

Implementación del codificador de voz G.722 sobre el DSP TMS320F206

Ana Isabel González-Tablas Ferreres (*), Belén Ruiz Mezcuca (*),
María Soledad Torres Guijarro(**), Pablo Fernández-Cid (**)

(*) IRIS: Laboratorio de Integración de Sistemas, Departamento de Informática,
Universidad Carlos III, c/ Butarque, 15 - 28911 Leganés, Madrid (ESPAÑA).
Teléfono: (34-1) 624 91 04. Email: aigonzal@inf.uc3m.es

(**) Dpto. Electrónica y Teoría de Circuitos, Universidad Europea de Madrid – CEES
Urbanización el Bosque, s/n – 28672 Villaviciosa de Odón, Madrid

Abstract

En este trabajo se ha implementado el codificador de voz especificado en la Recomendación ITU-T G.722 sobre un módulo hardware basado en el DSP TMS320F206.

Para evaluar la implementación se han realizado pruebas subjetivas y objetivas con señales de voz, y las pruebas con vectores digitales propuestas por el ITU-T, obteniéndose resultados muy satisfactorios.

1 Introducción

Actualmente el campo de las telecomunicaciones está experimentando una gran revolución. Una de las tecnologías base en esta revolución es el procesado de señal. Dentro de este campo, se ha desarrollado un interés creciente sobre las tecnologías de codificación de voz, fomentado por la disminución de la capacidad disponible para las comunicaciones en ciertos mercados y la cada vez más imprescindible exigencia de alta calidad.

2 Breve descripción del codificador ITU-T G.722

El estándar del ITU-T (Recomendación ITU-T G.722) de comprensión de voz de 7 KHz a 64 Kbps codifica voz de banda ancha a 64 Kbps con una alta calidad para aplicaciones tales como telefonía de banda ancha (RDSI), teléfonos manos libres, teleconferencia y canales de comentaristas para radiodifusión. Aunque no con la misma calidad, también pueden comprimirse y descomprimirse otras señales de audio tales como música.

Se utiliza un algoritmo de codificación en subbandas con ADPCM (Codificación por modulación de pulsos adaptativa y diferencial). Para separar las subbandas se utilizan filtros especulares en cuadratura (QMF).

Para adecuarse a las aplicaciones sobre RDSI, el estándar G.722 proporciona la posibilidad de transmitir datos de baja velocidad sobre el canal de 64 Kbps. Para ello define tres modos de operación del codificador: modo 1, con 64 Kbps de voz; modo 2, con 56 Kbps de voz y 8 Kbps de datos; y modo 3, con 48 Kbps de voz y 16 Kbps de datos.

3 Descripción de la plataforma: hardware y software

La tarjeta de depuración sobre la que se ha desarrollado el codificador es el módulo Easy-AA10ⁱ. Este módulo está diseñado para desarrollar aplicaciones de medio coste siendo su corazón el DSP TMS320F206 con 4K5 Words de memoria RAM y 32 KWords de memoria FLASH

de programa. El procesador trabaja en coma fija de 16 bits a 40 MHz (20 MIPS). La tarjeta incluye un CODEC de un canal de banda ancha (50 a 12.500 Hz) para permitir la integración en entornos analógicos y 64 KWords de memoria externa de datos.

La ubicación final del software desarrollado es el módulo codificador de voz easy chip® ECH8110-20ⁱⁱ.

El software utilizado para el desarrollo del código fuente en alto nivel (lenguaje C) es Microsoft Developer Studio. Para la integración del código en alto nivel se ha utilizado el entorno de desarrollo integrado CodeComposer. Finalmente, las pruebas se han evaluado con Matlab 5.0.

4 Procedimientos para la realización del trabajo

El trabajo llevado a cabo se puede desglosar en los siguientes pasos:

- i) Estudio del algoritmo y obtención de un ejecutable de referencia en coma flotanteⁱⁱⁱ.
- ii) Desarrollo del software en coma fija en alto nivel.
- iii) Desarrollo del software en bajo nivel.
- iv) Optimización del código en alto nivel.
- v) Integración.

5 Resultados y conclusiones

Se han realizado tres tipos de pruebas para evaluar la implementación:

- i) Pruebas objetivas de medidas de relación señal ruido (SNR) y señal ruido segmental (SnrSeg) con ocho ficheros de voz y música muestreados a 16 KHz y procesados por el conjunto del codificador y decodificador encadenados en cada uno de los tres modos de funcionamiento posibles. La *Tabla I* proporciona una relación de los ficheros utilizados. Las *Figuras 1* y *2* muestran los resultados obtenidos en dBs.
- ii) Pruebas subjetivas. Las pruebas subjetivas realizadas han consistido en hacer escuchar a una conjunto de sujetos^{iv} una combinación de las dos señales de voz existentes (señal de voz original, A, y señal de voz sintetizada, B) resultantes del proceso en modo 1; es decir, una secuencia AAB ó ABA. El sujeto debe decidir cuál es la señal de voz sintetizada conociendo que la primera es la señal original y que las dos siguientes pueden estar en cualquier orden. En

la *Figura 3* se muestran los resultados obtenidos.

iii) Pruebas del ITU-T. El ITU-T define en el Apéndice II de la Recomendación G.722 los vectores de prueba de entrada y salida así como su formato para dos configuraciones. En ambas configuraciones se eliminan los filtros QMF. En la configuración 1 se prueba el codificador y en la configuración 2 el decodificador, en los tres modos. En la *tabla II* se muestran los resultados de las pruebas en configuración 1 para el Test1 (tonos, DC, y ruido blanco) y Test2 (control de la saturación). En configuración 2: en la *Tabla III* se muestran los resultados tomando como entrada la salida de Test1; en la *Tabla IV* se muestran los resultados tomando como entrada la salida de Test2; y en la *Tabla V* se muestran los que toman como entrada una secuencia de test artificial considerada salida no habitual del codificador.

FICHERO	CONTENIDO	DURACIÓN
Spch.wav	Voz femenina de habla inglesa	6,48 seg.
Habla.wav	Voz femenina de habla inglesa	3,84 seg.
Dtfm.wav	Tonos	2,2 seg.
Clicks.wav	Voz femenina de habla inglesa con clicks de fondo	6,48 seg.
Clgt.wav	Rasgueo de guitarra	11,94 seg.
Bat.wav	Sonido de batería	2,1 seg.
Abba.wav	Música instrumental	29,73 seg.
Agustin.wav	Voz masculina de habla española	11,62 seg.

Tabla I: Relación de ficheros utilizados para las pruebas

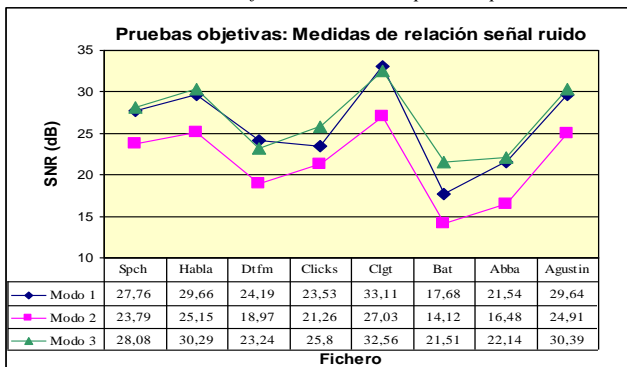


Figura 1: Medidas de SNR

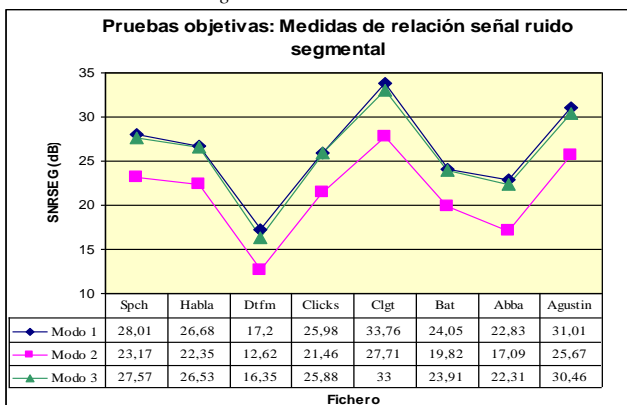


Figura 2: Medidas de SNRSEG

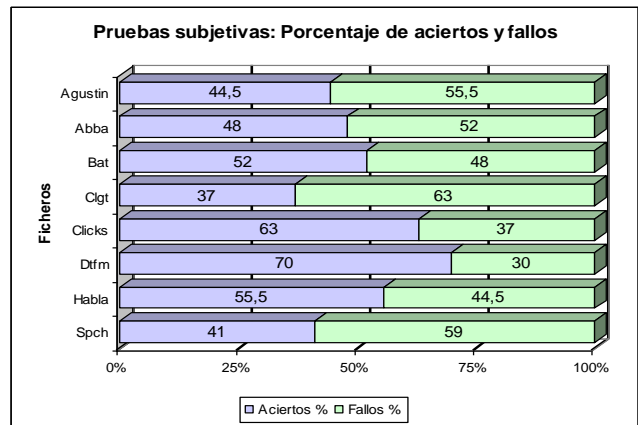


Figura 3: Relación de Aciertos y Fallos ante decisión AAB-ABA

	Test1	Test2
Snr	7,59 dB	73,06 dB
SnrSeg	9,55 dB	64,80 dB

Tabla II: Resultados de SNR y SNRSEG para la configuración 1.

	Modo1 BB	Modo2 BB	Modo3 BB	Banda Alta
Snr	8,38 dB	7,83 dB	9,17 dB	9,61 dB
SnrSeg	20,73 dB	18,40 dB	24,45 dB	19,45 dB

Tabla III: SNR y SNRSEG para la configuración 2 (Test1)

	Modo1 BB	Modo2 BB	Modo3 BB	Banda Alta
Snr	10,27 dB	10,07 dB	8,20 dB	9,24 dB
SnrSeg	51,15 dB	46,80 dB	32,09 dB	39,99 dB

Tabla IV: SNR y SNRSEG para la configuración 2 (Test2)

	Modo1 BB	Modo2 BB	Modo3 BB	Banda Alta
Snr	14,04 dB	14,77 dB	13,81 dB	13,36 dB
SnrSeg	61,03 dB	61,63 dB	61,96 dB	49,47 dB

Tabla V: SNR y SNRSEG para la configuración 2 (Test artificial)

El codec implementado obtiene unos muy buenos resultados tanto en las pruebas objetivas como las subjetivas. Las medidas de señal ruido y señal ruido segmental son excelentes para señales de voz tanto femeninas como masculinas. En el caso de los resultados de las pruebas subjetivas se puede observar que para las señales de voz se han llegado a valores cercanos al 50% en aciertos y fallos, indicando que resulta casi indistinguible la señal original de la señal sintetizada. Para las pruebas de señales no estrictamente de voz también se han obtenido muy buenos resultados, ligeramente más bajos, siendo el resultado coherente con el diseño del codificador. Por último, también pasa con éxito las pruebas del ITU-T, siendo peores los resultados referentes al Test1 por el gran número de transiciones existentes en el vector de prueba.

6 Referencias

- [1] González-Tablas Ferreres, A. I. "Implementación del compresor de voz G.722 sobre el DSP TMS320C206F", octubre 1999
- [2] "ITU-T Recommendation G.722" Fascículo III.4 del "Blue Book".
- [3] Maitre, X. "7 KHz Audio Coding within 64 kbit/s" IEEE Journal on Selected Areas in Communications, Vol. 6, Nº 2, Feb. 1988

ⁱ El modulo Easy-AA10 ha sido desarrollado por DATATECH Sistemas Digitales Avanzados.

ⁱⁱ El módulo easy chip® ECH8110-20 ha sido desarrollado por easy tools, s.l.

ⁱⁱⁱ El código en coma flotante se obtuvo a partir de uno desarrollado por el Grupo de Teoría de la Señal del Departamento de Tecnologías de las Comunicaciones de la Universidad de Vigo.

^{iv} Población de 27 sujetos de edades entre 12 y 80 años.