ISCA Archive
http://www.isca-speech.org/archive

4th European Conference on
Speech Communication and Technology
EUROSPEECH '95
Madrid, Spain, September 18-21, 1995

IMPROVED CELP ALGORITHM SUITED FOR VARIOUS SPEECH CODING APPLICATIONS

Sofía Moreno Pérez*, Ramón García Gómez**.

*Universidad Alfonso X El Sabio. Villanueva de la Cañada. 28691-Madrid
e-mail: sofiam@uax.es
**Universidad Politécnica de Madrid

10.21437/Eurospeech.1995-159

[1]ABSTRACT

This is another "*CELP based fantastic speech coding scheme*", and in this case we have solved a problem that will make your speech coding real time implementation really easier. We combine the efficiency of the open loop analysis with the robustness of the closed loop one. In this way, we obtain a significant computational save in the search of the optimum excitation vector with no loss of speech quality. We developed a real time implementation of this coder for low bit rate transmission (from 4.8 to 14.4 KBPS) and for wideband speech coding (from 16.0 to 28.8 KBPS).

1. INTRODUCTION

Speech coding trends nowadays proceeds in two basic avenues: one, low bit rate coding for transmission over wireless channels that requires the use of low complexity algorithms while maintaining the appropriate toll quality; and two, wideband speech coding that provides increased speaker presence and naturalness compared to traditional telephony due to the wider bandwith, 7 KHz.

During the past years a strong research effort has been done in the field of low bit rate speech coding [1,2,3,4,5,6,7,8,9]. In this article we present a new procedure for computing the excitation signal of a CELP type speech coder. Considering that searching for the optimum excitation signal in CELP coding is one of the largest computational load, our scheme offers a substantial computational reduction, suited for real time implementation in portable equipments where computational complexity, power consumption and volume are subject to severe constraints. Besides reducing the computational complexity associated with CELP algorithms with no loss of speech quality, our coding scheme, that we call "Signs Modulated-Regular Pulse Excitation-CELP" (SM-RPE-CELP), is capable of working at a continuous range of bit rates, with the same algorithm supporting a soft degradation of speech quality from 16 KBPS to as low as 4 KBPS. The proposed coder is already implemented in real time using a single Digital Signal Processor, a TMS320C31 working at 33 MHz.

The ITU-T recommendation G722 [14] defines a bit rate for coding wideband speech in the range of 48 to 64 KPBS, suited to basic ISDN channels, but reducing this bit rate offers the possibilities of new applications like G722 quality over analog subscriber loops using a V.34 MODEM that can work up to 28.8 KBPS, or video telephony over S0 channels with enhanced voice quality preserving the highest bit rate for coding the video signal. As in the low bit rate coding field, here CELP schemes provide efficient solutions at the expense of a heavy computational load. We handle this problem with the same algorithm, SM-RPE-CELP, obtaining a significant computational saving with almost no loss of speech quality at bit rates as low as 16 KBPS.

2. THE BASIC CODER

The SM-RPE-CELP coder, as shown in figure 1, is used for low bit rate coding, while for the wideband case we introduce some modifications that will be explained below. The main components of the coder are:

- A tenth order Short Term Predictor is computed over a frame of 20-40 ms, depending on the available bit rate; this frame is divided into 4 subframes [4] to get the other parameters. An interpolated short term predictor is computed for every subframe in order to obtain a smooth spectral tracking. The LPC parameters are transformed into Line Spectral Pairs for quantization (scalar and differential), interpolation and transmission. Thirty-four bits per frame are employed.

- On the basis of this short term predictor, a perceptual weighting filter is obtained for every subframe, with a spectral expansion of 15 Hz.

[1]This work has been partly supported by grantt TIC92-0269

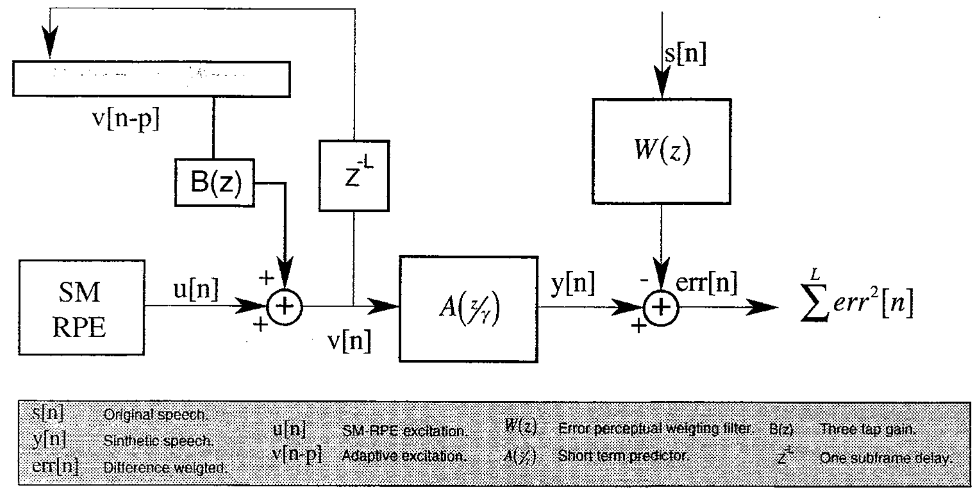

Figure 1: block diagram of the coder

- The adaptive excitation signal is 147 samples long and optimum delay is searched in the interval of 20 to 147 samples delay in odd subframes and in a subset of 32 lags centred in the last obtained position for even subframes [4]. Later, a three tap long term predictor is computed to reduce the roughness of the synthetic speech. These coefficients are vector quantized with 8 bits.

- The excitation signal u[n], the innovation of our proposal, is composed by the signs, modulated in amplitude, of a decimated subsequence of the low pass filtered linear prediction residual. The computation of this signal is described in detail in the following paragraph.

3. THE EXCITATION: THE SM-RPE-CELP

The excitation signal u[n] is decomposed as the product of two other signals, this is how we get a substantial computational reduction. In this way, we take profit of the robustness of closed loop analysis and the computational efficiency of open loop analysis.

First, we compute a sign sequence (sd[n], taken value +1 or -1) according to figure 2. The rationale for this approach is the following:

If we built two excitation code vector libraries, one with constant spectral module and another with constant spectral phase [2], with the same number of components, the synthetic speech quality we obtain with a CELP coder using the second one is much better than with the first one. In fact, the majority of CELP schemes take advantage of this property, for example with isolated or ternary codebooks. Therefore, the stochastic excitation tries to preserve the spectral phase of the original speech.

The linear prediction residual (signal e[n] in figure 2) retains the spectral phase information that we must preserve. We extract part of this information in an open loop analysis; because low frequency phase distortion is more audible than high frequency, the signal e[n] is low pass filtered with a linear phase filter.

As in the RPE-LTP speech coder of the European standard GSM [10], after being filtered a decimated sequence is extracted from ef[n], called d[n]. Given a decimation factor D, the subsequence with largest energy is adopted (parameter M in figure 2). This is done in the "Grid Selection" box of the figure.

Taking the sign information is sufficient for efficiently encoding (preserving the phase) d[n]: the spectral phases of signals ef[n], d[n] and sd[n], for a decimation factor of three, remain very close in the low frequency region of the spectrum.

Finally, a closed loop analysis by synthesis procedure (figure 2) is performed to get a sequence of amplitudes that modulates the signs sequence d[n], from a positive random codebook. A small codebook is necessary, with a number of vectors in the range of 16 to 64. This codebook is overlapped, being each vector a 1 sample shifted version of the previous one. The overlapping does not introduce degradation because the signals sd[n]a_i[n] are uncorrelated (consider that $a_i[n] = a[n+i]$).

The structure of the codebook allows the introduction of an efficient algorithm for computing the energies and correlation of filtered signals sd[n]a_i[n].

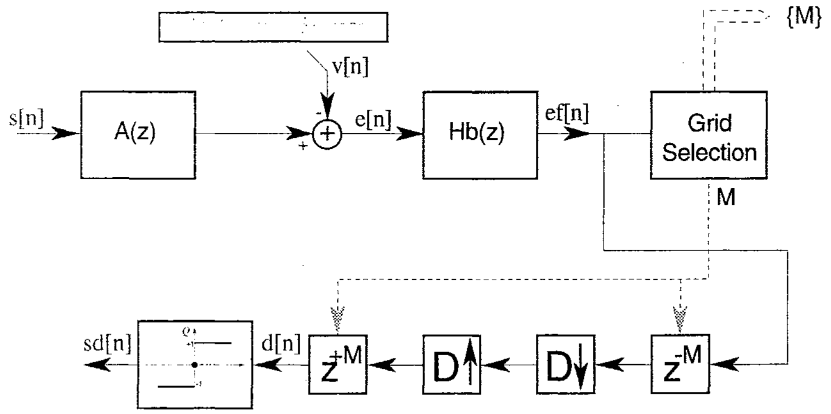

Figure 2: Signs sequence extraction

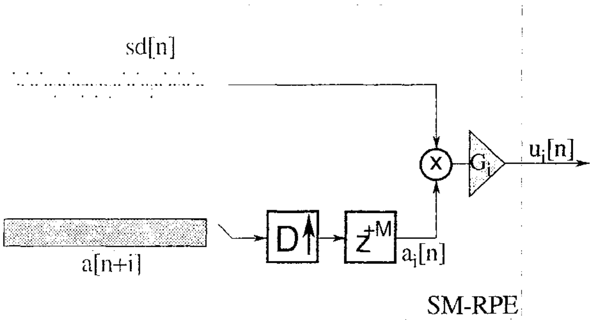

Figure 3: SM-RPE excitation.

4. WIDEBAND CODING

Based on the basic coding scheme as shown in figure 1, we derive a wideband coder introducing two quadrature mirror filters. The low frequency band is coded the same way as the low bit rate coder described above, while for the high frequency band the adaptive excitation is suppressed and the order of the short term predictor is reduced to 6 or 8.

Good results are obtained with a 9.6 KBPS coder for the low band and a 6.2 one for the high, for a 16.0 KBPS full coder. Almost clear quality is obtained with a 14.4 KBPS coder for the low band and a 11.2 one for the high, with a 26.0 KBPS whole rate.

5. RESULTS

For the low bit rate case we have obtained a coder, working in the range of 14.4 to 4.8 KBPS, implemented in real-time using a single TMS320C31 DSP at 33 MHz. For 9.6 KBPS rate the speech quality is similar to full-rate GSM coding, the Pan European standard for digital mobile telephony.

To assess the computational efficiency of our coding scheme, we compare it against the VSELP concerning computing the excitation signal, supposing that in both cases the coder structure is the same. For a subframe length of 5 ms and the VSELP with 2 codebooks of 7 vectors, the computational load of VSELP is of the order of 15000 FLOPS, while for the SM-RPE_CELP is 3000 FLOPS (the bit assignment can be seen in the table shown below, for a bit rate of 9.6 KBPS).

Bit Rate *(frame length)*	**Short Term Predictor**	**Long Term Predictor**	**SM-RPE**
14400 *20 ms* *(288 bits)*	34 bits	7+5+7+5 Gain 8* 4	D = 1, N = 2^4 bs = 40,bp = 0,bg = 5 total = 49*4
12000 *20 ms* *(240 bits)*	34 bits	7+5+7+5 Gain 8* 4	D = 2, N = 2^5 bs = 20,bp = 1,bg = 5 total = 31*4
9600 *(20 ms)* *192 bits*	34 bits	7+5+7+5 Gain 8* 4	D = 3, N = 2^5 bs = 13,bp = 2,bg = 5 total = 25*4
7200 *(30 ms)* *216 bits*	34 bits	7+5+7+5 Gain 8* 4	D = 3, N = 2^4 bs = 20,bp = 2,bg = 5 total = 31*4
4800 *(40 ms)* *192 bits*	34 bits	7+5+7+5 Gain 8* 4	D = 6, N = 2^4 bs = 13,bp = 3,bg = 5 total = 25*4

D Decimation factor.

N Number of shifted vector in the amplitude's codebook.

bs Number of bits employed in the sign's sequence.

bp Number of vector employed in the M coefficient of figure 2.

bg Number of vector employed in the gain.

In the case of wideband coding, our algorithm is able to work at bit rates in the range of 14.4 to 28.8 KBPS, drastically reducing the computational burden associated with high quality CELP coding at high sampling frequencies. The perceptual quality is comparable to that of G.722 standard at 48 KBPS.

The bit assignment for the high band is:

Bit Rate (*frame length*)	Short Term Predictor	SM-RPE
7600 *20 ms* *(152 bits)*	order = 8 28 bits	$D = 2, N = 2^4$ bs = 20,bp = 1,bg = 5 total = 30*4
6400 *20 ms* *(128 bits)*	order = 8 28 bits	$D = 3, N = 2^4$ bs = 13,bp = 2,bg = 5 total = 24*4

REFERENCES

[1] B. Atal y M.R. Schreder. "STOCHASTIC CODING OF SPEECH SIGNALS AT VERY LOW BIT RATES". *IEEE Int. Conf. Commun.-ICC'84, pag 1610 a 1613.*

[2] I. Trancoso and B. Atal, " EFFICIENT SEARCH PROCEDURES FOR SELECTING THE OPTIMUM INNOVATION IN STOCHASTIC CODERS," IEEE Trans. on ASSP, Vol. 38, n° 3, pp. 385-396, March 1990.

[3] P. Kroon and E. Deprettere, "A CLASS OF ANALYSIS-BY-SYNTHESIS PREDICTIVE CODERS FOR HIGH QUALITY SPEECH CODING AT RATES BETWEEN 4.8 AND 16 KBPS," IEEE Journal on Selected Areas in Communication, Vol. 6, Febr 1988.

[4] J. P. Campbell, Jr. y T. E. Tremain y V. C. Welch. "THE PROPOSED FEDERAL STANDARD 1016 4800 BPS VOICE CODER: CELP". *Speech Technology April-May 1990.*

[5] W. B. Kleijn, D. S. Krasin ki, R. H. Ketchum, "FAST METHODS FOR THE CELP SPEECH CODING ALGORITHM," IEEE Trans. on ASSP, Vol. 38, n° 8, pp. 385-396, August 1990.

[6] I. Trancoso, "AN OVERVIEW OF DIFFERENT TRENDS ON CELP CODING," NATO.ASI, New Advances and Trends in Speech Recognition and Coding, pp. 227-236. Bubión, Granada, Spain. June 1993.

[7] P. Kroon and E. Deprettere, "A CLASS OF ANALYSIS-BY-SYNTHESIS PREDICTIVE CODERS FOR HIGH QUALITY SPEECH CODING AT RATES BETWEEN 4.8 AND 16 KBPS," IEEE Journal on Selected Areas in Communication, Vol. 6, Febr 1988.

[8] Bishnu S. Atal, Vladimir Cuperman y Allen Gersho, editors. "ADVANCES IN SPEECH CODING". *Kluwer Academic Publishers, 1991.*

[9] I. Gerson, M. Jasiuk. "VECTOR SUM EXCITED LINEAR PREDICTION (VSELP) SPEECH CODING AT 8 KBPS". Proceedings of ICASSP'90.

[10] "GSM FULL RATE SPEECH TRANSCODING". GSM recommendation 06.10. European Telecommunications Standards Institute 1991.

[11] Bishnu S. Atal, Vladimir Cuperman y Allen Gersho, editors. "SPEECH AND AUDIO CODING FOR WIRELESS AND NETWORK APPLICATIONS". *Kluwer Academic Publishers, 1993.*

[12] R. Lefebre et alt., "HIGH QUALITY CODING OF WIDEBAND AUDIO SIGNALS USING TRANSFORM CODED EXCITATION". Proceedings of ICASSP'94, Adelaide, 1994.

[13] E. Harborg et alt., "A REAL-TIME WIDEBAND CELP CODER FOR A VIDEOPHONE". Proceedings of ICASSP'94, Adelaide, 1994.

[14] Christian G. Gerlach "CELP SPEECH CODING WITH ALMOST NO CODEBOOK SEARCH". *Procc. ICASSP 1994, PP II.109-112.*

[15] CCITT Recomm. G722.