CODIFICACION APVQ DE VOZ EN BANDA ANCHA USANDO ASIGNACION DINAMICA DE BITS

Josep M. SALAVEDRA, Enrique MASGRAU**, Antoni CERVANTES*.*

** Departament de Teoria del Senyal i Comunicacions. Universitat Politècnica de Catalunya.*
c/ Gran Capità s/n , Campus Nord , mòdul D5. 08034 BARCELONA
Tfno: +34.3.4017404. Fax: +34.3.4016447. E-mail: mia@tsc.upc.es

*** Departamento de Ingeniería Eléctrica e Informática. C.P.S. Ingenieros. Universidad de Zaragoza.*
Tfno: 34.76.519892. Fax: 34.76.512932. E-mail: masgrau@mcps.unizar.es

ABSTRACT

This paper describes a coding scheme for broadband speech. It can be seen as a vectorial extension of a conventional ADPCM encoder. In this scheme, signal vector is formed with one sample of the normalized prediction error of each subband and then it is vector quantized. It combines the advantages of the scalar prediction and those of vector quantization (VQ). We handle the high vector dimensionality by using a multi-VQ. It requires a previous subvector division and an adequate bit assignment among them. This scheme shows a high capacity to drive large dynamic range signals like broadband speech. Predictor and codebook dessigns are discussed. Some results about speech prediction and coding are reported.

1. INTRODUCCIÓN

La combinación de las técnicas de división en subbandas con cuantificación vectorial y predicción adaptativa proporciona muy buenos resultados en codificación de señal de voz de banda estrecha (4kHz) a velocidades medias de 1 bit/muestra (8 Kbps). Un ejemplo de este tipo de codificadores es el denominado APVQ (Adaptive Predictive Vector Quantization) [1] que consiste, básicamente, en una división de la señal de voz en 8 subbandas de 500 Hz, cada una mediante un banco de filtros QMF, seguido de una cuantificacion ADPCM "backward" de cada una de las bandas, con la particularidad de que la cuantificación del error de predicción en cada una de las subbandas se realiza mediante una cuantificación vectorial (VQ), de tal modo que cada uno de estos errores de predicción constituye una de las componentes del vector de entrada al VQ. Es decir, en vez de cuantificar el error de predicción de cada una de las subbandas de forma independiente mediante un cuantificador escalar, se cuantifican en bloque mediante un VQ. Además, este VQ es adaptativo, en el sentido de que los errores de predicción son previamente normalizados en ganancia mediante una estimación "backward" de ésta, o lo que es lo mismo, se hace uso de una cuantificación VQ ganancia-forma adaptativo.

Como se detalla en la referencia [1], a esta velocidad de transmisión moderada de 1 bit/muestra, la predicción adaptativa no aporta ninguna ventaja en las bandas 5 a 8 (por encima de 2 kHz), con lo que puede prescindirse de ella en estas bandas. Ello es debido a que el error de cuantificación producido en la representación de la señal en cada una de las subbandas enmascara el potencial de blanqueado en tiempo proporcinado por la predicción, ya bastante reducido debido a la división en subbandas, que como es bien sabido proporciona un blanqueado en frecuencia.

En este trabajo se presenta la extensión de este codificador al caso de voz de ancho de banda de 7 kHz, es decir, de calidad conversacional, adecuada para aplicaciones multimedia. En este caso, los requerimientos de calidad obligan a trabajar a velocidades de 1,5 a 2 bits/muestra (de 24 a 32 Kbps con una frecuencia de muestreo de 16 kHz). En este caso, el número de subbandas en que se divide el margen de 0 a 8 Khz de la señal es de 16, siendo todas ellas de 500 Hz de anchura, y despreciándose las dos subbandas superiores debido a su despreciable contenido energético. En este caso, la mayor precisión de representación de las muestras en cada subbanda proporciona un mejor aprovechamiento del potencial de blanqueado de la predicción adaptativa, lo que aconseja el uso de ésta por lo menos en las 8 primeras subbandas, las de mayor contenido energético. Una descripción más detallada del codificador/decodificador APVQ puede encontrarse en [3].

2. LA PREDICCION ADAPTATIVA

Los predictores utilizados en la codificación de cada una de las subbandas son de tipo FIR, adaptativos y de tipo "backward" tanto porque el algoritmo de adaptación se basa en la muestras reconstruidas

Este trabajo ha sido financiado por el Plan Integrado de Banda Ancha (PLANBA) perteneciente al Plan Nacional de I+D.

como porque la predicción misma se basa en esas mismas muestras. Por ello, no es necesario transmitir información lateral sobre el cálculo o valores de los coeficientes de los predictores. Los algoritmos de adaptación connsiderados son el conocido LMS y el GAL (Gradient Adaptive Lattice) [1]. El segundo, más complejo que el LMS y, por consiguiente, de mayor costo computacional, proporciona unas prestaciones mucho mejores debido a su mayor velocidad de convergencia, sobre todo trabajando con alta calidad de codificación [1], como es el caso que nos ocupa. Por ello, este algoritmo GAL es el elegido, pues la mejora de calidad proporcionada compensa el mayor costo computacional requerido. Respecto al orden de los predictores a utilizar, cabe distinguir entre las bandas inferiores, donde la estructura de tipo periódico (rayas espectrales) del espectro en los intervalos sonoros de la voz es nítida, y las superiores, donde esta estructura se pierde en gran medida. En el primer caso, es adecuado utilizar longitudes mayores, de tal modo que se pueda aprovechar la periodicidad de la señal en los intervalos sonoros. Esta longitud no tiene por qué ser demasiado alta, debido a que el diezmado de las subbandas reduce la longitud del periodo de pitch por el mismo factor de diezmado.

Según un estudio realizado para las primeras 14 subbandas (0-7kHz) se ha concluido que la predicción escalar comporta beneficios sólamente durante las 10 primeras subbandas. Para cada subbanda se ha encontrado un valor óptimo del parámetro de memoria ß y el orden **p** óptimo del predictor GAL, según la correlación presente en cada subbanda. En la Tabla 1 se muestran las ganancias de predicción global (GP_G) y segmentada (GP_{Seg}) resultantes para cada subbanda a partir de los valores anteriores de ß y **p**. Se ha considerado una base de datos formada por 16 locutores (8 masculinos y 8 femeninos). Inicialmente se ha evaluado la estructura 'forward' del codificador APVQ y posteriormente se ha procedido a la realización de pequeños ajustes cuando se procesa dicho codificador en su configuración 'backward'. Para las subbandas superiores el orden se reduce drásticamente, llegándose a eliminar la existencia del predictor a partir de la décima subbanda.

3. INICIALIZACION DEL CODEBOOK

La utilización de la técnica de Splitting para obtener el codebook inicial de tamaño 2^T comporta un inconveniente importante a tener en cuenta: para llegar a la obtención del codebook óptimo de tamaño 2^T se han de diseñar los codebooks óptimos de tamaños 2 hasta 2^{T-1}, lo que comporta un coste computacional enorme, especialmente para codebooks de tamaño elevado (512 o 1024). Recientemente ha aparecido un algoritmo propuesto por Katsavounidis, Kuo y Zhang [2] que permite un importante ahorro de cálculo sin perder prestaciones. Por simplicidad se ha denominado este algoritmo como algoritmo de Kuo.

Subbanda (kHz)	Orden p	Factor ß	GP_G (dB)	GP_{Seg} (dB)
0.0 - 0.5	9	0.97	23.72	20.97
0.5 - 1.0	8	0.97	9.50	6.83
1.0 - 1.5	7	0.97	4.12	2.33
1.5 - 2.0	5	0.97	5.18	2.09
2.0 - 2.5	3	0.97	3.87	1.24
2.5 - 3.0	2	0.97	2.62	2.06
3.0 - 3.5	1	0.97	2.15	1.72
3.5 - 4.0	1	0.97	1.58	0.89
4.0 - 4.5	1	0.97	1.82	0.92
4.5 - 5.0	1	0.97	2.30	1.90

Tabla 1 : *Ganancias de predicción correspondientes a cada subbanda para los órdenes de predicción óptimos (se han representado los valores correspondientes al mejor locutor).*

Primeramente se actúa sobre la secuencia de entrenamiento al completo, y se elige el vector de norma máxima como primer centroide. Una vez obtenido este primer vector de referencia, se cuantifica toda la secuencia de entrenamiento a partir de este VQ de tamaño 1, escogiendo como segundo centroide el vector de la secuencia que presenta mayor distorsión, es decir, el que tiene distancia máxima respecto al único vector que conforma el VQ. Nuevamente, se procede a la codificación de toda la secuencia de entrenamiento, hasta hallar el vector de entrada al VQ que presenta un mayor distancia y éste se elige como el nuevo centroide. Así, sucesivamente se obtiene el codebook inicial de tamaño 2^T deseado. La inicialización a que conduce esta nueva técnica, conduce a una distribución homogénea e inteligente del codebook inicial, puesto que está ubicando las diferentes celdas de forma que la clasificación de la secuencia de entrenamiento en el VQ inicial se mantenga dentro de unos límites de distorsión razonables. Después de la obtención de los centroides iniciales se aplica el algoritmo LBG para la consecución del codebook óptimo.

La aplicación de este nuevo algoritmo de inicialización, comporta relevantes ventajas en el proceso de diseño del cuantificador vectorial en cuanto a coste computacional, pero se deben verificar todavía sus prestaciones en términos de conformación de ruido de cuantificación. Al comparar las prestaciones de ambos algoritmos, Splitting y Kuo, se observa como la estrategia de Kuo es mucho más rápida pero no siempre conduce a valores de SNR del codebook mejores. En [2] los mencionados autores muestran como se obtienen SNR globales del codebook mejores tras aplicar el algoritmo LBG, en comparación a la inicialización por la técnica de Splitting. En nuestro caso también la técnica de KUO conduce a SNR globales ligeramente mejores, pero al mismo tiempo las SNR segmentadas del codebook final son ligeramente peores. Su principal problema radica en la existencia de vectores de norma alta, mal normalizados por el predictor de ganancia durante las transiciones de energía de la señal de voz. Por esta razón estos vectores son elegidos como centroides por el algoritmo de Kuo y, a veces, dan lugar a celdas vacías o semivacías tras aplicar el algoritmo LBG. La existencia de estos vectores mal normalizados proviene de cambios bruscos en la potencia de la señal (paso de silencio a señal activa), con lo que el estimador de ganancia antes de llegar a estabilizarse realiza una normalización deficiente de algunos vectores, y muestras pertenecientes a dichas transiciones tendrán una potencia indeseada muy superior a la unidad. Por este motivo se han considerado dos posibles modificaciones: el algoritmo de Kuo por porcentaje KUO_P y el algoritmo de Kuo por umbral KUO_U.

La primera estrategia KUO_P consiste en una primera inicialización mediante la técnica de KUO y posteriormente se cuantifica toda la secuencia de entrenamiento, de manera que se obtiene un histograma sobre la distribución de los distintos vectores de entrada en las diferentes celdas del codebook inicial. Aquellas celdas que reciben muy pocos vectores de entrada (tienen un número de vectores inferior a un cierto porcentaje del total) son eliminadas y de nuevo se aplica el algoritmo hasta que exista una distribución de la secuencia más coherente en las diferentes regiones o celdas. Esta nueva estrategia presenta la ventaja de su sencillez y ser bastante intuitiva, pero presenta el inconveniente de que pueden precisarse un número de iteraciones demasiado elevado, perdiéndose parcialmente la ventaja de reducción de complejidad asociada con el algoritmo de Kuo.

El método de Kuo por umbral KUO_U es parecido al anterior. KUO_U. Primeramente se procede de la misma forma que en al primer método, es decir, se eliminan los centroides cuyas celdas no superan un cierto umbral de vectores de entrada y sus vectores se reparten entre los centroides restantes. Entonces, se elige

Técnica	SNR_{seg} (dB)				
Kuo	5.78	5.03	5.50	5.43	3.99
KUO_P	6.06	5.20	5.51	5.65	4.30
KUO_U (g=.16)	5.75	5.20	5.45	5.68	4.34
KUO_U (g=.36)	6.03	5.25	5.57	5.77	4.32
Splitting	6.09	5.21	5.69	5.80	4.33
Subbanda	1	2	3	4	5

***Tabla 2** : Comparación en términos de SNR del codebook final de las diferentes técnicas de obtención de un codebook inicial de dimensión cinco.*

aquél que presenta mayor distancia (d_{max}) respecto a su centroide asignado. Esta distancia se pondera por un factor de reducción de distancia **g**. Se define un umbral **U** de la forma $U = d_{max} \cdot g$ y seguidamente se buscan todos los vectores de la secuencia de entrada cuya distancia a su centroide sea superior al umbral de distancia **U**. Para cada celda se suman las distancias de todos los vectores que superan este umbral y se selecciona la celda que contiene mayor distancia acumulada. Como nuevo centroide se selecciona el vector de mayor distancia perteneciente a esta celda. Este procedimiento finaliza momentáneamente cuando se llega al tamaño deseado. De este modo, las técnicas de inicialización KUO_P y KUO_U eliminan los centroides indeseados según los mismos criterios pero generan los nuevos centroides de forma diferente.

En un principio podría pensarse que este segundo algoritmo comporta un tiempo computacional elevado en relación al algoritmo de KUO, pero en realidad no es así puesto que el número de iteraciones a seguir es muy reducido debido a la forma de seleccionar los nuevos centroides. Al elegir la celda que presenta una mayor suma de distancias respecto a los vectores que superan el umbral **U**, el algoritmo se asegura la existencia de una cierta densidad de éstos en la periferia de dicha celda. Mediante el control de este umbral **U** se consigue que el nuevo centroide elegido tenga un número de vecinos lo suficientemente grande para que no sea eliminado posteriormente durante el primer paso del algoritmo.

El control del umbral de distancia **U** se regula mediante el parámetro **g**. Un valor elevado (en torno a la unidad) conduce a la primera estrategia de KUO_P. Un valor de **g** más pequeño conduce a una convergencia más rápida pero empeora las prestaciones del proceso de cuantificación, ya que no conduce a una inicialización homogénea del codebook, sino que tiende a agrupar los centroides en constelaciones, separadas entre sí. Esto conduce a tener un mal comportamiento tanto en términos de SNRov como de SNRseg. En la Tabla 2 se han representado los valores de SNR del codebook obtenidos por cada uno de los métodos de inicialización anteriormente comentados. Los resultados obtenidos muestran como técnica más óptima la de KUO_U cuando se considera un valor del parámetro de reducción de distancia g=0.36.

Así pués a modo de resumen, el método a seguir para el diseño de codebooks es el siguiente:

- Inicialización mediante el algoritmo de Kuo por umbral KUO_U considerando g=0.36
- Aplicación del algoritmo LBG para la obtención del codebook óptimo.
- Depuración automática de celdas vacías (si existen).
- Aplicación del algoritmo LBG.

4. CONCLUSIONES

El codificador aquí presentado, denominado APVQ-extendido, resulta adecuado para la codificación de calidad de señal de voz de banda ancha de 7 kHz, cuya característica principal es el alto margen dinámico espectral que presentan este tipo de señales. Este sistema, al incluir división en subbandas y cuantificación vectorial adaptativa, presenta muy buenas propiedades para manejar con altas prestaciones este tipo de señales. Los resultados previos obtenidos, junto a las buenas prestaciones obtenidas con este sistema para señales de voz de banda estrecha (4 Khz) y a las propiedades intrínsecas comentadas del sistema, garantizan un muy buen comportamiento con señales de voz de banda ancha. En este trabajo se ha presentado el diseño efectuado para los distintos predictores correspondientes a las distintas subbandas. Asimismo, se han comentado algunos métodos de inicialización del codebook para reducir el enorme coste de cálculo asociado con la técnica de Splitting.

5. REFERENCIAS

[1] E.Masgrau, J.B.Mariño. "Subband splitting, adaptive scalar prediction and vector quantization for Speech Encoding". Proc. EUSIPCO, pp.1035-1038. Grenoble, Francia..Septiembre 1988

[2] I.Katsavounidis, C.C.J.Kuo, Z.Zhang. "A new Initialization Technique for generalized Lloyd Iteration". IEEE Sig. Proc. Letters, Vol. 1, No. 10. Octubre 1994.

[3] E.Masgrau, J.M.Salavedra. "Codificación APVQ-Extendida de voz de Banda Ancha". Proc. Congreso URSI, pp. 17-21. Gran Canaria. Septiembre 1994.