IOSR Journal of Electronics and Communication Engineering (IOSRJECE)
ISSN : 2278-2834 Volume 2, Issue 1 (July-Aug 2012), PP 40-46
www.iosrjournals.org

SERA Noise Estimation for Speech Enhancement

B. Ravi Teja[1], Dr.B. Ananda Krishna[2]
[1,2]*M.Tech, Gudlavalleru Engineering College, Gudlavalleru, Krishna Dt. AP, India*

Abstract: *A practical single channel speech enhancement system consists of two major components, estimation of noise power spectrum and the estimation of speech. Therefore, a crucial component of any algorithm is the estimation of the noise power spectrum for highly non stationary noise environments. The performance of noise estimation algorithm is usually a tradeoff between speech distortion and noise reduction. In existing methods, noise is estimated only during speech pauses and these pauses are identified using Voice Activity Detector (VAD). This paper describes novel noise estimation method SERA (Spectral Entropy Recursive Averaging) to estimate noise in highly non stationary noise environments. In SERA, noise estimation is updated in both speech pauses and also speech present frames. Speech presence is determined by computing the ratio of the noisy speech power spectrum to its local minimum, which is computed by averaging past values of the noisy speech power spectra with a look-ahead factor. Environmental noise is present in all the frames of the noisy speech signal and if the speech/silence detection is not accurate, then it yields speech echoes and residual noise in the enhanced speech. In this paper, noise estimation is updated by dividing speech signal into pure speech, quasi speech and non-speech frames based on adaptive multiple thresholds without using of VAD. The proposed method is compared with weighted average noise estimation method in terms of segmental SNR. The simulation results of the proposed algorithm shows better performance over a system that uses VAD in noise estimation.*

Keywords: *Entropy, Noise Estimation, Quasi Speech, SERA, Smoothing Constant, Speech Enhancement, Voice Activity Detector (VAD)*

I. INTRODUCTION

Speech enhancement has found many applications particularly multimedia, wireless communications, communications between pilot and air traffic control tower, speech recognition, speech coding, etc. are affected by noise. The presence of noise in speech signals can result in appreciable degradation in both the quality and intelligibility. In automatic speech recognition systems, the performance degrades badly in the case of adverse environments with very low SNR. In the case of mobile communication, the speech signal is degraded by different types of noise in the communication channel. Noise is an unwanted signal and there are many types such as background noise, vehicle noise, etc. Unless, the nature of noise is known, it is difficult to enhance the speech. Due to random nature and inherent complexities of various types of noises, it is therefore Noise Spectrum Estimation is an important aspect of speech enhancement.

In most situations we have only the noisy speech signal available while the noise may be non-stationary and its power is unknown. Noise information has to be extracted from the noisy speech signal alone. Noise power estimation is crucial to effective speech enhancement .If Noise Estimate is too low, annoying residual noise will be audible, while if the noise estimate is too high, speech will be distorted resulting possibly in intelligibility loss.

In single channel speech enhancement systems there will be access only to noisy speech and hence the noise statistics have to be estimated from the noisy speech itself. The Main objectives of speech enhancement techniques are to improve quality, intelligibility, robustness and to increase the accuracy of the speech Recognition [1]. Speech enhancement techniques are concerned with algorithms that mitigate these unwanted noise effects and thus improve signal quality. Many speech enhancement systems have been developed based on spectral subtraction and Wiener filtering principles. The common features of all these methods are to estimate the power spectrum of clean speech using the power spectrum of noisy speech. Speech enhancement is an extremely difficult problem if we don't make any assumptions about the nature of the noise signal we aim to remove, since it is difficult to extract the information from noisy speech signals. Usually the noise spectrum estimate is obtained from the first few milli-seconds of noisy speech which are silence regions. This assumption is valid for the case of stationary noise in which the noise spectrum does not vary much over time. Traditional VADs also track the noise only frames of the noisy speech to update the noise estimate [4]. But the update of noise estimate in those methods is limited to speech absent frames. This is not enough for the case of non-stationary noise in which the power spectrum of noise varies even during speech activity. Hence there is a need to update the noise spectrum continuously over time [9]. Since it is difficult to extract the information from noisy speech signals, many noise estimation algorithms were proposed.

II. RELATED WORKS

Usually the noise spectrum estimate is obtained from the first few milli-seconds of noisy speech which are silence regions. This assumption is valid for the case of stationary noise in which the noise spectrum does not vary much over time [10]. Traditional VADs also track the noise only frames of the noisy speech to update the noise estimate. But the update of noise estimate in those methods is limited to speech absent frames. This is not enough for the case of non-stationary noise in which the power spectrum of noise varies even during speech activity. Hence there is a need to update the noise spectrum continuously over time and this is done by a noise estimation algorithms. Several noise-estimation algorithms have been proposed for speech enhancement applications.

The author Martin[6] proposed minimum statistics algorithm based on tracking of minima of noise power and this algorithm failed to predict rise in speech power and rise in noise power during voiced speech intervals and it requires large window length to encompass long segment of speech to work effectively. The authors Cohen I. and Berdugo[12] proposed Minima controlled recursive algorithm (MCRA) which updates the noise estimate by tracking the noise-only regions of the noisy speech spectrum. These regions are found by comparing the ratio of the noisy speech to the local minimum against a threshold. The noise estimate, however, lags by at most twice that window length when the noise spectrum increases abruptly.

In the improved MCRA approach, a different method was used to track the noise-only regions of the spectrum based on the estimated speech-presence probability. The noise estimated by continuously tracking the minimum of the noisy speech in each frequency bin. However, it fails to differentiate between an increase in noise floor and an increase in speech power updated the noise estimate by comparing the noisy speech power spectrum to the past noise estimate. Their method is also simple to implement, however it fails to update the noise estimate when the noise floor increases abruptly and stays at that level. Hirsch and Ehrlicher [13] proposed weighted average algorithm based on smoothing the spectral values of noisy speech. Noise estimation will never be updated, if SNR (Signal to Noise Ratio) is at high level.

To overcome this drawback, this paper addresses a reliable and fast noise estimation technique –SERA for speech enhancement in real time environment. The section 3 describes the proposed noise estimation algorithm, implementations and results in section 4 and section 5 concludes the work.

III. PROPOSED WORK

In this Paper, original signal consists of speech and noise, is given as input to the SERA algorithm. In SERA, noise estimation is updated in both speech pauses and also speech present frames. Speech presence is determined by computing the ratio of the noisy speech power spectrum to its local minimum, which is computed by averaging past values of the noisy speech power spectra with a look-ahead factor. The output of SERA algorithm is given as input to the Wiener filter is shown in Figure 1. The Wiener filter, filter out the noise based on posterior SNR and a prior SNR through gain function.

Figure 1: over view of proposed work

In this work, the estimation of noise algorithm is improved in the following aspects.
Update of the noise estimate without using voice activity decision.
Estimate of speech-presence probability exploiting the correlation of power spectral components in neighboring frames.
The proposed algorithm updates the noise estimate in each frame using a time–frequency dependent smoothing factor computed based on the speech-presence probability with the help of spectral entropy.
Let the noisy speech signal is denoted as

$$y(n)=x(n)+d(n) \quad \text{.....(1)}$$

Where x(n) is the original speech and d(n) is the noise.
The Fourier transform of y(n), x(n) and d(n) in the l^{th} frame and at the k^{th} frequency bin are expressed by

$$Y(l,k)=X(l,k)+D(l,k) \quad \text{.....(2)}$$

The smoothed power spectrum of noisy speech is computed using the following first - order recursive equation

$$\widehat{N}(l,k) = \alpha\widehat{N}(l-1,k) + (1-\alpha)|Y(l,k)|^2 \quad \text{...(3)}$$

Where $\widehat{N}(l,k)$ is the smoothed power spectrum, l is the frame index, k is the frequency index, $|Y(l,k)|^2$ is the short time power spectrum of noisy speech and α is a smoothing constant [7]. Smoothing constant is not fixed but varies with time and frequency. The above recursive equation provides a smoothed version of periodogram $|Y(l,k)|^2$[11].

3.1 SERA (Spectral Entropy Recursive Averaging Noise Estimation Algorithm)

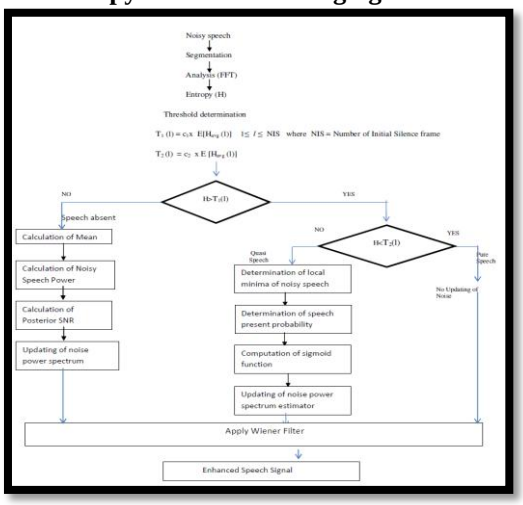


Figure 2: SERA Algorithm.

Figure.2 is the algorithm of our work and in this the noise signal is segmented into number of frames and FFT is performed on those frames [8]. To discriminate various frames of noisy speech signal entropy is calculated. Based on threshold determination, classification of noisy speech signal is done.

3.2 Determination of Entropy

The proposed noise estimation method classifies the noisy speech into three categories as pure speech, non-speech and quasi speech precisely. For this purpose, two thresholds are introduced for entropy H (*l*).

$$H(l,k) = \sum_{k=1}^{2M} S(l,k) \log(S(l,k)) \text{..... .(4)}$$

H (*l*) is called entropy of the noisy speech signal, which is a quantitative measure of how certain the outcome of a random noisy speech signal.

Where

$$S(l,k) = \frac{Y_{energy}\ (l,k)+R(l)}{\sum_{k=1}^{2M} Y_{energy}\ (l,k)+R(l)} \text{ ……... ….(5)}$$

$Y_{energy}(l, k)$ is the energy of noisy speech.

$R(l) = \max_k\{Y(l,k)\}$is a constant used to stabilize the S(*l*, k).

3.3 Determination of threshold conditions

It is highly dependent on SNR of noisy speech and controlled by max (y (*l*, k)). The stabilization parameter R (*l*) is adjusted in each frame in order to rapid change in noise power spectrum.

Let $T_1(l)=r_1E[H_{avg}(l)]$ …….....(6)

$T_2(l)=r_2E[H_{avg}(l)]$…….(7)

WhereT_1 (*l*), T_2 (*l*) are thresholds to classify noisy speechinto Non speech, original speech & quasi speech. r_1 r_2 are 0.98, 0.95 respectively which are determined by experiment.

E [$H_{avg}(l)$] means an average over the recent number of initial silence frames including lth frame.

If H_{avg} (*l*) > T_1 (*l*-1) thenT_1 (*l*), T_2 (*l*) are updated by (6) and (7).

3.4 Classifying noisy speech into speech present / absent frames

The power spectrum of the noisy speech is equal to the sum of the speech power spectrum and noise power spectrum since speech and the background noise are assumed to be independent. Also in any speech sentence there are pauses between words which do not contain any speech. Those frames will contain only background noise. The noise estimate can be updated by tracking those noise-only frames. To identify those frames, a simple procedure is used which calculates the ratio of noisy speech power spectrum to the noise power spectrum at three different frequency bands.

3.5 Update of noise estimate for Noise estimation for Non – speech

The noise estimate is then updated with a constant smoothing factor if the frame is classified as speech absent frame. This rule can be stated as follows

If $H_{avg}(l) > T_1(l)$ then

$$\widehat{N}(l,k) = \alpha\widehat{N}(l-1,k) + (1-\alpha)|Y(l,k)|^2 \text{ ....… (8)}$$

N (*l*, k) is the noise spectrum estimated in non-speech frame. α is known as forgetting factor (or) look – ahead factor (or) smoothing factor lies between 0.7 to 0.9 [9],[10].

3.6 Update of noise estimate for Noise estimation for Quasi – speech

The proposed algorithm for updating noise spectrum in speech present frames was based on classifying speech present or absent frequency bins in each frame. This was done by tracking the local minimum of noisy speech and then deciding speech presence in each frequency bin separately using the ratio of noisy speech power to its local minimum. Based on that decision a frequency-dependent smoothing parameter was calculated to update the noise power spectrum.

The purpose of introducing quasi – speech frame is to analyze noisy speech signal accurately.

If T_2 (l) <H_{avg} (l) < T_1 (l) then

$$\widehat{N}(l,k) = P(l,k)\widehat{N}(l-1,k) + \left(1 - P(l,k)\right) \text{……... ………} (9)$$

3.7 Tracking the minimum of noisy speech

For tracking the minimum of the noisy speech power spectrum over a fixed search window length, various methods were proposed [5].These methods were sensitive to outliers and also the noise update was dependent on the length of the minimum-search window. For tracking the minimum of the noisy speech by continuously averaging past spectral values, a different non-linear rule is used.

If $P_{min}(l-1,k) \le P(l,k)$ then

$$P_{min}(l,k) = \gamma P_{min}(l-1,k) + \frac{1-\gamma}{1-\beta}\left(P(l,k) - \beta P(l-1,k)\right) \text{..} (10)$$

If $P_{min}(l-1,k) > P(l,k)$ then

$$P_{min}(l,k) = P(l,k) \text{……………………................…………} (11)$$

γ=0.998, β=0.96 & **ξ**=0.6 to 0.7 were determined experimentally. In practical implementation smoothing parameter in (11) whose maximum value is 0.96 to avoid deadlock for $\check{r}(l,k) = 1$.

3.8 Speech presence probability

Let the ratio of noisy speech power spectrum and its local minimum be defined as

$$P_{sp}(l,k) = \frac{|Y(l,k)|^2}{P_{min}(l,k)} \text{. …………………………………………..}(12)$$

This ratio is then compared with a frequency-dependent threshold, and if the ratio is greater than the threshold, it is taken as speech present frequency bin else it is taken as speech absent frequency bin. This is based on the principle that the power spectrum of noisy speech will be nearly equal to its local minimum when speech is absent. Hence smaller the ratio defined in Eq. (12) the higher the possibility that it will be a noise region or vice versa. This can be summarized as follows.

If $S_r(\lambda,k) > \delta(k)$, then

$I(\lambda,k) = 1$ Speech present

Else

$I(\lambda,k) = 0$ Speech absent

End

Where δ(k) is the frequency dependent threshold whose optimal value is determined experimentally. Note that, a fixed Where δ(k) is the frequency dependent threshold whose optimal value is determined experimentally. Note that, a fixed threshold was used in place of δ(k) for all frequencies. From the above rule, the speech-presence probability $p(\lambda,k)$ is updated using the following first-order recursion.

3.9 Calculating frequency dependent smoothing constant

By using the above speech-presence probability estimate, the time–frequency dependent smoothing factor is computed as follows.

$$P(l,k) = \alpha_d + (1-\alpha_d)\, P_{sp}(l,k) \text{ ……………………………}(13)$$

Where α_d is a constant.

P_{min} (l, k) is minimum noisy speech spectrum and it is updated by the following equation.

$$P(l,k) = \xi P(l-1,k) + (1-\xi) \text{………….........................} (14)$$

Where **ξ** is smoothing factor, P (l, k) is average noise spectrum.

$$p(\lambda,k) = \alpha_p\, p(\lambda-1,k) + (1-\alpha_p) I(\lambda,k) \text{ ………}(15)$$

Where α_p is a smoothing constant, the above recursive implicitly exploits the correlation for speech presence in adjacent frames.

In practical implementation smoothing parameter in (11) whose maximum value is 0.96 to avoid deadlock for $\check{r}(l,k) = 1$.

$$\check{r}(l,k) = \widehat{N}(l-1,k)/\sigma_{N^2}(l,k) \text{……………………………..} (16)$$

Eq. (16) is a smoothed version of posterior SNR.

Wiener filter is used to produce estimated pure signal from a given noise speech signal [3]. Wiener filter is formulated to map an input signal to an output that is as close to a desired signal as possible and its structure is illustrated in Figure 3.

It is a class of optimum linear filter, involves linear estimation of desired signal by minimizing minimum mean square error shown in Figure4. We consider some applications of the Wiener filter in reducing broadband additive noise, in time-alignment of signals in multichannel or multisensory systems, and in channel equalization. In the frequency domain, the Wiener filter output $\hat{X}(f)$ is the product of input signal Y(f) and the filter frequency response W(f) .

$$\hat{X}(f)=Y(f).W(f) \quad \text{.....(17)}$$

The estimation error signal E(f) is defined as the difference between the desired signal X(f)and the filter output $\hat{X}(f)$

$$E(f)= X(f)-\hat{X}(f) \quad \text{.....(18)}$$

The mean square error at a frequency f is given by

$$[|E(f)|^2]=E[|X(f)-\hat{X}(f)|^2] \quad \text{.....(19)}$$

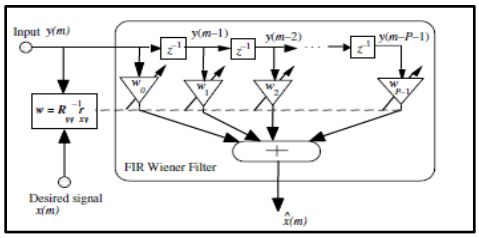


Figure 3: Illustration of Weiner structure.

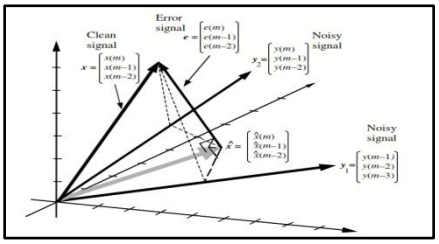


Figure 4: The least square error projection of a desired signal vector x onto a plane containing the input signal vectors y_1 andy_2.

IV. IMPLEMENTATION & RESULTS

The proposed algorithm implemented and analyzed by MATLAB. The performance of the proposed algorithm is compared with weighted average algorithm with the following spectral gain function [2].

$$G(\lambda,k)=\frac{C(\lambda,k)}{C(\lambda,k)+\mu_k D(\lambda,k)} \quad \text{.....(20)}$$

Where $C(\lambda,k)$is the estimated clean speech spectrum computed from the noisy speech and noise estimates as follows

$$C(\lambda,k)=\max\left\{|Y(\lambda,k)|^2-D(\lambda,k),\nu D(\lambda,k)\right\} \quad \text{.... (21)}$$

$G(\lambda,k)$Represents estimated clean speech spectrum, μ_k represents adaptation factor determined from the posteriori segmental SNR. Where $\nu=0.001$is a small positive number. The max operation is used to ensure positive values for the estimated clean speech spectra.

The over subtraction factor μ_k in Equation (16) is determined from the a posteriori segmental SNR according to Figure.5.

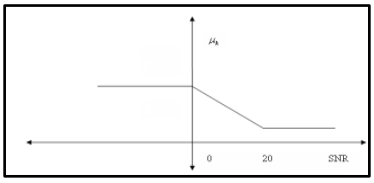


Figure5: Plot of the multiplication factor μ_k for different values of a posteriori SNR of noisy speech.

4.1 Time-domain SNR measures

The time-domain segmental SNR (SNR seg) measure was computed is given by

$$SNR_{seg} = \frac{10}{M}\sum_{m=0}^{\omega-1}\log\frac{\sum_{n=N_m}^{N_m+N-1}x^2(k)}{\sum_{n=N_m}^{N_m+N-1}\left(x(k)-\hat{x}(k)\right)^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots(22)$$

Where $x(k)$ the input is (clean) signal, $\hat{x}(k)$ is the processed (enhanced) signal, N is the frame length and M is the number of frames in the signal.

Experimental values of segmental SNR are given in the Figure5 which shows higher values for the proposed algorithm and also evaluated spectrograms and timing waveforms. Timing wave forms and spectrograms of noise corrupted speech by vehicle noise and enhanced speech signal with weighted average technique & proposed algorithm (SERA) are shown in the Figure6 & Figure7 respectively. The Experimental results of proposed noise estimation algorithm works effectively compared with weighted average algorithm.

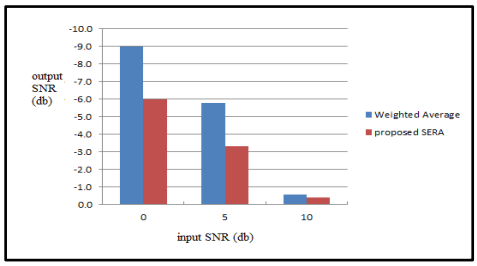
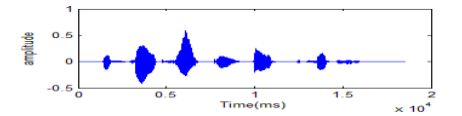
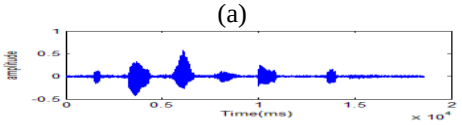


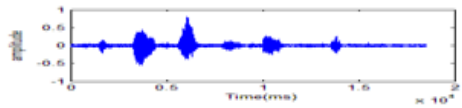
Figure 5: Segmental SNR Values (db)



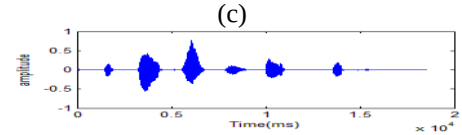
(a)



(b)

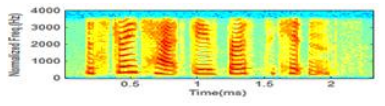


(c)

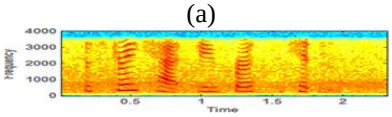


(d)

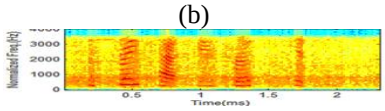
Figure6: Timing wave forms of (a) clean speech signal (b) noise corrupted speech signal and enhanced signals with (c) weighted average method (d) proposed SERA algorithm.



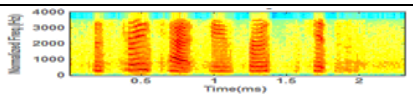
(a)



(b)



(c)



(d)

Figure 7: Spectrograms of (a) original speech signal (b) noise corrupted speech signal and enhanced signals with (c) weighted average method (d) proposed method.

V. CONCLUSION

This paper focused on the issue of noise estimation for enhancement of noisy speech signal in real time environments. Noisy speech signal is divided into non-speech, quasi-speech, & pure speech based on entropy. Initial noise power spectrum is updated and estimated in each frame based on classification. The noise estimate was updated continuously in every frame using time–frequency smoothing factors calculated based on speech-presence probability in each frequency bin of the noisy speech spectrum. The speech-presence probability was estimated using the ratio of noisy speech power spectrum to its local minimum. Unlike other methods, the update of local minimum was continuous over time and did not depend on some fixed window length. Hence the update of noise estimate was faster for very rapidly varying non-stationary noise environments.

Performance of the proposed algorithm is better compared to existing algorithms. This was confirmed by formal listening tests that indicated significantly higher preference for my proposed algorithm compared to the other existing noise-estimation algorithms.

REFERENCES

[1]. R.SundarRajan and C.L.Philipos, "A Noise Estimation Algorithm for Highly Non-stationary environments," speech communication, Vol.48, PP.220-231, 2006.
[2]. Ch.V.RamaRao, "Noise Estimation for Speech for Enhancement in Non-stationary environments – a new method", World Academy of Science, Engineering and Technology, Vol.70, PP.739-740, 2010.
[3]. T.Lalith Kumar and R.SundarRajan, "Speech Enhancement using Adaptive Filters", VSRD-JJEEE, Vol.2 (2), PP.92-99, 2012.
[4]. C.GaneshBabu and P.T.Vanathi, "Performance Analysis of Voice Activity Detection Algorithm for Robust SpeechRecognition System under Different Noisy Environment", Journal of Scientific & Industrial Research, Vol.69, PP.515-522, July 2010.
[5]. G.Poblinger, "Computationally Efficient Speech Enhancement by Spectral Minima Tracking in Subbands", Proc. Euro Speech 2, PP.1513-1516, 1995.
[6]. R.Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics", IEEE Trans speech Audio Process, PP-504-512, 2001.
[7]. P.Loizou, R.Sundarajan and Huy, "Noise Estimation Algorithm with Rapid Adaption for Highly Non-stationary Environments", prec. IEEE international conference on acoustic speech signal Proc, 2004.
[8]. J.Sohn and N.Kim, "Statistical Model Based Voice Activity Detection", IEEE signal ProcLett, 6(1), PP-1-3, 1999.
[9]. S.Tanyer and H.Ozer, "Voice Activity Detection in Non – Stationary Noise", IEEE Speech Audio Proc. 8(4), PP.478-482, 2000.
[10]. P.Loizou, "A Noise Estimation Algorithm with Rapid Adaption for Highly Non-stationary environments, Speech Communication Science direct, PP-220-231, 2006.
[11]. Anu Radha and R.Fuknu, "Noise Estimation Algorithms for Speech Enhancements in Highly Non-stationary Environments", IJCSI, Vol.8, PP.39-44, 2011.
[12]. Cohen, I. and Berdugo, B., "Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement," IEEE Signal Proc. Letters, vol. 9, no. 1, pp. 12-15, Jan. 2002.
[13]. Hirsch, H. G. and Ehrlicher, C., "Noise Estimation Techniques for Robust Speech Recognition," in Proc. 20th IEEE Int. Conf. Acoustics, Speech, Signal Processing, Detroit, MI, pp. 153-156, May 8-12, 1995.