COMPLEXITY REDUCTION FOR FS-1016 WITH MULTISTAGE SEARCH

M. MAUC, G. BAUDOIN and M. JELINEK

ESIEE, B.P. 99 - 93162 NOISY-LE-GRAND CEDEX - FRANCE

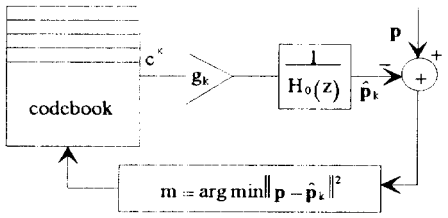
ABSTRACT

CELP speech coders have proved to be efficient for the coding of speech at medium and low bit rates. The NSA has recently introduced the Federal Standard 1016, CELP speech coder at 4800 bps. We propose two algorithms that reduce the complexity of this coder by a factor 2. The first algorithm uses the particular structure of the standardised code book to simplify the calculation of the cross-correlation term. The second algorithm is a multistage method by subsampling that eliminates non relevant code book sequences for the calculation of the energy term.

1. INTRODUCTION

Since their introduction by M.R. Schroeder and B.S. Atal [1], CELP coders have proved to be very efficient for the coding of speech at medium and low bit rates. Several methods have been proposed in order to reduce the complexity of these hybrid coders [2] [3] [4].

Figure 1 represents one analysis by synthesis loop.



We use the following notations :

N is the frame size. N' is the sub frame size with N' =N/4. c^k is the k^{th} code book sequence of length N'. g_j is the optimal corresponding gain. p is the perceptual memoryless speech signal of length N'. $\hat{p}_k$ is the synthetic perceptual memoryless signal corresponding to the response of the synthesis filter $1/H_0(z)$ to the code book sequence c^k. T is the number of sequences of the code book. P is the prediction order. H the impulse matrix response.

The optimum CELP sequence c^j minimises the Least Square Error E(k) between p and $\hat{p}_k$:

$$E(k) = \left\| \mathbf{p} - \hat{\mathbf{p}}_k \right\|^2 \quad (1)$$

or equivalently maximises the following performance criterion :

$$PC(k) = \frac{\left\langle \mathbf{p}, \mathbf{H}\mathbf{c}^k \right\rangle^2}{\left\| \mathbf{H}\mathbf{c}^k \right\|^2} = \frac{\beta_k^2}{\alpha_k^2} \quad (2)$$

where (3)

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{n=0}^{x} u_n v_n \quad \text{with } x = N'-1$$

Let j be the index of the best code book sequence then :

$$j = \arg(\max PC(k)) \quad (4)$$

The term β_k is the cross-correlation between **p** the vector of length N' of the original speech and $\hat{p}_k$ the synthetic speech. It can be written :

$$\beta_k = \left\langle \mathbf{p}, \mathbf{H}\mathbf{c}^k \right\rangle = \left\langle \mathbf{q}, \mathbf{c}^k \right\rangle \quad (5)$$

where **q** is a sequence calculated once per frame and does not depend on the code book sequence c^k .

The computation task of the best code book sequence search can be lessen combining two different methods: The first method concerns the calculation of the cross-correlation term [7]. The second one allows significant reduction in the computation of the energy term [8][9].

0-7803-1775-0/94 $3.00

2. CALCULATION OF βK TERMS

We have to calculate the scalars products β_k between a real sequence **q** of size N' and T ternary code book sequences of size N'.

If c^k is one code book sequence then :

$$\beta_k = \langle \mathbf{q}, c^k \rangle = \sum_{n=0}^{N'-1} q_n c_n^k \quad (6)$$

The **q** sequence can be segmented in N'/ L parts of length L then the first segment is :

$$sq_1 = [q_0 q_1 \cdots q_{L-1}] \quad (7)$$

and the j^{th} is :

$$sq_j = [q_{jL} \cdots q_{(j+1)L-1}] \quad (8)$$

and q is the concatenation of the sq_i 's.

$$\mathbf{q} = \left[sq_1 \; sq_2 \; \cdots sq_{\frac{N'}{L}} \right] \quad (9)$$

In the same way, each code word can be segmented in N'/L parts of size L. Let note sc_j^k the j^{th} segment of code word c^k, then we have :

$$c^k = \left[sc_1^k \; sc_2^k \cdots sc_{\frac{N'}{L}}^k \right] \quad (10)$$

The scalar product β_k (6) is thus given by :

$$\beta_k = \sum_{n=1}^{N'/L} sq_n \, sc_n^k \quad (11)$$

If we consider now one particular segment, say sc_1^k (see (10)), knowing that the code book does only contains -1, 0, +1 values, it exists $M=3^L$ different segment configurations for sc_1^k.

Let note l_m (m = 1 ... M) the M different segment configurations. For example, for L = 2, we have : l_1 = [-1,-1]

l_2 = [-1,0] , l_3 = [-1,1] , l_4 = [0,-1] , l_5 = [0,0] , l_6 = [0,1] ,

l_7 = [1, -1] , l_8 = [1,0] , l_9 = [1,1].

The idea is to pre-calculate the partial scalar products (see (3) with x = L) :

$$\langle sq_n, sc_n^k \rangle \quad \text{where} \quad sc_n^k = \{l_1, l_2, \ldots, l_M\}$$

for n = 1... N'/L

Then, for sq_1, we calculate M partial scalar products :

$$b_{1,m} = \langle sq_1, l_m \rangle \quad m = 1 \ldots M$$

For sq_2, we calculate M partial scalar products :

$$b_{2,m} = \langle sq_2, l_m \rangle \quad m = 1 \ldots M$$

and so on until $b_{N'/L,m}$.

The $b_{i,m}$ being calculated, the scalar product (6) for a code word c^k can be evaluated by adding the corresponding partial scalar products :

$$\beta_k = b_{1,m(1)}^k + b_{2,m(2)}^k + \ldots + b_{N'/L,m(N'/L)}^k \quad (12)$$

The overall complexity to calculate the scalar product β_k (6) for k = 1... T is then given by :

- Calculation of the partial scalar products. Considering the fact that just one operation is needed to evaluate the $b_{l,m}$ terms [7] (and developed independently from the authors by [10]) and that each $b_{l,m}$ term has a symmetric term but [0 0 ...0] the complexity is (N'/L) $(3^L - 1) / 2$.
- The scalar product β_k is the sum of the N'/L partial scalar products see (12). For k = 1 ... T, we have (N'/L - 1) T additions/ subtraction.

And the complexity for the calculation of the β_k^2 is :

$$C'_{\beta^2} = \frac{N'}{L}\left(\frac{3^L - 1}{2} + T\right) \quad (13)$$

3. CALCULATION OF THE ENERGY TERM BY SUBSAMPLING.

3.1 Principle:

First : compute a simplified performance criterion

$PC_q(j) = \beta_j^2 / \alpha_{jq}^2$, α_{jq}^2 requiring less computation than α_j^2.

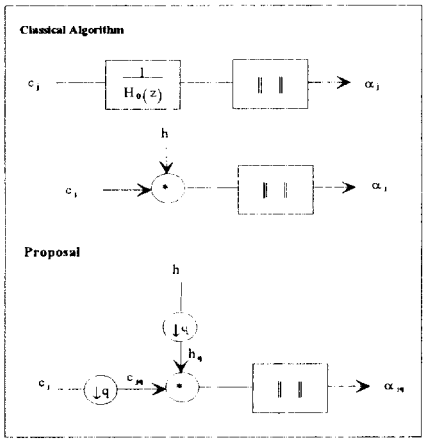
Second : Keep the subset of the original code book that contains the T_q best $PC_q(j)$ sequences.

Third Find in this subset the best sequence for the original criterion.

3.2 method

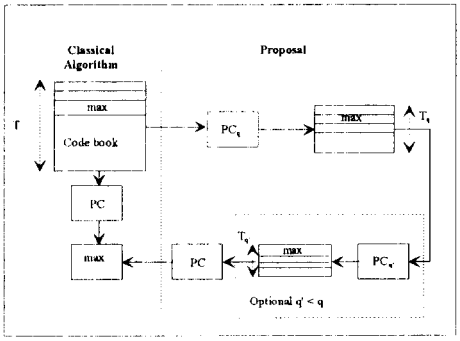
α_j^2 is the energy term of the synthetic speech vector $\hat{p}_j$. $\hat{p}_j$ is the convolution of the code word c_j and the impulse response h of the synthesis filter.

α_{jq}^2 is the energy term of the simplified synthetic speech vector $\hat{p}_{jq}$. We define this simplified vector as the convolution of the subsampled code word c_j of length N'/q that will be noted c_{jq} and the subsampled impulse response h_q. The subsampling is done by a simple averaging of q samples.



T_q size is fixed to make sure with 99% probability that the best code book sequence will be in the subset.

The method can be optionally iterated by taking a first subsampling factor q and then a second one q' (< q).



4. APPLICATION TO THE FEDERAL STANDARD 1016.

4.1 Results for the cross-correlation term

The FS-1016 [5] uses two code books to model the excitation. One is an adaptive code book of size T'= 256 to model long-term signal periodicity. The second is a fixed stochastic code book of size T = 512. The stochastic code book contains sparse, overlapped (shift by -2) and ternary valued samples (+1,0, -1) of zero-mean, unit variance white gaussien sequences, centre clipped at 1.2 resulting approximately 77% sparsity (zero values).

- Calculation of the cross correlation term

For the calculation of the β_k term, a signification computational saving can be obtained taking groups of size L=4 or 5 or 6. The calculation gain is respectively 4 and 4 and 5.

Since we have T scalar products of vectors of size N' to evaluate, the original complexity is (N'+1) T multiplications/ additions if we take account of the calculation of the square.

Thus $C_{\beta^2} = (N'+1)T = 61*512 = 31\,232 = 4$ MFlops.

With the method from [7], this can be reduced to $C'_{\beta^2} = 0.8$ MFlops with L = 6.

4.2 Results for the method by subsampling

4.2.1 Complexity

For the calculation of the energy term, we can use the subsampling method in its simplest way. The original code book is shifted by -2, we can choose a subsampling factor of q=2. Then, the subsampled code book contains code words that are just shifted by 1 if the subsampling is just done averaging two consecutive samples. More, we have experimented that taking a impulse response of just $N'_h = 20$ for subframe of size 60 didn't change anything in the sorting before taking out the T_q candidate code words of the subset. The subsampled code book obtained from the standardised code book of [6] contains 56 % of zero's samples. Our experiments have shown that $T_q = 8$ was sufficient to be sure to 99% of probability to have the best CELP sequence contained in the subset. Taking all these parameters into account, we can evaluate the complexity of the method.

- *PC_q Criterion*

Convolution $N'/2q(N'/q+1)$

Upgrade $N'_h/q\,(T-1)$

Energy N'/q T

- *Sorting* T_q T
- *PC Criterion*

Filtering T_q (N' P - P(P-1)/2)

Energy N' T_q

Total :

$$C'_{\alpha^2} = \frac{1}{2}\left(\frac{N'^2}{q^2} - \frac{N'}{q}\right) + \frac{(N'_h + N')}{q} T + T_q\left(T + N'(P+1) - \frac{P^2}{2} + \frac{P}{2}\right)$$

Taking N' = 60, q = 2, N'h = 20, T_q = 8, P = 10, we obtain: C'_{α^2} = 4 MFlops to compare with the classical algorithm :

$$C_{\alpha^2} = N'P - P\frac{(P-1)}{2} + \left(2N'_h + N'\right) T = 6{,}9 \text{ MFlops}$$

Taking account of the sparsity of the code book where no upgrade is made for a 0 sample, we have :

C'_{α^2} = 3,6 MFlops and C_{α^2} = 4.8 MFlops

4.2.2 Subjective tests

Informal listening tests have been realised in order to compare the quality between the FS-1016 and the FS-1016 with the subsampling method. We asked 30 listeners to judge the quality of 25 pairs of male and female French phonetically balanced sentences. They had choice between 3 answers : No difference or preference, coder 1 and coder 2. where coder 1 or 2 were randomly FS-1016 and FS-1016 with the subsampling method.

The following results :

No difference or preference 45 %

FS-1016 coder : 27 %

FS-1016 with the subsampling method : 28 %

show that the subsampling method does not degrade the perceptual quality of the reference coder where it is implemented.

4.3 Overall complexity

The overall complexity is thus

C' = 3,6 + 0.8 = 4.4 MFlops to be compared with

C = 4.8 + 4 = 8.8 MFlops.

The reduction of calculation is 2.

CONCLUSION

The computation task can be reduced combining two different methods.

- Using the ternary nature of the code book, the cross-correlation term can be calculated in two steps. Firstly, calculation of the partial scalar products, secondly, sum of the partial scalar products to obtain the value of βk.
- The cross-correlation term being calculated, the performance criterions are evaluated in two steps. Firstly, we eliminate non relevant code book sequences using a rough performance criterion where the energy is approximated. Secondly, the best CELP sequence is obtained in the subset of remaining code book sequences.

REFERENCES

[1] Schroeder M.R., Atal B.S. "Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates ". Proc. ICASSP, vol. 3,pp. 937-940 1985

[2] Kleijn W.B., Krasinski D.J., Ketchum R.H. "Fast Methods For The CELP Speech Coding Algorithm ". IEEE Trans. on Acous.,Sp.,and Sig. Proc,Vol. 38,No. 8,pp. 1330-1342,August 1990

[3] Gerson I.A., Jasiuk M.A. "Vector Sum Excited Linear Prediction (VSELP) Speech Coding At 8Kbps ". IEEE-ICASSP,pp. 461-464, 1990

[4] Trancoso I.M , Atal B.S. "Efficient Procedures For Finding The Optimum Innovation In Stochastic coders ". ICASSP,tokyo,pp.2375-2378, 1986

[5] Campbell Jr. J.P., Tremain T.E., Welch V.C. "The Federal Standard 1016 4800 Bps CELP Voice Coder ". Digital Processing I,pp. 145-155, 1991

[6] Fenichel R., Bodson D. "Details to Assist in Implementation of Federal Standard 1016 CELP ". Technical Information Bulletin 92-1,National Communication system 1992

[7] M. Mauc, G. Baudoin, M. Jelinek " Complexity Reduction For FS-1016 at 4800 bps CELP Coder", EUROSPEECH 93, Berlin September 93, pp. 245-248.

[8] M. Mauc, G. Baudoin, " Reduced Complexity CELP coder", Proceedings IEEE ICASSP, Mars 1992, San Francisco, Vol I., p. 53 - 56

[9] M. Mauc, G. Baudoin, M. Jelinek, P. Jardin, "Reduced complexity CELP Coder with Multistage Search", pp. 523-526, EUSIPCO 92, Bruxelles

[10] Di Francesco R., "Codage algébrique de la parole : prediction linéaire à Excitation par code ternaire". Ann. Télécom. 47.,n°5-6, 1992