ISSN 2229-5518

Effect of MMSE- STSA Algorithm in CELP and MELPSpeech Coders

Ancy S. Anselam, Sakuntala S. Pillai

Abstract— The role of speech coding is to reduce the bit rate by maintaining good speech quality. In order to improve the perceptual quality of degraded speech, different speech enhancement methods can be used. So, it is worthwhile to do research in joint systems (Speech Enhancement and Low bit rate speech coders). The work reported in this paper shows the improvement in the perceptual quality of speech coder outputs by incorporating speech enhancement technique. The simulation results of Code Excited Linear Prediction (CELP) and Mixed Excited Linear Prediction (MELP) speech coders with Minimum Mean Square Error Spectral Amplitude Estimator (MMSE-STSA) enhancement technique is analyzed in terms of objective quality measures and using PRAAT software.

Index Terms— Speech Enhancement, CELP, MELP, Perceptual Evaluation of Speech Quality, MMSE-STSA.

1 Introduction

SPEECH is a non-stationary signal but it can be divided into segments which have some common acoustic properties for a short time period. Therefore, it is easy to analyze the signal for a short interval. Idea of speech processing is manipulating the signal depending on the applications. Applications of speech processing include automatic speech recognition, speaker identification, speech coding [1], speech synthesis, speech enhancement etc. The concept of speech coding is quite different. Since there is limited allocated bandwidth, we need to reduce the number of bits required to represent the speech signal. So, coding is the process of representing the speech signal with minimum number of bits for transmitting the signal through wired or wireless channels.

Speech enhancement [2] is a necessary process in noisy systems to enhance the degraded speech signal. Different approaches are available in the literature to enhance the degraded speech. One approach is the preprocessing technique, where the signal is preprocessed before it gets degraded. Another one is the post processing approach [3]. In post processing method the degraded signal is further processed to enhance its quality. Nowadays, the combination of speech coding and speech enhancement is a hot research area.

In this paper Minimum Mean Squared Error Space Time Spectral Amplitude Estimator (MMSE STSA) [4] is applied to enhance the quality of speech decoded using low bit - rate speech coders like Code Excited Linear Prediction (CELP) [5] and Mixed Excited Linear Prediction (MELP) [6]. This paper is organized as follows: section II, III and IV describe the CELP, MELP and MMSE – STSA algorithms respectively.

- *Authors are in the Electronics and Communication Engineering Department of Mar Baselios College of Engineering and Technology, Thiruvanathapuram, Kerala, India. E-mail: ancy_anselam@yahoo.co.in and sakuntala.pillai@gmail.com.*

2 SPEECH CODING ALGORITHMS

2.1 CELP

CELP is a compression technique based on linear prediction method. The idea behind linear prediction is to predict the present sample of the signal from the previous samples. In CELP there is a look up table named code book to obtain the best match for the signal. In speech coders, speech signal can be uniquely represented using different parameters such as gain, linear prediction coefficients etc. The parameters extracted from the signal, which are quantized and then used to reconstruct the signal.

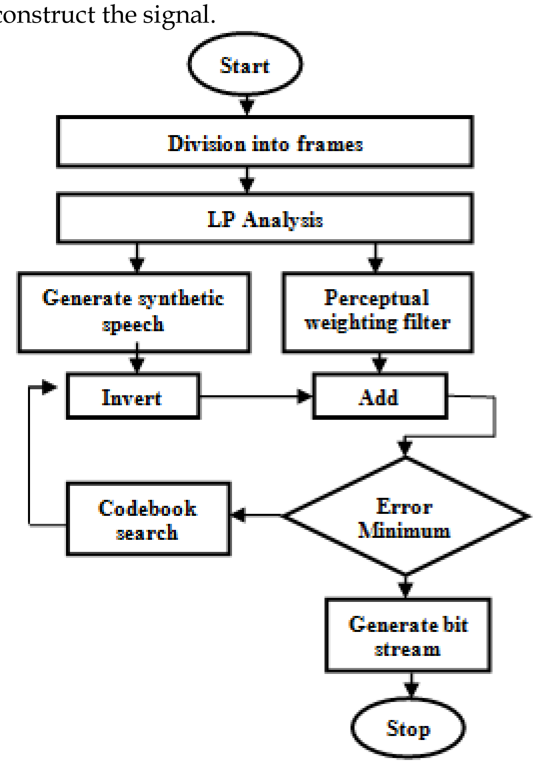


Fig. 1. CELP Encoder: Design flow

ISSN 2229-5518

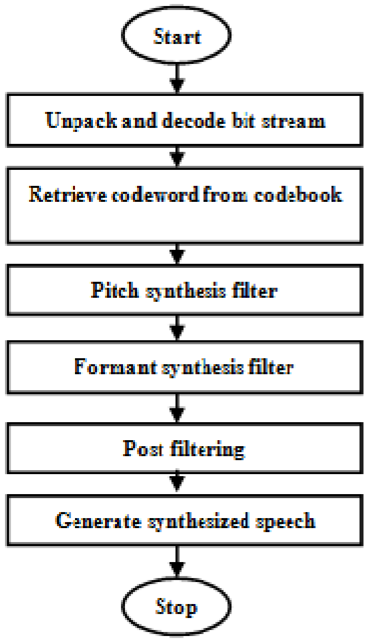


Fig 2. CELP Decoder: Design flow

Fig. 1 and Fig. 2 represent the design flow of CELP encoder and decoder respectively. At the encoder side, the speech signal is divided into different frames. The formant synthesis filter function and weighting filter functions can be defined as follows;

$$H_f(z) = \frac{1}{A(z)} = \sum_{k=1}^{N} a_k z^{-k} \quad (1)$$

Where a_k is the linear prediction coefficient (LPC).

$$W(z) = \frac{A(z)}{A\left(z/\gamma\right)} = \frac{1+\sum_{k=0}^{N} a_k z^{-k}}{1+\sum_{k=0}^{N} a_k \gamma^k z^{-k}} \quad (2)$$

2.2 MELP

Fig. 3 and Fig. 4 show the design flow of MELP encoder and decoder. The input speech is first given to the postfiltering block. Then the pitch is estimated from the postfiltered output by computing the pitch lag between the speech samples. In bandpass voicing analysis step the speech signal is divided into different frequency bands. After LP analysis vector quantization is performed.

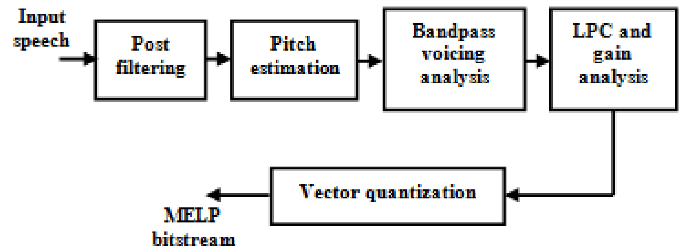


Fig. 3. MELP Encoder: Design flow

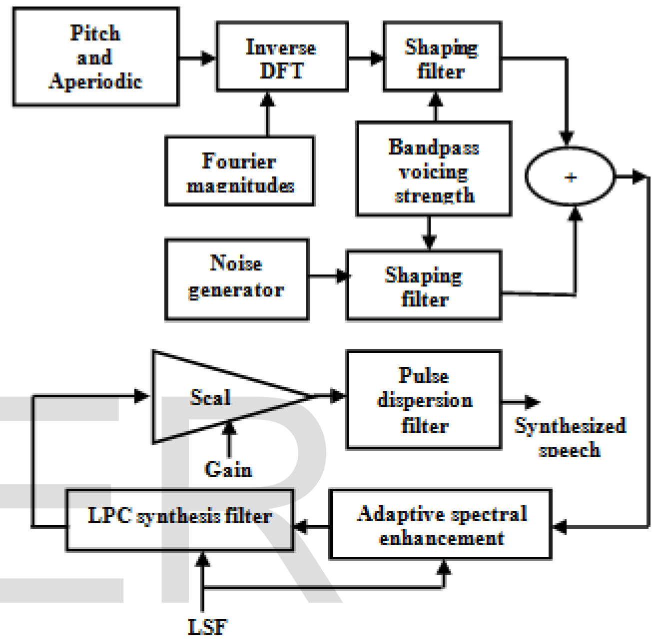


Fig. 4. MELP Decoder: Design flow

At the decoder end, the retrieved bits are unpacked and converted into parameter codeword. In MELP, the mixed excitation signal is obtained by summing the pulse and noise excitation signals. The advantage of this mixing is that it helps in countering the effect of buzz and tonal thumps in the decoded output. Before synthesizing the speech, the mixed excitation is enhanced first and then LP coefficients of this enhancement stage are generated.

3 SPEECH ENHANCEMENT ALGORITHM

3.1 MMSE – STSA ALGORITHM

The minimum mean square error estimate can be defined as the mean of the a posteriori density function.

$$\hat{A}_k = E\left[A_k / Y_k\right] = \int_0^{\infty}\int_0^{2\pi} a_k\, p\left(Y_k / a_k, \alpha_k\right) p\left(a_k \alpha_k\right) d\alpha_k\, da_k \quad (3)$$

where p(.) is the probability density function.

ISSN 2229-5518

$$p(Y_k / a_k, \alpha_k) = \frac{1}{\pi \lambda_d(k)} \exp\left\{ \frac{1}{\lambda_d(k)} \left| Y_k - a_k e^{j\alpha_k} \right|^2 \right\} \quad (4)$$

$$p(\alpha_k, a_k) = \frac{a_k}{\pi \lambda_x(k)} \exp\left(- \frac{a_k^2}{\pi \lambda_x(k)} \right) \quad (5)$$

where $\lambda_x(k)$ and $\lambda_d(k)$ represent kthspectral component of speech and noise respectively. Then the gain function of MMSE – STSA estimator can be computed as;

$$G_{MMSE}(v_k) = \Gamma(1.5) \frac{\sqrt{v_k}}{\lambda_k} \exp\left(-\frac{v_k}{2} \right) \left[(1 + v_k) I_0\left(\frac{v_k}{2} \right) + v_k I_1\left(\frac{v_k}{2} \right) \right] \quad (6)$$

where $\Gamma(1.5) = \sqrt{\pi}/2$ is the gamma function and I_0 is the zeroth order Bessel function. v_k is defined as;

$$v_k = \frac{\varepsilon_k}{1 + \varepsilon_k} \gamma_k \quad (7)$$

where ε_k and γ_k are the a priori SNR and a posteriori SNR values.

4 Results and Discussions

CELP and MELP coders were simulated using MATLAB-13. Different speech files were given as the input to the coders. Fig. 5 and fig. 6 show the comparison of clean speech (blue) and reconstructed speech (green). From the time domain representation we can definitely notice the degradation in the amplitude values. This clipping of amplitude reduces the perceptual quality of speech. So speech enhancement method can be used to improve the perceptual quality.

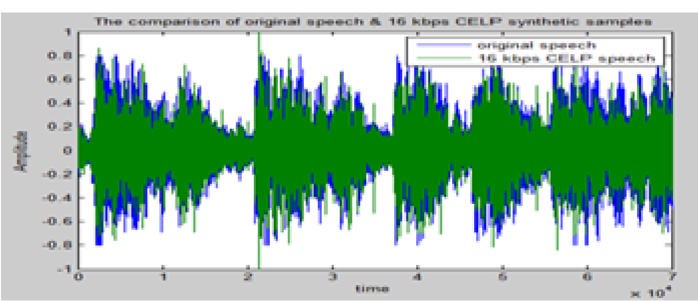


Fig. 5.Clean and CELP synthesized speech waveforms

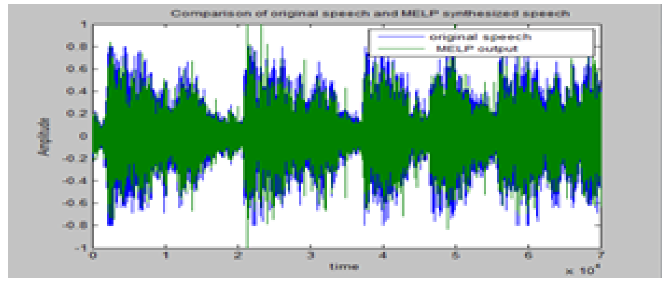
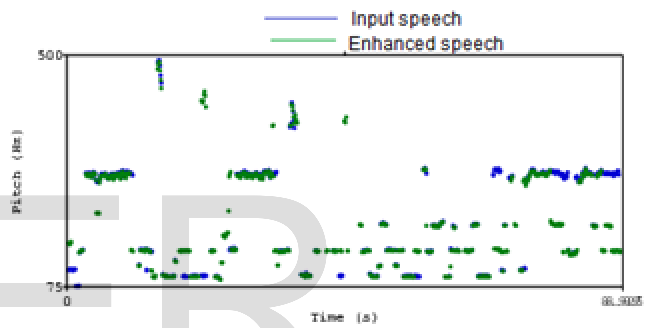


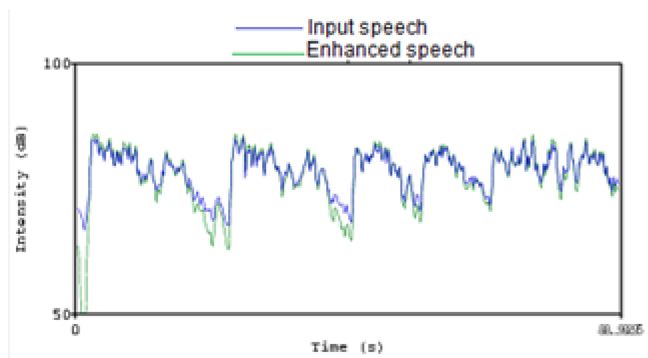
Fig. 6.Clean and MELP synthesized speech waveforms

4.1 Waveform analysis using PRAAT software

The results obtained during simulation were analyzed using PRAAT software to see the effect of MMSE STSA algorithm in CELP and MELP coders. Fig. 7 and 8 show the time domain representation, pitch and intensity contours of the clean speech and the enhanced decoded speech using CELP and MELP respectively.



(a)Pitch contour

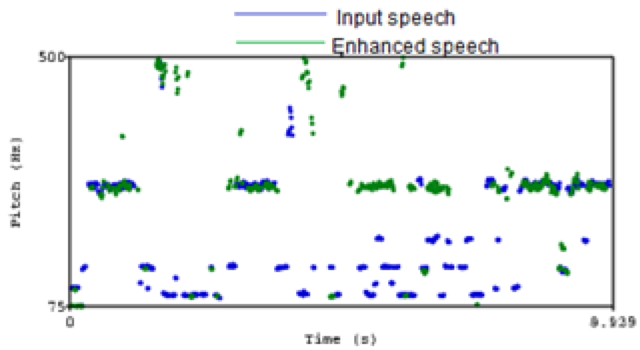


(b)Intensity contour

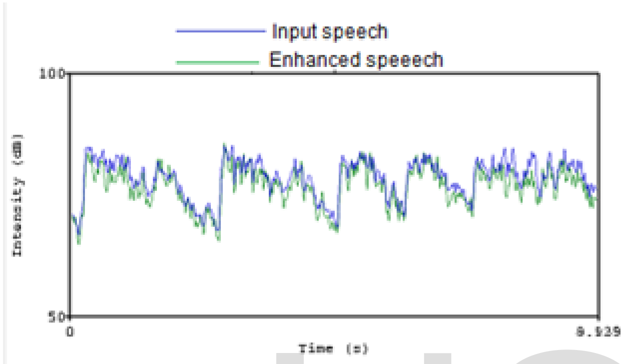
Fig. 7. (a) Pitch and (b) Intensity contours of clean speech and CELP synthesized speech enhanced by MMSE – STSA Algorithm

Pitch contour explains the voiced and unvoiced information present in the speech. Energy variations in the speech are specified with intensity contour. From the figures it is clear that, closest matching of the signal peaks is in CELP coder output. Variations are also there in the pitch and intensity contours of both the coders. But more pitch information are missing in the contour of MELP compared to CELP.

ISSN 2229-5518



(a) Pitch contour



(b) Intensity contour

Fig. 8.(a) intensity and (b) pitch contours of clean speech and MELP synthesized speech enhanced by MMSE – STSA Algorithm

4.2 Analysis using Objective Quality Measures

Objective testing was performed using MATLAB.

TABLE: 1
OBJECTIVE QUALITY MEASURES OF CELP - WITHOUT MMSE-STSA ALGORITHM

Speech files	Objective measures			
	SNRseg	*WSS*	*LLR*	*PESQ*
English_Conversation.wav	1.74	36.77	0.18	2.43
English_male.wav	6.55	24.99	0.22	3.14
English_female.wav	4.72	32.48	0.77	2.83
English_noisy.wav	3.01	41.90	0.16	2.52

TABLE: II
OBJECTIVE QUALITY MEASURES OF CELP - WITH MMSE-STSA ALGORITHM

Speech files	Objective measures			
	SNRseg	*WSS*	*LLR*	*PESQ*
English_Conversation.wav	3.42	26.09	0.15	3.00
English_male.wav	7.94	20.66	0.19	3.22
English_female.wav	7.63	28.6	0.54	3.15
English_noisy.wav	5.07	23.21	0.12	3.41

Different types of sample speech files were used for objective testing. The results of objective evaluation for Segmental SNR (SNRseg), Weighted Spectral Slope (WSS), Log Likelihood Ratio (LLR) and Perceptual Evaluation of Speech Quality (PESQ) are shown in table I to IV.

TABLE: III
OBJECTIVE QUALITY MEASURES OF MELP - WITHOUT MMSE-STSA ALGORITHM

Speech files	Objective measures			
	SNRseg	*WSS*	*LLR*	*PESQ*
English_Conversation.wav	1.33	38.14	0.23	1.35
English_male.wav	2.56	35.83	0.45	2.5
English_female.wav	3.54	43.97	0.95	1.54
English_noisy.wav	2.08	42.50	0.52	2.31

TABLE: IV
OBJECTIVE QUALITY MEASURES OF MELP - WITH MMSE-STSA ALGORITHM

Speech files	Objective measures			
	SNRseg	*WSS*	*LLR*	*PESQ*
English_Conversation.wav	2.44	30.21	0.20	2.51
English_male.wav	4.26	25.5	0.32	4.32
English_female.wav	5.21	30.61	0.71	2.25
English_noisy.wav	3.64	26.31	0.42	2.56

ISSN 2229-5518

It is observed from the tables that, CELP and enhanced CELP have high Segmental SNR compared to MELP and is highly desirable because the high SNRseg value imply better speech quality.

Enhanced CELP has the lowest WSS value, which is highly acceptable. The value of LLR should be closest to zero for a good communication channel and here LLR of CELP is more close to zero than MELP. Higher PESQ value also imply good speech quality.

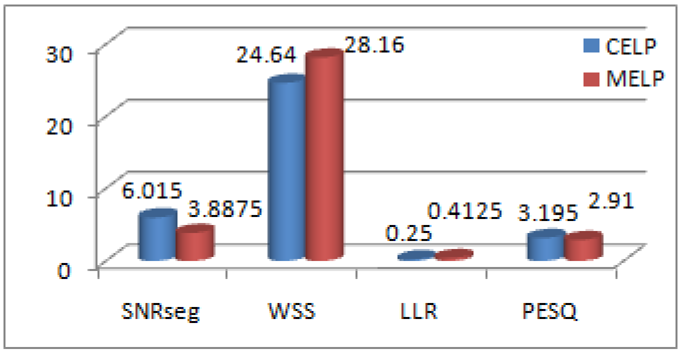


Fig. 9.Objective measures: CELP and MELP with MMSE – STSA Algorithm

4 Conclusion

Software implementation of MMSE - STSA Speech Enhancement Algorithm in MELP and CELP speech coders has been performed. For the applications where focus is on better quality at low bit rate this combined system can be used. Quality testing and waveform analysis have been performed using quality measurement and PRAAT software. From the results we can conclude that CELP coder gives better perceptual quality at the expense of high bit rate. MELP coder can be used with speech enhancement technique at low bit rate with better quality.

References

[1] [1] Wai C. Chau, "Speech coding algorithms: Foundation and evolution of standardized coders", Wiely Interscience 2003.

[2] [2] Hu, Y. and Loizou, P. (2008). "Evaluation of objective quality measures for speech enhancement," IEEE Transactions on Speech and Audio Processing, Vol. 16, Issue 1, pp. 229–238, 2008.

[3] [3] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Transaction on Acoustics, Speech and Signal processing, vol. ASSP 27, no.2, pp. 113-120, April 1979.

[4] [4] 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, December 1984.

[5] [5] Milind Tandel, Vandana Shah and Bhavina Patel, "Implementation of CELP coder to evaluate performance in terms of bit rate, coding delay and quality of speech", Third Int. Conf. on Electronics Computer Technology, vol.6, pp.86-89,2011.

[6] [6] Alan McCree, Kwan Truong, E. B. George, T. P. Barnwell and V. Viswanathan, "A 2.4 kbit/s MELP coder candidate for the new U.S. Federal Standard," IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, pp. 200–203, 1996.

[7] [7] Schroeder, Manfred and Atel, "Code Excited Linear Prediction (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.

[8] [8] Ancy S. Anselam and Sakuntala S. Pillai, "Performance evaluation of Code Excited Linear Prediction coders at various bit rates", IEEE Int. Conf. on Computation, Power, Energy, Information and Communication, 2014.

[9] [9] A. V. McCree and T.P. Barnwell, "Improving the performance of a mixed excitation LPC vocoder in acoustic noise", in Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, USA,vol.2,pp.137-140, March 1992.