MELECON '85 / Volume II: Digital Signal Processing
A. Luque, A.R. Figueiras Vidal, V. Cappellini (eds.)
Elsevier Science Publishers B.V. (North-Holland) / © IEEE 1985

ADAPTIVE PREDICTION AND BIT-ASSIGNMENT IN SUBBAND CODING OF SPEECH

José B. Mariño-Acebal, Jaume Marti-Ros

E.T.S.I. Telecomunicación
Universidad Politécnica de Cataluña
C/Jorge Girona Salgado, s/n
08034 Barcelona - Spain

The combination of time-domain harmonic scaling (TDHS) and sub-band coding (SBC) provides an encoding approach which allows 9.6 Kb/s speech encoding with good communication quality. Starting from this structure, this paper focuses the improvement of earlier designs |1|. It is shown that adaptive prediction and bit-assigment enhances the sub-band signal coding and, hence, the performance of the overall system.

The prediction is realized by an adaptive lattice, the algorithm being GAL2 |2|. The dynamic bit allocation takes place from the step-sizes of the backward adaptive quantizers (Jayant, |3|) in each sub-band. Improvements as high as 5 dB can be achieved for the average segmented signal to noise ratio.

I. CODER STRUCTURE

Figure 1 shows the basic block diagram of the coder. The 8KHz sampled speech is compressed in bandwidth and sampling rate by the TDHS algorithm |4|. This takes advantage of the inherent periodicity of speech by a pitch synchronous process; two pitch periods of speech are interpolated into one period to realize an effective signal compression factor of two. So, the signal after the TDHS block is sampled to 4 KHz. The compressed speech is now codified by the subband coder; the signal is

Band	2(frequency range)/f_s
1	0 - 0.125
2	0.125 - 0.25
3	0.25 - 0.5
4	0.5 - 0.75

f_s = sampling rate

Table I.- Normalized frequency range for each band.

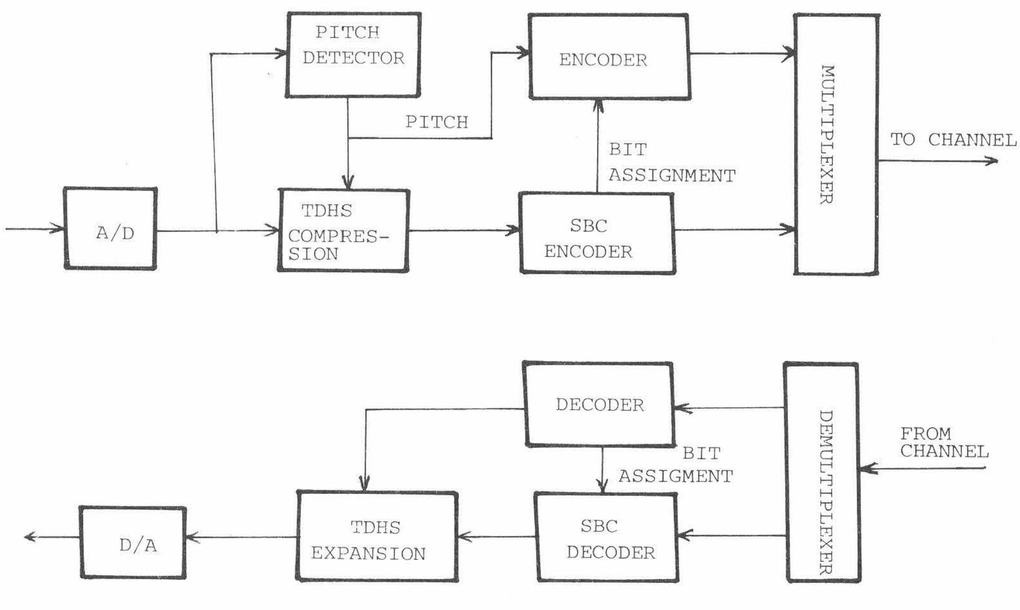

Figure I.- Block diagram of the coding system divided by QMF filters into four bands, whose frequency ranges are pointed out in Table I; then, each band is coded by an ADPCM coder; the bits assigned to each band are adapted according its power level, so that we can optimize the signal quantization. Finally, the generated bit stream is multiplexed with the side information: the pitch values and and the utilized bit-allocation.

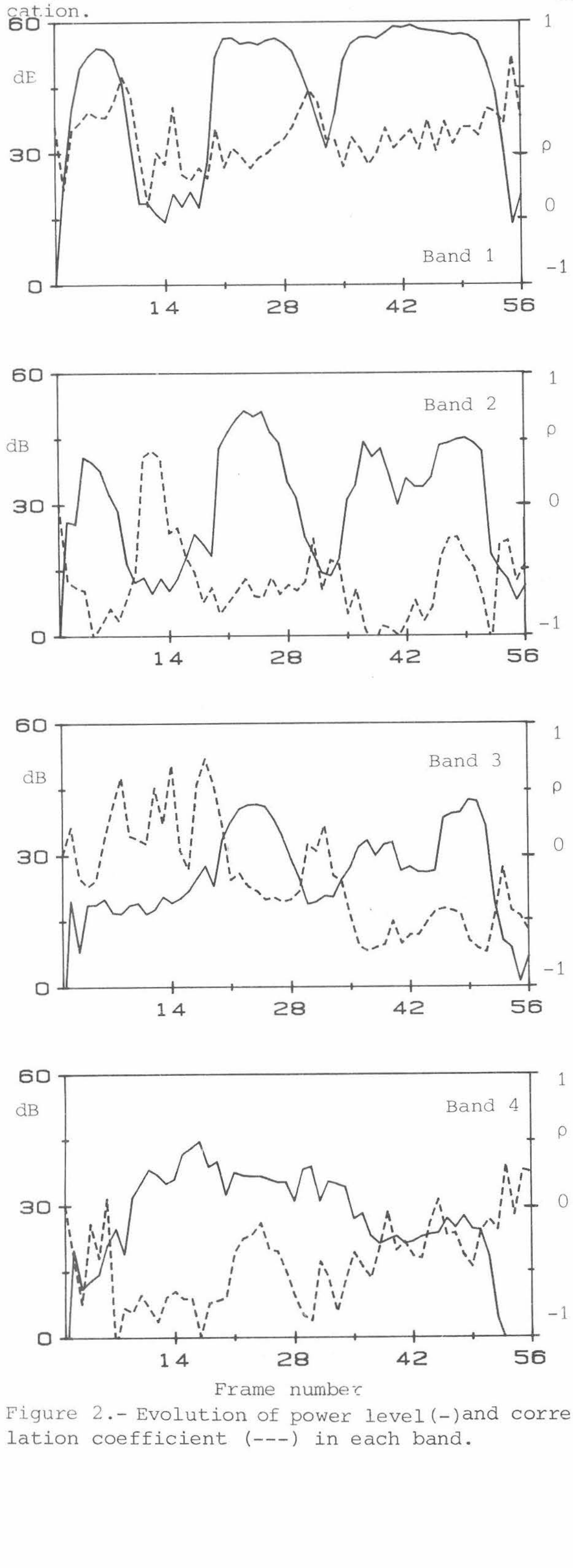

Figure 2.- Evolution of power level (-) and correlation coefficient (---) in each band.

This structure profits from both the pseudoperiodicity and the spectral propierties of speech, in order to achieve a good quality coding with a low bit rate stream.

II.SUBBAND SIGNAL CODING

As we can see in Figure 2, in each band the power level and correlation coefficient

$$\rho = R(1)/R(0)$$

suffer great variations with time. Then, it appears convenient to use adaptive quantization and prediction in order to implement each subband signal coder. We are adopted a backward configuration, which is shown in Figure 3. The quantizer adapts its step-size according to the robust version |5| of the one-word memory algorithm from Jayant |3|; the most important parameters of the quantizers are summarized in Table II. The predictor is implemented by a lattice network, that adapts its constants by means of the GAL 2 algorithm; this choice is motived by the results from the comparative study of different algorithms reported in |2|.

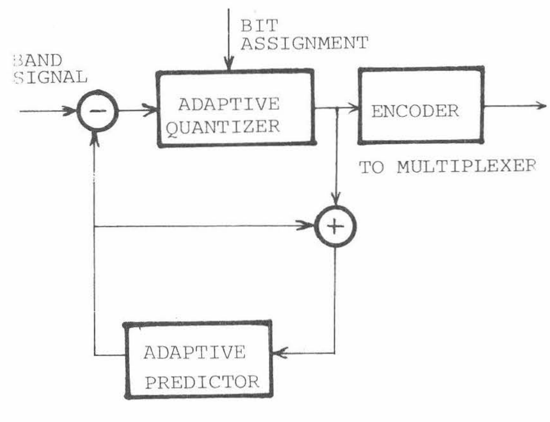

Figure 3.- Backward adaptive ADPCM

attack-time: 4.3 dB/sample
decay-time: -1.7 dB/sample
dynamic range: 42 dB
leaky factor: 63/64
relative power level

band 1: 0 DB
band 2: -3 dB
band 3: -6 dB
band 4: -9 dB

Table II.- Adaptive quantizer parameters

If n_i is the number of bits assigned to the quantizer of the i-th band and f_s is the sampling frequency of the signal entering the SBC coder, the bit rate originated by the speech quantization is

$$R = \left[(n_1+n_2)+2(n_3+n_4)\right]\ 0.125\ f_s \quad b/s$$

In our design we have taken

$$n_1+n_2+2(n_3+n_4) = 16 \qquad (1)$$

This means

$$R = 8\ \text{Kb/s}$$

For voiced segments of speech, the first and second bands have much more power than the third and fourth bands; so, n_1 and n_2 have to be greater than n_3 and n_4, in order to control properly the spectrum shape of the quantization noise. Unvoiced segments signify frequently an opposite situation (Figure 2 includes an example); hence, we will take higher values for n_3 and n_4 than for n_1 and n_2. As a consequence, we have incorporated adaptive bit-allocation to the coder; the following restriction:

$$2 \leqslant n_i \leqslant 4 \qquad (2)$$

is added because:

a) it is necessary to insure that the backward adaptive quantizer works properly ($n_i>1$);

b) we have found that 4 bits offer in our case a ADPCM decodified signal indistinguishable from the original one at perceptive effects.

Table III gives all the possible sets of bit--assignment that satisfy conditions (1) and (2). The adaptive bit-allocation is carried out by a properly modified version of the algorithm introduced in |6|; initially, 2 bits are assigned to each band and each residual subband power is estimated by the step-size of the corresponding quantizer; then, a bit is added to the subband with the largest step-size and this step-size is divided by two; this operation is repeated until the bit-assignment is complet. The bits are allocated for each block of 40 samples that enters the SBC, corresponding to 10 ms. of speech; the quantizer step sizes at the end of a block provide the bit-assignment for the next block of samples.

III. FRAMING OF THE DATA

The quantization bit stream and the side information are multiplexed into frames of 96 bits; each frame corresponds to 10 ms of encoded speech. Figure 4 summarizes the structure for one frame data.

	Band 1	Band 2	Band 3	Band 4	%
Bit allocation	4	4	2	2	65.0
	4	2	3	2	5.6
	4	2	2	3	11.0
	3	3	3	2	3.7
	3	3	2	3	4.9
	2	4	3	2	2.1
	2	4	2	3	--
	2	2	4	2	0.3
	2	2	3	3	1.6
	2	2	2	4	5.8

Table III.- Allowed bit allocations.

The first six bits serve as a synchronization header. The next six bits carry the pitch value used by the TDHS algorithm. The next set of 4 bits codifies the current bit-assignment; because the step-sizes are known at the receiver side, it is not mandatory to send this information; however, it is included in the frame to prevent that one transmission error in the SBC data produces an incorrent bit-allocation at the receiver end; this error would be uncorrectable from the data. Finally, the five groups of 16 bits encode the SBC data.

IV. RESULTS

Two sentences have been utilized in order to test the encoder system: one from a Catalan female speaker ("Aixó es una prova de codificació de veu") and another from a Spanish male speaker ("Esto es una prueba de codificación de voz").

The performance of the SBC encoder have been analyzed by codifying speech directly; a 16Kb/s rate stream results. TableIV summarizes the found segmented signal to noise ratio for different proves: without predictor, adaptive predictor for bands one and two, adaptive predictor for all the bands, without adaptive bit--assignment, adaptive bit-assigment from the step-size values (Δ) and from the estimate

$$\hat{\sigma} = \sqrt{\frac{1}{N}\sum_i x_i^2}$$

for each subband power level. For every case, the reached SNR for female/male sentences are given. In Table III is reported the utilization percentage for each bit-allocation set when the adaptive bit-assigment is active; as a consequence of this result, when the adaptive bit-alloca-

96 bit frame

6	6	4	16	16	16	16	16
SYNC	PITCH	BIT ASSIG:	SBC	SBC	SBC	SBC	SBC

Figure 4.- Frame structura for the 9.6 Kb/s stream.

Bit Assignment	Predictor		
	None	Bands 1 and 2	Band 1 to 4
None	16.4 / 15.6	21.8 / 17.4	22.2 / 17.6
Δ	17.2 / 15.9	22.4 / 17.7	22.9 / 17.8
$\hat{\sigma}$	17.0 / 15.9	22.2 / 17.6	22.7 / 17.8

Table IV.- Segmented SNR of the SBC encoder obtained in our experiments.

tion is not used, the sellected set is the first one. From table III, table V shows the statistics of the dynamic bit-assignment. Some conclusions can be drawn out:

a) Adaptive prediction provides a more significant improvement than adaptive bit-allocation. This result has a simple explanation: the predictor works for all the speech segments (voiced, unvoiced, loud and soft); however, although the dynamic bit-allocation is able to afford gains as high as 10 dB for same speech segments, it really works a small percentage of time.

b) The quantizer step-sizes constitute a correct basis to implement the adaptive bit assignment.

c) Female speech is more efficiently encoded than male speech.

d) Predictors for three and four bands provide a small encoder quality enhancement.This is because the low average power level of these bands.

At the time of writing this paper, we are evaluating the overall system by informal listening tests. The results will be reported at the conference.

Band	Percentage of bit allocat. 2	3	4	Average bit assign.
1	9.8	8.6	81.6	3.72
2	24.3	8.6	67.1	3.43
3	86.7	13.0	0.3	2.13
4	76.7	17.5	5.8	2.30

Table V.- Dynamic bit assignment percentages.

REFERENCES

|1| R.E. Crochiere et al. "A 9.6-Kb/s DSP speech coder" BSTJ, vol 61, pp 2263-2288 (November 1982)

|2| E. Masgrau et at. "Performance comparison of 32-Kb/s ADPCM coders in noise free and noisy environments" Proc. Melecon'85

|3| N.S. Jayant "Adaptive quantization with a one-word memory" B.S.T.J., vol 52, pp. 1119-1144 (September 1973)

|4| D. Malah "Time domain algorithms for harmonic banwidth reduction and time scaling of speech signals" IEEE Trans. ASSP-27, pp 121-133 (April 1979)

|5| D.J. Goodman, R.M. Wilkinson "A robust adaptive quantizer" IEEE Trans., CDM-23, pp. 1362-1365 (November 1975)

|6| T.A. Renstad "Subband coder with a simple adaptive bit allocation: a possible candidate for digital mobile telephony" Proc. ICASSP, 1982, pp. 203-207.