A Wide Band Speech Coding Technique using Low Delay Code Excited Linear Predictive Algorithm (LD-CELP)

Swati Joshi[1]
Rajasthan Technical University, Kota (Rajasthan)
Rajasthan, India
joshi92swati@gmail.com[1]

Hemant Purohit[2]
Jodhpur Institute of Engineering and Technology
Deptt. of Electronics and Communication JIET, Jodhpur,
Rajasthan, India
hemant.purohit@jietjodhpur.ac.in[2]

Abstract – **A fair level of speech quality is desired in speech transmission for mobile voice services. The effective utilization of bandwidth and higher bit rate is must for a best quality speech coder. But at a time the both requirements are not fulfilled in desired format. The research is ongoing in the area of designing speech coder's. In general the CELP is an algorithm to design a good quality speech coder. From 80's to present the advancement in this technique is going on. In this paper a wide band speech coding technique is proposed using LD-CELP algorithm. The overall performance of LD-CELP (16Kbps) is summarized and computed on MATLAB version R2016a with parameters MSE and SNR. In conclusion we observe that SNR for LD-CELP is not much better and enhancement in this is necessary.**

Key Words — Speech coding, CELP, LD-CELP, Perceptual weighted filter.

I.   INTRODUCTION

In present era, digital representation of voice signals is necessary for transmission of those signals over wireless channels. The limited bandwidth and more power requirements are major issues for enhancement in this field of technology. After converting the voice data into time domain representation the process of speech coding is applied on the data. Basically there are two types of speech coding technique-

1. Waveform Coding Techniques (Time Domain) - (PCM- Pulse code, DPCM-Differential Pulse Code ADPCM- Adaptive Differential Pulse Code) Modulation.
2. Parametric Method- LPC (Linear Predictive coding), RELP, MELP, CELP (Residual excited, Mixed excited, Code excited)

The quality of speech is provided by the encoded bit rate of the speech signal. According to bit rates the classification of speech coders is as follows –
   a) High bit rate (HBR) coders: bit rate >16 kbps.
   b) Medium bit rate (MBR) coders : bit rate 5-16kbps
   c) Low bit rate (LBR) coders : bit rate 2-5kbps
   d) Very Low bit rate (VLBR) coders : <2kbps

The concept for LD-CELP is based on the parametric coding. In parametric coding the Pulse code modulated data is excited to fetch the filter coefficients at encoder side. The filter coefficients are extracted with the help of linear predictive filter which works on forward and backward error prediction method. After that the Levinson-Durbin method is used to reduce the errors and complexity of the filter. The block diagram representation of linear predictive filter is shown in Fig.1 given below.

The algorithm is given by the formula given below:-

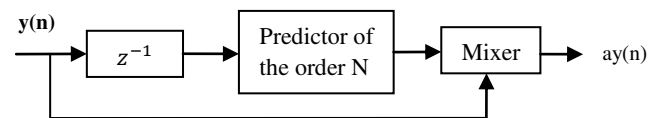$$y(n) = \sum_{i=1}^{N} a y(n-i) \qquad (1)$$

Figure 1: Block Diagram of Linear predictive filter

II.   CODE EXCITED LINER PREDICTION (CELP)

CELP is originally proposed by M. R. Schroeder and B. S. Atal in 1985. The basic principle of CELP is based on linear prediction, the digitized voice signals are highly correlative waveform. Each sample is represented as the combination of previous samples. The coefficients a_1, a_2 ….. a_N are linear prediction coefficients which are basically generated by Levinson Durbin algorithm and abbreviated as LTP (Long Term Predictor) coefficients. The CELP coder is a parametric coder hence it is work on the principle of 'Analysis by Synthesis' [2]. A Codebook is used to track the variation in input coded speech and analyzed by the fixed generated code whereas in waveform coding techniques the synthesis is done on the real data sequence and thus it required large bandwidth and time as per compared to CELP coders or parametric coders. A fixed codebook provide initial code vectors for data bit comparison and hence the high quality of speech is attained

at much lower bit rate then waveform coders thus the bandwidth is optimized as compared to waveform coders.

A. Low Delay - CELP

The total coded delay for CELP is 20ms which is not acceptable for high speed voice decoding thus, in an advancement for this a new version of CELP was designed in may 1992 and officially adopted as the CCITT G.728 standard for 16kbps speech data transfer, which is called LD-CELP (Low delay Code excited linear prediction) basically the delay for this coder is less than 5ms and the transmission rate is 16kbps [3].

The next target for researchers is LD-CELP for delay less than 2ms and speech data transfer rate will be 8Kbps, which is again the high quality bandwidth optimization.
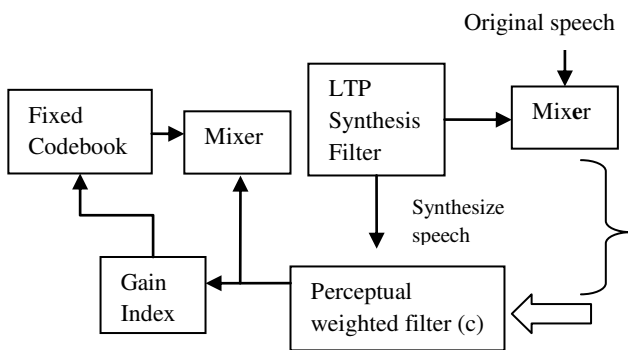

Figure 2: CELP Coder Block Diagram

B. Perceptual weighted filter

The perceptual weighted filter technique is usually used to provide improved speech quality and for minimizing the MSE (mean square error) in CELP coder [4]. This weighted filter used LP (Linear Predictive) analysis for generate autocorrelation function between noise and coded speech.

This process is also called as noise shaping of the signal. The perceptual optimization is a close loop system in which all possible combinations of speech signal is checked for every decoded bit sample with the help of the generated Fixed Gaussian codebook, and only one of them is stored for output which is having the best quality decoded speech among all of them.

The value for the perceptual weighted filter is set as constant value (c) the appropriate time delay is generated by the system in the output stream for best noise level detection and to get the efficient results for overall synthesis of the signal. The typical values for c are in between 0.5 to 0.9. The perceptual weighted filter is a linear filter and the time domain representation of this filter is given by the formula –

$$W(n) = \sum_{i=1}^{N} \frac{a(ni)}{a(ni/\gamma)} \qquad (2)$$

$$W(z) = 1 - \sum_{i=1}^{N} \left(\frac{aiz^{-1}}{ai\gamma z^{-1}}\right) \qquad (3)$$

Here the $a(ni)$ represent the nth order LP coefficients and the γ is a perceptually weighting factor, the factor γ is taken as constant value depending on the channel conditions. This factor is usually used for enhancing the bandwidth for some particular instants.

III. RESULT AND DISSCUSSION

Here, a detail performance analysis of LD-CELP 16kbps with perceptual weighted value c=0.85, and c=0.65 (constant) is presented. These simulation based comparative analyses illustrate the output speech quality in terms of SNR (signal to noise ration) and MSE (mean square error) of proposed speech coding technique.

IV. EVALUATION AND ANALYSIS

Analysis of 16kbps LD-CELP is done with the MATLAB simulating software version R2016a. The coder is designed to take audio speech samples at 8 kHz and output is observed in 16kbps. The extension for audio input file is (.wav) and total duration of this sample sound is 8sec. The audioread command is used in MATLAB to read audio voice sample as the (.wav) command is not working properly with MATLAB R2016a version. The x axis is used for time intervals and y axis is used for amplitude of coded audio data. The audio wave is generated 73113 sample points which are further compressed to 100 samples to find out the LP coefficients. The 'hello' file is taken as input audio and 'xhat1'is decoded sound file in 16kbps sampled format for CELP. Finally the experiment is performed for different values of c.

Firstly the LP analysis is done by the Levinson Durbin algorithm and the LP coefficients were calculated. The coefficients are real value samples. The graph for original speech is shown in Figure 3(a) and graph of LP coefficients is given in Figure 3(b). The comparison graph between original speech and 16kbps LD-CELP is given in Figure 3(c) and finally the input and output waveform shown together in graph with perceptual weighted constant value (c) is 0.85 on Figure 3(d) and 0.65 on Figure 3(e).

V. COMPARISON OF VARIOUS PARAMETERS

A. SNR

SNR is basically signal to noise ratio. The two signals 'hello' (original) and 'xhat1' (16kbps CELP) are the signals for comparison. Firstly mean square values are calculated then directly the command applied for SNR value
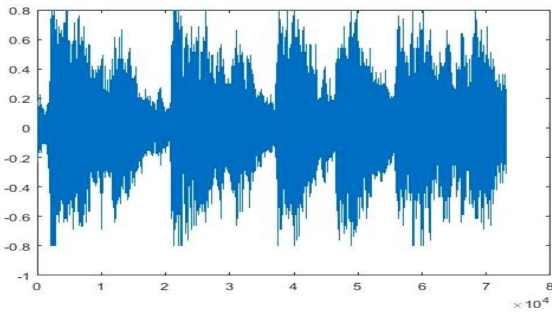
Figure 3(a): Original Speech signal sampled at 8 KHz
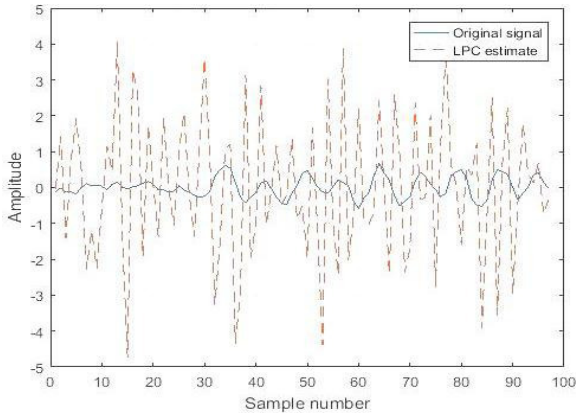

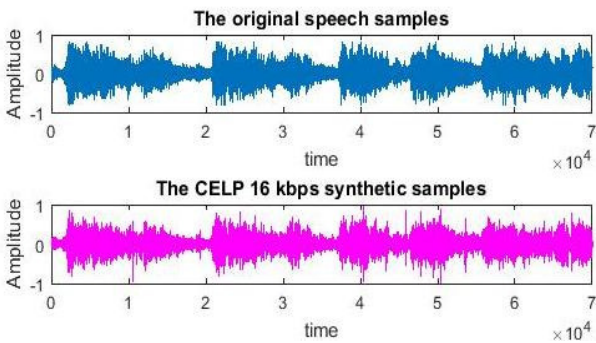Figure 3(b): LP coefficient estimate

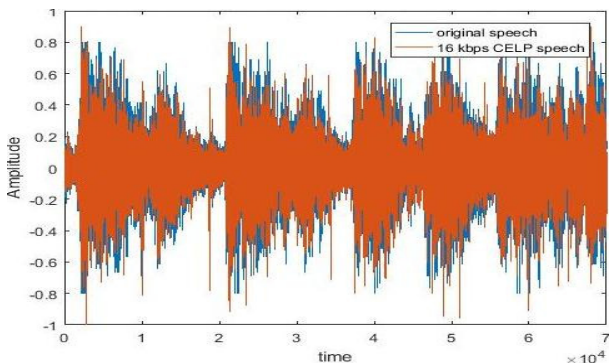Figure 3(c): Graph between 16kpbs LD-CELP and original signal

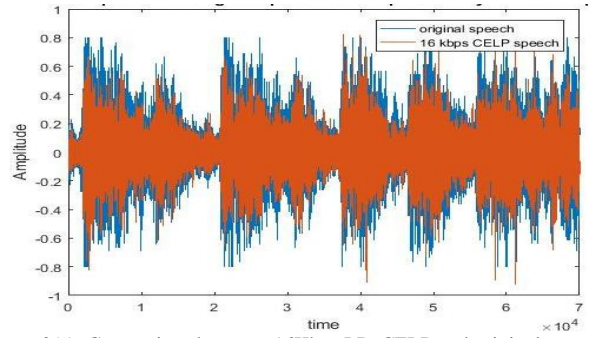Figure 3(d): Comparison between 16Kbps LD-CELP and original speech in one graph with c=0.85

Figure 3(e): Comparison between 16Kbps LD-CELP and original speech in one graph with c=0.65

B. MSE

The MSE is mean square error estimation of speech signals. In MATLAB MSE is calculating with the command -: mean ((desired - mean). ^2).

Table I shows the MSE (mean square error) of the 16 kbps LD-CELP compared with original 'hello' signal. From Fig. 4 it is concluded that for higher value of c the MSE is higher.

Table I

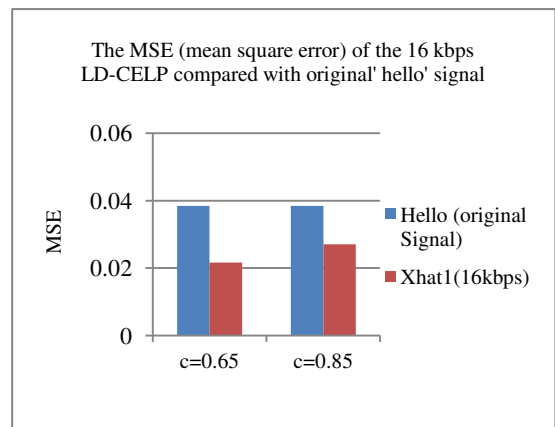| Speech signal	MSE estimation in dB		Comment
	c=0.65	c= 0.85	
'hello'	0.0385	0.0385	Original signal
'xhat1'	0.0217	0.0274	16 kbps decoded speech

Figure 4: Graphical representation of the MSE (mean square error) of the 16kbps LD-CELP compared with original 'hello' signal.

Table II shows the Signal to noise ration of the 16 kbps LD-CELP compared with original 'hello' signal. From Fig.5 it is concluded that for higher value of c lower the SNR.

Table II

Speech signal	SNR estimation in dB		Comment
	c=0.65	c= 0.85	
'hello'	109.704	109.662	Original signal
'xhat1'	98.367	85.009	16kbps decoded speech

The SNR (signal to noise ratio) of the 16 kbps LD-CELP compared with original' hello' signal
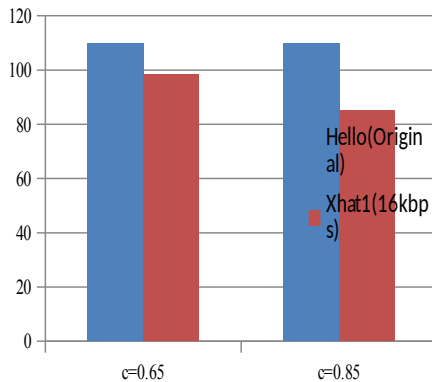

Figure 5: Graphical representation of the SNR (Signal to noise ratio) of the 16kbps LD-CELP compared with original ' hello' signal

VI.   CONCLUSION

In this paper, the Low Delay Code excited linear prediction (LD-CELP) algorithm for 16kbps is simulated. The 16kbps coder is designed using MATLAB simulation software. The linear prediction technique is used to find the coefficients for generation of fixed Gaussian code book and the whole parametric coding technique works on the 'Analysis –by- Synthesis' concept. Finally the parameters for coder comparison are evaluated and from graphical as well as from estimation values it is clearly shown that the coder is totally depend on the value of perceptual weighted constant (c). The comparison shows the results that the lower value of (c) is desired for better reconstruction of original speech signal for LD-CELP. Still the observations for SNR are not much better and enhancement in this technique is required for best speech quality.

VII.   FUTURE SCOPE

LD-CELP for 16 kbps is a parametric speech coder based on LP analysis the calculated SNR value for output speech is not good as per requirements of present day Digital Telephony. So advancement in this technique is necessary.

The other coders are also designed for better voice quality for speech as well as data services like iLBC and EVS. The frame independency is achieved with the help of adaptive codebook rather than fixed code book in iLBC. In 2014 recently 3Gpp standardized a new codec 'Enhance Voice Services' (EVS) codec which is the latest advancement in speech coding [7],[8]. The work for better quality speech and bandwidth optimization is continuing for the best outcomes, as the Digital Telephony expanding day by day.

REFERENCES

[1]   T. Ogunfunmi and M.J.Narasimha, "Speech Over VoIP Networks: Advanced Signal Processing and System Implementation", *IEEE Circuits and Systems Magazine*, pp.35-55, Second Quarter, 2012.
[2]   T. Ogunfunmi and M.J. Narasimha, "Principles of Speech Coding", CRC Press, 2010.
[3]   Prof. David L. Neuhoff Philip I.Tsai University of Michigan EECS 651 Project Report "Design of a CELP coder and analysis of various quantization techniques" Winter 2005.
[4]   Niranjan Shetty and Jerry D. Gibson "Perceptual Pre-weighting and Post-inverse weighting for Speech Coding" *Signals, Systems and Computers, ACSSC 2007, Conference Record of the Forty-First Asilomar Conference* 4-7 Nov. 2007.
[5]   Anssi Ramo and Henri Toukomaa Report **"**On comparing speech quality of various NB and WB speech coders". Multimedia Technologies laboratory, Nokia Research Center, Tampere, Finland.
[6]   B. Edler and G. Schuller, "Perceptual Audio Coding using Adaptive Pre and Post-Filters and Lossless Compression," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 10, pp. 881–884, 2000.
[7]   M.Dietz etal, "Overview of the EVS Codec Architecture", *Proc. IEEE ICASSP*, 2015.
[8]   Qualcomm Inc., IWAENC 2014 Presentation, "Enhanced Voice Services", Sept. 2014.
[9]   Prof. Tokunbo Ogunfunmi, Director Signal Processing Lab, Santa Clara University, California, USA, Mr. Shivakumar Mathapathi, Co-Founder and CTO Dew Mobility, Fremont California USA "Introduction to speech coding with internet of things (IOT) Applications" Course material   November 2016 NIT, Agartala (Tripura) India, by MHRDA India under GIAN (Global Initiative of Academic Networks) Courses.
[10]   Rhutuja Jage1, Savitha Upadhya "Implementation of CELP and MELP Speech Coding Techniques" *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 5, Issue 6, June 2016.