International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

ACOUSTIC ECHO CANCELLATION FOR E-LEARNING PLATFORM

G.S NITHYASHREE[1], ASHWIN VENKATAKRISHNAN[2], S. KARTHIK[3], ANEESH MOHAN KUMAR[4], DR. R.J ANANDHI[5]

[1,2,3,4] Students, Department of Information Science and Engineering, New Horizon College of Engineering, Bangalore, India
[5]Head of Department, Department of Information Science and Engineering, New Horizon College of Engineering, Bangalore, India

---

Abstract - *Acoustic echo is a common issue with audio conferencing systems. It begins with a local audio loop-back, which occurs when your microphone picks up audio signals from your speaker and transfers them to the other participant along with your speech. The use of teleconferencing services such as Microsoft Teams, Skype, WebEx, Zoom, and others has risen exponentially as the popularity and necessity for working remotely has grown. One of the major causes of poor speech quality ratings in voice and video calls is call quality degradation caused by acoustic echoes. Existing acoustic echo cancellation methods are either not open source or not cross-platform compatible. Multiple algorithms have been implemented using Machine Learning and Deep learning, but the accuracy levels have not met expectations, and deep learning approaches for acoustic echo cancellation are significantly less. This project is implemented using Python. It aims to remove as much acoustic echo as possible in order to improve audio quality for e-learning platforms by employing the Kalman Filter, Normalised Least Mean Square (NLMS), Spectral Gating, and Autoencoder algorithms, as well as demonstrating the accuracies of the various algorithms, so that conversations on online learning and meeting platforms can be more seamless.*

Key Words: Echo Cancellation, Online Learning Platforms, Audio Quality, NLMS, Kalman Filter, Spectral Gating, Autoencoders.

1. INTRODUCTION

With the growing popularity and need for working remotely, the use of teleconferencing systems such as Microsoft Teams, Skype, WebEx, Zoom, etc., has increased significantly. It is important to have optimal quality of communication to make the users' experience satisfying.

The degradation of the quality of digital communication due to acoustic echoes is one of the major sources of poor speech quality ratings in voice and video calls. While digital signal processing (DSP) based acoustic echo cancellation (AEC) models have been used to remove these echoes during calls, their performance can degrade given devices with poor physical acoustics design or environments outside their design targets and lab-based tests.

The current system has many problems and flaws. Existing acoustic echo cancellation solutions are either not open source or incompatible with multiple platforms. The existing system employs the Least Mean Square (LMS) and Recursive Least Squares (RLS) algorithms, both of which don't produce great accuracy levels. The drawbacks of the LMS Algorithm are that it has a defined step size for each iteration. Slow and data-dependent convergence is a problem with the LMS algorithm. Before delivering the data to the adaptive filter, it will be necessary to have prior knowledge of the statistics. Due to the extreme frequent changes in signal power in the audio stream, LMS does not perform well for echo cancellation. The RLS has a fast convergence rate, but this comes at the expense of enormous computational complexity and hence is a very complicated algorithm. To eradicate the issues mentioned above, the approaches we've chosen are NLMS, Autoencoders, Spectral Gating and Kalman Filter. Using these four algorithms we are going to be evaluating the working of different algorithms and to evaluate their performance through mean square error graph.

1.1 Echo Cancellation Algorithms

Normalised Least Mean Square (NLMS)

LMS Algorithm has a fixed step size for each iteration which adds upon to one of its disadvantages. NLMS algorithm is an extension to the least mean square algorithm. It calculates the maximum step size that will help in eradicating the issue found in LMS algorithm. The step size is calculated using the input vector x(n) as follows:

$$Step\ size = \frac{1}{x(n) \cdot x(n)}$$

Kalman Filter

Kalman filter is a recursive algorithm that provides estimations of unknown variables given previously observed measurements. In a system with variables from

different sources that provide information with some uncertainty or inaccuracy, the Kalman filter can help combine the uncertain information from various sources to provide an educated estimate of the next state of the system, which is easier as it incorporates almost all linear calculation except a matrix inversion. The filter can significantly reduce the steady state error level associated with random plant and measurement noise.

Spectral Gating

Spectral gating works based on the principle that states that the input signal is divided into two different ranges of frequency in such a way that one is above and the other is below the central frequency band that is specified using central frequency and also the bandwidth controls. The two signals that are specified above and below the band can be processed individually with the two controls, i.e., low- and high-level control, and the super and sub energy control.

Autoencoder

An autoencoder is a feed-forward neural network which attempts to understand the salient features of the input to approximate an identity function using back propagation. It is an unsupervised learning technique that leverages neural networks for the task of representation learning. The objective of the autoencoder is to reconstruct the input using a condensed, latent representation of the input features. By design, autoencoders reduce data dimensions by learning to ignore the noise in the data.

1.2 Evaluation Metric – Mean Square Weight Error

Mean square weight error (MSWE) is an objective metric used to measure the convergence of the echo path coefficients of the mixed signal with that of the enhanced signal. The coefficients are iteratively updated based on the parameters provided to each algorithm, i.e., NLMS and Kalman filter. The parameters of each algorithm are further discussed in Section 4.

2. LITERATURE REVIEW

Hadei et al. [1] propose the use of the FAP and FEDS algorithms in noise cancellation for speech enhancement, and compare the results of the proposed method with classic adaptive filter algorithms such as Least Mean Square (LMS), Normalised Least Mean Square (NLMS), Affine Projection (AP) and Recursive Least Square (RLS) algorithms. Through the use of optimum algorithm parameters and simulations, the proposed method was observed to perform well in attenuating the noise compared to the adaptive filter algorithms. Simulations were conducted by corrupting an original speech signal with office noise and passing this signal to each algorithm. The corrupted signal had a signal-to-noise ratio (SNR) of -10.218 dB. Using SNR improvement (SNRI) as the metric to compare the performance of the proposed algorithms and adaptive filter algorithms, the results obtained were:

1. LMS – 13.5905
2. NLMS – 16.8679
3. AP – 20.0307
4. **FEDS – 22.2623**
5. **FAP – 24.9078**
6. RLS – 29.7355

It was also observed that FEDS and FAP had faster convergence than LMS, NLMS, AP, and comparable to RLS algorithm. The parameters used to achieve these results were L = 25, P = 8 and u = 0.002.

Zhou et al. [2] propose a real-time residual acoustic echo suppression (RAES) method using a convolutional neural network (CNN) where a double-talk detector is used as a supplementary task to improve the performance of the RAES. In order to preserve the near-end signal while suppressing residual echo, a suppression loss is applied to the signal. The proposed method is also designed to work with adaptive filter algorithms. The results obtained for a double-talk scenario are summarized below using PESQ and STOI as metrics:

1. Speech - 2.816 PESQ – 0.875 STOI
2. Speech+Music – 2.864 PESQ – 0.872 STOI

Another method based on a neural network is proposed by Zhang et al. [3]. The proposed model is a recurrent neural network (RNN) with a bi-directional long short-term memory (BLSTM). Features extracted from the near-end and far-end signals are fed to the BLSTM, which is trained to estimate the ideal ratio mask. This is then used to suppress echo. The advantage of the proposed method is that the AEC problem is addressed without performing any double-talk detection. The evaluation metric used are echo return loss enhancement (ERLE) for single-talk and perceptual evaluation of speech quality (PESQ) for double-talk. A comparison of the performance of the proposed method and NLMS in double-talk scenario is shown below:

1. NLMS – 34.63 ERLE – 4.02 PESQ
2. BLSTM – 51.61 ERLE – 2.74 PESQ

Kothandaraman et al. [4] propose an improvement to AEC using Polynomial Eigen Value Decomposition (PEVD) based adaptive Kalman filter. PEVD is a pre-processing step that produces a strong de-correlation of the signal and removes a part of the noise. To further remove acoustic echo, an adaptive Kalman filter is applied to the signal. This improves the efficiency of AEC for signals with more than 30 dB noise. Compared to applying only the

Kalman filter, which resulted in an average ERLE of 34.2 dB, the proposed PEVD based adaptive Kalman filter achieves 36.7 dB for in-built speech.

3. PROPOSED SYSTEM

3.1 Normalised Least Mean Square

Since NLMS is an extension of LMS, the implementation of the is similar. The following steps take place with each iteration of the algorithm:

1. Adaptive Filters output is calculated using the below formula:

$$y(n) = \sum_{i=0}^{N-1} w(n)x(n-i) = \mathbf{w}^T(n)\mathbf{x}(n)$$

2. The difference between the filter output and the feedback signal gives us the error signal:

$$e(n) = d(n) - y(n)$$

3. Step size is calculated using the following equation:

$$\mu(n) = \frac{1}{\mathbf{x}^T(n)\mathbf{x}(n)}$$

4. The filter weights are calculated for each iteration using:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(n)e(n)\mathbf{x}(n)$$

5. *X(n)* is s the input vector of time delayed input values, *x(n) = [x(n) x(n-1) x(n-2). x(n-N+1)].T*
6. *w(n) = [w0(n) w1(n) w2(n). wN-1(n)].T* represents the coefficients of the adaptive FIR filter tap weight vector at time *n*.
7. The parameter μ is known as the step size parameter and is a small positive constant

3.2 Kalman Filter

The filter is based on five equations, whose one-dimensional form is listed below:

1. *State Update:* $\hat{x}_{n,n} = \hat{x}_{n,n-1} + K_n(z_n - \hat{x}_{n,n-1})$
   Where:
   n is the current iteration
   $\hat{x}_{n,n}$ is the estimate of the current state
   $\hat{x}_{n,n-1}$ is the predicted value of the current state
   K_n is the Kalman Gain
   z_n is the measurement

2. *State Extrapolation:* Depends on the dynamic model of the system. For the example of predicting the position of a truck, the dynamic model is determined by Newton's equation of motion, applied to three-dimensions.

3. *Kalman Gain:* $K_n = \frac{p_{n,n-1}}{p_{n,n-1} + r_n}$
   Where:
   n is the current iteration
   K_n is the Kalman Gain
   $p_{n,n-1}$ is the extrapolated estimate uncertainty
   r_n is the measurement uncertainty

4. *Covariance Update:* $p_{n,n} = (1 - K_n)p_{n,n-1}$
   Where:
   n is the current iteration
   K_n is the Kalman Gain
   $p_{n,n-1}$ is the extrapolated estimate uncertainty

5. *Covariance Extrapolation:* Depends on the dynamic model of the system.

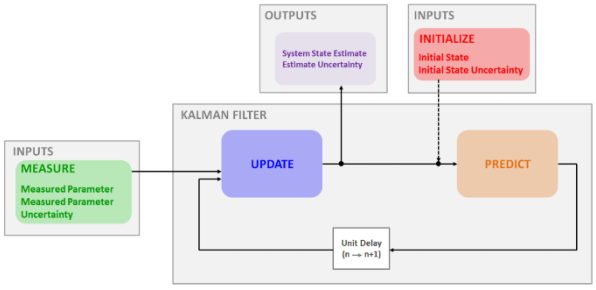

Figure 1: Kalman Filter Low-level Schematic

3.3 Spectral Gating

The first step is to calculate the Fast Fourier Transform (FFT) of the input signal. Fourier transform states that each non-linear function can be considered as the sum of the sine waves. It breaks the time signal and it returns the frequency of all the sine waves needed to simulate that time signal.

Over the FFT of the echo signal, statistics are calculated in terms of frequency such as mean power of the noise, standard power of the noise and noise threshold. Based on the computed statistics, threshold is calculated.

Next, FFT is calculated over the signal that has to be processed and compared to the threshold in order to determine mask. Using filters, the mask is smoothed over time and frequency.

Finally, the mask is applied to the FFT of the signal and then it is inverted. This will help in suppressing the echo from the input signal.

3.4 Autoencoder

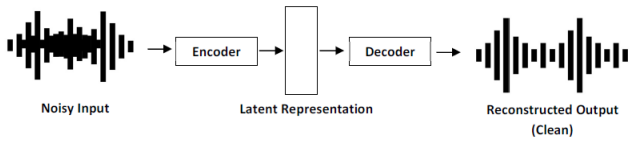

Figure 2: Autoencoder Architecture

Convolutional autoencoders are constructed with convolution layers and deconvolution layers in the encoder and decoder layers respectively. The convolution and deconvolution layers are used to utilize the automated feature learning capability of convolutional neural networks to exploit the structures in input data while preserving its salient information.

Undercomplete autoencoders constrain the number of nodes present in the hidden layers of the neural network in order to limit the amount of information that can flow through the network. Without this constrain, the network would simply copy the features of the input data and provide the output, which is not the goal for any solution. Along with constraining the nodes in the hidden layers, the network is also penalised according to the reconstruction error so that it learns the most important attributes of the input data and reconstructs with a level of high accuracy and precision, without noise.

In the proposed method, the concepts of convolutional and undercomplete autoencoders are applied to the acoustic echo cancellation problem. To begin with, the noisy audio signal is converted to a logarithmic scale using Short-time Fourier Transform (STFT). The STFT of the noisy and clean signals are passed to the autoencoders. The autoencoder comprises of encoder and decoder. The encoder is implemented using 2 convolutional layers and the decoder is implemented using 2 transpose convolutional layers. The signal obtained from the auto encoder is passed to an adaptive filter to further reduce the residual echo.

4. METHODOLOGY

I. SYSTEM ARCHITECTURE

In the proposed method, the far-end and near-end signals are mixed with echo path coefficients using the convolve operation. The mixed signal is passed to the algorithms mentioned in the previous section in order to retrieve the enhanced signal. The signal and its corresponding mean square weight error (MSWE) graph are obtained to understand the echo cancellation performance of each algorithm.

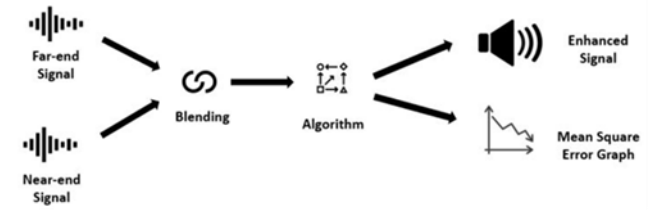

Figure 3: System Architecture

As mentioned in Section 1, MSWE measures the convergence of the echo path coefficients of the mixed signal and the enhanced signal. In each algorithm (NLMS and Kalman filter), the parameters provided to control the iteration and consequent update of coefficients are – 1000 for NLMS and 64 for Kalman filter.

II. WORKFLOW

The application when launched finds the 3D elements within its camera frame. These 3D elements, have properties that allow them to be dragged and dropped along the axes. The active elements in the environment when interacts with each other, like the test tubes and the flame, a change in the color of the content of the test-tube occurs. Until this change is noticed the elements have to be brought near each other. The change in the color of the elements is not spontaneous and is only triggered when it comes in contact with the flame.

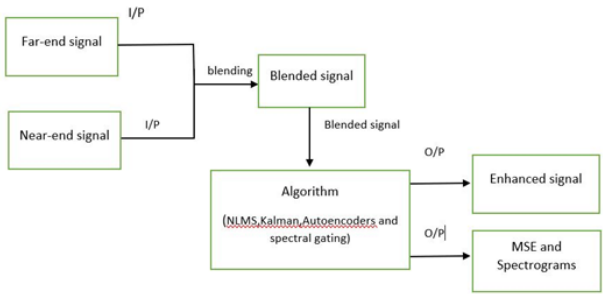

Figure 9: Work flow

5. EXPERIMENTAL RESULT

Figure 10 and 11 depicts the MSWE graph obtained for the audio enhanced using NLMS algorithm and Kalman filter, respectively.

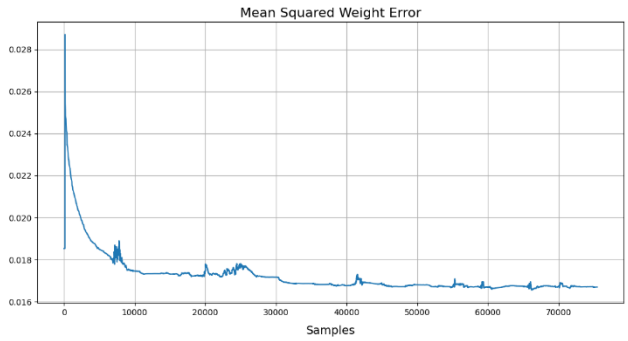

Figure 10: MSWE Graph for NLMS Algorithm

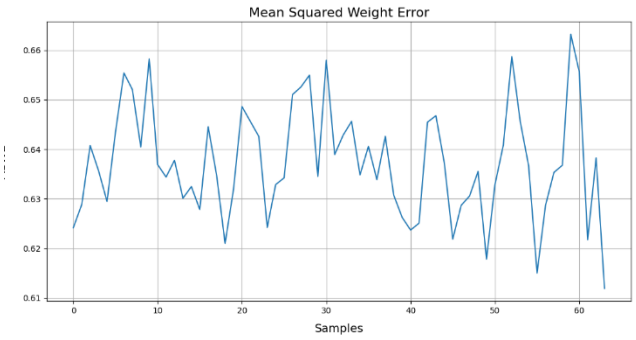

Figure 11: MSWE Graph for Kalman Filter

Figure 12 is an overlapping plot of the noisy and enhanced signal obtained using Spectral Gating. The grey signal is the enhanced signal.

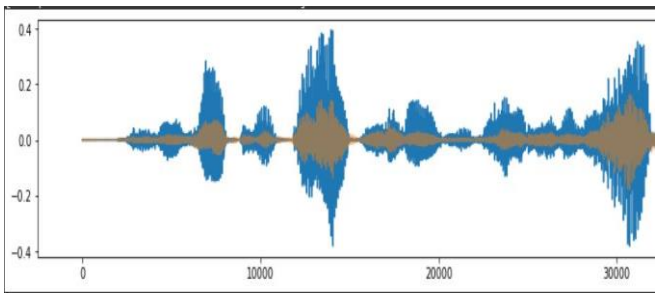

Figure 12: Noisy+Enhanced Signal Plot for Spectral Gating

Figure 13 depicts the MSWE graph obtained for the audio enhanced using the proposed autoencoder.

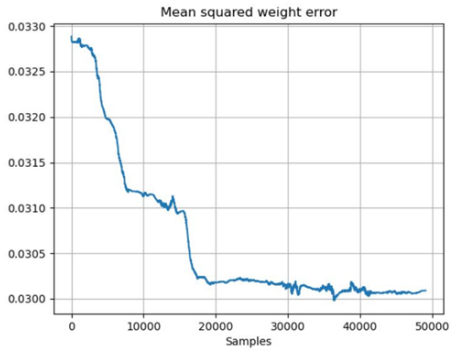

Figure 13: MSWE Graph for Autoencoder

6. CONCLUSIONS

The usage of conferencing devices, wireless audio, meeting platforms and electronic speakers are only gaining popularity day by day. Acoustic echo cancellation is very important in making sure that there is a seamless interaction between the speaker and the listener. Acoustic echo cancellation removes the echo, reverberation and unwanted noise caused by acoustic coupling between the microphone and loudspeaker.

The proposed system has potential for future enhancement. It can be integrated to e-learning and online meeting platforms, so that telephonic conversations and meetings can be more seamless and students can focus and learn much more effectively without distractions. The performance of the algorithm can be improved to work during real time conversation. Alternate implementations using deep learning techniques such as Generative Adversarial Network (GAN) and Long Short-term Memory (LSTM) can be explored. Access to subjective metrics such as Surreal, DECMOS can help in better evaluation of our model. To facilitate other platforms to easily integrate our application, an API can be built.

7. REFERENCES

[1] Hadei, S. (2011). A family of adaptive filter algorithms in noise cancellation for speech enhancement.

[2] Zhou, X., & Leng, Y. (2020). Residual acoustic echo suppression based on efficient multi-task convolutional neural network.

[3] Zhang, H., & Wang, D. (2018). Deep learning for acoustic echo cancellation in noisy and double-talk scenarios.

[4] Kothandaraman, M., Pawani, J. K., Pachaiyappan, A., & Sankaran, S. G. (2017). Acoustic echo cancellation using PEVD based adaptive Kalman filter.