# Hybrid Coding of Speech Signals

**Amera.Abdul.Wohid Funjan AL.Tayee**

**Department of Mathematics, Education College**

## 1. Abstract:

In this paper we present a novel approach for hybrid coding of speech signals which is code excited linear predictive (CELP) codec, this hybrid codec based on analysis-by-synthesis (AbS) loop. Method for updating the short-term synthesis filter once the excitation parameters have been determined is examined. We show that significant improvements can be achieved by updating the synthesis filter, similar to those obtained using the well known methods of interpolation and bandwidth expansion. However our proposed method of update avoids the increase in the delay of a codec that is usually associated with interpolation. Furthermore the method of determining the adaptive and fixed codebook parameters is examined. The closed loop technique is proposed for improving the performance of the codebook search while maintaining a reasonable level of complexity.We implemented the hybrid codec at bit rate (4.8 Kbps) and the performance of the proposed system was evaluated through signal to noise ratio (SNR) performance measure.

**Keywords:** Hybrid Codec, Speech Signals, Analysis by Synthesis.

**الخلاصة:**

تم في هذا البحث تقديم طريقة جديدة لترميز الأشارات الكلامية بطريقة الترميز الهجين وباستخدام مرمز ـفاك ترميز (ترميز اثارة التنبؤ الخطي)، ويعتمد هذا المرمز ـ فاك الترميز الهجين على دوارة التحليل بالتركيب. لقد درست الطرق لتعديل مرشح التركيب عندما تحدد معاملات الإثارة، وبالاعتماد على هذه الطرق فقد استطعنا انجاز تعديل جيد لمرشح التركيب مشابه للتعديل الذي نحصل عليه باستخدام الطرق المعروفة للتوليد والتوسيع للنطاق الترددي. على الرغم من ان الطريقة المقترحة للتعديل قد تجنبت الزيادة بالتأخير للمرمز ــ فاك الترميز المرافقة مع التوليد. بالأضافة الى ذلك فقد تم تطبيق طريقة لتحديد معاملات كل من كتاب الرموز الثابت وكتاب الموز المتكيف. لقد تم في هذا البحث اقتراح تقنية الحلقة المغلقة لتحسين كفاءة بحث كتاب الرموز مع مستوى معقول من التعقيد. لقد تم تنفيذ المرمز ــ فاك الترميز الهجين عند معدل نقل بيانات (4.8 Kbps) وقد قيمت كفاءة النظام المقترح من خلال مقياس كفاءة حسابي وهو (نسبة الأشارة الى الضوضاء).

## 2. Introduction:

The past decade has witnessed substantial progress towards the application of low-rate speech coders to civilian and military communications as well as computer-related voice applications. Central to this progress has been the development of new speech coders capable of producing high-quality speech at low data rates [1]. Most of these coders incorporate mechanisms to represent the spectral properties of speech, provide for speech waveform matching, and optimize the coder's performance for the human ear. A number of these coders have already been adopted in national and international cellular telephony standards [2] [3]. Speech coding techniques can be broadly divided into two classes waveform coding that aims at reproducing the speech waveform as faithfully as possible and vocoders that preserve only the spectral properties of speech in the encoded signal [4]. The waveform coders are able to produce high-quality speech at high enough bit rates; vocoders produce intelligible speech at much lower bit rates, but the level of speech quality in terms of its naturalness and uniformity for different speakers is also much lower. The applications of vocoders so far have been limited to low bit rate digital communication channels. The combination of the principles of waveform coding and vocoding has led to significant new codec in recent compression technology which is hybrid coding [1] [5].In our simulations we have used a 4.8 kbits/s CELP codec, with the synthesis filter parameters determined every 20ms frame, and the excitation parameters determined every 10ms sub-frame.

## 3. The 4.8 Kbps Codec:

In This paper, we present a novel hybrid model for speech codec which is code excited linear predictive (CELP) codec, the hybrid scheme is based on linear prediction (LP) and excitation signal. The 4.8 Kbps hybrid codec uses a frame size of 20 ms and each frame is divided in to 10 ms sub-frames.The block diagram of the encoder of such a codec is shown in Figure (1). The excitation signal $u(n)$ is given by the sum of a scaled adaptive codebook signal (which adds long term periodicities during voiced speech) and a scaled signal from a large fixed codebook $c_k(n)$. This excitation is used to drive a synthesis filter which models the effects of the vocal tract. At the decoder the excitation signal is passed through the synthesis filter to produce the reconstructed

Typically the filter parameters are determined first and then the $\hat{S}(n)$ speech signal

and $k$ as well as the gains $G1$ and $G2$ are found. The codebook $\alpha$ codebook indices parameters are chosen to minimize 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 closest to the input speech in the perceptually weighted sense. This largely closed-loop structure is used in order to produce a reconstructed signal which is as close as possible to the original speech.
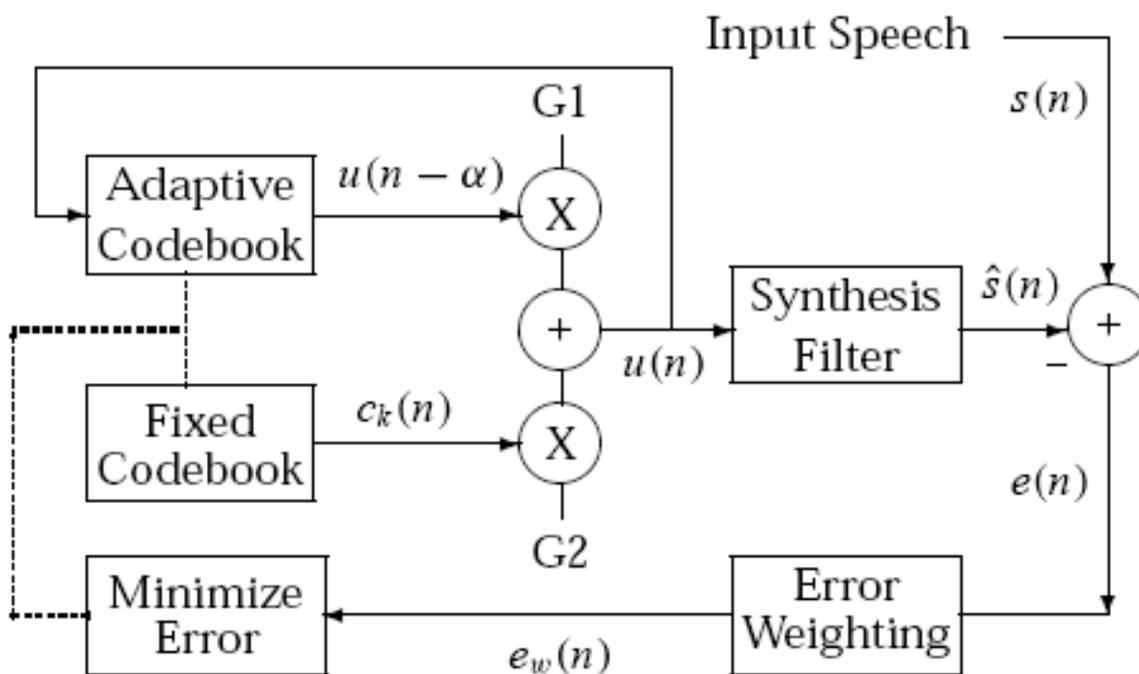


## 4. The linear prediction of speech signal :

Linear predictive method is considered one of the efficient method for speech signal analysis [1].Linear predictive method become common technique that used for representing speech signal at low bit rate and provide efficient computing of speech signal parameters [3].The basic idea of linear predicting coding (LPC) is that a sample of speech can be approximated as a linear combination of the past (M) speech samples. By minimizing the square difference between the original speech samples and the linearly reconstructed speech samples we can determine the predictor coefficients [6] [7].The linear predictive analysis is based on the reconstruction of speech samples,

and a driving signal $\hat{S}(n)$ which can be formed by a reconstruction valve

$$e(n).$$

$$(1) \ \hat{S}(n) \ S(n) = \ + e(n)$$

consists of M previous samples [5] [8] $\hat{S}(n)$ Where reconstruction valve

$$(2) \ \sum_{i=1}^{M} a_i * S(n-i) = \hat{S}(n)$$

Where $a_i$ representing predictor coefficients which can be obtained by the autocorrelation method [7] [9].The driving signal $e(n)$ can now be interpreted as the prediction error between the original data samples $S(n)$ and the reconstructed

$$\hat{S}(n) \ \text{samples}$$

$$(3) \ \hat{S}(n) \ e(n) = S(n) -$$

Where predictor coefficients $a_i$ are designed so that total squared prediction error $E$ is minimized over the data frame Length $L$. The total squared error $E$ is defined by [10] [11]:

$$(4) \ E = \sum_{n=0}^{L} (e(n))^2$$

## 5. Calculation of the Excitation Parameters:

In order to reduce the complexity of the encoder the error weighting filter in Figure (1) is usually moved so that the input and the reconstructed speech signals $s(n)$ and

are separately weighted before their difference is found. For an all-pole synthesis $\hat{S}(n)$ filter of the form $H(z)=1/A(z)$, where $A(z)=1 - a_1 z^{-1} - a_2 z^{-2} \ldots a_M z^{-M}$, and an error weighting filter $A(z)/A(z/y)$ where y is a constant set, the cascade of the synthesis filter and the error weighting filter is equivalent to using a weighted synthesis filter of the form $1/A(z/y)$. The weighted error $e_w(n)$ is then given by [12]: $e_w(n) = S_w(n) -$

$$(5) \ \hat{S}_w(n)$$

is the weighted reconstructed $\hat{S}_w(n)$ where $S_w(n)$ is the weighted input speech, and speech.The codebook search procedure attempts to find the values of the adaptive

as well as the fixed codebook index $k$ and gain $G_2$, which $\alpha$ codebook gain $G_1$ and delay minimize the mean square error $E_w$ taken over the sub-frame length $N$. This error can be writtenas:Where

$$(6)\; E_w = \frac{1}{N}\left( \sum_{n=0}^{N-1} X^2(n) - T_{\alpha k} \right)$$

$$T_{\alpha k} = 2*(G_1 * C_\alpha + G_2 * C_k - G_1 * G_2 * Y_{\alpha k}) - G_1^2 * \zeta_\alpha - G_2^2 * \zeta_k \qquad (7)$$

is the target $X(n) = S_w(n) - \overset{\wedge}{S_0}(n)$ is the term to be maximized by the codebook search,

represented zero-input response of the $\overset{\wedge}{S_0}(n)$ signal for the codebook search, and

weighted synthesis filter due to it's input in previous sub-frames,Here

$$(8)\; \zeta_\alpha = \sum_{n=0}^{N-1} \left[ (u(n)-\alpha)*h(n) \right]^2$$

is the energy of the filtered adaptive codebook signal and

$$(9)\; C_\alpha = \sum_{n=0}^{N-1} X(n) * \left[ (u(n)-\alpha)*h(n) \right]$$

is the correlation between the filtered adaptive codebook signal and the codebook target is the energy of the filtered fixed codebook signal $[c_k(n) * h(n)]$, $\zeta_k X(n)$. Similarly, and $C_k$ is the correlation between this and the target signal. Finally,

$$(10)\; Y_{\alpha k} = \sum_{n=0}^{N-1} \left[ (u(n)-\alpha)*h(n) \right] * \left[ c_k(n)*h(n) \right]$$

$\alpha$ The optimum values of the gains ($G_1$ and $G_2$) for a given pair of codebook indices and $k$ are expressed in equations (11) and (12)

$$(11)\; G_1 = \frac{C_\alpha * \zeta_k - C_k * Y_{\alpha k}}{\zeta_\alpha * \zeta_k - Y_{\alpha k}^2}$$

$$(12)\; G_2 = \frac{C_k * \zeta_\alpha - C_\alpha * Y_{\alpha k}}{\zeta_\alpha * \zeta_k - Y_{\alpha k}^2}$$

The full search procedure computes the terms ($C_k$ ,$C_\alpha$ , $\zeta_k$ , $\zeta_\alpha$ and $Y_{\alpha k}$) for every , $k$ and uses these to calculate the gains $G_1$ and $G_2$ [13] [14]. $\alpha$ pair of codebook indices

## 6. Results:

The performance of our (4.8 kbps) CELP codec, expressed in terms of the signal to noise ratio (SNR) performance measure, as shown in Table (1) for the following Arabic statements:

1.  Man say " صباح الخير ".
2.  Woman say " صباح الخير ".
3.  Man say " واحد ".
4.  Woman say " واحد ".

**Table (1) Signal to Noise Ratio (SNR) Performance Measure**

| Statements | SNR(dB) |
|---|---|
| 1 | 10.96 |
| 2 | 10.86 |
| 3 | 8.80 |
| 4 | 7.62 |

## 7. Conclusions:

In this work we have studied, implemented and improving hybrid coding of speech signals which is code excited linear predictive (CELP) codec. Several papers have appeared in the past considering updating the synthesis filter parameters once the excitation signal is known. We found the maximum SNR gain possible through such an update, and proposed a method which obtains almost this full improvement but has a lower complexity than other methods which have been proposed.Secondly we studied ways of improving the joint adaptive and fixed codebook closed loop searches. Again, initially we found the maximum improvement possible by carrying out a full joint codebook search. We then suggested method which almost improvement as possible the performance of the codebook search.Finally the results achieved from hybrid codec are high quality speech signal at low bit rate (4.8 Kbps) with small difference between original and synthesis speech signal.

## 8. References:

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

[2] B. S. Atal , and N. S. Jayant , " Transmission and Storage " , AT & T Bell Laboratories , Murray Hill , New . http:// cslu. cse.ogi.edu/HLTsurvey/ch10node4.htmlJersey , USA ,

[3] J. P. Woodard, "Digital Speech Coding", Mini-Thesis, Department of Electronics and Computer Science, University of Southampton, June 1994.

[4] W. B. Kleijn , and J. Haagen , " Transformation and Decomposition of The Speech Signal For Coding " , IEEE Signal Processing Lett. ,Vol. 1 , No. 9 , PP. 136-138 , 1994.

[5] A. Kiviluoto , " Speech Coding Standards " , http ://mia .ece. uic.edu /~papers/WWW/MultimediaStandards/chapter3.pdf.

[6] O. Ozun , P. Steurer , and D. Thell , " Wideband Speech Coding With Linear Predictive Coding (LPC) ", Department of Electrical Engineering , University of California at Los Angeles , pp. 1-18 ,Winter 2002.

[7] W. Pereira , " Modifying LPC Parameter Dynamics to Improve Speech Coder Efficiency " , M.Sc. Thesis , Department of Electrical & Computer Engineering , Faculty of Graduate Studies and Research , McGill University, Sept. 2001.

[8] 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 .

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

[10] R. D. Degroat , and E. M. Dowling, "The Data Least Squares Problem and Channel Equalization", IEEE Transactions on Signal Processing, pp. 407–411, 1993.

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

[12] B. S. Atal , and J. R. Remde, "A New Model of LPC Excitation for Producing Natural-Sounding Speech at Low Bit Rates", Proc. ICASSP, pp. 614–617, 1982.

[13] J. P. Woodard , and L. Hanzo ," Improvements to The Analysis –by– Synthesis Loop in CELP Codecs " , Southampton University ,http://www.ecs.soton.ac.uk/publications/rj/1995-1996 /comms/jpw91r .pdf , Jun. 1996 .

[14] H. Kumar , and K. Sundaresan , " Implementation of  The  Code Excited Linear Predictive (CELP) Codec For VOIP ", Department of Electrical Engineering , State University Of New York at Buffalo , 1996 .