ISSN 2347 - 3983

Volume 8. No. 9, September 2020

International Journal of Emerging Trends in Engineering Research

Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter232892020.pdf

https://doi.org/10.30534/ijeter/2020/232892020

# Estimation of Reverberation Time by Performing Acoustic Echo Cancellation Considering Near-end and Far-end Speech Signals

**Shreevalli S.**[1]**, Premananda B.S.**[2]

[1]RV College of Engineering, Bengaluru, India, shreevalli95@gmail.com

[2]RV College of Engineering, Bengaluru, India, premanandabs@gmail.com

## ABSTRACT

In teleconferencing, there will be both Near-End (NE) speech and Far-End (FE) speech signals. The FE speech signal bounces inside the room and mixes with the NE speech signal and creates echo during the conferencing. This echo is known as the acoustic echo. There are many methods to mitigate the acoustic echo. This paper adopts LMS algorithm for the acoustic echo cancellation (AEC). The echo return loss enhancement is also performed to find out the quality of echo present in the speech signal after the AEC. The reverberation time of the residual echo embedded in the speech signal is calculated using two methods and those values are compared. The methods adapted are maximum likelihood detection, also known as the online estimation method and the Schroeder method. Various cases of the room impulse response characteristics are considered from Aachen Impulse Response (AIR) database. This paper brings out the idea of estimation of reverberation time by performing AEC. It proves that the error percentage obtained after performing AEC is less than that obtained before performing AEC. The proposed logic gives the reverberation time value of 1.13 s with AEC, which was previously 1.20 s without AEC. This clearly shows that there is 0.07 s reduction in the reverberation time.

**Key words:** acoustic echo cancellation, LMS algorithm, maximum likelihood detection, reverberation time.

## 1. INTRODUCTION

Communication in the real world occurs in a noisy and echoic environment. This echo or reverb builds up due to numerous reflections of sound and decays gradually as it will be absorbed by the surrounding objects such as furniture, water, air, animals etc. The amount of the speech signal reverberated and its duration it lasts mainly depends on the kind of environment along with the objects associated with it. The concept of Reverberation Time (RT) estimation was coined in 1922 as given in [7]. For an explicit determination of RT that is solely dependent on the geometry of the area, used to characterize the quality of an auditory space. RT referred as $T_{60}$, is the time taken by the signal to deteriorate to 60 dB under the value of cessation. RT is used in applications such as designing sound reinforcement systems, sound recording and reproduction. Reverberation is a process of disturbance of the speech signal which is different from additive noise. This leads to a spreading of energy of speech over time, which results in a highly non stationary disturbance. Room's acoustic efficiency's boon or bane is the reverberation. There will be more echoes in the speech material when the RT is low. There are several methods of masking the reverb as mentioned in [11]. Intervals of one second or less is recommended for the classrooms and small lecture halls. Long RTs are suitable for music; the space and the style of music are the dependent factors for the optimum duration.

Reverberation is classified based on the requirement. Based on the area under observation, reverb is classified as rooms, halls, chambers, plates and ambiences. [6] Provides different types of algorithms such as Schroeder's reverb algorithm, Moorer reverb algorithm, etc. that can be chosen to simulate delay networks, computational acoustics and virtual analog modes. These algorithms are expected to have the ability of characterization of listening environment. The RT estimation is designed to be between 0.2 s and 1.2 s as mentioned in [9]. The reason behind the time range for the estimation of RT is that, any room with RT less than 0.3 seconds is called as an acoustically dead room and that room with RT more than 2 seconds is called an acoustically echoic room.

According to the literature survey carried out, it is found that the RT of a reverberant speech signal is calculated without acoustic echo cancellation. In [8], the acoustic echo cancellation is performed using LMS algorithm. [3] Considers Aachen Impulse Response (AIR) database for different cases of room characteristics. RT estimation is done using interrupted noise method, which gives narrow band noise signal as the output. Hence this method is not feasible. The speech signal for calculation of RT is taken from AIR database. In [5], aNE speech and a FE speech signals are considered. They are added and implemented to the LMS algorithm. [1] Uses maximum likelihood detection method for the estimation of RT. Schroeder method of estimation of RT is the tradition way to estimate the RT. This paper combines all these methods, calculates the RT and compares the values from two methods.

This paper aims at calculating the RT value using online method and maximum likelihood estimator (MLE) method, by performing acoustic echo cancellation for the both NE and FE speech signals and tries to reduce the RT value by performing AEC. The organization of the paper is as follows: The methodology is explained in the section 2. Sections 3, 4 and 5 explain the concepts of AEC, estimation of RT and MLE function in detail. Section 6discusses the simulated results. The conclusions drawn arediscusses in the section 7.

## 2. METHODOLOGY

The proposed model of the implementation is as given in Figure 1. A NE and FE speech signal along with echo are combined and passed into the microphone. The far-end speech signal contains echo as it bounces all around the room before it reaches the microphone. The microphone output signal is fed into the LMS algorithm with step size µ=0.22. The absolute value LMS algorithm output is taken and echo return loss enhancement is performed. This signal is fed into RT calculation block and RT is estimated using online and maximum likelihood method.

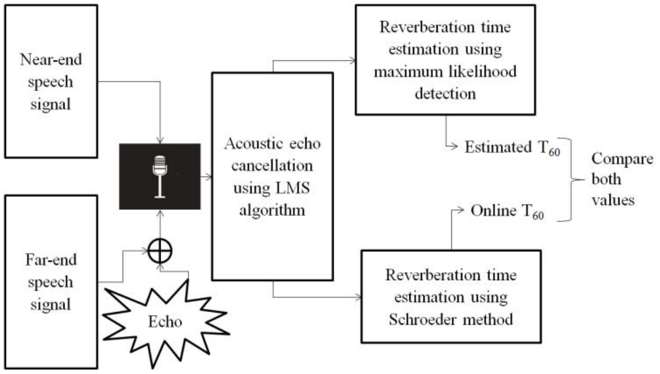

Near-end speech signal
Far-end speech signal
Echo
Acoustic echo cancellation using LMS algorithm
Reverberation time estimation using maximum likelihood detection
Estimated $T_{60}$
Online $T_{60}$
Compare both values
Reverberation time estimation using Schroeder method

**Figure 1:** Block Diagram for Estimation of RT by performing AEC

The RT is estimated in two methods; Schroeder method and maximum likelihood estimator method [8]. Energy decay curve is plotted initially for both the methods. Maximum likelihood estimator uses discrete random process for the estimation of RT. The Online method or Schroeder method uses least square fitting approach for the computation of RT. In the energy decay curve, a line is drawn at -60 dB. The time instant at which the maximum likelihood curve and the Schroeder method curve meets -60 dB line, the instant is considered as the RT for both the methods. RTs with and without AEC is calculated. Both the values are compared and it is proved that RT with AEC is much less than that obtained without AEC.

## 3. ACOUSTIC ECHO CANCELLATION

The first step in AEC is loading of signals.NE and FE speech signalsfrom an audio teleconferencing are considered. As the far-end signal travels all around the room and then mixes with near signal before getting into the microphone, it is said to have echo [7]. This echo caused because of the room impulse characteristics is said as the acoustic echo. If the room has good acoustics then the duration of the echo will be within 0.3 s and 2 s. The duration of speech signal considered is 30 s.
The room impulse response is calculated for the sampling frequency is 8000 Hz. The signals are loaded in .wav format to the function and then converted into discrete time samples. This loading of input signals is illustrated by the flowchart as given in Figure 2.

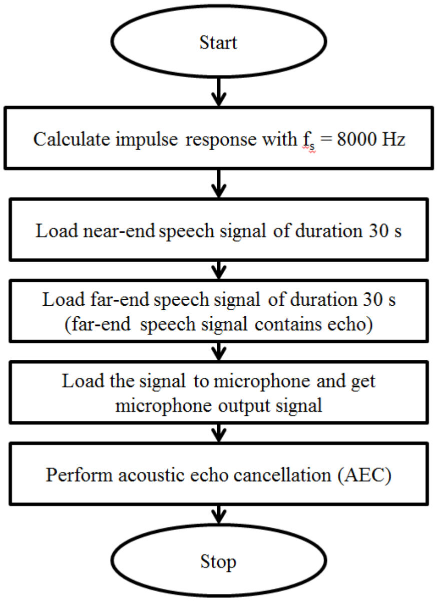

Start
Calculate impulse response with $f_s$ = 8000 Hz
Load near-end speech signal of duration 30 s
Load far-end speech signal of duration 30 s
(far-end speech signal contains echo)
Load the signal to microphone and get microphone output signal
Perform acoustic echo cancellation (AEC)
Stop

**Figure 2:** Flowchart for loading input signals

A very well-known conventional method for AEC is by using LMS algorithm. The step size considered is $\mu$ = 0.22. The flowchart of LMS algorithm is as shown in Figure 3. The reason behind using the LMS algorithm is that it uses negative feedback system to decrease the error signal as given in [2]. The corresponding mathematical equations for LMS algorithm are as given:

$$\mathrm{t(a)} = x^{H}(a-1).i(a) \quad (1)$$

$$\mathrm{f(a)} = \mathrm{s(a)} - \mathrm{t(a)} \quad (2)$$

$$x(\mathrm{a}) = x(\mathrm{a}-1) + \frac{i(a)}{b+i^{H}(a).i(a)} \cdot \mu i^{*}(a) \quad (3)$$

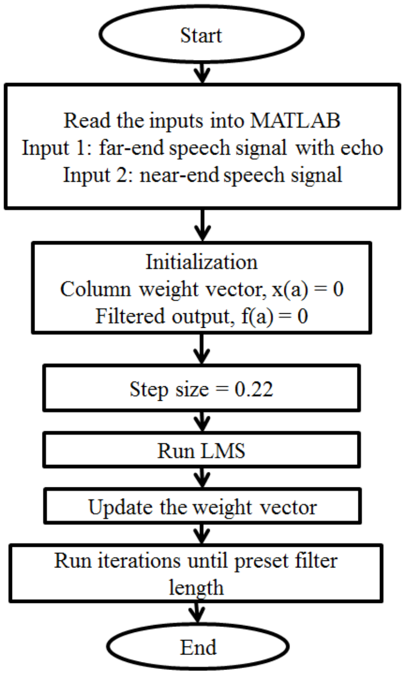

Start
Read the inputs into MATLAB
Input 1: far-end speech signal with echo
Input 2: near-end speech signal
Initialization
Column weight vector, x(a) = 0
Filtered output, f(a) = 0
Step size = 0.22
Run LMS
Update the weight vector
Run iterations until preset filter length
End

**Figure 3:** LMS Algorithm

## 4. ESTIMATION OF RT

The estimation of RT can be done in two methods viz., Schroeder method or online estimation method which uses least square fitting and MLE method which uses frame-wise processing of signals. The detailed descriptions of these methods are given in this section.

### 4.1 Schroeder Method

There are many methodsfor calculating RT on the basis of the sound decay curves. An illustration is the interrupted noise approach, where a narrow band noise is radiated and the recorded decayed curve estimates the RT. But, due to variation inthe excitation noise signal for different trials, we get a large number of decay curves and it has to be averaged to get a reliable estimate. To overcome this problem, the Schroeder method is proposed in this paper, which estimates RT using least square fitting method.

The flow chart depicting Schroeder method is as shown in Figure 4.Sound propagation delay is modeled for the fixed length of time and energy decay curve is plotted. The value of RT is at the -60 dB point on the energy decay curve. Schroeder's method has an immense practical utility; hence there is an improvisation over the years [5]. [9] The online time estimation can be made by using this method. This paper compares the value of RT obtained by Schroeder method and maximum likelihood method.

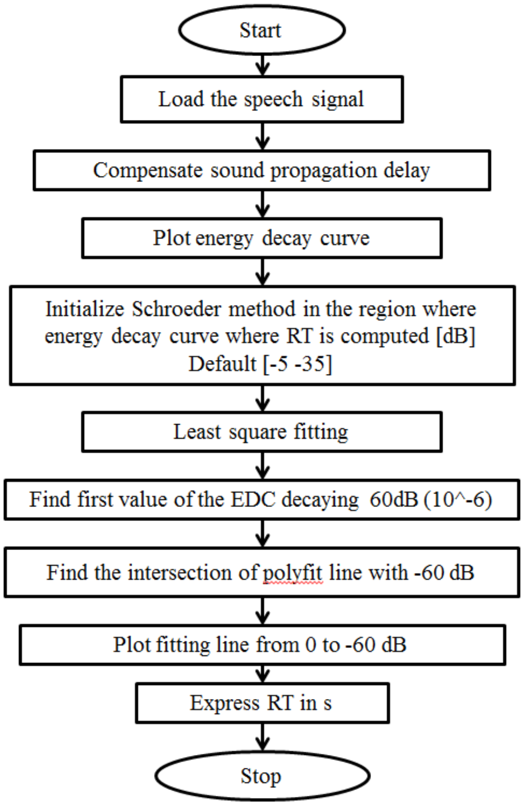

Start
Load the speech signal
Compensate sound propagation delay
Plot energy decay curve
Initialize Schroeder method in the region where energy decay curve where RT is computed [dB] Default [-5 -35]
Least square fitting
Find first value of the EDC decaying 60dB (10^-6)
Find the intersection of polyfit line with -60 dB
Plot fitting line from 0 to -60 dB
Express RT in s
Stop

**Figure 4:** Flowchart of Schroeder method

### 4.2 Maximum Likelihood Estimation

The RT for a single echo-free speech signal is estimated using different room impulse response characteristics taken from AIR database. The proposed maximum likelihood estimator uses frame-wise processing of speech signals for RT estimation. The steps to explain the flow of the algorithm is as shown in the flowchart in Figure 5. A discrete random process of the sound decay modeling of speech signals and the energy decay curve of the continuous time sound decay model is used. The RT obtained should be approximately equal to the time at 60 dB. RT is always estimated at -60 dB point on the energy decay because the measure defines time taken by the intensity of the sound to reduce to 60 dB.

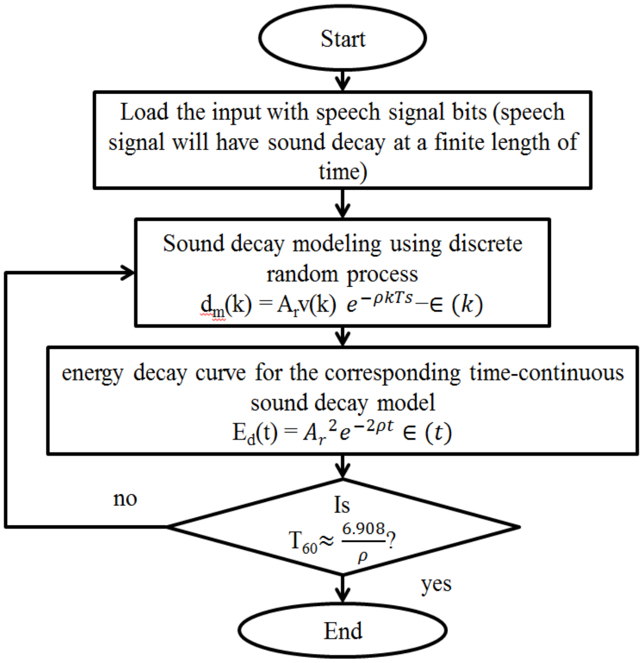

Start
Load the input with speech signal bits (speech signal will have sound decay at a finite length of time)
Sound decay modeling using discrete random process
$d_m(k) = A_r v(k)\ e^{-\rho k T s} - \in (k)$
energy decay curve for the corresponding time-continuous sound decay model
$E_d(t) = A_r{}^2 e^{-2\rho t} \in (t)$
Is
$T_{60} \approx \frac{6.908}{\rho}$?
no
yes
End

**Figure 5:** Flowchart of maximum likelihood estimation method

Let us denote the fine structure by a random sequence x(a), a>0, for each n, a sequence that is deterministic is defined by u(a)>0 as in [4]. The setup for room decay specifies the observations t as t(a) = u(a). x(a). The t(a) is independent due to the time-varying term u(a), but are not distributed identically, and their probability density function is A(0, u(a)) [9]. For a countable sequence of observations, a=0...A-1, the likelihood function of t or the joint probability density, represented by a and s, is given by,

$$M(t;\ u, \sigma) = \frac{1}{\{u(0).....u(A-1)\}} \left(\frac{1}{2\pi\sigma^2}\right)^{A/2} .\exp\left(-\frac{\sum_{a=0}^{A-1}\left(\frac{t(a)}{u(a)}\right)^2}{2\sigma^2}\right) \quad (4)$$

## 5. SIMULATED RESULTS

The algorithms discussed in the previous sections are implemented in the MATLAB. A random room impulse is calculated with sampling frequency of $f_s$= 8000 Hz as shown in Figure 6, whose room impulse response value is considered along with speech signals as input to the LMS algorithm. The magnitude response of room impulse response is shown in the Figure 7.

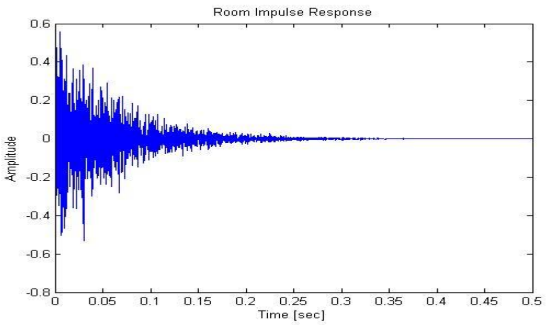


**Figure 6:** Room impulse response with $f_s$=8000 Hz

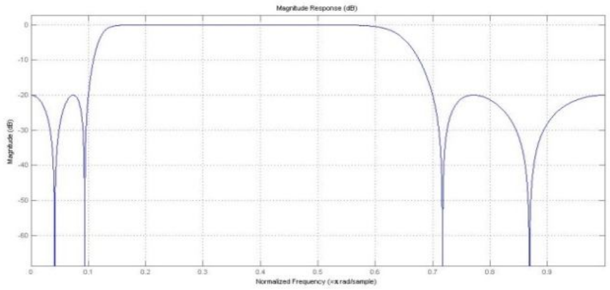
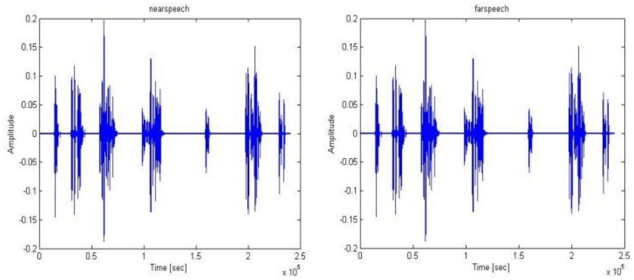


**Figure 7:** Magnitude response of room impulse response with $f_s$=8000 Hz

The input signal is obtained from a duplex teleconferencing conversation. The NE and FE input speech signals are illustrated in Figures 8a and 8b respectively. The graph shows the signals that are converted from .wav file to a discrete time signal format. These two discrete signals are combined. The far-end speech signal contains echo since it bounces inside the room and then reaches the microphone. This combined speech signal is as shown in the Figure 9.



**(a)** **(b)**

**Figure 8:** (a) NE (b) FE speech signal

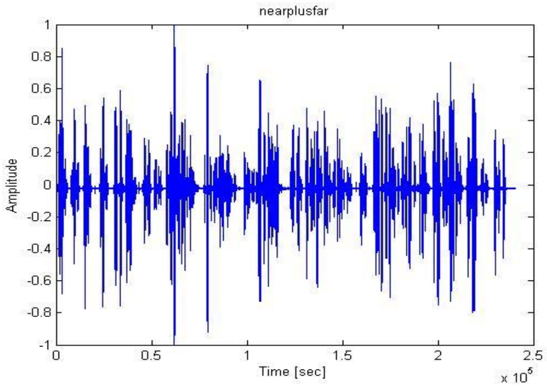


**Figure 9:** CombinedNE and FE speech signals

These two (NE and FE) signals are passed through the microphone and the resulting microphone output signal is as shown in Figure 10. Random bits are mixed with the microphone output signal to reduce the clumsiness from the signal mixture. LMS algorithm is implemented for acoustic echo cancellation. The step size considered is μ = 0.22. The LMS algorithm gives the desired signal and error signal separately. The mean square error, which is the absolute value of the output obtained from LMS algorithm is as shown in Figure 11. The error signal contains residual echo after AEC. The desired signal is free from echo. The quality of echo present in the speech signal is estimated with the help of Echo Return Loss Enhancement (ERLE), and output graph is as shown in Figure 12. The higher the value of ERLE, the better is the cancellation of echo. From the graph, it is seen that the ERLE obtained is 45 dB, which is the highest value obtained by the combination of two signals.

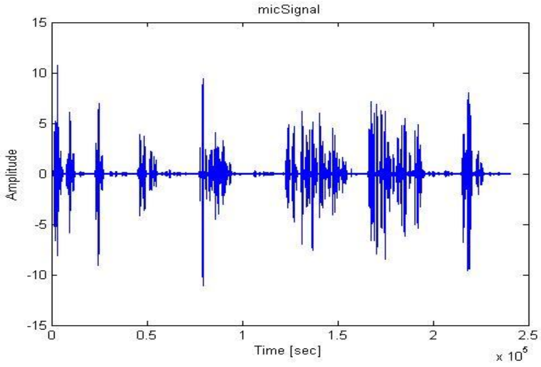


**Figure 10:** Microphone output signal

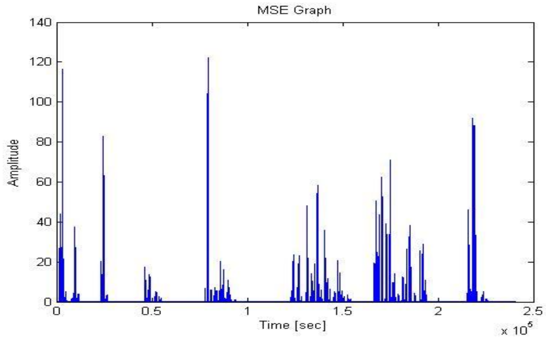


**Figure 11:** Mean square error obtained after AEC

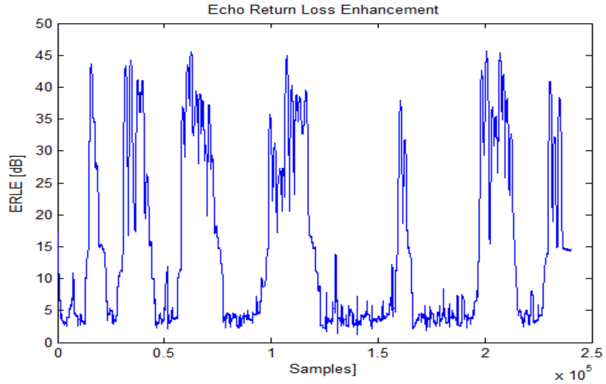


**Figure 12:** Echo return loss enhancement graph

The simulation is carried for several trials. According to the existing system, the speech signal from AIR database is considered. The RT for ML method is 1.08 s, when the simulation is done without AEC. As per the proposed methodology, the simulation is carried out in two attempts, the one with AEC and another trial without AEC. Consider a single speech signal, which has both FE and NE signals, combine both, pass through microphone and estimate RT. Consider the same signal, perform AEC, and then calculate RT. As a comparison, the estimated value matches with online value more appropriately in the latter case then the former one with a time difference of 0.07 s. Hence it can be proved that estimation of RT along with performing AEC is more recommendable.

The RT estimation is carried out in a similar fashion for several other signals such as, NE, FE, and speech signal from AIR data by performing AEC. These readings are tabulated in Table 1. The RT obtained by ML estimator method is 0.97631. The Schroeder method is designed so as to match the ML method estimated values. Table I represents the comparison results of the different trials carried out using speech signals with and without acoustic echo cancellation estimated using Schroeder method. It can be observed the estimated time obtained before AEC is more than the time obtained after AEC. The speech signals considered are of duration 30 s. These signals are divided frame by frame for the calculation of ERLE, after performing AEC. The estimation curve for both the methods is given by Figure 13. Since ML method gives a constant value of 0.97631, there is a straight line and online method calculates the RT frame by frame, different times are obtained. However, the value of online RT is decided by the energy decay curve at -60 dB point. The energy decay curve is shown in Figure 14.

**Table I:** Estimation Table of Simulated Results

| Input signals | RT estimated |
|---|---|
| Signal from AIR database without acoustic echo cancellation [3] | 1.08 s |
| Signal from AIR database with acoustic echo cancellation | 0.91 s |
| NE + FE signals without AEC | 1.20 s |
| NE + FE signals with AEC | 1.13 s |

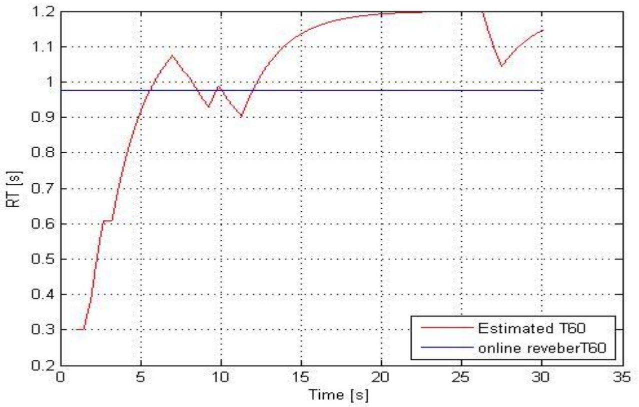


**Figure 13:** RT estimation curves

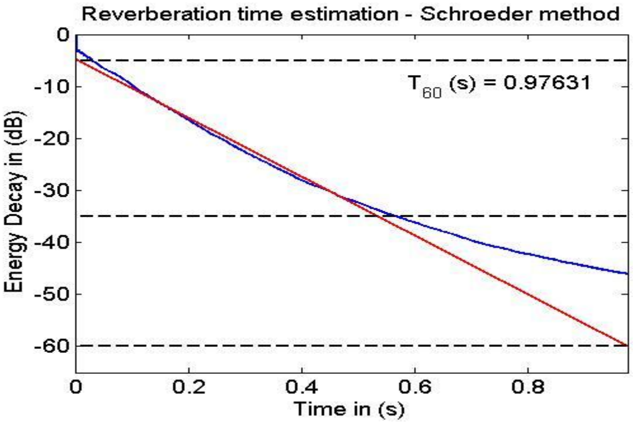


**Figure 14:** Energy decay curve

## 6. CONCLUSION

The proposed system provides an idea for calculating RT using NE and FE speech signals by performing AEC. The existing system estimates the RT using Schroeder method and maximum likelihood detection method. The speech signal considered is from the AIR database. The speech signal used, contains echo. The separation of FE and NE speech signals is performed through Audacity software. LMS algorithm is used for AEC. RT estimation is performed both before and after echo cancellation. RT computed using Schroeder method is 0.97631 s. The value before AEC gives 1.20 s and that after AEC gives 1.13 s. This shows that the error percentage before AEC was 22.369% and the proposed methodology reduces the error percentage to 15.369%.

As a future scope, the residual echo present in the speech signal, after doing AEC, can be calculated and RT estimation can be performed for the signals without residual echo.

## REFERENCES

1. Desiraju N.K., Doclo S., Buck M. and Wolff T. **Online Estimation of Reverberation Parameters for Late Residual Echo Suppression**, *inProceedings of IEEE Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 77-91, 2020.
2. Gulbadhar, NitikaBahel, Shalini Kaur and Harmeet. **Adaptive Algorithms for Acoustic Echo Cancellation: A Review**, *International Journal of Engineering Trends and Technology*, vol. 32, pp. 240-242, 2016.
3. Ratnam. **Blind Estimation of Reverberation Time**, *Journal of the Acoustical Society of America*, vol. 114(5), pp. 2877-2892, 2003.
4. ChulM.Lee, Jong W.S., and Nam S.K. **DNN-Based Residual Echo Suppression**, *in Proceedings of International Science Community Association Conference*, vol. 6(10),pp. 1775-1779, 2015.
5. F. Xiong, S. Goetze, B. Kollmeier and B.T. Meyer. **Joint Estimation of Reverberation Time and Early-To-Late Reverberation Ratio From Single-Channel Speech Signals**, *inProceedings of IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27(2), pp. 255-267, 2019.

6. Kendrick P., Li F.F., and Cox T.J. **Blind Estimation of Reverberation Parameters for Non-Diffuse Rooms**, *Journal of Acoustical Society of America*, vol. 93, pp. 760-770, 2007.
7. Li S., Schlieper R., and Peissig J. **A Hybrid Method for Blind Estimation of Frequency Dependent Reverberation Time Using Speech Signals**, *in Proceedings of International Conference on Acoustics, Speech and Signal Processing*, vol. 72, pp. 211-215, 2019.
8. Lolaee H., and Akhaee M.A. **Robust Stochastic Maximum Likelihood Algorithm for DOA Estimation of Acoustic Sources in the Spherical Harmonic Domain**, *in Proceedings of European Signal Processing Conference*, vol. 26, pp. 351-355, 2018.
9. Wu M., and Wang D. **A Pitch-Based Method for the Estimation of Short Reverberation Time**, *in Proceedings of Acta Acustica United With Acustica*, vol. 92, pp. 337–339, 2006.
10. L¨ollmann H.W., and Vary P. **Estimation of the Reverberation Time in Noisy Environments**,*inProceedings of International Workshop on Acoustic Echo and Noise Control*, vol. 62, pp. 1-6, 2008.
11. Venkataet al.**Estimation of Quality and Intelligibility of a Speech Signal with varying forms of Additive Noise,** *International Journal of Emerging Trends in Engineering Research*, vol. 7(11), 2019.