

Performance Analysis of Adaptive Filtering Algorithms for Acoustic Echo Cancellation in Teleconference System

Udofia, K.M., Ibanga, E. & ¹Ezenkwu, C.P.
 Department of Electrical Electronics and Computer Engineering
 University of Uyo
 Uyo, Akwa Ibom State, Nigeria
¹E-mail: chineduezenkwu@uniuyo.edu.ng
¹Phone: +234(0)7034471962

¹Corresponding Author

ABSTRACT

This paper presents the performance evaluations of different adaptive algorithms usages in acoustic echo cancellation. The acoustic echo canceller is built using Least Mean Square (LMS) and Normalized Least Mean Square (NLMS) with normalized cross correlation (NCC) algorithm double talk detector. The mirror-source model technique is adopted in the modelling of the Room Impulse Response (RIR) of the acoustic echo environment for this work. The adaptive algorithms LMS and NLMS are implemented using MATLAB. These algorithms are tested with the simulation of echo occurring environment. The performance of the LMS and NLMS are evaluated in terms of the echo return loss enhancement (ERLE) and mean squared error (MSE). The results show that NLMS algorithm achieves a better performance compared to the LMS algorithm.

Keywords: Performance Analysis, Adaptive Filtering Algorithms, Acoustics, Echo, Cancellation & Teleconference System.

CISDI Journal Reference Format

Udofia, K.M., Ibanga, E. & Ezenkwu, C.P. (2016): Performance Analysis of Adaptive Filtering Algorithms for Acoustic Echo Cancellation in Teleconference System. Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 7 No 4. Pp 217-228 Available online at www.cisdijournal.net

1. INTRODUCTION

In teleconference systems, providing a good free voice quality communication between two or more people from different locations is very paramount. This requires the use of hands-free communication technique which comprises a loudspeaker and a high-gain microphone, in place of a telephone receiver. As shown in figure 1, the audio signal from the far-end caller is broadcast by the loud speaker of the near-end. The waves from the audio signal are fed back to the remote user (far-end room) through the near-end microphone after reflection from the wall, floor and other objects inside the near-end room. The signal at the microphone consists of the original intended signal, time-delayed and distorted version of an original signal. The reverberation of the audio signal is called an echo [1-3]. The echo then creates a feedback loop and the far-end caller hears an echo of his or her own voice. The acoustic echo disturbs the conversation and reduces the quality of the system [4]. Furthermore, the acoustic system could become instable, which would produce a loud howling noise to occur.

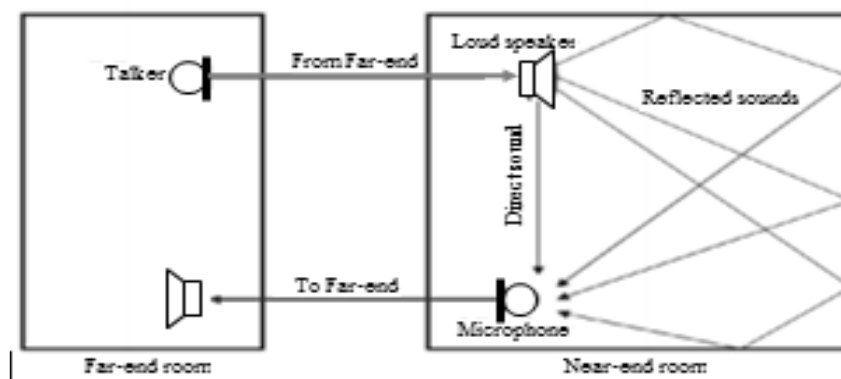


Figure 1: A teleconference system with echo paths of room

The acoustic echo generated in teleconference systems could be eliminated with an echo suppressor or canceller [5]. The echo suppressor offers a simple but effective method to counter the echo problem for only half-duplex communication. Half-duplex communication permits only one speaker to talk at a time. This drawback led to the invention of acoustic echo canceller (AEC) which allows full-duplex communication allowing both speakers to talk at the same time [5].

2. ADAPTIVE ECHO CANCELLATION

Adaptive echo cancellation (AEC) uses adaptive filters at both ends to estimate the transfer function, $h(n)$, between each loud speaker and its corresponding microphone. As shown in figure 2, an echo situation and an echo cancellation system using adaptive filters is illustrated. The far-end and near-end denote the transmitting and receiving ends over a communication channel, such as a teleconferencing system, where two users, one at the far-end and the other at the near-end side, are communicating.

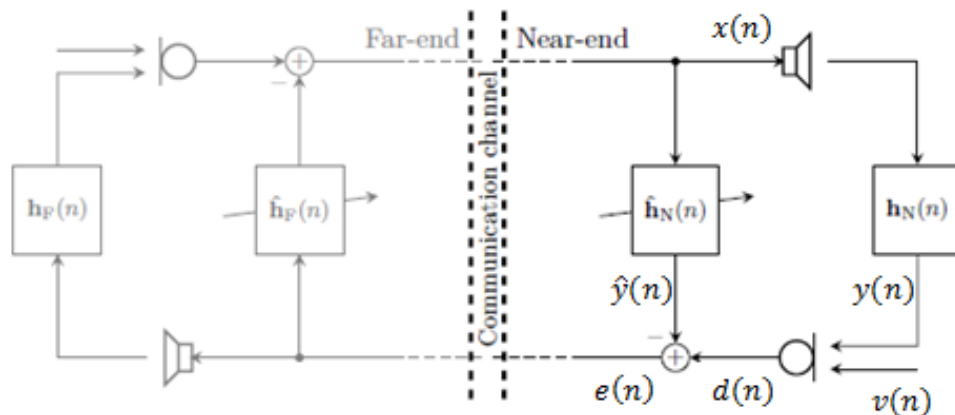


Figure 2: An acoustic echo cancellation system

Ideally, only the near-end speech signal $v(n)$ should be transmitted to the far-end, but in practice, the signal $v(n)$ is also transmitted to the far-end where it is perceived as an echo. Considering the near-end loudspeaker signal $x(n)$ as an unprocessed speech signal spoken by the far-end user. $x(n)$ is modified by the near-end echo path $h_N(n)$ to produce an echo signal $y(n)$ which is received by the microphone in addition to the near-end signal $v(n)$. The microphone signal $d(n)$ is transmitted back to the far-end. Hence, the far-end user would hear a delayed and distorted version of his/her own voice as an echo. This is the basic acoustic echo problem. The echo path $h(n)$ depends on the acoustic properties of the room such as reflective surfaces and movements of the user [6].

Echo cancellation using adaptive filters is an effective method to manage acoustic echoes [7]. In figure 2, an adaptive filter $\hat{h}_N(n)$ is used to model the echo path $h_N(n)$ and create the cancellation signal $\hat{y}(n)$. The adaptive filter is made up of an echo estimator and a subtractor. The echo estimator monitors the received path and dynamically builds a mathematical model $\hat{h}_N(n)$ of the line that creates the returning echo. The model of the line $\hat{h}_N(n)$ is convolved with the received signal $x(n)$. This yields an estimate of the echo signal $\hat{y}(n)$, which is applied to the subtractor. The subtractor eliminates the estimated echo signal from the line in the send path. The acoustic echo canceller is said to converge on the echo signal as an estimate of the line is built through the adaptive filter.

2.1 Double-Talk Problem In Echo Cancellation

A serious challenge associated with echo cancellation is the double-talk situation [8]. It occurs when both the far-end and the near-end users speak simultaneously, such that the near-end loudspeaker signal $x(n)$ and the near-end speech signal $v(n)$ are active at the same time. Unfortunately, adaptive algorithms adjusted to a high convergence rate usually diverge quickly in this situation [9]. A renowned technique to limit the divergence of adaptive filters is based on a double-talk detector, which controls the adaptive algorithm by freezing or slowing down the adaptation when a double-talk situation is detected [10].

3. METHODOLOGY

The diagram of the acoustic echo cancellation setup for teleconferencing is as shown in figure 3. The main components are the adaptive filters, double-talk detector, echo path, non-linear processor.

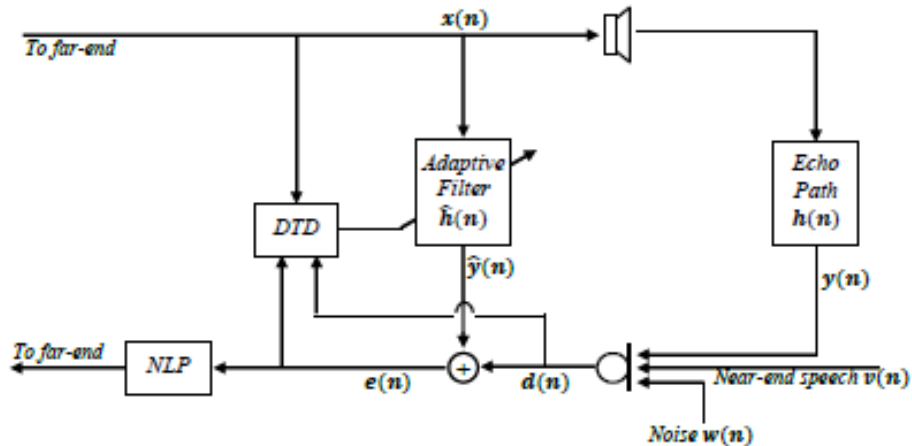


Figure 3: Diagram of the acoustic echo cancellation setup for teleconference system

3.1 Adaptive Filter Algorithms

Various adaptive filter algorithms have been developed for acoustic echo cancellation. The aim of an adaptive filter is to compute the difference between the desired signal $d(n)$ and the adaptive filter output, $y(n)$ to obtain $e(n)$. This error signal is fed back into the adaptive filter and its coefficients are iteratively changed in order to minimize a function of this difference. In this work, three of these algorithms are considered, namely Least Mean Square (LMS) and Normalized Least Mean Square (NLMS).

3.2 LMS Algorithm

The LMS algorithm is a stochastic gradient-based algorithm that utilizes the gradient vector of the filter tap weights to converge on the optimal wiener solution. Using equations (1) – (3), the filter tap weights of the adaptive filter are updated with each iteration of the LMS algorithm.

$$\hat{y}(n) = \hat{k}^T(n)x(n) \tag{1}$$

$$e(n) = d(n) - \hat{y}(n) \tag{2}$$

$$\hat{k}(n + 1) = \hat{k}(n) + 2\mu x(n)e(n) \tag{3}$$

where

$\hat{y}(n)$ is the adaptive filter output, $\hat{k}(n)$ represents the adaptive filter weight vector at time n , $x(n)$ represents time delayed input signal samples, $d(n)$ is the microphone output, $e(n)$ represents error signal to be minimized and μ represents step size or convergence factor between $0 < \mu < 2/\lambda_{max}$ where λ_{max} maximum eigenvalue of autocorrelation matrix.

3.3 NLMS Algorithm

In LMS algorithm, the adjustment applied to $\hat{k}(n)$ varies directly with the input signal $x(n)$. Hence, when $x(n)$ is large, the LMS algorithm experiences the problem of gradient noise amplification. To take care of this problem, the correction applied to the tap weight vector $\hat{k}(n)$ at each iteration is normalised by $\|x(n)\|^2$.

The NLMS algorithm is a time varying step-size algorithm that calculates the convergence factor $\mu(n)$ as in Equation (4).

$$\mu(n) = \frac{\alpha}{c + \|x(n)\|^2} \quad (4)$$

where

α is the NLMS adaption constant, which optimizes the convergence rate of the algorithm and should satisfy the condition $0 < \alpha < 1$, and c is a small positive constant in order to avoid division by zero when the values of the input vector are zero.

The filter tap weights using NLMS algorithm are updated as given in Equation (5).

$$\hat{h}(n+1) = \hat{h}(n) + \frac{\alpha}{c + \|x(n)\|^2} x(n)e(n) \quad (5)$$

3.4 Double Talk Detection Algorithm

DTD detectors produce a detection statistic ξ , which is a function of the input signals i.e. the received signal $x(n)$ and the microphone signal $d(n)$. The detection statistic ξ is compared with a predefined threshold T ; doubletalk is declared if $\xi < T$. Commonly a hold feature is used, i.e. if doubletalk is declared for a sample, the detector continues to declare doubletalk for the next N hold samples, no matter the value of ξ . When $\xi > T$, the filter resumes adaptation. Examples of DTD detectors are the Geigel detector, cross correlation calculations (Benesty and Normalized Cross-Correlation algorithms) and the Variance Impulse Response.

In this work, the Normalized Cross-Correlation (NCC) method is adopted in detecting the existence of the double-talk.

The NCC algorithm computes the decision statistic depending on the relations of microphone signal $d(n)$ and error signal $e(n)$. It can be approached by considering the values of variance of near-end signal $v(n)$, and cross-correlation between error signal and microphone signal.

The decision statistic is given by

$$\xi = \frac{\sqrt{h^T R_{xx} h}}{\sqrt{h^T R_{xx} h + \sigma_v^2}}$$

3.5 Echo Path

The room impulse response (RIR) $h(n)$ of the echo path is determined using the image-source model technique (ISM). The image source method allows the calculation of sound behaviour using ray propagation assumptions, travelling directly from the source to receiver but also indirectly via reflections. It is extrapolated as sound reaching the receiver from two sources, the second source being the image source behind the mirror. The image source position is calculated from the source position and reflector position and angle. The straight line distance between image source and receiver contains the information required to model the actual reflected sound path. Each reflective surface itself produces an image source and a second-order image source is produced by the combined reflection off the two surfaces giving us four effective sources instead of one. This can be extended to a desired number of image sources; the greater the number of image sources the greater degree of accuracy in reverberation time estimate calculation.

The whole early and late reverberation source image positions (x_i, y_j, z_k) are calculated as follows:

$$\begin{aligned} x_i &= (-1)^i x_s + \left[i + \frac{1 - (-1)^i}{2} \right] x_r - x_m \\ y_j &= (-1)^j y_s + \left[j + \frac{1 - (-1)^j}{2} \right] y_r - y_m \\ z_k &= (-1)^k z_s + \left[k + \frac{1 - (-1)^k}{2} \right] z_r - z_m \end{aligned}$$

where

(x_m, y_m, z_m) is distance between the microphone and origin, (x_s, y_s, z_s) is the distance between source and the origin and (x_r, y_r, z_r) is the reflecting wall distance from the origin.

Each echo path distance d_{ijk} is measured from Pythagorean Theorem.

$$d_{ijk} = \sqrt{x_i^2 + y_j^2 + z_k^2}$$

The unit impulse response function of each virtual source is given as

$$a_{ijk}(u_{ijk}) = \begin{cases} 1 & \text{if } u_{ijk} = 0 \\ 0 & \text{otherwise} \end{cases}$$

where

$$u_{ijk} = t - \frac{d_{ijk}}{c}; \quad t \text{ is time and } c \text{ is the speed of sound.}$$

The acoustic properties of the room are characterized by means of a reflection coefficient β for each of the six enclosure surfaces:

$$\beta = [\beta_{x=0}, \beta_{x=x_p}, \beta_{y=0}, \beta_{y=y_p}, \beta_{z=0}, \beta_{z=z_p}].$$

The combined reflection coefficient of surfaces in the x-axis is given as

$$\beta_{x_i} = \beta_{x=0} \left| \frac{z_i - \frac{z_i}{4}(t-1)^2}{z_i + \frac{z_i}{4}(t-1)^2} \right|, \beta_{x=x_p} \left| \frac{z_i - \frac{z_i}{4}(t-1)^2}{z_i + \frac{z_i}{4}(t-1)^2} \right|$$

The combined reflection coefficient of surfaces in the y-axis is given as

$$\beta_{y_j} = \beta_{y=0} \left| \frac{y_j - \frac{y_j}{4}(t-1)^2}{y_j + \frac{y_j}{4}(t-1)^2} \right|, \beta_{y=y_p} \left| \frac{y_j - \frac{y_j}{4}(t-1)^2}{y_j + \frac{y_j}{4}(t-1)^2} \right|$$

The combined reflection coefficient of surfaces in the z-axis is given as

$$\beta_{z_k} = \beta_{z=0} \left| \frac{z_k - \frac{z_k}{4}(t-1)^2}{z_k + \frac{z_k}{4}(t-1)^2} \right|, \beta_{z=z_p} \left| \frac{z_k - \frac{z_k}{4}(t-1)^2}{z_k + \frac{z_k}{4}(t-1)^2} \right|$$

The total reflection coefficient for every virtual source is given as

$$\beta_{ijk} = \beta_x \cdot \beta_y \cdot \beta_z$$

The total magnitude of each echo is

$$g_{ijk} = \frac{\beta_{ijk}}{d_{ijk}}$$

The room impulse response (RIR) $h(t)$ is obtain from the summation of the total magnitude of each echo and unit impulse response function as given

$$h(t) = \sum_{i=-n}^n \sum_{j=-n}^n \sum_{k=-n}^n a_{ijk} \cdot g_{ijk}$$

where

n is the total number of reflection of the sound signal.

3.6 Nonlinear Processor (NLP)

A nonlinear processor is a signal processing circuit or algorithm that is placed in the speech path after echo cancellation in order to provide further attenuation or removal of residual echo signals that cannot be removed completely by an echo canceller. NLP evaluates the residual echo. It removes all signals below some threshold and replaces them with simulated background noise which seems like the original background noise without the echo.

3.7 Flowchart of the Adaptive Acoustic Echo Cancellation System

The flowchart of the adaptive echo cancellation system for teleconferencing is shown in figure 4.

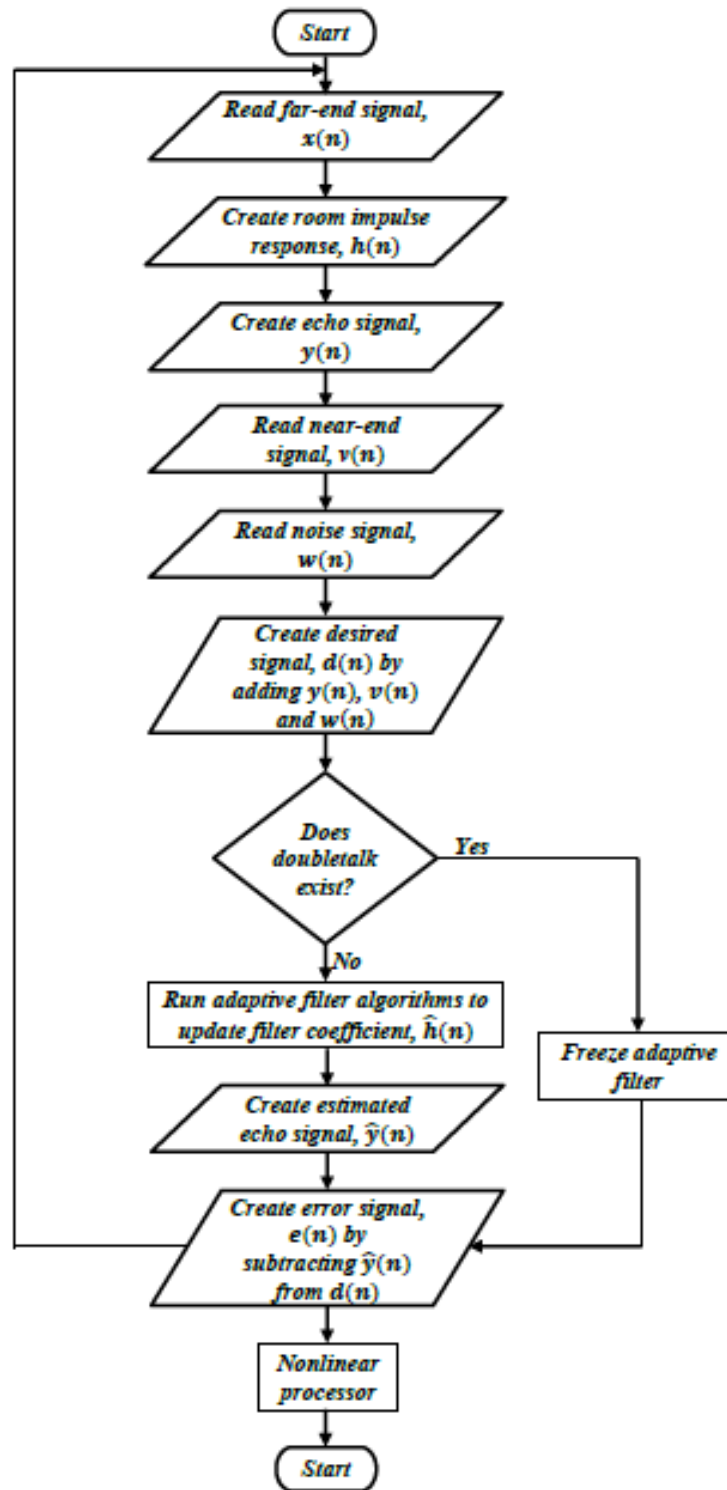


Figure 4: Flowchart of the adaptive echo cancellation system for teleconference system

3.8 Performance Evaluation Parameters

Mean Square Error (MSE): This is the sequence of mean squared error. This column vector contains predictions of the mean squared error of adaptive filter at each time instant.

The mean squared error is calculated as

$$MSE = \frac{\sum e^2}{n}$$

where

e is the error signal and n is number of samples.

Echo Return Loss Enhancement (ERLE): In order to evaluate the quality of the echo cancellation algorithm echo return loss enhancement (ERLE) is used. It measures how much echo attenuation the echo canceller removed from the microphone signal. ERLE is defined as the ratio of the instantaneous power of the signal, $d(n)$, and the instantaneous power of the residual error signal, $e(n)$, immediately after cancellation.

Mathematically, it can be expressed as

$$ERLE (dB) = 10 \log \frac{P_d(n)}{P_e(n)}$$

4. SIMULATION RESULTS

In this work, acoustic echo cancellation using adaptive algorithms LMS and NLMS are implemented using MATLAB. The image source modelling method is used for the calculation of room impulse response (RIR) with the microphone located at (1m, 3m, 2m) and loud speaker located at (5m, 2m, 5.5m) in a 6m x 5m x 6m room size. A predefined far-end speech input signal of sampling frequency 8000Hz is convoluted to the RIR, to generate an acoustic echo signal which serves as the input to the echo canceller and results are taken and plotted for each adaptive algorithm. Near-end speech signal and White Gaussian noise are also introduced into the input signal. Results are taken for the three algorithms.

The results are plotted to demonstrate the performance of the acoustics echo canceller with different algorithms. Plots of far-end signal $x(n)$ echo signal $y(n)$, near-end speech signal $v(n)$, and microphone signal $d(n)$ are first presented. Secondly, plots of the output signal adaptive filter $\hat{y}(n)$ and the error signal $e(n)$ for each algorithm are presented. Plots of performances of the adaptive filter (MSE, ERLE) are also presented.

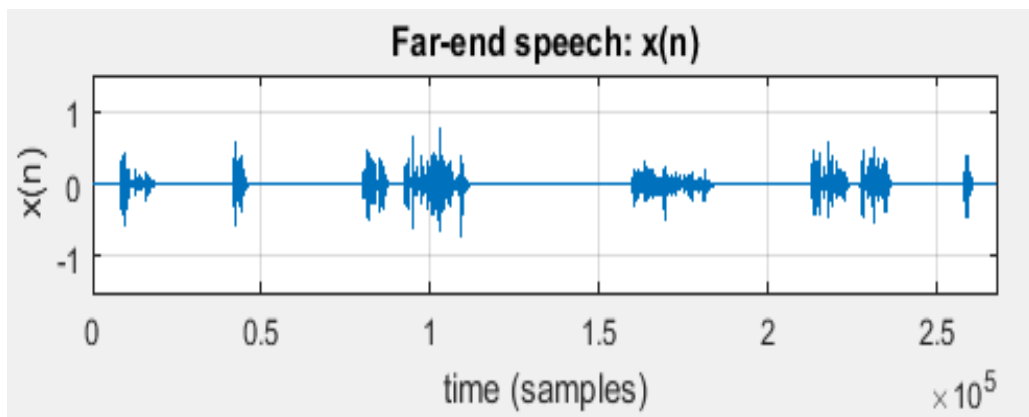


Figure 5: Plot of far-end speech signal

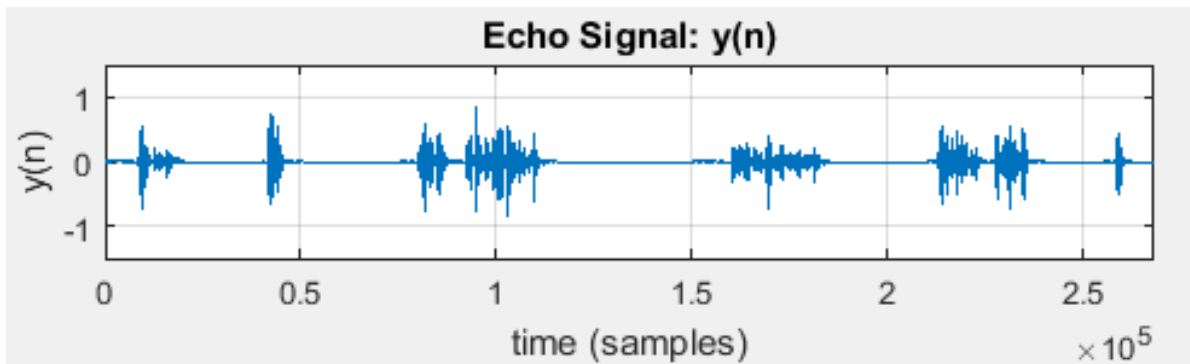


Figure 6: Plot of echo signal

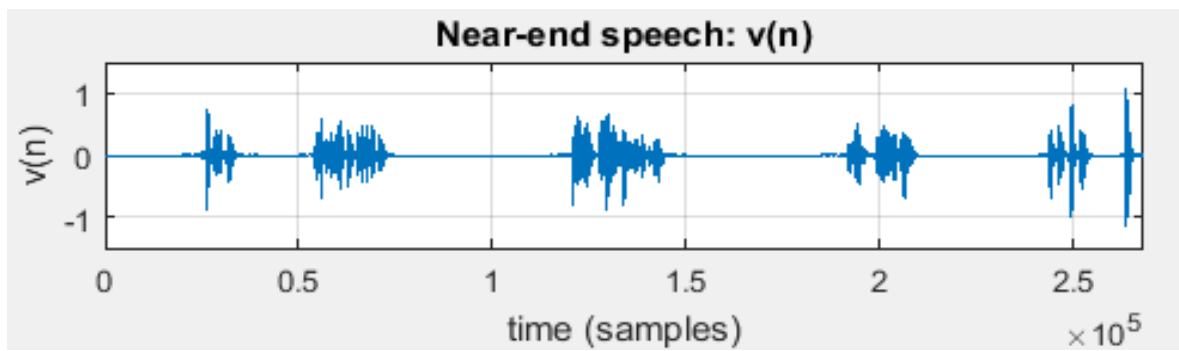


Figure 7: Plot of near-end speech signal

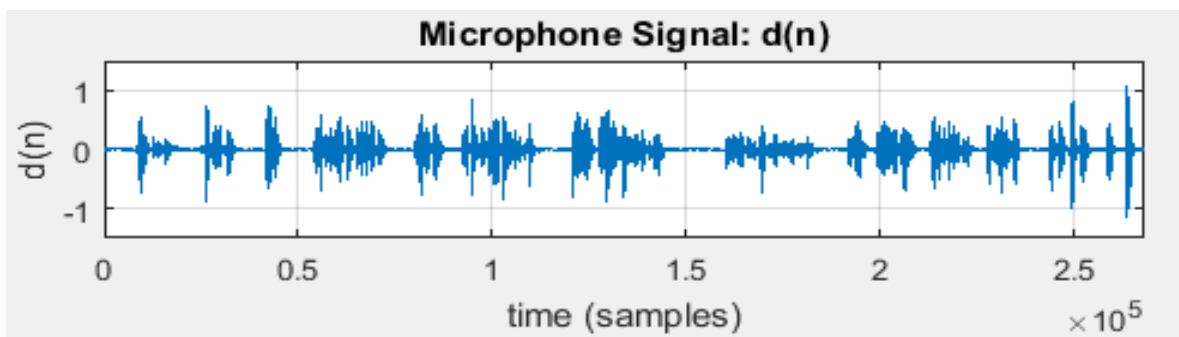


Figure 8: Plot of microphone signal

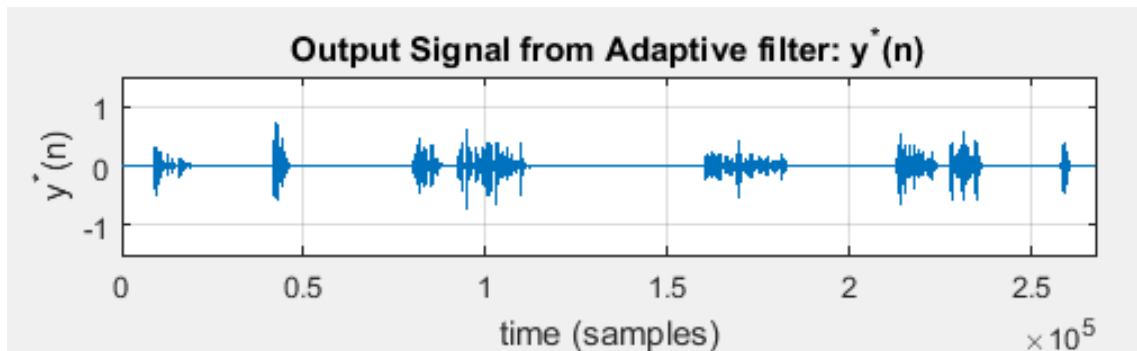


Figure 9: Plot of output signal from LMS adaptive filter

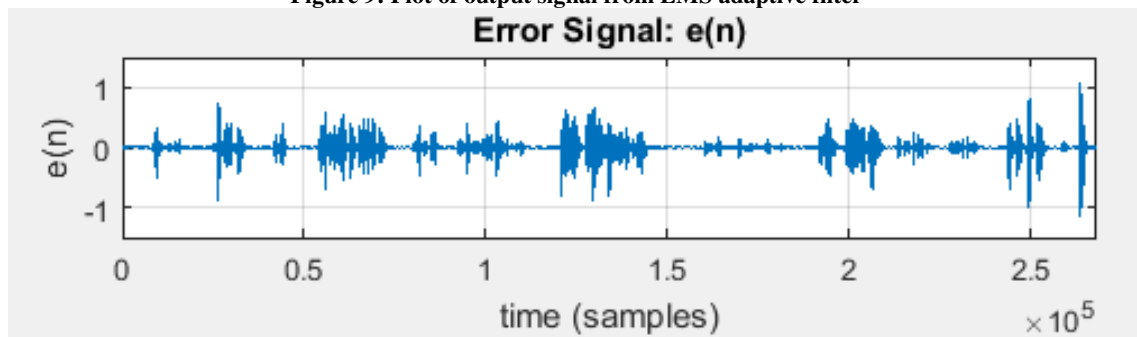


Figure 10: Plot of error signal for LMS

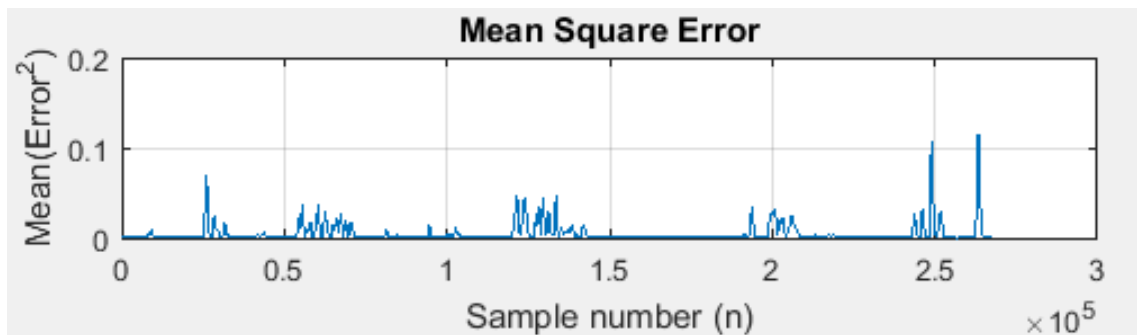


Figure 11: Plot of mean square error (MSE) for LMS

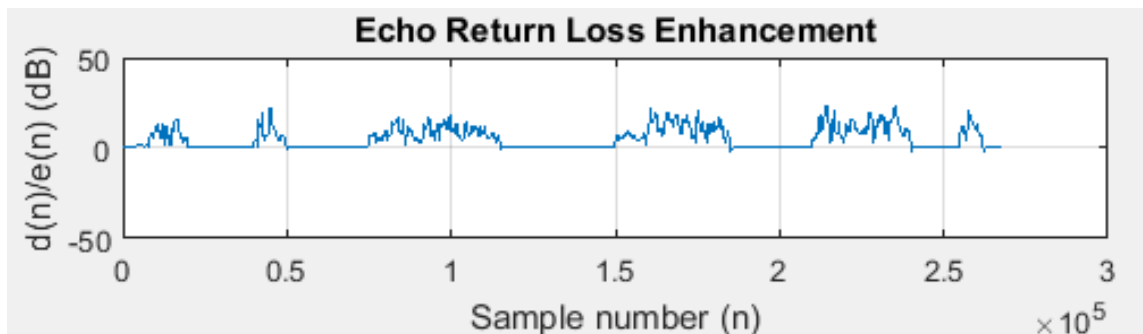


Figure 12: Plot of echo return loss enhancement (ERLE) for LMS

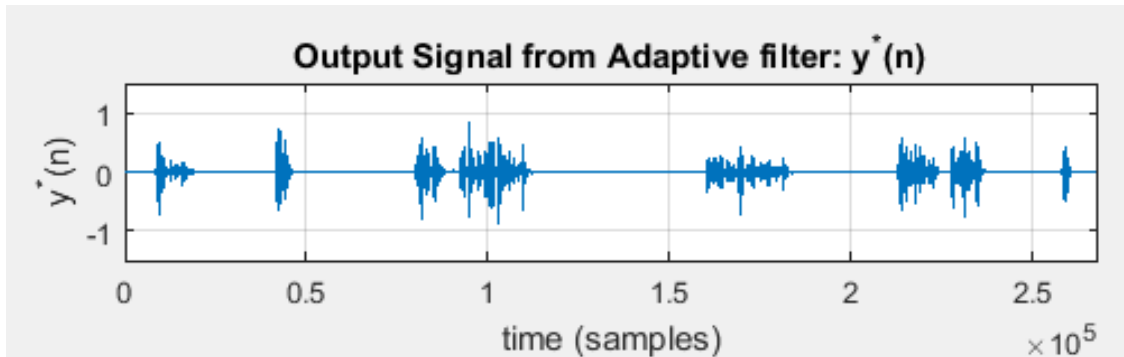


Figure 13: Plot of output signal from NLMS adaptive filter

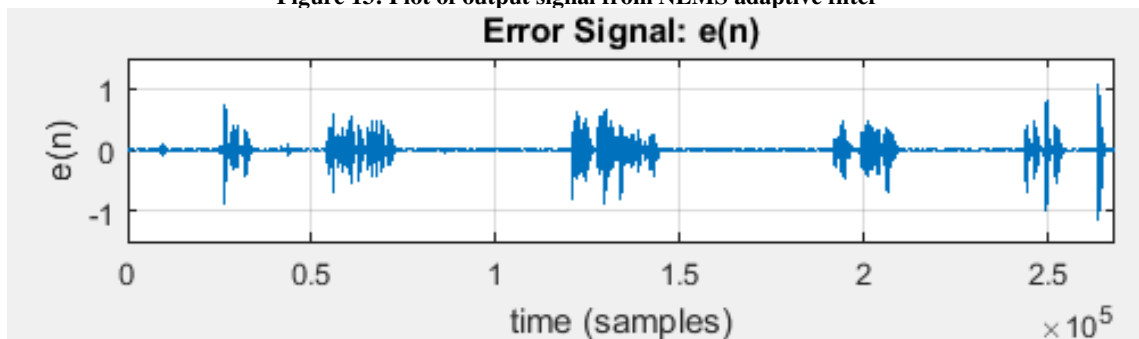


Figure 14: Plot of error signal for NLMS

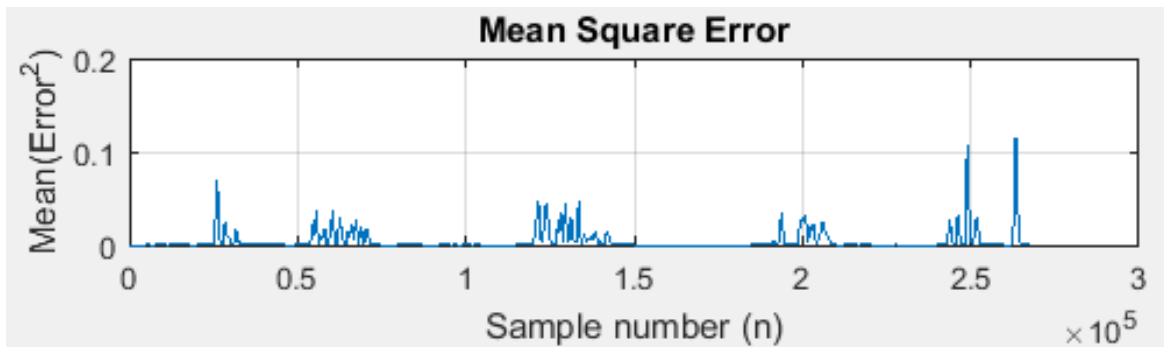


Figure 15: Plot of mean square error (MSE) for NLMS

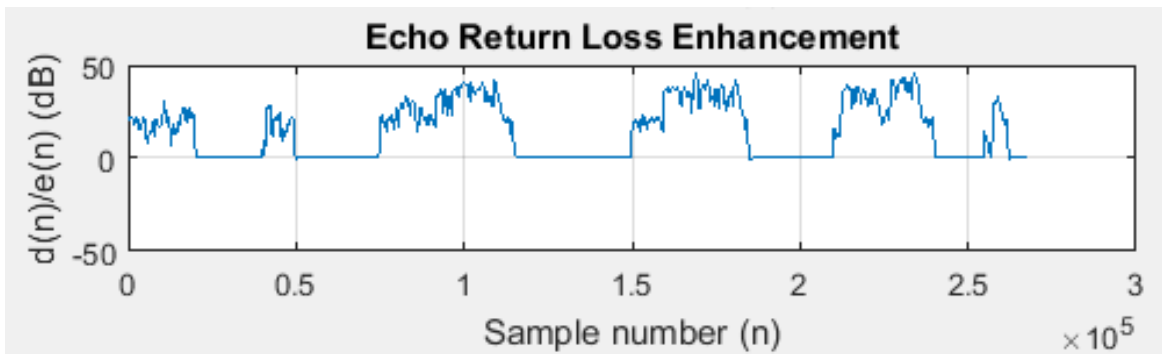


Figure 16: Plot of echo return loss enhancement (ERLE) for NLMS

Considering resultant error signals in figures 10 and 14, it is observed that NLMS cancels the echo signals to a maximum extent, whereas in LMS algorithm, echo signals are cancelled out moderately. The ERLE plots for LMS and NLMS algorithms represented in figures 12 and 16 respectively show that for LMS algorithm, ERLE value lies in the range [-2.0dB, 24dB]; and for NLMS algorithm, ERLE value lies in the range [-1dB, 45dB]. But, ERLE value has to stabilize in the range [- 40dB, 45dB] at 45dB SNR for better performance. Hence, NLMS algorithm offers better performance when compared to LMS. The MSE plots for LMS and NLMS algorithms represented in figures 11 and 15 respectively show that NLMS has a lower value of MSE (0.0042) than LMS (0.0512).

5. CONCLUSION

In this work, an acoustic echo cancellation system for teleconferencing using LMS and NLMS adaptive filtering algorithms with normalized cross correlation (NCC) double talk detection algorithm is presented. The Echo cancellation algorithms were successfully implemented to find a software solution for the problem of echoes in the teleconferencing system. Considering the analysis of the plots for various parameters, echo return loss enhancement (ERLE) and mean square error (MSE) NLMS is observed to have out-performed LMS. The NLMS algorithm has higher the value of ERLE and lower value of MSE than LMS algorithm. Thus, NLMS algorithm is the better the echo canceller and has minimum residual error signal and maximum reduced echo signal to train by using NLP.

REFERENCES

1. Pradhan, S. S. and Reddy, V. U. 1999. A New Approach to Subband Adaptive Filtering, *IEEE Transactions on Signal Processing*, Vol. 47, No. 3, pp. 655-664,
2. K. Sonika and S. Dhull .2011. Double Talk Detection in Acoustic Echo Cancellation based on Variance Impulse Response, *International Journal of Electronics and Communication Engineering*, Vol. 4, No. 5, pp. 537-542.
3. P. Rajesh and A. Sumalatha. 2012. A Novel Approach of Acoustic Echo Cancellation Using Adaptive Filtering, *International Journal of Engineering Research & Technology*, Vol. 1 Issue 5, pp. 1-10.
4. J. Lee and H. C. Huang. 2010. A Robust Double-Talk Detector for Acoustic Echo Cancellation, proceeding of the international Multi Conference of Engineers and Computer Scientists, Vol. 2, pp. 1-4.
5. Raghavendran, S. 2010. Implementation of an acoustic echo canceller using MATLAB. Graduate Theses and Dissertations, University of South Florida.
6. Lu Lu. 2007. Implementation of Acoustic Echo Cancellation For PC Applications Using MATLAB. Stockholm. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.222.2711&rep=rep1&type=pdf> on December 12, 2016.
7. Pushpalatha.G.S and Mohan K. N. 2014. Echo Cancellation Algorithms using Adaptive Filters: A Comparative Study, *Int. J. on Recent Trends in Engineering and Technology*, Vol. 10, No. 2.
8. Constantin P, J and Benesty, S.2010 . *Sparse Adaptive Filters for Echo Cancellation*, Morgan and Claypool Publishers, Georgia. Pg. 3.
9. Guo, M. (2013). *Analysis, Design, and Evaluation of Acoustic Feedback Cancellation Systems for Hearing Aids: A Novel Approach to Unbiased Feedback Cancellation*, Aalborg, Denmark
10. T.A. Vu, H. Ding, and M. Bouchard. 2004. A survey of double-talk detection schemes for echo cancellation applications, *Canadian Acoustics*, vol. 32, no. 3, pp. 144–145.