



Review of MMSE Estimator for Speech Enhancement

Savita Hooda and Smriti Aggarwal

Maharishi Markandeshwar University,
Mullana (Ambala), INDIA

Abstract: *Speech enhancement aims to improve speech quality by using various techniques and algorithms. The MMSE estimator is one of the algorithms proposed for removal of additive background noise. It is a single channel speech enhancement technique for the enhancement of speech degraded by additive background noise. Background noise can affect our conversation in a noisy environment like in streets or in a car, when sending speech from the cockpit of an airplane to the ground or to the cabin and can affect both quality and intelligibility of speech. With the passage of time Spectral subtraction has undergone many modifications. This is a review paper and its objective is to provide an overview of MMSE estimator that have been proposed for enhancement of speech degraded by additive background noise during past decades. Section I gives the Introduction to Speech enhancement. Section II gives the various speech enhancement methods. Section III gives a Literature review on MMSE estimator speech enhancement. Section IV gives proposed method which the author proposed after a literature survey. .*

Keywords: *MMSE estimator; Speech enhancement; Speech enhancement methods and Speech signals SNR estimation.*

I. INTRODUCTION

Speech is associated with many definitions. However in general, it can be defined as a mechanism of expressing thoughts and ideas using vocal sounds [1,2]. In humans, speech sounds are produced when breath is exhaled from the lungs & causes either a vibration of the vocal cords (when speaking vowels) or restriction in the vocal tract (when speaking consonants) [1]. In general, speech production and perception is a complex phenomenon and uses many organs such as lungs, mouth, nose, ears & their associated controlling muscles and brain. To produce a variety of speech signals, the shape of vocal tract is varied in accordance with the vibrations of vocal cords. More specifically, the speech is produced by the cavity between the vocal cords & lips and acts as a resonator that spectrally shapes the periodic input, much like the cavity of a musical wind instrument. The resonator is formed by combining the oral and pharyngeal cavities, in situations when the velum is closed. The tongue is used to change the shape of vocal tract as it can be moved up, down, forward and back. Thus, it can also be used to construct the tract for the production of consonants. By moving the lips outward, the length of the vocal tract can be increased. Vocal tract is situated in larynx called Adam's apple. The vocal tract is at rest when open. Its tension and elasticity can be varied; can be made thicker and thinner, shorter or longer & can be either closed, open wide or held in some position. The oral tract is highly mobile & the position of the pharynx, palate, lips affect the speech sound made [3,4].

The bandwidth of speech signals is around 4 KHz. However, the human ear can perceive sounds, with frequencies in between 20 Hz to 20 KHz. The signals with frequencies below 20 Hz are called subsonic or infrasonic sounds, and above 20 KHz are called ultrasonic sounds. The noise produced by various ambient sources such as vehicles also lies in this frequency range. Therefore, speech signals get easily distorted by the ambient noise or AWGN. These distorted or degraded speech signals are called noisy speech signals. This paper focuses on speech processing (in particularly speech enhancement) of the noisy speech signals. The key speech processing techniques include spectral subtraction approach [5,6,7], signal subspace approach [8,9], adaptive noise cancelling and iterative Wiener filter. The performances of these techniques depend on the quality and intelligibility of the processed speech signal. The prime focus of all these techniques is to improve speech signal-to-noise ratio. Among the above mentioned technologies, spectral subtraction is the earliest method for enhancing speech, degraded by noise. This technique estimates the spectrum of the clean (noise-free) signal by subtracting the estimated noise magnitude spectrum from the noisy signal magnitude spectrum; while, keeping the phase spectrum of noisy signal. These techniques are reported to have several drawbacks such as residual noise and musical sounds. Therefore this paper, is a review paper and its objective is to provide an overview of MMSE estimator that have been proposed for enhancement of speech degraded by additive background noise during past decades.

II. SPEECH ENHANCEMENT METHODS

1. Model based speech enhancement

This approach is applied as a two-step procedure: a) the statistics of signal and noise are first estimated from training data of speech and noise, and b) then use this estimated statistics along with currently available distortion measures to address the speech enhancement problem.

2. Subtractive type algorithms

In these speech processing algorithms, the input to the system is the noisy speech signal. The frame-by-frame analysis is performed and the Short-term Fourier Transform of the signal with Overlap and Add (OLA) is usually the most commonly used method to determine the estimate of speech signal. In these methods, the magnitude of speech spectrum is usually modified according to the estimated noise signal measured during speech pauses/silences period.

3. Voice activity detection

The process of decimating between voice activity (i.e. speech presence) and silence (i.e. speech absence) is called voice activity detection. VAD algorithms extract features (e.g. short-time energy, zero crossings, periodicity measure) from the input signal and compares against a threshold value, usually determined during speech absent periods. VAD algorithms generally output a binary decision on a frame by frame basis, where a frame may last approximately 20-40 msec. The VAD is mostly used in telephonic communication, audio conferencing and digital cordless telephone system. However, these are not suitable in low SNR conditions

4. Minimal tracking algorithms

These algorithms are used to estimate the power spectral density of non-stationary noise, when a noisy speech signal is given. These algorithms can be combined with any speech enhancement algorithm, requiring noise power spectral density estimate. It tracks spectral minima in each frequency band without any distinction between speech activity and speech pause. In these algorithms, an unbiased noise estimator is developed, based on the optimally smoothed power spectral density estimate and the analysis of statistics of spectral minima.

5. Minimum statistics noise estimation

Minimum statistics is used to track the minimum of noisy speech power spectrum within a finite window (analysis segment). These are better than VAD algorithms, as these yields better quality and improvement in speech intelligibility. In addition, tracking of minima in each frequency bin helped preserve the weak voiced consonants (e.g. m and n), which might otherwise be classified as noise by most VAD algorithms as their energy is concentrated in a small number of frequency bins (low frequencies). However on its downside, it is unable to respond to fast changes of the noise spectrum.

6. Continuous spectral minimum tracking algorithm

In this method, in contrast to using a fixed window for tracking the minimum of noisy speech, the noise estimate is updated continuously by smoothing the noisy speech power spectra in each frequency bin using a non-linear smoothing rule. For minimum tracking of noisy speech power spectrum, a short time smoothed version of periodogram of noisy speech is computed. Its key advantages over minimum statics algorithm are its low computational cost and the non-linear tracking used, maintains the continuous psd smoothing without making any distinction between speech present and absent segments. The noise estimate increases whenever the noise speech power spectrum increases, irrespective of the changes in noise power level. The key disadvantages include very narrow peaks in speech spectrum resulting in overestimation of noise during speech activity.

7. Weighted spectral average algorithms

A different and simple approach to recursive averaging noise estimation is defined by the fact that each spectral component is having a different effective SNR. Consequently, estimation and update of individual frequency band of the noise spectrum can be done whenever the effective SNR at a particular frequency band is extremely low. In this algorithm, noise spectrum is estimated as a weighted average of past noise estimates and the present noisy speech spectrum. In this approach, the smoothing factor is kept fixed and the decision as to whether the noise spectrum should be updated or not is based on the comparison of estimated posteriori SNR to a threshold. The weighted spectral average algorithm performs moderately well compared to continuous spectra algorithm. However, its disadvantages are that it occasionally overestimates the noise level particularly when low SNR segments preceded by high energy segments.

8. Minima controlled recursive algorithm

In minima controlled recursive algorithm (MCRA), the noise estimation is given by averaging past spectral power values and using a smoothing parameter that is adjusted by the signal probability in sub-bands. Presence of speech in sub-bands is determined by the ratio of the local energy of the noisy speech and its minimum within a specified time window. The noise estimate is computationally efficient, robust with respect to signal to noise ratio and type of underlying additive noise and

characterized by the ability to quickly follow abrupt changes in the noise spectrum. MCRA noise estimation is formulated using a detection theory framework. Its key advantages include that the time smoothing factor take binary values either 0 or 1 and the estimated noise psd follow the spectral minima. The key disadvantage include that the noise psd estimate may lag, particularly when the noise power is rising, by as many as D frames from the true noise psd.

9. Improved minima controlled recursive algorithm

In this algorithm, further refinements are done to the MCRA algorithm. The conditional speech presence probability $p(\square, k)$ is obtained after substituting log likelihood ratio. This algorithm uses posteriori and priori SNRs. It is advantageous as the delay may be smaller because the recursive averaging is carried out instantaneously. However, it is disadvantageous as the improved MCRA yields smaller estimation errors for several types of noise.

III. LITERATURE REVIEW

Cohen et. al. [10] proposed an improved MCRA noise variance estimator improvements. For objective results, the improvement in segmental SNR was reported for white Gaussian noise, car interior noise and F16 cockpit noise for various noise levels from -5 to 10 dB. In all the cases, the MCRA approach showed a higher performance compared to weighted averaged method. Also, the methods were compared with a subjective study of spectrogram of enhanced speech and informal listening tests. The tracking ability of the algorithms was tested by authors by comparing the spectrograms of enhanced speech for a signal recorded in a car by suddenly turning on the defroster in full.

Berdugo, B. et. al. [11] proposed a new approach called minima controlled recursive averaging (MCRA) for noise estimation. The noise estimate was updated by averaging the past spectral values of noisy speech which was controlled by a time and frequency dependent smoothing factors. These smoothing factors were calculated based on the signal presence probability in each frequency bin separately. This probability was in turn calculated using the ratio of the noisy speech power spectrum to its local minimum calculated over a fixed window time.

R. Martin et. al. [12] described that Gaussian statistical model provides a good approximation for the noise DFT coefficients. For speech signals, however, where typical DFT frame sizes used in mobile communications are short (10ms -40ms) this assumption is not well fulfilled. It is valid only if the DFT frame size is much longer than the span of correlation of the signal under consideration.

Cohen et. al.[13] presented methods that incorporated the fact that speech might not be present at all frequencies and at all times. Authors provided an estimate of the probability that speech is absent at a particular frequency bin. In this research, MMSE magnitude estimator under the assumed Laplacian model and uncertainty of speech presence has been described & considered a two-state model for speech events. According to this two state model, either speech is present at a particular frequency bin (hypothesis H1) or not (hypothesis H0).

R. Martin et. al. [14] proposed a new estimator, in which the real and imaginary parts of the clean signal were estimated in the MMSE sense conditional on the real and imaginary parts of the observed noisy signal. This estimator, however, is not the optimal spectral amplitude estimator but clean signal & noise were modeled by a combination of Gaussian, Gamma and Laplacian distributions.

C. Breithaupt et. al. [15] described that the real and imaginary part of the speech coefficients are better modeled with Laplacian and Gamma densities. This observation led researchers to derive a similar optimal MMSE STSA estimator but based on more accurate models, Laplacian and/or Gamma. However, the derivation of such an estimator is complicated leading some people to seek alternative techniques to compute the MMSE STSA estimator.

Malah et. al. [16] derived the MMSE STSA estimator, based on modeling speech and noise spectral components as statistically independent Gaussian random variables. Authors analyzed the performance of the proposed STSA estimator and compared it with a STSA estimator derived from the Wiener estimator. Authors also examined the MMSE STSA estimator under uncertainty of signal presence in the noisy observations.

Y. Ephraim et. al. [17] derived a short-time spectral amplitude (STSA) estimator for speech signals which minimizes the mean square error of the log-spectra (i.e., the original STSA and its estimator) and examined it in enhancing noisy speech. This estimator is also compared with the corresponding minimum mean-square error STSA estimator derived previously.

Xuchu et. al. [18] have proposed an algorithm (fast noise tracking algorithm) improved over MMSE-LSA algorithm. It suits non-stationary noise environments better than the traditional algorithm. The main part of this method is the estimation of noise, which is updated using time-frequency smoothing factors calculated based on speech-present probability in each frequency bin of the noisy speech spectrum. According to authors, it can keep up with the noise change closely. Authors mentioned that their objective tests showed that the proposed algorithm is superior to the traditional methods in noise-tracking and mean opinion score.

Israel Cohen et. al. [19] proposed a minima controlled recursive averaging (MCRA) approach for noise estimation. 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 sub bands The noise estimate is computationally efficient, robust with respect to the input signal-to-

noise ratio (SNR) and type of underlying additive noise, and characterized by the ability to quickly follow abrupt changes in the noise spectrum.

Israel Cohen et. al. [20] described noise spectrum estimation as a fundamental component of speech enhancement and speech recognition systems. Authors presented an Improved Minima Controlled Recursive Averaging (IMCRA) approach, for noise estimation in adverse environments involving non-stationary noise, weak speech components, and low input signal- to- noise ratio. The noise estimate is obtained by averaging past spectral power values, using a time-varying frequency-dependent smoothing parameter that is adjusted by the signal presence probability.

IV. PROPOSED METHOD

The proposed extended version of minimum mean-square error (MMSE) algorithm to reduce noise is described below.

A. The Gaussian Based MMSE STSA Estimator

In order to derive the MMSE STSA estimator, the *a priori* probability distribution of the speech and noise Fourier expansion coefficients are assumed, as these are unknown in practice. Let $y(n) = x(n)+d(n)$ be the sampled noisy speech signal consisting of the clean signal $x(n)$ and the noise signal $d(n)$. Taking the short-time Fourier transform of $y(n)$, to have:

$$Y(\omega_k) = X(\omega_k) + D(\omega_k) \quad (1)$$

Where, $\omega_k = 2\pi k/N, k=0,1,2,\dots,N-1$, and N is the frame length. The above equation can also be expressed in polar form as

$$Y_k e^{j\theta_y(k)} = X_k e^{j\theta_x(k)} + D_k e^{j\theta_d(k)} \quad (2)$$

As, the spectral components are assumed to be statistically independent, the MMSE amplitude estimator \hat{X}_k can be derived from $Y(\omega_k)$ only. That is,

$$\begin{aligned} \hat{X}_k &= E\{X_k | Y(\omega_0), Y(\omega_1), \dots\} \\ &= E\{X_k | Y(\omega_k)\} = \frac{\int_0^\infty \int_0^{2\pi} x_k p(Y(\omega_k) | x_k, \theta_k) p(x_k, \theta_k) d\theta_k dx_k}{\int_0^\infty \int_0^{2\pi} p(Y(\omega_k) | x_k, \theta_k) p(x_k, \theta_k) d\theta_k dx_k} \end{aligned} \quad (3)$$

Where, $\theta_k = \theta_x(k)$. Under the assumed Gaussian model $p(Y(\omega_k) | x_k, \theta_k)$ and $p(x_k, \theta_k)$ are given by

$$p(Y(\omega_k) | x_k, \theta_k) = \frac{1}{\pi \lambda_d(k)} \exp \left\{ -\frac{1}{\lambda_d(k)} |Y_k - X_k e^{j\theta_x(k)}|^2 \right\} \quad (4)$$

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

Where, $\lambda_x(k) \triangleq E\{|X_k|^2\}$, and $\lambda_d(k) \triangleq E\{|D_k|^2\}$ are the variances of the k th spectral component of the speech and noise respectively. Substituting Eq. 4 and Eq. 5 into Eq. 3 gives

$$\hat{X}_k = \Gamma(1.5) \frac{\sqrt{v_k}}{\gamma_k} \exp \left(-\frac{v_k}{2} \right) \left[(1 + v_k) I_0 \left(\frac{v_k}{2} \right) + v_k I_1 \left(\frac{v_k}{2} \right) \right] Y_k \quad (6)$$

Where $\Gamma(\cdot)$ denotes the gamma function, with $\Gamma(\cdot) = \sqrt{\pi}/2$, and $I_0(\cdot)$ and $I_1(\cdot)$ denote the modified Bessel functions of zero and first order, respectively. The variable, v_k is defined by

$$v_k \triangleq \frac{\xi_k}{1 + \xi_k} \gamma_k \quad (7)$$

Where ξ_k and γ_k are interpreted as the *a priori* and *a posteriori* signal-to-noise ratio (SNR), respectively and are defined by

$$\xi_k \triangleq \frac{\lambda_x(k)}{\lambda_d(k)} \quad (8)$$

$$\gamma_k \triangleq \frac{Y_k^2}{\lambda_d(k)} \quad (9)$$

At high SNR, $\xi_k \gg 1$ and $\gamma_k \gg 1$; therefore, the estimator can be simplified as:

$$\hat{X}_k \triangleq \frac{\xi_k}{1 + \xi_k} \gamma_k \quad (10)$$

The above is called as Wiener estimator. Because of its several advantages, the MMSE estimation of speech spectrum have received considerable attention; however, the existing related methods have been reported several limitations either on the underlying assumptions or derivation of the estimators. Therefore, a Laplacian based MMSE estimator is presented below.

B. The Laplacian based MMSE STSA estimator

The basic idea of Laplacian based MMSE STSA estimator is to find the optimal estimate of the modulus of speech signal DFT components. It is based on the assumption that the real and imaginary parts of these components are modeled by a Laplacian distribution. The noise signal DFT components are assumed to be Gaussian distributed. The Laplacian estimator

has been discussed before; however, it is presented here to determine the speech-presence uncertainty. It is because in a typical speech signal, it is very likely that speech is not present at all times. It is also because running speech contains a great deal of pauses, even during speech activity. The stop closures, for example, which are brief silence periods occurring before the burst of stop consonants, often appear in the middle of a sentence. Also, speech might not be present at a particular frequency even during voiced speech segments. Therefore, a two-state model for speech events is considered, which is based on the fact that either speech is present at a particular frequency bin (hypothesis H_1) or that is not (hypothesis H_0). This is expressed mathematically using the following binary hypothesis model:

$$H_0^k: \text{speech absence: } Y(\omega_k) \quad (11)$$

$$H_1^k: \text{speech present: } Y(\omega_k) = X(\omega_k) + D(\omega_k) \quad (12)$$

To incorporate the above binary model to an MMSE estimator, a weighted average of two estimators is used. So, if the original MMSE estimator had the form $\hat{X}_k = E(X_k|Y(\omega_k))$, then the new estimator, has the form:

$$\hat{X}_k = E(X_k|Y(\omega_k), H_1^k)P(H_1^k|Y(\omega_k)) + E(X_k|Y(\omega_k), H_0^k)P(H_0^k|Y(\omega_k)) \quad (13)$$

Where $P(H_1^k|Y(\omega_k))$ denotes the conditional probability that speech is present in frequency bin k , given the noisy speech spectrum. Similarly $P(H_0^k|Y(\omega_k))$ denotes the conditional probability that speech is absent given the noisy speech spectrum. The term $E(X_k|Y(\omega_k), H_0^k)$ in the above equation is zero since it represents the average value of X_k given the noisy spectrum $Y(\omega_k)$ and the fact that speech is absent. Therefore, the MMSE estimator mentioned above reduces to

$$\hat{X}_k = E(X_k|Y(\omega_k), H_1^k)P(H_1^k|Y(\omega_k)) \quad (14)$$

The term $P(H_1^k|Y(\omega_k))$ can be computed using Bayes rule. The MMSE estimator of the spectral component at frequency bin k is weighted by the probability that speech is present at that frequency:

$$P(H_1^k|Y(\omega_k)) = \frac{p(Y(\omega_k)|H_1^k)P(H_1)}{p(Y(\omega_k)|H_1^k)P(H_1^k) + p(Y(\omega_k)|H_0^k)P(H_0)} = \frac{\Lambda(Y(\omega_k)|q_k)}{1 + \Lambda(Y(\omega_k)|q_k)} \quad (15)$$

Where $\Lambda(Y(\omega_k), q_k)$ is the generalized likelihood ratio defined by:

$$\Lambda(Y(\omega_k), q_k) = \frac{1 - q_k}{q_k} \frac{p(Y(\omega_k)|H_1^k)}{p(Y(\omega_k)|H_0^k)} \quad (16)$$

where $q_k = P(H_0^k)$ denotes the a priori probability of speech absence for frequency bin k . The a priori probability of speech presence i.e. $P(H_1^k)$ is given by $1 - q_k$. Theoretically, the optimal estimate under hypothesis H_0^k is identical to zero but a small nonzero value might be preferable for perceptual purposes. Under hypothesis H_0 , $Y(\omega_k) = D(\omega_k)$, and given that the noise is complex Gaussian with zero mean and variance $\lambda_d(k)$; it follows that $p(Y(\omega_k)|H_0^k)$ will also have a Gaussian distribution with the same variance, i.e.,

$$p(Y(\omega_k)|H_0^k) = \frac{1}{\pi\lambda_d(k)} \exp\left[-\frac{Y_k^2}{\lambda_d(k)}\right] \quad (17)$$

If $X(\omega_k)$ follows a Laplacian distribution, it is required to compute $p(Y(\omega_k)|H_1^k)$. Assuming independence between real and imaginary components, we have:

$$p(Y(\omega_k)|H_1^k) = p(z_r, z_i) = p_{z_r(k)}(z_r)p_{z_i(k)}(z_i) \quad (18)$$

where $z_r(k) = \text{Re}\{Y(\omega_k)\}$ and $z_i(k) = \text{Im}\{Y(\omega_k)\}$. Under hypothesis H_1 , the pdf of $Y(\omega_k) = X(\omega_k) + D(\omega_k)$ needs to be derived, where $X(\omega_k) = X_r(\omega_k) + jX_i(\omega_k)$ and $D(\omega_k) = D_r(\omega_k) + jD_i(\omega_k)$. The pdfs of $X_r(\omega_k)$ and $X_i(\omega_k)$ are assumed to be Laplacian and the pdfs of $D_r(\omega_k)$ and $D_i(\omega_k)$ are assumed to be Gaussian with variance $\sigma_d^2/2$ and zero mean

V. REFERENCES

- [1] R.C. Nongpiur, "Impulse Noise Removal In Speech Using Wavelets" ICASSP, IEEE, 2008.
- [2] X. Hou, S. Guo, H. Cui, K. Tang and Ye Li, "Speech Enhancement for Non-Stationary Noise Environments", ISBN 978-1-4244-4994.
- [3] Meng Joo Er, "Adaptive Noise Cancellation Using Enhanced Dynamic Fuzzy Neural Networks", IEEE Trans. Fuzzy Systems, vol. 13, No. 3, June 2005, pp 331-342.
- [4] C. Plapous, C. Marro and P. Scalart, "Improved signal-to-noise ratio estimation for speech enhancement". IEEE Trans. Audio Speech Lang. Process., pp. 2098-2108, 2006.
- [5] C. Plapous, C. Marro and P. Scalart, "A two-step noise reduction technique," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process, Montréal, QC, Canada, May 2004, vol. 1, pp. 289-292.
- [6] Z. Chen, "Simulation of Spectral Subtraction Based Noise Reduction Method", International Journal of Advanced Computer Science and Applications, vol. 2, No. 8, 2011.
- [7] M. Hasan, S. Salahuddin and M. Khan, "A modified a priori SNR for speech enhancement using spectral subtraction rules," IEEE Signal Processing, vol. 11, No. 4, pp. 450-453, Apr. 2004.
- [8] C. Avendano "Acoustic Echo Suppression in STFT Domain" IEEE WASPA/OI, Mohonk, 2001.
- [9] S. Ou, X. Zhao and Y. Gao, "Speech Enhancement Employing Modified a Prior SNR Estimation", pp. 827-831, 2007.
- [10] Cohen, I., "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging", IEEE Trans. on speech and audio processing, vol. 11, no. 5, pp. 466-475, Sept. 2003..
- [11] Berdugo, B. and Cohen, I., "Noise estimation by minima controlled recursive averaging for robust speech enhancement", IEEE Signal Proc. Letters, vol. 9, no. 1, pp. 12-15, Jan. 2002.
- [12] R. Martin, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimation," in Proc. IEEE Int. Conf. Acoust.,

Speech, Signal Processing, pp. 504–512, 2002.

- [13] I. Cohen, “Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator,” *IEEE Signal Processing Lett.*, vol. 9, pp. 113-116, Apr. 2002.
- [14] R. Martin and C. Breithaupt, “Speech enhancement in the DFT domain using Laplacian speech priors,” in *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, pp. 87–90, Sept. 2003
- [15] C. Breithaupt and R. Martin, “MMSE estimation of magnitude-squared DFT coefficients with super-gaussian priors,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 848-851, 2003.
- [16] Malah, D., Cox, R.V. and Accardi, A.J., “Tracking speech-presence uncertainty to improve speech enhancement in non-stationary noise environments,” *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp.789-792, 15-19 Mar 1999
- [17] Y. Ephraim,” *Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator*”, *IEEE Transactions On Acoustics, Speech and Signal Processing* 0096-35 18/8S/0400-0443, 1985
- [18] H. Xuchu, S. Guo, H. Cui, K. Tang and Y. Li, “Speech Enhancement for Non-Stationary Noise Environments”, *International Conference on Information Engineering and Computer Science, ICIECS 2009*, pp. 1-3, 19-20 Dec. 2009.
- [19] I. Cohen and B. Berdugo, “Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement”, *IEEE Signal Processing letters*, vol. 9, No. 1, January 2002.
- [20] I. Cohen, “Noise Spectrum Estimation in Adverse Environments: Improved Minima Controlled Recursive Averaging”, *IEEE Transactions on Speech and Audio Processing*, 2003.