iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 Issue: XII Month of publication: December 2017

DOI:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

Non-Stationary Noises Suppression through an Adaptive EMD Combined with DCT for Speech Enhancement

Mr. S. Nageswara Rao[1], Dr. K. Jaya Sankar[2], Dr. C.D. Naidu[3]
[1]*Assistant Professor, Department of ECE, Sri Venkateswara Engineering College, Suryapet, Telangana, India*
[2]*Prof. & Head, Department of ECE, Vasavi College of Engineering, Hyderabad, India*
[3]*Professor, Department of ECE, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India*

Abstract: Suppression of non-stationary noise from a noisy speech is very important for speech oriented applications like speech recognition, speech communication etc. This paper proposes a new speech enhancement approach to suppress the non-stationary noises like Babble noise, car noise etc., based on Discrete Cosine Transform and Empirical Mode Decomposition. The complete accomplishment of proposed approach is carried out in two phases, DCT based soft thres holding and Adaptive Empirical mode Decomposition based denoising. Initially the noise speech is filtered through DCT based soft thres holding and the partially enhanced speech is processed through Adaptive Empirical mode Decomposition to reduce the residual noise further. An adaptive noise cancellation procedure is accomplished which considers the correlation between the noise and the reference signal to filter the noise. The extensive simulations are carried out over the proposed approach through different speech signals with different noises at different signal strengths and the performance is evaluated through Output SNR, Output Avg Seg SNR and PESQ test.
Keywords: Speech enhancement, Discrete Cosine Transform, Empirical Mode Decomposition, Babble Noise, Output SNR, PESQ.

I. INTRODUCTION

Speech enhancement is very important in many speech related applications like speech recognition, mobile communications, coding and communication applications. The main aspect of any speech enhancement technique is to filter the noise from a noise contaminated speech signal to improve the intelligibility and perceptual quality of a signal, through noise reduction algorithms [1]. Due to the substantial improvement of this process in today's information technology, there have been so many approaches proposed for speech enhancement. A common problem in all existing algorithms is that the recovered speech signal has lot of distortions when compared it with original speech signal, resulting in the loss of some speech details. Thresholding [2], [3] is one of the powerful approach, removes the noise components from a noise contaminated speech by subtracting a constant value from every sample coefficients obtained after the analysis of transformation. Though, the subtraction is a simple approach, the direct subtraction of coefficients results in the speech components degradation. Instead of direct subtraction criteria, in [4], a new soft thresholding based strategy was proposed based on frequency frames. This thresholding based speech enhancement is able to remove noise components with a slight damage to the speech signal. This approach is also able to process signals with high SNRs. Though, this approach effectively removes the noise components from a noise contaminated speech, a significant amount of noise still remains in the enhanced speech signal. One more disadvantage with this algorithm is that this is not robust for various types of noises. In [5], a hybrid approach was proposed by combining Discrete Cosine Transform (DCT) with Empirical Mode Decomposition. [5] Divides a signal into speech dominant and noise dominant and then applies linear thresholding for both segments to suppress the noise. The complete process was carried out in two stages, DCT stage and EMD stage. The residue noise present in the enhanced speech after DCT will be filtered by EMD. The main drawback of [5], due to the application of linear thresholding in EMD stage, some speech samples are also missed, reduces the quality and intelligibility of speech. The linear thresholding filters the noise samples based on a constant noise variance. The noise variance is varying in nature for non-stationary signals. By deriving a constant noise variance, the speech samples are also getting suppressed.

In this paper, an adaptive speech enhancement technique is proposed to suppress the non-stationary noises from the noise contaminated speech signal. Similar to the conventional this approach also combines the DCT and EMD and named s Hybrid DCT and Adaptive EMD (HDAEMD). The complete suppression is carried out in two phases, DCT based filtering and EMD based filtering. In the first phase, DCT based filtering, the suppression of noises is accomplishes through the extended DCT, i.e., the noise is filtered from the DCT coefficients of signal through an adaptive thresholding. Further in the second phase, the first phase output is processed through Adaptive Empirical Mode Decomposition (AEMD) to remove the residual noise. Here the proposed AEMD accomplishes an extra noise cancellation mechanism to filter the noise. Various speech signals with various signal strengths are considered for performance evaluation.

Rest of the paper is organized as follows: section II illustrates the details of related work. Section III illustrates the details of conventional approach. The complete details about the proposed approach are represented in section IV. Simulation results are shown in section V and finally the conclusions are given in section VI.

II. RELATED WORK

Several approaches have already been proposed to improve the speech enhancement results. The spectral subtraction is one of the early methods to reduce the noise effects from the observed speech signals. In this method, the noise reduction is achieved by appropriate adjustment of the set of spectral magnitudes [6]. Its basic requirement is the noise spectrum which is determined from the non-speech segments [7]. In such single channel speech enhancement system, the residual noise is a usual issue. It decreases the speech intelligibility and hence further processing is required to reduce the residual noise.DWT [8]-[10] was the main approach in the speech enhancement methods under time-frequency domain category.DWT is more appropriate technique to reduce various types of non-linear and non-stationary noises that corrupt the clean speech. DWT is superior alternative noise reducing techniques for the methods based on short time Fourier transform (STFT). The main challenge in DWT based denoising techniques is to estimate a threshold value that provides a perfect distinction between wavelet coefficients of the desired speech and also that of wavelet coefficients of noise. Then, by using the derived threshold, design of a denoising scheme to minimize the effect of wavelet coefficients of noise to the wavelet coefficients of speech is anther task. In fact, the conventional DWT based de-noising works effectively at relatively high Signal-to-noise (SNR) ratio. In [2], a universal thresholding based speech denoising approach was proposed by for a speech signal contaminated with zero-mean normally distributed white noise. For a noise contaminated speech, applying a common threshold for all DWT coefficients irrespective of silence segments and speech segments will suppress noise to some extent, but also removes silence segments, reducing the intelligibility and quality of speech. In [10], a new thresholding approach was proposed based on statistical modeling, where the new threshold was derived by measuring the distance similarities between the probability distributions of signals. The thresholding was applied in time domain because, there may be no existence of speech in all samples, and this will be larger for the case of without speech and is small for the case of with speech. This approach was succeeded in eliminating maximum noise, still maintaining speech intelligibility. However, the approach in [10] needs a noise variance estimation to find the noise frames from that of the speech frames with knowledge to derive two different thresholds for them. So, to derive a new time adaptive threshold with a view of silence and speech frames, some signal energy over time is necessary. Recently, a new Empirical mode decomposition (EMD) based approaches were proposed in [11-15] for signal enhancement and noise suppression. These methods also included single –channel speech enhancement in stationary fractional Gaussian environments (FGN) [16]. Even though number of attempts was made in order to enhance the way of understanding of EMD operation on speech enhancement, it still lacks a perfect mathematical theory and an essential description about the algorithm. In [17], an adaptive soft thresholding algorithm was developed based on EMD. Here the variance is derived for every IMF not for speech signal by which this algorithm achieved better performance. However the main drawback of this approach is in finding the speechless part to determine the variance of noise. In [18] a new speech enhancement approach is developed based on the EMD and Hurst component to suppress the white noise as well non-stationary acoustic noises. The main contribution is focused for the adaption of Hurst exponent in the selection of IMFs to reconstruct the speech. EMD is used as a pre-processing filter to decompose the noisy speech signal into IMFs. However it is not considering any signal or noise variance in the speech enhancement. Further a combined speech enhancement algorithm was developed in [19] to suppress the non-stationary noises. It employs the combination of variational mode decomposition (VMD) [20, 21] and empirical mode decomposition (EMD) methods. This approach tried to filter the noise components efficiently in both low frequencies and high frequencies.

III. CONVENTIONAL APPROACH [5]

In [5], a hybrid was proposed algorithm which will include a two-stage soft thresholding. In the first stage, baseband approach DCT domain soft thresholding is adapted to the noisy speech. The remaining noise in the enhanced speech looks like random tones and results in an irritating sound. Hence further denoising should be applied to get ridof this artifact. However, it is not an easy task to identify and remove these noise components without degrading the speech signal. Due to the frequency characteristics of the IMFs, further enhancement is achieved in the second stage through an EMD based soft thresholding strategy.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

A. *DCT based De-noising*

In this stage, the complete noise speech is divided into two segments, signal dominant segment and noise dominant segment. Then the linear thresholding was applied on the noise dominant segments and signal dominant segments are left as it is. Because, the thresholding of signal dominant segments will suppress the desired speech samples. The noisy speech is first segmented into frames and then DCT is applied on each frame. The DCT coefficients of the frames are further divided into frequency bins, each containing DCT coefficients. For adaptive thresholding, each bin is categorized as either signal- or noise-dominant. The classification pertains to the average noise power associated with that particular bin. If the i_{th} bin satisfies the following inequality:

$$\frac{1}{N}\sum_{k=1}^{N}\left|X_i^k\right|^2 \geq \sigma_n^2 \qquad (1)$$

Where σ_n^2 denotes the variance of the noise,X_i^k is the k_{th}DCT coefficient of the i_{th} frequency bin, and N is the number DCT coefficients of the bin; then the bin is characterized as signal-dominant, otherwise as noise-dominant. The signal dominant bins are not thresholded, since it is highly possible to degrade the speech signal, especially for high SNRs. In the case of a noise-dominant frequency bin, the absolute values of the DCT coefficients are sorted in ascending order and ali near thresholding is applied:

$$\widehat{X_k} = sign(X_k)[\max\{0, (|X_k| - n_j)\}] \qquad (2)$$

Where $n_j$is the linear threshold function obtained as

$$n_j = j\frac{\lambda\sigma_n N}{\sum_{k=1}^{N}k^2} \qquad (3)$$

Where $j$is the index of sorted$|X_k|$. The value of λ will be calculated by dividing the variance of particular frequency bin by the total variance and it will vary from bin to bin.

B. *EMD based De-noising*

A significant amount of the noise components is reduced in the first stage. How ever, there is still remaining noise from both the thresholded noise dominant and un-thresholded signal-dominant frequency bins. It is possible to extract a considerable amount of this residual noise in the second stage from the IMFs of the enhanced speech. Due to the frequency characteristics of EMD, the noise and speech signals mostly dominate indifferent IMFs. The IMFs are in time domain and may have frequency overlaps. However, at any time instant, the instantaneous frequency represented by each IMF is different. That is why, although the IMFs are in time domain, they have spectral difference at time instances. Therefore, the DCT soft thresholding algorithm can be applied to the IMFs as given in [22]. First, the EMD is applied to the enhanced speech. The obtained IMFs are divided into frames. Due to the decomposition characteristics, the IMFs differ in terms of noise and speech energy distribution. Therefore the specific noise variance of each IMF is estimated from the speechless parts. As, in the DCT bin categorization case, the frames are characterized as either signal- or noise-dominant frames with the novel categorization limit given in (1). The noise dominant frames are thresholded using (2), while the signal dominant frames are not.

IV. PROPOSED APPROACH

In this approach, an adaptive hybrid speech enhancement algorithm was proposed by combining the DCT and Adaptive Empirical Mode Decomposition (AEMD). In this approach, a new threshold was derived for DCT and the EMD filtering was also done in an adaptive way. The complete process was carried out in two stages, DCT and AEMD. Figure.1 shows the overall block diagram of proposed approach.

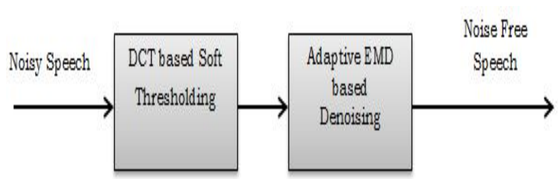


Figure.1. Overall block diagram of proposed approach

A. *DCT based De-noising*

The major problem with the proposed DCT based de-noising in [5] is that is not able to de-noise all types of noises. Since all the frequency in the time domain, the algorithm is mainly applicable to white noise which has a flat spectrum. The method fails for other noise types that show different spectral distribution within the frequency bins. Therefore, it is important to have a sub Band approach where a specific noise varianceis calculated for each frequency band. Here a new mechanism is proposed based on DCT and soft thresholding and the block diagram is shown in figure.2.Here the index of the frequency bins represents the index of the sub band. For instance, the first frequency sub band consists of the first frequency bins of each frame. The variance of each sub band is calculated through a minimum statistics approach from the frequency bins. With this sub band approach, each band will have an effective bin categorization. Therefore, the algorithm will be robust to different noise types. Apart from the sub band approach, a novel strategy is introduced here for the bin categorization. The limit given in(1),which is set to noise variance, is not efficient to identify all the noise-dominant bins. Since the variance of the noisy bins will have fluctuations, there will be many noise-dominant bins which will be identified as signal-dominant. Therefore, the limit for bin categorization should have a larger value than the noise variance, in order to guarantee that all the noisy binsare thresholded. A novel limit relies on the idea that a bin can be defined as noise-dominant, if the noise power in that binis higher than the speech power.

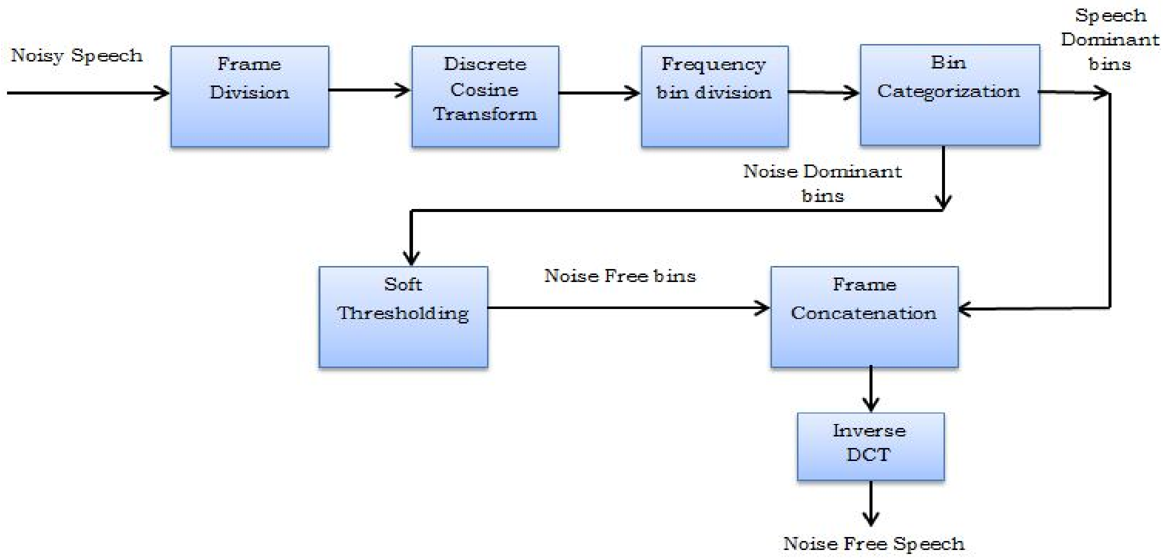


Figure.2 Block diagram of DCT based soft thresholding

Therefore, the limit should be set to the case where the noise and speech variances $\sigma_n^2$and σ_s^2 , respectively, are equal. The variance $\sigma^2$of the noisecontaminated speech for any frequency bin is represented as

$$\sigma^2 = \sigma_s^2 + \sigma_n^2 + 2\varphi(s,n) \qquad (4)$$

Where $\varphi(s,n)$is the covariance term of signal and noise. Ifthe signal and noise are independent, the covariance function gives zero; thus we have

$$\sigma^2 = \sigma_s^2 + \sigma_n^2 \qquad (5)$$

For frame categorization (into signal- and noise-dominant frames), the threshold is considered with equal noise and speech power, and hence $\sigma^2 = 2\sigma_n^2$.Therefore, in case of equalnoise and speech power, the variance of the bin is equal to$2\sigma_n^2$. The variance of a speech segment directly correspondsto its power.The equal variance of speech and noise exhibitsthe equilibrium contribution of speech noise power to thenoisy speech frame. Hence such level of power is considered as the threshold for speech frame categorization. It is treated as the minimum power level of noise-free speech frame. Any frame with power higher than such threshold exhibits that the speech power is dominating. Otherwise, the noise power dominates the analyzing frame. That is why the limit for the categorization of the bins in (1) should be set to this value. With the proposed strategy, if

$$\frac{1}{N}\sum_{k=1}^{N}\left|X_i^k\right|^2 \geq 2\sigma_n^2 \qquad (6)$$

Where σ_n^2 denotes the variance of the noise for the i_{th}sub band and $X_i^k$is the k'$_{\text{th}}$ sample of the i_{th} bin, then this bin is categorized as signal-dominant, otherwise as noise dominant. Noise-dominant frequency bins are thresholded asin (2). This approach can be further improved by finding an optimum value for λ. As per [5], the value of λ depends on the input SNR. The higher value of λ is optimal for low SNRs and the lower value of λ is better for high SNRs. Thus, the λ value greatly depends on the SNR. The input SNR can be estimated as

$$SNR_i = 10log\left(\frac{\sigma_s^2}{\sigma_n^2}\right) \qquad (7)$$

For each and every frequency bin, the noise variance and signal variance are varying, thus the input SNR is also varying. From (7), the optimal value of λ can be formulated as

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

$$\lambda \propto SNR_i (8)$$

$$\lambda = \alpha SNR_i \ (9)$$

Where $\alpha$is an arbitrary constant and the range was evaluated as $(0.59 < \alpha < 0.77)$ from simulations.

B. AEMD based De-noising

In the first stage, a significant amount noise was reduced and the remaining residual noise present in the noise-dominant and signal dominant bins will be reduced using AEMD. This AEMD use an adaptive noise cancellation technique based on the correlation of noise variance. Due to the decomposition characteristics of EMD, the obtained IMFs having different variances. The block diagram of AEMD is shown in figure.3.

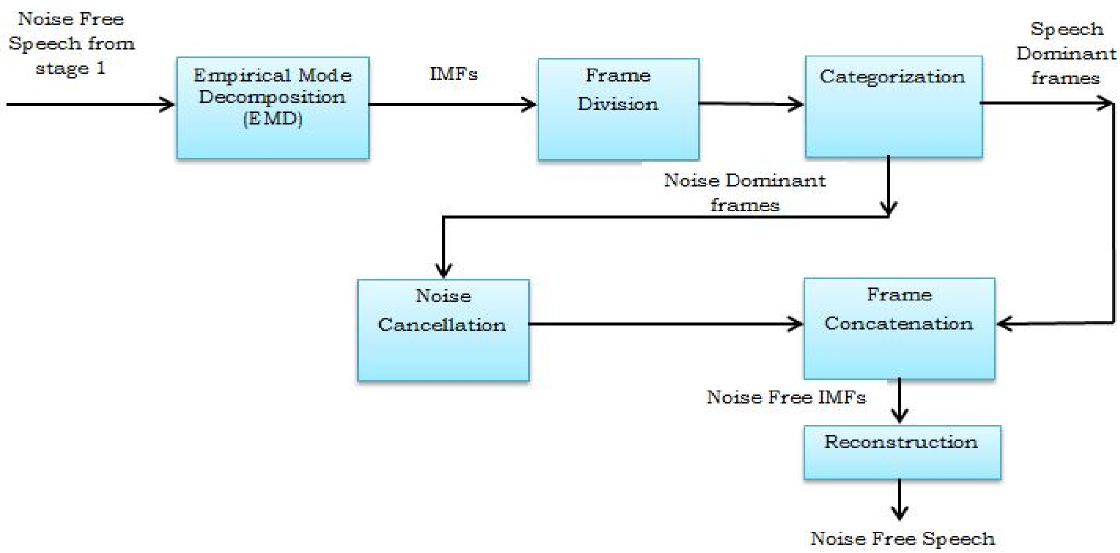


Fig.3. Block diagram of AEMD based de-noising

In Fig.3, first the EMD is applied to enhanced speech, output of first stage. For the obtained IMFs, categorization is done through (6) to separate noise-dominant IMFs and signal-dominant IMFs. In general, the maximum noise present in the initial IMFs, thus the initial IMFs are noise-dominant IMFs. Then these noise-dominant IMFs are subjected to noise cancellation through an adaptive noise cancellation technique according to the Fig.4. The output of noise cancellation block is an enhanced IMF. For each and every noise dominant IMF, the noise cancellation block will be applied to suppress the noise.

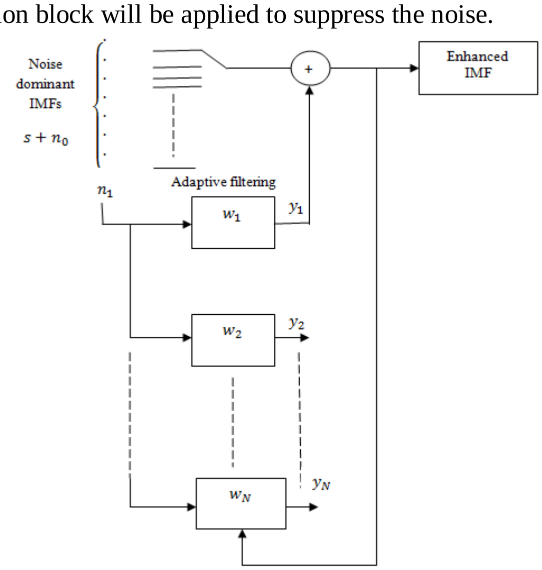


Fig.4. Noise cancellation block

The new EMD based adaptive technique for speech enhancement(AEMD) is illustrated in Fig. 4, where ‘s’ is the original speech, ‘n_0’ is the contaminating noise and ‘n_1’ is the reference noise signal. The input noise-dominant IMFs contain the noise in dominant manner and the signal in less and are adaptively filtered using the noise reference. The obtained IMF at the output of noise cancellation block are combined with signal dominant IMFs to reconstruct the enhanced speech. Finally the noise filtered IMFs and signal dominant IMFS are concatenated and applied for reconstruction to obtain final enhanced speech.

C. Analysis

It is assumed in Fig.2 that the contaminating noise is correlated to the reference and the original signal is uncorrelated with these noises. Therefore, the k_{th} IMF of the noisy signal can be described below by:

$$I_k = s_k + n_0 \quad (10)$$

The error signal e can be obtained by subtracting the received IMF from the noise speech, as

$$e = I_k - y$$

$$E[e^2] = E[(I_k - y)^2]$$

$$E[e^2] = E[((s_k + n_0) - y)^2]$$

$$E[e^2] = E[((s_k) + (n_0 - y))^2]$$

$$E[e^2] = E[(s_k)^2] + E[(n_0 - y)^2] + 2E[s_k(n_0 - y)] \qquad (11)$$

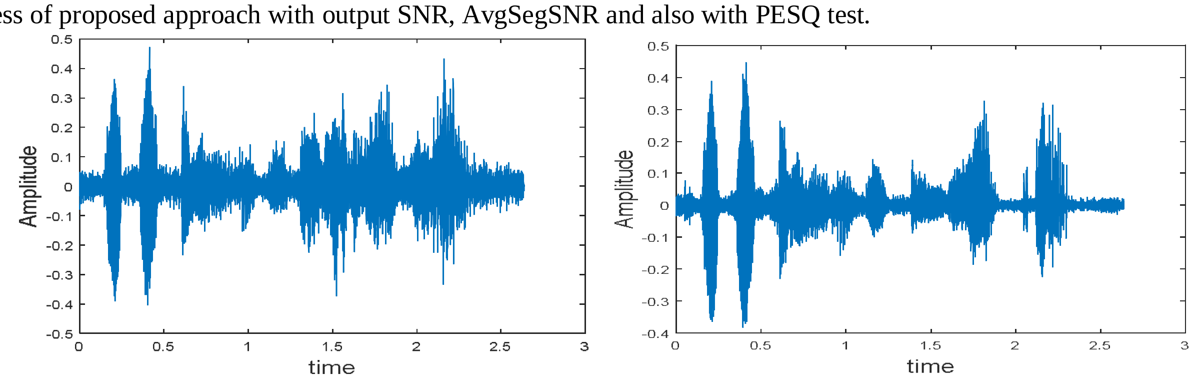
As discussed above, there is no correlation for between the reference signal and noise but the correlation exists between reference noise and contaminated noise, thus the equation. (11) becomes

$$E[e^2] = E[(s_k)^2] + E[(n_0 - y)^2] \qquad (12)$$

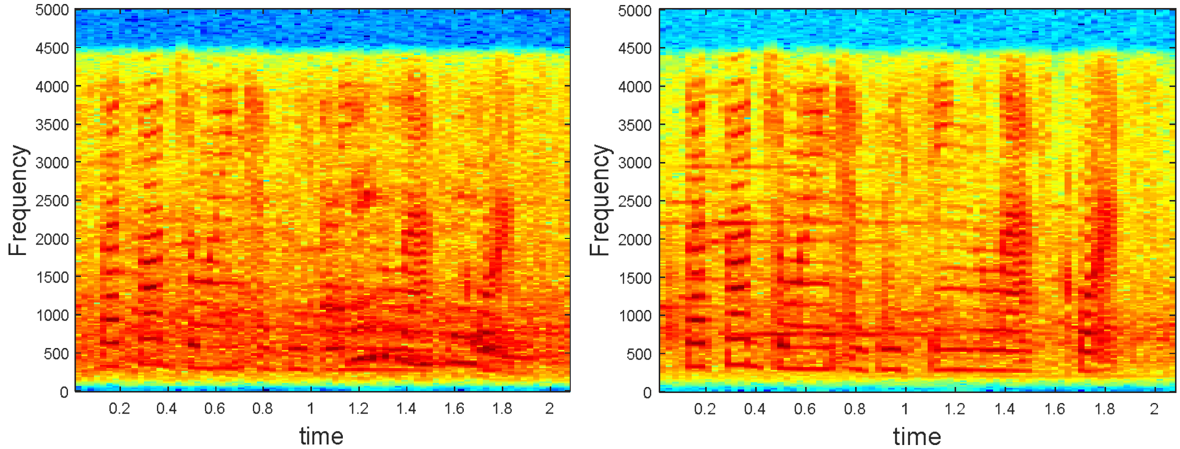
Where the speech component in that IMF is unaffected and the other term is minimized to lower the MSE signal. The noise will be adaptively filtered to feed into the system, to produce an error signal with which it is uncorrelated. Hence, the error signal will be an enhanced version of the k_{th} IMF. In this way, all the noise-dominated IMFs are subjected to noise cancellation to produce a noise free IMF, enhanced IMF. Finally, the enhanced IMFs obtained from noise cancellation block and the signal dominant IMFs are combined to reconstruct an enhanced speech.

V. SIMULATION RESULTS

This section describes the performance evaluation details of the proposed approach. To illustrate the effectiveness of the proposed Hybrid DCT & AEMD (HDAEMD) speech enhancement algorithm, extensive computer simulations were carried out with 10 male and 10 female utterances, randomly selected from the TIMIT database. To evaluate the performance of proposed approach, Overall SNR and average segmental SNR improvements are measured. The quality of the enhanced speech signal through the proposed approach was evaluated through Perceptuale valuation of speech quality (PESQ) test. To show the robustness of proposed approach, various noise types like Babble noise, Car-interior Noise, Restaurant noise, Train Noise and Airport Noise are used at various signal strengths. The proposed approach also compared with earlier Soft DCT [22], AEMD [23] and HDEMD [5] to illustrate the effectiveness of proposed approach with output SNR, AvgSegSNR and also with PESQ test.

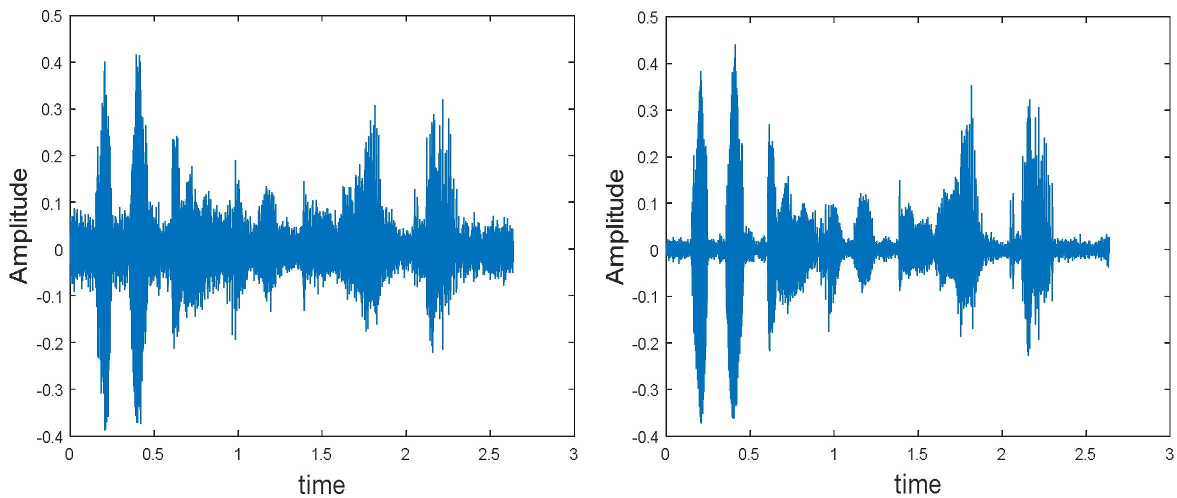


(a) (b)

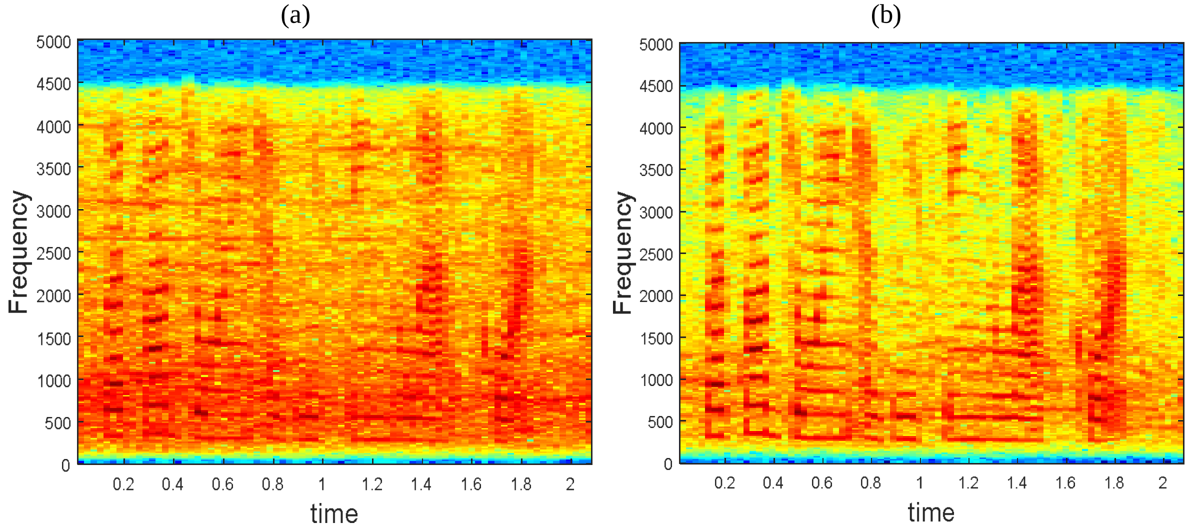


(c) (d)

Figure.5 (a) Clean female speech with airport noise at 0dB, (b) Enhanced speech of (a), (c) Spectrogram of noisy speech, (d) Spectrogram of enhanced speech



(a) (b)



(c) (d)

Figure.6 (a) Clean female speech with airport noise at 5dB, (b) Enhanced speech of (a), (c) Spectrogram of noisy speech, (d) Spectrogram of enhanced speech

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

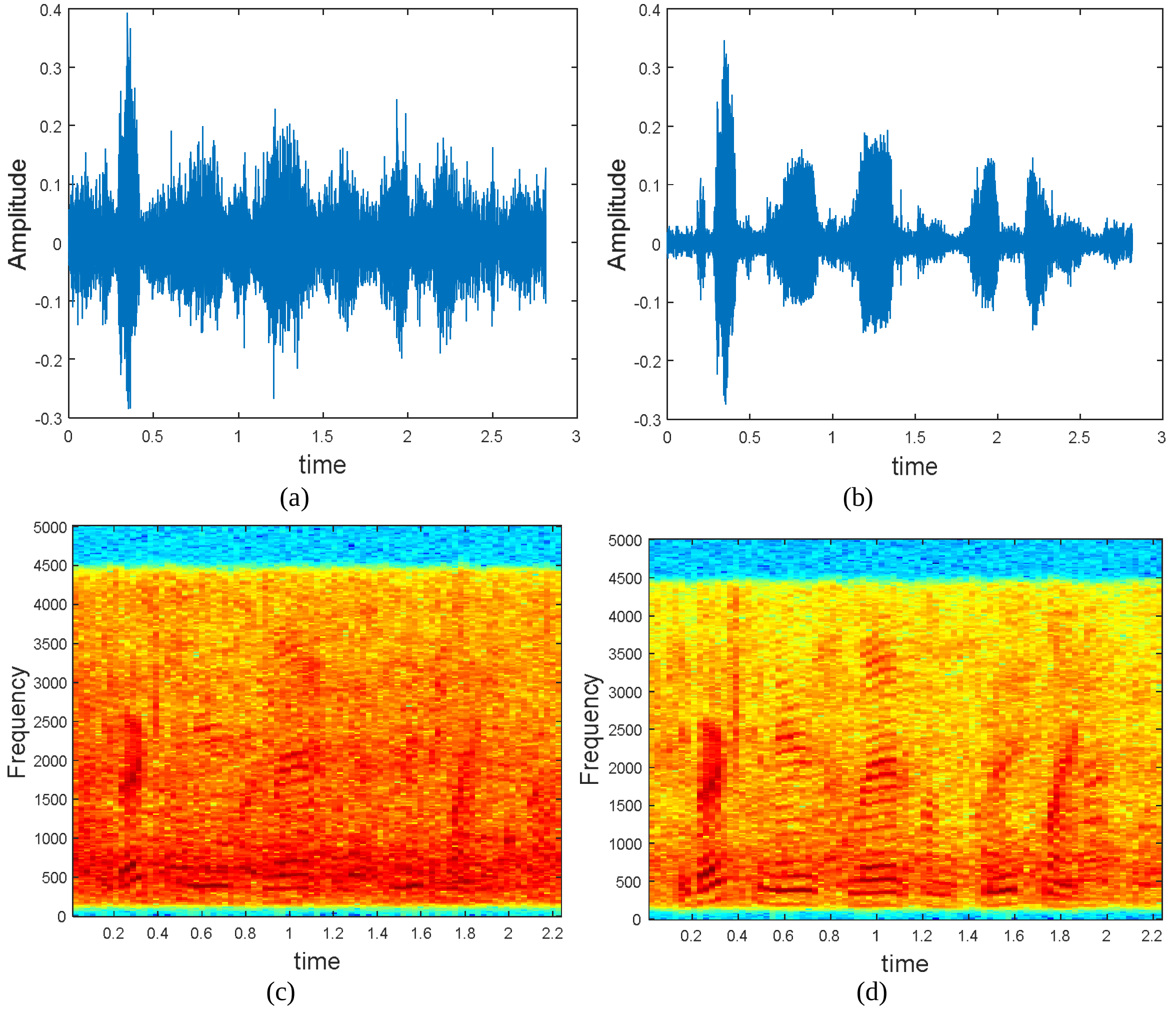
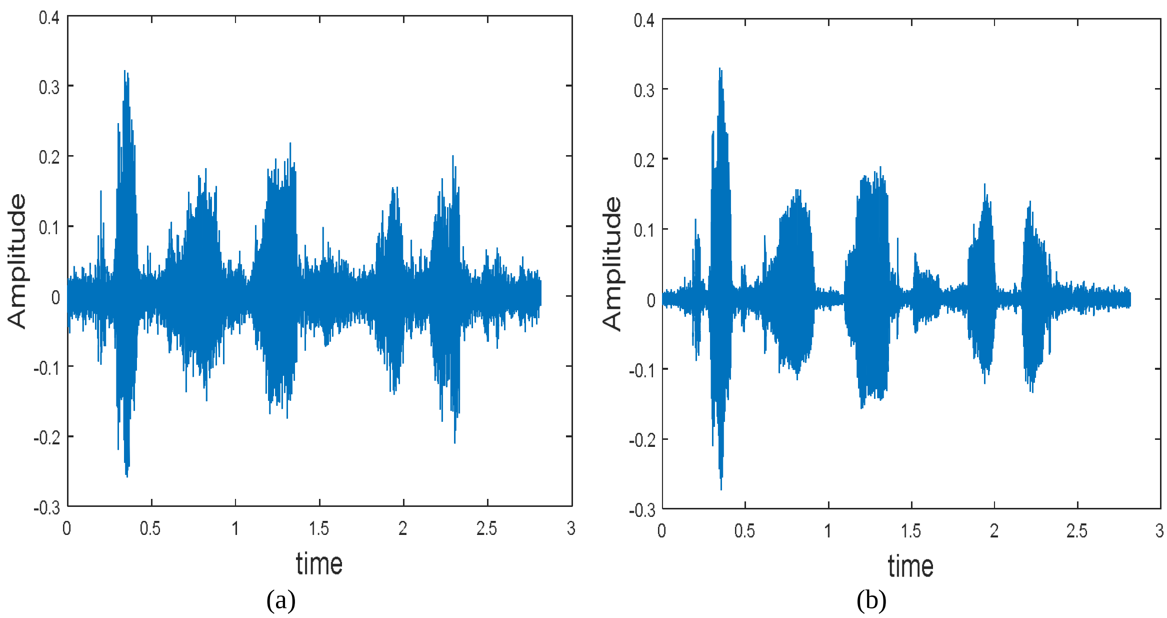


Figure.7 (a) Clean male speech with Babble noise at 0dB, (b) Enhanced speech of (a), (c) Spectrogram of noisy speech, (d) Spectrogram of enhanced speech



International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

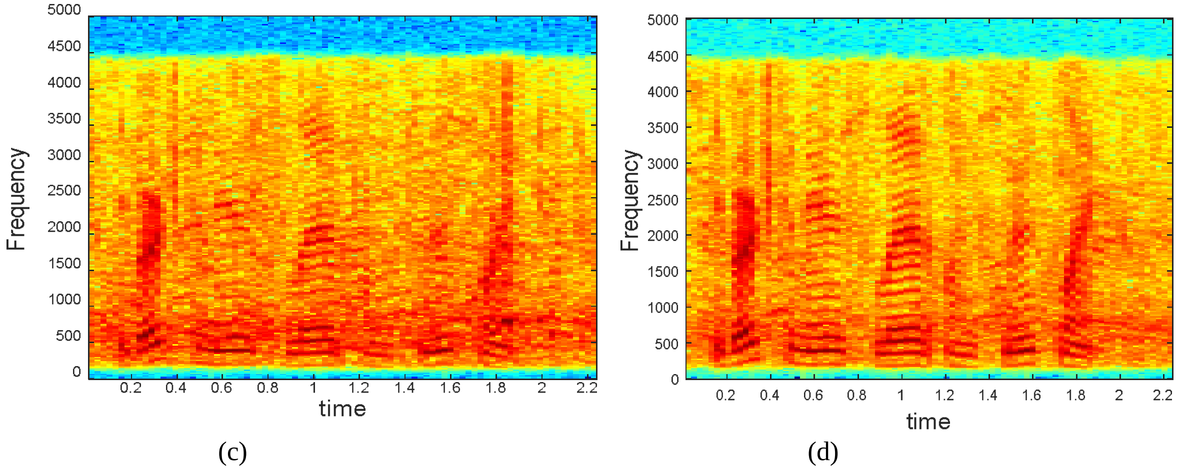


Figure.8 (a) Clean male speech with Babble noise at 5dB, (b) Enhanced speech of (a), (c) Spectrogram of noisy speech, (d) Spectrogram of enhanced speech

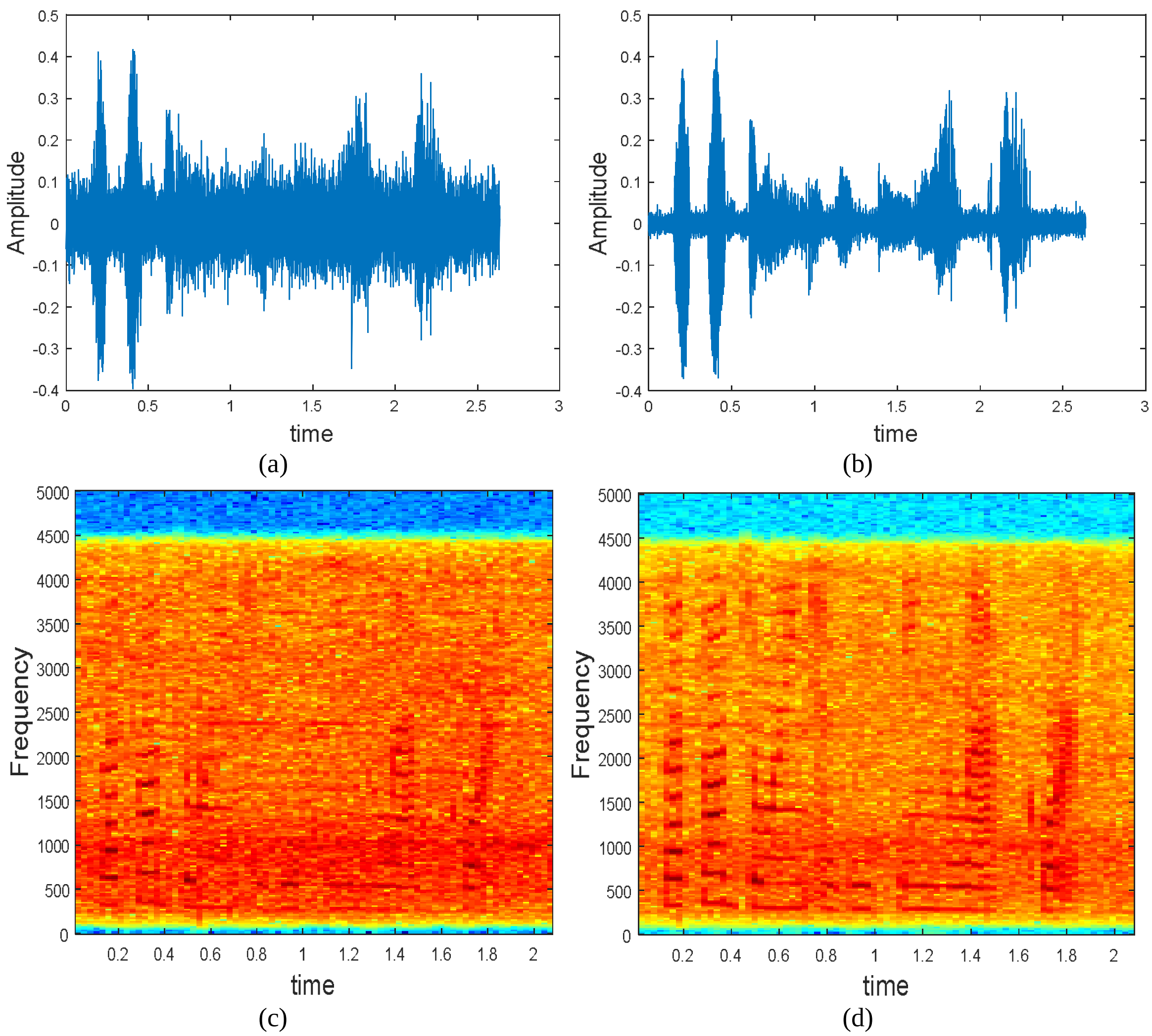
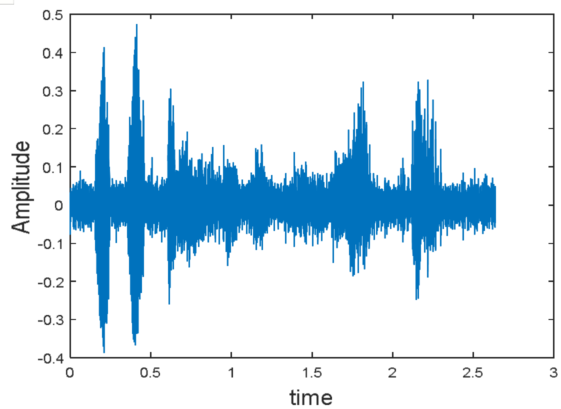
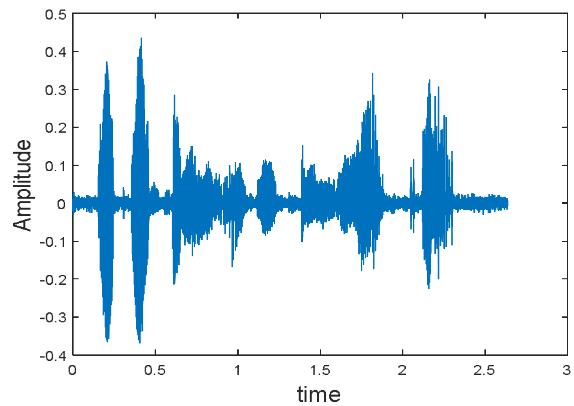


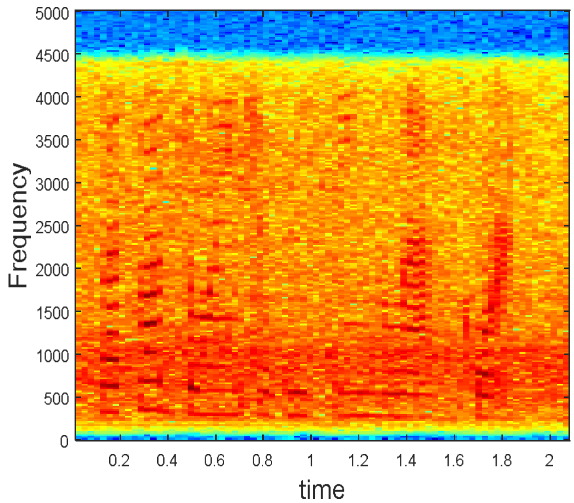
Figure.9 (a) Clean female speech with Car-Interior noise at 0dB, (b) Enhanced speech of (a), (c) Spectrogram of noisy speech, (d) Spectrogram of enhanced speech



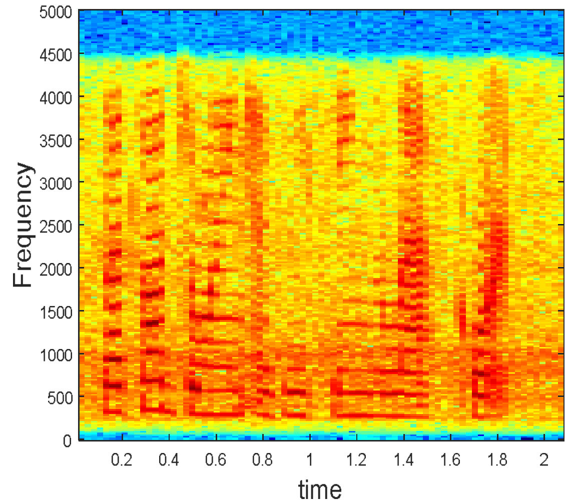
(a)



(b)

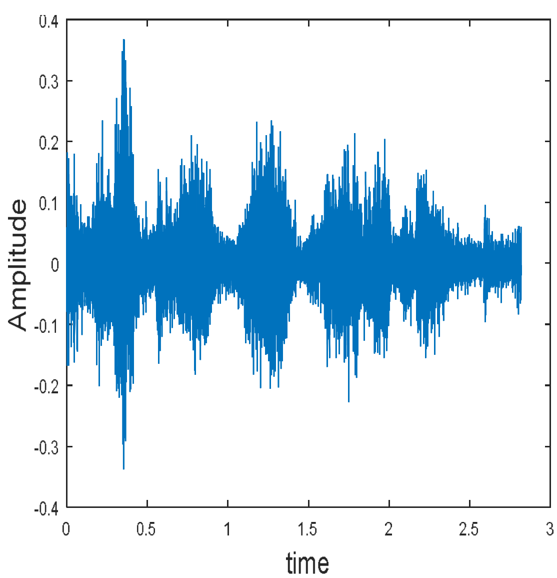


(c)

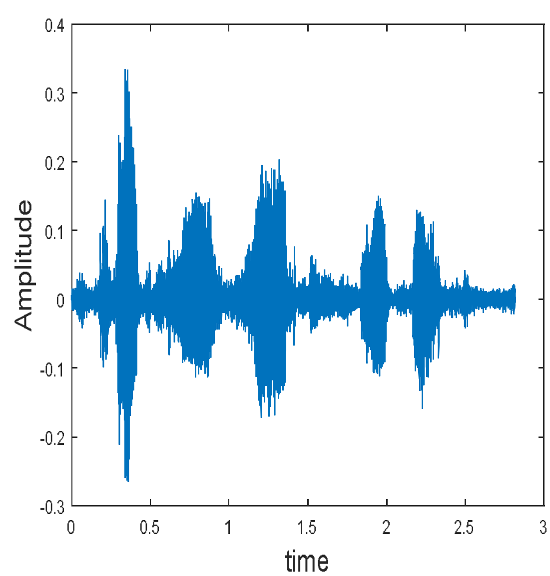


(d)

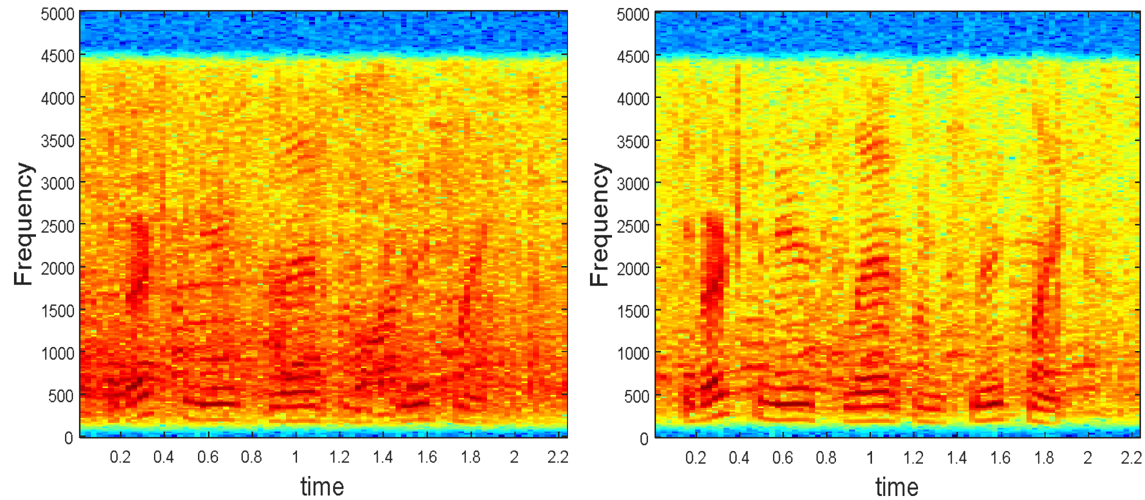
Figure.10 (a) Clean female speech with Car-Interior noise at 5dB, (b) Enhanced speech of (a), (c) Spectrogram of noisy speech, (d) Spectrogram of enhanced speech



(a)

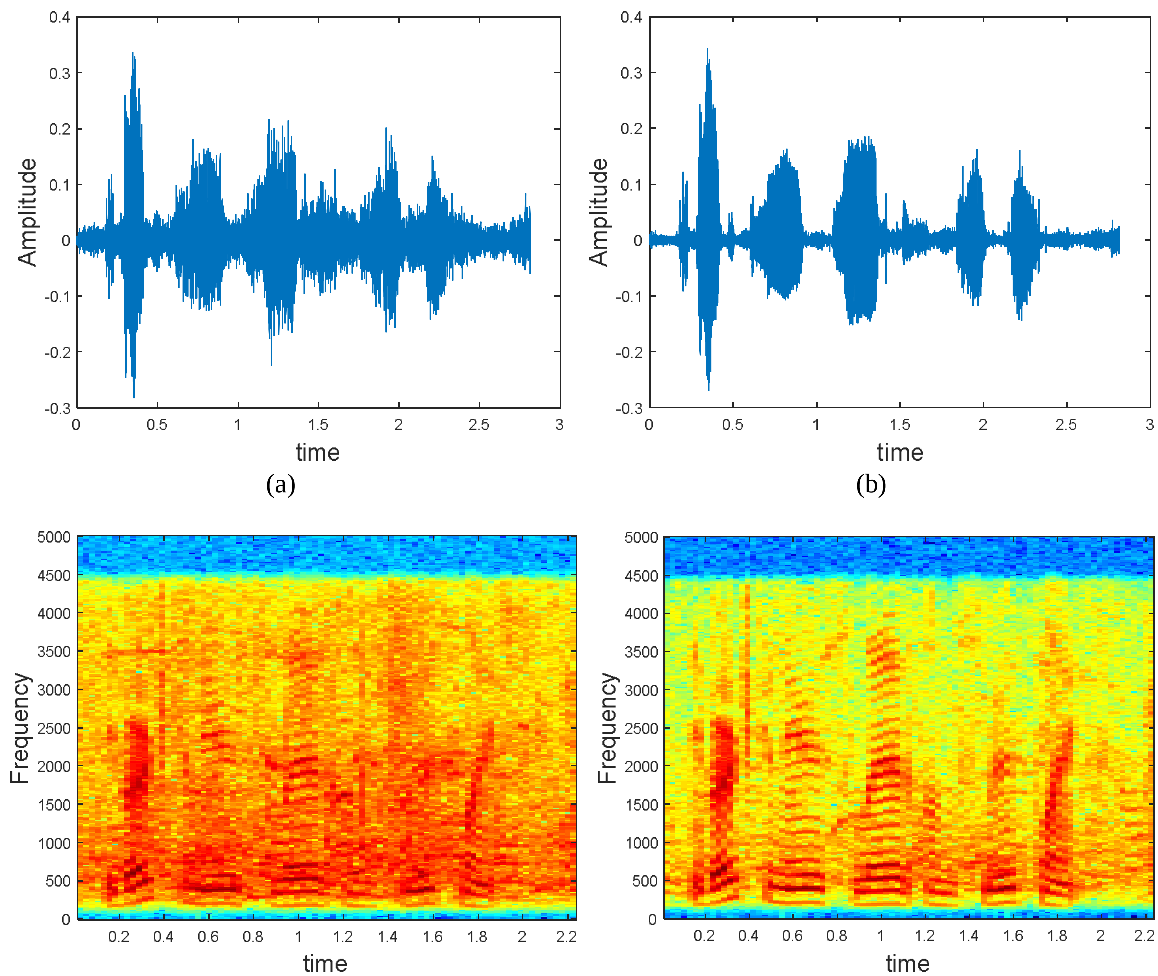


(b)



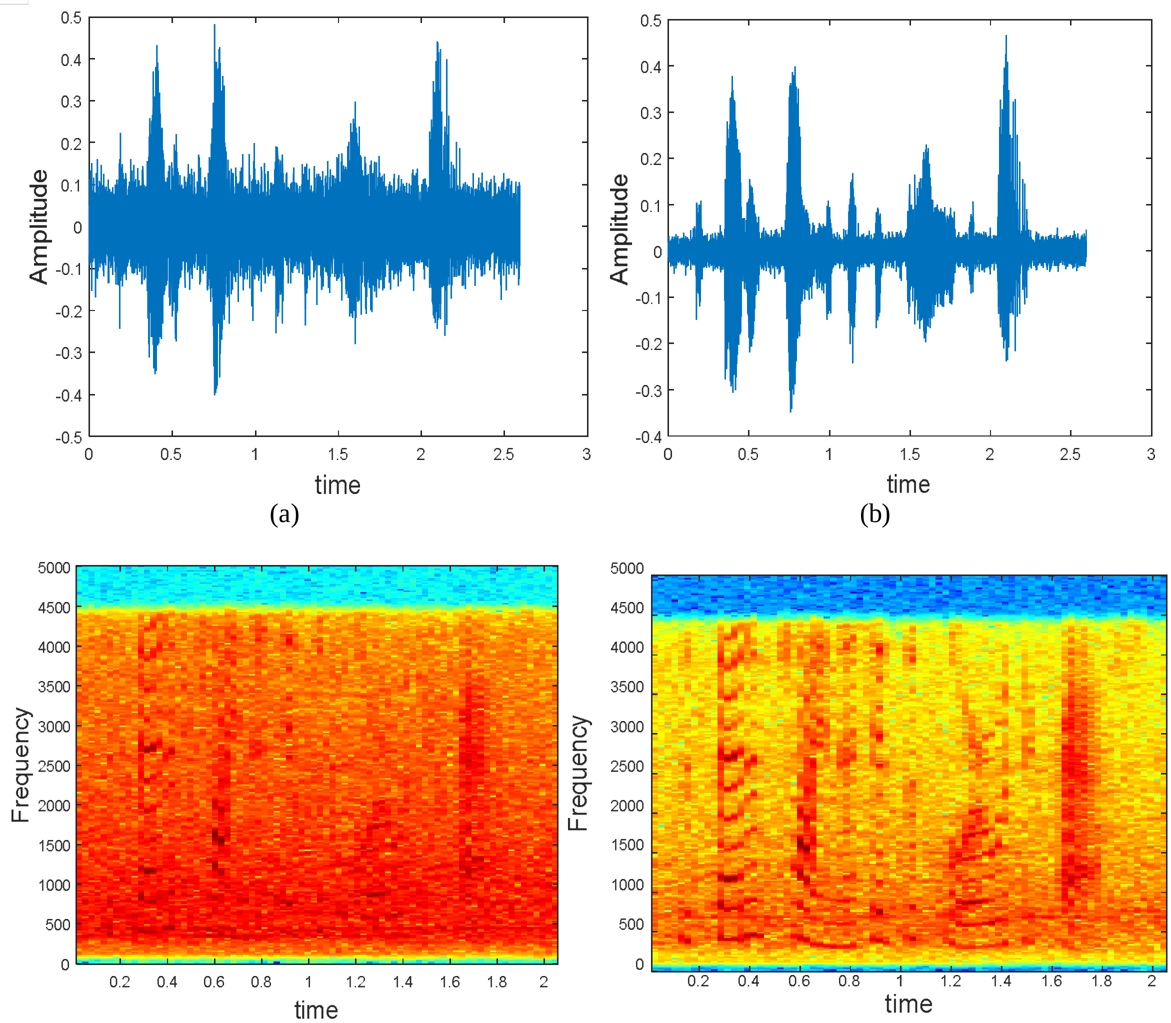
(c) (d)

Figure.11 (a) Clean male speech with Restaurant noise at 0dB, (b) Enhanced speech of (a), (c) Spectrogram of noisy speech, (d) Spectrogram of enhanced speech



(c) (d)

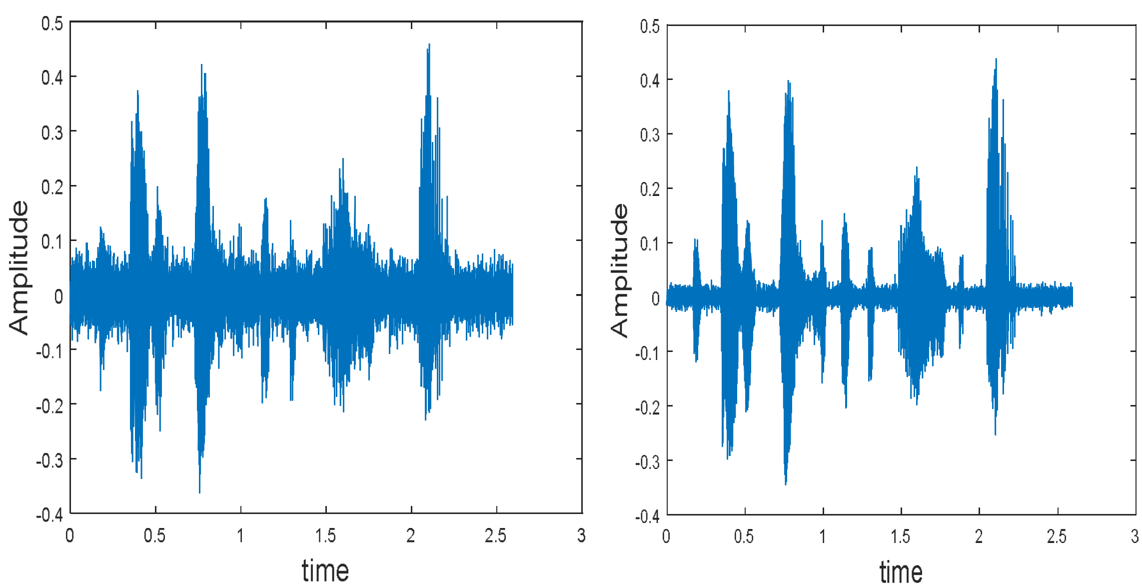
Figure.12 (a) Clean male speech with Restaurant noise at 5dB, (b) Enhanced speech of (a), (c) Spectrogram of noisy speech, (d) Spectrogram of enhanced speech



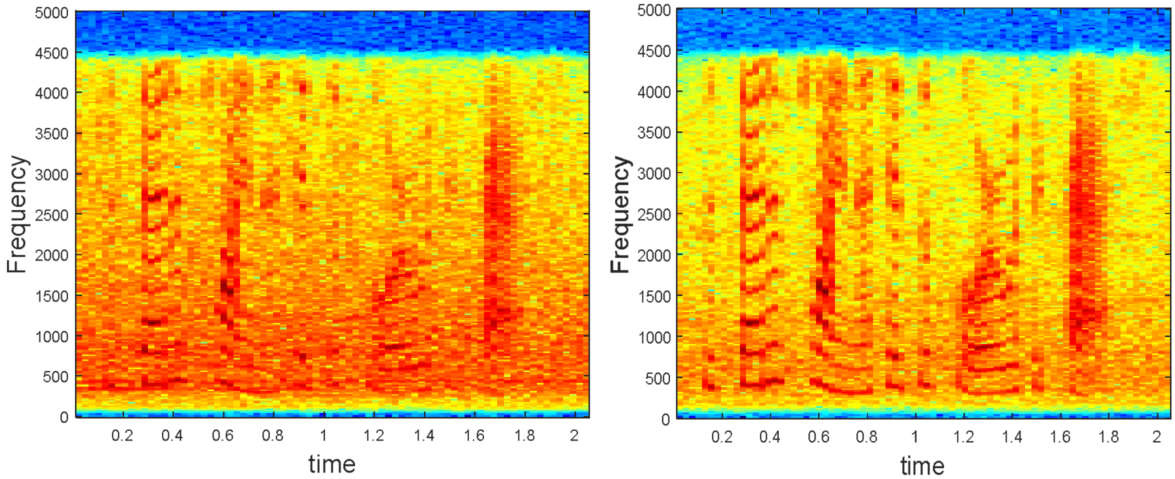
(a) (b)

(c) (d)

Figure.13 (a) Clean female speech with Train Station noise at 0dB, (b) Enhanced speech of (a), (c) Spectrogram of noisy speech, (d) Spectrogram of enhanced speech



(a) (b)



(c) (d)

Figure.14 (a) Clean female speech with Train Station noise at 5dB, (b) Enhanced speech of (a), (c) Spectrogram of noisy speech, (d) Spectrogram of enhanced speech

Figures.5-14(a) represents the clean speech signals contaminated with Airport noise, Babble Noise, Car-interior noise, Restaurant Noise andTrain station noise respectively. The respective enhanced speech signals are shown in figures.5-14(b). Further the spectrogram representation of both noisy and a noise free signal is represented in (c) and (d) respectively. Here the proposed approach is tested through different types of noises with different signal strengths. On an average, this simulation considered 10 male and 10 female samples at different signal strengths like 0dB, 5dB, 10dB and 15dB and the performance is measured the performance metrics and are represented in the following figures.

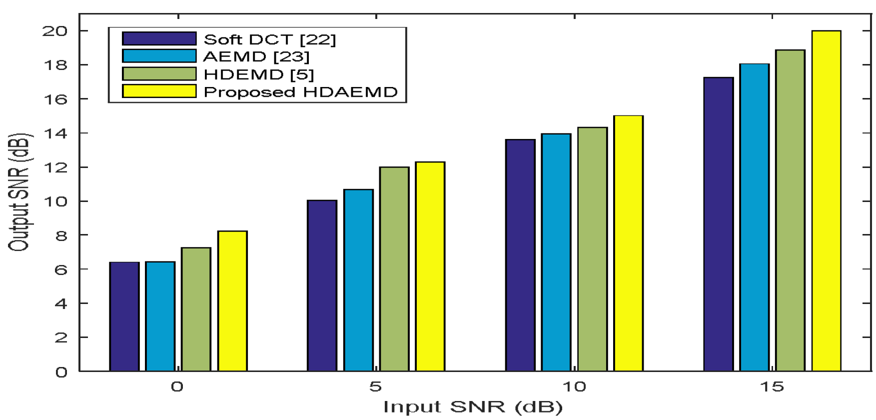


Figure.15 Average output SNR for varying inout SNR for different methods

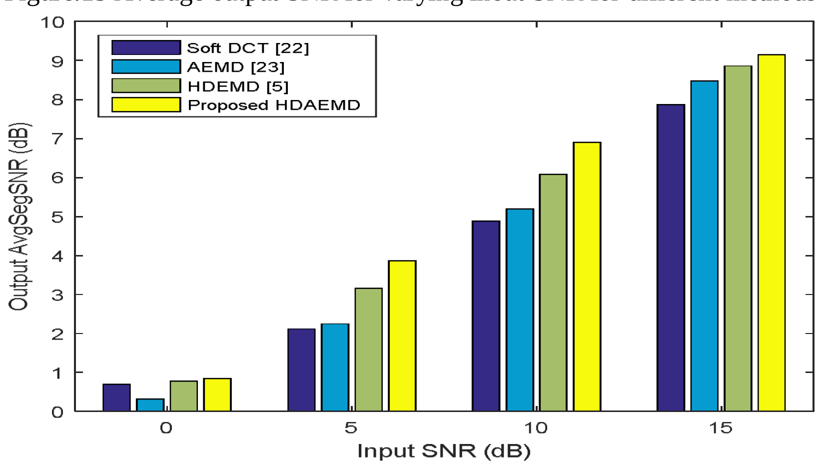


Figure.16 Average output AvgSegSNR for varying inout SNR for different methods

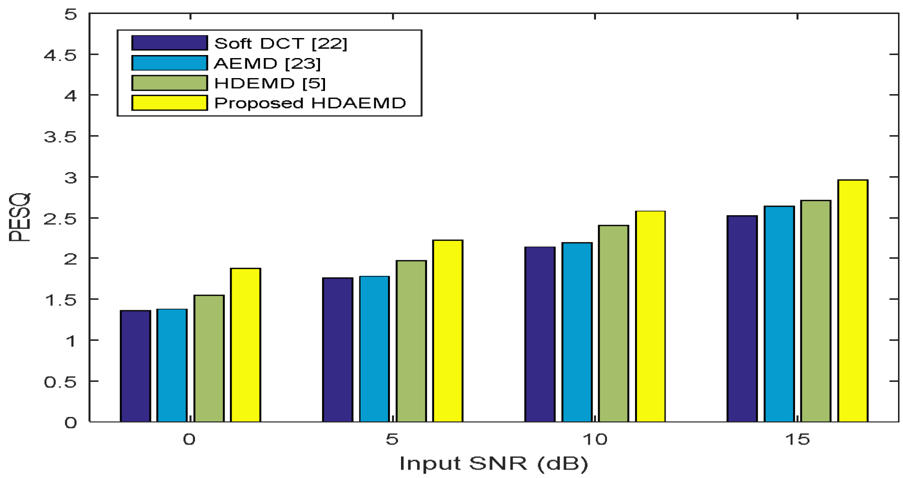


Figure.17 Average PESQ for varying inout SNR for different methods

The average results of the computer simulationsfor various speech utterances for a wide range of SNRvalues with a comparison of different denoising methods are represented in the above figures.15-17. The superiority of the proposed HDAEMD was clearly observed from these figures. The output SNR is a perfect measure for speech quality assessment but it can't assess the performance at all SNR levels, thus another measure, AvgSegSNR was also evaluated for all levels of SNR and also for various denoising methods. The obtained results of AvgSegSNR for various denoising methods shows the superiority of proposed HDAEMD approach. The details of obtained output AvgSegSNR for different methods are represented in figure.16. Further to provide a better idea about the speech quality of enhanced speech signals, the enhanced speech signals are subjected to Perceptualevaluation of speech quality (PESQ) test. Here, the PESQ test was conducted by considering the scores of listeners. For this purpose, the original and enhanced speech signals were listened by a randomly selected 20 listeners and then asked them to give the score from 0 to 5. The PESQ is a subjective measure, gives the score from 0 to 5, with higher scores indicating better quality. The details of PESQ are shown figure.17, shows the enhancement of proposed approach. As discussed in earlier, the soft DCT thresholding is not applicable for non-stationary noises, the performance decreases gradually as the noise variance increases. In our approach, due to the subband nature of first stage, the HDAEMD is robust for such type of noises. In the second stage, the proposed AEMD always performs improvement for all SNR levels.

VI. CONCLUSIONS

This paper proposed a novel speech enhancement technique to suppress the non-stationary noises from a noise contaminated speech signal. The complete accomplishment of proposed approach is carried out in two phases, DCT based soft thresholding and Adaptive EMD based denoising. This is a hybrid approach which was formulated by combining the DCT with EMD. Instead of applying a linear thresholding, this approach derives an adaptive thresholding for the noise suppression from DCT coefficients. Further in the second phase, the proposed approach accomplished a recursive noise cancellation block to further reduce the residual noise present in the first stage output. Simulation is carried out over different speech samples contaminated with different noises at different signal strengths and the performance is measured through the performance metrics. The obtained performance metrics are compared with the earlier approaches and proved to be efficient. Thus the proposed approach can achieve an optimal performance in the speech enhancement.

REFERENCES

[1] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, "Discrete-Time Processing of Speech Signals," IEEE Press, New York, NY, USA, 2000.
[2] D. L. Donoho, "De-noising by soft-thresholding," IEEE Transactions on Information Theory, vol. 41, no. 3, pp. 613–627, 1995.
[3] M. Bahoura and J. Rouat, "Wavelet speech enhancement based on the Teager energy operator," IEEE Signal Processing Letters, vol. 8, no. 1, pp. 10–12, 2001.
[4] S. Salahuddin, S. Z. Al Islam, M. K.Hasan, and M.R.Khan, "Soft thresholding for DCT speech enhancement," Electronics Letters, vol. 38, no. 24, pp. 1605–1607, 2002.
[5] ErhanDeger, "Subband DCT and EMD Based Hybrid Soft Thresholding for Speech Enhancement", Hindawi Publishing Corporation Advances in Acoustics and Vibration, Volume 2014, Article ID 765454.
[6] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 27, no. 2, pp. 113–120, 1979.

[7] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 32, no. 6, pp. 1109–1121, 1984.
[8] S. Chang, Y. Kwon, S.-i.Yang, and I.-j. Kim,. Speech enhancement for non-stationary noise environment by adaptive wavelet packet,.In Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), vol. 1, 2002, pp.I.561 .I.564.
[9] Y. Hu and P. Loizou, .Speech enhancement based on wavelet thresholding the multitaper spectrum,.IEEE Trans.Speech and Audio Processing, vol. 12, pp. 59.67, 2004.
[10] S. Tabibian, A. Akbari, and B. Nasersharif, .Anew wavelet thresholding method for speech enhancement basedon symmetric kullback-leibler divergence,.inProc. 14th Int. CSI Computer Conference, (CSICC), Oct. 2009, pp. 495.500.
[11] K. Khaldiet al., “Speech enhancement via EMD,” in Proc. EURASIP J. Adv. Signal Process., 2008, vol. 2008, p. 8.
[12] Y. Kopsinis and S. McLaughlin, “Development of EMD-based denoising methods inspired by wavelet thresholding,” IEEE Trans. Signal Process., vol. 57, no. 4, pp. 1351–1362, Apr. 2009.
[13] G. Rilling and P. Flandrin, “One or two frequencies?The empirical mode decomposition answers,” IEEE Trans. Signal Process., pp. 85–95, Jan. 2008.
[14] Y. Kopsinis and S. McLauglin, “Investigation and performance enhancement of the empirical mode decomposition method based on a heuristic search optimization approach,” IEEE Trans. Signal Process., pp. 1–13, Jan. 2008.
[15] Y. Kopsinis and S. McLauglin, “Improved EMD using doubly-iterative sifting and high order spline interpolation,” EURASIP J. Adv. Signal Process., vol. 2008, 2008.
[16] NavinChatlani and John J. Soraghan, “EMD-Based Filtering (EMDF) of Low-Frequency Noise for Speech Enhancement”, IEEE Transactions on Audio, Speech, and Language Processing, Vol. 20, No. 4, May 2012
[17] M. E. Hamid, S. Das, K. Hirose, and M. K. I. Molla, “Speech enhancement using EMD based adaptive soft-thresholding (EMD-ADT),” International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 5, No. 2, pp. 1–16, 2012.
[18] A.P. Soman, “EMD and Hurst based mode selection for speech enhancement”, International Journal of Advanced Research in Electronics and Communication Engineering, Vol.4, No.2, pp.175-182, 2015.
[19] A. Upadhyay, “Speech enhancement based on mEMD-VMD method”, IEEE Electronics Letters, Vol.53, No.7, pp.502-504, 2017.
[20] A. Upadhyay, M. Sharma, R. B. Pachori, “Determination of instantaneous fundamental frequency of speech signals using variational mode decomposition”, Journal of Computers & Electrical Engineering, Vol.62, pp.630-647, 2017.
[21] K. Dragomireisky and D. Zosso, “Variational Mode Decomposition”, IEEE Transactions on Signal Processing, Vol.62, No.3, pp.531-544, 2014.
[22] M. K. Hasan, M. S. A. Zilany, and M. R. Khan, “DCT speech enhancement with hard and soft thresholding criteria,” Electronics Letters, vol. 38, no. 13, pp. 669–670, 2002.
[23] John J Soraghan, “Adaptive EMD for signal enhancement with application to speech”, IEEE Conference Paper, JULY 2008.