ANALISIS DE UN ALGORITMO DE COMPRESION DE AUDIO EN EL ESPACIO TRANSFORMADO

Ing. Oscar N. Bria[1]
Lic. Ramiro F. Sosa[2]
Lic. Pablo J. Thomas[3]
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 el algoritmo de una técnica de Compresión de Audio en el Espacio Transformado.

El trabajo es continuación de una serie de investigaciones y desarrollos que los autores han realizado desde 1995 y aquí se pone énfasis en el planteo y discusión del algoritmo seleccionado.

Se analizan detalles del algoritmo, a la vez que se muestran resultados de mediciones de calidad de la señal recuperada en el decompresor, tanto de carácter objetivo como subjetivo.

Para finalizar se realiza una breve discusión de las limitaciones prácticas de ejecución sobre microcomputadora y se discute las posibles implementaciones en Tiempo Real.

[1]Profesor 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
[2]Ayudante Diplomado. Dpto. de Informática. Facultad de Ciencias Exactas. Univ. Nacional de La Plata. Lugar de investigación: LIDI. Email: rfs@info.unlp.edu.ar
[3]Jefe 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
[4] 48 y 116 . 2do piso. La Plata (1900). Bs.As. Argentina
[5]50 y 115 1er piso. La Plata (1900). Bs. As. Argentina. Tel : 54 - 021 - 227707

Introducción

Los algoritmos de codificación en el dominio del tiempo, tratan la señal en forma completa eliminando la redundancia antes de la codificación mediante la aplicación de algún método predictivo. [JAY94] 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.[BST95]

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.[BST95]

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].

Método empleado

BTC (Block Transform Coding) 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] La principal ventaja de BTC es su inherente resolución en frecuencia.

Varias técnicas pueden ser usadas para la codificación de señales de audio en el dominio espectral con diferente performance y complejidad de implementación. La elegida es *Adaptative Transform Coding* (ATC). en la cual no se consideran las líneas con energías menores que un "Threshold" mínimo., y usa un algoritmo específico para alocación de bits de modo que, tal alocación se acerque lo más posible a una asignación óptima, dada teóricamente [DEL93].

Codificación

El proceso de codificación y decodificación, dada una señal, utilizando la metodología descripta previamente es [BST95]:

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

Primeramente se divide la señal de audio de entrada en **bloques o frames** de longitud fija N. La longitud de cada bloque es de 1025 muestras, de las cuales 25 son solapadas entre bloques consecutivos para atenuar los problemas de distorsión de lados. En nuestro caso la señal está digitalizada con una resolución de 16 bits utilizando 1 o 2 canales (mono o estéreo) y muestreada a 44.100 Hz. Para cambiar de dominio se debe emplear una transformación

lineal adecuada ya que el objetivo es "decorrelacionar" las muestras. A fin de lograr tal objetivo, adoptamos el uso de **DCT** (Discrete Cosine Transform), ya que atenúa los efectos de bordes en los límites de bloques.

La decorrelación es necesaria ya que nos permite eliminar algunas líneas sin causar una pérdida significativa en la información de la señal. Por lo tanto de aquí en más la consideramos como hipótesis de trabajo para las etapas siguientes.

ii - Codificación de las líneas espectrales

La técnica de codificación que utilizamos es **ATC**, ya que es la que más aproxima la forma de onda original, al establecer umbrales diferentes para cada bloque, o sea , se adapta a los cambios de la señal, y usa una cantidad fija de bits para cada frame, los cuales se distribuyen en forma proporcional al valor real discreto de cada una de las muestras.

Esta etapa se subdivide en 2 pasos que, si bien no son totalmente disjuntos, clarifican el proceso de codificación. Dichos pasos son:

a. Asignación de bits a las líneas espectrales de un frame

b. Cuantización de líneas espectrales de un frame

Teniendo en cuenta que nos surgieron diferentes ideas para lograr mayor compresión durante la codificación, podemos establecer una serie de variantes que marcaron el progreso de nuestro trabajo [TSO96] . Tales variantes fueron:

1) Codificar cada línea espectral de un frame con un número de bits específico

2) Dividir cada bloque en bandas de long fija (con diferentes alternativas)

3) Dividir cada bloque en bandas de longitud variable

4) Establecer cantidades variables de bits para codificar cada frame según la tasa de compresión deseada, combinado con **3.**

1) Debido a que BTC proporciona teóricamente digitalizaciones de alta calidad a una tasa promedio de 2 bits por muestra, utilizamos 2050 (2*1025) bits para codificar todo el frame. A partir de esta hipótesis inicial, establecimos el *umbral* del bloque y la *cantidad de bits* para codificar cada coeficiente del mismo, del siguiente modo:

Proceso 1 : Establecer Umbral y Asignación de Bits

N = *1025 (tamaño del bloque)*

i = $1...N$

$Sumalg2$ = $\sum lg_2(Muestra_i)$

TotB = *cantidad de bits para codificar todo el bloque*

$Umbral = (Sumalg2 - TotB) / N$

Int (x) = *parte entera más cercana superior(x)*

Bits[i] = *cantidad de bits para codificar la* $Muestra_i$

Proc 1.1

Comienzo ***Proc 1.1***

$Bits[i] = Int\ (lg_2(Muestra_i) - Umbral)$ *si* $lg_2(Muestra_i) - Umbral > 0$

$Bits[i] = 0$ *si* $lg_2(Muestra_i) - Umbral <= 0$

$Signo[i] = 1$ *si* $Muestra_i < 0$

$Signo[i] = 0$ *si* $Muestra_i >= 0$

$SumaB = \sum Bits[i]$

Mposit = cantidad de muestras con $Bits[i] > 0$

Fin ***Proc 1.1***

Mientras (SumaB > TotB)
  Umbral = Umbral + (SumaB - TotB) / Mposit {distribuir la diferencia}
  Proc 1.1
Fin Mientras

Proceso 2 : Cuantización
$Diferencia_i = lg_2(Muestra_i) - Umbral$
Si Bits[i]>0
  $MuestraCuantizada_i = Int ((Bits[i] - Diferencia_i) * 2^{Bits[i]})$
sino
  $MuestraCuantizada_i = 0$
Fin Si

2) Como se ve en el punto **1)**, cada coeficiente espectral tiene una cantidad de bits asociada. Este tipo de codificación es óptima en lo referido a calidad de reproducción, pero no lo es desde el punto de vista del espacio requerido. Entonces vimos la necesidad de cambiar la codificación de manera tal que, varias muestras usen la misma cantidad de bits para ser representadas (la cantidad indicada por la banda).
De este modo surgió la idea de subdividir cada bloque en subbloques de longitud fija, los cuales denominamos *bandas*. Obviamente a menor cantidad de bandas menor pérdida, pero luego de varias pruebas deducimos que usando entre 20 y 27 la degradación era aceptable. Por lo tanto establecimos 25 bandas por frame. Cada una de ellas con 41 muestras.
Para establecer el umbral y realizar la asignación de bits agregamos al ***Proceso 1*** :
Proc 1.2 : Establecer las Bandas de longitud fija
K = cantidad de bandas = 25
k = 1...25
Mueband = cantidad de muestras por banda = 41
Bandas[k] = cantidad de bits para codificar las muestras de la Banda$_k$
$Bmax_k$ = *máx(Bits[i])* $\forall$ *i tal que k*Mueband <= i < (k+1)*Mueband*
$Bprom_k$ = *promedio(Bits[i])* $\forall$ *i tal que k*Mueband <= i < (k+1)*Mueband*
DM_k = *Desviación Media para la k-ésima banda*
DS_k = *Desviación Standard para la k-ésima banda*
$DM_k = (\sum$ *valor absoluto (Bits[i] -* $Bprom_k$ *)) / Mueband*
DS_k = *raiz cuadrada (* $(\sum (Bits[i] - Bprom_k)^2)$ */ Mueband)*
$Bandas[k] = Bmax_k$ *alternativa.1*
$Bandas[k] = Bprom_k$ *alternativa.2*
$Bandas[k] = Bprom_k + DM_k$ *alternativa.3*
$Bandas[k] = Bprom_k + DS_k$ *alternativa.4*

Proceso 2 : Cuantización
$Diferencia_i = lg_2(Muestra_i) - Umbral$
Si $Diferencia_i > 0$ *.y* $Bandas[k] > 0$
  $MuestraCuantizada_i = Int ((Diferencia_i / Bandas[k]) * 2^{Bandas[k]})$
sino
  $MuestraCuantizada_i = 0$
Fin Si

3) Con el mismo propósito de codificación en bandas quisimos obtener mayores ventajas de esta técnica. Una alternativa fue cambiar el número de bandas, pero vimos que la ganancia no era lo suficiente como para justificar dicho cambio. Por lo tanto pensamos en bandas de

longitud variable que, combinadas con la Desviación Media o Standard nos diera mayores beneficios[SPI94]; entonces adecuamos esta idea a los principios psicoacústicos. Con esta meta agrupamos los coeficientes transformados en bandas no uniformes que reflejan el sistema auditivo humano.
Este tipo de análisis indica cuáles frecuencias son críticas y deben ser codificadas con alta precisión, y cuáles son menos sensitivas y pueden tolerar algún ruido de cuantización sin degradación de la calidad percibida del sonido. La sensibilidad del oído varía con las frecuencias, es más sensible a frecuencias cercanas a 4 kHz. [TSU92]. Las bandas no uniformes surgen naturalmente a partir de experiencias en audición humana, y pueden también derivarse a partir de la distribución de las células sensoriales en el oído interno. Por lo tanto las bandas pueden pensarse como la escala de frecuencias usada por el oído.
Dichas bandas son más estrechas en frecuencias bajas que en frecuencias altas, en efecto, el 75% de ellas están ubicadas por debajo de los 5 kHz. Esto indica que el oído recibe más información de las bajas frecuencias que de las altas.
Con el esquema hasta aquí descripto los ***Procesos 1, 2 y 3*** son similares al punto **2)**, con la excepción que :
Mueband = depende de la cantidad de muestras por banda

4) Los mejores resultados, que veremos en detalle más adelante, los obtuvimos con la codificación con bandas de longitud variable. A partir de ello, buscamos aumentar la tasa de compresión y que el algoritmo sea capaz de recibir un parámetro con la tasa deseada (obviamente con restricciones). Hasta este punto la cantidad inicial de bits para cada frame fue uniforme, de aquí en más dependerá del nivel de compresión recibido como parámetro [TSO96].

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.**

Criterios de Fidelidad

Cualquier evaluación de la calidad de señal implica una medición de la fidelidad. Para la mayoría de los sistemas de comunicación es difícil de especificarla cuantitativamente, debido a que esto involucra la percepción humana. Por lo tanto, la calidad de audio es tradicionalemente evaluada por el criterio de quien escucha.

Mientras la calidad de cada técnica de compresión es subjetiva, es implícito que a medida que el radio de compresión aumenta la calidad disminuye, para una frecuencia de muestreo establecida.

Básicamente, cada técnica de compresión tiene un límite de calidad para diferentes radios de la misma. Esto significa que no existe una única técnica de compresión que sea la ideal para todas las aplicaciones potenciales. La selección de una técnica apropiada es un elemento clave de cualquier aplicación en la que participen señales de audio.[SOS95]

Resultados Obtenidos

Las pruebas realizadas nos indicaron que este método de codificación no es aplicable con los mismos resultados a cualquier tipo de señal de audio. Específicamente, a un nivel de compresión deseado, la calidad de reproducción obtenida no es igual en todos los casos.

El estudio de la performance del método se realizó con los siguientes tipos de señales de audio [TSO96] :

- música pop **(i)**
- música clásica **(ii)**
- speech **(iii)**
- efectos sonoros **(iv)**

Los resultados se reflejan en la siguientes tablas, y valen tanto para señales mono como estéreo:

Cuadro 1 : Asignación de bits individual a cada línea espectral

	2 bits promedio
Compresión	1:2
Calidad **(i)**	Excelente
Calidad **(ii)**	Excelente
Calidad **(iii)**	Excelente
Calidad **(iv)**	Excelente

Cuadro 2 : 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 **(i)**	Mala	Excelente	Muy Buena	Muy Buena
Calidad **(ii)**	Mala	Excelente	Buena	Muy Buena
Calidad **(iii)**	Mala	Excelente	Muy Buena	Muy Buena
Calidad **(iv)**	Mala	Excelente	Muy Buena	Muy Buena

Cuadro 3 : 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 **(i)**	Mala	Excelente	Muy Buena	Muy Buena
Calidad **(ii)**	Mala	Excelente	Buena	Buena
Calidad **(iii)**	Mala	Excelente	Buena	Muy Buena
Calidad **(iv)**	Mala	Excelente	Buena	Muy Buena

Cuadro 4 : Bandas de longitud variable, con variación de la cantidad de bits por frame

Tasa	Cantidad de Bits	Calidad (i)	Calidad (ii)	Calidad (iii)	Calidad (iv)
3	4150	Excelente	Muy Buena	Excelente	Excelente
4	2800	Muy Buena	Buena	Muy Buena	Muy Buena
5	2050	Muy Buena	Regular	Buena	Buena
6	1450	Buena	Mala	Regular	Buena
7	1100	Buena	Mala	Regular	Regular
8	800	Buena	Mala	Mala	Regular
9	500	Buena	Mala	Mala	Mala
10	290	Regular	Mala	Mala	Mala
11	150	Regular	Mala	Mala	Mala
12	100	Regular	Mala	Mala	Mala

Obviamente estos resultados dependen del oído subjetivo del receptor y de las muestras analizadas dentro de cada tipo de señal; por lo tanto no los consideramos absolutos sino como un referente o patrón. Aunque, cabe destacar que los mejores logros los encontramos codificando música pop y según el ejemplo usado, la compresión 1:10 o 1:12 es aceptable en lo referido a calidad de reproducción [TSO96].

Para mostrar *objetivamente* que tales logros fueron exitosos, calculamos la relación señal ruido total y la relación señal ruido segmental, para los resultados obtenidos en el cuadro 4, con lo que se observa que la codificación de música pop fue la más aceptable para los distintos niveles de compresión:

SNR para Bandas de longitud variable, con variación de la cantidad de bits por frame.

Tasa	Cantidad de Bits	Calidad (i)	Calidad (ii)	Calidad (iii)	Calidad (iv)
3	4150	14.24509 dB	2.57614 dB	19.49338 dB	2.60884 dB
5	2050	14.68057 dB	2.58396 dB	20.13647 dB	3.53323 dB
10	290	13.46096 dB	2.58777 dB	18.34249 dB	1.92300 dB

SNR segmental para Bandas de longitud variable, con variación de la cantidad de bits por frame.

Tasa	Cantidad de Bits	Calidad (i)	Calidad (ii)	Calidad (iii)	Calidad (iv)
3	4150	20.91071 dB	2.39821 dB	20.72236 dB	2.57475 dB
5	2050	21.55829 dB	2.43880 dB	21.27642 dB	3.49948 dB
10	290	19.05656 dB	2.39646 dB	17.56410 dB	1.89441 dB

También realizamos la comparación con los resultados de un algoritmo de compresión sin pérdida tipo Pkzip como se muestra en la siguiente tabla :

Tipo de Señal	Compresión
Música Pop Mono	de 2 a 2.5 veces
Música Pop Estéreo	1.14 veces
Música Clásica	2.13 veces
Speech	1.96 veces
Efectos Sonoros	1.80 veces

Limitaciones Prácticas en PC's. Implementaciones en Tiempo Real

Por su naturaleza, la compresión de audio es un proceso de tiempo real. Históricamente, la arquitectura de las PC's (sistema operativo y hardware) no fue diseñada para soportar operaciones en tiempo real en forma óptima.

Con la demanda creciente sobre las PC's para soportar manejo sofisticado de gráficos y proveer acceso a sistemas de archivos, el procesamiento de audio fue relegado a un plano secundario. Con este panorama existen dos alternativas posibles: ajustar el procesamiento de audio en las MIPS (millones de instrucciones por segundo) restantes, o tener un coprocesador dedicado que pueda satisfacer los requerimientos de tiempo real para dicho procesamiento.

Aquí es donde se manifiestan los beneficios de los procesadores de señales digitales (DSPs). La arquitectura de los DSPs ha sido optimizada desde sus comienzos para soportar las necesidades de los sistemas de tiempo real. El DSP está diseñado para realizar una gran cantidad de operaciones en cada ciclo de reloj. Esta unidad de trabajo es consistente e incorpora la capacidad de realizar un cálculo acumulativo múltiple (MAC) y al mismo tiempo efectuar sumas u otro tipo de operaciones ALU.

Debido a esto, podemos decir que la esencia de las variadas técnicas de compresión de audio es un tipo de cálculo conocido como "suma de productos", y los DSP's proporcionan un soporte perfecto para ellas [REI94].

Conclusiones

La Codificación Digital 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: división de la señal de audio en frames, separación del espectro en bandas de longitud variable, y asignación de bits según la longitud de las bandas y principios psicoacústicos, logramos obtener una relación de compresión entre 3 y 12 veces. La calidad obtenida depende del tipo de información que contiene la señal que se comprime, siendo la música pop la que soporta la tasa más alta (entre 10 y 12) sin pérdida significativa para el oído humano.

Bibliografía

[BRI95] Oscar N. Bria, **"Block Transform Coding for Audio Signals"**, Reporte Técnico CETAD, Marzo1995.

[BST95] Oscar N. Bria, Ramiro F. Sosa y Pablo J. Thomas, **"Compresión de Audio en el Espacio Transformado"**, 1er Congreso Argentino de Ciencias de la Computación (CACIC '95). Octubre 1995.

[DEL93] J.R. Deller et. al., ***Discrete-Time Processing of Speech Signals***, Macmillan, 1993.

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

[REI94] J.B. Reimer, **"DSP and Audio Compression"**, ***Dr. Dobb's Multimedia Sourcebook***, Winter 1994.

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

[SOS95] Ramiro F.Sosa, Oscar N.Bria y Pablo J. Thomas, **"Audio Compression in the Transformed Space"**, ICIE '95 (II International Congress on Information). Noviembre 1995.

[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.

[TSU92] K. Tsutsui et. al. , **"ATRAC: Adaptative Acoustic Coding for Minidisc"**, *Audio Eng. Soc. Preprint,* presented at the 93rd Convention AES, October 1992.

[TSO96] Pablo J. Thomas, Ramiro F. Sosa, **"Compresión Digital de Audio en el Espacio Transformado"**. Trabajo de Grado de la carrera Licenciatura en Informática, Departamento de Informática, Facultad de Ciencias Exactas, U.N.L.P. Abril 1996.