MEDIUM-RATE CODING WITH VECTOR QUANTIZATION

José A. Rodríguez Fonollosa, Enrique Masgrau, José B. Mariño, Asunción Moreno

Vector quantization (VQ) in opposition with the independent o scalar quantization of each sample of a waveform, is a simultaneous quantization of a sequence of samples.

The theoretical foundations of vector quantization lie in a branch of information theory known as rate-distortion theory, originally set forth by [Shannon 1948]. The fundamental results of this theory is the optimality of VQ, i.e., performance arbitrarily close to the ultimate rate-distortion can be achieved by VQ if the vector dimension is large enough.

Nevertheless, the main problem of VQ is the exponential growth in complexity as the vector dimension increases. This fact forces the use of low-dimensionality VQ in practical systems. However, direct waveform coding of speech with VQ of low dimension does not give adequate performance.

To cove this problem two ways can be followed. The first one is to use VQ-based structures with a reduced computational and/or storage cost. The other way is to combine VQ with other techniques that exploit the correlation of samples from different vectors.

In this part of the session we will present examples of both alternatives. All the systems are also adaptive. Speech can be considered as a process locally stationary with slowly varying statistics, therefore, a coder with the possibility of being continuously adapted to this statitistics is expected to perform better than a non-adaptive one.

In first place we will present an Adaptive Vector Predictive Coder (AVPC) that combine VQ with a vector predictor and it can be seen as a vector version of the scalar ADPCM.

Then two speech coders that include also a Subband Splitting are described. In the first SBC scheme each of the bands is coded with the AVPC system. This allows to control the bit assignment to each band and the subjective quality increases. Nevertheless the complexity of this coder is quite high because some vector quantizers and predictors ares used. The second developed SBC scheme uses scalar predictors and only one vector quantizer. This scheme allows to obtain high quality at rates between 10 and 8 Kbps.

The last scheme we will present is a new adaptive vector quantizer based on the multistage structure. Its main advantage over other VQ is its robustness against different speakers and environments, and it has been applied to quantize the LPC parameters with very good results.

1.- Adaptive Vector Predictive Coding

The scheme of the AVPC system can be seen in figure 1. The vector quantizer and the gain estimation form a gain-adaptive vector quantizer [Chen 1987], and the other main block is the vectorial predictor. Both the gain estimation block and the vector predictor work in a backward mode, i. e., no side information transmission is needed. The first AVPC system was introduced by [Cuperman 1985] and its continuously adaptive version by [Masgrau, Mariño 1986].

We have tested some adaptive algorithms to predict the signal. The VGAL algorithm is based on a previous orthogonalization by means of a lattice structure. It gives the best prediction gain in open loop (without quantization) but its performance is similar to the VLMS gain in close loop.

We have also used some methods to estimate the gain. The best results are obtained with a scalar GAL predictor.

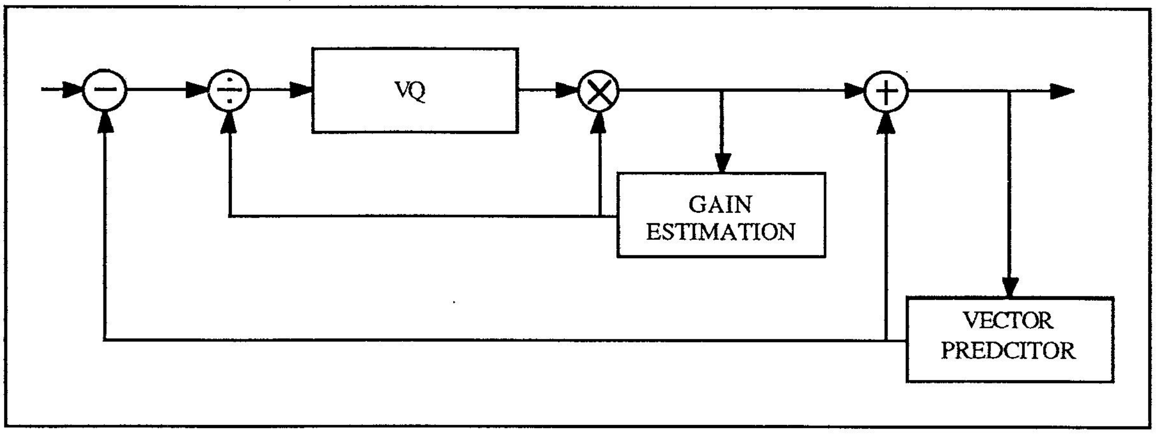

Figure 1. AVPC system with backward prediction and gain estimation.

2.- The SBC-AVPC System

The previous AVPC system works well at rates about 2 bits/sample . In this case, the number of codewords in the VQ is enough to code the vector shape even though the gain estimation was not exact. Test carried out at rates about 1 bit/sample had shown the inability of the codebook for coding the vector shape and even in the case of perfect gain estimation the decoded speech has a "metallic" quality.

The previous subband splitting of the speech allows to improve this shape coding meanwhile the transmission rate is maintained. Thus the smooth and rough shape are addressed by different coding paths, and therefore, a grater fidelity is obtained. The proposed coder splits the speech signal into four bands (0-0.5, 0.5-1, 1-2 and 2-3 KHz) by using a QMF bank, [Johnston 1980], and then each band output is coded by an AVPC system. The bits are assigned to each band to maximize the subjective quality. The results can be seen in table I.

SPEAKER	INSIDE						OUTSIDE		
CODER	F1	F2	F3*	H1	H2	H3	F4	H4	F3*
CCITT 32Kbps	31.2	29.7	26.5	27.6	23.6	21.7	32.2	30.3	29.0
AVPC 16Kbps	21.0	17.9	--	16.2	15.0	--	18.9	18.5	18.1
AVPC 9,6Kbps	13.1	11.2	--	10.3	9.4	--	11.8	12.3	12.4
AVPC 8Kbps	11.5	9.8	--	8.1	7.4	--	11.6	11.0	10.1
AVPC-SBC 12Kbps	14.6	13.4	11.5	10.5	10.0	10.2	12.2	12.6	12.5
AVPC-SBC 10Kbps	14.1	12.4	10.7	9.6	9.1	7.4	12.0	12.3	11.9
AVPC-SBC 8Kbps	13.0	11.1	10.8	8.2	7.4	6.1	11.6	11.3	10.4

Table I. SEGSNR results for several coders.

3.- SBC with Adaptive VQ and scalar prediction

The SBC-AVPC system described above has two drawbacks: it uses some vector predictors and vector quantizer with a significant increase in complexity over the AVPC system, and the prediction gain of the vector quantizers of in each subband is limited due to the spectral whitening effect caused by the filtering and decimation process.

To solve this problems a new scheme with scalar predictor and only one vector quantizer is proposed [Masgrau, Rodríguez 1987], (Figure 2). Each band has gain estimator and four of them uses a scalar predictor. The gain estimation is used to normalize each component (band) of the vector to quantize and in the process of choosing the optimum quantized vector, (Weighting).

The filter bank is a tree structure of QMF filters that gives 8 bands of 500 Hz each. The order of the backward adaptive predictors in the bands 1 to 4 are 9, 9, 4 and 2. In the bands 5, 6, 7 no predictor is used because simulations showed no improvement in quality when they were included. This fact reduces the number of taps to be updated to 24 each 8 samples of speech (3 per sample). For the estimation of the standard deviation of each a simple recursive estimation is used.

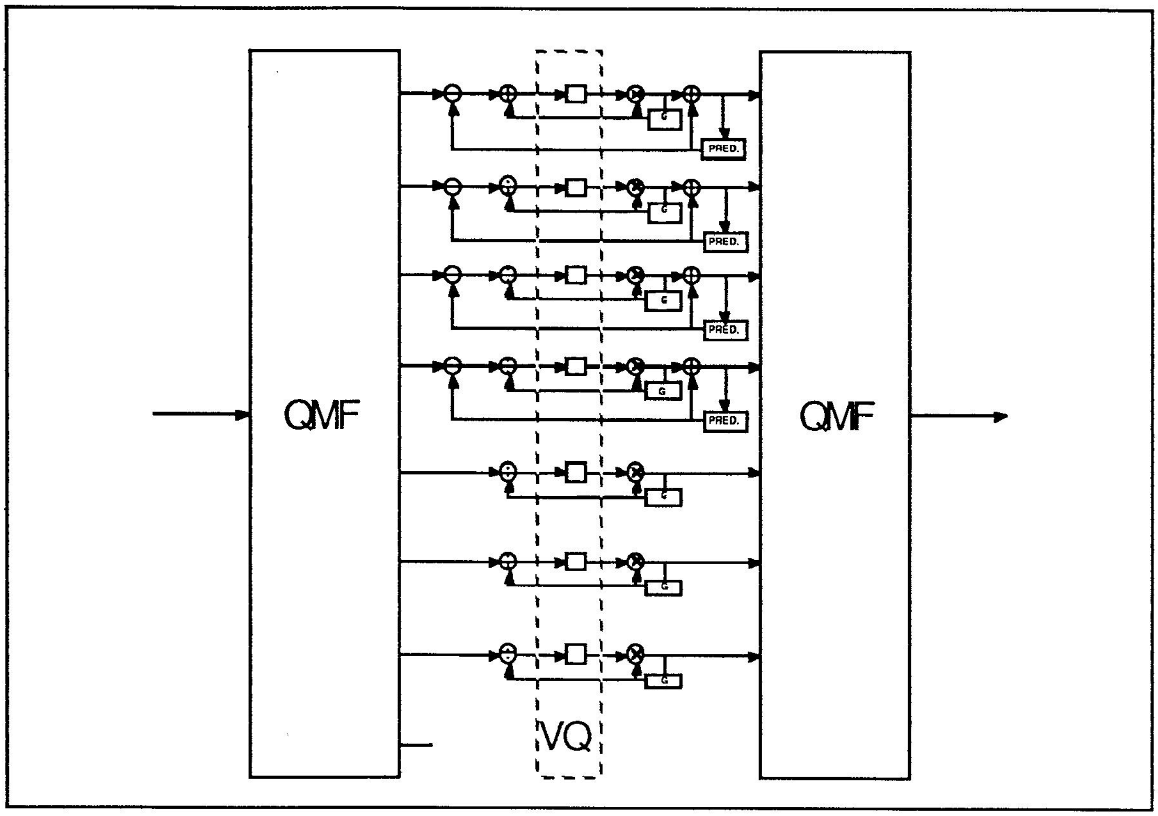

Figure 2. General scheme of the SBC system with scalar prediction and vector quantization.

Figure 3 shows the results at 8 and 10 Kbps in each band. The overall results are a SEGSNR of 14 dB at 8 Kbps and 16 dB at 10 Kbps.

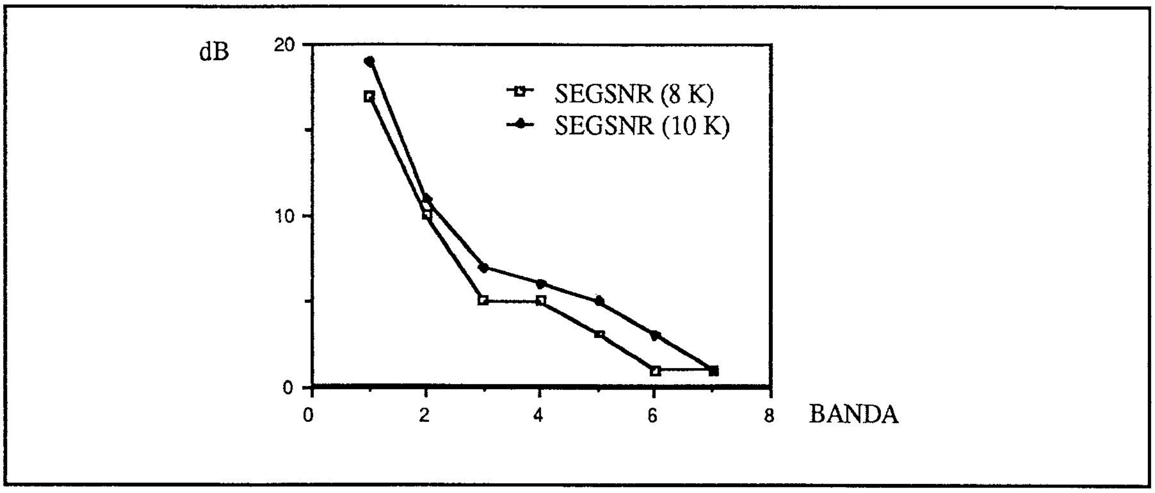

Figure 3. Segmented SNR as a function the subband

4.- Adaptive Multistage Vector Quantization.

In this section, we present a new multiple stage vector quantization method that allows the adaptation of the quantizer to signal to be coded. This adaptation is computationally very simple and is made with no increase in bit-rate. The resulting quantizer provides a robust performance across different speakers and environments. It has been applied to the quantization of the LPC parameters and the results show a significant increase in performance respect to conventional multi-stage vector quantizers [Rodríguez 1989].

Most of the low-bit-rate coders that have been developed recently include one or various vector quantizers in their scheme. Previous work on VQ usually employed trained codebooks or codebooks populated with randomly generated samples. In the first case, it is important to use long and representative training sequences to obtain a good performance "out" of the training sequence. Even though, the performance of any codebook is expected to deteriorate if used with different speakers and environments. In the second case, the quantizer is not taking advantage of the four interrelated properties and the performance reduces.

It is well known that, for a given bit rate, a speaker-dependent codebook, i.e., designed for the specific speaker whose speech is being coded, would work better than a speaker-independent codebook. This fact has motivated our research of a VQ system that adapts its codebook as it is used. This system would have the extra advantage of also automatically adapting to the acoustic environment of the speaker and to the recording conditions.

This idea is not completely new and, for example, in [Paul 1983] a system that changes the code-vectors in time is described. Nevertheless this change creates the necessity of transmitting the new vectors to the receiver with a significant increase in bit-rate. In this section, we describe a simple algorithm for updating the codebook that eliminates the necessity of very long and representative training sequences, and allows the quantizer to be continuously adapted to the signal statistics with no increase in bit-rate. This adaptation provides a better signal to noise ratio and robust performance across differents speakers and speech recording conditions.

In the last part of the paper, it is shown that the application of this Adaptive Multi-Stage Vector Quantization (AMSVQ) technique to the quantization of the log-area ratios (LAR) give very good results. Accurate quantization of LPC parameters is necessary for obtaining high quality speech at low bit-rates and AMSVQ is shown to be a good candidate to quantize the short-term predictor coefficients and other parameters in multipulse excitation coders, regular pulse excitation and codebook excited coders.

Adaptation Algorithm. The main idea of the adaptation algorithm is to update the preceding codebooks taking into account the information given by the rest of the quantizers. In the most simple scheme, with 2 stages, a vector x is quantized to give the quantized vector z

$$z = c + e_q \quad (1)$$

where c is the output of the first quantizer and e_q the contribution of the second codebook. (See figure 4).

$$c = q1(x) \quad (2)$$

$$e_q = q2(x - c) \quad (3)$$

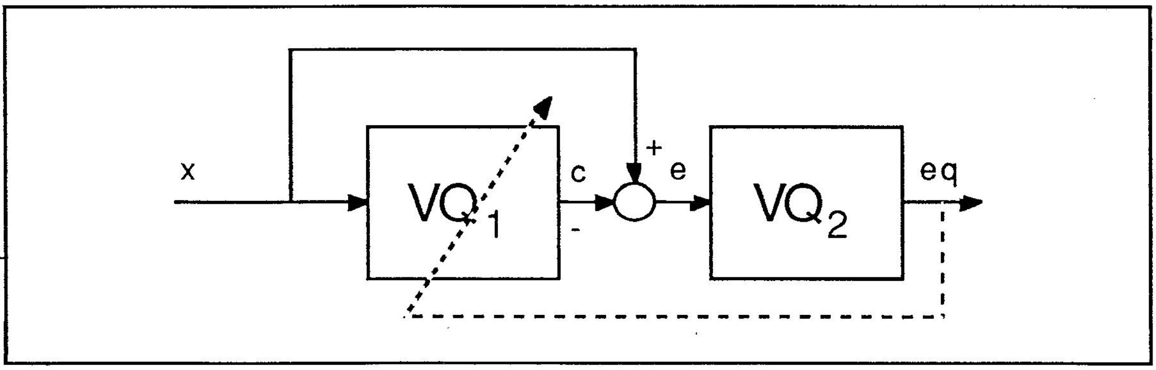

Figure 1. AMSVQ system with 2 stages

Then eq, that is an estimation of the quantization error of the first codebook, can be used to adapt the quantizer in the following way:

$$c = c + \mu\, e_q \quad (4)$$

where μ is the adaptation factor. This adaptation is made without the need of increasing bit-rate and is computationally very simple. It can be seen as a modification of the centroid in the direction of the new quantized vector and, in this sense, its closely related with the LBG algorithm [Linde 1980]. The difference between both algorithms is that the AMSVQ system is continuously changing the centroids and uses an estimation of the error (or input vector), while the LBG algorithm changes the centroid after classifying all the training sequence and uses the actual input vector during this training phase.

Results. The AMSVQ system has been applied to the quantization of LPC parameters with important improvements respect to the conventional multi-stage structure. Quantization of the LPC coefficients has been extensively studied. Generally, the inverse sines of the reflection coefficients or the log-area ratio values are quantized. We chose the log-area ratio values with the mse distortion for our experiments.

The results have been obtained using a codebook of 8 codewords (3 bits/frame) in the first stage and 3 codebooks of 256 codewords (8 bits/frame) in the second, third and fourth stage. The number of bits for each frame is thus 27 (3+8+8+8) and at 50 frames/s, the bit rate equals 1350 bps. To obtain a fast adaptation and robustness against channel errors, only the 8 codewords of the first quantizer are adapted.

The designed codebooks have been used to quantize a test sequence with different values of the step size μ. In figure 3 it is shown the increment in SNR respect to 19 dB of the conventional, ($\mu = 0$), multistage vector quantizer. The best results are obtained with μ=0.4 and the AMSVQ system designed for μ = 0.1. Nevertheless smaller values of the step size as μ=0.1 or μ=0.01 give also a significant increase of the performance and can be a good choice for providing robustness against channel errors.

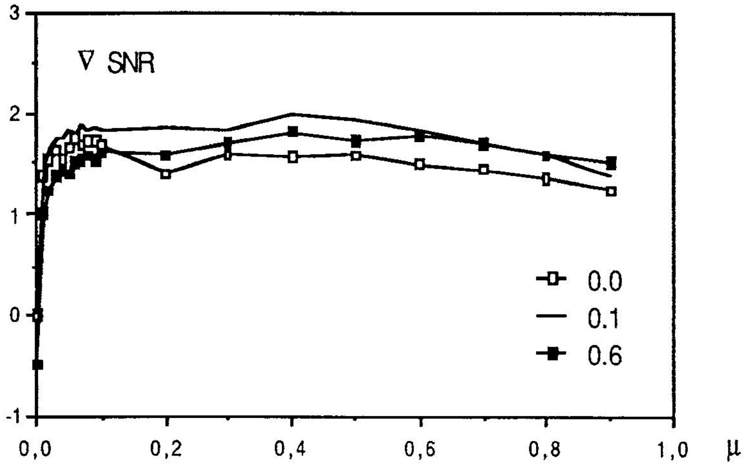

Figure 5. Increase in LAR SNR for the three different designed AMSVQ systems and step sizes between 0.0 and 1.

This quantization of the LPC parameters has been included in a multipulse coder at 9.6 Kbps. The results show that the coded speech can hardly be distinguish from the one obtained with unquantized coefficients.

REFERENCES

[CHEN 1987] "Gain Adaptive Vector Quantization with application to Speech Coding". Juin-Hwey C. CHEN, Allen GERSHO. IEEE Trans. on COM, No. 9, Vol. 35, p. 918-930, September.

[CUPERMAN 1985] "Vector Predictive Coding of Speech at 16 Kbps". Vladimir CUPERMAN, Allen GERSHO. IEEE Trans. on COM, No. 7, Vol. 23, p. 685-696, July.

[JOHNSTON 1980] "A Filter Family Designed for use in Quadrature Mirror Filter Banks". J. D. JOHNSTON. Proc. ICASSP, p 291-294, Denver, April.

[LINDE 1980] "An algorithm for Vector Quantizer design". Yoseph LINDE, Andrés BUZO, Robert M. GRAY. IEEE Trans. on COM, No. 1, Vol. 28, p. 84-95, January.

[MASGRAU 1986] "A continuosly Adaptive Vector Predictive Coder (AVPC) for Speech Encoding". Enrique MASGRAU, José B. MARIÑO, Francesc VALLVERDU. IEEE Proc. ICASSP, No. 56.01, p. 3027-3030, April.

[MASGRAU 1987] "AVPC-Subband Coding system for Speech Encoding". Enrique MASGRAU, José B. MARIÑO, José A. RODRIGUEZ FONOLLOSA, Josep SALAVEDRA MOLI. Proc. , p. 189-192, September.

[MASGRAU 1988] "Subband splitting, Adaptive scalar Prediction and Vector Quantization". Enrique MASGRAU, José A. RODRIGUEZ FONOLLOSA, José B. MARIÑO. EURASIP Proc. EUSIPCO, Vol. 2, September.

[PAUL 1983] "An 800 bps Adaptive Vector Quantization Vocoder using a perceptual distance measure". Douglas B. PAUL. IEEE Proc. ICASSP, No. 02.04, p. 73-76, April.

[RODRIGUEZ 1989] "Adaptive Multistage Vector Quantization" José A. RODRIGUEZ FONOLLOSA, Enrique MASGRAU GOMEZ. IEEE Proc. MELECON, Lisboa, April.