

EVALUACION DE LA CALIDAD SUBJETIVA DEL CODIFICADOR DE BAJO RETARDO G.728 EN APLICACIONES DE AUDIO

Angel Fernández Herrero

Departamento de Ingeniería Electrónica
Universidad Politécnica de Madrid
angelfh@die.upm.es

Fco. Javier Casajús Quirós

Dpto. de Señales, Sistemas y Radiocomunicaciones
Universidad Politécnica de Madrid
javier@gaps.ssr.upm.es

ABSTRACT

In this work we apply the Perceptual Audio Quality Measure (PAQM) to the objective evaluation of the perceptual quality obtainable from the G.728 low-delay CELP voice coder in its application to audio coding.

1. INTRODUCCIÓN

La proliferación de las aplicaciones de comunicación digitales ha traído consigo una intensa investigación en algorítmica de codificación para aplicaciones de voz y audio, que resulta cada vez más compleja a medida que los medios de implementación disponibles aumentan su capacidad de proceso.

En este contexto, los métodos tradicionales de evaluación de calidad están basados en extensas pruebas subjetivas con oyentes humanos. Pero estos procedimientos son caros, difíciles de reproducir y de larga ejecución. Por ello, los últimos diez años han visto un importante trabajo en el campo del desarrollo de medidas de evaluación de la calidad subjetiva.

El objetivo de este trabajo ha sido la aplicación de la medida presentada en [2], conocida como PAQM (*Perceptual Audio Quality Measure*), a la evaluación de la calidad del codificador de voz G.728 en su empleo para aplicaciones de audio de baja calidad, como puede ser el hilo musical.

2. EL CODIFICADOR G.728

A pesar de ser un codificador diseñado en torno al modelo CELP (*Code Excited Linear Prediction*), el G.728 tiene varias características diferenciadoras del mismo, que surgen del principal requerimiento para el que fue concebido [3], su bajo retardo de codificación (2 ms).

En el G.728 se ha eliminado el predictor de tono (redundancia a largo plazo) tradicional del modelo CELP, aumentando en su lugar muy significativamente la longitud del análisis LP (*Linear Prediction*), que modela habitualmente la redundancia a corto plazo, desde 10 a 50 coeficientes. Este análisis se realiza de forma *backward-adaptive*, es decir, con base en muestras previamente cuantificadas, de manera que no es necesario enviar

los coeficientes resultantes, porque el decodificador puede reproducir el proceso de forma idéntica.

De este modo, el G.728 resulta menos específico para voz, ya que no asume ninguna periodicidad en la señal de entrada, y es en gran medida un codificador de forma de onda. Esto hace que funcione muy satisfactoriamente en aplicaciones no estrictamente relacionadas con la voz [3], como pueden ser las señales musicales, si bien con la restricción fundamental de su ancho de banda, limitado a menos de 4 KHz.

3. LA MEDIDA PAQM

Las medidas clásicas de diferencia entre señales, como por ejemplo el MSE (*Mean Squared Error*), no son adecuadas en el marco de sistemas no lineales, como sucede con los actuales codificadores de voz y audio, y tampoco si el destinatario de las señales es un receptor tan complejo como el oído humano.

Lo que se necesita en este caso es un procedimiento algorítmico de comparación que, evitando las baterías de tests subjetivos, proporcione medidas de diferencia o parecido entre señales muy correladas con aquéllos.

Ese es el objetivo de la medida PAQM [2], que se emplea para evaluar la calidad subjetiva de dispositivos de audio mediante la realización de una comparación entre las señales de entrada y salida del mismo en base a criterios basados en los principios de funcionamiento del sistema auditivo humano.

La medida se construye como un proceso en dos etapas. Primero la construcción para cada señal de una representación interna o psicoacústica de la misma y, a continuación, la evaluación de una fórmula comparativa entre representaciones. Se determina de este modo la diferencia perceptual entre las señales y no se precisa para nada la señal diferencia.

La representación interna se construye en varios pasos que definen un modelo simplificado de audición:

- 1) Enventanado de la señal en tramos de unos 40 ms con ventanas de Hanning solapadas un 50% y cálculo de la densidad de potencia para cada una de ellas.
- 2) Cambio de la escala de frecuencia f (Hz) a escala de tonalidad z (bark) empleando la transformación:

$$f = \sinh(z/6)$$

- 3) Aplicación de una dispersión temporal, para dar cuenta de la influencia causal de unas ventanas sobre las siguientes.
- 4) Aplicación de una dispersión espectral, como la que tiene lugar en la recuperación de la información auditiva en la membrana basilar, mediante convolución con una función apropiada, cuya forma tiene dependencia con la tonalidad y la intensidad de la excitación.
- 5) Aplicación de una fórmula de compresión específica basada en el umbral de audición y dependiente de la tonalidad.

Para la comparación de representaciones se escala una de ellas (de forma independiente en tres rangos diferentes de la tonalidad) buscando la mayor coincidencia posible, se obtiene la diferencia entre ambas, se integra en el eje de tonalidad (para obtener la potencia total de excitación en cada ventana) y se pondera en el tiempo. El logaritmo del valor obtenido es la PAQM.

4. CALIBRADO DE LA MEDIDA

Para evaluar la medida PAQM y calibrar su relación con el MOS (*Mean Opinion Score*) de acuerdo con las curvas presentadas en [2], se aplicó a una señal de audio [1] mezclada con ruidos blanco y rosa de diversas potencias. Los resultados se muestran en la figura 1, donde queda patente la menor influencia perceptual del ruido coloreado para igual relación señal-ruido.

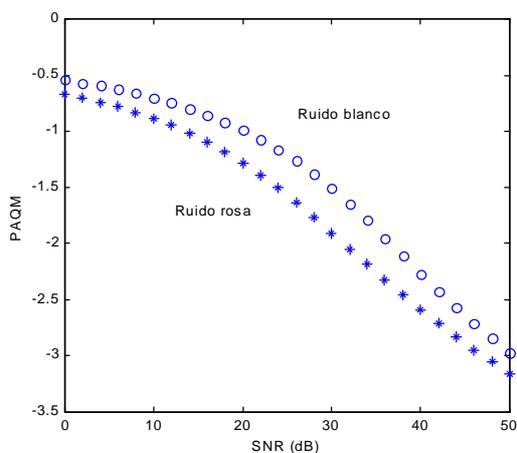


Figura 1. Evaluación de la medida PAQM.

Es también destacable el hecho de que las curvas representadas muestren una tendencia a la saturación para valores pequeños de la relación señal-ruido, como corresponde a una medida perceptual correlada con las respuestas reales en los tests subjetivos, donde el nivel de desagrado alcanza un máximo superada una cierta cantidad de perturbación.

5. RESULTADOS

El procedimiento de medición obtenido se aplicó a la misma muestra de audio [1] codificada mediante una implementación

del algoritmo G.728. Para comparar los resultados se realizaron también medidas con otros codificadores de voz estandarizados, como son el G.721 a 32 Kbit/seg. y el G.729 (8 Kbit/seg) [5].

El G.721 es un algoritmo de baja complejidad basado en técnicas ADPCM (codificación de forma de onda), por lo cual son de esperar prestaciones satisfactorias con señales arbitrarias. Entre los requerimientos iniciales del G.728 se encontraba la obtención de una calidad de voz comparable a la del G.721 a 32 Kbit/seg en similares condiciones.

El G.729 [4] se adhiere estrictamente al modelo CELP, lo que le convierte en poco adecuado para la codificación de señales que no se ajusten a ese modelo de producción, como sucede con las señales musicales.

Los resultados obtenidos pueden verse en la tabla 1, donde se muestran también valores aproximados de MOS extraídos mediante las curvas presentadas en [2].

	PAQM	MOS
G.721	-1.88	4.5
G.728	-1.84	4.4
G.729	-1.37	3.5

Tabla 1. Evaluación de codificadores.

6. CONCLUSIONES

Se ha realizado en este trabajo una evaluación del estándar de codificación de voz G.728 (bajo retardo, 16 Kbit/seg.) para su empleo en aplicaciones de audio de baja calidad (limitadas a una banda menor de 4 KHz), como puede ser el hilo musical.

Las medidas se han efectuado de forma objetiva con el procedimiento PAQM, y los resultados obtenidos demuestran que la calidad perceptual que se puede obtener con el G.728 es similar a la de un codificador de forma de onda (G.721) en las mismas condiciones, estando alejada hasta en un punto MOS de un codificador CELP típico (G.729).

7. REFERENCIAS

- [1] Gyorgy Ligeti, *Lux Aeterna*.
- [2] John G. Beerends, Jan A. Stemerdink, "A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation", *J. Audio Eng. Soc.*, December-1992.
- [3] Juin-Hwey Chen, Richard V. Cox, Yen-Chun Lin, Nikil Jayant, Melvin J. Melchner, "A Low-Delay CELP Coder for the CCITT 16 Kb/s Speech Coding Standard", *IEEE J. on Selected Areas in Comm.*, June-1992.
- [4] R. Salami et al., "Design and Description of CS-ACELP: A Toll Quality 8 Kb/s Speech Coder", *IEEE Trans. on Speech and Audio Processing*, March-1998.
- [5] Richard V. Cox et al., "Scanning the Technology. On the Applications of Multimedia Processing to Communications", *Proc. of the IEEE*, May-1998.