A COMPARISON OF NOISE CANCELLATION OF SPEECH SIGNALS USING ADAPTIVE LMS ALGORITHM AND DEEP LEARNING ALGORITHM

Anand Suresh[1], Farhana S[2], Amjatha Ali[3], Asif Khan[4] ,Sajini T[5]

[1] *Student, Electronics and Communication Engineering, K.M.E.A Engineering College, Kerala, India*
[2] *Student, Electronics and Communication Engineering, K.M.E.A Engineering College, Kerala, India*
[3] *Student, Electronics and Communication Engineering, K.M.E.A Engineering College, Kerala, India*
[4] *Student, Electronics and Communication Engineering, K.M.E.A Engineering College, Kerala, India*
[5] *Professor, Electronics and Communication Engineering, K.M.E.A Engineering College, Kerala, India*

ABSTRACT

In the perspective view of communication system, noise can be defined as any undesired signal that is present in the system. The amount of noise should be reduced to a significant extend for proper transmission and reception of the signals. There are various types of methods and algorithms to remove different types of noises in which the background noise cancellation algorithms are of a major concern. This project proposes a novel comparison of the two majorly used noise removal techniques namely Adaptive noise cancellation using Least Mean Square filters and noise cancellation using Deep Learning Networks (Convolutional Neural Networks).The paper also estimates which method is more suitable for noise cancellation. Stimulation results shows that noise cancellation using deep neural networks is found to retrieve the messages more efficiently than that compared to the traditional Least Mean Square filters. The comparative analysis is done on the basis of parameters like Signal to Noise Ratio, Average power and Frequency response.

Keyword: - *LMS Filter ,Neural Networks, Deep Learning, Convolutional Neural Networks, SNR and Spectrogram*

---

1. INTRODUCTION

The major factor affecting a cellular system or any other wireless communication system is the presence of background noise in the received signal. This background noise can come under a wide range of frequency spectrum and may include noises like environmental noise, industrial noise, traffic noise, certain white Gaussian noise etc. Since these noises fall in a wide spectral range it is insufficient to use the traditional low pass filters to remove these type of noises therefore to resolve this problem we require a filter that adjust itself in accordance to the input signal, these type of filters are known as the adaptive filters. The filtering of the signals by adaptive means is done by adjusting the coefficients of the filter in accordance to the input signals. There are various types of adaptive filters however the commonly used adaptive filter for noise cancellation is the Least Mean Square adaptive filter. As stated in [1] the adaptive filter works by computing the error signal of the input signal corrupted by the noise and a reference noise signal with zero mean and correlation. The major part of any adaptive filter is the algorithm which is used to change the coefficients in accordance to the input signal. The inputs to this adaptive algorithm are the reference noise and the error signal. Depending on the nature of the algorithm it can be LMS, Hybrid-LMS, VSS-LMS etc., [4]. The major advantage of the LMS filter over other filters is the ease of computational complexity for example the RLMS algorithm has a high convergence rate but the computational load is quite high. The major application of a typical adaptive filter includes system identification, noise cancellation, etc. In the view of noise cancellation, adaptive filtering algorithms are mainly used in cellular phones, headsets [5], video conferencing [6], etc.

During the time of digital evolution the demand for data has gained an exponential interest in the field of data analytics, computer communications, statistics etc. these factors give rise to the evolution of machine learning to a great extent. Simply stated machine learning is the process of training a computer or a device to attain a suitable objective. The sole part of any machine learning method is the data on which the machine learns, depending on the nature of data, it can be either labeled or unlabeled data. There are various types of machine learning depending upon the implementation, data, and training process [8], [15]. The most popular way in implementing a machine learning process is by the help of neural networks. The neural networks are analogous to the neurons in our brain, the major part of the neural network is the input layer, hidden layer and output layer [16]. The basic working of the neural network involves adjusting the weight of nodes in a neural network in accordance to the error signal which is formed as a result of the difference between input and output. The input is normally multiplied by a weight factor and is added with a bias factor. Depending upon the number of layers in the network it can be classified either as a single or multi-layer neural network. The applications and scope of machine learning is high it varies from healthcare to defense fields [15]. Deep learning is a modified version of the machine learning process, as the name indicates the concept of deep learning arises when the number of hidden layer increases. Similar to machine learning there are various types of deep learning algorithms [9].

A Convolutional Neural Network (CNN) is a class of deep neural network in deep learning [10]. It is mainly used for applications which involve identification and classification of images. The working of the CNN includes two major phases. They are convolution and pooling. The major function of the convolutional layer is to extract the unique features of the images like size, contrast, brightness, etc. The convolution is done with the help of specialized filters. The function of the pooling layer is to reduce the size of the image. A more detailed insight to the convolutional layer is given in [13].

2. LITERATURE REVIEW

There are various algorithms and approaches to eliminate the background noise in a received signal. The article [1] illustrates a typical background noise cancellation process. Here the author uses a typical LMS adaptive filter to eliminate the noise with the help of a known reference sound. The paper also put forward a hardware implementation of the proposed method using an embedded processor.

Several other algorithms like the spectral subtraction method [2] & [3] also provide an optimum cancellation of the noise. In [2] author Firdauzi proposes spectral subtraction implementation schemes using a typical FPGA board. Their stimulation results show that a Signal to Noise Ratio of approximately 71dB is obtained from the recovered signal. It is further more observed that their experiment improves the SNR of the recovered signal to a great extend such that it further enhances the quality of the received signal. A. L. L. Ramos [3] provides an alternative use of spectral subtraction method in the field of gunshot acoustics. The major objective of their paper is to model and eliminate a noise spectrum of the gunshot sound, furthermore it has been observed that the proposed algorithm can be used to improve the position of the sniper gun systems.

There are various types of adaptive algorithms that are in use. A performance comparison regarding their performance, computational power, efficiency, noise cancellation capability etc. are discussed in [4], moreover the paper also put forward a new type of algorithm similar to CS-LMS algorithm. The results of stimulation of their work prove that the new algorithm outperforms the existing algorithms like N-LMS and CS-LMS. More practical applications of the adaptive filters are pointed out in [5] & [6]. In [5] the author applies the adaptive filtering technique to develop a noise cancellation headset. The paper makes use of the Variable Step Size adaptive filter to cancel out the background noise in the headset. Furthermore the results show that headset is capable in cancelling out the extraneous noise to a wide extent. The application of adaptive filter in video conferencing use is illustrated in [6], in which the authors try to eliminate the background noise that is present during a video conferencing. Their results show that the proposed scheme is a cost effective method and is able to reduce the noise to a factor of 38dB. [7] Proposes a novel comparison of the software stimulation tools in Matlab in the stimulation of the adaptive noise cancellation. Their work suggested that both tools (Simulink and Data Acquisition toolbox) in Matlab show a good rejection of noise signals during stimulation process.

In [8], [9] and [15] the authors presents a briefing about the concepts of machine learning, deep learning and its applications. A more insight to the machine learning implementation is given in the IJCNN Conference in Neural networks [16], where the paper discusses different aspects of neural networks, its challenges and its types. One of the most commonly used neural network is the Convolutional Neural Network (CNN). The detailed analysis, working and performance of the CNN are illustrated in [10]. The various methods in speech de-noising is done with the help of deep learning [11], [12] and [13]. All of these papers utilize different methods of speech de-noising like deep cerebral articulation, hybrid DSP learning, convolutional network etc. Of all the above mentioned papers deep learning techniques is found to reduce the noise present in a speech signal to a very large extent. A more realistic approach towards deep learning is given in [14] where the authors have developed a smartphone application that enhance the speech signals thereby helping partially impaired people.

3. PROPOSED METHOD

The major objective of the project is to compare the performance of the existing Least Mean Square Filter and Convolutional speech de-noising algorithm. For the stimulation of the performance, Matlab signal processing toolbox and deep learning toolbox are utilized. Since either of the mechanisms would require a reference noise, we normally choose a washing machine noise as a reference noise. For the ease of simplicity the methodology is further divided into two parts namely LMS Filtering and speech de-noising using deep learning.

3.1 LMS Filtering

The simplified architecture of a typical LMS filtering mechanism is shown in Figure 1. Here the input signal consist of the mixture of the message signal and the known or unknown noisy signal. The signal is then given as the input to the adaptive algorithm (LMS). The reference noise is also given as input to the algorithm and the filtered signal is obtained at the output.

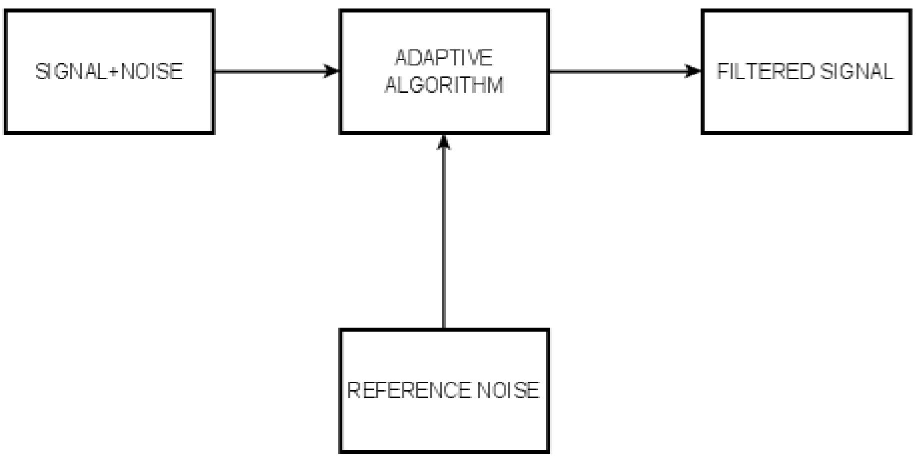

Fig -1 Basic adaptive filtering method

A more detailed architectural insight to the working of the LMS Adaptive filter is shown in Figure 2

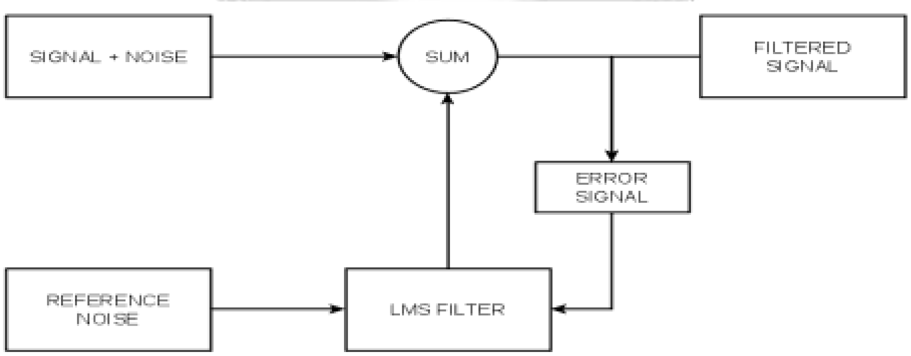

Fig -2 Detailed block diagram of the LMS Filter

As indicated in Figure 2 the reference noise is given to the LMS filtering algorithm and the output of the filter is added with the input signal which comprises of the message signal and the noise. The error signal is obtained as the difference of the input signal and the filter output. This value of this error signal is then used to adjust the coefficients of the LMS Filter. This process repeats itself till the noise is completely eliminated.

3.2 De-Noising Using Deep Learning

The details to the working of the deep learning networks and convolutional neural network are given in [8], [9] and [10]. To reduce the complexity in analysis of the networks we normally choose Convolutional Neural Network (CNN) as a basis for our experiment. The overall block diagram of the speech de-noising process is shown in Figure-3

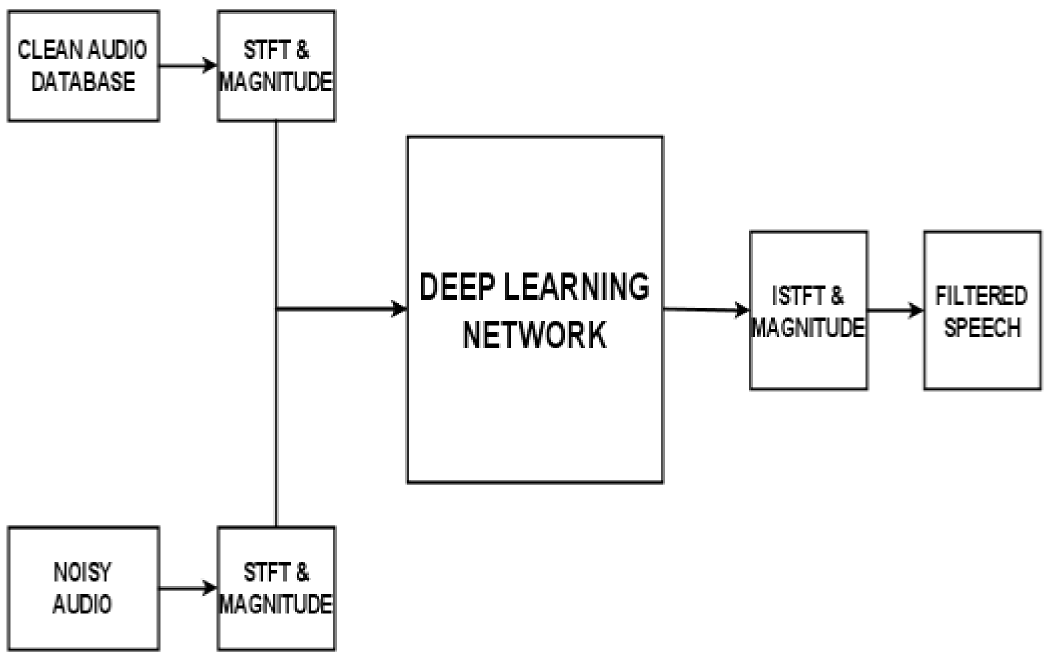

Fig -3 Audio de-noising process

The overall process of the noise cancellation in speech or audio signals is shown in the above figure. Here initially we create a typical data base consisting of clean audio samples spoken by several people in different localities. The Short Time Fourier Transform (STFT) of these audios are taken and plotted. The magnitude of the STFT is also computed. The processed data then acts as a target to the network. The target acts as the expected output of the neural network. In a similar manner the STFT and magnitude of the reference noise is also computed and it acts as the predictor. The predictor usually has a relation or mapping to the target variable. These data's are then fed as input to the deep learning network.

In our analysis the type of deep learning network used is the Conv-Net. The CNN used in our analysis contains 16 convolutional layers out of which the first 15 layers are repetitions of three groups of the convolutional layer. There are a total of 18, 30 and 8 filters with a filter width of 9, 5 and 9 respectively. The last CNN has one filter having a width of 129. Like all other neural nets a typical activation function like ReLu is used.

4. RESULTS

To analyze the performance of the proposed system parameters like SNR, Frequency Response and RMS values are chosen. For the ease of analysis the parameters are characterized into two sections

4.1 Performance of LMS Filter

Following Figure 4 shows the typical results obtained for noise cancellation using LMS Filter. The results basically involves the time domain representation of the clean audio, noisy audio, audio corrupted with the reference noise and the filtered signal

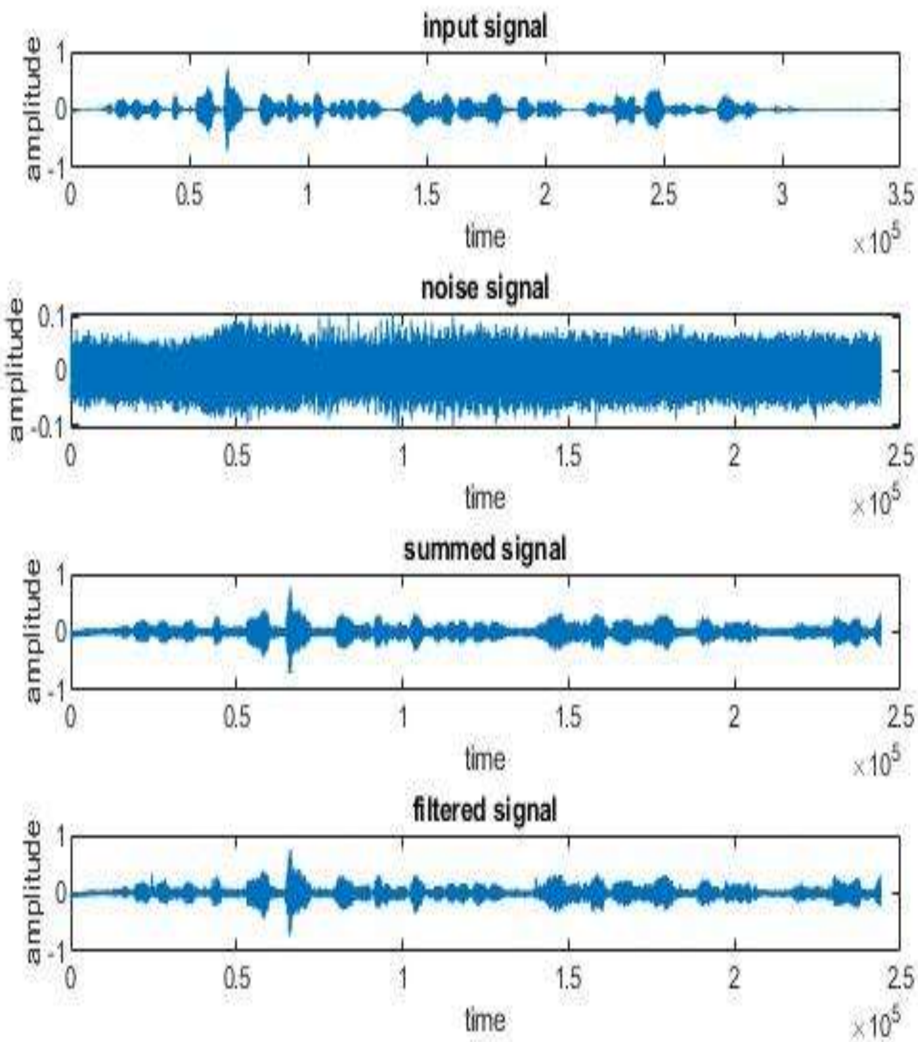

Fig -4 Experimental outputs of LMS Filter

As it's clear from the figure 4 that using LMS filter we are able to retrieve the original message signal, but however there is a small fraction of the reference noise in the output signal. Therefore we cannot completely state that the output signal is devoid from the reference noise. An analysis of the signal to noise ratio (SNR) of the input and output signal indicates that the input signal (clean audio) is having SNR of approximately 4.5dB. After addition and filtering of the reference noise is done, the filtered output is found to have SNR of approximately 4dB. Thus there is an approximate loss of 0.5dB in the message signal after filtering. Figure 5 and Figure 6 shows the frequency response plot of the input and output signals using LMS Filtering.

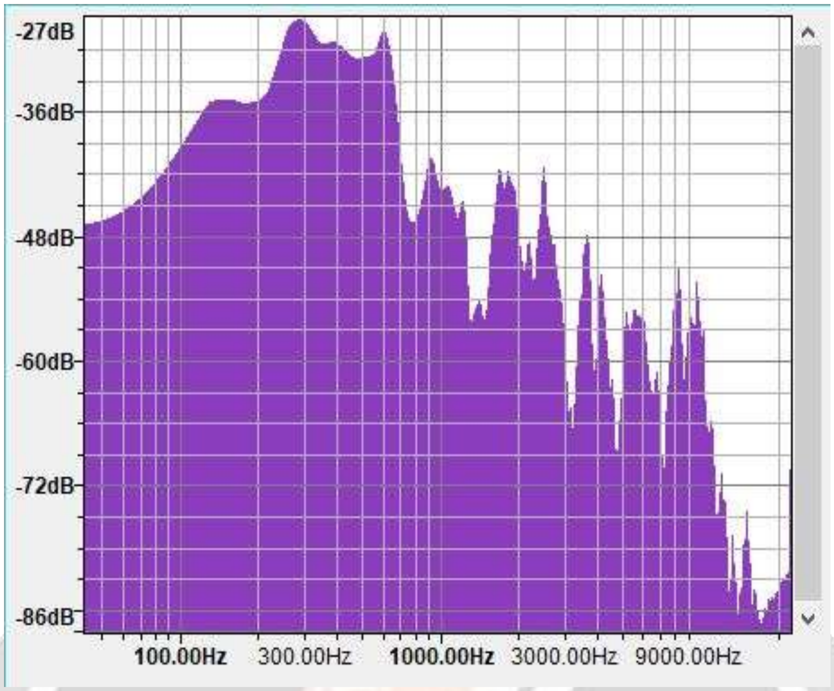

Fig -5 Response of input signal (audio)

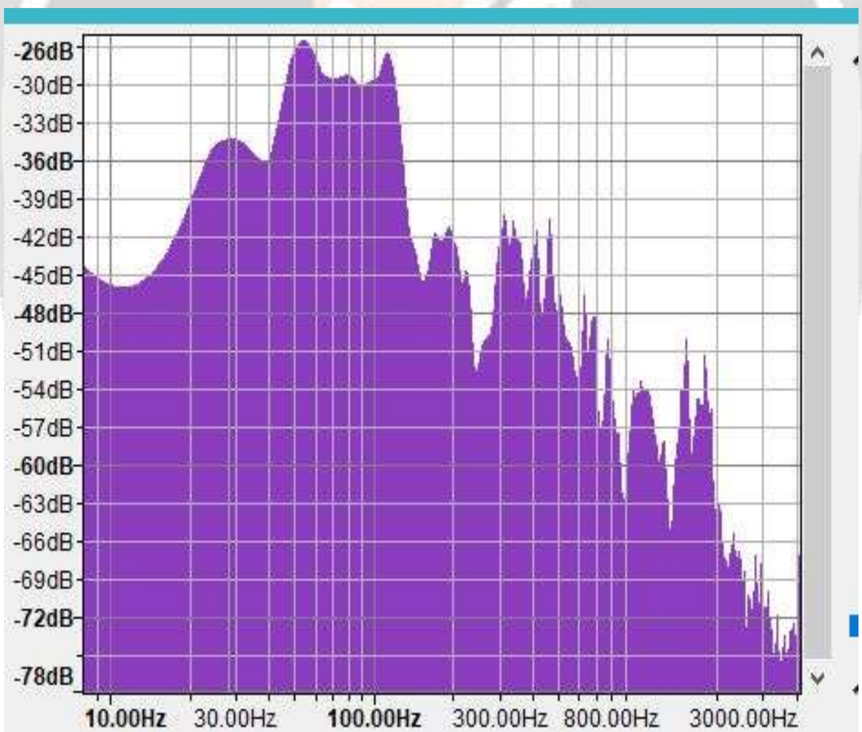

Fig -6 Response of filtered output signal (audio)

It is obviously clear from the response that both the responses vary slightly in their respective amplitudes. Also a measure of Root Mean Square value (RMS) can be computed from the above two responses. It has been found that the input audio is having an RMS value of -23.79dB whereas the filtered output is having an RMS value of --22.36dB. Thus both of the RMS values are different which further more proves that there's some amount of reference noise present in the output.

4.2 Performance of Deep Learning Network

Figure 7 shows the results obtained for deep learning networks in filtering a noised signal (audio). The figure includes the plot of clean audio, noisy audio and filtered audio along with their respective spectrograms.

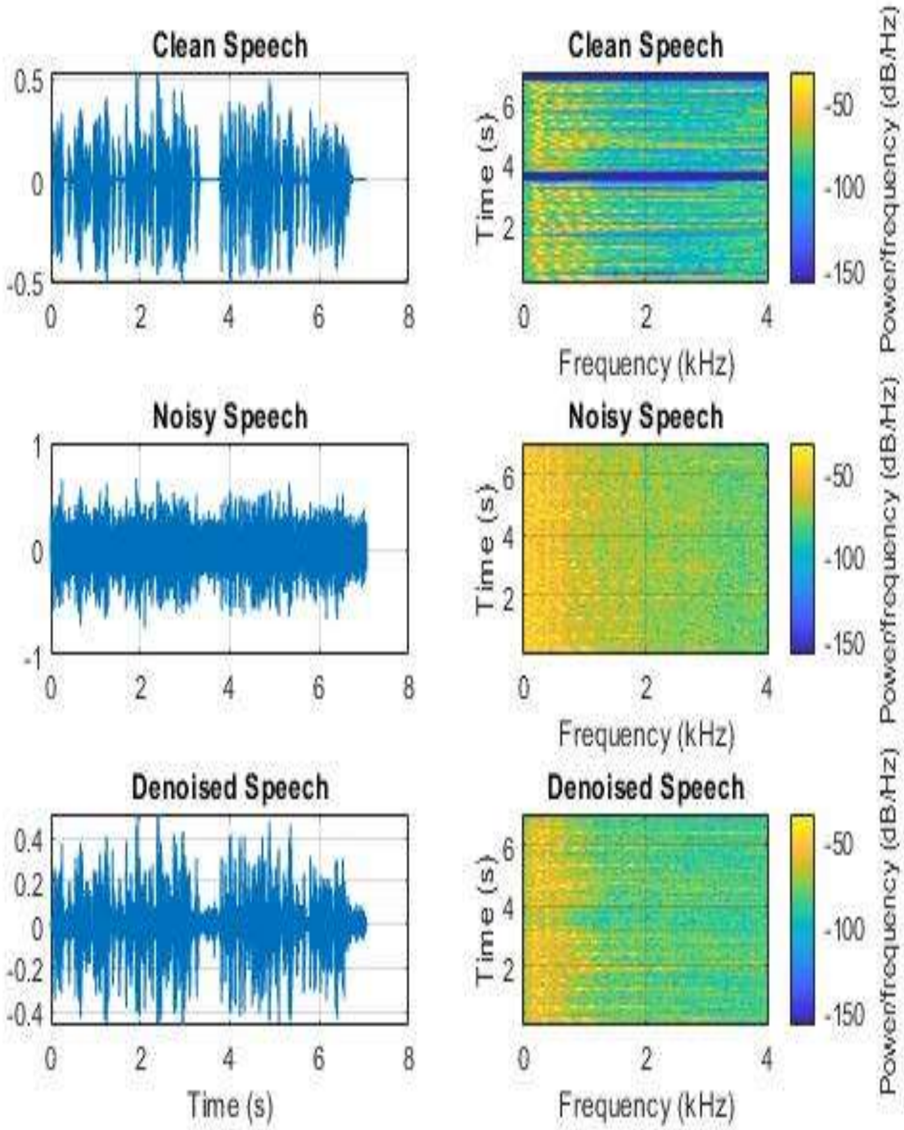

Fig -7 Results of deep learning network

From the figure it's clear that the deep learning network is capable in removing the reference noise from the clean speech. Furthermore it's evident that the filtered output comprises of less noise as compared with the noisy speech. A typical analysis of the SNR of the input and outputs of the CNN layer shows that, the SNR of the clean audio is approximately 14.05dB and that of the signal after the filtering process is approximately 14.0dB. Thus we can say that the input and the output signal is almost same and only a factor of 0.05Db of interference is present in the filtered output.

Figure 8 and Figure 9 shows the respective frequency response plot of the input and the filtered signal.

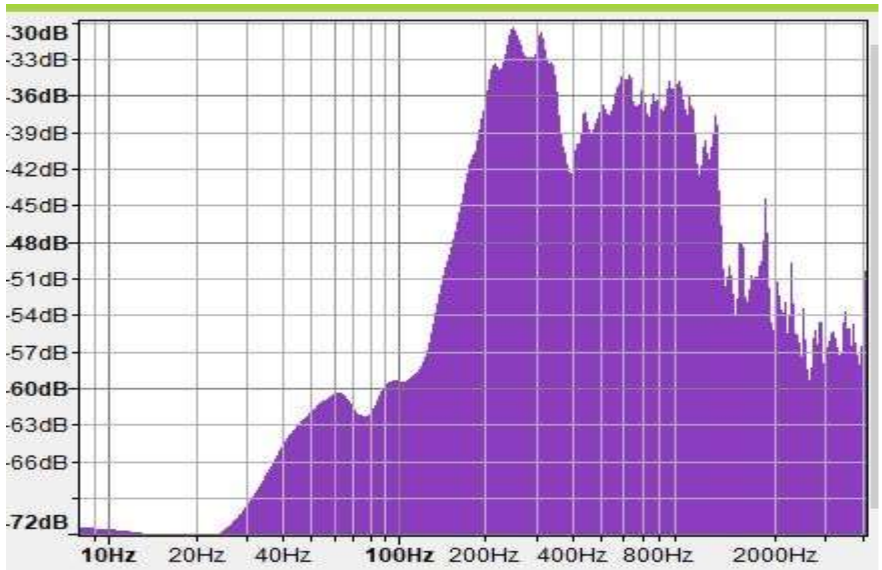

Fig -8 Response of clean audio signal

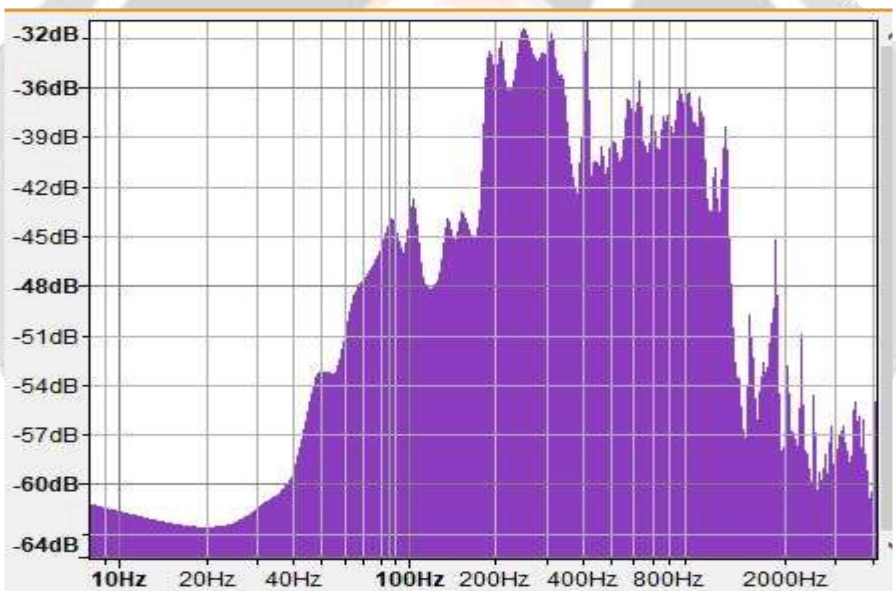

Fig -9 Response of filtered audio signal

From the frequency response plot of the two signals it is clear that both the signals contain less amount of noise. The RMS values of both the signals can be analyzed from the response graph and it is observed that the RMS value of the input signal is approximately-19.34dB and that of the filtered signal is approximately –19.20dB. thus it's clear that the amount of noise in the filtered signal is less as compared to that of LMS Filter.

5. CONCLUSIONS

This paper proposes a comparison of the existing LMS Filter and the CNN networks and it has been analyzed based on the results and performance parameters that deep learning network like CNN is capable in eliminating the noise to a large extent. It's also observed that even with the same reference noise the LMS Filter is not capable in eliminating the noise present in the input signal. Therefore it can be concluded that Deep Learning Network (CNN) is best suited for noise cancellation in audio signals.

6. REFERENCES

[1] P. Mahanty, B. Firdous and R. Swarnkar, "Real Time Background Noise Cancellation in End User Device," 2007 IEEE International Conference on Portable Information Devices, Orlando, FL, 2007, pp. 1-4.

[2] Anugerah Firdauzi, Kiki Wirianto, Muhammad Arijal, Trio Adiono Design and Implementation of Real Time Noise Cancellation System based on Spectral Subtraction Method

[3] A. L. L. Ramos, S. Holm, S. Gudvangen and R. Otterlei, "A modified spectral subtraction algorithm for real-time noise reduction applied to gunshot acoustics," 2012 International Conference on Signals and Electronic Systems (ICSES), Wroclaw, 2012, pp. 1-5.

[4] L. Bai and Q. Yin, "A modified NLMS algorithm for adaptive noise cancellation," 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, 2010, pp. 3726-3729.

[5] Say-Wei Foo, T. N. Senthilkumar and C. Averty, "Active noise cancellation headset," 2005 IEEE International Symposium on Circuits and Systems, Kobe, 2005, pp. 268-271 Vol. 1.

[6] S. M. Kuo, Y. C. Huang and Zhibing Pan, "Acoustic noise and echo cancellation microphone system for videoconferencing," in IEEE Transactions on Consumer Electronics, vol. 41, no. 4, pp. 1150-1158, Nov. 1995.

[7] Narula, Vaibhav & Sagar, Mukul & Joshi, Pranab & Mehta, Puneet & Tripathi, Sudhanshu. (2011). Real-Time Active Noise Cancellation with Simulink and Data Acquisition Toolbox.

[8] P. Louridas and C. Ebert, "Machine Learning," in IEEE Software, vol. 33, no. 5, pp. 110-115, Sept.-Oct. 2016.

[9] X. Du, Y. Cai, S. Wang and L. Zhang, "Overview of deep learning," 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, 2016, pp. 159-164.

[10] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), Antalya, 2017, pp. 1-6.

[11] Y. Tsa, H. Chu, S. Fang, J. Lee and C. Lin, "Adaptive Noise Cancellation Using Deep Cerebellar Model Articulation Controller," in IEEE Access, vol. 6, pp. 37395-37402, 2018.

[12] J. Valin, "A Hybrid DSP/Deep Learning Approach to Real-Time Full-Band Speech Enhancement," 2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP), Vancouver, BC, 2018, pp. 1-5.

[13] Y. Shi, W. Rong and N. Zheng, "Speech Enhancement using Convolutional Neural Network with Skip Connections," 2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP), Taipei City, Taiwan, 2018, pp. 6-10.

[14] G. S. Bhat, N. Shankar, C. K. A. Reddy and I. M. S. Panahi, "A Real-Time Convolutional Neural Network Based Speech Enhancement for Hearing Impaired Listeners Using Smartphone," in IEEE Access, vol. 7, pp. 78421-78433, 2019.

[15] S. Angra and S. Ahuja, "Machine learning and its applications: A review," 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), Chirala, 2017, pp. 57-60.

[16] "IJCNN International Joint Conference on Neural Networks (Cat. No.90CH2879-5)," 1990 IJCNN International Joint Conference on Neural Networks, San Diego, CA, USA, 1990, pp. .