

EVALUACIÓN OBJETIVA DE CODIFICADORES DE VOZ MEDIANTE EL MÉTODO PSQM

Referencia pacs: 43.71.+m

Héctor Gustavo Giuliano
Gerencia Laboratorios
Telefónica de Argentina
Reconquista 179
1702 Ciudadela
República Argentina
Tel: 54 11 4332-8380
Fax: 54 11 4653-3382
E-mail: giulianh@telefonica.com.ar

ABSTRACT

Conventional methods utilized for the evaluation of the voice quality in telephony networks are not applicable to the new equipment that possess speech codecs with voice compression, due to inherent non-linear behavior, time-variant characteristics and loss of redundancy. An objective method that permits to evaluate this type of systems necessarily must take into account perceptual criteria in the design and must permit to utilize actual signals to perform the tests. The method PSQM "Perceptual Speech Quality Measure" developed by Institute KPN of Holland was adopted by International Telecommunication Union for the elaboration of Recommendation P.861, later of an analysis of others methods.

RESUMEN

Los métodos clásicos utilizados para la evaluación de la calidad de voz de los sistemas convencionales de telefonía no son aplicables a los nuevos equipos que poseen codificadores con compresión de voz, debido a sus inherentes características de no linealidad, variancia temporal y pérdida de redundancia. Un método objetivo que permita evaluar este tipo de sistemas, necesariamente deberá tener en cuenta criterios perceptuales en su diseño y deberá permitir utilizar señales reales como señales de prueba. El método PSQM "Perceptual Speech Quality Measure", desarrollado por el Instituto KPN de Holanda, fue adoptado por la Unión Internacional de Telecomunicaciones, luego del análisis de varios métodos, para la elaboración de la Recomendación P.861.

ANTECEDENTES

Las redes telefónicas han ido creciendo en complejidad con el transcurso de los años. Primero fue la introducción de los códecs PCM de 8 bits con compresión logarítmica (leyes μ y A) en las tarjetas de abonado de las centrales de conmutación, luego la incorporación de la compresión ADPCM para ciertos trayectos de transmisión, y en la actualidad, se está pensando en utilizar codificadores del tipo híbridos, como los de la familia CELP, para lograr un mejor aprovechamiento del ancho de banda disponible. La incorporación de estos sistemas conlleva una reelaboración de las técnicas de medición utilizadas con el fin de garantizar la calidad de servicio en las comunicaciones establecidas entre dos extremos. Estas técnicas están en su gran mayoría en etapa de estudio para demostrar su validez, con el agravante de que mucho de los trabajos realizados no son de aplicación general debido a que este tipo de codificadores son

dependientes de las características del idioma, imposibilitando la extrapolación de los resultados obtenidos.

Los factores de degradación de la calidad en este tipo de redes se relacionan básicamente con el tipo de códec empleado y con la tecnología de transmisión adoptada para el transporte.

La codificación introduce un tipo de distorsión que debe ser evaluada a través de un método que tenga en cuenta las características perceptuales de la audición, existiendo consenso en que el método PSQM posee una buena correlación con las determinaciones subjetivas del MOS. Por el contrario, la utilización de una red de transmisión del tipo de "paquetes" (ATM o IP) introduce retardos fijos y variables que, sumados a la pérdida de paquetes, los errores de bit y el recorte silábico, generan un impacto sobre la intelegibilidad del habla que aún está siendo evaluado. (El método MNB, "Measuring Normalizing Blocks", está siendo estudiado en estos momentos para evaluar la degradación introducida por esta etapa.)

EL MÉTODO PSQM

El método PSQM, "Perceptual Speech Quality Measure", fue desarrollado por el Instituto KPN de Holanda y fue adoptado por la Unión Internacional de Telecomunicaciones para la elaboración de la Recomendación P.861 en el año 1998 [1]. Se analizaron en su oportunidad diferentes métodos, siendo el adoptado el que mejor resultados obtuvo en el estudio de correlación con las pruebas subjetivas de calidad de códecs vocales.

En la figura 1 se muestra un diagrama conceptual del método PSQM, en el que se observa la utilización de un *modelo perceptual* y de un *modelo cognitivo* para la determinación de la calidad de voz. La señales de entrada del modelo consisten en una señal de referencia y la correspondiente señal codificada (las señal obtenida a la salida del códec cuando se lo excita con la señal de referencia). Como señal de entrada se puede utilizar tanto voz artificial (Recomendación UIT-T P.50) como voces reales almacenadas, por ejemplo, en archivos .WAV.

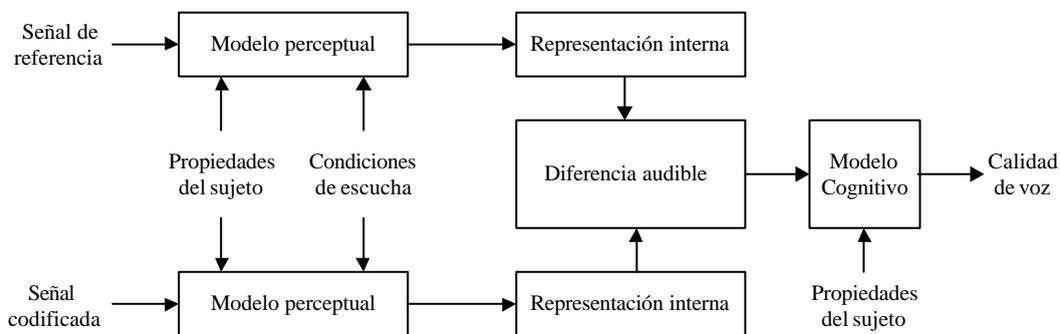


Figura 1: esquema conceptual del método PSQM.

Como modelo perceptual se utiliza el modelo ampliamente aceptado desarrollado por E. Zwicker [2], que transforma la representación del fenómeno en el mundo físico (dBspl vs. Hz) al sensorial interno (sones vs. bark). Como única diferencia, se ha disminuido el exponente de compresión para que el modelo se ajuste mejor a las señales de voz.

El modelo cognitivo adquiere por lo general la topología de una red neuronal, pero en este caso se adoptó un esquema más sencillo basado en una única entrada. Esta entrada es la diferencia de las representaciones internas de las señales de referencia y de prueba. A esta diferencia (asimetría) se le asigna un factor de peso de acuerdo al criterio de que resulta más molesta la aparición de energía adicional en una determinada banda (distorsión) que la pérdida de ella, ya que en este último caso la señal resultante sigue siendo un fenómeno auditivo coherente. Matemáticamente, y teniendo en cuenta la nomenclatura del esquema de la figura 2 y de la referencia 1:

$$N_i = \sum_{j=1}^{N_b} N_i[j] \cdot C_i[j] \cdot \Delta z$$

$$N_i[j] = |Ly_i'[j] - Lx_i[j]| - 0,01$$

siendo:

Cuando $PHx_i[j]$ y $PHy_i[j]$ (potencias corregidas de la trama i en la banda j de las señales de

$$C_i[j] = \left(\frac{PHy_i[j] + 1}{PHx_i[j] + 1} \right)^{0,2}$$

referencia y de prueba) no alcanzan los 20 dB por encima del umbral absoluto de audición en dicha banda, $C_i[j]$ se fija en 1. Por otro lado, el valor máximo se debe limitar a 2.

OPTIMIZACIÓN DEL ALGORITMO. EL MÉTODO PSQM+

Con el fin de optimizar el comportamiento del algoritmo para señales que poseen alta distorsión y saturación por recorte, se ha propuesto una modificación del modelo cognitivo utilizado [3]. Esta modificación consiste en el agregado de un nuevo factor de peso que tiene en cuenta la relación entre las potencias totales de las señales de referencia y de prueba. Matemáticamente:

$$N_i' = (S_i^+)^{0,25} N_i$$

siendo:

$$S_i^+ = \text{MIN} \left(\frac{Px_i' + 1}{Py_i' + 1}, 10000 \right)$$

RELACIÓN CON LA ESCALA DE EVALUACIÓN SUBJETIVA

Es posible encontrar una función de transferencia que permita vincular el resultado del algoritmo con el valor MOS obtenido de las pruebas subjetivas. Sin embargo esta función no es única ya que depende en gran medida del idioma utilizado para la evaluación. No se dispone aún de los datos suficientes que permitan determinar esta función de transferencia para el idioma castellano.

RESULTADOS Y CONCLUSIONES

Actualmente están siendo realizadas en el laboratorio pruebas de calidad de servicio sobre sistemas telefónicos que utilizan diferentes algoritmos de compresión de voz (Recomendaciones UIT-T G.711, G.726, G.728 y G.729). Se espera para la fecha de presentación de la comunicación, disponer de los resultados de las mismas y de algunas conclusiones preliminares sobre la validez del método sobre redes reales con tráfico en idioma castellano.

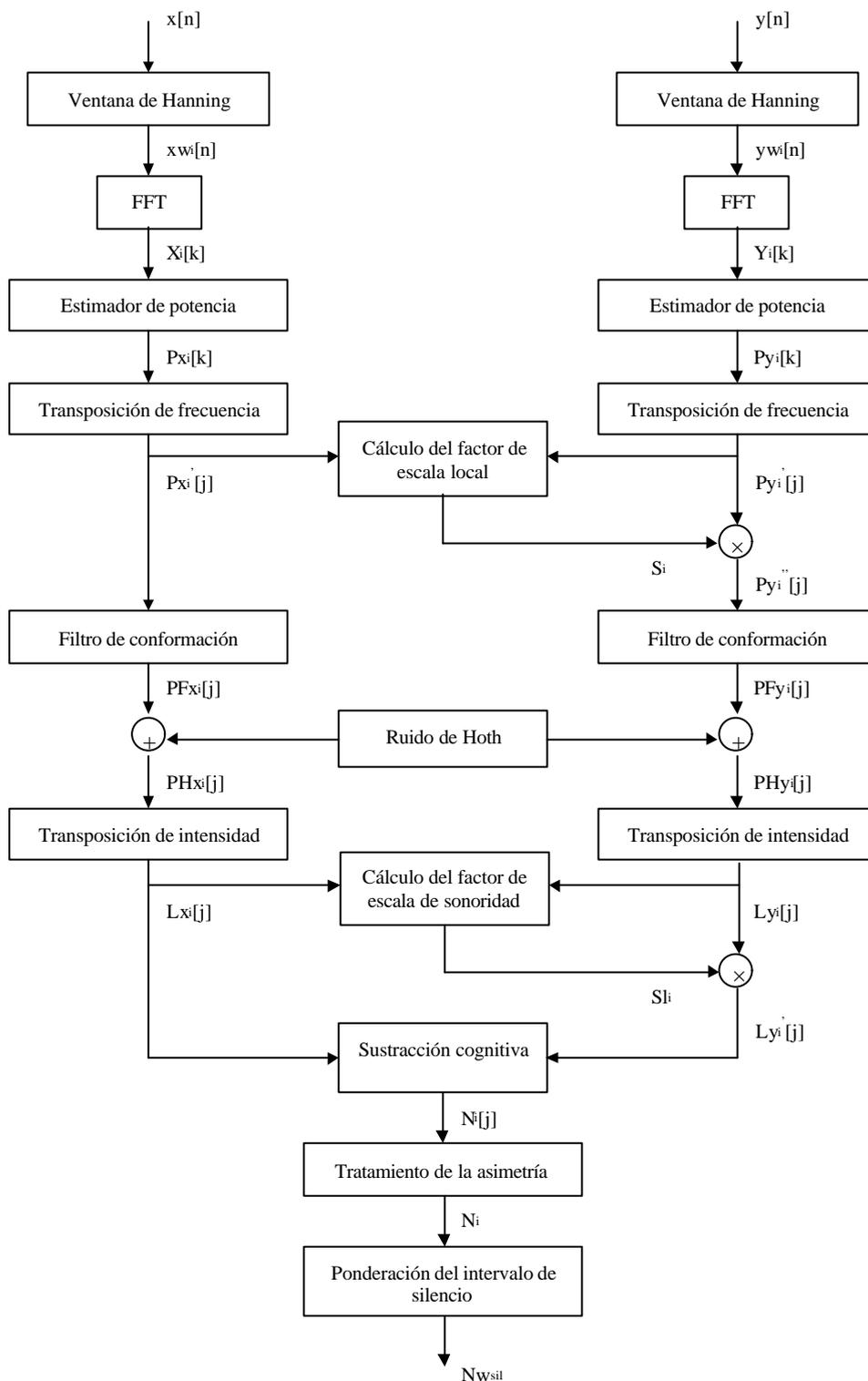


Figura 2: diagrama de flujo del algoritmo PSQM.

REFERENCIAS

[1] Recomendación UIT-T P.861 "Medición objetiva de la calidad de los códecs vocales de banda telefónica (300-3400 Hz)", febrero de 1998.
 [2] E. Zwicker, H. Fastl, "Psychoacoustics, Facts and Models", segunda edición, Springer, 1999.
 [3] J. G. Beerends, "Improvement of the P.861 perceptual speech quality measure", UIT-T, Study Group 12, Contribution 20, diciembre de 1997.