SIGNAL PROCESSING IV: Theories and Applications
J.L. Lacoume, A. Chehikian, N. Martin, and J. Malbos (Editors)
Elsevier Science Publishers B.V. (North-Holland)

SUBBAND SPLITTING, ADAPTIVE SCALAR PREDICTION AND VECTOR QUANTIZATION FOR SPEECH CODING

E. Masgrau-Gómez, J. A. Rodríguez-Fonollosa and J. B. Mariño Acebal

Dept. T.S.C. E.T.S. Ingenieros de Telecomunicación.
Apdo. 30002. 08080 BARCELONA. SPAIN

This paper describes a new coding structure based on the combination of Vector Quantization, Linear Prediction and Subband Splitting that achieves high quality speech at rates below 10 Kbit/sec. In this scheme, a vector is formed with one sample of the normalized prediction error of each band and then a vector quantizer is applied to it. This quantization of the prediction error allows to use scalar adaptive predictors while conserving the advantages of the vector quantization. The necessary noise shaping for achieving high subjective quality is obtained by the use of a Frequency-Weighted distance in the vector quantizer

1. INTRODUCTION

Vector Quantization and Linear Prediction have been shown as interesting techniques for speech coding at rates below 10 Kbit/sec and most of the current research in speech coding is based on them. In the scheme we present these techniques are combined with a previous subband splitting.

Subband Coding (SBC) of speech was introduced by Crochiere et al [1] in 1976 and its known to produce good quality at medium bit rates in combination with linear prediction and dynamic bit allocation [2]. However the performance of the adaptive predictors using scalar quantizers deteriorate rapidly at bit rates below 16 Kbit/sec and the speech quality becomes inadequate for many applications.

Some schemes using vector quantization to encode the prediction error have been presented in the last few years [3,4,5,6]. The scheme we present in this study has been developed with the idea of improving the perceptual quality while conserving a reasonably complexity.

In contrast with other high quality schemes [7] that work at a forward block mode and need an important transmission of side information, in our scheme a backward sample by sample adaptation is used and only the codeword index is transmitted.

Figure 1 shows the scheme of the proposed coder. Each band has a predictor and a gain estimator operating in a backward mode as in most ADPCM systems. 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).

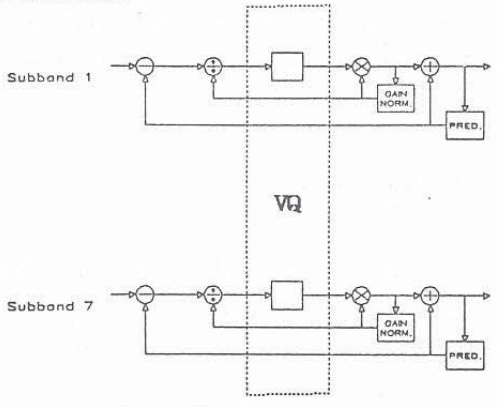

Figure 1.- General scheme of the coder.

2. CODER STRUCTURE

2.1. Subband Filtering

Subband coding offers several advantages. It removes part of the redundancy in the speech and provides a set of uncorrelated signals to quantize. As a frequency domain coding technique arbitrary forms of noise shaping can be obtained and subjective quality optimized.

Normally the reconstruction error variance of each band is controlled by the bit allocation that can be static (fixed) or dynamic (adaptive). In the scheme introduced in this paper the problem of dynamic bit allocation is avoided by the simultaneous vector quantization of all the bands to be transmitted. The distance used in the search of the optimum vector allows to control the variance of each band and the fidelity of each band is not restricted to a finite set of bit rates.

The filter banks used are tree-structures of quadrature mirror filters and all the bands are sampled at the same rate to permit their simultaneous quantization. The input signal is sampled at 8 Khz and each subband output is sampled at the Nyquist rate corresponding to its baseband representation. If we want to use Vector Quantizers with a number of codewords (vectors) in the range from 256 to 1024 (8-10 bits) then we have to split the signal in 8 bands to obtain transmission rates in the range from 8 to 10 Kbit/sec. The highest frequency band (3500-4000 Hz) which is highly damped due to anti-aliasing filtering is not transmitted.

2.2. Linear Prediction

Since the subband splitting of the speech signal does not remove completely the autocorrelation of each band, we can improve the performance of the coder by using linear prediction. This is specially true in the lower bands where the pitch and formant structure are better defined.

We have chosen adaptive linear prediction operating in a backward mode with the GAL algorithm [4]. We tried different lengths for the predictors of each band and we found that predictors of order 9 were suitable for the bands 1 and 2 (0-0.5, 0.5-1 KHz) while for the bands 3 and 4 (1-1.5, 1.5-2 KHz) the performance do not improve for orders greater than 4 and 2 respectively.

In the bands 5, 6, 7 (2-2.5, 2.5-3, 3.5-4 Khz) no predictor is used because simulations showed that no improvement in quality was obtained when it was included. This fact reduces the number of taps to be updated to 24 each 8 samples of speech (3 per sample).

2.3. Gain normalization

The prediction error of each band is normalized previously to its quantization to reduce its dynamic range. For the estimation of the standard desviation a simple recursive estimator is used. As it works in a backward mode there is no need of side information transmission. The update of the estimation is made in the following form:

$$s_i(n) = \beta s_i(n) + (1-\beta)|eq_i(n-1)| \quad (1)$$

where $s_i(n)$ is the estimation of the standard desviation of band i, $eq_i(n-1)$ the last quantized prediction error of band i and β a parameter that control the memory or window lenght of the estimator. Simulations show that, in our case, β=0.8 is a good value in all the bands.

Then the prediction error is normalized by a division.

$$g_i(n) = s_i(n) + mg_i \qquad (2)$$

$$z_i(n) = e_i(n) / g_i(n) \qquad (3)$$

where the term mg_i prevent z_i from becoming too large in silence periods. After the quantization $zq_i(n)$ is obtained and the quantized prediction error is denormalized by a multiplication

$$eq_i(n) = zq_i(n)\, g_i(n) \qquad (4)$$

If the objective were to minimize the variance of the reconstruction error then the distance used by the quantizer to select the optimum vector had to be

$$d = \sum_{i=1}^{M} (eq_i(n) - e_i(n))^2$$

$$= \sum_{i=1}^{M} g_i^2(n)\, (zq_i(n) - z_i(n))^2 \qquad (5)$$

where M is the number of bands to quantize.

This distance would produce a flat reconstruction error spectrum and would maximize the signal to noise ratio, but not the subjective quality. To produce some noise masking, an error weighting is added to the distance definition as it is usually done in the process of finding the optimum bit allocation in Subband and Transform Coding. It consists in replacing g_i^2 by $w_i g_i^2$ to obtain an error spectrum that is now described by a constant $w_i d_i^2$ rule.

In our case the distance is defined as

$$d = \sum_{i=1}^{M} w_i(n) g_i^2(n)\, (zq_i(n) - z_i(n))^2 \qquad (6)$$

where the weighting w_i has the form

$$w_i(n) = wo_i (g_i^2(n))^{\lambda} \qquad (7)$$

The fixed term wo_i is chosen to enhance the bands where the error perception is greater and λ is chosen in a manner such the quantization noise is more efectively masked by the speech signal.

2.4. Vector Quantizer

The codebook design is carried out by the LBG algorithm with the splitting technique to obtain the starting codebook. Taking into account the distance (6) the centroid is given by

$$c_i = \frac{\sum_{n=1}^{N} W_i(n)\, z_i(n)}{\sum_{n=1}^{N} W_i(n)} \qquad (8)$$

$$W_i(n) = w_i(n)\, g_i^2(n)$$

The training set used by the LBG algorithm ($W_i(n)$ and $z_i(n)$ in our case) is obtained making the system work without quantization. The input of the quantizer is different when the encoder works with quantization because of the error feedback, but if this feedback is small the training set obtained by the aboved method is representative.

3. RESULTS

3.1. Data Base

In order to carry out our experiments we select 40 fonetically balanced spanish utterances from male and female speakers. 18 of them were used to carry out the design of the Vector Quantizer and the others to test the coder.

The results presented are the average over the data base and they were obtained with codebooks

of 256 vectors (8 Kbit/sec) and 1024 vectors (10 Kbit/sec).

3.2. Prediction Gain

In Table I is shown the prediction gain and the segmented prediction gain for each of the 4 lower bands. Although the prediction gain is very low in the bands 3 and 4, the simulations show that the subjective quality improves when they are used.

Band	Pred.Gain	SEG. Pred.Gain
1	8.0/8.2 dB	6.8/7.4 dB
2	4.8/4.9 dB	2.3/2.8 dB
3	1.7/1.8 dB	0.1/0.3 dB
4	1.1/1.2 dB	0.1/0.3 dB
	6.0/6.2 dB	5.6/5.9 dB

Table I. Prediction Gain (8/10 Kbps)

3.3. Frequency Weighting

Several simulations were carried out with differents weightings and the best subjective results were obtained with

$$wo = (1.0,1.0,1.3,1.5,1.5,1.0,1.0)$$
$$\lambda = -0.3$$

Table II shows the SNR and SEGSNR of the whole signal and of each of the seven transmitted subbands, achieved with the above weighting.In figure 2 the original signal and the coded one at 8 Kbit/s can be compared in the beginning of a voice segment.

Band	SNR	SEGSNR
1	18 / 20 dB	17 / 19 dB
2	14 / 17 dB	10 / 11 dB
3	9 / 12 dB	5 / 7 dB
4	8./ 10 dB	5 / 6 dB
5	6 / 8 dB	3 / 5 dB
6	2 / 4 dB	1 / 3 dB
7	2 / 4 dB	1 / 1 dB
	14 / 17 dB	14 / 16 dB

Table II. SNR results at 8/10 Kbps

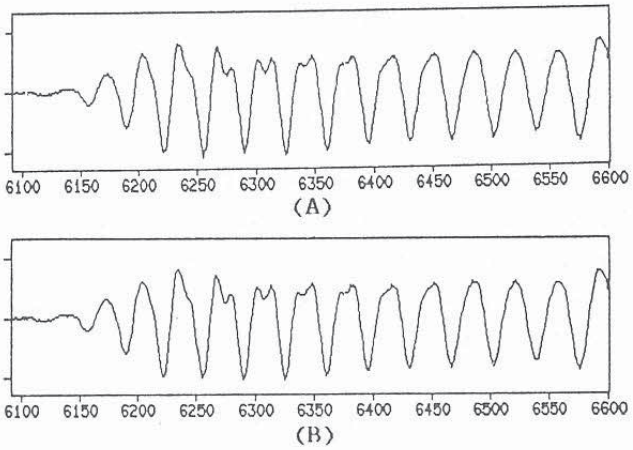

Figure 2. Original (A) and coded signal (B) at 8 Kbit/sec.

4. . CONCLUSIONS

The proposed coder achieves high quality a 8 Kbit/sec and very high quality at 10 Kbit/sec. It reduces the complexity of the AVPC-SBC [6] and removes the vectorial predictor used in previous AVPC systems [4], [5], [6] while conserving a backward sample by sample operation mode.

REFERENCES

[1] R.E. Crochiere, S.A. Webber and J.L. Flanagan, "Digital Coding of Speech in Sub-Bands", Bell System Technical Journal, vol 55, OCT-1976

[2] F.K. Soong, R.V. Cox and N.S. Jayant, "Subband Coding of Speech Using Backward Adaptive Prediction and Bit Allocation", Proc. ICASSP, paper 43.1, 1985.

[3] Vladimir Cuperman and Allen Gersho, "Vector Predictive Coding of Speech at 16 Kbit/sec" IEEE Trans. on Com., Vol COM-33, No. 7 JUL-1985.

[4] E. Masgrau, J. Mariño and F. Vallverdú, "Continuosly Adaptive Vector Predictive Coder (AVPC) for Speech Encoding", Proc. ICASSP, paper 56.1, APR-1986.

[5] Juin-Hwey Chen and Allen Gersho, "Vector Adaptive Predictive Coding of Speech at 9.6 Kb/s", Proc. ICASSP, paper 33.4, APR-1986.

[6] E.Masgrau Gómez, J.B.Mariño Acebal, J.A. Rodríguez Fonollosa and J. Salavedra Moli,"AVPC-Subband coding System for Speech encoding" Proc. Europ. Conf. Speech Technology , Edinburg 1987.

[7] B.S. Atal, "High-Quality Speech at low bit rates: Multi-Pulse and Stochastically Excited Linear Predictive Coders" Proc. ICASSP, paper 33.1, APR-1986.