# Hybrid Coding Method of Speech Signals Using Conjugate Structure-Algebraic Code Excited Linear Prediction (CS-ACELP) Codec

Lamis Hamood Mohaissn Al-Saadi

Department of Mathematics, Education Collage

## Abstract:

A compression algorithm for high quality speech signal using prediction coding techniques is developed. In this paper we present an improvement scheme for hybrid coding of speech signals based on linear prediction (LP) analysis and excitation signal which is Conjugate Structure–Algebraic Code Excited Linear Prediction (CS-ACELP) codec this codec is one of the techniques to compress speech signal to a bit rate (6.4 Kbps). The five important stages associated with the encoding principle of CS-ACELP include the pre-processing stage, the LP analysis stage, the synthesis filter, the algebraic codebook search, and the adaptive codebook search. The speech signal is analyzed for speech frames of 10 ms corresponding to 80 samples at a sampling rate of 8000 samples per second. Subjective evaluation experiment was conducted to test the performance of the hybrid (CS-ACELP) codec which is Mean Opinion Score (MOS) evaluation. The experimental results showed the quality of the reconstructed speech signals using hybrid coding technique was almost the same as that of the original speech signals.

# طريقة ترميز هجينة للإشارات الكلامية باستخدام مرمز – فاك ترميز الهيكل المترابط لترميز إثارة التنبؤ الخطي الجبري

لميس حمود محيسن السعدي

جامعة بابل - كلية التربية

## الخلاصة:

لقد طورت خوارزمية الضغط للإشارة الكلامية ذات النوعية العالية باستخدام تقنيات الترميز التنبؤي. وفي هذا البحث قدمنا طريقة محسنة للترميز الهجين للإشارات الكلامية بالاعتماد على تحليل التنبؤ الخطي وإشارة الإثارة وهي مرمز – فاك ترميز (الهيكل المترابط لترميز إثارة التنبؤ الخطي الجبري) ويعتبر هذا المرمز – فاك الترميز احد التقنيات لضغط الإشارة الكلامية عند معدل نقل بيانات (Kbps 6.4). يتضمن المرمز للهيكل المترابط لترميز إثارة التنبؤ الخطي الجبري خمسة مراحل مهمة وهي مرحلة قبل المعالجة ومرحلة تحليل التنبؤ الخطي ومرشح التركيب وبحث كتاب الرموز الجبري وبجث كتاب الرموز المتكيف. تحلل الإشارة الكلامية إلى إطارات وكل إطار ذا فترة زمنية (ms 10) يحتوي على (80) عينة عند تردد نمذجة (8000) عينة لكل ثانية. ولاختبار كفاءة المرمز – فاك الترميز للهيكل المترابط لترميز إثارة التنبؤ الخطي الجبري الهجين فقد استخدم اختبار تقويم حسي وهو مقياس (علامة معدل الرأي)، وبينت نتائج الاختبار بان نوعية الإشارة الكلامية المسترجعة باستخدام تقنية الترميز الهجين تكون غالباً مشابهة للإشارة الكلامية الأصلية.

## 1. Introduction:

Speech coding is very important area of research in digital signal processing. It is a fundamental element of digital communications and has progressed at a fast pace in parallel to the increase of demands in telecommunication services and capabilities [1].

The subject of speech coding has been an area of research for several decades. Speech coders are present in our everyday life and their use is often taken for granted. For example, speech coding is present in most digital telephone systems and in every cellular application [2][3].

Speech coders, whose goal is to represent the analog speech signal in as few binary digits as possible, can be described as belonging to one of three fundamentally different coding classes: waveform coders, vocoders, and hybrid coders [2][4]. A waveform coder attempts to mimic the waveform as closely as possible by transmitting actual time or frequency domain magnitudes. Speech quality produced by waveform coders is generally high, although at high bit rates [5]. Vocoders, or parametric coders, analyze the waveform to extract parameters that in some cases represent a speech production model. The waveform is synthetically reproduced at the receiver based on these quantized parameters. Vocoders can generally achieve higher compression ratios than waveform coders; however, they provide more artificial speech quality [6]. In hybrid coders, the high compression efficiency of vocoders and high quality speech reproduction capability of waveforms are combined to produce good quality speech at medium to low bit rates. The so called analysis by synthesis coders, such as the Coded Excited Linear Prediction (CELP) and Mixed Excitation Linear Prediction (MELP) are all hybrid coders [7].

The structure of a (6.4 Kbps) speech codec, based on the hybrid model, is outlined in this paper. The new codec shows promise of achieving high quality at a bit rate of (6.4 kbps).

## 2. Coder Description:

Most of the speech coders are based on linear prediction (LP) analysis. CS-ACELP coder is a typical and popular example of this class of coders [8][9]. In this section, the 6.4 Kbps CS-ACELP speech coding algorithm is described, Figure (1) illustrates the principle of the encoding algorithm. It follows the Linear Prediction Analysis by Synthesis (LPAS) principle [10]. The coder operates with a frame size of 10 ms which consists of two 5 ms subframes. The main building blocks are LP analysis and quantization for the short-term spectral envelope, synthesis filter, an adaptive codebook for long term (pitch) prediction and an algebraic codebook for innovation coding. This coder performs LP analysis of speech for extracting LP coefficients and employs an analysis by synthesis procedure to search algebraic and adaptive codebooks to compute the excitation signal. The method used for performing LP analysis plays an important role in the design of a CS-ACELP coder. The autocorrelation method is conventionally used for LP analysis.

The Linear Predictive Coding (LPC) spectrum calculated from these proposed methods is shown to be more robust. These methods work as well as the conventional methods when the speech signal is clean or has high signal to noise ratio. Also, these robust methods give less quantisation distortion than the conventional methods. The application of these robust methods for speech compression using the CS-ACELP coder provides better speech quality when compared to the conventional LP analysis methods.

In CS-ACELP we build two codebooks which are algebraic and adaptive codebooks to obtain the excitation signal which is given by the sum of a scaled adaptive codebook signal and a scaled signal from algebraic codebook. This excitation is used to drive a synthesis filter which models the effect of the vocal tract. At the decoder the LP coefficients and excitation signal are passed through synthesis filter to produce the reconstructed speech signal. Typically the filter parameters are determined first and then codebook indices as well as the adaptive codebook gain and algebraic codebook gain are found. The codebook parameters are chosen to minimise the weighted error between the reconstructed and the original speech signals. In effect each possible codebook entry is passed through the synthesis filter to test which gives an output close to the input speech in the perceptually weighted sense. This largely close loop structure is used in order to produce a reconstructed speech signal which is as close as possible to the original speech signal. A brief description of all the CS-ACELP codec stages is provided in the subsequent sections.

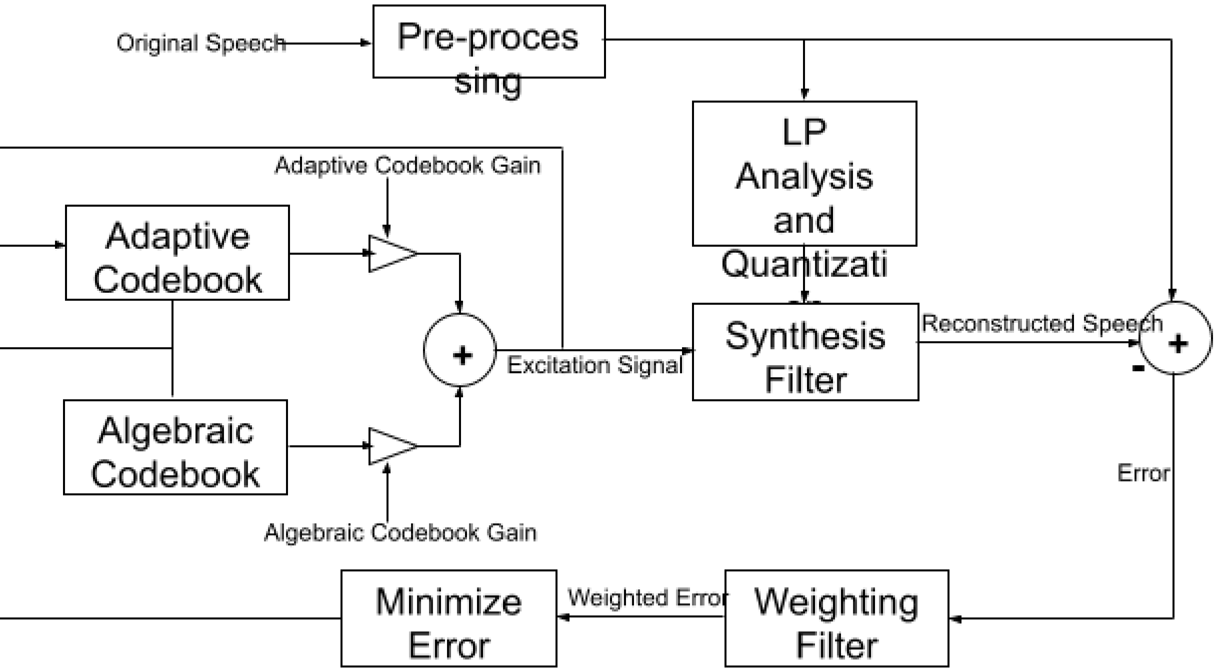

**Figure (1) Principle of encoding**

## 2.1 Pre-processing:

The input signal is high pass filtered and scaled in the preprocessing stage. A second order pole-zero filter with a cutoff frequency of 140 Hz is used to perform the high-pass filtering. To reduce the probability of overflows in the fixed-point implementation, the scaling operation is performed.

## 2.2 LP analysis and quantization:

The pre-processed signal, *s(n),* is windowed using a 30 ms (240 samples) asymmetric window. The LP analysis window consists of half a hamming window and quarter of a cosine function cycle. The window operates on 120 samples from the past speech frame, 80 samples from the present speech frame and 40 samples from the future speech frame (a total of 240 samples). Using the Levinson-Durbin algorithm, the LP coefficients are computed from the autocorrelation coefficients corresponding to the windowed speech. The LP coefficients obtained are converted to line spectral pairs (LSP). These are later quantized and interpolated. A two stage vector quantizer is used to quantize the LSP coefficients based on the minimization of the weighted mean squared error.

## 2.3 Synthesis Filter:

The synthesis filter is usually simply assumed to be the inverse of the prediction error filter $A(z)=1 - a_1z^{-1} - a_2z^{-2}$ ... $a_Mz^{-M}$ which minimizes the energy of the prediction residual for the input speech signal $S(n)$. Here $M$ is the order of the filter, which we took to be equal to ten. Once the excitation signal $u(n)$ has been determined it is possible to recalculated the synthesis filter coefficients in order to maximize quality of the reconstructed speech [11][12]. The filter coefficients were performed using the Levinson-Durbin algorithm. The autocorrelation function is computed from the speech signal frame. The resulting coefficients were converted to Line Spectrum Frequencies (LSFs) [13] for quantization. Given an excitation signal $u(n)$ and a set of filter coefficients $a_i$, $i$ =1, 2…$M$, the reconstructed speech signal $\hat{S}(n)$ will be given by:

$$\hat{S}(n) = u(n) + \sum_{i=1}^{M} a_i \hat{S}(n-i) \qquad (1)$$

We wish to minimize $E$, the energy of the error signal $e(n) = S(n) - \hat{S}(n)$ over the frame length $L$. $E$ is given by:

$$E = \sum_{n=0}^{L-1} (S(n) - \hat{S}(n))^2$$

$$= \sum_{n=0}^{L-1} \left( S(n) - u(n) - \sum_{i=1}^{M} a_i \hat{S}(n-i) \right)^2 \qquad (2)$$

## 2.4 Perceptual Weighting Filter:

The perceptual weighting filter is computed from the following equation:

$$W(z) = \frac{A(z/y_1)}{A(z/y_2)} \qquad (3)$$

The weighted speech signal, $s_w(n)$ is computed as follows:

$$S_W(n) = S(n) + \sum_{i=1}^{10} a_i y_1^{\,i} S(n-i) - \sum_{i=1}^{10} a_i y_2^{\,i} S_W(n-i) \qquad n = 0, \ldots, 39 \quad (4)$$

where, $s(n)$ is the pre-processed speech, $\gamma_1$ and $\gamma_2$ are the adaptive weights, and $a_i$, i = 1, 2, . . . , 10 are the unquantized LP coefficients. The weighted error $e_w(n)$ is then given by:

$$e_W(n) = S_W(n) - \hat{S}w(n) \qquad (5)$$

where, $\hat{S}w(n)$ is the weighted reconstructed speech signal.

## 2.5 Algebraic (Fixed) Codebook Search:

The name Algebraic in the codec implies the structure of the codebook used to select the excitation codebook vector. The codebook vector consists of a set of interleaved permutation codes containing few nonzero elements [14][15]. The fixed codebook structure is given by Table (1).

**Table (1): Fixed codebook structure**

| Track ($k$) | Signs | Pulse Positions ($p_k$) |
|---|---|---|
| $i_0 = 0$ | $S_0$: ±1 | $P_0$:0,5,10,15,20,25,30,35 |
| $i_1 = 1$ | $S_1$: ±1 | $P_1$:1,6,11,16,21,26,31,36 |
| $i_2 = 2$ | $S_2$: ±1 | $P_2$:2,7,12,17,22,27,32,37 |
| $i_3 = 3$ | $S_3$: ±1 | $P_3$:3,8,13,18,23,28,33,38 |
| | | 4,9,14,19,24,29,34,39 |

where, $p_k$ is the pulse position and k is the pulse number. The codebook vector, $c(n)$, is determined by placing the 4 unit pulses at the found locations ($p_k$) multiplied with their signs (±1).

## 2.6 Adaptive Codebook Search:

The impulse response, *h(n)* of the weighted synthesis filter, $W(z)/\hat{A}(z)$ is computed for each subframe. The target signal *x(n)* is obtained by filtering the residual signal *r(n)*, through the combination of synthesis filter ($1/\hat{A}(z)$) and perceptual weighting filter ($W(z)$).

Closed loop pitch analysis is performed on a subframe (40 samples) basis.

The adaptive codebook search (closed loop search) is shown in Figure (2).

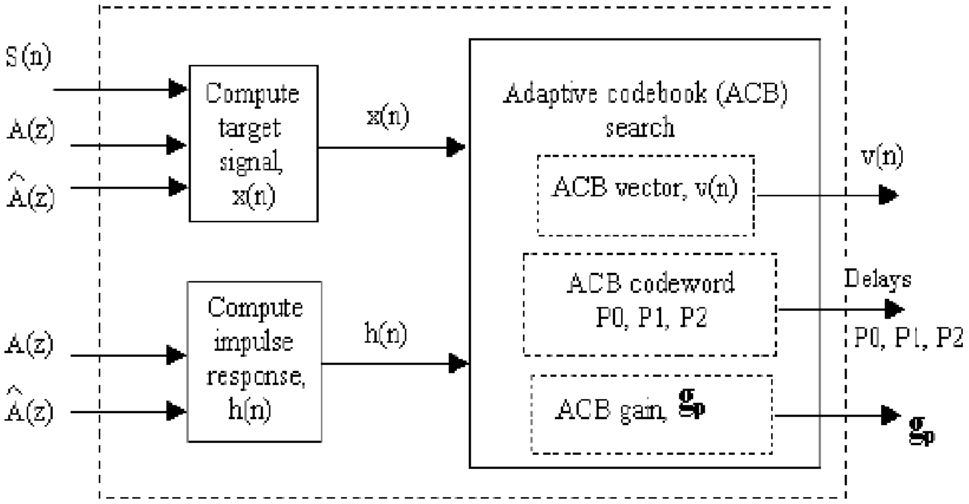

**Figure (2) Adaptive codebook (Closed loop pitch) search block**

The adaptive codebook vector, *v(n)*, is computed by interpolating the past excitation signal. The adaptive codebook gain, $g_p$, is computed as follows:

$$g_p = \frac{\sum_{n=0}^{39} x(n)y(n)}{\sum_{n=0}^{39} y(n)y(n)} \qquad (6)$$

$$y(n) = \sum_{i=o}^{n} v(i)h(n-i) \qquad n=0, \ldots, 39 \qquad (7)$$

where, *x(n)* is the target signal, *y(n)* is the filtered adaptive codebook vector, *v(n)* is the adaptive codebook vector and *h(n)* is the impulse response of the weighted synthesis filter.

## 3. Results:

The proposed CS-ACELP codec has been computer simulated and its performance was evaluated using subjective performance evaluation which is Mean Opinion Score (MOS) evaluation. We measure the (MOS) of the original speech file compared to the reconstructed speech file in subjective performance evaluation, result for this measure are shown in table (2) for the following sentences:

1. A male speaker saying "Nice Day".
2. A female speaker saying "Nice Day".
3. A male speaker saying "Thank".
4. A female speaker saying "Thank".

**Table (2) The MOS Measure**

| Sentences | MOS |
|---|---|
| 1 | 4.17 |
| 2 | 3.98 |
| 3 | 4.82 |
| 4 | 4.41 |

## 4. Conclusions:

In this paper we introduce an improved method that approaches optimal coding efficiency by implementation of the CS-ACELP codec. The developed low bit rate speech coding based on CS-ACELP codec was proposed, analytically studied and simulated. Due to strong interactions of the closed loop encoding structure in CS-ACELP codec, a careful design is necessary for a given applications and its particular requirements in terms of bit rate, delay, robustness and speech quality under different operating conditions. The results achieved from the CS-ACELP codec are high quality speech at low bit rates of (6.4 Kbps) with very small difference in the speech before coding and after decoding and this means the reconstruction speech signal as close as possible to the original speech signal.

The 6.4 kbps CS-ACELP employs new fixed and adaptive codebooks. Under most conditions, the coder exceeds the requirements, providing high quality for bandwidth limited systems and achieved high quality comparable to the low rate codecs.

**5. References:**

[1] K. N. Prasetiyo, "Robust Linear Prediction Analysis for Low Bit-Rate Speech Coding", M.Sc. Thesis , Griffith University, 2002.

[2] A. P. Bernard, " Source and Channel Coding for Speech Transmission and Remote Speech Recognition ", PhD Thesis, Department of Electrical Engineering, California University, 2002.

[3] A. Gersho, "Advances in Speech and Audio Compression", IEEE Transactions on Speech and Audio Processing, vol. 82, no. 6, pp. 900–918, June 1994.

[4] N. Jayant, "Signal Compression: Technology Targets and Research Directions", IEEE Journal on Selected Areas in Communications, vol. 10, no. 5, pp. 796–818, June 1992.

[5] M. R. Schroeder, "A Brief History of Speech Coding", Proceedings International Conference on Communications, pp. 26.01.1–4, Sept. 1992.

[6] P. Alku , and T. Backstrom , " Linear Predictive Method For Improved Spectral Modeling of Lower Frequencies of Speech with Small Prediction Orders " , IEEE Transactions on Speech and Audio Processing , 2004.

[7] EE5401 Cellular Mobile Communications, Institute for Infocomm Research, National University of Singapore, 2007.

[8] ITU-T Recommendation G.729, "Coding of Speech at 8 kbit/s Using Conjugate Structure-Algebraic Code Excited Linear Prediction (CS-ACELP).

[9] A. S. Spanias, "Speech Coding: A Tutorial Review", Department of Electrical Engineering, Arizona State University , http://www.fulton.asu.edu/~spanias/E607S05/Speech Coding.pdf, pp.1-95, Jan. 2002.

[10] W. B. Kleijn and K. K. Paliwal, "Speech Coding and Synthesis", Amsterdam, Holland: Elsevier, 1995.

[11] F. F. Tzeng, "Near – Optimum Linear Predictive Speech Coding" IEEE Global Telecommunication Conference, pp. 508.1.1-508.1.5, 1990.

[12] M. Niranjan, "CELP Coding with Adaptive Output-Error Model Identification", Proc. ICASSP, pp.225-228, 1990.

[13] R. Steele, "Mobile Radio Communications", Pentech Press, 1992.

[14] A. Kataoka et al., "An 8-kb/s Conjugate Structure CELP (CS-CELP) Speech Coder", IEEE Trans. Speech and Audio Processing, vol. 4, no. 6, pp. 401-411, Nov 1996.

[15] R. Salami et al., “Design and Description of CS-ACELP: A Toll Quality 8 kbps Speech Coder”, IEEE Trans. Speech and Audio Processing, vol. 6, no. 2, pp.116-130, Mar 1998.