A CONTINUOUSLY ADAPTIVE VECTOR PREDICTIVE CODER (AVPC) FOR SPEECH ENCODING

Enrique Masgrau, José B. Mariño and Francesc Vallverdú

DPTO. TRATAMIENTO Y TRANSMISION DE LA INFORMACION
E.T.S.I. TELECOMUNICACION - U.P.C.
Apdo. 30.002
08071 Barcelona. SPAIN

ABSTRACT

In this work we present a waveform speech coding system including vector quantization. This system can be seen as a vector version of the scalar ADPCM speech coder. In such system the speech samples are grouped in vectors that are coded by a vector quantizer when its prediction subtraction has been made. This prediction is obtained in the coder by a vector predictor. Due to the non-stationarity of the speech signal, the code must be continuously adapted to the local characteristics of the current input signal of the speech. We propose the use of an adaptive vector predictor that follows the speech statistics variations so the prediction error to be coded presents the minimum dynamic range. On the other side it is wellknown that the prediction error is proportional to the signal energy. To compensate this effect a vector quantizer "gain-shape" model has been proposed, so vectors gain and its shape are separately coded. The obtained empirical results are very promising and exhibit good competitivity with other solutions existing in the literature.

INTRODUCTION

Although the application of Vector Quantization (VQ) is only beginning to be used in speech encoding systems, the obtained results up to now shows good possibilities and in a near future those methods will be of a great use in transmission systems with a low and medium bit rate. As a matter of fact, after CCITT having adopted an APDCM standard system at 32 Kbps, the issue is being focused at 8-16 Kbps rates. It may be possible that by using a VQ together with a vector predictor in the feedback loop of an ADPCM the bit rate will be reduced meanwhile the quality is maintained. The initial works in this direction have shown high competitivity in relation to other encoding methods as subband coding |1,2|.

These coding systems are named Vector Predictive Coding (VPC) and its general scheme is shown in figure 1. Basically is composed of two main blocks: the vector quantizer and the vector predictor.

The way it works is as follows: the prediction of current input vector available from the output of vector predictor is subtracted from the input vector. The resulting error of prediction is coded by the vector quantizer. The vector

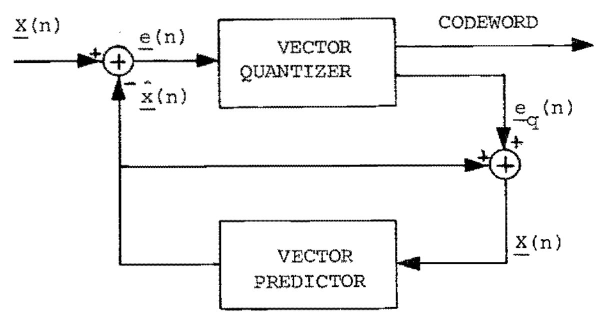

Figure 1.- General scheme of a VPC system.

predictor works the reconstruted input signal rather than the exact input signal, because of the latter is not available at the receiver.

Due to the non-stationary character of the speech signal the use of permanent blocks for the vector quantizer and the predictor give a low performance . The solution lies on adapting somehow these blocks "to the local characteristics of the current input speech". In the previously mentioned works, the adaptation process is reduced to the classification of the input speech frames (containing a low number of vectors) among a small number of classes by means of assigning a particular invariant predictor and a codebook to each one of them. The main problems of using this structure are two: 1) the redundance removal capacity of the predictor is not wholy used, because this is not fitted in a continuously way to the non-stationary characteristic of the input speech; and 2) the dynamic range of the vector quantizer is fitted to the dynamic range or gain of the error signal in a rude way. Moreover, a third drawback is the side information that is required to be sent to the receiver has to be related to the codebook and the predictor selected by each frame. This is a minor drawback as it represents a low increase on the transmission rate.

Our aim is to reduce the effects of these drawbacks. In order to achieve this we use a serie of blocks, continuously adapted to the signal for both the "vector quantizer" and "vector predictor".

The "Adaptive Vector Predictor" (AVP) contains an adapted algorithm which constantly renews the coefficients of the matrices of the Vector Predictor, using a minimizing approach, which will be discussed in the following section. This AVP can be thought of being the vector equivalent to an adaptive scalar predictor.

The Adaptive Vector Quantizer (AVQ) can be split up into two different blocks: Firstly, an estimator-quantizer of the gain (the square-root of the energy) of the prediction error vectors; and secondly, we have an invariant VQ that codes these error vectors previously normalized by its gain-estimation quantized. The estimator-quantizer and the subsequent error vectors normalization provide the adaptation from the VQ dynamic range to the error dynamic range. The VQ essentially consists of a time-invariant codebook which codes the shape of the prediction error vectors.

Both the AVP and AVQ adaptive characteristics allow for the use of just one invariant block for the implementation of each one, instead of a series of invariant blocks together with a speech signal weft clarificator, as discussed in solutions |1, 2|.

We shall call the whole system, containing both adapted blocks, the Adaptive Vector Predictive Coder (AVPC).

THE AVPC SYSTEM

The Adaptive Vector Predictor (AVP)

The predictor function is used to obtain the most accurate prediction of the actual input vector so that the prediction error be a minimum. The dynamic range reduction of the signal to be quantized, allows a more accurate representation by taking the mean of the VQ, and therefore, a smaller quantizer error occurs.

The predictor is of the backward type, so the input vector prediction must be done on the basis of the previous reconstructed vectors which differ from the real ones in the quantization error. The prediction coefficients, $\{A(n)\}$, are matrices KxK , where K is the dimension of the vector whilst n is the current time.

Then, the expression for the vector prediction is:

$$\hat{\underline{X}}(n) = \sum_{i=1}^{P} \underline{\underline{A}}_i(n)\ \tilde{\underline{X}}(n-i) \qquad (1)$$

where P is the order of the predictor and $\tilde{\underline{X}}(n)$ is the reconstructed signal vector.

The coefficients $\underline{\underline{A}}_i(n)$ are calculated so that the prediction error $\underline{e}(n)=\underline{X}(n)-\hat{X}(n)$ be in a certain way minimized. The most used criterium is to minimize the matrix trace of the mean square error.

$$Tr\left[E\{\underline{e}(n).\underline{e}^T(n)\}\right] \qquad (2)$$

where the supraindex T means transposed vector. The minimization of (2) conveys to the vectorial version of the wellknown normal equations.

$$\sum_{i=1}^{P} \underline{\underline{A}}_i(n).\underline{\underline{R}}(j-i) = \underline{\underline{R}}(j) \quad ; j=1,\dots P \qquad (3)$$

where $\underline{R}(j)$ are the correlated signal matrices $\tilde{\underline{X}}(n)$. $\tilde{\underline{X}}(n)$ — signal is only avalaible once the predictor coefficient estimation has been done. The coefficients are calculated from the $\tilde{\underline{X}}(n)$. This problem requires an iterative solution, in which both, the predictor estimation and the signal $\tilde{\underline{X}}(n)$ are alternately obtained. This solution will be discussed further on; we will now simplify the problem by designing the predictor from the real speech signal $\underline{X}(n)$ that only differs from $\tilde{\underline{X}}(n)$ in the quantizer error being that a small one. In this case the predictor coefficient calculation from (3) requires the previous estimation of $\underline{X}(n)$ correlation matrices. Due to the non-stationary character of the speech signal the matrices should be update in certain periods of time and being the predictor coefficient re-calculated. The predictor coefficients must be transmitted to the receiver, highly increasing the transmission rate.

This drawback can be avoided by using an adaptive predictor, where the coefficients are constantly renewed. Different ways of minimizing expression (2) give several adaptive predictor algorithms. Three algorithms have been considered: The transversal LMS (Least Mean Square), and two lattice algorithms in which the speech signal has been previously ortogonalized by means of an adaptive lattice predictor with a gradient type algorithm GAL2 (Gradient Adaptive Lattice) or by the ELSAL (Exact Least Square Adaptive Lattice).

Algorithm Vector LMS

In this case, similarly to the scalar case, the predictor coefficients are calculated sample by sample and changing in the direction of minus gradient.

$$\underline{\underline{A}}_i(n+1)=\underline{\underline{A}}_i(n)+\mu(n)-\left[-\underline{\underline{\nabla}}_i(n)\right] \qquad (4)$$

where $\underline{\underline{\nabla}}(n)$ is the gradient of (2) and $\mu(n)$ is a time variant escalar parameter to match the speech energy.

The gradient accurate expression requires the previous estimation of the signal correlation matrices. We can avoid this drawback by taking the instantaneous gradient as gradient estimation $\hat{\underline{\underline{\nabla}}}_i(n)$. Replacing the statistic function by instantaneous values in expression (2) and deriving the function we obtain:

$$\hat{\underline{\underline{\nabla}}}_i(n)= \sum_{j=1}^{P} \underline{\underline{A}}_j(n)\underline{X}(n-j)\underline{X}^T(n-i)-\underline{X}(n)\underline{X}^T(n-i)=$$

$$= -\underline{e}(n).\underline{X}^T(n-i) \qquad (5)$$

To reproduce the predictor coefficient estimation is necessary to work with avalaible signals once in the receiver. The $\underline{X}(n)$ signal is indeed the reconstructed $\tilde{\underline{X}}$ signal. The available error signal is the quantized error $\underline{e}_q(n)$ being equivalent to minimize the error trace rather than the $\underline{e}(n)$. Replacing (5) in (4) and taking into account the

56. 1. 2

abovementioned, the final recursion of the coefficients is obtained.

$$\underline{\underline{A}}_i(n+1) = \underline{\underline{A}}_i(n)+\mu(n)\underline{e}_q(n)\tilde{\underline{X}}^T(n-i) \qquad (6)$$

The $\mu(n)$ parameter is taken as time variant to follow the signal energy as the gradient term is proportional to this value.

$$\mu(n) = \frac{\alpha}{P.\sigma(n)} \; ; \quad 0 < \alpha < 2 \qquad (7)$$

where

$$\sigma(n) = \beta\sigma(n-1)+(1-\beta)\tilde{\underline{X}}^T(n).\underline{X}(n) \quad 0<\beta<1 \qquad (8)$$

is an estimation of the input predictor signal energy.

Vector GAL2 Algorithm

In this predictor the speech signal is previously orthogonalized by means of a lattice predictor. After this process, the signal vectors exhibit orthogonal components. The complet predictor structure is shown in figure 2 for the usual case with P=1 and dimension K.

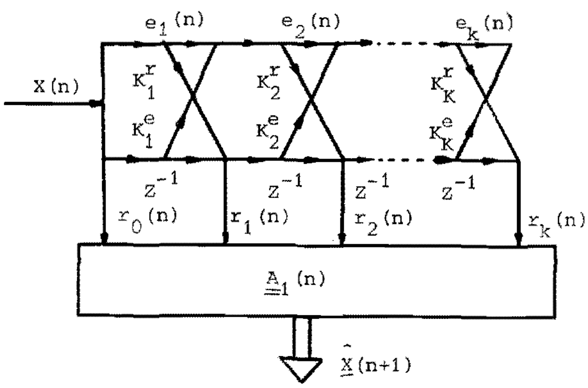

Figure 2. Vector Predictor Structure with previous orthogonality with order P=1 and dimension K.

The lattice structure provides an orthogonalized version of the input signal. Grouping in vectors the backward error sequence $r_i(n)$ we can define a vector process $\underline{r}(n)$. The new process presents a diagonal correlation block-matrix, and, therefore, a nulle eigenvalues-spread. The $X(n)$ prediction is obtained as the LMS case by weighting the new vectors

$$\hat{\underline{X}}(n) = \sum_{i=1}^{P} \underline{\underline{A}}_i(n).\underline{r}(n-i) \qquad (9)$$

The lattice predictor order is KxP. The equations of the GAL algorithm can be seen in reference |3|. The update recursion of the $\underline{\underline{A}}_i(n)$ coefficients is the same that the LMS (6) replacing $\tilde{\underline{X}}(n)$ by $\underline{r}(n)$.

In this algorithm the evolution of the matrices $\underline{\underline{A}}_i(n)$ components are independents between them because they weight orthogonal signals. This property provides a fast speed of convergence and therefore, a best tracking of the speech statistic.

Vector ELSAL Algorithm

This predictor is similar to the GAL2 one, but in this case the lattice Parcor coefficients are controlled by ELSAL algorithm which has better convergence properties than GAL2. The equations of this algorithm can be seen in reference |3|.

To evaluate the applications of the three proposed algorithms we start to measure the prediction gain of every one of them. A learning set was used as a signal test of signal speech and the same one will be used for the complet dessign of AVPC system and it is described on the results section. In order that these measurements be independent from the VQ design the predictors worked in a forward sense, out of the AVPC feedback loop.

The experiments were carried out between 1 and 5 for the dimension vectors K. The case K=5 will be chosen for the complet AVPC system evaluation. For this dimension predictors with range up to 1 do not result suitable because the correlation among samples speech signal decreases fast with the distance. So we take P=1 as a predictor order. The algorithm convergence parameter values were empirically optimized.

In figure 3 the prediction gains are shown for both LMS and GAL2 predictors. The predictors gains obtained with ELSAL are very similar to the ones obtained in GAL2 and are not drawn. This approaching to the results between the algorithms lattice gradient pattern GAL and the ELSAL ones have been also tested by the authors on the ADPCM scalar context |3|.

The predictor gains in the lattice cases are greater than in the LMS one, as it was hoped. This increase is about 10% (0.6dB) in the vector dimension K=5 case, decreasing for smaller dimensions. This is due to the orthogonality effects that are more important for higher range correlation matrix (equal to K.P).

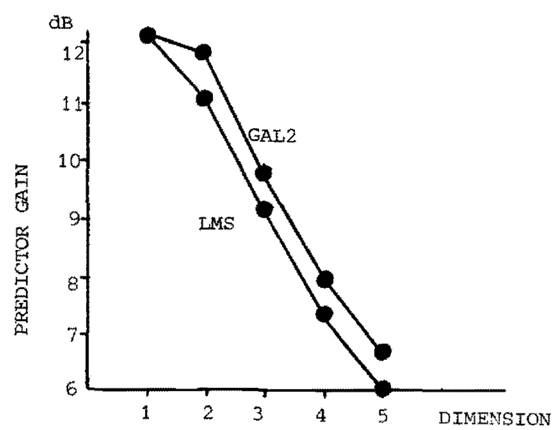

Figure 3.- Predictor gains plot versus K dimension for GAL2 and LMS Predictors.

The Adaptive Vector Quantizer (AVQ)

The AVQ codes the error prediction by assigning to each error vector a codeword.

It is composed of a Gain Estimator Quantizer and a VQ. The former provides a gain estimation of the error vectors by windowing the exact gain sequence. This smoothing is necessary to reduce the gain rate transmission. The normalization of the error vectors by the quantized gain gives the adaptive character to the AVQ. The VQ is an usual quantizer working with these normalized vectors.

Gain Estimator-Quantizer

The use of a window in the gain estimation causes a smoothing of the accurate gain. A long window provides a higher smoothing as well as a lower gain rate transmision. We have studied the gain sequence spectrum for the window design resulting that a rectangular window of 8-10 vectors length (K=5 dimension) is a good election. The estimation is carried out every 6 vectors and the gain is 6 bits quantized.

It has been recently published a work |4| that proposes several solutions for the Gain-Shape quantization. These are forward (as the one presented here) and bacward solutions. These latter do not need to transmit the gain being this reconstructed in the receiver.

Vector Quantizer VQ

The VQ consists of a time invariant codebook. The codebook design requires a learning set and it is carried out by the use of a modified LBG algorithm |4|. This design takes into account the gain value of every vector for the centroids estimation. The expresion of the centroids is

$\Sigma\, \sigma^2_i y_i \,/\, \Sigma\, \sigma^2_i$, where σ_i is the vector gain and y_i the normalized vectors.

An initial VQ design is carried out in a forward sense: the learning set goes through the adaptive predictor obtaining an error vectors sequence. The gain vectors are estimated by the gain estimator-quantizer, and the error vectors are normalized. By using the gain and the normalized error sequences the VQ is designed.

In the real case, the predictor works in a backward sense and the VQ design can be modified. Working with the complet AVPC system and the initial codebook a new error sequence is presented to the VQ. With this sequence a second VQ design is carried out. By repeating this cycle we obtain the best VQ design. The authors experiences realize that two or three iterations are sufficient. The final enhancement in SEGSNR respect to the initial VQ is about 0.5 dB.

EMPIRICAL RESULTS

To evaluate the whole AVPC system we have used a learning set containing about 5 sec of speech signal sampling to 8 KHz. This learning set contains two utterances: a female Catalan utterance of "Aixo es una prova de codificacio de veu" and a male Spanish utterance of "Esto es una prueba de codificación de voz".

The vector dimension is K=5 and the order predictor is P=1. The parameter convergence of LMS algorithm are β=0.98 and α=0.06. In K=5 case, the vectors number in the learning set is about 8.000. It was considered the case with 16,32 and 64 codewords in the codebook, this is to say, 0.8, 1 and 1.2 bits/sam. of codeword transmision. This rate is increased by the gain rate transmision.

A first experience was carried out to evaluate the assymthotic best performance. It did consist on the use of the gain of every error vector (not smoothing) quantized by 6 bits for vector normalization.The SEGSNR obtained in the three cases is shown in figure 4.

By the use of the gain quantizer solutions proposed in the above section and assigning 1 bit/vector for gain quantization as well as transmiting every 6 vectors (6bits/30 samples=1,6 Kbs/sec), the SEGSNR obtained for every case is shown in figure 4. As a test out of learning set we have used the Spanish word "dos" with female utterance. We have acted in the same way with the gain quantizer proposed in |4|. It consists of the quantization of the gain every 25 vectors using 4 bits (256bits//sec). The results are also shown in Fig. 4. The results for the lattice GAL2 predictor are also expected to be up to 0.5-1 dB.

At present, the AVPC system is being simulated on Numerix MARS 432 array processor with a VAX 11/750 host computer. This computational power will allow us to work with greater data base and greater codebooks. Further results for transmission rates up to 16 Kbps will present in the paper exposition.

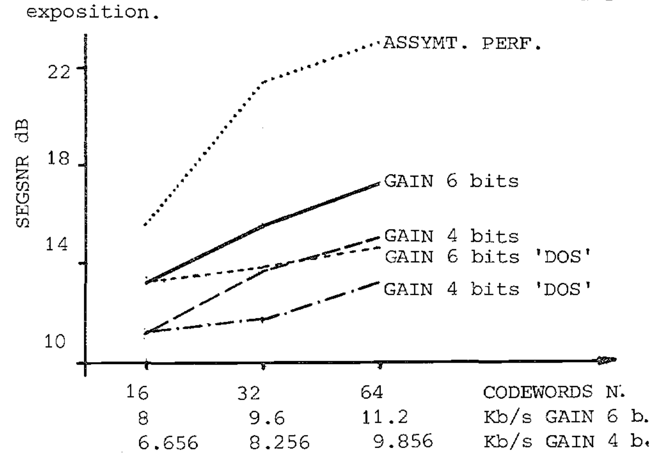

Figure 4.- SEGSNR for several AVPC.

|1| A. Gersho, V.Cuperman "Vector Quantization: A Pattern Matching Technique for Speech Coding". IEEE Comun. Magaz, pp. 15-21, December 1983.

|2| V. Cuperman, A. Gersho. "Vector Predictive Coding of Speech at 16 Kbits/s".IEEE Trans. on Commun. Vol. COM-33, pp 685-696, July 1985.

|3| E. Masgrau et al. "Performance Comparison of 32 KBPS ADPCM Coders in Noise-Free and Noisy Environments". Proc. on MELECON-85, pp. 179-182 Madrid Oct. 1985.

|4| Juin-Hwey Chen, A.Gersho "Gain-Adaptive Vector Quantization for Medium-Rate Speech Coding".Proc. ICC-85, pp. 1456-1460, 1985.

This work is supported by CAICYT - Plan n.2906/83C3