Suppression of Non-Stationary Noises Through the Generalized Signal Detector

Nageswara Rao Saggurti[1]   Jaya Sankar Koti Reddy[2]   Dhanunjaya Naidu Challa[3]

[1]Department of Electronics and Communication Engineering,
Sri Venkateswara Engineering College, Suryapet, Telangana, India
[2]Department of Electronics and Communication Engineering,
Vasavi College of Engineering, Hyderabad, India
[3]Department of Electronics and Communication Engineering,
Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India
* Corresponding author's Email: nageshwarrao135@gmail.com

Abstract: This paper presents an adaptive speech enhancement approach to suppress non-stationary noises form a noisy speech signals. This approach is based on the Empirical Mode Decomposition and Signal Uncertainty. The EMD is just used as pre-processor for signal decomposition into Intrinsic Mode Functions. Further The IMFs are processed for noise suppression through a recursive smoothing based on a smoothing factor which was decided based on the probability of speech presence. A new signal detector is proposed here to measure the probability of speech presence. This approach mainly focused on the optimization of probability of detection through the newly proposed signal detector. The simulation is carried out through various speech signals contaminated with different noise types like White noise, Babble noise and Airport noise at various SNR levels reveals the outstanding performance of proposed approach. The performance evaluation is carried out by measuring the performance metrics, Overall output SNR, Output AvgSegSNR and Perceptual Evaluation of Speech Quality (PESQ). The evaluation is carried out for varying noise strengths and for every test case all these metrics are evaluated and compared with conventional approaches.

Keywords: Speech enhancement, Non-stationary noise, Speech uncertainty, EMD, PESQ.

1. Introduction

Recently, the drastic growth in the market of speech communication applications like hearing-aids, teleconferencing, speech recognition and hands free telephony motivated the research on the speech enhancement. Particularly in the hands free telephony speech communication systems, the microphones are typically placed at certain distant from the speaker's mouth. In such cases, various noise sources makes the speech signal corrupted, by which the performance of a speech oriented devices are not able to process the signal effectively. For example the performance of an automatic speech recognition (ASR) system becomes worst in the case of adverse acoustic environment [1]. Hence there is a necessity of speech preprocessing to remove the unnecessary noise before transferring it to the final peripheral devices.

Several approaches have already been proposed to improve the speech enhancement results. Although the microphone array based approach exhibits better results, at the same time speech processing research community is trying to reduce the number of microphones (channels). 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 [2]. Its basic requirement is the noise spectrum which is determined from the non-speech segments [3]. In such 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. The subband

approach to speech enhancement is another potential method. Fourier transform and wavelet transform are dominating methods widely used in subband based speech enhancement techniques. But the Fourier transformation is not suitable to analyze non-stationary signals like speech. There are several approaches of using wavelet transformation in subband decomposition. The decomposition results are varied with the different parameters, for example, the basis wavelet, the number of decomposition levels, and so forth. Moreover, the selection of parameters also depends on the analyzing data. Therefore, a data adaptive tool for analyzing non-stationary and nonlinear signal is highly desirable [4].

This paper proposes new speech enhancement approach by focusing the signal uncertainty as a main critic. EMD is used as a pre-processing filter to decompose the noisy speech signal into IMFs. Further a new signal detector approach is proposed here to improve the probability of detection and to reduce the probability of false alarm. Based on the hypothesis framework, the segments of IMFs are decided as speech or noise dominant. For the noise dominant segments the proposed approach applies a recursive smoothing to remove the noise. The proposed approach applied over the noisy speech signals taken from TIMIT data base for performance evaluation. Non-stationary noises like car noise, babble noise, airport noise are considered at three SNR levels.

Rest of the paper is organized as follows; Section II illustrates the details of Literature survey. Section III illustrates the details of proposed speech enhancement approach. Section IV illustrates the simulation results and section V concludes the paper.

2. Related work

Several methods have been used to reduce the noise existing in degraded speech signals. Spectral subtraction is one of these methods, and it is easily implemented. In this method, the noise spectrum is estimated first from the silence periods of the noisy signal, and then this estimated noise spectrum is subtracted from the spectrum of degraded speech producing an estimate of the original signal with less noise. Although, this method is easy to implement, it submits some source of musical noise to the enhanced signals [5].

Empirical Mode Decomposition (EMD) is observed to be a better technique to study the characteristics of non-stationary signals. EMD is developed by Huang et.al. [7] to decompose a non-stationary signals into a finite set of functions called

Intrinsic Mode Functions (IMFs). Several approaches are developed in earlier based on EMD [6]. In [6], 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 [8] 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 [9] to suppress the non-stationary noises. It employs the combination of variational mode decomposition (VMD) [10, 16] and empirical mode decomposition (EMD) methods. This approach tried to filter the noise components efficiently in both low frequencies and high frequencies. All these approaches tries to do the speech enhancement through the thresholding, i.e., based on particular static measure the noise components are filtered from the noisy speech signal. However all these approaches introduce a musical noise after the speech reconstruction. The musical noise has similar characteristics with noise and needs to be reduced further. Some approaches also focused to reduce the musical noise by considering the signal uncertainty problem.

In [11], minima controlled recursive averaging (MCRA) approach was developed to reduce the noise from a noisy speech. The noise estimate is given by averaging past spectral power values and using a smoothing parameter that is adjusted by the signal presence probability in subbands. Presence of speech in subbands is determined by the ratio between the local energy of the noisy speech and its minimum within a specified time window. However the probability of speech presence considered a constant value of 2 for all types of noises. This constant value does not suites for all types' non-stationary noises. One more approach is developed in [12] by assuming that the noisy speech signal follows a chi-square distribution and the signal uncertainty is measured based on two hypotheses. This approach combines the DCT and EMD for speech enhancement, DCT as a first stage filter and EMD is for reducing the musical noises existing in the output of first stage. However in the chi-square

distribution the signal variables are assumed to be independent. But there exists a strong correlation between the samples in a noisy speech.

3. Proposed approach

3.1 Speech enhancement

Let y(n) ne a noisy speech signal that is the sum of clean speech signal s(n) and a noise signal d(n); y(n)=s(n)+d(n). Apply empirical mode decomposition (EMD) over the noisy speech and divide the every IMF into K segments with every segment having equal number of samples.

$$Y(m,k) = S(m,k) + D(m,k) \qquad (1)$$

Where $m = 1,2, \dots, M$ is the frequency bin and the $k = 1,2, \dots, K$ is the segment index. $S(m,k)$ and $D(m,k)$ denotes the speech signal and noise signal of k_{th} frequency bin in the m_{th} segment, respectively. Consider two hypotheses $H_0(m,k)$ and $H_1(m,k)$ which indicates the absence of speech and presence of speech respectively, as

$$\begin{aligned} H_0(m,k): Y(m,k) &= D(m,k) \\ H_0(m,k): Y(m,k) &= S(m,k) + D(m,k) \end{aligned} \qquad (2)$$

In the case of $H_0(m,k)$, the speech signal is absent and only the noise signal is present. Whereas in the case $H_1(m,k)$ denotes the presence of both speech signal and noise signal.
Let $\sigma_d(m,k) = E[|D(m,k)|^2]$ be the noise variance in the m_{th} segment, the speech signal can be detected by the conventional signal detector (CSD) [13] as

$$W = \sum_{i=1,k=constant}^{N} \left(\frac{Y(i,k)}{\sigma_d(k)}\right)^2 \underset{H_0}{\overset{H_1}{\underset{<}{>}}} \delta \qquad (3)$$

Where N is the total number of samples in k_{th} segment and δ is the threshold to be determined. In the above expression, the presence or absence of speech signal is detected by simply dividing the amplitudes of all speech samples with noise variance σ_d and then squaring the result. Then the obtained result is compared with a predefined threshold. If the result is greater than the threshold, then it is declared as the presence of speech signal otherwise it is declared as the absence of speech signal.
Form the above Eq. (3), it can be observed that the CSD decides the k_{th} segment is belongs the Hypothesis H_1 or H_0 just by squaring the amplitudes

of all samples divided by the noise variance σ_d. However a simple squaring operation cannot give much efficient results of probability of detection (P_d), i.e., it can't increases the probability of detection or reduces the probability of false alarm (P_F). In this case P_d and P_F are defined as

$$P_d = pr(H_1|H_1) \qquad (4)$$
$$P_F = pr(H_1|H_0) \qquad (5)$$

Here the P_d is defined as the probability of correct detection, i.e., the result shown by Eq. (3) and the assumption are same. Otherwise it results under P_F. Since the characteristics of non-stationary noises like Babble noise, Airport noise etc., are varying in nature, the signal detector has to consider this varying nature also. Then only the signal can be detected exactly. For every noise, there are unique characteristics. Thus a common threshold cannot satisfy the detection performance.
The noise variance $\sigma_d(k)$ involved in the Eq. (3) reflects the varying nature of non-stationary noises by a simple squaring operation can't differentiate the entire segment is noise dominant or signal dominant. Hence the signal detector must be an optimal and varying in nature according to the varying statistics of noise. The statistics of noise are different from noise to noise, like car noise, babble noise, aircraft noise, fan noise etc. hence this paper proposed a new signal detector called generalized signal detector (GSD) [14] and it is formulated as

$$W' = \sum_{i=1,k=constant}^{N} \left(\frac{Y(i,k)}{\sigma_d(k)}\right)^p \underset{H_0}{\overset{H_1}{\underset{<}{>}}} \delta' \qquad (5)$$

Where $p > 0$ is an arbitrary constant and δ' is a new threshold to be determined by fixing the probability of false alarm, mean and variance values of two hypotheses H_1 and H_0. The proposed approach obtains a noise power estimate by applying a temporal recursive smoothing to the noise in the absence of speech and the new hypotheses are as follows;

$$\begin{aligned} H_0'(m,k): \quad & \hat{\sigma}_d(m,k+1) = \alpha_d \xi + \zeta \\ H_1'(m,k): \quad & \hat{\sigma}_d(m,k+1) = \alpha_d \xi \end{aligned} \qquad (6)$$

Where $\zeta = (1 - \alpha_d)|Y(i,k)|^2$, $\xi = \hat{\sigma}_d(m,k)$ and $\alpha_d \ (0 < \alpha_d < 1)$ is a smoothing parameter. Here the above eq.(6) tries to optimize the detection rate by detecting whether the k_{th} segment is noise dominant or signal dominant. But the notation m specifies the frequency bin. For every segment,

there are M frequency bins and for every frequency bin there is N number of samples. In the m_{th} frequency bin of k_{th} segment, amplitude of every sample is considered for evaluation.

Form [11] a clear distinction was provided between the two hypotheses (2) and (6). Deciding that the speech is not present (H_0) when there is a presence of speech (H_1) makes the speech enhancement not to achieve an optimal performance. Thus the H_1' is chosen to achieve higher confidence levels, i.e., $P(H_1(m,k)|Y(m,k)) \geq P(H_1'(m,k)|Y(m,k))$ [11].

Where the $P(H_1'(m,k)|Y(m,k))$ denotes the posteriori probability of speech presence. Then the presence of speech in the equation.(6) becomes

$$\hat{\sigma}_d(m, k+1) = \varsigma\xi + [1-\varsigma]|Y(m,k)|^2 \quad (7)$$

Where $\varsigma = \hat{\alpha}_d(m,k)$ is a time varying smoothing parameter that is adjusted by the probability of speech presence as follows;

$$\hat{\alpha}_d(m,k) = \alpha_d + (1-\alpha_d) \geq$$
$$P(H_1'(m,k)|Y(m,k)) \quad (8)$$

Further the conditional probability $P(H_1'(m,k)|Y(m,k))$ is decides based on the test static defined in the equation.(5), as

$$P\big(H_1'(m,k)\big|Y(m,k)\big)$$
$$= \alpha_p P\big(H_1'(m,k-1)\big|Y(m,k-1)\big)$$
$$+\big(1-\alpha_p\big)I(m,k) \quad (9)$$

Where $\alpha_p\big(0 < \alpha_p < 1\big)$ is a smoothing factor and I is an indicator function obtained as

$$I(m,k) = \begin{cases} 1 & if \ Q_r(m,k) > \delta' \\ 0 & if \ Q_r(m,k) < \delta' \end{cases} \quad (10)$$

Here defining one new parameter called $Q_r(m,k)$ to measure the statistics of segment. It is simply obtained by dividing the local energy of a noisy speech ($Q(m,k)$) by its determined minimum $Q_{min}(m,k)$.

$$Q_r(m,k) = Q(m,k)/Q_{min}(m,k). \quad (11)$$

Where
$$Q(m,k) = \beta Q(m,k-1) + (1-\beta)|Y(m,k)|^2 \quad (12)$$

Where β is a smoothing parameter.

3.2 Threshold (δ') determination

The mathematical formulation for the threshold derivation is carried out in this section. In the conventional signal detector [13] the threshold has a simple relation with the shape parameter (μ_0') and scale parameter (σ_0'). In the generalized signal detector the threshold is determined through a complex process. According to the Neyman-Pearson rule, the expression for threshold [13] is as

$$\delta' = F^{-1}(1 - P_F, \mu_0', \sigma_0') \quad (13)$$

Where

$$\mu_0' = \frac{M^2\{(W`|H_0)\}}{V\{(w`|H_0)\}} = n.\frac{\Gamma^2(\frac{p+1}{2})}{\Gamma(\frac{2p+1}{2})\sqrt{\pi}-\Gamma^2(\frac{p+1}{2})} \quad (14)$$

$$\sigma_0' = \frac{V\{(W`|H_0)\}}{M\{(w`|H_0)\}} = \frac{2^{p/2}}{n}.\frac{\sqrt{\pi}\Gamma(\frac{2p+1}{2})-\Gamma^2(\frac{p+1}{2})}{\Gamma(\frac{p+1}{2})\sqrt{\pi}} \quad (15)$$

Where

$$M\{(w`|H_0)\} = \frac{2^{p/2}}{\sqrt{\pi}}\Gamma(\frac{p+1}{2}) \quad (16)$$

$$V\{(W`|H_0)\} = \frac{2^p\Gamma(\frac{2p+1}{2})}{n\sqrt{\pi}} - \frac{2^p}{n\pi}\Gamma^2(\frac{p+1}{2}) \quad (17)$$

$$F(x, v_1, v_2) = \int_0^x \frac{1}{v_1^{v_2}\Gamma(v_2)}t^{v_2-1}e^{\frac{-t}{v_1}}dt \quad (18)$$

The μ_0' and σ_0' denotes the shape parameter and scale parameter for the hypothesis H_0. In the conventional signal detector, these parameters are defined as just functions of the number of samples by which the probability of detection won't be that much effective. For that reason, here the proposed approach defines the shape and scale parameters as a function of mean and variances, as shown in Eq.(14) and Eq.(15).

In the CSD, the decision variable W does not follow the gamma distribution, however the modified decision variable W' can be well approximated as a Gamma random variable by matching the mean and the variance. This approximation enables us to determine the detection threshold δ' for the new detector in Eq. (5), which is otherwise difficult to obtain without the distribution of W'.

Based on the above illustration, the probability of detection is defined as

$$P_D = 1 - F(\delta', \mu_1', \sigma_1') \quad (19)$$
$$\sigma_1' = \frac{V\{(W`|H_1)\}}{M\{(w`|H_1)\}} = \frac{2^{p/2}(1+\gamma)^{p/2}}{n}.\frac{\sqrt{\pi}\Gamma(\frac{2p+1}{2})-\Gamma^2(\frac{p+1}{2})}{\Gamma(\frac{p+1}{2})\sqrt{\pi}} \quad (20)$$
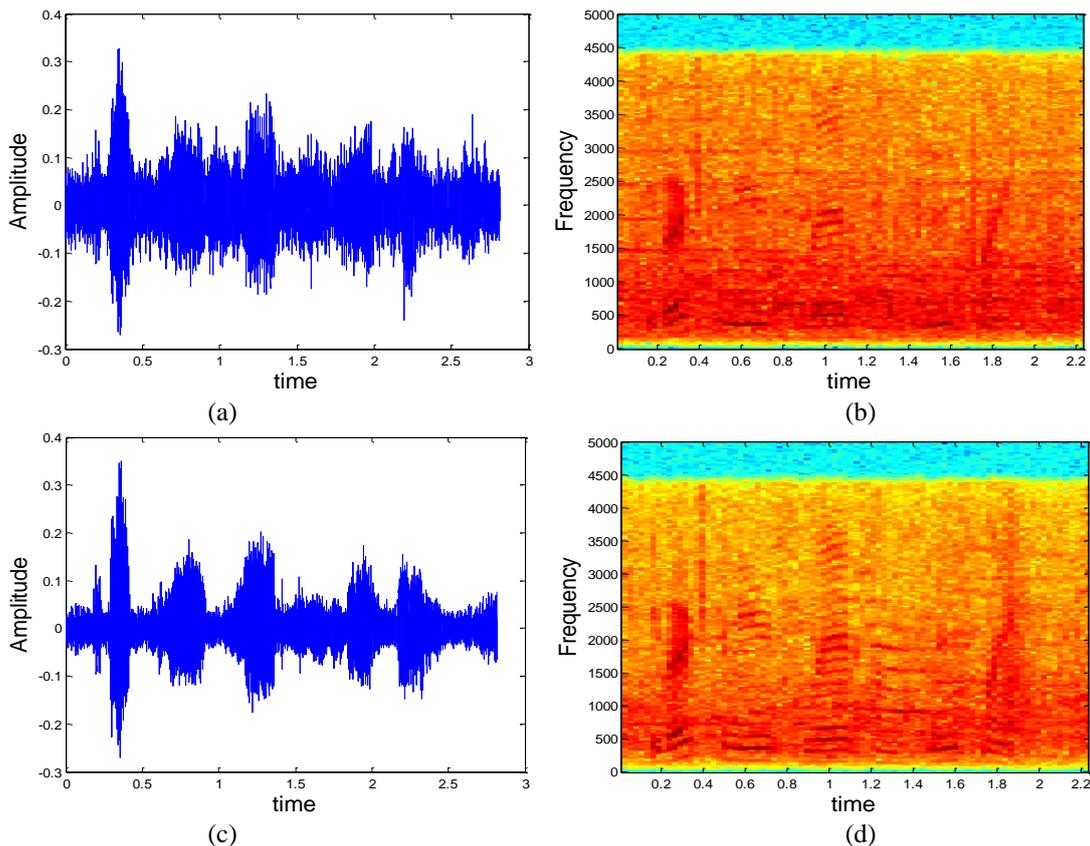
Figure.1 Obtained results: (a) clean speech of a male with airport noise at 0dB SNR, (b) spectrogram of (a), (c) enhanced speech signal of (a), and (d) spectrogram of (c)
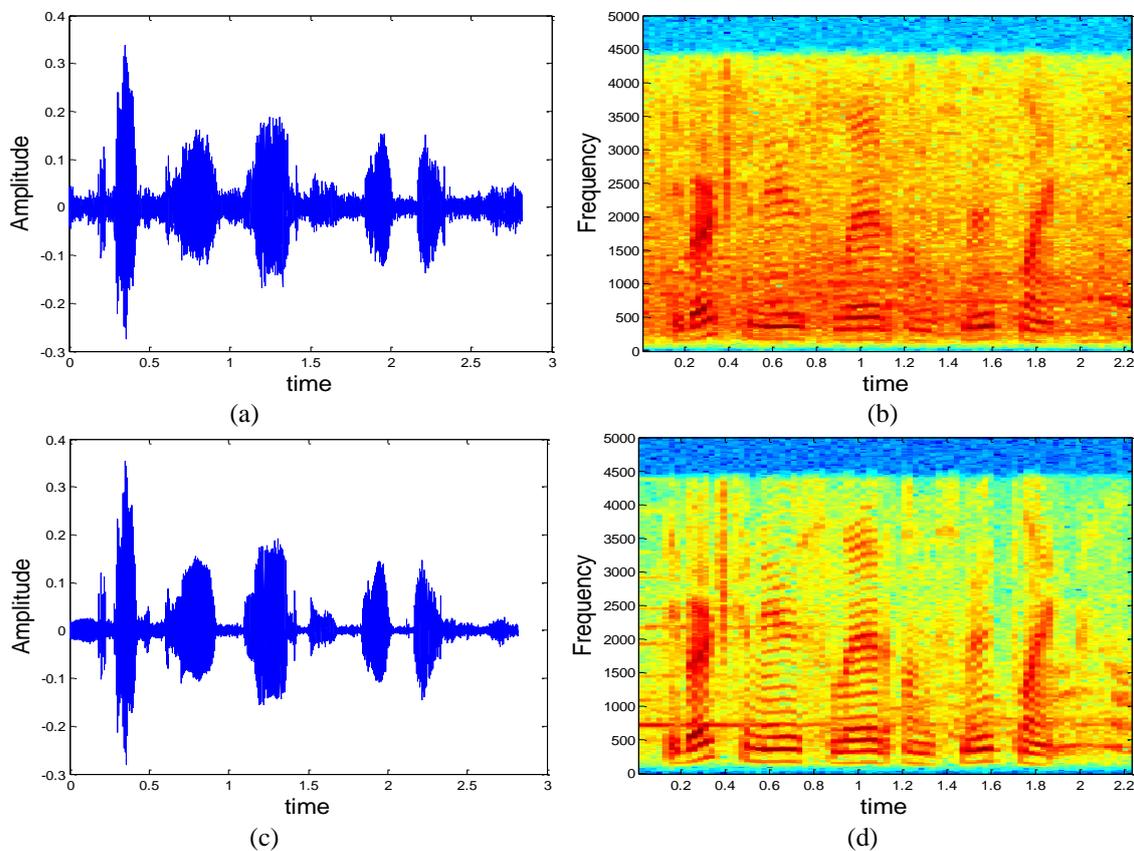

Figure.2 Obtained results: (a) clean speech of a male with airport noise at 10dB SNR, (b) spectrogram of (a), (c) enhanced speech signal of (a), and (d) spectrogram of (c)
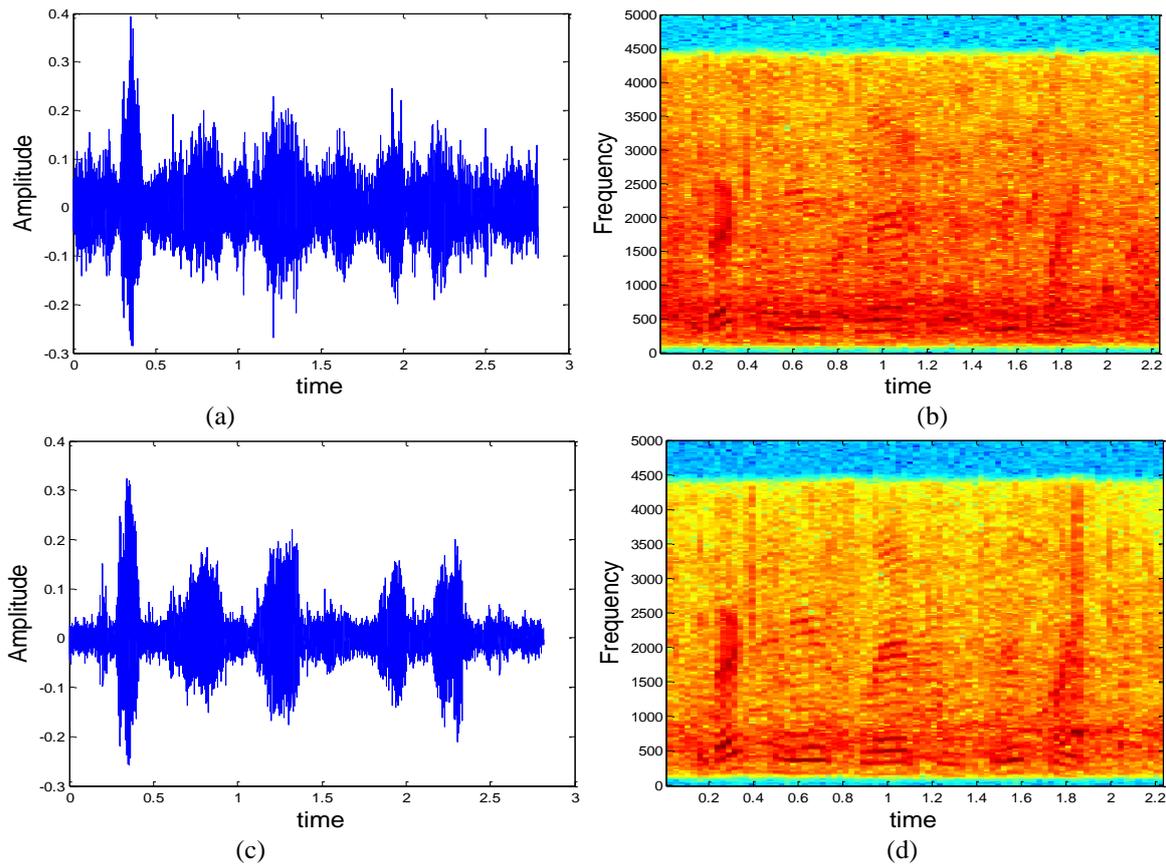
Figure.3 Obtained results: (a) clean speech of a female with babble noise at 0dB SNR, (b) spectrogram of (a), (c) enhanced speech signal of (a), and (d) spectrogram of (c)

$$\mu'_1 = \frac{M^2\{(W^`|H_1)\}}{V\{(w^`|H_1)\}} = n.\frac{\Gamma^2(\frac{p+1}{2})}{\Gamma(\frac{2p+1}{2})\sqrt{\pi}-\Gamma^2(\frac{p+1}{2})} \quad (21)$$

Where γ is the average signal to noise ratio (ASNR) of the total speech signal.

Form the above expressions, we can obtain an optimal values for probability of detection through the detection threshold. Here the detection threshold is determined through the scale and shape parameters of the speech signal when the speech signal is assumed to be absent by which the complete characteristics of noise can be obtained which gives more clarity about the features of clean speech and noise. Here the arbitrary constant p is implicitly related to the false alarm, the probability of detection and ASNR and the number of samples. Thus we can obtain an optimal value of p that maximizes the probability of detection at fixed values of Probability of false alarm, ASNR and N. Thus, Eq. (19) is a very general expression that can be used in different applications. An analytical expression for the optimum value of p is difficult to obtain, if not impossible.

4. Simulation results

This section describes the performance evaluation of the proposed approach. To illustrate the effectiveness of the proposed speech enhancement algorithm, extensive computer simulations were carried out with 10 male and 10 female utterances, randomly selected from the TIMIT database. Totally four types of noses are considered for the evaluation. They are White Noise, Babble Noise, Car Interior Noise and Airport Noise.

To evaluate the performance of proposed approach, overall and average segmental SNR improvements are measured. The quality of the enhanced speech signal through the proposed approach was evaluated through Perceptual evaluation of speech quality (PESQ) test. To show the performance improvement of proposed speech enhancement algorithm, it was compared with conventional approaches proposed by Tafiq Hasan et.al [12] and the S.N.Rao et.al [15].

The obtained results after simulation are represented in figures 1-6. Figure 1 and figure 2 shows the results of a clean speech with airport noise at 0dB and 10dB SNR respectively. Similarly
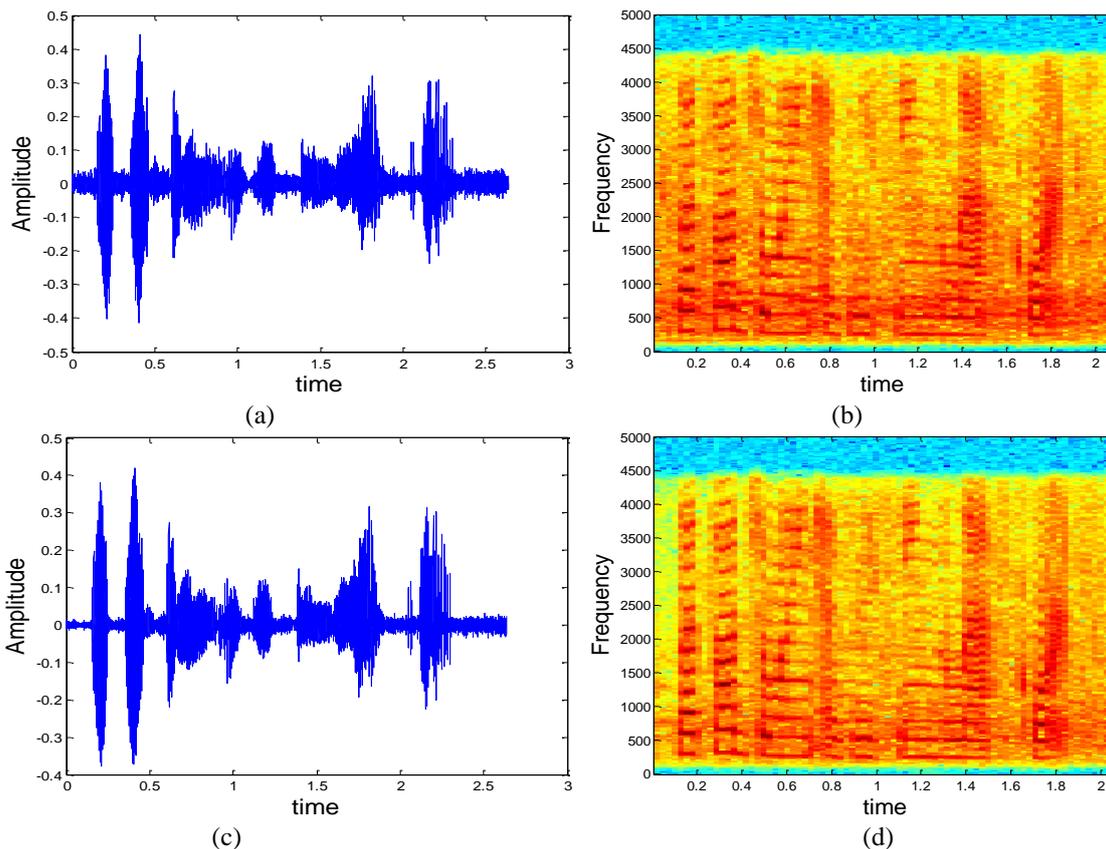
Figure.4 Obtained results: (a) clean speech of a female with babble noise at 10dB SNR, (b) spectrogram of (a), (c) enhanced speech signal of (a), and (d) spectrogram of (c)
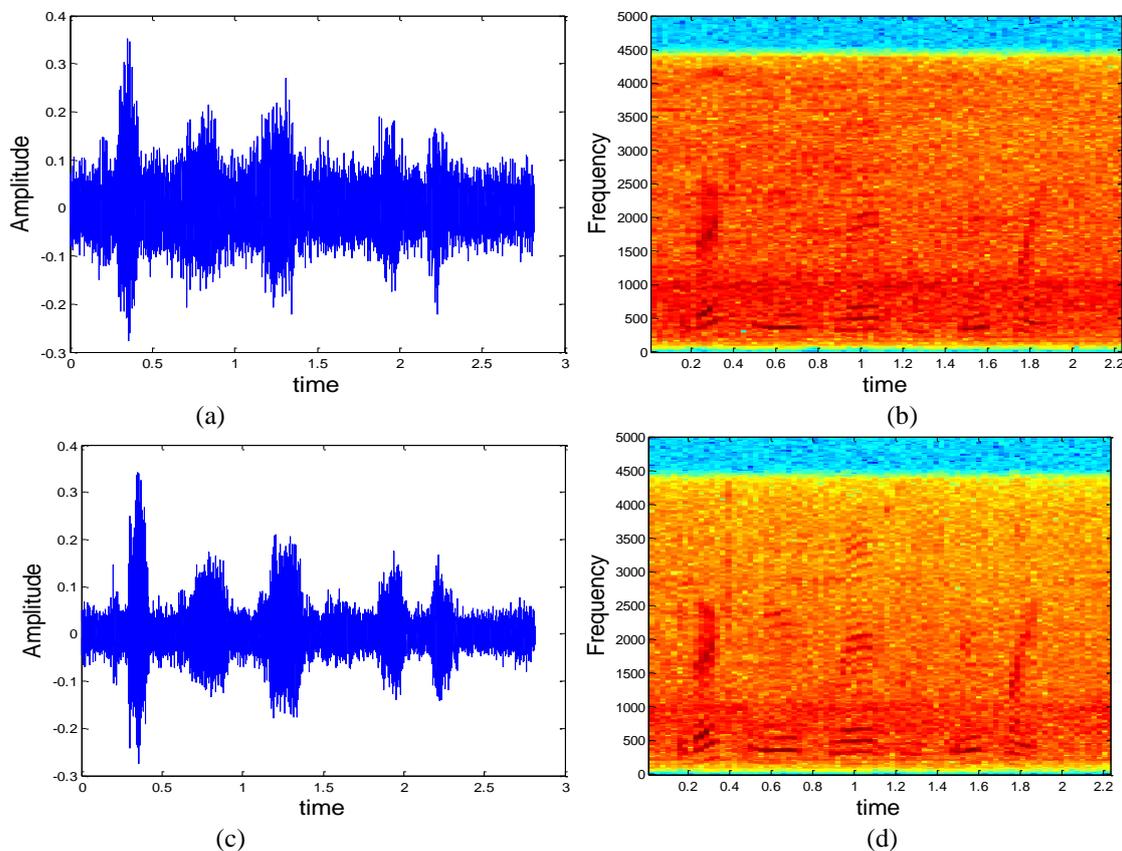

Figure.5: Obtained results: (a) clean speech of a male with car interior noise at 0dB SNR, (b) spectrogram of (a), (c) enhanced speech signal of (a), and (d) Spectrogram of (c)
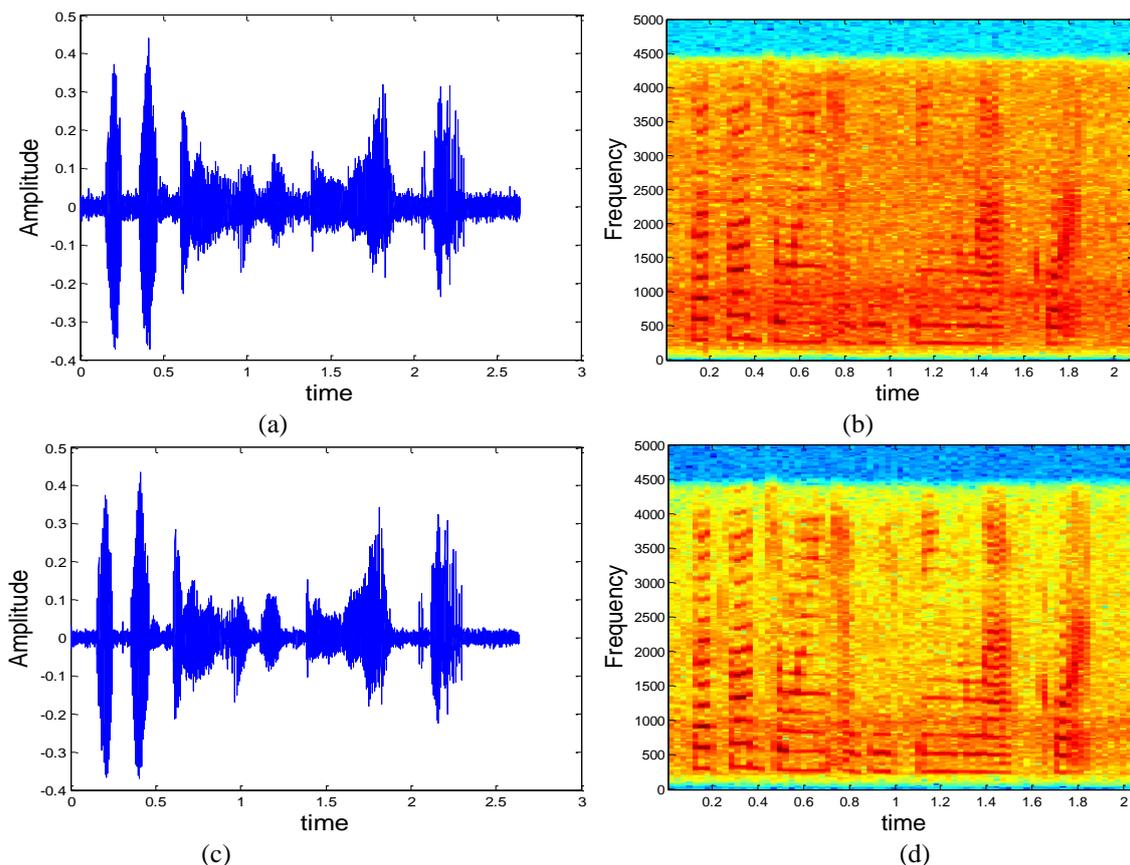
Figure.6 Obtained results: (a) clean speech of a female with car interior noise at 10dB SNR, (b) spectrogram of (a), (c) enhanced speech signal of (a), and (d) spectrogram of (c)

Table 1. Comparison of the output SNR improvements of different denoising methods

Noise Type	Input SNR (dB)	Tafiq.et.al [12]	S.N.Rao et.al [15]	Proposed
Babble	0	2.0345	3.8663	5.2345
	5	5.6685	7.2493	8.8847
	10	11.0027	12.5812	13.3386
	15	16.0037	18.2214	18.7711
Airport	0	2.8492	4.4986	6.1920
	5	6.5743	7.3468	9.8495
	10	11.1297	12.8592	13.4965
	15	16.9171	18.7682	19.7416
Car Interior	0	2.9916	4.2880	5.8902
	5	6.1538	8.1650	8.9204
	10	11.802	13.3732	14.1873
	15	16.1455	19.1808	19.7050

Figure 3 and figure 4 shows the results of a clean speech with Babble noise at 0dB and 10dB SNR respectively and Figure 5 and figure 6 shows the results of a clean speech with Car Interior noise at 0dB and 10dB SNR respectively. The respective spectrograms are also shown in the every figure for both noisy speech and the enhanced speech. Further the performance of proposed approach is evaluated through the performance metrics such as Output SNR and Average SegSNR. The following tables describe the details of performance measures for all test cases.

Tables 1-3 describe the details of Output SNR, Output AvgSegSNR and PESQ for varying noise environments. The complete results are evaluated for average results of 10 male and 10 female utterances for every instant of SNR and noise type. The superiority of proposed approach is observed in the all tables. Form table.1 it can be observed that the proposed approach obtained an increased Output SNR for every Input SNR; compared to the conventional approach it is observed to be high. Similarly the output AvgSegSNR of proposed

Table 2. Comparison of the output AvgSegSNR improvements of different denoising methods

Noise Type	Input Avg SegSNR (dB)	Tafiq.et.al [12]	S.N.Rao et.al [15]	Proposed
Babble	-4.1482	-1.1788	0.0526	0.0556
	-1.2744	0.0144	2.0585	2.3377
	2.2858	2.6682	3.3983	3.4085
	5.6931	6.0244	7.9849	8.2963
Airport	-4.1041	-1.1593	0.0805	0.1215
	-1.2358	0.0629	2.1334	2.3532
	2.3615	2.7125	3.4256	3.6698
	5.7725	6.0886	8.0519	8.3458
Car Interior	-4.1332	-1.0897	0.1177	0.1209
	-1.2486	0.1103	2.0840	2.3710
	2.3698	2.7229	3.4488	3.4665
	5.7185	6.0382	8.0548	8.3183

Table 3. Comparison of the PESQ improvements of different denoising methods

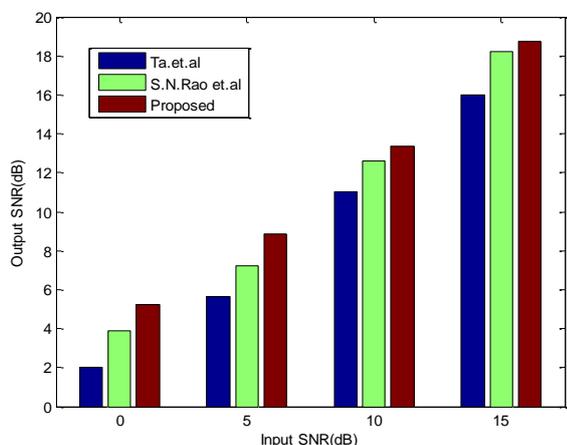| Noise Type	Input SNR (dB)	Tafiq.et.al [12]	S.N.Rao et.al [15]	Proposed
Babble	0	2.8502	3.2620	3.3356
	5	2.9986	3.7414	3.8574
	10	3.0014	3.8332	3.9223
	15	3.1047	3.9223	3.9984
Airport	0	2.7687	3.2423	3.3004
	5	2.9742	3.7162	3.7743
	10	2.9084	3.7715	3.8637
	15	3.0697	3.8749	3.9434
Car Interior	0	2.7971	3.0339	3.1438
	5	2.9206	3.2846	3.3288
	10	2.9079	3.4256	3.5565
	15	3.0917	3.5369	3.7130

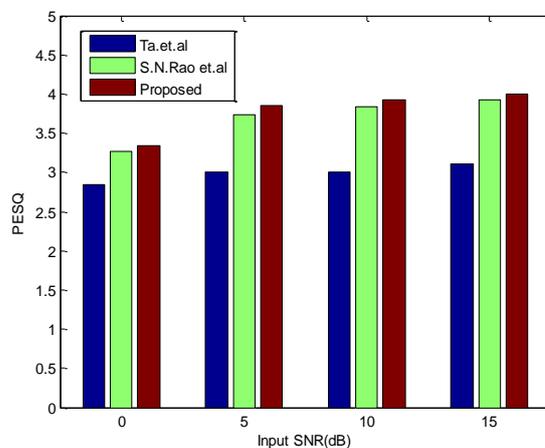Figure.7 Output SNR comparison for different methods

Figure.8 PESQ comparison for different methods

approach is also observed to be high compared with Tafiq.et.al [12] and S.N.Rao et.al [15].

Further to provide a better idea about the speech quality of enhanced speech signals, the enhanced speech signals are subjected to Perceptual evaluation 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 in table.3. The PESQ values obtained through the proposed approach are observed to be high when compared with the PESQ values of conventional approaches. Further the obtained results for white noise at various input SNR levels are represented in the following figures.

Figure 7 and figure 8 shows the comparative analysis of Output SNR and PESQ for various values of Input SNRs, for both proposed and conventional approaches. From the above figures, the proposed approach is observed to achieve an outstanding performance compared with conventional approaches.

5. Conclusion

A new speech enhancement algorithm was proposed in this paper with an aim of reducing the non-stationary noises added on the clean speech signals. Suppression of non-stationary noise in a critical issue. Because, the characteristics of noises are different for different noises. A constant suppression criterion can't afford to achieve better performance for all types of noise. Thus, this paper proposed an adaptive technique which considers the internal characteristics of noises to suppress them from a noisy speech signal. Most of the earlier approaches focused on the thresholding in which the signal gets cropped. Even though there are adaptive thresholding techniques they are not able to suppress the noises, especially the non-stationary noises. Since the proposed approach tried to suppress the noise features based on the mean and variance of signal, the proposed approach achieved better results compared to conventional approaches. further the proposed approach test the signal both segment and frequency bins wise, the noise features are detected more precisely. Here the speech presence uncertainty is mainly focused to detect whether the segment of a speech is noise dominant or signal dominant. Further the noise dominant segments are filtered through a recursive smoothing. The obtained results for various noises at various SNR levels reveal the outstanding performance of proposed approach. In the case of Babble noise, on an average the proposed approach achieved an improvement of 2.8789 dB and 1.0777 dB in the Output SNR compared with the Tafiq.et.al [12] and S.N.Rao et.al [15] respectively. Similarly the improvement is observed to be 7.1816 dB and .4517 dB for Airport Noise. Further the improvement for Car interior noise is 2.9025 dB and 0.9240 dB.

From the obtained Output AvgSegSNR values, the proposed approach is observed to be achieved an improvement of 4.9626 dB and 0.4935 dB from the conventional approaches Tafiq.et.al [12] and S.N.Rao et.al [15] respectively. Further the obtained results through PESQ test also reveal an enhancement of 2.0861 and 0.2732 approximately from the conventional approaches.

In the case of deep analysis about the signal uncertainty, the future work can be extended through the design of an adaptive threshold which confines to the exact detection of signal and noise with less probability of false alarm.

References

[1] A. Sankar, S. F. Beaufays, and V. Digalakis, "Training data clustering for improved speech recognition", In: *Proc. of European Conference on Speech Communication and Technology (EUROSPEECH)*, Madrid, Spain, pp.1-4, 1995.

[2] 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.

[3] 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.

[4] E. Deger and M.K. Hasan, "Speech enhancement using soft thresholding with DCT-EMD based hybrid algorithm", In: *Proc. of the European Signal Processing Conference (EUSIPCO)*, September 2007.

[5] A. Pawar and P. Choudhari, "Enhancement of speech in noisy conditions", *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, Vol.2, No.7, pp.3385-3391, 2013.

[6] 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.

[7] N. E. Huang, Z. Shen, and S. Long, "The empirical mode decomposition and Hilbert spectrum for non-linear and non-stationary time series analysis", In: *Proc. of The Royal Society London A*, Vol. 454, No.1971, pp. 903–995, 1998.

[8] 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.

[9] A. Upadhyay, "Speech enhancement based on mEMD-VMD method", *IEEE Electronics Letters*, Vol.53, No.7, pp.502-504, 2017.

[10] 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.

[11] I. Cohen and B Bergudo "Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement", *IEEE Signal Processing Letters,* Vol.9, No.1, pp.12-15, 2002.

[12] T. Hasan and Md Kamrul Hasan "Suppression of Residual Noise from Speech Signals Using Empirical Mode Decomposition", *IEEE signal processing letters,* Vol. 16, No. 1, pp.2-5, 2009.

[13] Y. Chen, "Improved Energy Detector for Random Signals in Gaussian Noise", *IEEE Transactions on Wireless Communications,* Vol. 9, No. 2, pp.558-563, 2010.

[14] S. S. Kalamkar and A. Banerjee, "On the Performance of Generalized Energy Detector under Noise Uncertainty in Cognitive Radio", In: *Proc. of National Conference on Communications (NCC),* 2013.

[15] S.N, Rao, K.J. Shankar, and C.D. Naidu, "An adaptive speech enhancement approach based on DCT and empirical mode decomposition", In: *Proc. of International Conference on Communication and Signal Processing (ICCSP),* 2016.

[16] K. Dragomireisky and D. Zosso, "Variational Mode Decomposition", *IEEE Transactions on Signal Processing,* Vol.62, No.3, pp.531-544, 2014.