

AVPC-SUBBAND CODING SYSTEM FOR SPEECH ENCODING

E. Masgrau-Gómez*, J.B. Mariño*, J.A. Rodríguez-Fonollosa*,
J.Salavedra-Molí.

ABSTRACT

The combination of Vector Quantization (VQ) and predictive techniques -named AVPC systems- has been shown as interesting systems for speech waveform coding at medium-high rates (ref. 1 to 3). In this work we present a such system including a previous four subbands splitting. This allows good quality of speech at low-medium rates (1-1.25 bits/sample). A comparative study show that the AVPC-SBC outperforms the simple AVPC coder.

INTRODUCTION

By using the data compression ability of the VQ, it is possible to reduce the transmission rate meanwhile the speech quality is maintained. The Adaptive Vector Predictive Coder (AVPC) is a system that uses the vectorial quantization; it was previously introduced by Cuperman and Gersho (ref. 1), and it is basically a vector version of a DPCM system. The authors had proposed in some previous works (ref. 2 and ref. 3) a continuously and fully adaptive version consisting in the use of a vector-by-vector adaptive vector predictor in the feedback loop and an adaptive gain-shape VQ. The gain or energy of the prediction errors vectors was estimated by frames, and quantized as side information. The vector shape was quantized by a fixed codebook. The continuous adaptation to the variant statistic of the input speech provides good results at 16 Kbps as in SEGSNR as subjective quality. The obtained quality is not enough at lower rates, for example at 9.6 Kbps.

The improvement of the speech quality at these rates requires in the first place a better gain estimation. In this paper we propose the use of a backward adaptive predictor working on the gain sequence of the past error vectors, in order to implement a vector-by-vector gain estimation. The backward characteristic of this estimation allows the saving of the side transmission rate.

In the second place, vector shape is badly reproduced due to the small number of codeword, producing a poor coding of the high frequency frames. The use of a previous subband splitting affords the same coding fidelity for both high and low frequency frames, providing a higher speech quality. This system, named AVPC-Subband Coding (AVPC-SBC), will be described later.

* Dpto. Teoría de la Señal y Comunicaciones. E.T.S.I. Tele
comunicación (U.P.C). Apdo. 30002, 08080 Barcelona, SPAIN.

This work is supported by CAICYT - Plan No 2906/83C3.

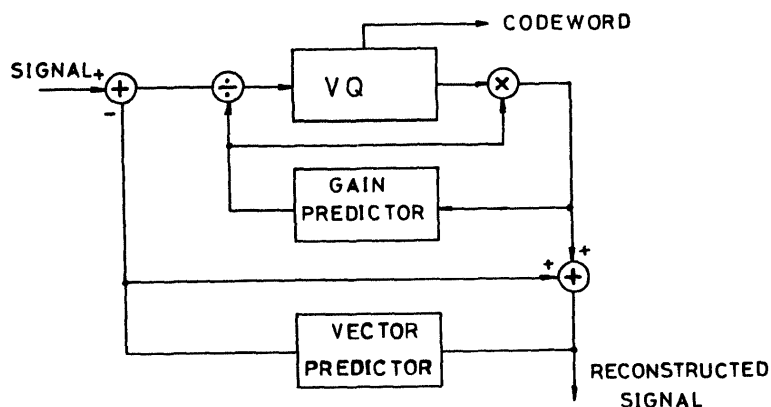


Figure 1. Transmitter scheme of the AVPC system.

THE AVPC SYSTEM

Figure 1 shows the scheme of the proposed AVPC system. We can clearly distinguish two main blocks: the Adaptive Vector Predictor (AVP) and the Adaptive Vector Quantizer (AVQ). The predictor goal is to obtain a prediction error as small as possible. The so-reached reduction of the dynamic range of the signal to be quantized allows a more accurate representation at the VQ output. The predictor operates in a backward configuration. Thus, the vector by vector coefficient adaptation can be reproduced at the receiver side without sending any side information. Two algorithms for predictor adaptation were proposed in (ref. 2 and 3): the VLMS and the VGAL (see these ref. for details). The inherent data orthogonality provides a better prediction performance for VGAL (+1.5dB in forward SEG prediction gain over the VLMS) (ref. 3).

The AVQ is a gain-shape product quantizer. The gains or L_2 norms of the vectors to be quantized are estimated by means of an adaptive backward predictor, using the gain of the output quantized vectors in place of the original vectors. The predictor update can be driven by any adaptive algorithm, i.e. the LMS or the GAL. The prediction error vectors are normalized by the gain estimation and the resultant vectors are quantized by a fixed VQ. The codebook design is accomplished by the LBG algorithm from a training data-base. In order to maximize the signal-to-noise ratio, the centroids are calculated taking into account the estimated norm of every vector (ref. 2).

THE AVPC-SBC SYSTEM

The previous AVPC system works well at rates about 2 bits/sample. In this case, the number of codewords in the codebook is enough for the vector shape coding 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 shape vector. This is true even though the gain was estimated without error. In this case the decoded speech showed 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 path, and, therefore, a greater fidelity is obtained. This subband splitting can not be carried out on the predictor error because the bank filter delay. So, 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 (ref. 4); then, each band output is coded by an AVPC system similar to that we have described before. The bits are assigned to each band according to its energy. The predictivity of the signal in every subbands is limited due the spectral withening effect caused by the filtering and decimation process. This effect is greater in third and fourth bands.

RESULTS

In order to carry out our experiments fonetically balanced Spanish utterances from male (M_) and female (F_) speakers were selected. Every speaker uttered several phrases. Part of them made up the data base for designing the codebooks; the rest was used as outside test for the different coding schemes considered in this paper. As it is known, in order to design codebooks providing a good representation of signals it is necessary to take a great amount of signal vectors for every codebook. This condition requires a greater data base for designing the AVPC-SBC system than for designing the AVPC coder.

The following coders were designed: AVPC system at 1, 1.2 and 2 bits/sample (8, 9.6 and 16 Kbps) and AVPC-SBC coder at 1, 1.25 and 1.5 bits/sample (8, 10 and 12 Kbps). The dimension vector was fixed at $k=4$ except for the AVPC coder at 9.6 Kbps where it was taken $k=5$. The algorithm of predictors was chosen the VLMS one with order $p=2$; however the VGAL algorithm was implemented in the AVPC system working at 16 Kbps. The reason of this modification was because the VGAL algorithm exhibits a threshold effect that allows to take advantage of this algorithm only when the segmented signal-to-noise ratio rises over 15 dB (ref 3). The gain predictor was LMS with order $p=10$. In table I the performance of these coders is assessed by means of SEGSNR for every couple of coder and speaker; scores for both inside and outside the data base are included. As we can see, the coder performance depends greatly on the speaker; this variability is typical in coders that use signal prediction. In order to asses the predictibility of the speech material we have in our data bases, table I includes the SEGSNR afforded by the standard CCITT-ADPCM at 32 Kbps.

The AVPC system at 2 bits/sample provides a good subjective speech quality for every speaker, even outside the data base. The masking effect that the signal exercises over the noise, makes the subjective quality prtactically independent over the SEGSNR. However, at 1 and 1.2 bits/sample the AVPC degrades noticeably its performance.

The AVPC-SBC coder behaves at low bit rate better than the previous system, and provides a much better subjective quality than the AVPC coder at comparable bit stream. The AVPC-SBC behaviour depends greatly on the bit assignment for each subband. In our experiments the bit assignment was the following: 2,2,1,1 bits/sample at 8Kbps; 2,2,2,1 bits/sample at 10Kbps and 2,2,2,2 bits/sample at 12Kbps. In the first case, subband 3 and 4 were badly codified because they could be quantized only with sixteen code words. At 10Kbps the third subband is correctly codified, being the subjective quality very good; at 12Kbps this quality is excellent. In order to optimize the subband bit assignment were carried out several informal test; the assignment 2,1.75,1.75 and 1.25 bits/sample shows a good compromise between bit rate (9.75 bit/sample) and subjective quality.

SPEAKER CODER	INSIDE			OUTSIDE					
	F1	F2	F3*	M1	M2	M3	F4	M4	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.

CONCLUSIONS

The AVPC-SC coder affords a good quality of speech at 8-10 Kbps rates. The subbands splitting provides a good fidelity of high frequency frames, unlike of the AVPC case, where the VQ design method causes a great smoothing of these. An effort must made in order to reach dynamic bit allocation strategies. In other hand, the vector predictor has shown a bounded performance. At present, we are experimenting a new and simpler subbands vector predictive coder, including scalar prediction and dynamic bit allocation. The results will be published in further papers.

REFERENCES

- /1/ V. Cuperman et al. COM-33, pp.685-696, (july 1985)
- /2/ E. Masgrau et al. ICASSP-86, pp.3027-3030. Tokyo, (1986).
- /3/ E. Masgrau et al. EUSIPCO-86, pp.431-434. The Hague (1986).
- /4/ J.D. Johnston. ICASSP-80, pp.291-294. Denver, (1980).

* Different sentences corresponding to the same speaker.