ALGORITHMS FOR THE ENHANCEMENT OF SPEECH SIGNAL RECORDED IN ADVERSE CONDITIONS

Apoorva A, Ajish K. Abraham

AbstractT- The performance of a system can have negative effect by the influence of ecological noise, speech distortions, noise induced by microphones, limited frequency bandwidth etc. In this paper various noise reduction algorithms are studied and implemented in MATLAB and performance of each algorithm is evaluated on the basis of SNR.

Index Terms- Adaptive Spectral subtraction, Kalman , Noise reduction, NLMS , RLS, SNR, Spectral Subtraction.

————————————   ◆   ————————

1 INTRODUCTION

The speech recorded through microphone is generally contaminated by noise. Noise is present in all environments. Noise is nothing but contamination. Such contamination changes the quality of the signal recorded and inhibit the speech quality and intelligibility, hence incurring harm to human to machine interface. Noise detection and reduction being a digital filtering problem, clean speech is deduced by passing noisy speech through linear filter. Thus designing an optimal filter that can significantly suppress contamination without affecting speech itself becomes major issue.

The various types of noise reduction algorithms in literature are classified in main three categories, namely; filtering/estimation based noise reduction, beam forming and active noise cancellation (ANC) techniques. Vineela Murikipudi et al proposed design and analysis of speech processing using Kalman filtering. The filtering ability is tested by reconstructing skills of the algorithm based on accurate estimation[1]. The spectral subtraction technique put forth by Kotta Manohar et al, a pre-processor noise reduction technique for digital voice processors, the performance of both basic method and modified method is evaluated and the criteria used is described and concluded.[2].

————————————————

- *Apoorva A, M.Tech BSP&I, , Department of Instrumentation Technology Sri Jayachamarajendra College of Engineering, Manasagangothri, Mysuru, Karnataka, India.*
- *Ajish K. Abraham, Head, Department of Electronics, All India Institute of Speech and Hearing, Manasagangothri, Mysuru, Karnataka, India.*

A review on Noise Cancellation by Lakshmikanth.S et al gives information about need of noise cancellation and speech enhancement. The study briefly explains the nature of speech and the noise and the type of filters used in time doiman, spatial domain and frequency domain. The paper introduced to the techniques of filtering and portrays the present technique and future scope of the technique used for filtering [3].

New technique for SNR estimation proposed by Jurgen Tchorz et al has single microphone noise supression algorithm to estimate signal to noise ratio. The study deals with explanations of SNR estimation and feature extraction, neural network classification and noise supression. The performance depends on type of noise. The algorithm was proved to be effective in case of gaussian noise [4].

A technique to filter the time sequences of spectral parameters proposed by Climent Nadeu et al comprises of details regarding steps involved in filtering process, introduction to TSSP detailing its spectrum and its design. The experiment is done on TSSP CDHMM digit recognition and cepstral parameters and results are achieved by removing dc component and adding differential parameters [5]. Noise reduction algorithm based on LMS adaptive filter of audio filters was proposed by Yang Liu et al effectively solves the filtering problem by NLMS. Improved performance and quality is seen. The technique is found effective in case of white Gaussian noise [6]. Allam Mousa, Marwa Qados, Sherin Bader [10] proposed Speech Signal Enhancement Using Adaptive Noise Cancellation Techniques Based on the simulation results, it was deduced that the following order of performance for the 4 algorithms,

RLS> NLMS > LMS> LPC

There is a slight deterioration in performance of LMS for near 0 SNR because of the estimation errors introduced by the noise PSD estimator.

2 NOISE CANCELLATION METHODS

2.1 Normalized LMS Algorithm

The normalized LMS is based on the principle of minimal disturbance which states the change in weight vector of an adaptive filter for each iteration should be minimum, with constrained filter output. NLMS has adaptive constant with no dimension. For both correlated and uncorrelated data NLMS converges faster than LMS[7]. The learning rate μ which decides the stability of the system is hard to choose in case of LMS since the working of the algorithm depends on scaling of the input x(n). This problem is solved by normalizing with the power of the input in case of NLMS.

Gradient noise amplification problem is introduced by choosing large convergence factor μ in case of LMS . NLMS eases out this difficulty. At n+1th iteration the change in w(n) is normalized with respect to squared Euclidian norm of x(n). The NLMS algorithm is a time-varying step-size algorithm, calculating the convergence factor μ as in (6).

$$g(n)=x(n)+v(n) \tag{1}$$
$$w(n)=[w_1(n)w2(n).......w3(n)]T \tag{2}$$
$$x(n)=[x(n)x(n-1).....x(n-N+1)]T \tag{3}$$
$$g(n)=wT(n)x(n) \tag{4}$$
$$e(n)=d(n)-g(n) \tag{5}$$
$$\mu=$$
$$w(n+1)=w(n)+\mu e(n)x(n) \tag{7}$$
$$w(n+1)=w(n)+(\mu\hat{}/xT(n)x(n))e(n)x(n) \tag{8}$$

2.2 RLS Algorithm

Recursive Least Squares (RLS) algorithm converges much faster than the LMS algorithm, as it considers all inputs from start of adaptation to present value of input[9]. The design and flow of algorithm is done on the basis of following equations refer (9) to (12). The mathematical representation in the equations given below are same as described in NLMS section Where g(n) is recorded speech, d(n) uncorrelated additive noise.

$$e(n) = g(n) - d(n) \tag{9}$$
$$w'(n+1) = w'(n) + e(n).k'(n) \tag{10}$$
$$k'(n) = P(n)u'(n)/(\lambda + u^T(n)P(n)u'(n)) \tag{11}$$
$$P(n + 1) = \lambda^{-1}P(n) - \lambda^{-1}k'(n)u^T(n)P(n) \tag{12}$$

2.3 Spectral Subtraction

Spectral Subtraction method works best to eliminate the background stationary noise. The major advantages of

algorithm are its simplicity, an estimate of the mean noise power which is whole and sole requirement and unlike other algorithms signal assumption is not at all required.

The greatest drawback being addition of negative estimates of the power spectrum occasionally within the framework. Artificial flooring enables consistent estimation and also other information bearing part of recorded speech yields characteristic musical noise in time frequency representation. But that can be minimized [8]. Digital voice processors use these as preprocessor for speech compression, recognition and validation. The noise here is uncorrelated additive noise. Hence the power spectrum of the contaminated speech is sum of the signal power spectrum and noise power spectrum.

$$g(n)=x(n)+d(n) \tag{12}$$

Where x(n) is the actual speech signal, d(n) is the uncorrelated additive noise and g(n) represents the degraded speech signal. The Fourier transform is given by,

$$G(\omega)=X(\omega)+D(\omega) \tag{13}$$

The processing is done on short time basis hence, $G_w(n)$, $S_w(n)$, $D_w(n)$ represents the windowed signal and its short-time Fourier transform is given by (15)

$$Gw(n)=Xw(n)+Dw(n) \tag{14}$$
$$|Gw(\omega)|2=|Xw(\omega)|2+|Dw(\omega)|2+Xw(\omega).Dw^*(\omega)+$$
$$Xw^*(\omega)Dw(\omega) \tag{15}$$

where $D_w^*(\omega)$, $X_w^*(\omega)$ represent complex conjugates of $D_w(\omega)$, $X_w(\omega)$. $|X_w(\omega)|2$ gives the short-time energy spectrum of the speech. In this method, the terms $|D_w(\omega)|2$, $X_w(\omega).D_w^*(\omega)$ and $X_w^*(\omega)D_w(\omega)$ cannot be computed thus they are assumed to be equivalent to $E[|D_w(\omega)|2]$, $E[X_w(\omega).D_w^*(\omega)]$, $E[X_w^*(\omega)D_w(\omega)]$

Since noise and signal are uncorrelated $E[X_w(\omega).D_w^*(\omega)]$, $E[X_w^*(\omega)D_w(\omega)]$ are zero.

$$|X_w(\omega)|2= |Gw(\omega)|2 - E[|Dw(\omega)|2] \tag{16}$$

Hence mathematical representation of power spectrum subtraction method given by (17)

$$|X_w(\omega)|2=Y_w(\omega)|2-|D_w(\omega)|2 \text{ if } |X_w(\omega)|2>0$$
$$=0 \quad \text{otherwise} \tag{17}$$

Assuming short time phase to be relatively unimportant, R $Xw(\omega)$, is approximated as the phase of $Xw(\omega)$, given by $RGw(\omega)$ so that,

$$|X_w(\omega)|2=|X_w(\omega)|.exp[jRG_w(\omega)] \tag{18}$$

$$X_w(n)=F^{-1}\{X_w(\omega)\} \tag{19}$$

$$K_k = P_{k-1} \ H_{k-1} \ [H_{k-1} \ P_{k-1} \ H_{k-1} + R]^{-1} \qquad (27)$$

Thus the enhanced signal is reconstructed as per (20).

2.4 Adaptive Spectral Subtraction

This is a new technique which combines both spectral subtraction and adaptive method of noise reduction. The adaptively noise reduced signal is given as input to perform spectral subtraction. The Fig.1.1 shows block diagram of proposed method. Even though the algorithm has high computational complexity, it can reduce the background noise effectively. However, experimental results show that there is some residual noise in the processed signal, which affects the hearing effect. To reduce the influence of the background noise and increase the definition of the speech and also to resolve the problem of slow learning rate of spectral subtraction method, the algorithm based on the adaptive spectral subtraction is introduced in this paper.

2.4 Kalman Filtering

The Kalman Filter formulates the present "state" of a linear dynamic system contaminated by white noise. This system has optimal characteristics even in the case of disturbance caused by estimation errors and their quadratic functions. It is a Recursive Data Processing Algorithm. Speech is a largely correlated signal. This correlation is useful in case of coding or filtering. The all pole, or autoregressive (AR), signal model is often used for speech. The AR signal model is introduced as y_k, refer (20).

$$y_k = b_1 y_{k-1} + b_2 y_{k-2} + -- + b_N y_{k-N} + w_k \qquad (20)$$

Where,
k = Number of iterations
y_k = current input speech signal sample;
y_{k-N} = (N-1)th sample of speech signal;
b_N = Nth Kalman filter coefficient and
w_k = excitation sequence

The state space form of speech is given according to (21) and (22).

$$H_k = XH_{k-1} + Wz \qquad (21)$$

$$Y_k = gH_k \qquad (22)$$

Where X is the system matrix, H_k has speech samples; wk is the excitation vector and g, the output vector. (k-N+1)th iteration is due to the state vector.
Following steps denote the implementation of Kalman filter within every single loop of the filter. The row vector is defined as $H_{k-1}{}^T$.

$$H_{k-1}{}^T = [y(k-1) \ y(k-2).......y(k-N)] \qquad (23)$$

$$Z_k = X_k H_{k-1}{}^T + W_k \qquad (24)$$

Each iteration updates X_k. At k=0 H_k is undetermined but when the time Z_k is encountered, the value in matrix H_{k-1} is known. Thus the above intention is adequate to define the filter , which consists of X_k

$$X_k = [I - K_k H_{k-1}{}^T] X_{k-1} + K_k Z_k \qquad (25)$$

Where K_k is the Kalman gain matrix, P_{k-1} is the a priori error covariance matrix, R is measurement noise covariance, and P_k is the a posteriori error covariance matrix.

$$P_k = P_{k-1} - P_{k-1} \ H_{k-1} [H_{k-1}T \ P_{k-1} \ H_{k-1} + R] - 1 \ H_{k-1}T \ P_{k-1} + Q \quad (26)$$
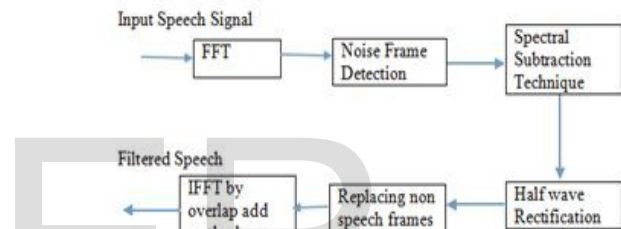

Fig. 1: Block diagram of Adaptive spectral subtraction

RLS method of filtering, refer (1) to (8) is followed by Spectral Subtraction method, refer (13) to (19). The non speech frames are detected using magnitude averaging and replaced. The output is reconstructed using overlap add method.

3 EXPERIMENTAL RESULTS

The performance of KALMAN filter, RLS filter, NLMS filter and Spectral Subtraction method of filtering is studied with the help of MATLAB simulations. Utterances were pronounced by male speaker. The speech data sampled at 8khz is used and the recorded time is 0.01s. The original clean speech is corrupted by random noise which is a type of noise generated by MATLAB using randn command . The proposed approaches improve noise audio signal quality and the performance of each algorithm is evaluated on the basis of SNR.

3.1 Kalman Filtering

The following results for Kalman filttering is obtained by choosing Kalman coefficients of a 5th order Kalman filter. -

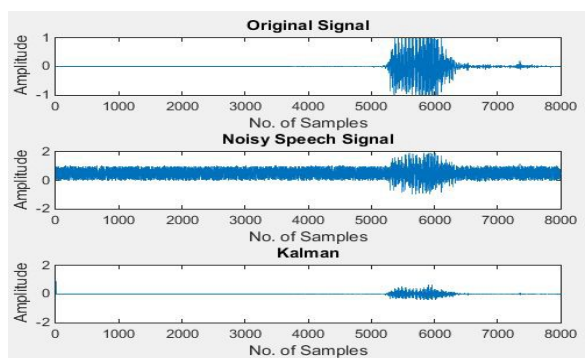0.8,0.2, -0.6, 0.7 and -0.4. Fig. 2 shows the enhanced speech output for Kalman algorithm.

Fig.2. Kalman Filtering

3.2 RLS Filtering

For RLS filtering the speech signal is segmented every 120ms and a frame includes 100 sampled data. The 256 point FFT is performed with the value of forgetting factor being 1 and length being 50. Fig. 3 shows combined plot of adaptive and required filter response of designed RLS filter. Fig. 4 shows the output of the designed RLS filter and the noise picked up by secondary microphone. Fig. 5 shows the time scope plot which actual shows the region of original speech data in filtered speech sample.

required filter response
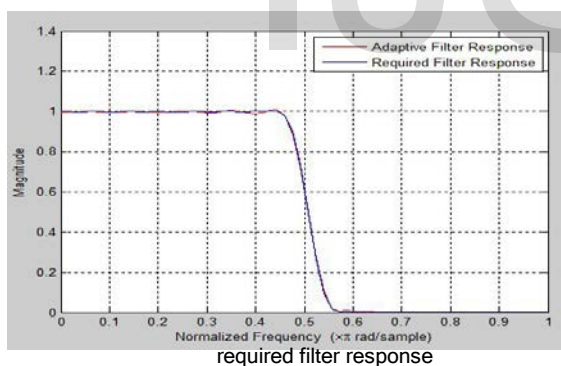
3.3 Normalized LMS Filtering

The audio signal noise reduction by NLMS technique is as shown in Fig. 6. The figure actually shows the true and estimated values of filtered speech. The time domain representation of original and filtered signal is almost similar. It is evident   from Fig. 6 that the noise level in the noise canceled signal is reduced clearly.

        It also indicates that the adaptive noise reduction method result in a noise variance that is more nearly equation to zero. The value of alpha being 0.95. By contrast, the variance of the noise from the original and filtered signal nearly equals to 9.3132e-11. The step size is 380. The spectrogram for noise eduction algorithms for the same randn noise is also shown in Fig. 7.

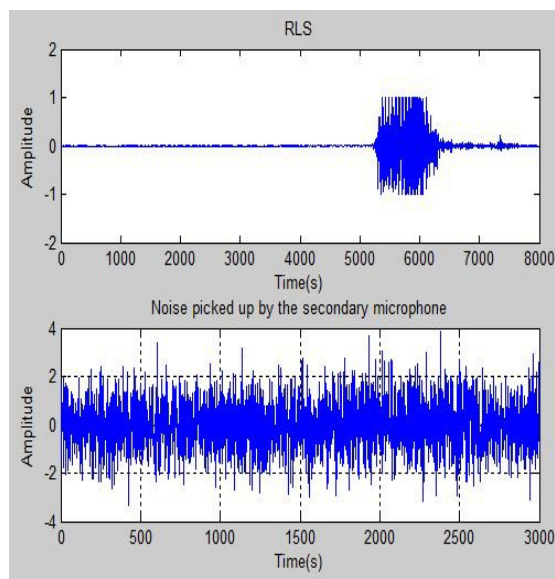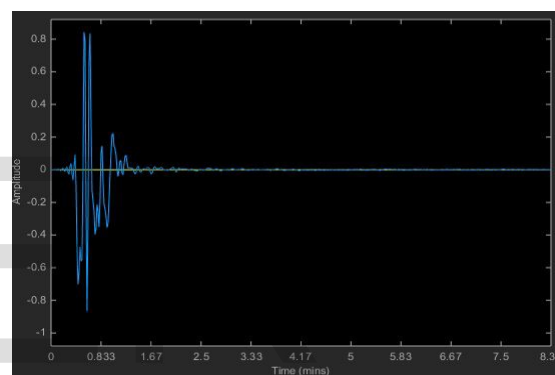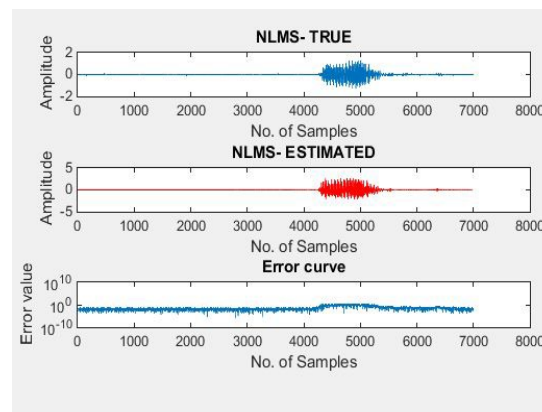Fig. 4. The RLS filter output and noise

Fig. 5. Time Scope Response

Fig. 6. The NLMS Filter output

3.4 Spectral   Subtraction   Technique

Specification of parameters chosen to evaluate the filter are as follows. The Frames are 50% overlapped and are of 60ms length. 4096 point FFT and Hamming window windows the frames .

The difference in performance of spectral subtraction technique can be seen on spectrogram of speech samples enhanced by the algorithm. Fig. 8 shows the time domain representation of the recorded and filtered speech along with their spectrographic representations.
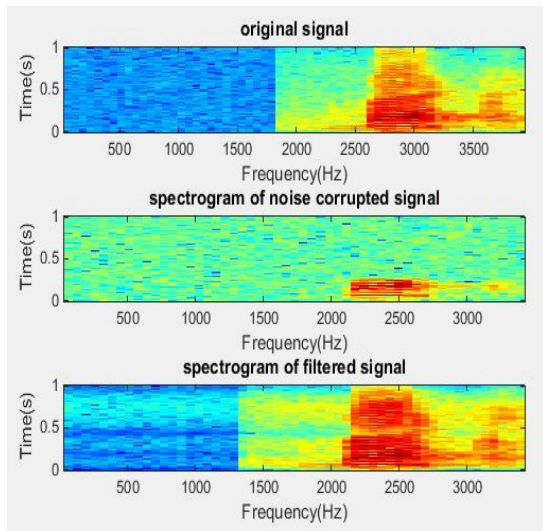

Fig. 7. Spectrogram of input speech,
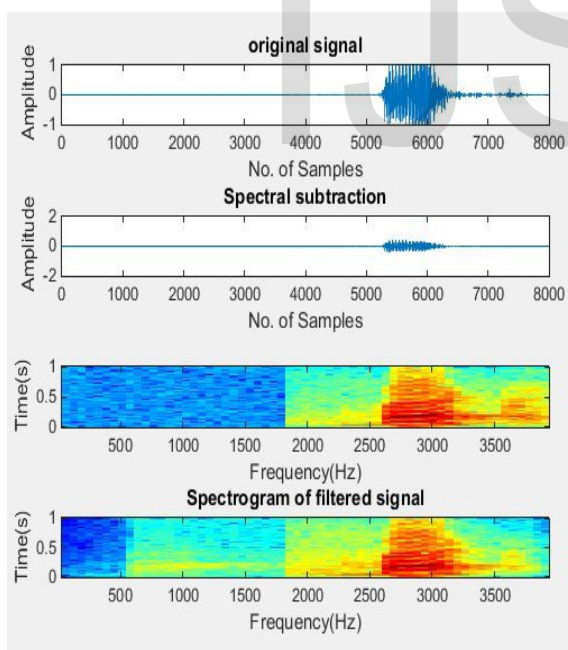noise and filtered speech

Fig. 8. Time domain and spectrogram representation

3.5   Adaptive Spectral Subtraction

Similar to spectral subtraction algorithm of the proposed algorithms implemented in MATLAB with parametrs chosn as follows. The frames here are also 50% overlapped and are of 60ms length. 4096 point FFT and Hamming window windows the frames of spectral subtraction part of the

programme. The Adaptive part is performed by 50 point FIR filter with a forgetting factor of value 0.001.

The performance of proposed alogithm is evaluated using spectrogram response and the SNR. Looking at the comparision it can be deduced that the proposed algorithm has higher learning rate than that of spectral subtraction. Fig. 1.9 shows the time domain representation of the recorded and filtered speech along with their spectrographic representations.

4   COMPARISION

Spectral Subtraction technique, NLMS ,RLS, Kalman filtering and proposed algorithms are compared on the basis of SNR, percentage noise reduction, complexity and stability. In the simulation for RLS and NLMS the reference input signal x(n) is a random noise of SNR -20dB generated using randn function in MATLAB, the required signal d(n) ,sum of a delayed version of x(n) into optimal signal x(n), d(n) = x(n) + g(n).
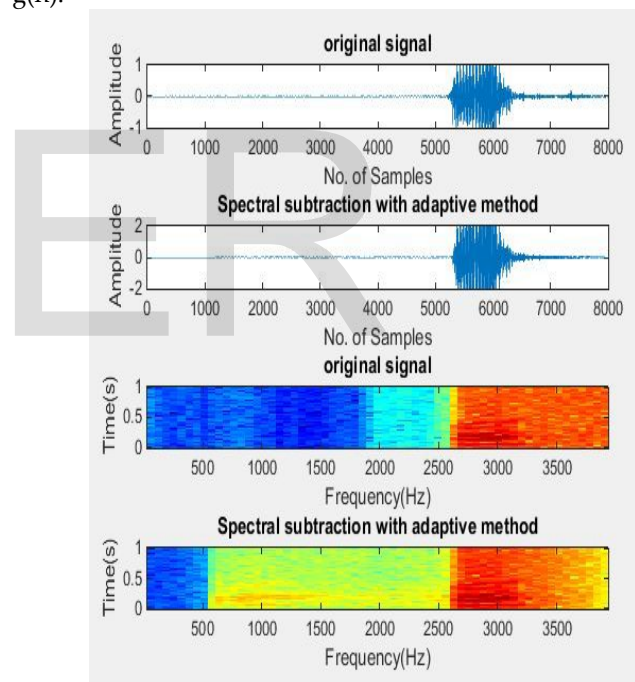

Fig. 9. The time domain representation and spectrogram of spectral
subtraction with adaptive method

Effectiveness of each algorithm is indicated in the spectrogram plot of each filter as shown in Fig. 10. On investigation it has been found that the filtered output of all algorithms, the adaptation of spectral subtraction, RLS > Adaptive Spectral Subtraction > Spectral Subtraction > Kalman > NLMS[11].

The table 1 shows the SNR obtained after implementing each type of filter in MATLAB Simulink. Thus by the look at the table 1 it is clear that the performance of algorithms goes in the order of RLS>Proposed algorithm>Spectral Subtraction>Kalman filtering>NLMS

Table 1: Table of SNR comparison

Sl. No	Type of Algorithm	Input speech (dB)	Noise (dB)	Filtered Speech (dB)
1	Kalman	-8.78	-20	1.0453
2	Spectral Subtraction	-8.78	-20	3.27
3	NLMS	-8.78	-20	0.0125
4	RLS	-8.78	-20	4.4175
5	Adaptive Spectral Subtraction	-8.78	-20	4.3440

Fig. 1 ... RLS ,

[1]   i   l   i   i   i   i   i   i   i   i   . abibulla Khan, G.Mrudula, K.Sudhakar Babu, T.Raghavendra Vishnu, "Design And Analysis of Speech Processing Using Kalman Filtering", K L University, Guntur DT, AP, India Head of the department, K L University, Guntur DT, AP, India . Journal of Theoretical and Applied Information Technology JATIT & LLS © 2005 - 2011.

[2]   Kotta Manohar (02307428) Supervisor: Prof. Preeti Rao ,"Single Channel Credit Seminar Report Electronic Systems Group EE Dept, IIT Bombay Submitted November, 2002.

[3]   ] Lakshmikanth.S, Natraj. K.R, Rekha.K.R Research scholar, EEE, Jain University, Bangalore, Karnataka, "Noise Cancellation in Speech Signal Processing-A Review ", International Journal of Advanced Research in Computer and Communication Engineering   Vol. 3, Issue 1, January 2014 .

[4]   Jürgen Tchorz and Birger Kollmeier, "SNR Estimation Based on Amplitude Modulation Analysis With Applications to Noise Suppression" IEEE transaction on speech and audio processing, VOL. 11, NO. 3, MAY 2003.

[5]   Climent Nadeu , Pau Paches-Leal , Biing-Hwang Juang   Dept. Teoria de Senyal i Comunicacions, UniÕersitat Politecnica de Catalunya, J. Girona,"Filtering the Time Sequences of Spectral Parameters for Speech Recognition" , NJ 07974, USA. Received 8 November 1996; revised 20 June 1997.

[6]   Yang Liu, Mingli Xiao and Yong Tie,"A Noise Reduction Method Based on LMS Adaptive Filter of Audio Signals", 3rd International Conference on Multimedia Technology ICMT, 2013

[7]   M.Kalamani,S.Valarmurthy,M.Krishnamoorthi , "Modified Noise Reduction Algorithm for Speech Enhancement", Applied Mathematical Sciences,HIKARI Ltd,vol.8,no.89,4447-4452,   2014.

[8]   Ekaterina Verteletskaya, Boris Simak ," Noise Reduction Based on Modified Spectral Subtraction Method ",IAENG International Journal of Computer Science, IJCS_38_1_10, 38:1, submitted February, 2011

[9]   Pogula Rakesh, T. Kishore Kumar," A Novel RLS Based Adaptive Filtering Method for Speech Enhancement", World Academy of Science, Engineering and Technology International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering Vol:9, No:2, 2015.

[10]   Allam Mousa, Marwa Qados, Sherin Bader ,"Speech Signal Enhancement Using Adaptive Noise Cancellation Techniques ", Canadian Journal on Electrical and Electro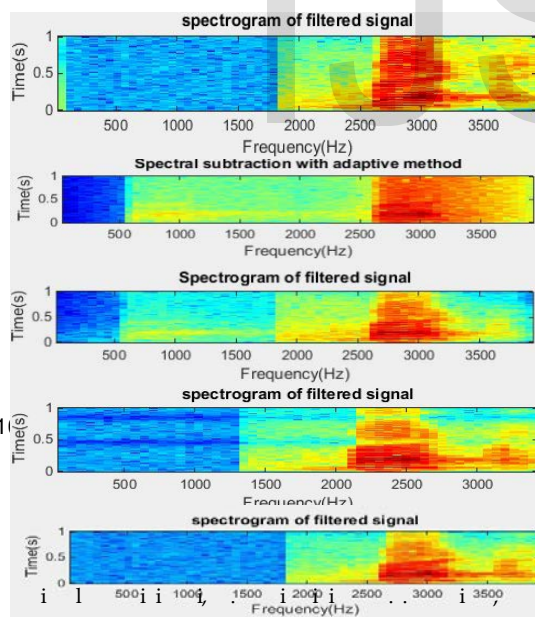nics Engineering Vol. 3, No. 7, September 2012