D.1.2.

RPCELP: A HIGH QUALITY AND LOW COMPLEXITY SCHEME FOR NARROW BAND CODING OF SPEECH

M. LEVER and M. DELPRAT

MATRA-Communication
Rue J.P. Timbaud, B.P. 26
78392 Bois d'Arcy Cedex
FRANCE

Abstract

Code-Excited Linear Prediction (CELP) is a powerful technique for low bit rate speech coding but the basic scheme leads to a huge computational load. The paper introduces a related scheme that replaces the conventional stochastic excitation by an efficient Regular Pulse excitation. This new method, called Regular Pulse CELP (RPCELP), takes advantage of the codebook structure together with the use of a convenient perceptual criterion to achieve a very low complexity while maintaining high quality output speech. Objective performances are reported for several configurations.

Introduction

Low bit rate coding techniques are of major interest for narrow band speech transmissions such as in mobile radio applications. For instance, the available bit rate for a digital transmission over a 12.5 KHz channel is around 8 Kbps. Because of transmission errors, a part of this rate is allocated to an error correcting code, so the speech signal must actually be coded with at most 6 Kbps. At such a low bit rate, some existing coders achieve good results [1], but the reconstructed speech quality must be improved for general public applications.

Code-Excited Linear Prediction (CELP) is a very attractive approach for low bit rate coding of speech [2]. For each block of samples, an innovation sequence is picked up in a codebook of waveforms and processed through a synthesis filter which exhibits both a short term predictor (LP filter) and a long term (pitch) predictor. The innovation sequences are optimally selected through an analysis-by-synthesis procedure according to a given perceptual criterion. The reconstructed speech quality is quite good, but the basic scheme leads to a huge computational load.

The paper presents a new CELP scheme called Regular Pulse Code-Excited Linear Prediction (RPCELP) because each innovation sequence is constituted of equidistant pulses separated by zeros. A very fast search procedure in a "binary" codebook is exposed, that takes advantage (in a similar way as in RPE coders [3]) of the regular pulse structure with a suitable choice of the perceptual weighting filter.

The first section describes the basic CELP scheme and present a new strategy for complexity reduction, based on the choice of a convenient perceptual weighting filter. Section II compares several types of excitation and a regular pulse codebook is shown to be a good choice. The RPCELP algorithm is detailed in section III and performances are reported for various bit rates. Finally the paper considers the design of a RPCELP coder in terms of quantization issues, real time implementation and protection against transmission errors.

I. CELP coding and excitation estimation

<u>I.1. CELP coding</u>

In CELP coding, the speech spectrum is modeled by a time-varying linear prediction filter and the residual signal is vector quantized using a codebook of waveforms (figure 1). Assuming the filters A(z) and B(z) have been computed and quantized, the coding operation consists of determining the optimum innovation sequence (codeword c_k and gain G_k) for each block of the frame through an analysis-by-synthesis procedure. for each codeword c_k, the resulting synthetic signal is compared to the original one and the difference signal is processed through the perceptual weighting filter W(z), which is expressed as $W(z)=A(z)/A(z/\gamma)$, with γ around 0.8. The codeword that minimizes the weighted error signal energy is then selected for the current block.

The basic CELP coder described above can easily be transformed in a more convenient (but equivalent) structure [4,5,6]. First W(z) is moved from the output of the error subtraction operation to both of its input branches. Then the contribution of

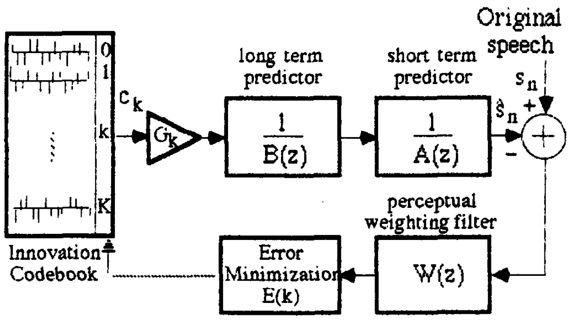

Figure 1: The basic CELP analysis scheme

88CH2607-0

the memory in both long term predictor $1/B(z)$ and weighted short term predictor $1/A(z/\gamma)$ is subtracted to the weighted original signal, yielding signal x_n, prior to the start of a codebook search. So during the search procedure, the codeword c_k is only processed by the memoryless weighted synthesis filter $1/\tilde{A}(z/\gamma)$, yielding signal $z_k(n)$. This memoryless filtering can be expressed with a convolution of two finite sequences and represented as a matrix-vector product:

$$z_k = G_k H c_k \qquad (1)$$

where H is a L x L lower triangular matrix whose elements are from the impulse response h(i) of $1/A(z/\gamma)$:

$$H = \begin{bmatrix} h(0) & 0 & 0 & \cdots & 0 \\ h(1) & h(0) & 0 & \cdots & 0 \\ h(2) & h(1) & h(0) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h(L-1) & h(L-2) & h(L-3) & \cdots & h(0) \end{bmatrix} \qquad (2)$$

In the same way, the vector $x=(x_n)$ is written as

$$x = Hr + x_0 - z_0 \qquad (3)$$

where r is the residual signal with the effects of the long term predictor subtracted, x_0 and z_0 represent the contribution of the filter memory in the computation of x and z_k respectively. Now the weighted error for the k^{th} codeword is expressed as

$$E(k) = \| x - z_k \|^2 = \| x - G_k H c_k \|^2 \qquad (4)$$

The search procedure to determine the optimum innovation sequence (c_k, G_k) is then derived from equation (4):

1) Find the index k_0 which maximizes the weighted inner product $P_\psi(k)$:

$$P_\psi(k) = (x^t H c_k) / \| H c_k \| \qquad (5)$$

2) Compute the related gain G_{k_0}:

$$G_{k_0} = P_\psi(k_0) / \| H c_{k_0} \| \qquad (6)$$

This inner product formulation is faster than the previous euclidean distance formulation, but the codebook search complexity remains very high.

I.2. Perceptual weighting filter

The major drawback of the basic CELP scheme described above is the huge amount of computations involved in the filtering of all the codewords by the time varying synthesis filter $1/A(z/\gamma)$. Recent research has focused on this complexity issue and several strategies have been proposed [4,5,6].

The complexity considerably decreases when the weighted synthesis filter is fixed. This is the case if the perceptual weighting filter has the form

$$W'(z) = A(z) / C(z/\gamma) \qquad (7)$$

where $1/C(z)$ is an average low order linear short term speech predictor. The weighted synthesis filter is then modified in $1/C(z/\gamma)$ whose coefficients are time invariant. Such a filter has already been used in multipulse coding of speech [3] and has proved its remarkable ability to provide almost equivalent subjective results.

Table 1 shows the comparative objective results we obtained in introducing a fixed weighted synthesis filter in a CELP coding scheme involving different codebooks. In spite of a relatively lower SNR, the perceived lost in quality due to a fixed filter is quite low and acceptable with regard to the enabled simplifications in the search procedure. Indeed, this procedure does not involve any repetitive filtering operation anymore since all excitation sequences may be pre-filtered and stored.

II. Design of the excitation codebook

II.1. Binary versus Stochastic excitation

In the initial CELP scheme [2], the innovation codebook is populated with i.i.d. Gaussian samples (stochastic coding). Recent works have shown that better performances can be achieved with statistical codebooks [6]. However stochastic or statistical codebooks are essentially not structured and the codebook structure is one of the keys to complexity reduction.

As pointed out in [7], sequences of +1 and -1 are just as good as stochastic sequences concerning vector quantization performances at low bit rate (1/2 bit per sample or below). Algebraic structures from Code Theory provide efficient binary codebooks. The related fast algorithms are used to speed up the search procedure and the codebook does not have to be stored anymore. It has been shown in [4] that such codebooks are quite effective even for small dimensions. Table 1 shows that a binary codebook derived from the Reed-Muller code and complemented with single-pulse sequences compares favourably with a stochastic codebook at rate 3/8 bit per sample in dimension 16.

II.2. Regular Excitation structure.

The good results achieved with codebooks including pulse sequences (cf. II.1 and [6]) suggest the use of excitation sequences of length L having a regular structure consisting in q equidistant pulses separated by D-1 zeros, the first pulse (initial phase p) being at one of the locations 0 to D-1. This approach attempts to better represent the phase information in the excitation signal.

The Regular Pulse (RP-) codebook, populated in a statistical or stochastic manner, may be constituted of K independent sequences or of the D possible shifts of a basic set of K/D sequences made of RP-squences with initial phase zero. In the latter case, each codeword c_k is expressed as

$$c_k = \Delta_p d_m \qquad (8)$$

where Δ_p is a L x q decimation matrix, function of the initial phase p and d_m is a q-dimensional vector with $k = p.(K/D) + m$.

The RP-structure can obviously be exploited to reduce both computational load and storing requirements. Furthermore, considering the efficiency of binary vectors at low bit rates (see II.1), a "binary" RP-codebook is of great interest. It is built from the 2^q binary words of length q (0 becoming -1). Table 1 reports the excellent results provided by such a RP-excitation codebook, which appears to be the best of all tested codebooks.

Table 1: Performances (average SNRseg in dB) of various codebooks with variable/fixed weighted synthesis filter for two speakers.

Codebook \ Filter \ Speaker	F		M	
	variable	fixed	variable	fixed
stochastic	8.57	6.86	10.62	8.17
algebraic	8.68	6.94	10.71	7.78
regular-pulse	8.90	7.15	11.03	8.63

III. Regular Pulse Code-Excited Linear Prediction (RPCELP)

III.1. Base Band CELP

Regular pulse sequences as defined in the last section resemble upsampled versions (in a D ratio) of vectors of length q=L/D. The same observation has already been made in RPE speech coding [3], derived from Multi-Pulse Linear Predictive Coding. RPE coding combines the use of RP excitation and of a suitable weighting filter in an efficient algorithm that may be considered as a generalization of Base Band speech coding with spectral folding as high frequency regeneration technique. A low complexity / high quality RPE coder at 13 Kbps has recently been chosen as a standard for the Pan European mobile radio system [8]. The optimal excitation sequence is here determined by down sampling the LP residual with a choice of the best decimation grid, involving a fixed smoothing filter. The RP excitation may be vector quantized to achieve lower bit rates [9]. Though, vector quantizing once the excitation has been determined does not lead to an optimal excitation sequence.

Introducing a RP codebook in a CELP coding algorithm is a different approach yielding an optimal choice of the excitation sequence. This section shows that similar transformations as in RPE can be performed in RPCELP analysis, providing a very fast search procedure. RPCELP can be considered as a RPE technique in which the pulse amplitudes are optimally vector quantized and in that sense may be viewed as a Base Band CELP coding technique.

III.2. RPCELP algorithm and performances

The error E(k) in the codebook search (see II.1) is traditionally minimized over the block (0,1... L-1). An other approach consists of minimizing E(k) with longer sequences x_n and $z_k(n)$, say L+J samples, such that the impulse response of the weighted synthesis filter W(z)/A(z) is practically zero after J samples. Signals r_n and $c_k(n)$ are set to zero for $n \geq L$ and the impulse response matrix H becomes a (L+J) x L Toeplitz matrix:

$$H = \begin{bmatrix} h(0) & 0 & 0 & \cdots & 0 \\ h(1) & h(0) & 0 & \cdots & 0 \\ h(2) & h(1) & h(0) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h(J-1) & h(J-2) & h(J-3) & \cdots & h(0) \\ 0 & h(J-1) & h(J-2) & \cdots & h(1) \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & & \cdots & h(J-1) \end{bmatrix} \quad (9)$$

In this "autocorrelation" method, the "memory error" $x_0 - z_0$ in equation (3) is minimal and will be approximated to zero. Thus, substituting equation (3) in equation (5) gives

$$P_W(k) = (r^t H^t H c_k) / \| H c_k \| \quad (10)$$

The matrix $R = H^t H$ is a L x L symetrical Toeplitz matrix built on the autocorrelations R(i) of the impulse response h(n). Then the vector $y^t = r^t H^t H$ is efficiently computed (once per block) as the result of a filtering operation (smoother, [3]). So the codewords are now filtered only once per frame to compute the weighting coefficients $\| H c_k \|$, which results in a low complexity algorithm.

In the case of a Regular Pulse excitation, the codebook structure can be exploited to speed up even more the search procedure. As a matter of fact, it comes from equation (8)

$$\| H c_k \|^2 = c_k{}^t H^t H c_k = d_m{}^t R_D d_m \quad (11)$$

where $R_D = \Delta_p{}^t H^t H \Delta_p$ (12)

is a q x q symetrical Toeplitz matrix whose i^{th} diagonal term is R((i-1)D). Note that R_D is independent of the phase p as a result of the autocorrelation method defined above.

Moreover, R_D can be forced to a diagonal matrix, using a reasonable approximation on the weighted synthesis filter: its impulse response is shortened in order that $h(n) = 0$ for $n \geq D$. In the case of a fixed weighted synthesis filter (see II.3), it can merely be designed such that $R(iD) = 0$ for $i > 0$. With the normalization by R(0), the matrix R_D becomes the identity matrix and equation (11) gives

$$\| H c_k \|^2 = \| d_m \|^2 \quad (13)$$

Assuming the codewords are normalized, the search procedure comes down to maximize the inner product $P(k) = y^t c_k$, which represents a small amount of computations (all the more because the codewords are sparse). The RPCELP algorithm is illustrated in figure 2.

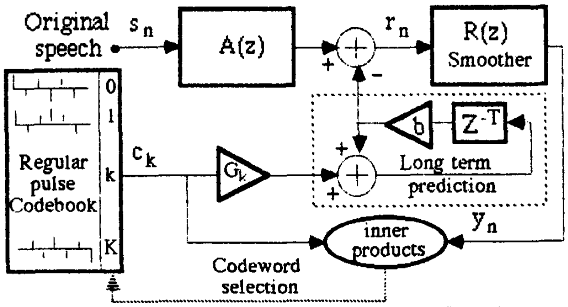

Figure 2: Block diagram of the RPCELP coder

Besides, with a binary RP codebook the optimum codeword is efficiently determined in a two-steps procedure:

1) Find the phase p_0 which maximizes M(p) with

$M(p) = \Sigma_i |y(p+iD)|$, sum over i=0,...,q-1 (14)

2) Choose the vector d_{m_0} such that $y^t c_{k_0} = M(p_0)$:

$d_{m_0}(i)$ = sign of $y(p_0+iD)$, for i =0,...,q-1.

The related gain is then given by $G_{k_0} = M(p_0)/q$, since $\|Hc_k\|^2 = q$ for any k.

The above procedure involves a very small amount of computations: around 20 instructions per sample, which is 2000 times less than in the basic scheme [5]. Moreover, this very low complexity is independent of the block length L.

Figure 3 shows the performances obtained with a binary RP codebook, a decimation factor D=4 and various block lengths: L=8,16,20,40. It can be seen that for L=16 (3/8 bit per sample) these results are very close to those presented in table 1 with a RP codebook and a fixed weighted synthesis filter. In fact, the approximations introduced in the design of the fast RPCELP algorithm do not produce any noticeable degradation.

III.3. Design of a RPCELP coder

The great advantage of the RPCELP technique is its very low complexity. The whole coding/decoding scheme, including protection against transmission errors, can easily be real time implemented on any commercially available DSP chip.

With a binary RP codebook, the bit rate R of the excitation is expressed as

$$R = (L/D + \log_2 D) / L \quad \text{bit per sample} \qquad (15)$$

The values D=4 and L=20 provide a suitable bit rate / quality trade off. In this case R=7/20, or equivalently 2800 bps with an 8 KHz sampling frequency, so the global rate may be around 6 Kbps or even lower if the LP filters are vector quantized.

Finally, it should be noted that the binary regular pulse excitation is theoretically robust against transmission errors. As a matter of fact, the codebook can be efficiently indexed so that each codeword c_k is directly deduced from the binary decomposition of its index k [10]. Then one error on a given bit of k (except for the phase information) produces only a single wrong pulse in the innovation sequence.

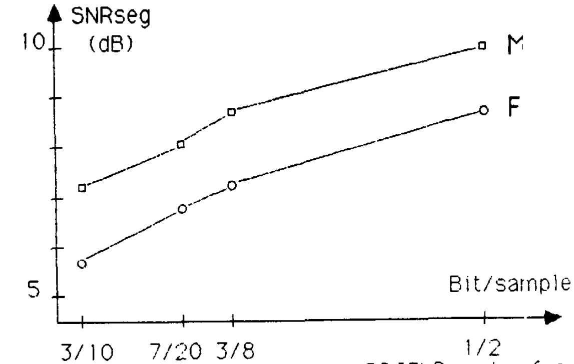

Figure 3: Performances of the RPCELP coder for 2 speakers, as a function of the excitation bit rate.

Conclusion

In this paper, we have formulated different ways of reducing the computational load of CELP coders without significantly degrading the reconstructed speech quality.

Furthermore, a very low complexity Base Band CELP algorithm, called RPCELP, has been presented. A suitable perceptual weighting filter together with a binary regular pulse codebook enables a very fast search procedure, which involves only 20 multiply-adds per sample instead of 40000 for the basic CELP scheme. This algorithm can easily be real time implemented and may provide a good communication quality at 6 Kbps, even over noisy transmission channels.

References

[1] I. Lecomte, M. Lever, L. Lelièvre, M. Delprat, A. Tassy, "Medium Band Speech Coding for Mobile Radio Communications", in Proc. ICASSP, Apr. 1988.

[2] M.R. Schroeder and B.S. Atal, "Code-excited linear prediction (CELP): high-quality speech at very low bit rates", in Proc. ICASSP, Mar. 1985.

[3] P. Kroon, E. F. Deprettere, R. J. Sluyter, "Regular Pulse Excitation: a Novel Approach to Effective and Efficient Multi-Pulse Coding of Speech", IEEE Trans. on ASSP, Vol. ASSP-34, Oct. 1986.

[4] J.P. Adoul, P. Mabilleau, M. Delprat and S. Morissette, "Fast CELP coding based on algebraic codes", in Proc. ICASSP, Apr. 1987.

[5] I.M. Trancoso and B.S. Atal, "Efficient procedures for finding the optimum innovation in stochastic coders", in Proc. ICASSP, Apr. 1986.

[6] G. Davidson, M. Yong and A. Gersho, "Real-time vector excitation coding of speech at 4800 bps", in Proc. ICASSP, Apr. 1987.

[7] J.P. Adoul, C. Lamblin, "A comparison of some algebraic structures for CELP coding of speech", in Proc. ICASSP, Apr. 1987.

[8] P. Vary, R.J. Sluyter, C. Galand, M. Rosso, "RPE-LPC Codec: the candidate for the GSM radio communication system", Int. Conf. on Digital Land Mobile Radio Communication, Venice, Jul. 1987

[9] P. Kroon and E.F. Deprettere, "Quantization Procedures for Regular-Pulse Excitation Coders", in Proc. 3rd European Signal Processing Conference (EUSIPCO-86), Sep. 1986.

[10] A. Le Guyader, P. Combescure, C. Lamblin, M. Mouly and J.F. Zurcher, "A robust 16 Kbits/s vector adaptive predictive coder for mobile communications", in Proc. ICASSP, Apr. 1986.