Performance Analysis of Speech Enhancement Techniques in Code Excited Linear Prediction Coders

Linta Raj, Ancy S. Anselam, Sakuntala S. Pillai

Abstract— Emerging digital speech applications in various fields like wireless and multimedia created a demand for high quality speech. Speech in uncontrolled environment contains degradation components like background noise, thus the ultimate aim of speech enhancement algorithm is to reduce the noise or to improve the quality and intelligibility of corrupted speech. Different methods have been proposed for enhancing the speech quality. These enhancement methods can be used to improve the quality of speech coded using low bit – rate speech coders like Code Excited Linear Prediction (CELP) coder. This paper reports the performance evaluation of Spectral subtraction and Wiener noise reduction methods in CELP coder. The obtained results explain the performance evaluation of the speech enhancement algorithms and show improvements both qualitatively and quantitatively.

Index Terms— CELP Coder, Noise reduction, Spectral subtraction, Speech enhancement.

I. INTRODUCTION

Speech coding techniques seek to reduce the bit rate without an objectionable loss of speech quality. Significant progress has been made in speech coding at low bit rates. We can combine the speech coding and speech enhancement to improve the speech quality at low bit rates. The term Speech Enhancement refers to the improvement in the overall quality and intelligibility of speech signal. The Speech Enhancement problem received attention in the literature since mid 1970s and still an active area of research. Speech Enhancement techniques are designed for enhancing the speech degraded by additive noise for human listening or as preparation for processing before human listening. The enhancement algorithm can be applied at the decoder end of low bit rate coders like Code Excited Linear Prediction (CELP), to improve the performance. Enhancement of noise corrupted speech signal is useful in many applications like, speech recognition and voice communication. Approaches for speech enhancement vary depend on the context in which the interference problem exists. Previous works to suppress the effect of noise using spectral subtraction algorithm have demonstrated quantitative improvement in the quality of speech [1], [2].

Revised Version Manuscript Received on June 29, 2015.

Linta Raj, Department of Electronics and Communication Engineering, Mar Baselios College of Engineering and Technology, Thiruvananthapuram, India.

Ancy S. Anselam, Department of Electronics and Communication Engineering, Mar Baselios College of Engineering and Technology, Thiruvananthapuram, India.

Sakuntala S. Pillai, Department of Electronics and Communication Engineering, Mar Baselios College of Engineering and Technology, Thiruvananthapuram, India.

The noise reduction algorithm in single microphone technique is based on the estimation of short time spectral gain. Since the spectral gain is a function of *a priori* SNR and *a posteriori* SNR, it matches the previous frame rather than the current frame. So it degrades the performance of noise enhancement system. Therefore, a two step noise reduction technique (TSNR) is proposed. The main advantage of this approach is the cancellation of annoying reverberation effect of Decision Directed (DD) approach proposed by Ephraim and Malah in [3]. One major limitation of TSNR algorithm is that some harmonics are considered as noise only components and consequently suppressed by the noise suppression technique. In order to overcome this limitation, a new method named harmonic regeneration noise reduction (HRNR) is introduced which takes into account the harmonic characteristic of speech. TSNR and HRNR methods are presented in [4] and [5]. This paper discusses the simulation and quality evaluation of existing methods for the enhancement of degraded speech. Section II, III, IV and V describe the Speech enhancement algorithms and CELP coder respectively. Finally discuss the results obtained.

II. NOISE REDUCTION PRINCIPLE

In all speech enhancement techniques, speech signal model is, $s(t)=m(t)+n(t)$ where $s(t)$, $m(t)$ and $n(t)$ denote noisy speech, speech and noise signal respectively. $S(p,k)$, $M(p,k)$ and $N(p,k)$ represent the k - th spectral component of frame p of the noisy speech, clean speech and noise signal respectively. The main objective of the enhancement algorithm is to find an estimator $\hat{M}(p,k)$ which minimizes the given distortion measure. The main idea is to derive the SNR estimate from the available noise characteristics. An estimate of k-th spectral component can be obtained by applying a spectral gain $G(p,k)$ in each spectral component $S(p,k)$. The behavior of the spectral gain component is determined by the given distortion measure. However, the estimated SNR determines the efficiency and effectiveness of speech enhancement method. Most of the speech enhancement algorithms are based on two evaluation parameters, *a priori* SNR and *a posteriori* SNR respectively defined by,

$$SNR_{post}(p,k)=\frac{|S(p,k)|^2}{E\left[|N(p,k)|^2\right]} \quad (1)$$

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication Pvt. Ltd.

$$SNR_{prio}(p,k) = \frac{E\left[|M(p,k)|^2\right]}{E\left[|N(p,k)|^2\right]} \quad (2)$$

Where E[.] is the expectation operator.

III. SPECTRAL SUBTRACTION

This is the earliest method for speech enhancement. It has less computational complexity and is easy to implement. This method is based on the fact that human perception is not sensitive to the phase of speech signal. So the spectral amplitude of speech signal must be extracted effectively from the corrupted speech signal amplitude to have acceptable quality. But in practical situations power spectral density is used instead of spectral amplitude. Simple subtraction method entails subtracting the estimated noise power spectrum from the power spectrum of noisy speech, setting negative values to zero, recombining new power spectrum with the corresponding phase values and finally reconstructing the time domain waveform. Fig. 1 shows the basic block diagram of spectra subtraction method. First of all, the speech signal is segmented and windowed using a window function like hamming window. Then, the Discrete Fourier Transform (DFT) of the segmented and windowed signal are taken. The following assumptions are used for the analysis of speech enhancement system. Generally speech and the noise signal are neither ergodic nor stationary processes. It is also assumed that the Fourier transform coefficients are independent Gaussian random variables with mean zero and time varying variance.

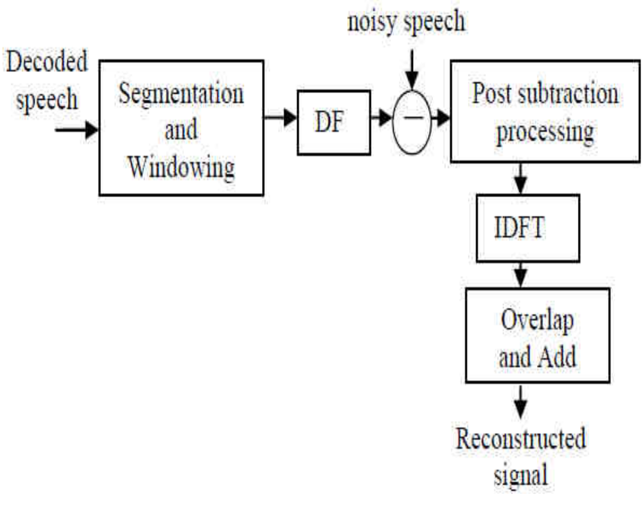

Fig. 1. Block diagram of Spectral subtraction

Due to the time - varying characteristics of speech signal, appropriate window is to be selected. Initial preprocessing is required to avoid the effect of excessive low frequency components and background noise.

IV. WEINER NOISE REDUCTION METHOD

A. Decision Directed Approach - Principle

Here the noise spectrum is estimated using voice activity detection. Then the *a priori* SNR and *a posteriori* SNR is computed based on the the estimated noise power spectral density.

$$\hat{SNR}_{post}(p,k) = \frac{|S(p-1,k)|^2}{\hat{\gamma}_n(p,k)} \quad (3)$$

$$SNR^{DD}_{prio}(p,k) = \beta\frac{|S(p,k)|^2}{\hat{\gamma}_n(p,k)} + (1-\beta)P[\hat{SNR}_{post}(p,k)-1] \quad (4)$$

Where P[.] denotes the half wave rectification and $\hat{M}(p-1,k)$ represents the estimated spectrum of previous frame. This a priori SNR estimator is called as DD approach. Te behavior of the estimator is controlled by the parameter β (typically value is 0.98) and then the chosen spectral gain function is the Weiner filter then,

$$G_{DD}(p,k) = \frac{SNR^{DD}_{prio}(p,k)}{1+SNR^{DD}_{prio}(p,k)} \quad (5)$$

This is the basic DD algorithm.

B. Two Step Noise Reduction

In this method there are two steps in order to enhance the performance of the noise reduction technique. When the parameter β close to one, the spectral gain computed for the present frame matches the previous one. The spectral gain of next frame can be computed using DD approach and applied to the current frame due to the frame delay. First of all, compute the spectral gain using DD approach as described in the previous section. In the second step, *a priori* SNR of the next frame is computed using the spectral gain obtained from DD approach. The *a priori* SNR can be computed as,

$$SNR^{TSNR}_{prio}(p,k) = \beta'\frac{|G_{DD}(p,k)S(p,k)|^2}{\hat{\gamma}_n(p,k)} + (1-\beta')P[SNR_{post}(p+1,k)-1] \quad (6)$$

The knowledge of future frame $S(p+1,k)$ is required for the computation of $SNR_{post}(p+1,k)$. In order to reduce the additional processing delay introduced due to the computation of *a posteriori* SNR of future frame, the parameter β' is selected as one. So the above equation can be written as,

$$SNR^{TSNR}_{prio} = \frac{|G_{DD}(p,k)S(p,k)|^2}{\hat{\gamma}_n} \quad (7)$$

Finally, the gain can be computed as,

$$G_{TSNR}(p,k) = h\left(SNR^{TSNR}_{prio}(p,k), SNR_{post}(p,k)\right) \quad (8)$$

And the estimated speech can be obtained as,

$$\hat{M}(p,k) = G^{TSNR}(p,k)S(p,k) \quad (9)$$

Thus, the chosen spectral gain of the Wiener filter is given by,

$$G_{TSNR}(p,k) = \frac{SNR^{TSNR}_{prio}(p,k)}{1+SNR^{TSNR}_{prio}(p,k)} \quad (10)$$

This algorithm in two steps defined by the above equations is called as the two step noise reduction technique.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

C. Harmonic Regeneration

Due to the inherent estimation errors, the output signal obtained by TSNR method suffers from distortions. In TSNR some harmonics are considered as noise only components and consequently suppressed by the algorithm. In order to prevent the distortions, we can take the advantage of the harmonic structure of speech. For that, the distorted signal is processed to get fully harmonic signal and this harmonic signal is used to regenerate the missing speech harmonics with biased amplitudes. A non-linear function is applied to the time signal enhanced with the classical noise reduction method. Then the restored signal is obtained by,

$$m_{harmo}(t) = NL(\hat{m}(t)) \tag{11}$$

$$M_{harmo}(p,k) = \frac{\rho|M(p,k)|^2+(1-\rho)|M_{harm}(p,k)|^2}{\hat{\gamma}_n(p,k)} \tag{12}$$

Where ρ is a constant used to control the mixing level of $|\hat{M}(p,k)|^2$ and $|M_{harm}(p,k)|^2$ (typically $0 < \rho < 1$). The speech signal can be estimated as,

$$\hat{S}(p,k)=G_{harmo}(p,k)X(p,k) \tag{13}$$

Where $G_{harmo}(p,k)$ is named as suppression gain. This gain has the ability to restore the harmonics suppressed by classic noise reduction methods. The important characteristic is that the artificially restored harmonics are generated at the same positions as the clean speech ones.

V. CELP CODER

In CELP coder, codebooks contain code vectors and the CELP algorithm is based on Analysis-by-Synthesis approach. The encoder (analyzer) analyzes the signal and produces linear prediction coefficients (LPCs). The LPCs are used to synthesis the signal. Same codebook is maintained on the encoder and decoder. Fig.2 and Fig.3 show the CELP encoder and decoder respectively. First of all, the encoder segments the speech into frames and then performs the linear prediction. At the decoder, it unpacks and decodes the bit stream. Two synthesis filters are there at the decoder part. First one is the pitch synthesis filter which inserts the pitch periodicities in the reconstructed signal. Second filter termed formant synthesis filter, introduces the frequency shaping effect based on the formant resonance present. Post filtering can also be used to enhance the quality of decoded speech.

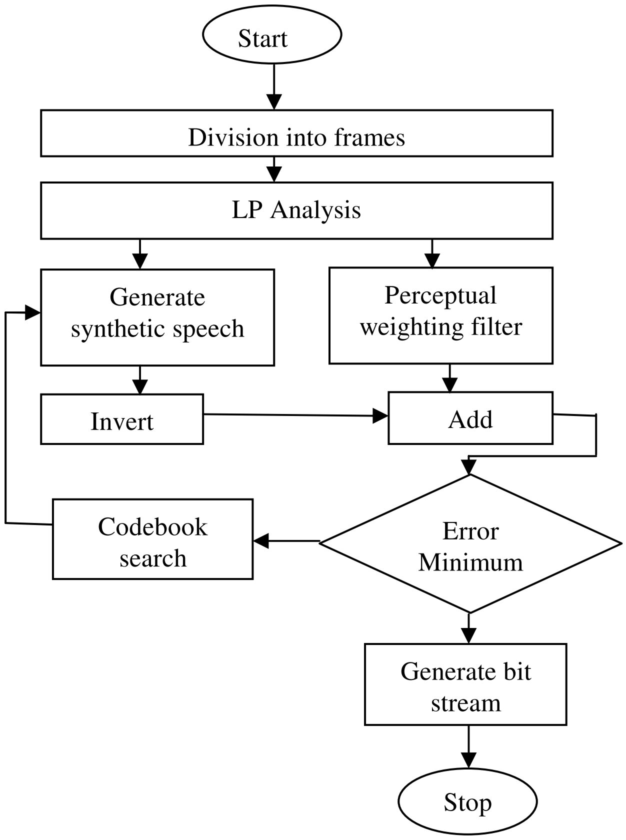

Fig. 2. CELP Encoder

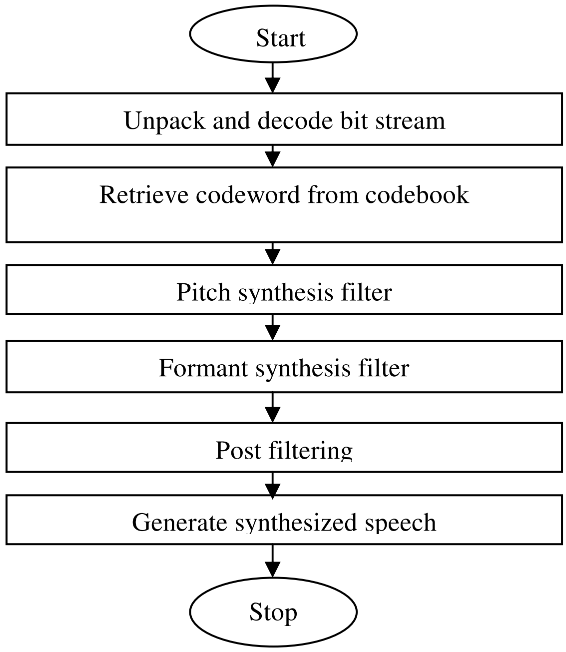

Fig. 3. CELP Decoder

VI. RESULTS

A. Waveform Analysis using PRAAT software

The simulation is performed in such a way that different speech files are given as input to the CELP encoder and then the synthesized speech obtained at the output of CELP decoder is given to the speech enhancement block for further quality improvement. Praat is a flexible tool for the analysis of speech signal. Different speech files are given as input to the CELP coder. And the decoded speech is given to the speech

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

enhancement block. Fig.4 shows the comparison of original speech and synthesized speech using 16kbps CELP coder. From the time domain waveforms it can be concluded that the overall shape has been preserved. But peaks have been clipped at some portions. Time domain representation of speech files as well as pitch and intensity contours were plotted using Praat object and picture windows.

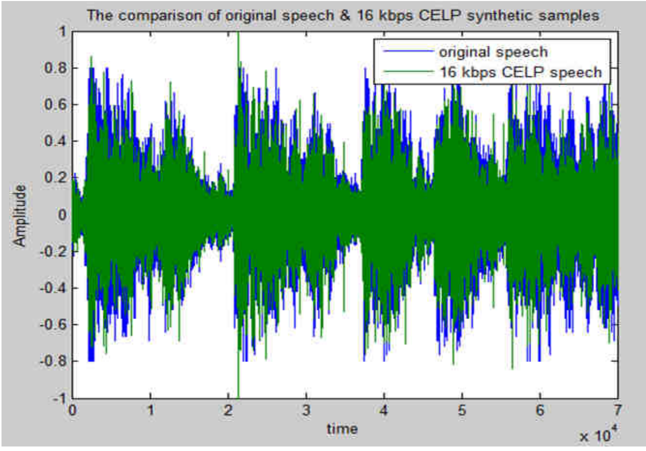

Fig. 4. Comparison of original speech and CELP synthesized speech

The results of waveform analysis using Praat software when handel.wav is given as input are shown in Fig. 5 and Fig. 6. In these figures comparison of the clean speech and speech enhanced using spectral subtraction and wiener noise reduction based on different waveforms are shown. The pitch and intensity contours show some variations from the clean speech.

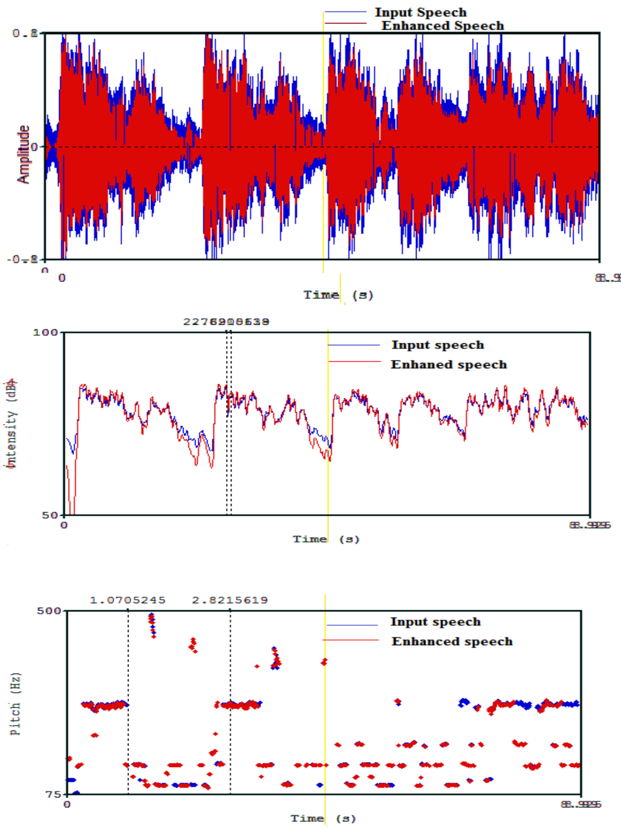

Fig. 5. Waveform, Intensity and pitch contours of input speech and CELP synthesized speech enhanced by Spectral Subtraction method

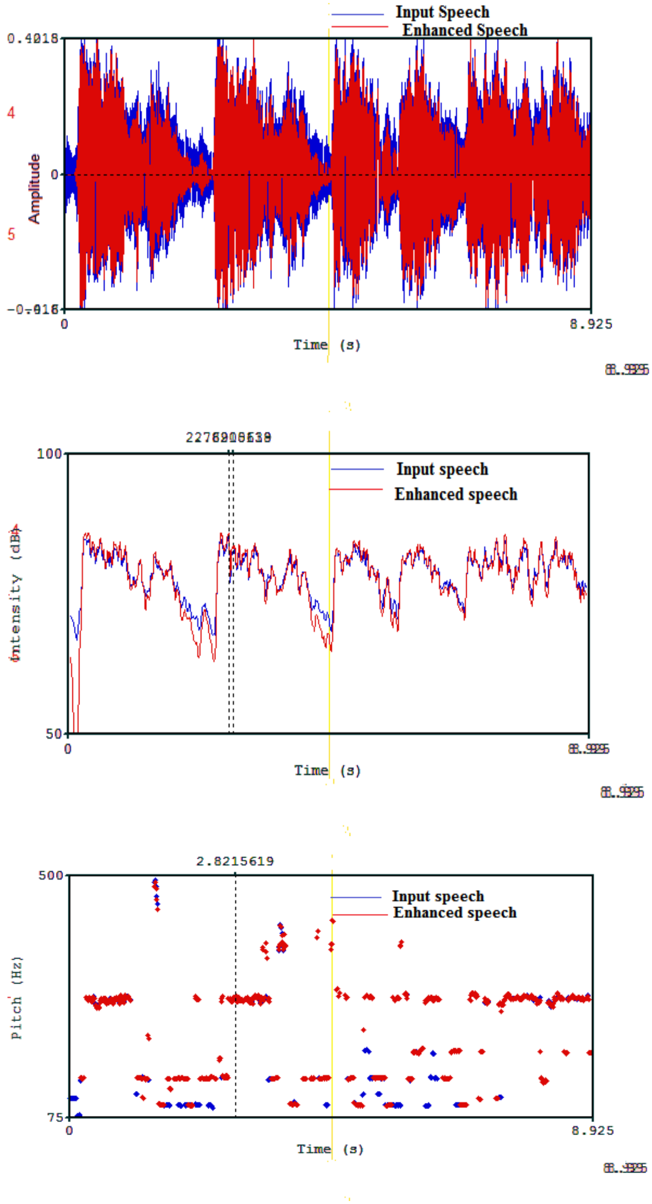

Fig. 6. Waveform, Intensity and pitch contours of input speech and CELP synthesized speech enhanced by Wiener noise reduction method

B. *Objective Testing*

The results of objective evaluation for Segmental SNR (SNRseg), Log Likelihood Ratio (LLR), Weighted Spectral Slope (WSS) and Perceptual Evaluation of Speech Quality (PESQ) are shown in table I and II.

TABLE I. OBJECTIVE QUALITY MEASURES OF SPECTRAL SUBTRACTION METHOD

File name	Objective measures			
	SNRseg	*WSS*	*LLR*	*PESQ*
handel.wav	-2.76	46.07	0.29	3.56
male.wav	8.05	11.91	0.29	3.00
women.wav	7.43	3.50	0.34	3.93
female.wav	8.01	8.23	0.11	3.52

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

TABLE II. OBJECTIVE QUALITY MEASURES OF WIENER NOISE REDUCTION METHOD

File name	Objective measures				
	SNRseg	*LRR*	*WSS*	*PESQ-TSNR*	*PESQ-HRNR*
handel.wav	4.34	0.09	15.16	3.53	3.56
male.wav	8.51	1.95	94.74	3.03	3.05
women2.wav	8.35	0.14	6.99	3.57	3.79
female.wav	8.81	1.18	14.45	3.77	3.90

TABLE III. PERFORMANCE COMPARISON OF SPEECH ENHANCEMENT METHODS IN CELP CODER

Input file name	Spectral Subtraction		Wiener Noise Reduction	
	PESQ Scores		PESQ Scores	
	Coder output	Enhanced coder output	Coder output	Enhanced coder output
handel.wav	2.43	3.56	2.43	3.67
male.wav	2.39	3.02	2.39	3.10
women.wav	2.99	3.12	2.99	3.48
female.wav	3.03	3.47	3.03	3.65

The Perceptual Evaluation of Speech Quality (PESQ) is an international standard for estimating the Mean Opinion Score (MOS) from both clean speech and degraded speech. Log Likelihood Ratio (LLR) value indicates the agreement between the spectral magnitudes of clean speech and degraded speech. Low LLR value indicates the close agreement between the spectral magnitudes. It is observed that the output files obtained from Wiener noise reduction method show higher SNRseg values compared to the other. Lower values of LLR and WSS are also desirable. Higher PESQ scores also imply better quality. HRNR method gives better PESQ scores. Table III shows the comparison of enhancement methods in CELP coders. It is very clear that the PESQ value after enhancement is higher than at the output of coder.

VII. CONCLUSION

The software simulation and comparison of spectral subtraction and wiener noise reduction method have been described in this paper. Simulation has been performed using different speech files. Since low bit rate and better perceptual quality are the two major constraints in the current scenario, it is more relevant to combine low bit rate speech coders with the speech enhancement techniques. From the results we can conclude that wiener noise reduction method gives better performance than spectral subtraction method. So these enhancement methods can also be applied to improve the performance of low bit rate coders like CELP.

REFERENCES

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. on Acoustics, Speech and Signal Processing, vol. ASSP 27, pp.113-120, April 1979.
[2] S. F. Boll, "Suppression of acoustic noise in speech using SABER method", Proceedings of IEEE International Conf. on Acoustics, Speech and Signal Processing, Tulsa, PP. 606-609, April 1978.
[3] Y. Ephraim and D. Malah, "Speech Enhancement using minimum mean square error spectral amplitude estimator", IEEE Trans. on Acoustics, Speech and Signal Processing, vol.ASSP-32, no.6,pp.1109-1121.
[4] C. Plapous, C. Marro, P. Scalart and L. Mauury, "A two-step noise reduction technique", in Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Montreal, Canada, Vol. 1,PP.289-292.
[5] C. Plapous, C. Marro, and P. Scalart, "Speech enhancement using harmonic regeneration",in Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Philadelphia, Vol.1,pp. 157-160, March 2005.
[6] Schroeder, Manfred and Atel,"Code Excited Linear aprediction (CELP): High quality speech at very low bit rates",IEEE Int. Conf. on Acoustics, Speech and Signal Processing(ICASSP), Vol.10, PP>937-940, April 1985.
[7] Ancy S. Anselam and Sakuntala S. Pillai,"Performane evaluation of Code Excited Linear Prediction coders at various bit rates", IEEE Int. Conf. on Computation, Power, Energy, Information and Communication, 2014.

Linta Raj received B.Tech degree in Electronics and Communication Engineering from Mar Baselios College of Engineering and Technology, Trivandrum, (Kerala university) in 2012. She is Currently doing M.Tech in Telecommunication Engineering.

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).

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.