

COMPRESION DE AUDIO EN EL ESPACIO TRANSFORMADO

*Ing. Oscar N. Briá¹
A.C. Ramiro F. Sosa²
A.C. Pablo J. Thomas³
C.E.T.A.D.⁴
L.I.D.I.⁵*

*Departamento de Informática. Facultad de Ciencias Exactas.
Universidad Nacional de La Plata.*

Resumen

Se presenta una Técnica de Compresión de Audio en el Espacio Transformado. La señal de entrada es dividida en bloques de longitud fija, los cuales son transformados al dominio frecuencial. Los coeficientes transformados son agrupados en bandas y luego cuantizados.

Se discuten aspectos de la codificación en bandas, así como los resultados obtenidos según el criterio empleado para aprovechar los beneficios de tal técnica.

Introducción

La manipulación de audio por medio de computadoras, es un desarrollo relativamente nuevo. Tal manipulación ha sido posible desde el nacimiento de las computadoras digitales, pero sólo durante los últimos cinco años, gracias a una disminución de los costos del hardware ha sido posible que esta tecnología llegue a los escritorios de los usuarios. Ahora es de esperar que la posibilidad de escuchar sonido digital se transforme en una parte integral de la inminente "Revolución Multimedia". [NEL94]

Como las tecnologías digitales evolucionan, y las economías de integración de gran escala tienden a ser alcanzadas, se renueva el interés por métodos eficientes de codificación digital de audio. [FLA79]

¹Profesor Adjunto, Dpto. de Informática, Facultad de Cs. Exactas; y Profesional de Apoyo del CONICET en el CETAD, Facultad de Ingeniería. Univ. Nacional de La Plata.

Email: onb@ada.info.unlp.edu.ar

²Ayudante Diplomado. Dpto. de Informática. Facultad de Ciencias Exactas. Univ. Nacional de La Plata. Lugar de investigación: LIDI. Email: rfs@info.unlp.edu.ar

³Jefe de Trabajos Prácticos con Semi dedicación. Dpto. de Informática. Facultad de Ciencias Exactas. Univ. Nacional de La Plata. Lugar de investigación: LIDI. Email: pthomas@info.unlp.edu.ar

⁴48 y 116 . 2do piso. La Plata (1900). Bs.As. Argentina

⁴Director: Ing. Antonio A. Quijano. Prof. Titular. Univ. Nacional de La Plata.

⁵50 y 115 1er piso. La Plata (1900). Bs. As. Argentina. Tel : 54 - 021 - 227707

⁵Director: Ing. Armando De Giusti. Investigador Principal del CONICET. Prof. Titular Dpto. de Informática. Vicedecano de la Facultad de Ciencias Exactas, Univ. Nacional de La Plata. Investigador en el LIDI.

Desde el punto de vista de los usuarios, un tema obligado en el uso de Multimedia es la Compresión de Datos. En el caso particular de audio digitalizado los requerimientos de almacenamiento son muy elevados. [THO94]

Para lograr calidad de audio CD se necesita muestrear 44100 veces por segundo, con una resolución de 16 bits por cada muestra. En stereo esto lleva a 1,4 Mbits por segundo, con lo cual se ve la imperiosa necesidad de comprimir. [THO94]

La codificación de audio en sucesivos bloques en el espacio transformado, es una de las técnicas más eficientes que se conoce en lo referente a buena calidad y alta compresión.

Codificación de Audio en el Espacio Transformado

i- Generalidades

El concepto de Entropía Condicional proveniente del marco de la "Teoría de Información", sugiere que la cantidad de información necesaria para representar una señal puede reducirse, si las muestras que la componen no son independientes[SIM86]. La dependencia o correlación es una función de la separación en el tiempo para la señales de audio.

Los algoritmos de codificación en el dominio del tiempo, tratan la señal en forma completa [JAY84] (sin división alguna) eliminando la redundancia antes de la codificación mediante la aplicación de algún método predictivo. Las principales diferencias entre los algoritmos utilizados están dadas por el grado de predicción empleado, y si esos esquemas son adaptativos o no.

Por otro lado, los algoritmos de codificación en el dominio frecuencial dividen la señal de entrada en componentes de frecuencias separadas, y codifican cada uno de ellos separadamente.

Para una digitalización eficiente se utiliza *Block Transform Coding* (BTC). La idea central subyacente de esta técnica en el tratamiento de señales no independientes es no codificar cada muestra individualmente, sino en grupos los cuales se llaman bloques. La correlación puede ser evaluada de algún modo, y mediante transformaciones las muestras pueden ser "decorrelacionadas" [BRI95].

Si se tiene en cuenta la energía de cada muestra dentro del bloque en el dominio transformado ("decorrelacionado"), generalmente es posible descartar algunas de ellas logrando así cierta compresión con mínima pérdida de información.

ii - Métodos

BTC es una técnica general para codificación de señales digitalizadas en una o más dimensiones. Para audio, la señal es muestreada en el tiempo, subdividida en bloques de N muestras y transformada al dominio espectral para codificación y transmisión o almacenamiento. En el decodificador, cada bloque de muestras espectrales se transforma nuevamente al dominio del tiempo, y la señal de audio es sintetizada concatenando los bloques reconstruidos. [BRI95]

BTC es una técnica de codificación en el dominio frecuencial, sin embargo, el problema de asignación de bits para cada línea espectral no es meramente simple.

La codificación en el dominio espectral puede ser justificada por una teoría "bien fundada". Puede demostrarse que la performance según el error cuadrático medio es teóricamente la misma para *Differential Pulse Code Modulation* (DPCM) que para BTC.

Varias técnicas pueden ser usadas para la codificación de señales de audio en el dominio espectral con diferente performance y complejidad de implementación. Entre ellas, *Zonal Transform Coding* (ZTC), *Threshold Transform Coding* (TTC) y *Adaptative Transform Coding* (ATC). En ZTC simplemente no se consideran las líneas de alta frecuencia. En TTC no se consideran las líneas con energías menores que un "Threshold" mínimo. ATC es similar a TTC en cuanto a la selección de líneas, pero usa un algoritmo específico para alocaión de bits de modo que, tal alocaión se acerque lo más posible a una asignación óptima, dada teóricamente.

Diferentes estrategias de alocaión de bits han sido sugeridas en la literatura del tema. Las tendencias del momento, tratan de usar modelos psicoacústicos para controlar el cuantizador y el codificador.

Utilizando valores apropiados para la longitud de los bloques y estrategias adecuadas de alocaión de bits, los procedimientos de "Transform Coding" proporcionan digitalizaciones de alta calidad a una tasa promedio de 2 bits/muestra, y digitalizaciones aceptables de 1 bit-muestra.

También surgen problemas empleando la técnica de BTC. La eficiencia dependerá del tipo de transformada lineal (Transformada Discreta de Fourier, Transformada Discreta Coseno, etc) utilizada, y la estrategia de alocaión de bits para los coeficientes transformados.

Y aparecen problemas de distorsión de bloque en cualquier implementación como consecuencia de la cuantización "grosera". Estos son: *Clicking*, *Burbling* y *Pre-echoes* [TRI79] [O'SH87]. Varias técnicas han sido sugeridas para resolver los efectos de bloque, llamadas genéricamente "transformadas solapadas" como por ejemplo la Transformada Discreta Coseno Modificada (MDCT) [IWA92].

Se concluye que, la principal ventaja de BTC es su inherente resolución en frecuencia.

Esquema de Codificación

El proceso de codificación y decodificación, dada una señal, utilizando la metodología descrita previamente es :

- i - cambio del dominio del tiempo al dominio frecuencial
- ii - codificación de las líneas espectrales
- iii - decodificación de las líneas espectrales
- iv - cambio del dominio frecuencial al dominio temporal

i - Cambio del dominio temporal al dominio frecuencial

En primer lugar se divide la señal de entrada en bloques o "frames" de longitud fija N. Para cambiar de dominio se debe emplear una transformación lineal adecuada, ya que el objetivo es "decorrelacionar" las muestras. A fin de lograr la "decorrelación" mencionada se analizan 3 alternativas:

- *Transformada de Karhunen-Loève* (KLT) : empaqueta la mayor parte de la energía de la señal dentro de una mínima cantidad de coeficientes para un nivel de error deseado, conduciendo teóricamente a la tasa mínima de bits.

- *Transformada Discreta de Fourier* (DFT) : su cálculo es más simple que KLT, y se expresa como una combinación de funciones sinusoidales [O'SH87]

- *Transformada Discreta Coseno* (DCT) : es la más cercana a KLT en cuanto a su optimalidad. Una ventaja importante de destacar es que atenúa el efecto de bordes en el solapamiento de bloques (clicking) [O'SH87]

ii - Codificación de las líneas espectrales

Se puede realizar una subdivisión en etapas que, si bien no son totalmente disjuntas, sirven para comprender el esquema de codificación.

- a - asignación de bits a cada línea espectral
- b - división de cada "frame" en bandas
- c - asignación de bits por banda
- d - cuantización de frames por banda

a - Cada muestra del espacio transformado debe representarse con una cantidad apropiada de bits, teniendo en cuenta una tasa de bits promedio predeterminada. Por lo tanto, esta asignación no debe exceder una cantidad de bits por frame, previamente establecida.

b - Se podría cuantizar cada una de las líneas espectrales usando la cantidad de bits determinada en a, pero es posible reducir ese número subdividiendo

cada frame en bandas, las cuales pueden ser de igual longitud o longitud variable. De este modo se optimiza el almacenamiento de la señal codificada, logrando mayor compresión en la misma.

c - A cada banda se le asigna una cantidad fija de bits con la cual se cuantizarán todas las muestras de la misma. Esa asignación puede variar según la relación de compresión deseada, teniendo en cuenta diferentes factores, entre ellos los principios psicoacústicos.

d - Una vez que cada una de las bandas tienen asignadas un número de bits determinado, se realiza la cuantización de cada línea espectral con la cantidad de bits indicada por la banda a la cual pertenece.

iii - Decodificación de las líneas espectrales

Se reconstruye cada una de las líneas espectrales en función de la asignación de bits de la etapa anterior, quedando así las muestras disponibles para ser cambiadas de dominio nuevamente.

iv - Cambio del dominio frecuencial al dominio temporal

En esta última etapa se aplica la transformación lineal inversa a la empleada en el paso i.

Resultados Obtenidos

Utilizando el esquema de codificación descripto, se obtuvieron diversos resultados según :

- transformada utilizada : DCT
- método de codificación : ATC
- la longitud de cada frame en líneas espectrales (BL)
- la cantidad de bits total por frame en una asignación inicial (SUMB)
- la cantidad bandas por frame (BANDAS)
- la longitud de las bandas (MUEBAND)
- la asignación de bits por banda (B)

Para todos los casos se utilizaron los siguientes parámetros:

BL = 1025

SUMB = 2050

BANDAS = 25

Se experimentó inicialmente con bandas de longitud fija (MUEBAND = 41), y luego con longitudes variables establecidas según la frecuencia ,

asignando longitudes pequeñas a las frecuencias bajas y longitudes grandes a las frecuencias altas ($4 \leq \text{MUEBAND} \leq 305$, monótono creciente).

La asignación de bits estuvo relacionada con la tasa de compresión y fidelidad deseada, por lo que se manejaron varias alternativas:

B = cantidad promedio de bits de las muestras de una bandas (BPROM)

B = cantidad máxima de bits para todas las muestras de la banda (BMAX)

B = (BDM) cantidad promedio de bits de las muestras de una banda + cantidad indicada por la Desviación Media [SPI94]

B = (BDS) cantidad promedio de bits de las muestras de una banda + cantidad indicada por la Desviación Standard [SPI94]

Estas dos últimas asignaciones con MUEBAND fijo fueron también fijas; pero con MUEBAND variable se estableció el siguiente criterio: *a medida que la longitud de la banda crece la cantidad de bits decrece, respetando de este modo los principios psicoacústicos; ya que la mayor pérdida se da en las bandas más anchas y esto es imperceptible por el oído humano.*

Considerando Bandas de longitud fija

	BPROM	BMAX	BDM	BDS
Compresión	de 1:3 a 1:4	de 1:2.5 a 1:3	de 1:3 a 1:4	de 1:3 a 1:3.5
Calidad	--	Ok	Ok	Ok

Considerando Bandas de longitud variable

	BPROM	BMAX	BDM	BDS
Compresión	de 1:3 a 1:4	1:2.5	1:5	de 1:3 a 1:4
Calidad	--	Ok	Ok	Ok

Referencias:

1:x a 1:y : significa que se comprime entre x e y veces

-- : significa mala calidad de audición

Ok : muy buena calidad de audición, la pérdida es imperceptible

Conclusiones

La Codificación de Audio en el Espacio Transformado, es una metodología muy útil para obtener excelentes resultados en lo referente a compresión sin degradación de la calidad sonora. Con dicha metodología a través de la combinación de varias técnicas incluyendo: bandas de longitud variable, y la asignación de bits según la longitud de las bandas, logramos obtener una relación de compresión de **5 veces** .

El resultado obtenido hasta el momento nos estimula a seguir avanzando en el estudio y desarrollo del tema. Cualquier sugerencia o duda al respecto serán bienvenidas por los autores de este trabajo.

Bibliografía

[BRI95] Oscar N. Bria, "Block Transform Coding for Audio Signals", Reporte Técnico CETAD, Marzo 1995.

[FLA79] J. Flanagan et. al., "Speech Coding", *IEEE Transaction on Communications*, April 1979.

[IWA92] M. Iwadare et. al., "A 128 Kb/s Hi-Fi Audio CODEC Based on Adaptative Transform Coding with Adaptative Block Size MDCT", *IEEE Journal on Selected Areas in Comm*, January 1992.

[JAY84] S.N. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice-Hall, 1984.

[NEL94] J. Nelson, *The Data Compression Book*, 1994.

[O'SH87] D. O'Shaughnessy, *Speech Communication: Human and Machine*, Addison Wesley, 1987.

[SIM86] J.-C. Simon, *Patterns and Operators*, McGraw-Hill, 1986.

[SPI94] M. R. Spiegel, *Estadística*, McGraw-Hill, 1991.

[THO94] Pablo J. Thomas, Ramiro F. Sosa y Rodolfo Bertone, "Análisis de Audio, Algoritmos de Compresión y Descompresión de Archivos de Sonido". Informe Técnico LIDI. Octubre 1994.

[TRI79] Tribolet and Crochiere, "Frecuency Domain Coding of Speech", *IEEE Transaction on Acoustic, Speech, and Signal Processing*, October 1979.