International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-4 Issue-5, June 2015

Performance Analysis of CS-ACELP Speech Coder

Nimisha Susan Jacob, Ancy S. Anselam, Sakuntala S. Pillai

Abstract*— *In modern communication systems, number of users to access the wired and wireless networks has increased rapidly. Consequently, the use of channel capacity has to be increased. Speech compression aims to compress the speech signal to attain maximum channel capacity with lower bit rate and highest quality. G.729 is one of the widely used standard in ITU-T for speech compression. This paper presents the analysis of CS-ACELP coder based on the objective measurements. Also the perceptual quality of the reconstructed speech was analyzed from the spectrogram based on the parameters pitch, intensity and formants.

Index Terms*—*Conjugate Structure Algebraic Code Excited Linear Prediction (CS-ACELP), International Telecommunications Union-Telecommunications (ITU-T), Linear Predictive Coding (LPC), Speech coding

I. INTRODUCTION

Speech coding is an art of creating minimally redundant representation of a speech signal. The goal of all speech coders is to minimize the distortion at a given bit rate, or minimize the bit rate to reach a given distortion with high quality. Analysis-by-synthesis method in speech coding produce good quality of speech[5]. In time domain analysis-by-synthesis coder, the excitation signal is chosen by attempting to match the reconstructed speech waveform as closely as possible to original speech waveform. G.729(8kbps) is one of the famous standard for speech compression by ITU-T in 1996[1]. ITU-T recommendation is based on Conjugate Structure Adaptive Code Excited Linear Prediction (CS-ACELP) algorithm, operating at a bit rate of 8Kbps for discrete speech samples sampled at a rate of 8000 samples per second. G.729 operates to take every 10ms speech frame, the input speech signal is analyzed to extract the parameters. These parameters are then coded. Standard G.729 with variable bit rate are available. The G.729 is a low complexity continuous data transmission scheme for VoIP applications and provide good synthesized speech quality at low bit rate. CS-ACELP can only be used for human voice (due to the model used) and is relatively complex. This paper focused on CS-ACELP algorithm standardized by ITU-T G.729[1]. In section II, an overview of complete coder described. Section III describes the method adopted for the analysis of coder. Section IV touches upon performance evaluation of the coder.

Manuscript published on 30 June 2015.
* Correspondence Author (s)
Nimisha Susan Jacob, Department of Electronics and Communication Engineering, Kerala University, Mar Baselios College of Engineering and Technology, Trivandrum, India.
Ancy S. Anselam, Department of Electronics and Communication Engineering, Kerala University, Mar Baselios College of Engineering and Technology, Trivandrum, India.
Sakuntala S. Pillai, Department of Electronics and Communication Engineering, Kerala University, Mar Baselios College of Engineering and Technology, Trivandrum, India.

II. CS-ACELP DESCRIPTION

A. Encoder

The CS-ACELP coder processes input signals on a frame-by-frame and subframe-by-subframe basis. The frame length is 10 ms and consists of two 5 ms subframes. The algorithm utilizes vector quantization method, both the adaptive codebook and fixed codebook are vector quantized to form conjugate structure. The 8kbps core speech coder is derived from G.729 coder[2] and the coder is based on Code-Excited Linear Predictive(CELP) coding model operating on speech frame using analysis-by-synthesis method. The encoding principle of CS-ACELP is shown in Fig. 1. The encoding stages of CS-ACELP mainly contain six blocks.

1. Pre Processing

Preprosessing block contains 2 stages scaling and highpass filtering .The input to the speech encoder is assumed to be a 16-bit PCM signal and it then undegos combained scaling and highpass filtering. The scaling means, dividing the input signal by a factor two to avoid the possibility of overflows in the fixed-point implementation of coder. For high pass filtering a second order pole/zero filter with a cut-off frequency of 140 Hz is used. Both the scaling and high-pass filtering are combined and the resulting filter is given by:

$$H_{h1}(z) = \frac{.46363718 - 0.927244705\,z^{-1} + 0.46363719\,z^{-2}}{1 - 1.92724705\,z^{-1} + 0.9114024\,z^{-2}} \quad (1)$$

The input signal filtered through $H_{h1}(z)$ is referred to as $s(n)$, and will be used in all subsequent coder operations.

2. LP Analysis

The linear prediction(LP) technique, taking the advantage of *Pth*-order linear prediction filter, is the most frequently used technique for speech analysis. LP analysis block is shown in Fig. 3. Reflection coefficients are obtained as by product of Levinson Durbin algorithm in LP analysis. The short term analysis and synthesis are based on 10th order LP filter.

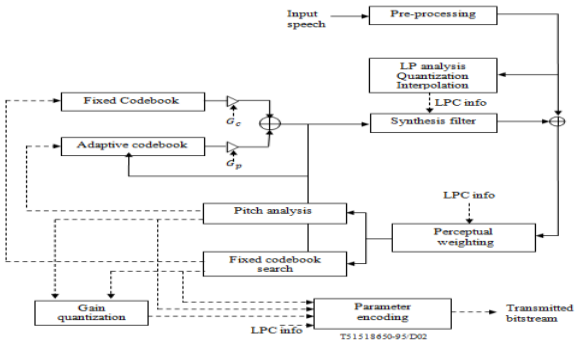


Fig. 1. CS-ACELP ALGORITHM

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

The LP synthesis filter is defined by:

$$\frac{1}{\hat{A}(z)} = \frac{1}{1+\sum_{i=1}^{10}\hat{a}_i z^i} \qquad (2)$$

where $\hat{a}i$, i = 1,...,10, are the (quantized) Linear Prediction (LP) coefficients.

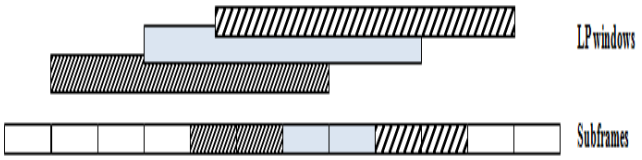


Fig. 2. WINDOWING PROCEDURE IN LP ANALYSIS

These LP coefficients are obtained by using Levinson-Durbin algorithm from the autocorrelation coefficients corresponding to the windowed speech. For obtaining windowed speech, the preprocessed signal, s(n) is windowed using a 30ms (240 samples) asymmetric window. The LP analysis window consists of half a hamming window and quarter of a cosine function cycle. The window is given by:

$$W_{lp}(n) = \begin{cases} .54 - \cos(\frac{2\pi n}{399}) & \text{n= 0,1..199} \\ \cos(\frac{2\pi(n-200)`}{159}) & \text{n=200,…239} \end{cases} \qquad (3)$$

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), shown in Fig. 2. Using the Levinson-Durbin algorithm, the LP coefficients are computed from the autocorrelation coefficients corresponding to the windowed speech. LP parameters are quantized in the line spectral pair (LSF) domain with 18 bits. The LSP paremeters are characterised with abetter interpolation and quantization. The quantization of LSP parameters are obatined by using predictive two-stage quantization[3].

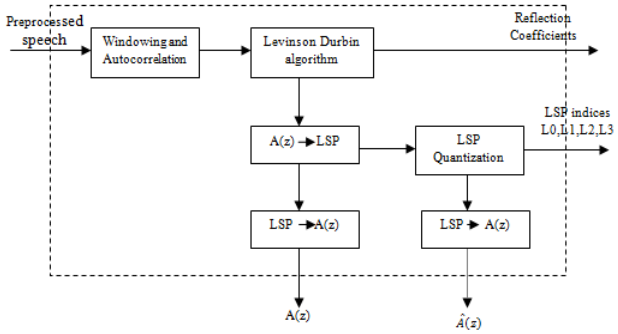


Fig. 3. LP ANALYSIS BLOCK

The interpolated quantized and unquantized filters are converted back to the LP filter coefficients (to construct the synthesis and weighting filters for each subframe).

3. Open-Loop Pitch Search

To reduce the complexity of the search for the best adaptive - codebook delay, search range is limited around a dealy T_{OP}. The pre-processed signal acts as an input signal for all the further analysis The excitation signal is chosen by an analysis-by-synthesis search procedure. For that, the weightened means squre error between the orginal and reconstructed speech is minimized acoording to percepually weightening filter. Percepual weightening is done for the sake of homology to the hearing effect of human eares. The amount of percetual weightening is made adaptive to improve the performance of input signal. The perceptual weightening filter is given by:

$$W(z) = \frac{A(z/\gamma_1)}{A(z/\gamma 2)} = \frac{1+\sum_{i=1}^{10}\hat{\gamma}_1 a_i z^{-i}}{1+\sum_{i=1}^{10}\hat{\gamma}_2 a_i z^{-i}} \qquad (4)$$

Where γ1 and γ2 are the Weight factor. The reflection coefficients k, obtained as a by-product from Levinson-Durbin algorithm are used to compute the adaptive weight factors, γ1 and γ2. By the proper adjusting of these variables it is possible to make the weighting more effective. The unquantized LP coefficients, a_i are used to perform the perceptual weighting. An open-loop pitch delay denoted by T_{OP} is estimated once per 10 ms frame by using the perceptually weighted speech signal Sw(n). Fig: 4 shows the open loop pitch analysis. Sw(n) is used for the open loop pitch lag estimation:

$$R(k) = \sum_{n=0}^{39} s_w(n)s_w(n-k) \qquad (5)$$

The three maxima of the correlation are found and they are in following three ranges; (20:39), (40:79), (80:143). The open loop pitch is obtained by taking the maxima of the three ranges by using the normalized autocorrelation function [1].:

$$R'(t_i) = \frac{R'(t_i)}{\sqrt{\sum_n s_w^2(n-t_i)}} \qquad \text{i=1,2,3} \qquad (6)$$

The winner among the three normalized correlations is selecting by favoring the delays with the values in the lower range. The computation of the pitch is dependent on the voiced and the unvoiced signal. The pitch contour lies in the voiced signal only.

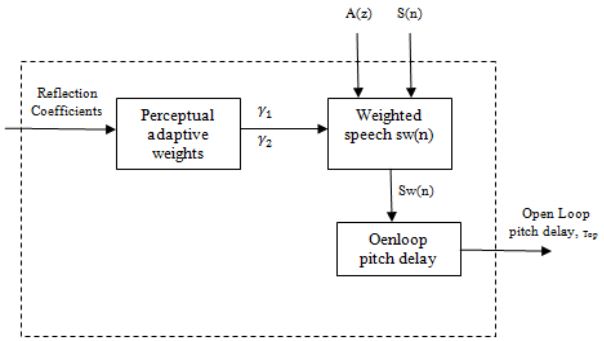


Fig. 4. OPEN-LOOP PITCH ANALYSIS

4. Closed-Loop Pitch Search

By through open-loop and closed loop pitch analysis the pitch delay is determined. The range of pitch are roughly obtained in openloop pitch search, a further precise pitch can be located when searching the closed-loop.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

For that the boundaries for the closed-loop pitch search are limited. Around the open-loop pitch delay, the closed-loop analysis is performed per 5ms sub-frame of the speech signal for better analysis with resolutions 1/3 in the range 19 1/3 – 84 2/3 and integer only in the range 85 – 143. For the first sub frame the pitch delay is encoded with 8 bits and for second subframe it is differentially encoded with 5 bits The optimal delay, closed loop pitch analysis, is determined by minimizing the mean-squared weighted error between the reconstructed and original speech. This is achieved by maximizing the term:

$$R(k) = 39\frac{\sum_{n=0} x(n)y_k(n)}{\sqrt{\sum_{n=0}^{39} x(n)y_k(n)}} \quad (7)$$

where $x(n)$ obtained by filtering the residual signal r(n), through the combination of 1/ $A\hat{}(z)$ and $W(z)$ and impulse response $h(n)$ of the weighted synthesis filter, by searching around and estimating the value of the open-loop pitch de-lay. Target signal computation is shown in fig:5. The signal to be matched is referred as the target signal $x(n)$. Then the adaptive codebook search is performed on a subframe (40samples) basis.

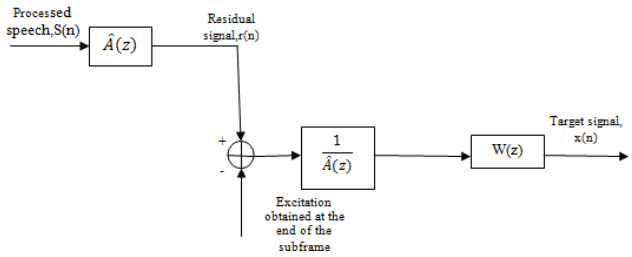


Fig. 5. COMPUTATION OF TARGET SIGNAL

5. Algebraic Codebook Search

With the determination of an optimal code vector as the goal of the algebraic codebook search. Fixed codebook is based an algebraic codebook structure (deterministic codebook), excitation code vector for the algebraic codebook is derived from the transmitted codebook index. In this codebook,each codebook vector usually contain four non zero impulsus with amplitude either +1 or -1[3]. 17 bits are usually occupy for fixedcodebook. the fixed codebook structure are shown in table 1. In the codebook the codebook vector, c(n), is determined by placing the 4 unit pulses at the found locations multiplied with their signs (±1) as follows:

$$c(n) = s_0\delta(n-m_0) + s_1\delta(n-m_1) + s_2\delta(n-m_2) + s_3\delta(n-m_3) \quad (8)$$

n=0...39

where δ_0 is a unit pulse.

Table 1. FIXED CODEBOOK STRUCTURE

Pulse	*Sign*	*Track*	*Positions*
i_o	$S_0 \pm 1$	T_0	m_0 0 5 10 15 20 25 30 35
i_1	$S_1 \pm 1$	T_1	m_1 1 6 11 16 21 26 31 36
i_2	$S_2 \pm 1$	T_2	m_2 2 7 12 17 22 27 32 37
i_3	$S_3 \pm 1$	T_3 T_4	m_3 3 8 13 18 23 28 33 38 4 9 14 19 24 29 34 39

The adaptive codebook contribution is subtracted from the target signal to obtain the new target signal x'(n)as follows:

$$x'(n) = x(n) - g_p y(n) \quad (9)$$

where y(n) is the filtered codebook vector obtained by convolving adaptive codebook vector v(n) and impulse response h(n) and g_p is the adaptive codebook gain. The correlation signal d(n) is computed from the target signal $x'(n)$ and the impulse response h(n), given by:

$$d(n) = \sum_{i=n}^{39} x'(i)h(i-1) \quad \text{n=0,...39} \quad (10)$$

If C_k is the fixed codebook vector , then the codebook is search by maximizing the term

$$\frac{C_k^2}{E_k} = \frac{(\sum_{n=0}^{39} d(n)c_k(n))^2}{c_k^t \Phi c_k} \quad (11)$$

The matrix Φ is calculated before the codebook search, contains correlation of h(n) and t denotes transpose.

III. METHOD

The CS-ACELP algorithm was simulated using MATLAB R2010. The MATLAB® application supports the import and export of data in various file format. The objective measurements like Segmented SNR (segSNR), Log Likelihood ratio (LLR), Weighted Spectral Slope Measures (WSS) and Perceptual Evaluation of Speech Quality (PESQ) of CS-ACELP was also calculated using MATLAB. Segmental SNR (SSNR) is defined as the average of SNR values over segments with speech activity. LLR which compares LPC vector of original speech signal with reconstructed speech. The Weighted Spectral Slope (WSS) distance measure is a direct spectral distance measure. It is based on comparison of smoothed spectra from the clean and distorted speech samples. The PESQ, is a family of standards comprising a test methodology for automated assessment of the speech quality as experienced by a user of a telephony system. PESQ values ranges from 0.5 to 4.5. Higher values of PSEQ provides better quality. For analysis of the output speech files, Praat software package was used. Praat consist of two part, Praat object and Praat picture. Spectrogram analysis of input and output speech of CS-ACELP was performed. Time domain representation of speech signal, information regarding pitch, intensity and formants was extracted. Formant frequencies f1, f2, f3, f4 and bandwidth information were also extracted. The Praat Objects window is used to open existing sound files or create.

IV. RESULTS AND DISCUSSION

This section describes the results of CS-ACELP algorithm obtained using the MATLAB R2010a. A handel.wav file is used as a test signal for coder implementation and handeldec is the reconstructed signal. First the file compression results are performed, then the analysis details and objective measurements were displayed. The objective measurements of are shown in Table 4. From the analysis of PSEQ value of speech signal, coder works with a better quality. From the Praat analysis of input and output of CS-ACELP coder, sound

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

pressure or amplitude of the sound waves in Pascal are shown in Table 2. The very slight differences in the minimum, maximum, root mean square (RMS) and mean values are representative of the good quality of output of the vocoder. The formant frequencies and corresponding bandwidths are listed in Table 3. The combined time domain representation of the input and output sound waves are shown in Fig. 6. From the time domain representation, it can be concluded that the overall shape has been preserved. But peaks have been clipped at some portions The spectrograms and waveforms of intensity, pitch and formants of the sound files are shown in Figs. 7 to 10

Table 2. CS-ACELP CODER AMPLITUDE (SOUND PRESSURE)

AMPLITUDE(in pa)	input (handel.wav)	output (handeldec.wav)
Minimum	-0.799	-0.863
Maximum	0.799	0.930
Mean	2.49×10^{-3}	5.05×10^{-3}
RMS	0.1962	0.124

Table 3. CS-ACELP CODER FORMANTS AND BANDWIDTH

PARAMETER(in Hz)	handel.wav	Handeldec.wav
F1	593.106	470.96
F2	1503.03	1050.91
F3	2281.968	2482.303
F4	2988.78	3027.77
BW1	381.71	136.74
BW2	269.57	1050.91
BW3	205.48	2482.30
BW4	206.714	280.97

Table 4. OBJECTIVE MEASUREMENTS

File Name	Objective Measures			
	SNRseg	WSS	LLR	PESQ
handel.wav	-1.81	47.75	0.309	2.36
male.wav	-1.66	42.58	0.430	2.27
child.wav	-1.97	89.80	0.530	2.11
female.wav	-0.90	80.40	0.670	2.17

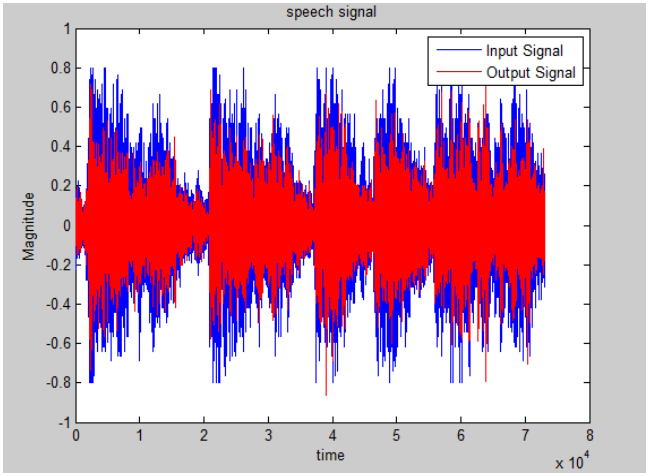
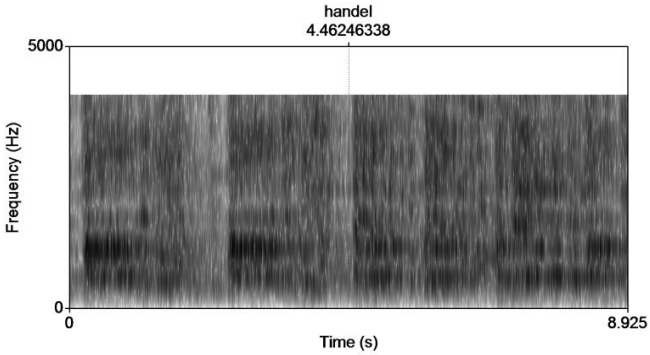
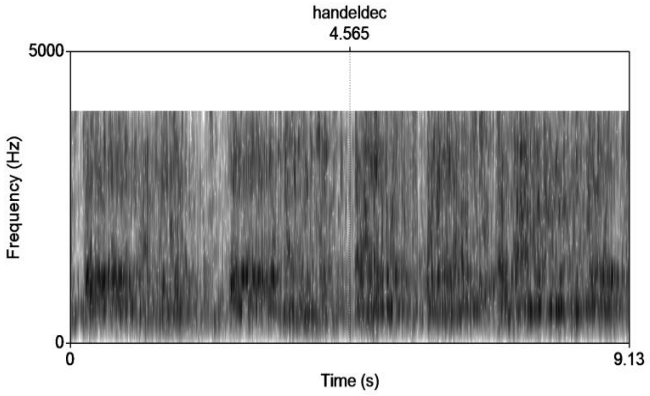


Fig. 6. Combined representation of original speech and reconstructed Speech of CS-ACELP vocoder

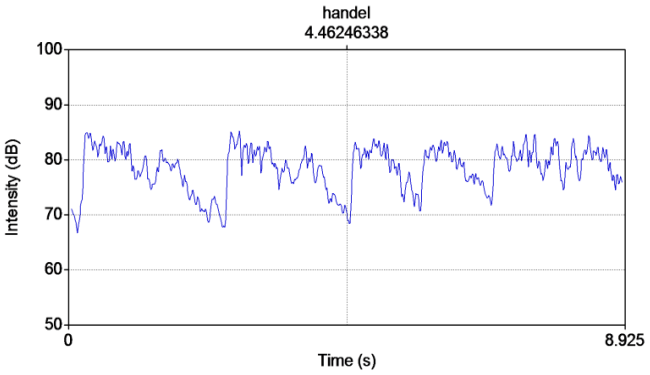


(a) Original sound file

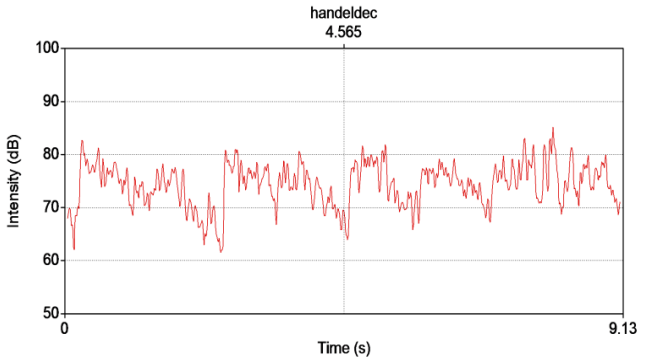


(b) CS-ACELP output file

Fig. 7. Spectrogram output



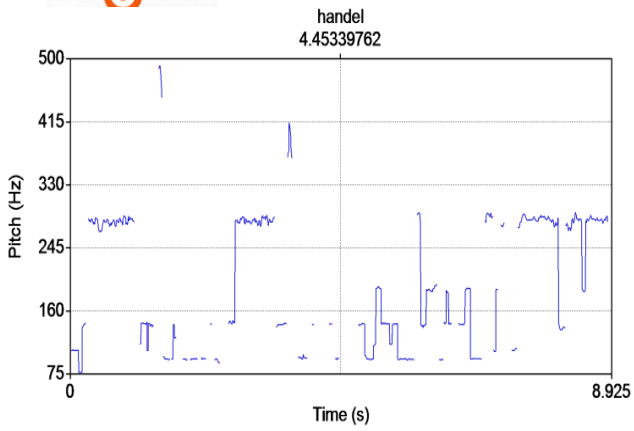
(a) Original sound file



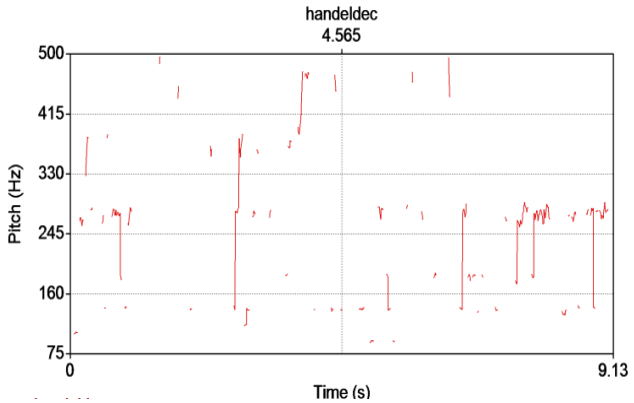
(b) CS-ACELP output file

Fig. 8. Intensity waveforms

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

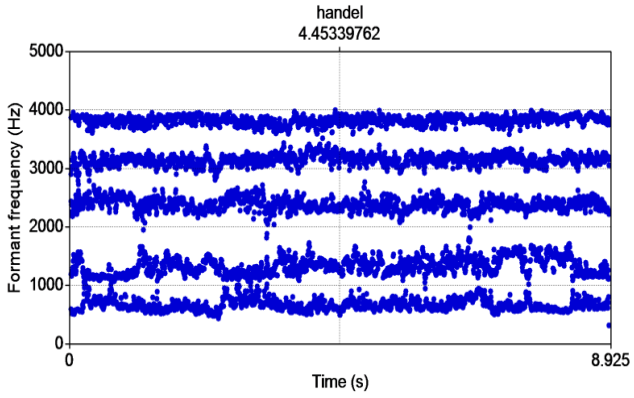


(a) Original sound file

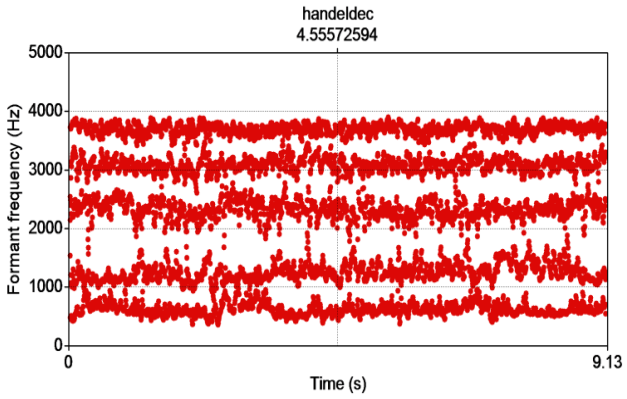


(b) CS-ACELP output file

Fig. 9. Pitch waveforms



(a) Original sound file



(b) CS-ACELP output file

Fig. 10. Formants of speech

V. CONCLUSION

The implementation and analysis of an efficient algorithm for providing secured speech transmission for various application with different speech input is described in this paper. Praat tools have proven to be very handy in speech file analysis. From the analysis of formant, pitch and intensity graphs of the input and output files clearly have very great similarity with the input wave form. From the experimental results, it is evident that the algorithm yields good compression and obtain very good perceptual quality

REFERENCES

[1] ITU-I Rec. G.729, " Coding of speech at 8kbps using Conjugate - Structure – Algebraic – Code - Excited Linear-Prediction(CS-ACELP" March 1996.

[2] Atti, V.; Spanias, A., "A simulation tool for introducing algebraic celp (ACELP) coding concepts in a DSP course," *Digital Signal Processing Workshop, 2002 and the 2ndSignal Processing Education Workshop. Proceedings of 2002 IEEE 10th* , vol., no., pp.306,311, 13-16 Oct. 2002

[3] Cheng-Yu Yeh; Yue-huan Zhong, "An efficient algebraic codebook search for G.729 speech codec," *Computer Applications and Information Systems (WCCAIS), 2014 World Congress on* , vol., no., pp.1,4, 17-19 Jan. 2014

[4] ITU-T Recommendation P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," Feb. 2001.

[5] R. Salami, et al., "Design and description of CS-ACELP: A toll quality 8 kb/s speech coder*", IEEE Trans. Speech and Audio Processing*, Vol.6, No. 2, pp. 116-130, March 1998.

[6] K. Ubul, A. Hamdulla and A. Aysa, "A Digital Signal Processing teaching methodology using Praat," *IEEE 4th International Conference on Computer Science and Education*, pp. 1804–1809, 2009.

[7] S. H. Hwang, "Computational improvement for G.729 standard," *Electronics Letters*, Vol. 36, No. 13, pp. 1163-1164, June 2000

[8] Chen, F.K.; Yang, J.-F.; Yan, Y.-L., "Candidate scheme for fast ACELP search," *Vision, Image and Signal Processing, IEE Proceedings* - , vol.149, no.1, pp.10,16, Feb 2002 doi: 10.1049/ip-vis:20020151

Nimisha Susan Jacob received B.Tech degree in Electronics and Communication Engineering from Thangal Kunju Musaliar Institute of Technology, Kollam, Kerala (Cochin University) in 2013. She is Currently doing M.Tech in Signal Processing.

Ancy S. Anselam received B.Tech degree in Electronics and Communication Engineering from N.S.S. College of Engineering, Palakkad (Calicut University) in 2000, M.Tech degree in Digital Systems and Communication Engineering from National Institute of Technology, Calicut in 2008. currently doing research under Kerala university in the field of Speech Coding. She has been working as faculty in Mar Baselios College of Engineering and Technology, Trivandrum from 2004 onwards. Her areas of interests include Cryptography, Information Theory, Coding Theory, Digital Communication, Fuzzy Logic and Speech Processing.

Sakuntala S. Pillai obtained from Ph. D Degree from University of Kerala 1989. She was the Head of the Department of Electronics and Communication from College of Engineering, Trivandrum from 1996-1998 later worked as Director, LBS Centre for Science and Technology, Trivandrum. She joined Mar Baselios College of Engineering and Technology, Trivandrum as Head the Department of Electronics and Communication in 2003. Currently she is working there as Dean (R&D). Her research interests include OFDM, MIMO wireless systems, Speech Coding etc. She is a senior member of IEEE, Fellow of IETE and Fellow of Institution of Engineers (INDIA).