International Journal of Engineering & Technology, 7 (3.29) (2018) 586-590

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

An efficient low-bit rate CELP speech coding algorithm

A. Satyanarayana Murthy [1] *, A. Harika [1], Ch. Nikhita [1]

[1] *B. V. Raju Institute of Technology, Narsapur, Medak, TS*
**Corresponding author E-mail: satyanarayanamurthy.a@bvrit.ac.in*

Abstract

Voice communication is one of the easiest and natural methods of communication between human beings. In recent times it became popular because of mobile communications. It contains information as well as personal information such as age, identity, emotions etc. The requirement for efficient coding method is still a research topic for noise environment. Speech coding technique reduces the bandwidth requirements and storage space for speech data. Efficient coding method helps to reduce the bit rate and at the same time keep the speech quality and intelligibility at reasonable rate. Many speech coders have been developed to compress effectively and to ensure optimized performance to the human ear. Service providers of mobile phone communications are looking for low bit rate coders to allocate more number of users in limited bandwidth. In this paper a mixed coding method i.e. Code Exited Linear Prediction (CELP) is implemented on noise free speech signals using MATLAB. With this more exact modeling of spectral zeros is possible compared to the linear predictive coding (LPC). Subjective tests indicated that the coder at 16kbps and 9.6kbps achieves a significant improvement in performance over LPC coder under the same coding framework and bit allocation.

Keywords: *Linear Prediction; Code-Exited Linear Prediction; Speech Coding; Bit-Rate.*

1. Introduction

Voice is a natural gift to humans, used to get communication between humans. There is rapid growth in speech technologies [1]; computers can interpret speech signals in many applications. Speech is the result of complex changes of air pressure inside the vocal tract controlled by motor signals from the brain. While speaking lungs will supply the required air pressure into vocal folds to open and close the vocal cords. More energy will be supplied for the voiced speech compared to unvoiced speech signals. Almost periodic nature appears in voiced signals compared to unvoiced signals. So pitch is associated with the voiced signals and no pitch for unvoiced signals.

The vocal tract system will create the poles and harmonics in the speech. The vocal tract system can be treated like an analog filer. The poles will distinguish the voiced sounds in the filter response. No such clarity of peaks in the frequency response for unvoiced sounds. The important observation from speech signals is that the pitch does not change very fast and so an analysis window of 20-50 milliseconds will be used for processing. In this paper, two state of the art speech compression methods implemented using MATLAB. The analysis and synthesis of these two methods [2], [3] have been performed. The synthesis results perfectly matched with the original speech signals in code exited linear prediction compared to the linear prediction technique. The subjective and objective analysis has been performed on these results.

2. LPC synthesis

Let us see how the speech signals are coded in the cell phone [4]. The software in the cell phone executed millions of instructions per second using input speech data. As speech signal is stationary for short period they are processed for every 30 milliseconds of speech segments. A set of parameters have been extracted from the processing. Linear prediction is one of the most important algorithms for speech coding. In this model parameters like formants, voiced or unvoiced nature, intensity and pitch parameters are extracted from each segment, which will be capable of producing the speech in the encoder.

From the modeling we can understand that the discrete-time speech signal, s(n) results due to convolution of excitation signal, e(n) and time varying filter coefficients, a_k. The excitation and filter coefficients are derived from the LPC algorithm. The predicted sample, $\tilde{s}$(n) can be written as in Eqn. (2.1).

$$\tilde{s}(n) = \sum_{k=1}^{p} a_k \, . \, s(n-k) \qquad (2.1)$$

Where a_k, represents linear prediction filter coefficients

The error estimation, e(n) obtained from the original sample, s(n) and predicted sample, $\tilde{s}$(n) by the Eqn. (2.2).

$$E(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{p} a_k \, . \, s(n-k) \qquad (2.2)$$

The filter coefficients are evaluated using minimum mean square error and Levinson-Durbin algorithm [5] which is most efficient to get efficient values. The error computation is performed using the block diagram as shown in Fig. 2.1. Where A(z) is inverse filter transfer function of all-pole filter.

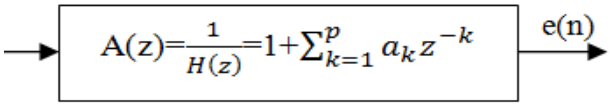

Fig. 2.1: Computing the LP Residual by Inverse Filtering.

The speech is sampled at 8 kHz frequency and is processed for every 30 milliseconds of the speech segments which consist of

240 speech samples. Speech is synthesized using a filter, whose transfer function, H (z) is given by Eqn. 2.3.

$$H(z) = \frac{1}{1+\sum_{k=1}^{p}(a_{k^2-k})} = \frac{1}{A(z)} \quad (2.3)$$

The details of the analysis and synthesis block diagrams are shown in Fig. 2.2. The following information is detected from every speech segment i.e., voiced or unvoiced, pitch period, filter coefficients. The data may be different for every segment and is to be updated from time to time.

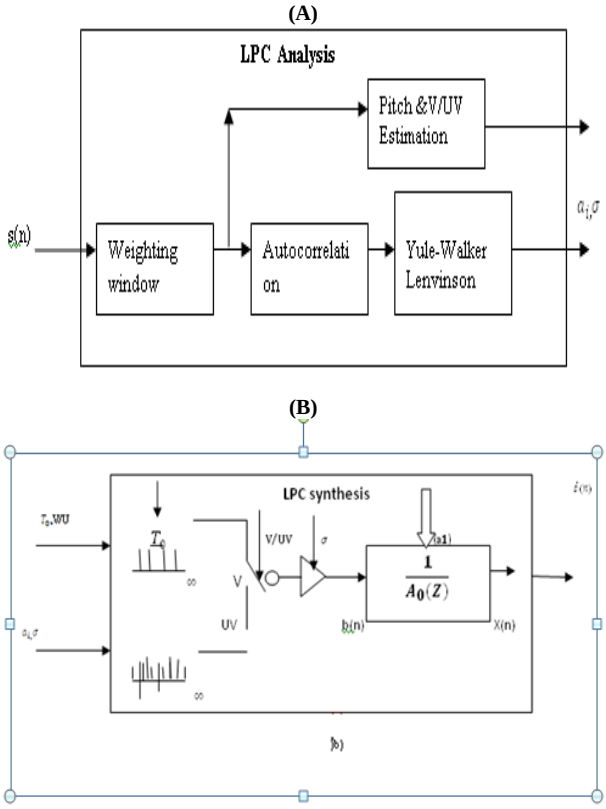

Fig. 2.2: Block Diagrams of LPC Analysis and Synthesis (Top to Bottom).

A speech synthesizer incorporates a model of vocal tract parameters to create the speech signal. Each segment is decoded individually and combined with the previous decoded elements to produce the continuity in the speech. The synthesized speech quality and intelligibility tested using subjective and objective analysis methods; a close to the original voice is the best synthesizer.

3. CELP synthesis

The LPC [6]-[8] is not an exact model of speech, but it produces satisfactory results. The model is limited to voiced and unvoiced, but cannot be modeled for fricative sounds. The fricative sounds having both the voiced and unvoiced characteristics which cannot be modeled appropriately. Also the LPC method is sensitive to the voice/unvoiced detection. The alternative to the LPC method is code-exited linear prediction (CELP) method [9]. The CELP is an improved method, generates quality and intelligent speech coding. In this approach one of the pre-determined fixed code-book is selected for excitation and sends only its index to the synthesizer, which has a similar code-book to generate the exact excitation. The block diagram for a CELP speech synthesizer at the encoder has shown in Fig. 3.1.

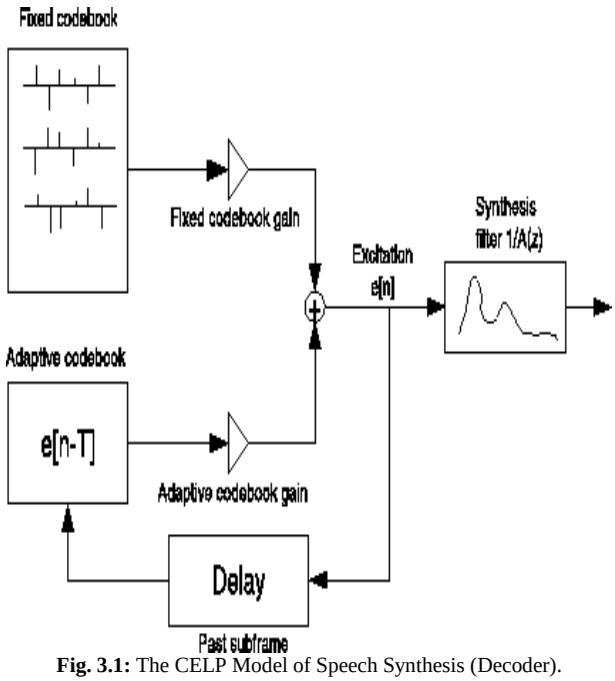

Fig. 3.1: The CELP Model of Speech Synthesis (Decoder).

Since the voiced speech has periodic in nature, the excitation sequence, e (n) at the synthesizer is determined by multiplying the past samples and gain i.e.$\beta e(n-T)$.

$$e(n) \cong p(n) = \beta e(n-T) \quad (3.1)$$

Where T is the pitch period of the voiced sequence.
Therefore the excitation signal for the synthesizer is sum of the pitch prediction and the code-book signal, c(n).

$$e(n) = p(n) + c(n) = \beta e(n-T) + c(n) \quad (3.2)$$

The Z-domain of the speech signal is given by the Eqn. 3.3

$$S(z) = \frac{C(z)}{A(z)(1-\beta z^{-T})} \quad (3.3)$$

The voiced and unvoiced sounds can be handled efficiently by the CELP technique compared to the LPC technique which only codes efficiently the voiced signals. In CELP coding a perfect excitation is generated to the filter.
In CELP a vector quantization is performed on the linear prediction error sequence; the error has a flat spectral response, makes it easier to produce the code-book. The predictor further improves its efficiency by using the adaptive vector quantization for voiced frames of speech.

4. Results and discussion

a) LPC:

Voiced speech, the female speech utterance “she has a smart way of wearing clothes” has taken for the experiment, the waveform and spectrogram has shown in

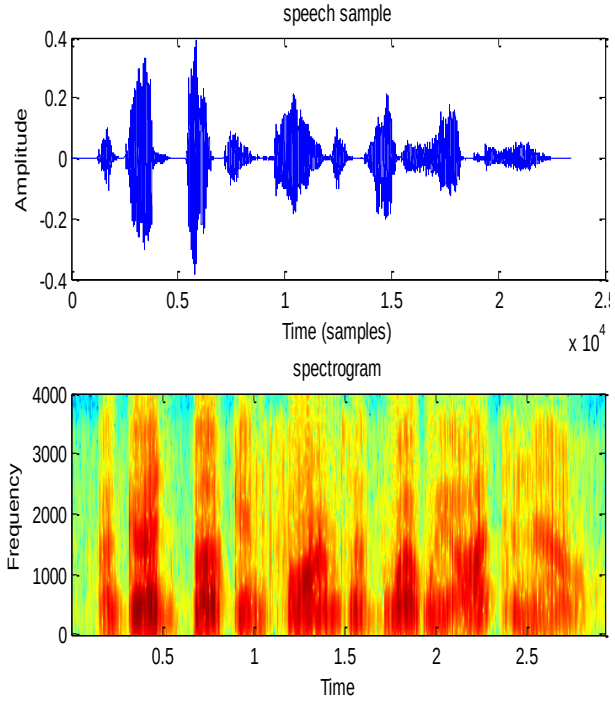

Fig. 4.1: Input Speech and Spectrogram (Top to Bottom).

The voiced segment is simulated by taking 30 ms frame which consists of 240 samples. The wave and its spectrum has shown in Fig. 4.2.

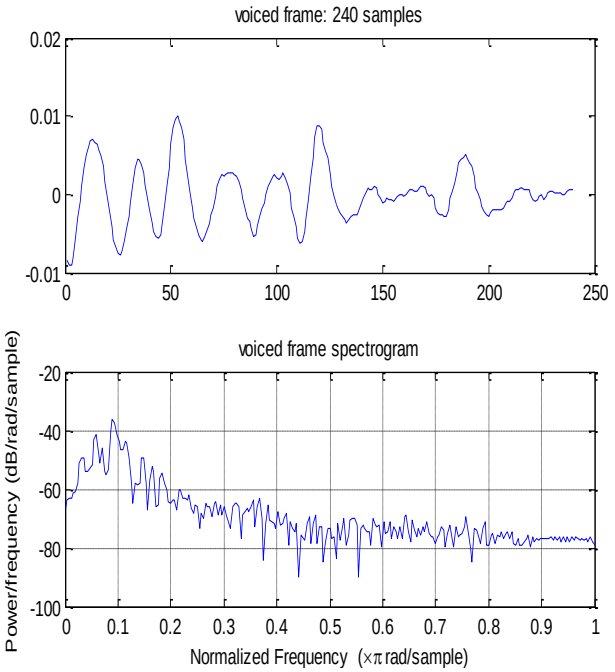

Fig. 4.2: A 30 Ms Long Voiced Speech and Spectrum (Top to Bottom).

The LP filter of order 10 has been chosen for the voiced frame, the filter coefficients (a_i) and variance of the residual signal (σ^2) has been obtained. The LP residual and its spectrum as shown in Fig. 4.3.

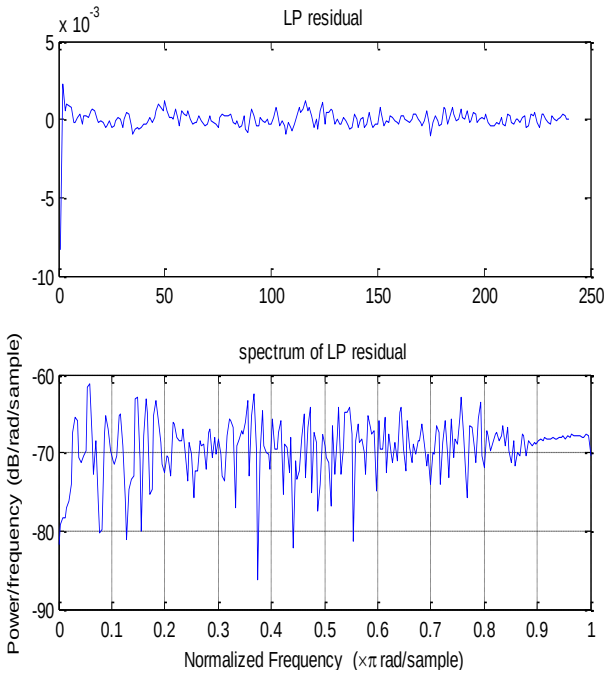

Fig. 4.3: The Residual Waveform and Spectrum (Top to Bottom).

The synthesis filter result has been in Fig. 4.4 which is exactly same as the original voiced frame.

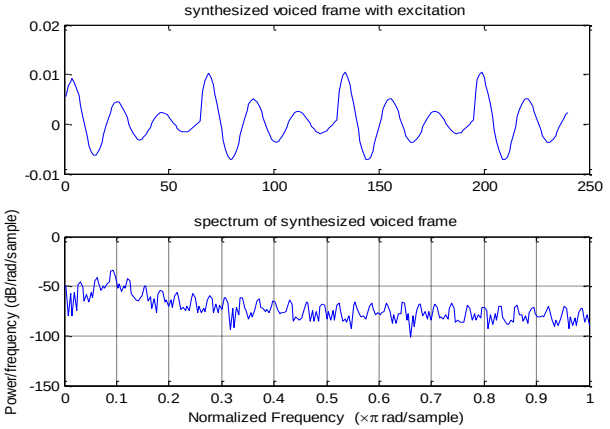

Fig. 4.4: Synthesized Speech Waveform and Spectrum (Top to Bottom).

b) Unvoiced speech

The unvoiced segment of 30 ms frame considered for simulation which consists of 240 samples. The original and synthesis wave spectrums are shown in Fig. 4.5-4.6.

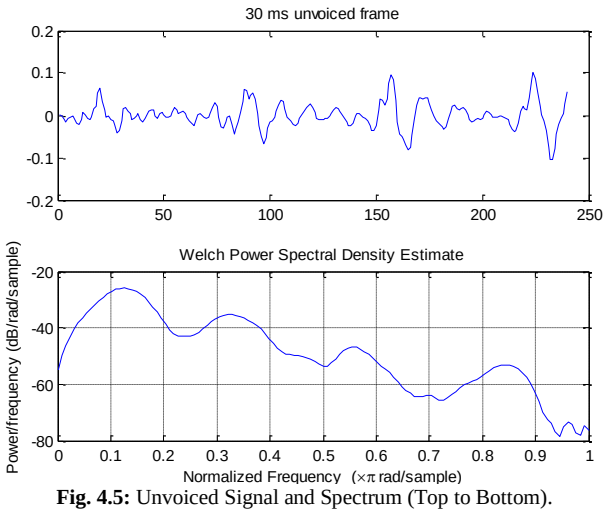

Fig. 4.5: Unvoiced Signal and Spectrum (Top to Bottom).

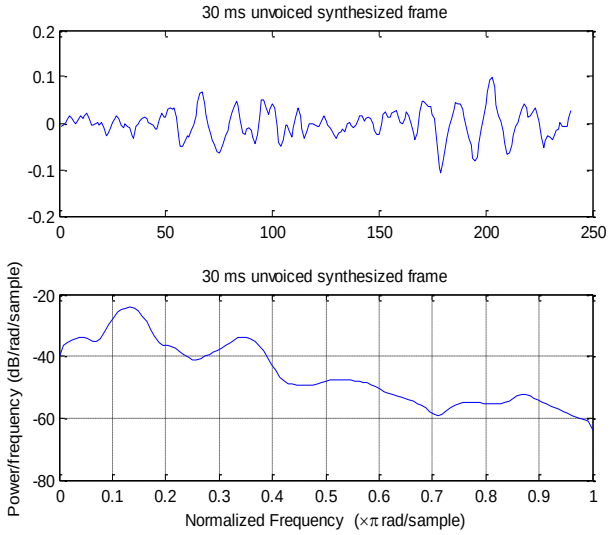

Fig. 4.6: Synthesized Unvoiced Signal and Spectrum (Top to Bottom).

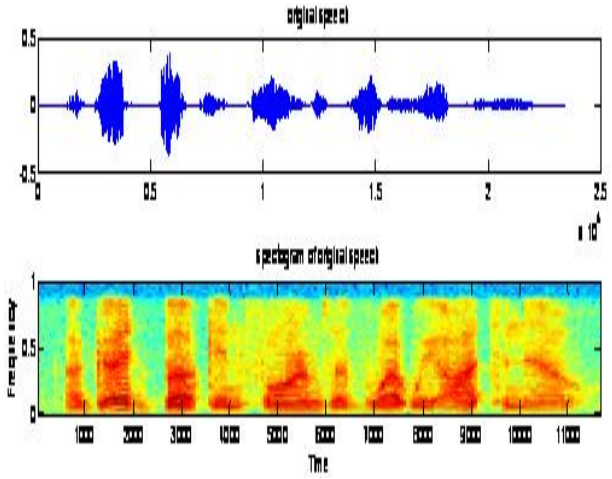

Fig. 4.7: Original Speech and Spectrogram.

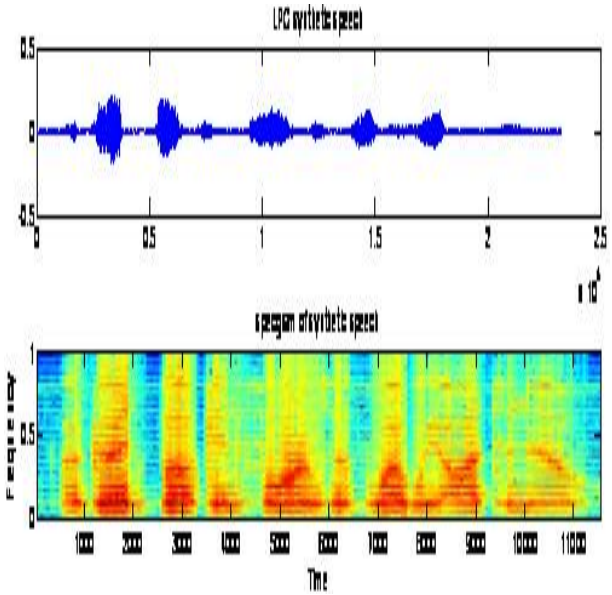

Fig. 4.8: LPC Synthetic Speech and Spectrogram.

The original and synthetic plots for the given speech utterance has shown in Fig. 4.7 and 4.8 using the LPC technique. It is found that the technique works well in reproducing the speech.

c) CELP Synthetic speech

The algorithm is executed on the same female voice which is used in the LPC synthesis. The simulations are performed using 16kbps and 9.6kbps coding techniques. The best linear combination of excitation components are selected from a codebook. The selection is performed in a closed loop so as to minimize the difference between the synthetic and original signals. The results are shown in Fig.4.7-4.8.

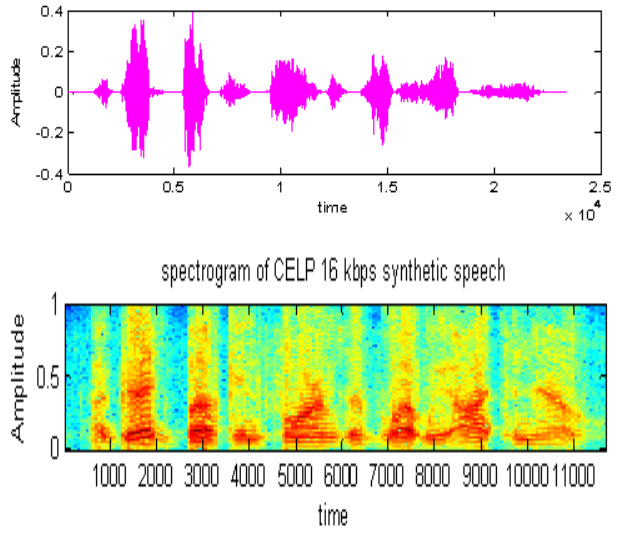

Fig. 4.8: 16 KBPS CELP: Signal and Spectrogram (Top to Bottom)

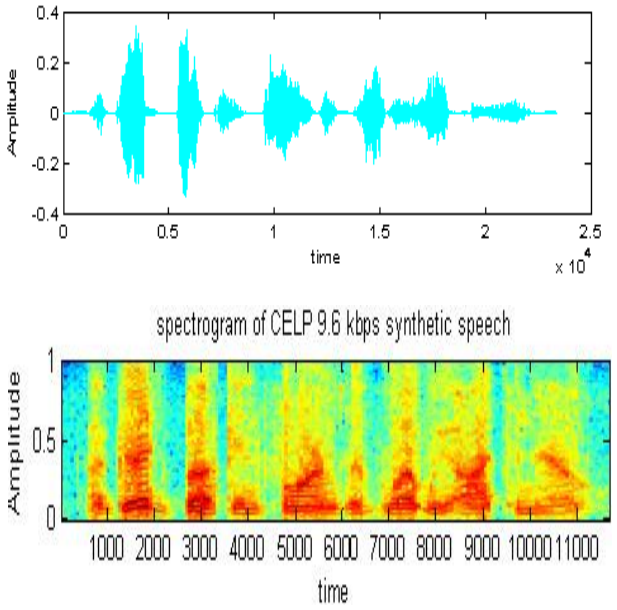

Fig. 4.9: 9.6 Kbps CELP: Signal and Spectrogram (Top to Bottom).

The synthetic speech from CELP technique is more natural than LPC technique. The plosive sounds are much better rendered and voiced sounds are no longer buzzy. In this method pitch and v/uv estimation are no longer required. The synthetic speech quality is maintained and is revealed by listening. The synthetic speech is similar to original speech waveform.

5. Conclusion

One of the most important communications between humans using speech signals. Humans produce the speech naturally where as technology produce the synthetic speech that is used to communicate between humans and machines. In this paper synthetic voice is generated by LPC and CELP techniques. The CELP methods generate the speech almost like nature speech. The results of these two experiments have been evaluated using listening and objective tests, which proven the two methods are working well to produce the synthetic speech.

References

[1] Ian Mcloughlin, Applied speech and Audio Processing, Cambridge University Press.

[2] B.S. Ttal and Suzanne and L.Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech wave", Bell Laboratories, New Jersey, 1971.

[3] Manfred R. Schroeder and Bishu S. Atal, "Code-exited linear prediction: High quality speech at low bit rates, AT&B Bell Laboratories, New Jercy.

[4] Klaus Fell Baum, Jorg Pichter, "Human speech production based on linear predictive vocoder, April, 1999.

[5] Digital Processing of Speech Signal by L.R.Rabiner and R.W. Schafer, Printice Hall, 1978.

[6] Florian Keiler, Daniel Arfib and Udo Zolzer, "Efficient Linear Prediction for Digital Audio Effects" Germany.

[7] Jani Nurminen, Victor Papa, "A Parametric Approach for Voice Conversion", Nokia Research Centre, Finland.

[8] Hwai-Tsuhu and Hss-Tsungwu, "A glottal- exited linear prediction model for low-bit – rate speech coding".

[9] Rhutuja Jage, Savitha Upadhya, "CELP and MELP speech coding techniques", International conference on wireless communications, signal processing and networking (WISPNET), 23-25, March 2016, Chennai, India.

[10] Theory Dutoit and Ferran Marquees, Applied Signal Processing. A Matlab based proof of concepts.

[11] Mahmoud A osman, Hussein M. Magboub Nasser, SA Alfandi, "Speech compression by LPC and Wavlet", 2nd International Conference in computer engineering and technology" (ICCET), 16-18, April 2010, Chengdu, China.

[12] Takehir, Moriya, "Progress in LPC-based frequency domain audio coding", Communication Laboratories", NTT, Atsugi, Japan, April 2016

[13] Kassim, Gunawan, "Development of low bit rate speech coder based on vector quantization and compressive sending", Journal of Applied Sciences, vol. 13, 49-59, 2013.