

ALGORITMO DE CODIFICACIÓN DE VOZ “CS-ACELP” PARA USARLO EN REDES DE DATOS

Karen Hernández R¹, Arturo Veloz G², José L Leyva M²
karenhr@cucsur.udg.mx / aveloz@gdl.cinvestav.mx / luis.leyva@cts-design.com

Introducción

La compresión de voz es un proceso que remueve información redundante para reducir el ancho de banda que se requiere para su transmisión. Algunos algoritmos, como el CS-ACELP, pierden parte de la información contenida en la señal original durante el proceso de compresión y descompresión. Sin embargo, la inteligibilidad del mensaje y el reconocimiento de la persona que habla se conservan. Podemos decir, más que perder información, con el algoritmo CS-ACELP obtenemos una buena relación entre calidad de la voz y ancho de banda (8 Kbps).

El codificador de voz CS-ACELP se basa en el modelo de codificación mediante la predicción lineal excitado por código (CELP) [1]–[2]. El proceso para el algoritmo CS-ACELP empezó en 1990 en el CCITT (hoy ITU-T) para aplicaciones inalámbricas. Se seleccionaron dos codificadores antes de crear el algoritmo en 1994: CS-CELP y ACELP. Ambos algoritmos se analizaron y de estos, se eligieron los mejores aspectos [3]–[5] para formar el algoritmo CS-ACELP. En el proceso de la estandarización del algoritmo CS-ACELP, se realizaron tres pruebas relacionados con la percepción del oído con la idea de caracterizar el desempeño subjetivo de la calidad de la voz. Los resultados indicaron que este codificador tiene un buen desempeño (alta calidad de voz sobre un canal de radio ruidoso y baja complejidad) y bajo retardo [6]–[8]. La calidad de la voz que se obtiene del algoritmo CS-ACELP es equivalente al producido por un codificador ADPCM de 32kbps.

Antecedentes

En la actualidad no se han negociado convenios entre fabricantes y proveedores de servicios de transmisión de datos para el transporte de voz sobre redes de datos; sin embargo, muchos fabricantes de equipos han desarrollado métodos propios para integrar voz sobre tramas de retransmisión (Frame Relay) [9] o redes IP. El algoritmo CS-ACELP (G.729) es un estándar recomendado por el Frame Relay Forum. Esta organización trabaja junto con los fabricantes de equipos de telecomunicación para asegurar la interoperabilidad de los equipos de los fabricantes.

Aunque ya existe una implementación de la recomendación G.729 un DSP C54xx ya se ha reportado en la literatura [10], decidimos implementarlo en un DSP con base en ciertas características para que se use en una red de datos. La idea es usar una compresión para que se atiendan a varios usuarios. Lo que se persigue con esto es reducir los costos de un sistema optimizando los recursos. Una tarjeta de compresión se puede instalar en una computadora que funcione como un punto de acceso o gateway entre varias computadoras y la red de paquetes, ver Figura 1. En este esquema, únicamente las comunicaciones externas de voz se comprimen (llamadas telefónicas sobre IP o sobre Frame Relay). La computadora gateway tiene las siguientes funciones: comprimir/descomprimir voz para/desde la red y controlar la comunicación entre las redes.

¹ Centro Universitario de la Costa Sur. <http://www.cucsur.udg.mx>. Av. Independencia Nacional 151, Autlán, 48900, Jal. México. TEL: 317-3825010. FAX: 317-3823200

² Centro de Investigaciones y de Estudios Avanzadas, Campus Guadalajara. <http://www.gdl.cinvestav.mx>. Av. Científica 1145 col. El Bajío, Zapopan, 45010, Jal. México. TEL: 33-37703700. FAX:33-37703709

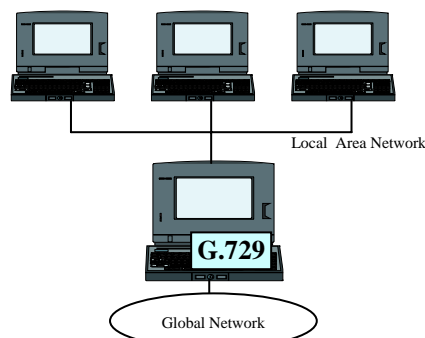


Figura 1. Diagrama con el sistema gateway

Funcionamiento del CS-ACELP

El codificador CS-ACELP maneja tramas de voz de 10 ms a una velocidad de muestreo de 8000 muestras por segundos (80 muestras por trama). Para realizar el análisis de las tramas, la trama se divide en 2 sub-tramas de 5 ms cada una. Las muestras de entrada y de salida se representan en PCM lineal de 16 bit. El método se basa en hacer una comparación entre la señal reconstruida y la señal original [11].

El principio de codificación se muestra en la Figura 2. La señal de entrada se introduce a un bloque de preprocesamiento donde es filtrada y escalada para evitar componentes indeseables de baja frecuencia o componentes de DC y reducir la posibilidad de desbordamiento en la aplicación de punto fijo (lenguaje C). El libro de análisis de predicción lineal LP calcula los coeficientes del filtro de síntesis (LPC). Este filtro es usado en la reconstrucción de un modelo del tracto vocal. Los coeficientes LPC son convertidos a coeficientes LSP (L_0, L_1, L_2, L_3) por una transformación lineal. Estos coeficientes son robustos al ruido y evitan la inestabilidad en los filtros del decodificador. La búsqueda de los parámetros para definir la señal de excitación se realiza en dos pasos: proceso de ciclo abierto y proceso de ciclo cerrado.

El objetivo en ciclo abierto es estimar una frecuencia fundamental preliminar (tono o pitch) por trama, basado en la señal de voz ponderada perceptualmente. Una vez encontrado el tono preliminar (T_{op}) se comienza el ciclo cerrado. El objetivo del ciclo cerrado es determinar el tono más exacto basándose en el T_{op} del ciclo abierto por medio (búsqueda del índice del libro) de minimización del error ponderado ϵ . Este procedimiento empieza ejercitando todos los índices del libro adaptable tratando de minimizar ϵ . El índice (P_0, P_1, P_2), que produce el mínimo error, es elegido y usado en el siguiente paso. En un segundo paso, el mismo tipo de procedimiento se lleva a cabo sobre el libro fijo (libro algebraico). De este procedimiento, se obtiene el índice del libro fijo (S_1, S_2, C_1, C_2). También son obtenidas las ganancias del libro adaptable y fijo basándose en sus índices y son cuantificados vectorialmente usando un libro de estructura conjugada (F_1, F_2, G_1, G_2). Los parámetros L_i, P_i, S_i, C_i, F_i y G_i son la salida de la información del CS-ACELP enviados cada 10ms con un retardo implícito de 15ms.

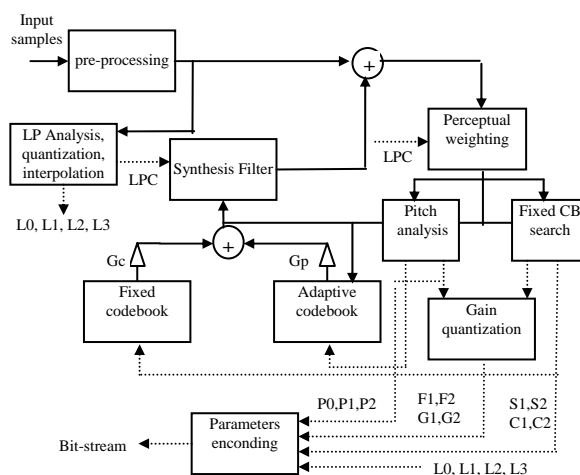


Figura 2. Codificador

El principio del decodificador se muestra en la Figura 3. Primero, se extraen los parámetros de la trama. Estos parámetros son decodificados para obtener la información correspondiente a los 10ms de trama de voz. Los coeficientes LSP cuantificados se interpolan y se convierten en coeficientes del filtro LP para cada sub-trama. Después se realizan los siguientes pasos:

- Se reconstruye la señal de excitación sumando los vectores de las tablas de código adaptable y fijo (escalados con sus respectivas ganancias).
- Se reconstruye la señal de voz filtrando la excitación obtenida por el filtro de síntesis LP.
- La señal de voz reconstruida pasa por una etapa de postprocesamiento (para mejorarla) y así obtener las muestras de entrada que tuvo el codificador.

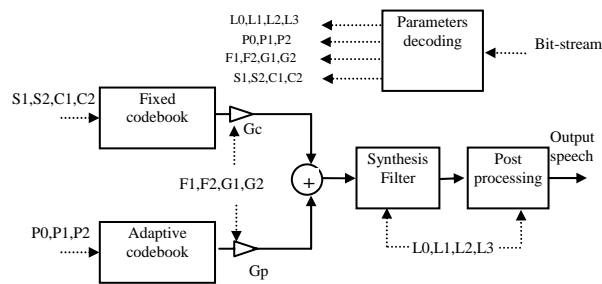


Figura 3. Decodificador

Implementación

El algoritmo se decidió implementar en el DSP C6201. El DSP seleccionado puede mantener 10 usuarios (canales de voz). La implementación consiste de dos partes: diseño de un DSP basado en una tarjeta y el software operacional (interfaz de usuarios, control de comunicación y el algoritmo G.279). La tarjeta está constituida por cinco bloques -ver Figura 4-; AD/DA converter, DSP processor, Memory (program and data), Address bus decoder y ISA bus interface (considering a dual port memory):

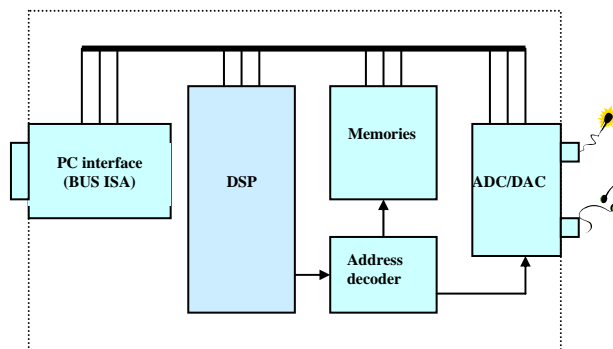


Figura 4. Diagrama de bloque de la tarjeta G729

El tiempo de ejecución del algoritmo G.279 se calculó por el número de operaciones que se realizaban para la ejecución de la comprensión y descomprensión de la voz en 10 ms (total 45500 operaciones) y con la frecuencia del DSP de 200MHz. Así que esto nos indica un tiempo de ejecución de 500 μ s y por lo tanto, que esta tarjeta puede atender hasta 20 usuarios (10ms).

El espacio de memoria requerido para instalar los programas G.279 se basó en las líneas de código que se obtuvieron después de la compilación del algoritmo (45000 líneas). Se consideraron dos bytes por línea, así que en total que se requieren de 88Kbytes para almacenar los programas de la recomendación G.279.

La interfaz del bus ISA usa una memoria de puerto dual. A través de esta memoria la PC y el DSP pueden intercambiar datos independientemente. La memoria se eligió considerando el número de bytes de voz comprimida (10 bytes) y el número de bytes para comprimir la voz (160 bytes) así como el tiempo requerido para la ejecución del proceso de compresión y descompresión. En total, se necesitan 1700 bytes para que se puedan atender a 10 usuarios.

Resultados

De las pruebas hechas en la PC se verificó el buen funcionamiento del algoritmo. Se introdujeron varios archivos de voz para comprimir; una vez que el algoritmo de compresión generó un archivo de información comprimida éste se utilizó para descomprimirlo. El tiempo de compresión fue de 10ms. Se escucharon ambos archivos de voz (los originales a comprimir y los descomprimidos) resultando iguales al oído. También se comparó el contenido de estos archivos mediante un programa en C; con el cual se comparó el archivo de voz original contra el reconstruido por el codificador G.729. Los archivos resultaron ser muy diferentes desde el punto de vista numérico. Sin embargo, el oído humano no percibe esta diferencia. Esto se concluye después de verificar las opiniones del grupo de personas que escucharon los archivos. Por otro lado, cuando la comparación se lleva a cabo desde el punto de vista espectral nos damos cuenta que los archivos son muy parecidos en vista de que la distribución de energía es similar. Esto se puede notar en los espectrogramas que se presentan en la Figura 5 (señal original) y en la Figura 6 (la señal reconstruida).

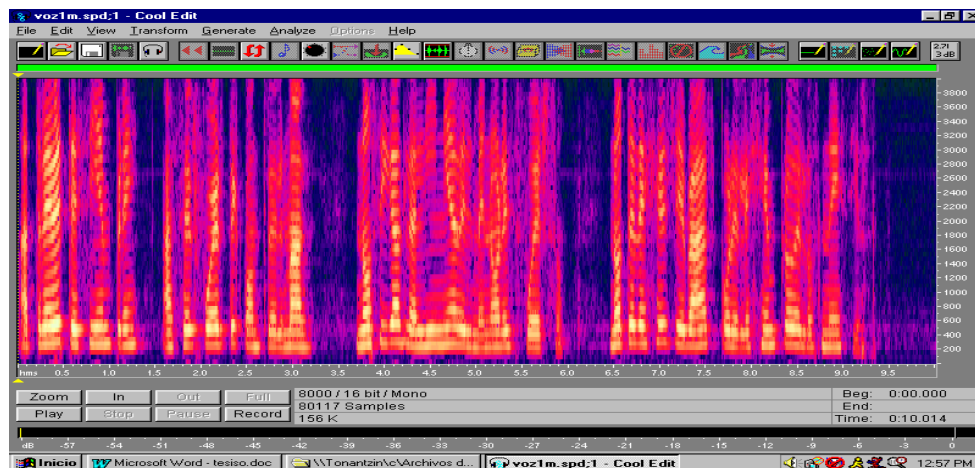


Figure 5. Espectrograma de la señal

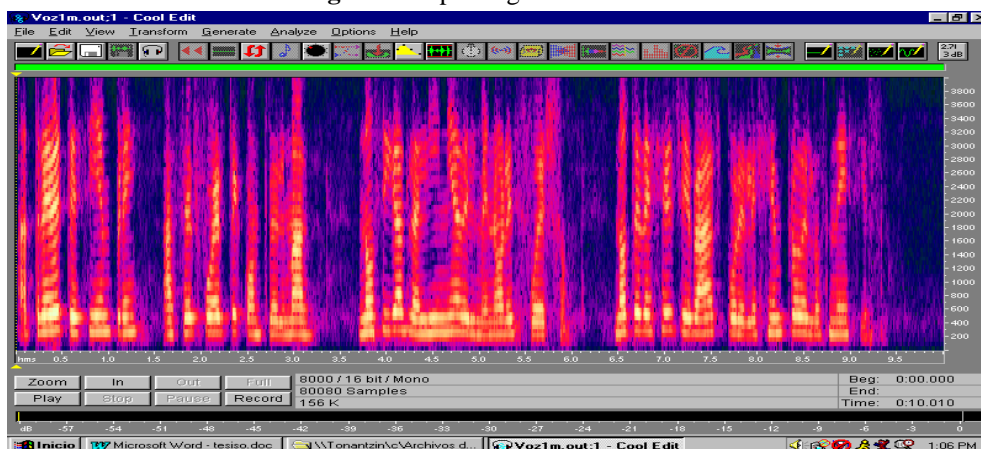


Figure 6. Espectrograma de la señal

Respecto al funcionamiento de la tarjeta, lo único que se pudo verificar fue la parte del DSP que va conectada al convertidor TLC32040; se simuló en Maxplus II para saber si se podían leer y escribir los datos del DSP sin problemas de tiempo.

Conclusiones

Fue necesario estudiar el DSP que se utilizaría para saber las partes que se debían considerar en el diseño de la tarjeta y poder especificar los requisitos de diseño. Una vez terminado el diseño y compilado el algoritmo, se procedió a adquirir los dispositivos que se iban a utilizar (memorias, osciladores, capacitores, etc.); se contactaron diversos distribuidores, nacionales e internacionales. Los dispositivos no siempre cumplen con todas las características requeridas para el diseño y se debe hacer una ponderación entre los posibles candidatos.

El diseño de la tarjeta se presentó para el uso del codificador CS-ACELP pero puede ser usada para cualquier otro tipo de codificador. Se concluye esto porque el algoritmo se carga en una memoria externa al DSP y aunque se consideró el tiempo de procesamiento de trama para 10 usuarios, puede ser modificado. La tarjeta también puede ser usada como un multicodecador; esto es, como un dispositivo que puede contener varios tipos de codificadores y dependiendo del sistema que se use, seleccionar alguno en particular. Es importante aclarar que para llevar a cabo lo anterior, se tendría que cambiar las dimensiones de las memorias que se consideraron en el diseño.

Trabajo futuro

La tarjeta se diseñó y se dejó lista con los archivos de fabricación para que se mande a fabricar. Sin embargo, algunos de los puntos que faltan por hacer son los siguientes:

- Hacer un programa en la PC para conectar la tarjeta (inicialización, interrupciones, etc).
- Modificar la tarjeta para que pueda ser insertada a una PC dedicada (sin convertidores A/D-D/A, sin micrófono y bocina).
- Fabricar la tarjeta diseñada, ensamblarla y verificarla.
- Verificar el tiempo de ejecución del algoritmo considerando los retardos que se generen en la red.
- Analizar si es posible reducir los libros de código, según el idioma seleccionado

Referencias

- [1] M. R Schroeder and B. S. Atal. (1985). Code-Excited Linear Prediction (CELP): high quality speech at very low bit rates. *Proc. IEEE Int. Conf. Acoustic, Speech, Signal Processing., Tampa, FL:* 937-940.
- [2] A.M Kondo. (), "Digital Speech: coding for low bit rate communication systems", John Wiley and Sons, pp. 174-209, 1994.
- [3] Kataoka, T. Moriya and S. Hayashi. (1993). An 8kbit/s speech coder based on conjugate structure CELP. *Proc. IEEE Int. Conf. Acoustic, Speech, Signal Processing II:* 592-595.
- [4] Kataoka, T. Moriya and S. Hayashi. (1996). An 8kbit/s Conjugate Structure CELP Speech Coder. *IEEE Trans. Speech Audio Processing.* **6:** 401-411.
- [5] J. Adoul, P. Mabillean, M. Delprat and S. Morissette. (1987,). Fast CELP coding based on algebraic codes. *Proc. IEEE Int. Conf. Acoustic, Speech, Signal Processing.* 1953-1956.
- [6] M. E. Perkins et al. (1996). Characterizing the Subjective Performance of the ITU-T 8 kbps Speech Coding Algorithm (ITU-TRec. G.729). *IEEE Communications Magazine.*
- [7] R.V. Cox (1996). Three New Speech Coders from the ITU Cover a Range of Applications. *IEEE Communication Magazine.*
- [8] V.K. Varma (1993). Testing Speech Coders For Usage in Wireless Communications Systems. *Proc. IEEE Wksp. Speech Coding for Telecommunication, Sainte-Adele, Québec, FL:Canada, October 13-15 pp.* 93-94.
- [9] The Frame Relay Forum, "A Discussion of Voice over Frame Relay October 1996", download: www.frforum.com/40000.

- [10] R. Salami, C. Laflamme, J-P. Adoul, A. Kataoka, S. Hayashi, T. Moriya, C. Lamblin, D. Massaloux, S. Proust, P. Kroon, Y. Shoham [1998]. Design and Description of CS-ACELP: A Toll Quality 8 kb/s Speech Coder. *IEEE Transactions on Speech and Audio Processing*, **6**(2)
- [11] ITU-T Recommendation G.729 (1996). Coding of Speech at 8 kbps using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP).