A Study on Performance Analysis of Tamil Speech Enhancement using Spectral Subtraction and Adaptive Techniques

J Indra
Assistant Professor (SG),
Department of Electronics and Instrumentation Engineering,
Kongu Engineering College

N.Kasthuri, Ph.D
Professor,
Department of Electronics and Communication Engineering,
Kongu Engineering College

S Navaneetha Krishnan
PG student (AE),
Department of Electrical and Electronics Engineering,
Kongu Engineering College

ABSTRACT
Speech is produced when air from the lungs passes through the throat, the vocal cords, the mouth and the nasal tract. Speech processing is the study of the speech signals and the processing methods of these signals. Speech enhancement is a technique used to reduce the background noise present in the speech signal. It simply means the improvement in intelligibility and quality of degraded speech. The need to enhance speech signal arises in many situations in which the speech signal originates from noisy locations. The aim of the proposed method is to reduce the background noise present in the Tamil speech signal by using spectral subtraction and adaptive techniques. There has been no such works or efforts in the past in the context of Tamil speech enhancement in the literatures. Fifty Tamil speeches are taken as sample speech from the Tamil database [1] [2]. Sample noises such as pink noise, white noise and Volvo noise are taken. By using the spectral subtraction techniques such as Non-Linear, Multiband and Minimum Mean Square Error spectral subtraction and adaptive techniques such as Least Mean Square and Recursive Least Square methods, enhanced Tamil speech is obtained. Performance of the above two techniques are compared based on their Signal to Noise Ratio and Log Spectral Distance.

Keyword
MBSS, MMSE, RLS, LMS, SNR, LSD

1. INTRODUCTION
Speech signal processing refers to the acquisition, manipulation, storage, transfer and output of speech signals or vocal utterances by a computer. The main applications are enhancement, recognition, synthesis and compression of human speech. Speech signals are degraded in ways that limit their effectiveness for communication[3]. In such cases, digital signal processing techniques can be applied to improve the speech quality. It includes the removal of reverberation (or echoes) from speech, or the removal of noise from speech, or the restoration of speech.

2. SPEECH ENHANCEMENT
Speech signals are corrupted by various types of degradations[4]. Background noise, reverberation and speech from competing speaker(s) are among the most common degradations. When looking into perceptual quality and intelligibility degraded speech is poor. Degraded speech, therefore needs to be processed for enhancing its perceptual quality and intelligibility. Various methods have been proposed in the literature for the enhancement of degraded speech [5]. They can be classified into spectral processing and temporal processing methods. Degraded speech is processed in the frequency domain and in the time domain respectively for spectral processing and temporal processing methods [6]. The way of approaching towards speech enhancement varies depending on the type of degradation. For example, the type of processing for speech enhancement varies for background noise and reverberation.

3. SPECTRAL SUBTRACTION
One of the earliest proposed algorithms for noise reduction is the spectral subtraction algorithm. It is very easy to implement. The basic principle behind spectral subtraction is that the noise spectrum can be estimated and updated when speech signal is not present. Then it can be subtracted from the noisy speech signal in order to obtain clean speech signal spectrum [7].

It is assumed that noise is additive and the noise spectrum does not change with respect to time. This indirectly means that noise is stationary or it’s slowly time varying signal. The spectrum of noise does not change significantly between the updating periods. Figure 1 shows the general block diagram of spectral subtraction technique [8].

Let y(n) be the noise corrupted input speech signal. y(n) is composed of the clean speech signal x(n) and the additive noise signal d(n). Then y(n) is given by

$$y(n) = x(n) + d(n) \qquad (1)$$

Fourier domain representation of y(n) is written as,

$$Y(\omega) = X(\omega) + D(\omega) \qquad (2)$$

Magnitude and phase representation of Y(ω) can be expressed as

$$Y(\omega) = |Y(\omega)|e^{j\emptyset y} \qquad (3)$$

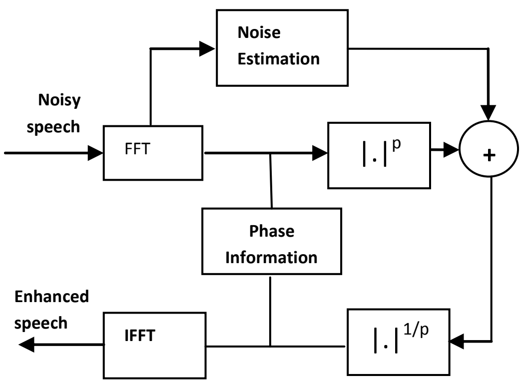

Fig 1: Block diagram of spectral subtraction

The different types of Spectral subtraction techniques are

1. Non-Linear Spectral Subtraction
2. Multiband Spectral Subtraction
3. MMSE Spectral Subtraction

1. Non-Linear Spectral Subtraction

This is proposed by Lockwood and Boudy. This method makes the over subtraction factor frequency dependent. So the subtraction process is non-linear. The assumption is all spectral components are not affected equally by noise. Certain types of noise may affect the low frequency region of the spectrum more than high frequency region [9]. So, it is suggested to use a frequency dependent subtraction factor for different types of noise. Due to this, subtraction process becomes nonlinear. Smaller values are subtracted at frequencies with high SNR levels and larger values are subtracted at frequencies with low SNR levels. The subtraction rule has the following form.

$$|X_e(\omega)| = |Y(\omega)| - \alpha(\omega)N(\omega) \qquad (4)$$

$$|Y(\omega)| > \alpha(\omega)N(\omega) + \beta|D_e(\omega)| \qquad (5)$$

2. Multiband Spectral Subtraction

In this approach the speech spectrum is divided into N overlapping bands. Then spectral subtraction is performed independently in each band. The processes of splitting the speech signal into different bands can be performed either in the frequency domain by using appropriate windows or in the time domain by using band pass filters [10]. The estimate of the clean speech spectrum in the ith band is obtained by

$$|X_{ei}(\omega_k)|^2 = |Y_i(\omega_k)|^2 - \alpha\delta_i|D_i(\omega_k)|^2 \qquad (6)$$

where

$$b_i < \omega_k < e_i$$

where $\omega_k = \frac{2\pi k}{N}$, k = 0, 1,…., N – 1 are the discrete frequencies, $|D_i(\omega_k)|^2$is the estimated noise power spectrum obtained during speech absent segment, α_I is the over subtraction factor of the ith band and δ_i is an additional band.

3. MMSE Spectral Subtraction

It is an effective method used for optimal selection of the subtractive parameters in the mean error sense [11]. At the same time it also minimizes the Mean Square Error (MSE). This approach optimizes the estimate of real spectral amplitudes. The spectral components are statistically independent and they follow Gaussian distribution. Consider the equation below:

$$|\boldsymbol{X}(\boldsymbol{\omega})|^2 = \boldsymbol{r_p}(\boldsymbol{\omega})|\boldsymbol{y}(\boldsymbol{\omega})|^{\boldsymbol{p}} - \boldsymbol{\alpha_p}(\boldsymbol{\omega})|\boldsymbol{D_e}(\boldsymbol{\omega})| \qquad \mathbf{(7)}$$

Where $r_p(\omega)$ and $\alpha_p(\omega)$ are the parameters of interest. The parameter $r_p(\omega)$ can be determined by minimizing the mean square error spectrum.

Signal- to- Noise Ratio (SNR)

One of the measures in noise reduction is the Signal-to-Noise Ratio (SNR). SNR is defined as the power ratio between speech signal (meaningful information) and background noise (unwanted information)[14].

$$SNR = 10\log_{10}\left(\frac{P_{Signal}}{P_{noise}}\right) \qquad (8)$$

Log Spectral Distance (LSD)

The Log-Spectral Distance (LSD) also referred to as log-spectral distortion, is a distance measure (expressed in dB) between two power spectrums [15]. The log spectral distance between spectra P(ω) and $\hat{P}$ (ω) is defined as

$$D_{LS} = \sqrt{\frac{1}{2\pi}\sum_{-\pi}^{\pi}\left[10\log_{10}\frac{p(n)}{\hat{p}(n)}\right]^2 d\omega} \qquad (9)$$

where,

P(n) – Power of Clean Speech Signal,

$\hat{P}$ (n) – Power of Enhanced Speech signal

Experimental Results

Fifty Pure Tamil speeches are taken from the database developed by R. Thangarajan et-al. Three type of noises at different dB's are used such as white noise (0 dB, 5 dB, 10 dB, 15 dB), Volvo noise (0 dB, 5 dB, 10 dB, 15 dB) and pink noise (0 dB, 5 dB, 10 dB, 15 dB). Volvo noise is a real time noise; the other two noises (White, Pink) are the default noises present in the software. Performance of the system is compared for both the default noises and real time noise, so as to justify that the system works well for both the types. The speech signal is sampled at the rate of 8 kHz. The speech enhancement output for one sample speech combined with all the three noises at different dB's (0 dB, 5 dB, 10 dB, 15 dB) is shown in the paper.

Input Speech Signal with noise

Figure 2 shows the input speech signal and the speech signal combined with noise.

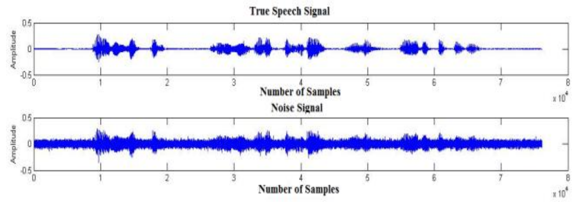

Fig 2: Input Speech Signal and noisy speech

Enhanced Output (5dB)

Figure 3 shows the enhanced speech obtained from white noise, pink noise and Volvo noise respectively by using Non-linear Spectral Subtraction technique.

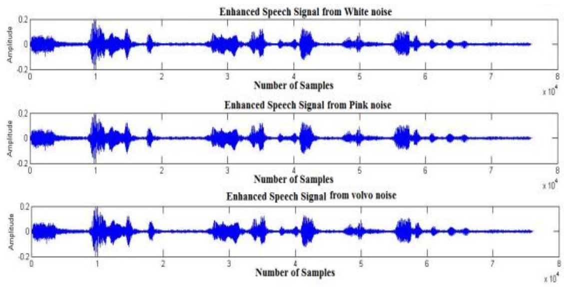

Fig 3: Enhanced output for Non-linear spectral subtraction technique

Figure 4 shows the enhanced speech obtained from white noise, pink noise and Volvo noise respectively by using Multiband Spectral Subtraction technique.

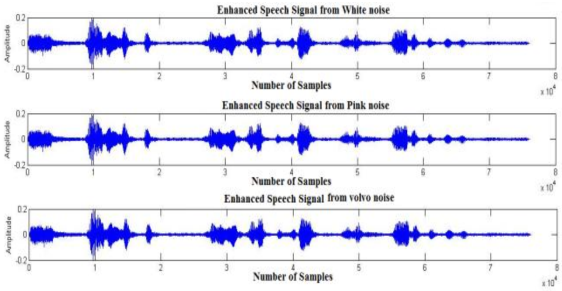

Fig 4: Enhanced output for Multiband spectral subtraction technique

Figure 5 shows the enhanced speech obtained from white noise, pink noise and Volvo noise respectively by using MMSE Spectral Subtraction technique.

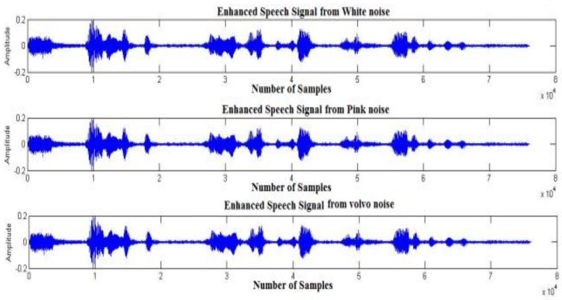

Fig 5: Enhanced output for MMSE spectral subtraction technique.

Comparison Based on SNR

Table 1. Comparison of SNR values

Input over all SNR (dB)		Nonlinear SS	Multiband SS	MMSE SS
White	0	3.207	2.863	3.126
	5	5.074	3.768	4.382
	10	6.266	5.516	5.293
	15	7.033	7.697	6.065
Pink	0	3.176	2.814	3.127
	5	5.111	4.473	4.377
	10	6.450	6.052	5.564
	15	7.307	8.286	6.472
Volvo	0	2.277	6.000	3.272
	5	4.834	7.477	5.369
	10	7.473	8.377	7.513
	15	8.415	8.850	8.496

Table 1 shows the comparison of SNR values for Nonlinear Spectral Subtraction, Multiband Spectral Subtraction and MMSE spectral subtraction techniques.

While comparing the SNR values it is evident that Nonlinear Spectral Subtraction technique offers better output for white and pink noise whereas Multiband Spectral Subtraction technique serves better for Volvo noise.

Comparison Based on LSD

Table 2: Comparison of LSD values

Input over all SNR (dB)		Clean speech	Non Linear SS	Multi band SS	MMSE SS
White	0	3.363	3.363	2.978	2.111
	5	3.063	3.063	2.485	2.057
	10	2.674	2.674	2.054	1.608
	15	2.262	2.262	1.653	1.276
Pink	0	2.843	2.843	2.385	2.050
	5	2.544	2.544	1.988	1.574
	10	2.160	2.160	1.570	1.242
	15	1.766	1.766	1.246	0.994
Volvo	0	0.9236	0.9236	0.868	0.868
	5	0.7414	0.7414	0.754	0.830
	10	1.1186	1.1186	0.903	0.915
	15	0.4185	0.4185	0.648	0.788

Table 2 shows the comparison of LSD values for Nonlinear Spectral Subtraction, Multiband Spectral Subtraction and MMSE spectral subtraction techniques. While comparing the LSD values it is evident that MMSE Spectral Subtraction technique offers better output for white and pink noise whereas Multiband Spectral Subtraction technique serves better for Volvo noise.

4.ADAPTIVE ALGORITHMS

Adaptive filtering is an important subfield of digital signal processing having been actively researched for more than five decades and having important applications such as noise cancellation, system identification, etc., [16] LMS and RLS algorithms are the most frequently and widely applied adaptive algorithms for noise cancellation.

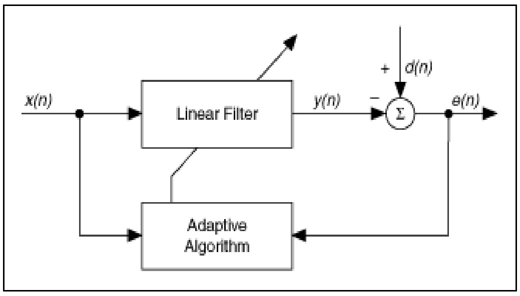

Fig 6: Block diagram for Adaptive Algorithm

Figure 6 shows the diagram of general adaptive filtering technique [17]. Digital filter carries out the filtering of the input signal X(n) and producing output signal Y(n). Adaptive algorithm adjusts the filter coefficient included in the vector W(n), in order to let the error signal e(n) to be the smallest. Error signal is the difference of useful signal d(n) and the filter output y(n). Therefore, adaptive filter automatically carries on a design based on the characteristics of the input signal x(n) and the useful signal d(n). Using this method, adaptive filter can be adapted to the environment set by these signals. When the environment changes, the adaptive filter through a new set of factors, adjusts for new features [18]. The most important properties of adaptive filter is that it can work effective in unknown environment, and can track the input signal with time varying characteristics.

Least Mean Square Algorithm

Widrow and Hoff were developed the Least Mean Square (LMS) algorithm which is the first and most used adaptive algorithm. The LMS, itself established as the workhorse of adaptive signal processing for two primary reasons: Easy to implement and computational efficiency. i.e., linear in the number of adjustable parameters robust performance.

LMS is a gradient descent algorithm and it modifies adaptive filter taps by an amount proportional to the instantaneous estimate of the gradient of the error surface. The following operations are performed in the standard LMS algorithm to update the filter coefficients [19]. Computes the output signal y(n) from the adaptive filter. Computes the error signal e(n) by the equation

$$e(n) = d(n) - y(n) \qquad (10)$$

Modifies or updates the filter coefficients by the equation

$$\overline{w}(n+1) = \overline{w}(n) + \mu(n)\overline{u}(n) \qquad (11)$$

Where $\overline{w}(n)$ is the filter coefficients vector, μ is the step size of the adaptive filter, and $\overline{u}(n)$ is the filter input vector [20]. To minimize the cost function, LMS algorithm adjusts the filter coefficients. They do not demand any matrix operations and therefore less computational resources and memory are required.

Recursive Least Squares Algorithm

RLS filter is the recursive implementation of wiener filter. It is used to find the difference between the desired and actual signals. At each instant an exact minimization of the sum of the squares of the desired signal estimation errors is performed [21].

The RLS approach offers faster convergence and smaller error, at the expense of requiring more computations when compared to LMS.

The steps involved in RLS algorithm are:

Initialize the algorithm by setting

$$\hat{w}(0) = 0$$
$$P(0) = \delta^{-1} I$$

and

$$\pi(n) = P(n-1)u(n),$$
$$k(n) = \frac{\pi(n)}{\lambda + u^H(n)\pi(n)},$$
$$\xi(n) = d(n) - \hat{w}^H(n-1)u(n),$$
$$\hat{w}(n) = \hat{w}(n-1) + k(n)\xi^*(n),$$
$$P(n) = \lambda^{-1}P(n-1) - \lambda^{-1}k(n)u^H(n)P(n-1)$$

Experimental Results

Three types of noises at different dB's are used such as white noise (0 dB, 5 dB, 10 dB, 15 dB), Volvo noise (0 dB, 5 dB, 10 dB, 15 dB) and pink noise (0 dB, 5 dB, 10 dB, 15 dB). The speech signal is sampled at the rate of 8 kHz.

Enhanced Output (5dB)

Figure7 shows the enhanced speech obtained from white noise, pink noise and Volvo noise respectively by using LMS technique.

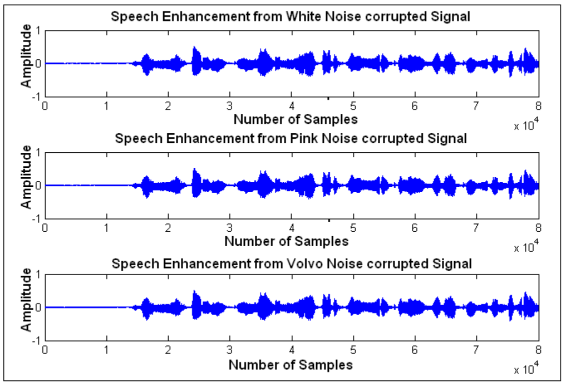

Fig 7: Enhanced output for LMS Algorithm.

Figure 8 shows the enhanced speech obtained from white noise, pink noise and Volvo noise respectively by using RLS technique.

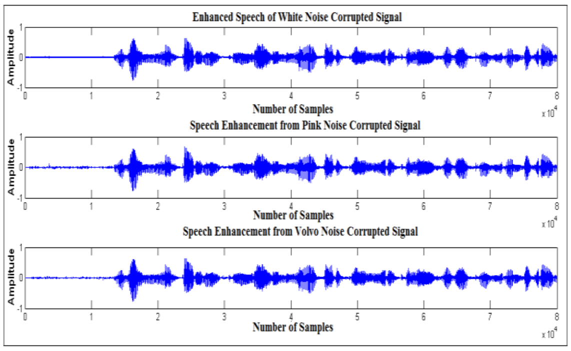

Fig 8: Enhanced output for RLS Algorithm.

Comparison Based on SNR

Table 3: Comparison of SNR values

Input over all SNR (dB)		LMS	RLS
white	0	13.4128	16.5207
	5	13.4365	21.5142
	10	13.4468	26.5180
	15	13.4515	31.5221
pink	0	13.4860	12.7656
	5	13.4892	17.8412
	10	13.4801	22.7905
	15	13.4714	27.8104
volvo	0	11.9737	13.0572
	5	12.0057	18.0533
	10	12.0126	23.0639
	15	12.0130	28.0400

Table 3 shows the comparison of SNR values for LMS and RLS Adaptive algorithms. While comparing the SNR values it is evident that RLS adaptive algorithm offers better output for all the three types of noises chosen.

Comparison Based on LSD

Table 4: Comparison of LSD values

Input over all SNR (dB)		Clean Speech	LMS	RLS
white	0	3.363	0.913	0.850
	5	3.063	0.922	0.814
	10	2.674	0.928	0.784
	15	2.262	0.932	0.704
pink	0	2.843	0.892	0.853
	5	2.544	0.946	0.798
	10	2.160	0.965	0.713
	15	1.766	0.964	0.653
volvo	0	0.9236	1.504	1.220
	5	0.7414	1.572	1.113
	10	1.1186	1.613	1.331
	15	0.4185	1.577	0.981

Table 4 shows the comparison of SNR values for LMS and RLS Adaptive algorithms. While comparing the SNR values it is evident that RLS adaptive algorithm offers better output for all the three types of noises chosen.

5.CONCLUSION

Speech enhancement is mostly done for English and other European languages. Less effort is made in the area of Tamil speech enhancement. Tamil language is the most ancient one compared to all the other languages. So an effort has been made to enhance Tamil speech by using the traditional techniques available for speech enhancement. Enhancement of Tamil speech signal is performed by using different spectral subtraction techniques and adaptive filtering techniques. The performance of these techniques is compared by their Signal to Noise Ratio values and Log Spectral Distance Values. The disadvantage of Spectral subtraction techniques is the addition of musical noise to the original speech signal [22]. When adaptive filtering techniques are used, the results are better RLS algorithm offers better speech enhancement. In future, wavelets may be used in combination with the adaptive filters to obtain enhanced Tamil speech.

6.REFERENCES

[1] M.Selvam, A.M.Natarajan, R.Thangarajan, “Structural parsing of natural language text in Tamil Language using dependency model ”, International Journal of Computer Processing of Languages, Volume 22, Issue 3, Page No:237-256, 2009.

[2] R.Thangarajan, A.M.Natarajan, M.Selvam, “Development of Large Vocabulary Speech Corpus for Tamil Language”, International conference on “Emerging technologies in Intelligent system and Control”, Volume 2,2005.

[3] S. Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction” *IEEE Transactions on Acoustics, Speech,Signal processing*, Volume 27, Issue 2, Page No: 113-120, 1979.

[4] Brady N. M. Laska, MiodragBolic and Rafik A. Goubran , “Particle Filter Enhancement of Speech Spectral Amplitudes”, *IEEE Transactions on audio, speech, and language processing*, Volume 18, Issue 8, Page No:2155-2167, 2010.

[5] M.Berouti, R. Schwartz and J Makhoul,”Enhancement of Speech corrupted by acoustic noise”, in Proceedings of IEEE International Conference on Acoustics, Speech and Signal processing , Volume 4, Page No: 208–211, 1979.

[6] Krishnamoorthy.P and MahadevaPrasanna.S.R ,” Temporal and spectral processing Methods for Processing of Degraded Speech: A Review”, IETE technical review, Vol-26, Issue-2, Page No:137-148, 2009.

[7] NavneetUpadhyay and AbhijitKarmakar, "The spectral subtractive-type algorithms for enhancing speech in noisy environments”, in Proceedings of *International Conferenc on Recent Advances in Information*

Technology, ISM Dhanbad, India, Page No: 841-847,2012.

[8] Paurav Goel1, Anil Garg, "Developments in Spectral Subtraction for Speech Enhancement", International Journal of Engineering Research and Applications (IJERA), Volume 2, Issue 1, Page No: 55 to 63, 2012.

[9] Anuradha R.Fukane, Shashikant L. Sahare "Different approaches of spectral subtraction method for enhancing speech signal in noisy environments", International Journal of Scientific and Engineering Research, Volume 2, Issue 5, 2011.

[10] S. Kamath and P. Loizou, "A multiband spectral subtraction method for enhancing speech corrupted by colored noise", '*IEEE International Conference on Acoustics, Speech and Signal processing*, 2002

[11] Y. Eprahim and D. Malah (2002), "Speech Enhancement using minimum mean- square error short- time spectral amplitude estimator",'*IEEE Transactions on on Acoustics,Speech,and Signal processing,* Volume 32, No:6, Page No:328–337, 1984.

[12] R. Martin "Speech Enhancement using MMSE short Time Spectral Estimation with Gamma Distributed Speech Priors", IEE*E International Conference on Acoustics, Speech,and Signal processing* , Volume 1,Page No: 253–256, 2002.

[13] Gang Wang, Chunguang Li, and Le Dong, "Noise Estimation Using Mean Square Cross Prediction Error for Speech Enhancement", *IEEE transactions on circuits and systems*, Volume 57, Issue 7, Page No: 1489-1499, 2010.

[14] C. Breithaupt and R. Martin,"Analysis of the Decision-Directed SNR Estimator for Speech Enhancement with respect to Low-SNR and Transient Conditions", *IEEE transactions on audio, speech, and language processing*, Volume 19, Issue 2, Page No:277-289, 2011.

[15] Jie Wang, Hao Liu, ChengshiZheng, Xiaodong Li " Spectral subtraction based on two-stage spectral estimation and modified cepstrum thresholding", *Applied Acoustics*, Volume 74, Issue 3, Page No: 450–458, 2013.

[16] John Hakon husoy, "Adaptive filters viewed as iterative linear equation solvers" Third International conference on "Numerical Analysis and Applications", 2004.

[17] B. Widrow, J. R.Glover, J. M.McCool, J.Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler,Eugene Dong, R. C. Goodlin: "Adaptive NoiseCancelling: Principles and Applications", Proceedings of the IEEE, Volume 63, Issue 12, Page No: 1692, 1716, 1975.

[18] Tekale P. B , 2 Kulkarni S. R, " Modified Kalman Based NLMS Algorithm For Noise Cancellation", International journal of Systems and Technologies, Volume 5 , Issue 1,Page No: 47-55,2012.

[19] Harjeet Kaur, Dr. Rahul Malhotra, Anjali Patki,"Performance Analysis of Gradient Adaptive LMS Algorithm", International Journal of Scientific and Research Publications, Volume 2, Issue 1, Page No: 1-4, 2012.

[20] Jasveen Kaur and RanjitKaur, "A Comparison Study of Power Spectrum Densities of Various Adaptive Algorithm Using Adaptive Filter", International Journal of Engineering Research and Applications (IJERA), Volume 2, Issue 3, Page No:2335-2341, 2012.

[21] Vimala C, Radha V, "optimal Adaptive filtering for Tamil Speech Enhancement", International Journal of Computer Applications, Volume 41, Issue 17, 2012.

[22] S. Ogata, T. Shimamura , "Reinforced Spectral Subtraction method to enhance speech signal" , *IEEE International Conference on Electrical and Electronic Technology*, Volume 1, Page No:242-245, 2001.