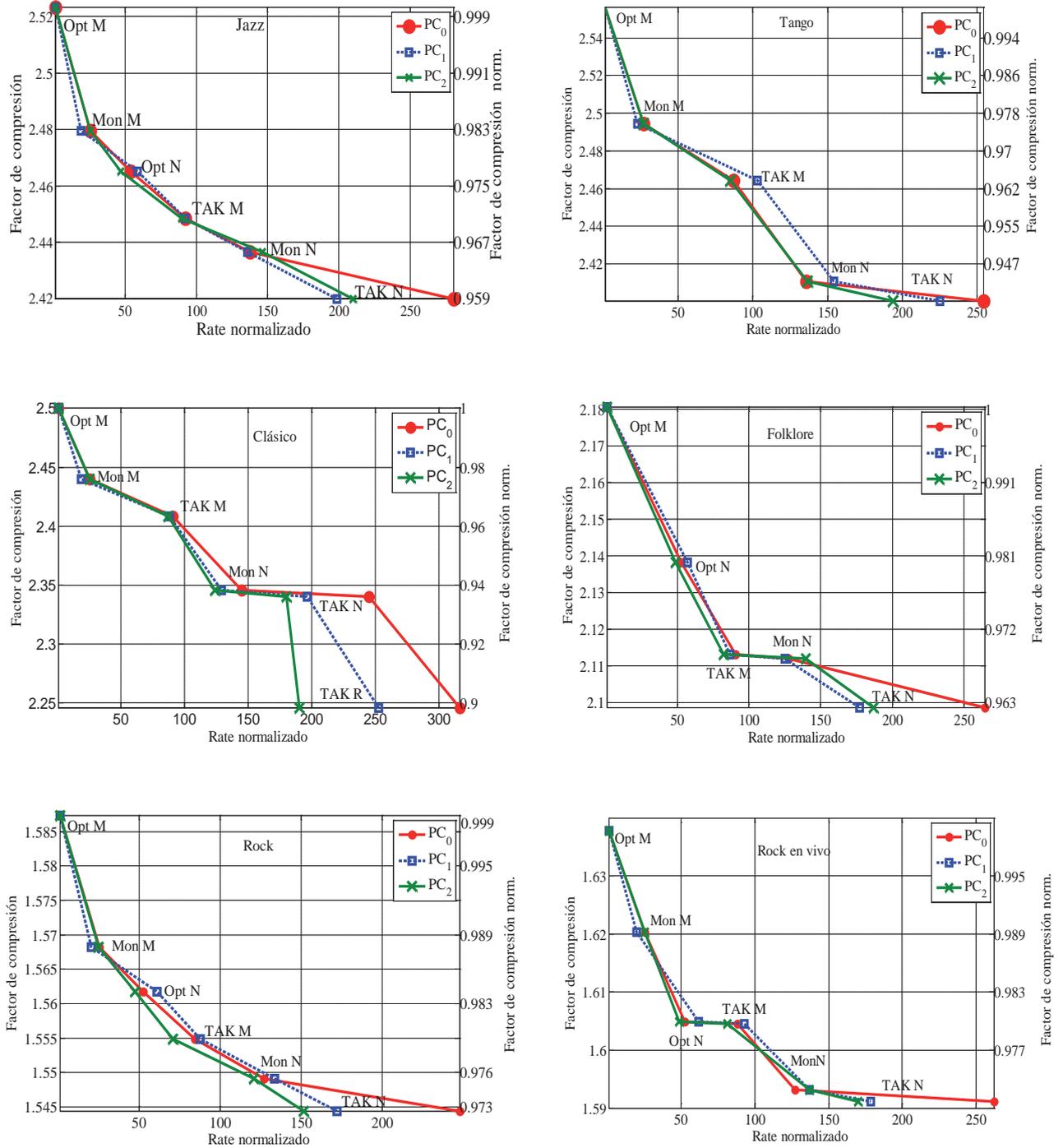


Figura 12: Fronteras normalizadas para cada género musical analizado.

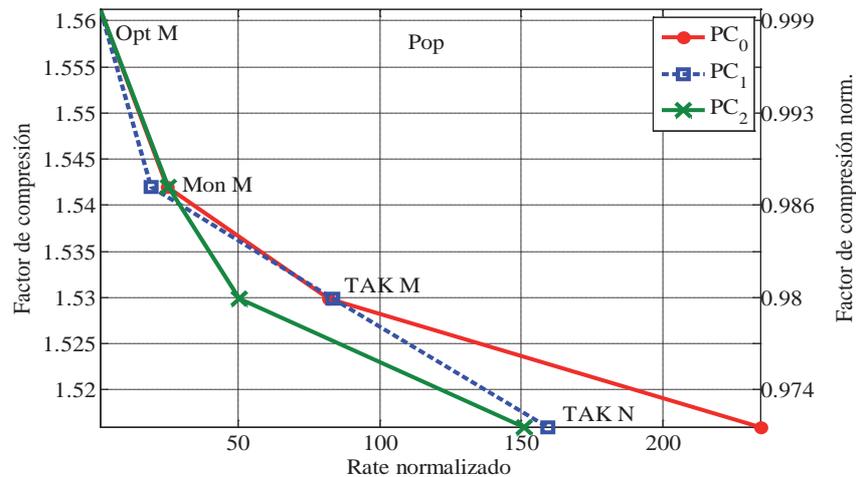


Figura 12: Fronteras normalizadas para cada género musical analizado (cont.).

6 HERRAMIENTA DE PREDICCIÓN

Si el usuario posee una colección musical de un género determinado y la desea codificar óptimamente, podrá elegir cualquiera de los C-M óptimos que aparecen resaltados con negrita en la fila correspondiente de la [Tabla 5](#). Si se necesita conocer *a priori* el tiempo de codificación y el tamaño de la colección de salida con cualquiera de los C-M óptimos, es necesario disponer de una herramienta cuantitativa. La misma es provista por las fronteras que se muestran en la [Figura 12](#), seguida de la denormalización de los resultados $\{R_{n\ col}, FC_{n\ col}\}$ a través de los parámetros de desempeño de la colección del género asociados con Opt M $\{R_{col\ Opt\ M}, FC_{col\ Opt\ M}\}$. Para este último paso, los parámetros de desempeño de los restantes C-M en la frontera (etiquetados como C-M x) se determinan como:

$$R_{col\ CMx} = R_{n\ col\ CMx} \times R_{col\ Opt\ M}, \quad (3)$$

$$FC_{col\ CMx} = FC_{n\ col\ CMx} \times FC_{col\ Opt\ M}. \quad (4)$$

Para estos C-M, el tiempo de codificación de la colección del género $T_{e\ col}$ y su tamaño codificado $S_{e\ col}$ se determinan como:

$$T_{e\ col\ CMx} = \frac{T_{p\ col}}{R_{col\ CMx}}, \quad (5)$$

$$S_{e\ col\ CMx} = \frac{S_{u\ col}}{FC_{col\ CMx}}, \quad (6)$$

donde $T_{p\ col}$ es el tiempo de reproducción de la colección del género y $S_{u\ col}$ es el tamaño de dicha colección sin comprimir. (Si el tiempo de reproducción de cada pieza es desconocido, el mismo se puede estimar multiplicando el tamaño del archivo de entrada (en MB) por $2^{20} / (4 \times 44\ 100) \approx 5,94$. Esto es válido para cualquier archivo PCM con pocos metadatos, tal como las piezas provenientes de CD de audio).

Se necesita aproximar los parámetros de codificación de la colección completa del género

con Opt M $\{R_{col\ Opt\ M}, FC_{col\ Opt\ M}\}$. Para ello, se propone codificar solamente unos pocos archivos de la colección seleccionados al azar. Nuestras pruebas numéricas sugieren que el rate de la colección de un género dado es próximo al rate de cualquiera de sus piezas. Este error se limita al 6 %, a excepción de unas pocas piezas, donde el error alcanza el 25 %. Por lo tanto, el rate de la colección con Opt M se puede aproximar como la mediana de los rates de pocos archivos codificados.

El factor de compresión de las diferentes piezas (con cualquier C-M incluyendo Opt M) se ve más afectado por la dispersión que el rate. Por lo tanto, para la determinación de la compresión de la colección se sugiere la aplicación de la Ec. (2) utilizando la información de los archivos procesados. Mientras más archivos se seleccionen para realizar este proceso, mejor será la aproximación, pero también aumentará el tiempo de procesamiento. Teniendo en cuenta este compromiso, se decidió seleccionar sólo unos pocos archivos para nuestro método, aunque la inclusión de más archivos permite una predicción más robusta.

Es importante señalar que si se seleccionan piezas largas (con respecto al tiempo medio de reproducción en la colección), el tiempo de procesamiento aumenta considerablemente. Por lo tanto, conviene que las piezas elegidas arbitrariamente no sean demasiado extensas.

Finalmente, se podría normalizar la frontera con cualquiera de las computadoras mencionadas en la [Tabla 2](#), pero normalizamos sólo con la PC₀ ya que tiene mayor poder de cómputo y por lo tanto se asemeja más a cualquier computadora nueva en manos del audiófilo, a quien va destinada esta herramienta.

6.1 Ejemplo numérico

A fin de ilustrar la herramienta de predicción, a continuación se ilustra un ejemplo de su aplicación. Se evaluaron dos colecciones de género en las computadoras PC₂ (mencionada en la [Tabla 2](#)) y PC₃, esta última con Windows XP Service Pack 2 y equipada con:

- Placa madre PCChips M909.
- Procesador Intel Pentium 4 de 2400 MHz.
- Memoria RAM PC2700 DDR SDRAM de 512 MB.
- Disco duro HDS728080PLAT20 de 80 GB, 7200 RPM, Ultra-ATA/133.

Las colecciones de prueba consistieron en ([Rodríguez Guerrero et al. 2011](#)): a) 20 piezas de música clásica (1147,4 MB de tamaño y 1 h 48 min 19 s de tiempo de reproducción) y b) 11 piezas de jazz (544 MB de tamaño y 51 min 19 s de tiempo de reproducción).

Las fronteras normalizadas para clásico y jazz mostradas en la [Figura 12](#) son denormalizadas de acuerdo a las Ecuaciones (3) y (4). Para cada colección, se calculó el rate y la compresión correspondiente a Opt M luego de procesar tres piezas seleccionadas al azar (piezas 5, 12, 18 para clásico y 2, 7, 11 para jazz), seguido de la determinación de la media del rate y de la compresión de la colección de acuerdo a la Ec. (2). El tiempo de codificación y el tamaño de la colección codificada se aproximan a través de las Ecuaciones (5) y (6) y se etiquetan como "Est 1" en la [Figura 13](#). Estos valores se acercan más a los verdaderos que las siguientes estimaciones marcadas en la [Figura 13](#): Est 2) promedio del rate y de la compresión, y Est 3) rate aproximado con la Ec. (1) y compresión aproximada como su media. Como se muestra en la [Figura 13](#), el error de predicción del tamaño de la colección comprimida es el mismo independientemente de la computadora. Esto es esperable pues la compresión no se ve afectada por la computadora utilizada. Además, estos errores son menores al 10 %. Se corroboró numéricamente que este límite es válido para el 65 % de todas las combinaciones posibles de tríadas de piezas.

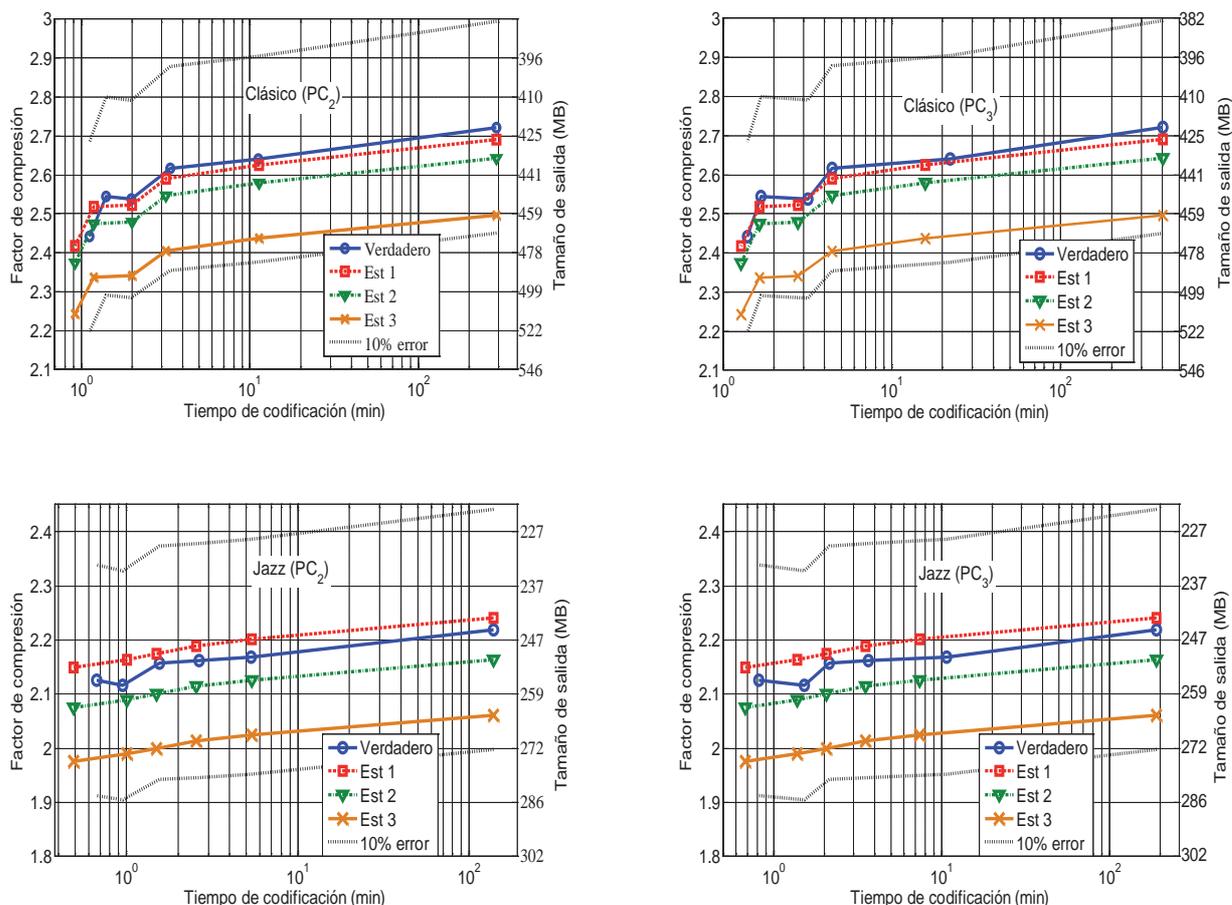


Figura 13: Resultados de la predicción del desempeño de los diferentes C-M óptimos para dos géneros distintos en dos computadoras diferentes. A fines comparativos se agregan los valores verdaderos de las colecciones analizadas.

Con el fin de evaluar la robustez del método propuesto, se determinó la cota del error de predicción para el 90 % de las combinaciones mencionadas. La misma fue de:

- 17,9 % (Est 1), 18,3 % (Est 2) y 21,7% (Est 3) para la predicción del tamaño de salida.
- 1,7 % en la PC₂ y 4,6 % en la PC₃ para la predicción del tiempo de codificación. Esto es válido para los tres métodos propuestos y no consideran los C-M más rápidos en cada frontera. Asimismo, el método Est 1 brinda el error más pequeño para los C-M más lentos.

Por lo tanto, el desempeño de cada C-M óptimo se predice con mayor precisión en cuanto al tiempo de codificación. Esto es válido incluso para un ordenador (PC₃) distinto a los evaluados.

En cuanto a la compresión, el error de predicción es similar para todos los C-M óptimos (en la Figura 13 hay un desplazamiento vertical para cada conjunto de valores estimados). Este error se puede minimizar mediante una mejor predicción de la compresión de la colección con Opt M. Para ello se deben adoptar más piezas, lo que aumenta la carga computacional.

Finalmente, la Figura 13 muestra que Mon N (segundo codec de izquierda a derecha en la colección de jazz) no es óptimo como se esperaba, pues se observa que comprime menos que TAK N y tarda más tiempo en hacerlo. Esto se debe a efectos de dispersión en la compresión

en las fronteras normalizadas, lo que significa que si dos C-M diferentes tienen una compresión muy similar, uno de ellos podría ser excluido del conjunto óptimo. Sin embargo, este C-M puede considerarse cuasi-óptimo, ya que su rendimiento es similar al esperado y es muy útil para el usuario que necesita información sobre su desempeño por anticipado.

7 CONCLUSIONES

El tiempo de codificación depende esencialmente del codificador y de la computadora utilizada. La compresión es dependiente en menor medida del codificador utilizado y en gran medida del género. Atribuimos esta dependencia a dos causas. Una de ellas es una estructura de codificación común en los algoritmos analizados, basado en la predicción y la codificación del residuo resultante. La otra causa tiene que ver con el origen de la señal codificada. En cada género se aplican técnicas de masterizado comunes que influyen sobre la varianza en cada canal del archivo resultante. Dicha varianza modifica la redundancia intracanal y por ende afecta el desempeño del predictor, ya que éste brinda un residuo cuya varianza es función de la varianza de entrada. Esto repercute en forma directa sobre la compresión alcanzada.

Se propuso un método que permite al usuario determinar las combinaciones de C-M óptimas en el dominio velocidad-compresión para un dado género musical. Más aún, esta herramienta permite predecir el desempeño de los C-M óptimos en forma robusta y eficiente. De esta manera, el audiófilo que desea codificar una colección musical puede conocer en forma anticipada no sólo los C-M más convenientes, sino también el tamaño que poseerá la colección comprimida y el tiempo necesario para llevar a cabo dicha compresión. Este método es flexible, dado que puede aplicarse para nuevos géneros y subgéneros, acorde a la necesidad del usuario.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), gracias a una beca de investigación postdoctoral otorgada al primer autor de este trabajo. Los restantes autores de este artículo son Adscriptos al Laboratorio de Acústica y Electroacústica de la Universidad Nacional de Rosario. Se agradecen las valiosas revisiones y observaciones realizadas por los Profesores Ernesto Accolti, Federico Miyara y Marcos Katz.

REFERENCIAS

- Ashland, M.T., *Monkey's Audio*. 2011. Disponible en: <http://www.monkeysaudio.com/index.html>
- Becker, T., *Tom's audio compressor (TAK)*. 2011. Disponible en: <http://thbeck.de/Tak/Tak.html>
- Bruickers, A., Oomen, A., van der Vleuten, R., Lossless coding for DVD audio. *Proc. 101st Audio Eng. Soc. Conv.*, 4358, L.A., 1996.
- Bryant, D., *WavPack - Hybrid Lossless Audio Compression*. 2009. Disponible en: <http://www.wavpack.com/>
- Coalson, J., *FLAC - Free lossless audio codec*. 2009. Disponible en: <http://flac.sourceforge.net/>
- Craven, P., Gerzon, M., Lossless coding for audio discs. *Journal of Audio Eng. Soc.*, 44:706–720, 1996.
- Craven, P., Law, M., Stuart, J., Lossless compression using IIR prediction filters. *Proc. 102nd Audio Eng. Soc. Conv.*, 4415, Munich, 1997.

- Djuric, A., *TTA Lossless audio codec*. 2010. Disponible en: http://en.true-audio.com/TTA_Lossless_Audio_Codec_-_Realtime_Audio_Compressor
- DVD Specifications for Read-Only Disc. *DVD Forum*. 2004. Disponible en: <http://www.dvdforum.com>
- Ghido, F., *Ghido's Data Compression Page*. 2011. Disponible en: <http://www.losslessaudio.org/>
- Hans, M.C., Schafer, R.W., Lossless compression of digital audio. *IEEE Sig. Proc. Mag.*, 8:21–32, 2001.
- Liebchen, T., MPEG-4 Audio Lossless Coding (ALS). 2011. Disponible en: http://www.nue.tu-berlin.de/menue/forschung/projekte/beendete_projekte/mpeg-4_audio_lossless_coding_als/parameter/en/
- Liebchen, T., *The LPAC homepage*. 2010. Disponible en: http://www.nue.tu-berlin.de/menue/mitarbeiter/ehemalige_mitarbeiter/tilman_liebchen/lpac_-_lossless_audio_codec_for_windows_and_linux/
- Malvar, H.S., Lossless and Near-Lossless Audio Compression Using Integer-Reversible Modulated Lapped Transforms. *Proc. IEEE Data Compr. Conf.*, 2007.
- Pawlowski, P., *Foobar2000*. 2010. Disponible en: <http://www.foobar2000.org/>
- Purat, M., Liebchen, T., Noll, P., Lossless transform coding of audio signals. *Proc. 102nd Audio Eng. Soc. Conv.*, 4414, Munich, 1997.
- Robinson, T., Shorten: Simple lossless and near-lossless waveform compression. Cambridge Univ. Eng. Dept., UK, Tech. Rep. 156, 1994.
- Rodríguez Guerrero, J.M., Graphic tool for the evaluation of lossless audio codecs. *Codificación de Audio Digital, Laboratorio de Acústica y Electroacústica, Univ. Nac. de Rosario*. Disponible en: <http://www.fceia.unr.edu.ar/acustica/codecdigital>
- Rodríguez Guerrero, J.M., Treffiló, M.A., Roveri, E.A., Audio tracks tested with lossless audio codecs. *Codificación de Audio Digital, Laboratorio de Acústica y Electroacústica, Univ. Nac. de Rosario*. Disponible en: <http://www.fceia.unr.edu.ar/acustica/codecdigital>
- Salomon, D., *Data Compression*, 4th ed. Springer-Verlag, 2007.
- Wikipedia, *Lossless comparison*. 2011. Disponible en: http://wiki.hydrogenaudio.org/index.php?title=Lossless_comparison