Volume: 15 Issue: 05 Pages: 6146 – 6150 (2023) ISSN: 0975-0290

Real-Time Noise Suppression Method Using Deep Learning Algorithm

Nayana Hegde
School of Electronics and Communication Engineering, REVA University, Bangalore-64
Email: nayana.srikanth@gmail.com
Ashwini P
School of Electronics and Communication Engineering, REVA University, Bangalore-64
Email: ashwinip@reva.edu.in

---ABSTRACT---
The capacity of deep learning-based speech enhancement algorithms to effectively eliminate both stationary and non-stationary noise components from noisy speech observations has been demonstrated. However, they frequently add false residual noise, particularly when the training goal lacks phase information, such as an ideal ratio mask or the magnitude and fluctuations of clean speech. It is widely known that the perception speech quality may deteriorate whenever the power of the residual noise components surpasses the noise masking threshold of the human auditory system. One logical approach is to use a post processing strategy to further attenuate the remaining noise components. However, estimating the noise power spectral density (PSD) is a difficult challenge due to the kind of residual noise's very non-stationary character. In order to address this issue, the research suggests three methods for estimating the noise PSD frame by frame. The remaining noise can then be efficiently eliminated by using a gain function developed using a decision-directed methodology.

Keywords – **Noise cancellation, Deep learning, Suppression, Algorithms**

Date of Submission: January 20, 2024 Date of Acceptance: February 27,2024

I. INTRODUCTION

The capacity to improve a noisy voice signal by isolating the main sound is known as background noise removal. Everywhere background noise elimination is utilized, including in video conferencing systems, audio/video editing software, and noise-canceling headphones. Background noise reduction is still a rapidly developing field of technology, and artificial intelligence has opened up a whole new range of methods for doing it better. An old concept, artificial neural networks, has just come to life as deep learning. Although there are various deep learning methods for removing noise, they all operate by learning from a training dataset.

Phones, laptops, conferencing systems, and other edge devices have all successfully used conventional noise suppression. Given that the edge device is the one that initially records the user's voice, this strategy appears logical. Once captured, the device filters the noise out and provides the result to the opposite end of the call. Consider stationary noise as having a pattern that is distinct from human speech but repeatable. When filtering such noises, conventional DSP methods (adaptive filters) can be fairly effective. Non-stationary noises feature complex patterns that make it challenging to tell them apart from human speech. The signal could be brief and flash quickly.

A. Traditional Approach to Noise Removal

The majority of noise-removal algorithms are subtractive, which means they find the frequencies with the most background noise and take those bands out of the original signal. In order to extract what is thought to be the dominant signal, several of these systems employ static filters with predetermined settings, such as low pass, high pass, and band pass filters.

B. Signal model

A signal is a numerical value or attribute that transmits data. The human brain is thought to send signals to various sections of the body on a regular basis, causing it to perform particular acts or react meticulously. As a result, the brain signal is referred to as an excitation, and the action taken by the specified organ is referred to as a reaction. There are many examples of signals in the physical world. Wireless transmission of television and radio, radio-astronomical signals, signals used for information transfer by satellites circling the earth, and sonar signals for underwater acoustics are all examples of man-made electrical signals. Signal transmission between two points can also be accomplished using cables or optical fibers, as with telephone lines. Signal transmission between distant sites can now be accomplished utilizing a combination of cable and wireless technologies.

There are mainly two categories of signals, analog signal and digital signal. The usual environment we live in generates analogue signals. The signal f(t) is specified for all values inside the continuous variable t, and its amplitude can be any value in a continuous range. An analogue signal is one that has no digital components. When describing analogue signals, the terms continuous-time or simply continuous might be used. Sound, temperature, and pressure, as well as voltage, current, and light, are examples of analogue signals that are used in everyday life. Other signals, which are largely created by humans, are defined solely for discrete values of time t, in contrast to analogue signals. Because the independent variable t takes discrete values that are often integral multiples of the same basic number T, it can be written as follows:

$f(nT) = f(t)$ n=0, 1, 2,… (1)

A discrete-time signal, or simply discrete signal, is one such signal.

A signal can be classified as either random or predictable. Random signals are those that can be expressed mathematically (as a function) or graphically (as an arbitrary graph). Random signals, on the other hand, are not explained by such approaches. As an alternative, statistical approaches can be used to describe them. Unwanted additive noise is responsible for a signal's randomness. When the signal is transmitted across noisy channels, noise may be introduced into the system. At the moment, a great deal of effort in the areas of signal transmission and detection is devoted to the problem of extracting the information contained in a signal contaminated by random noise. Both analogue and digital signals are covered by this idea.

Signals of various forms are transmitted from one location to another in order to transmit data. Over the years, a large number of strategies have been developed in order to perfect the method of revision, transmission, and detection of the essential information. Signal manipulation systems are required for these strategies. Amplification, differentiation, integration, addition, and multiplication as a more sophisticated operation are all examples of such manipulations. Furthermore, if the signal is analogue, it must be converted to digital before it can be processed by a digital processor.

Interference noise in speech signals is the most common cause of speech processing problems. Interference noise alters the spoken signal in a certain way, reducing its clarity. Interference noise can come from a variety of auditory sources, including ventilation equipment, echoes, crowds, and, more broadly, any signal that interferes with speech signals. The signal-to-noise ratio, which is measured in decibels, is a link between the strength of the voice signal and the modulation caused by noise. The signal-to-noise ratio should be greater than 0dB in an ideal scenario. This means the speech is louder than the background noise. High-frequency noise wraps only the constants, and as the noise's amplitude grows, its efficiency declines. When the noise signal is louder than the speech signal, however, low frequency noise is a considerably more effective modulation. At high pressure levels, the noise can engulf vowels and consonants.

The paper is structured as follows: Section II provides an overview of significant research conducted in the field of noise cancellation, while Section III outlines the existing system. Section IV presents detailed information about the proposed system, followed by the presentation of results in Section V. Finally, Section VI concludes the paper.

II. RELATED WORK

In this section we present some of the related works.

To provide better noise suppression and faster convergence, an enhanced least mean square method with configurable step length was employed for adaptive noise cancellation in [1]. Authors in [2] solely explore the effects of changing noise on LMS calculations. The goal of the work explained in [3] is to find a way to increase the noise canceller's performance in terms of filter settings. An adaptive noise canceller utilizing the LMS (least mean square) algorithm can recover the desired signal that has been contaminated by additive noise in this fashion. This adaptive noise canceller can help you increase your signal-to-noise ratio [4]. Various filter applications are discussed in work [5]. The Particle Swarm Optimization (PSO) and Ant Colony Optimization algorithms are used to optimize the Least Mean Square (LMS) algorithm (ACO). The advantages and disadvantages of merging the gradient based (LMS) method with Swarm Intelligence SI have been investigated (ACO, PSO). This improvement to the LMS algorithm will allow us to extend the applications of adaptive filtering to systems with multi-model error surfaces, which is still a grey area in adaptive filtering [6].

In work [7], authors use a fixed forward error correction (FEC) scheme to investigate the rate adaptability of quadrature amplitude modulation (QAM)-based probabilistic constellation shaping (PCS) over a wide range of information rates (IRs). It was decided to use blind adaptive equalization, which does not forfeit any of the IRs. In the framework of a state variable model [8], the authors create an efficient LMS method. In order to minimize system misalignment, the suggested solution follows an optimization problem and includes a variable step-size. Work in [9] provides an overview of optical fiber communications as well as the various sorts of catastrophes and mitigation strategies for dealing with them. The failure of a network caused by disasters leads to the collapse of multiple optical communication channels and massive data loss in an optical communication network.

The utilization of multiple deep neural networks (DNNs) in a regression-based speech enhancement model yielded favorable results, demonstrating substantial improvement across various objective metrics. Additionally, in a subjective evaluation, 76.35% of participants expressed a preference for this model over other conventional techniques [13]. The efficacy of the Redundant Convolutional Encoder-Decoder (RCED) model was showcased by revealing that employing a CNN with minimal parameters can yield superior outcomes, particularly in the context of embedded systems such as hearing aids [14]. The hybrid model successfully validates a noise suppression approach by synergizing DSP-based techniques with deep learning. Recognizing the challenge of fine-tuning noise suppression through deep learning alone, the integration of DSP-based methods enhances the model's efficacy. The proponents believe in the adaptability of this hybrid approach, foreseeing its seamless extension to residual echo suppression. Furthermore, its applicability extends to post-filtering purposes in microphones [15]. This study introduces a real-time speech enhancement method utilizing a streamlined recurrent neural network. The model showcases its ability to regulate speech distortion by incorporating fixed-weighted and SNR-weighted coefficients into the loss function. The research delves into different aspects of training an RNN model, with a particular emphasis on enhancing short-time speech through a single-frame-in, single-frame-out approach—a framework commonly employed in classical signal processing methods [16].

III. EXISTING SYSTEM

As a noise canceller, adaptive filters are commonly utilized. The Adaptive Noise Canceller (ANC) is a device that proves noise-polluted transmissions. It has an advantage over other signal processing approaches in that it does not require any prior signal or noise knowledge. Its cost is inextricably linked to the fact that it requires two input signals: the primary input signal plus noise, as well as a reference noise signal.

Adaptive filtering is a signal processing procedure in which the parameters used to process signals alter based on a set of criteria. The estimated mean squared error is commonly used as criteria. Because the adaptive filters' settings are always changing in order to meet a performance criterion, they are time-varying. An adaptive filter, in this sense, can be thought of as a filter that executes the approximation step in real time. The existence of a reference signal is frequently required for the determination of the performance requirement, which is usually hidden in the approximation step of fixed-filter design. The main inconvenience of the steepest-descent (MMS) gradient algorithm consists in the detail that exact measurements of the gradient vector are required at each step in the iteration process. This is not practical and one needs an algorithm for deriving estimates of the gradient vector only from the available data. This is achieved by using the least-mean-square (LMS) error gradient algorithm. Its advantages are the following: It is done straightforward, does not require matrix inversion, and it does not require correlation measurements. The MMS error gradient algorithm uses expectation values for calculating the gradient vector. The expression one uses is the following

$$\nabla(n) = -2E\{e(n)[x(n)]\}$$

Instead of using expectation values, the LMS algorithm uses instantaneous estimates of the vector based on sample values of the input [x(n)] and the error e(n). Therefore the following expression is used for calculating the instantaneous estimate of the gradient vector

$$\nabla(n) = -2e(n)[x(n)]$$

Such an estimate is clearly neutral since its expected value is the same as the value in the expression from the above given equation. Along the LMS algorithm, the filter coefficients are updated along the direction of the gradient vector estimate according to following expression

$$h(n+1) = h(n) + \frac{1}{2}\mu - \nabla(n) = \mu e(n)[x(n)]$$

This is much simpler than the expression used in the MMS algorithm for calculating the filter coefficients:

$$h(n+1) = h(n) + \mu E\{e(n)[x(n)] = h(n) + Rx(n)$$

The gradient vector estimate of the LMS algorithm only requires knowledge of the data[x(n)] and no cross-correlation estimate. The error signal is defined by the following expression:

$$e(n) = f(n) - x(n)h(n)$$

The adaptive filtering process ends up working very similar to the steepest-descent algorithm. One has to initialize the coefficient values as [h(0)] at n=0, therefore a value of zero is given as an initial guess, but as the algorithm is started the coefficient values of the filter will begin to converge to the correct value. At a time n, the coefficients are updated in the following way:

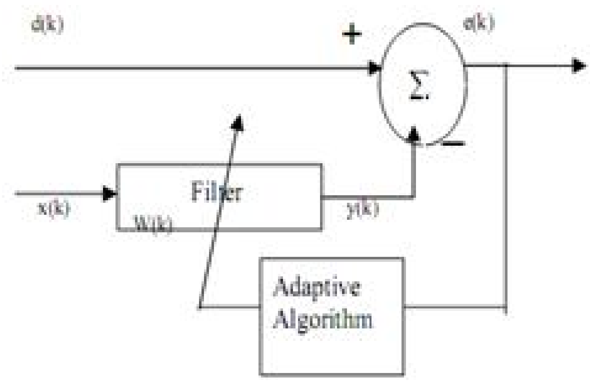

Fig. 1. General Setup of Adaptive filter

The general set up of adaptive filtering environment is shown in Fig. 1, where k is the iteration number, x(k) denotes the input signal, y(k) is the adaptive filter output, and d(k) defines the desired signal. The error signal e(k) is calculated as d(k)-y(k). The error is then used to form a performance function or objective function that is required by the adaptation algorithm in order to determine the appropriate updating of the filter coefficients. The minimization of the objective function implies that the adaptive filter output signal is matching the desired signal in some sense.

IV. PROPOSED SYSTEM

An old concept, artificial neural networks, has just come to life as deep learning. Although there are various deep learning methods for removing noise, they all operate by learning from a training dataset.

Step1: Pre-processing

This session focuses on separating the important information from the cluttered audio for use in training.

Step2: Training Sets for Noise Reduction

Creating a good training dataset is the first step in creating a precise noise removal model. Our dataset should include recordings of both clean speech and its noisy counterpart because our objective is to reduce background noise.

Step3: Recurrent Neural Networks (RNN)

Recurrent neural networks are models with sequential data recognition and comprehension capabilities. Things like audio, text, or the position of an object across time are examples of sequential data. RNNs are particularly good at eliminating background noise because they can recognize patterns over long periods of time, which is necessary for interpreting audio.

Step 4: Working of RNN

An RNN that has been trained to remove background noise from a noisy audio sample. The audio sample can be divided into a series of equally spaced time segments. The hidden state is updated throughout each iteration, keeping track of the prior steps each time, as each individual sample of the sequence is fed into the RNN. After each cycle, the output is routed via a feed-forward neural network to create a new audio stream that is completely free of background noise. Figure 2 illustrates the proposed system.

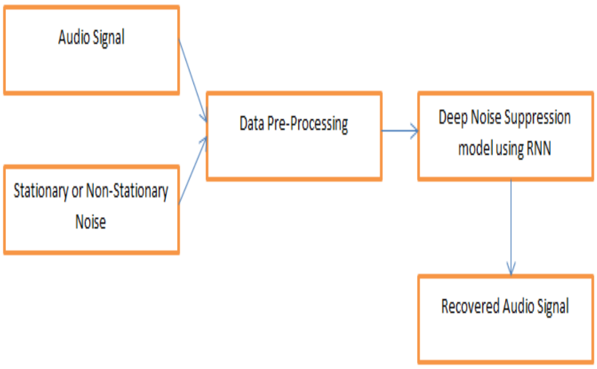

Fig. 2 Block diagram of the proposed system

RNNs excel in audio noise cancellation by leveraging their sequential processing ability. RNNs, with their recurrent connections, capture temporal dependencies within audio signals. In the context of noise cancellation, RNNs learn to distinguish between relevant audio features and undesirable noise patterns over time. This enables them to adaptively filter out noise, enhancing the clarity of the desired audio signal. By processing sequential information, RNNs can effectively model and suppress background noise, making them valuable tools for improving audio quality in various applications, such as speech recognition, music processing, and telecommunications.

V. RESULTS

The simulation of the work is done in system with following details. We employed a computer running a Windows operating system, equipped with an Intel Core i5-4200 processor operating at a frequency of 2.5 GHz for the simulations. The programming is done using Matlab programming tool.

Original signal and noisy reference signal are fed into adaptive filter. The output obtained in the filter are shown in the figure 3 and figure 4. The results illustrated in figure 3 highlight that time varying step size can improve the performance of the algorithm. The choice of step size in LMS algorithm defines the behavior of the output's convergence to the desired signal. Step size is chosen in adhoc way. A large step size will move the system close to the desired signal faster but will have the problems of overshooting the optimal solution when the error becomes small. Conversely a small step size will take a longer time to approach zero error but it will better behavior around the optimal solution. The signal tested is an audio signal. The format used is .wav format. The noise signal is used as reference signal with uniform distribution. The output is verified at different step size and different number of taps. The figure 4 demonstrates with step size 0.0001 and number of taps=5. Here the noise is not reduced to minimum. The fig 4 shows error is reduced to minimum with step size=0.0001 and number of taps =20.By adjusting both the parameters, the output is efficient where error is almost zero.

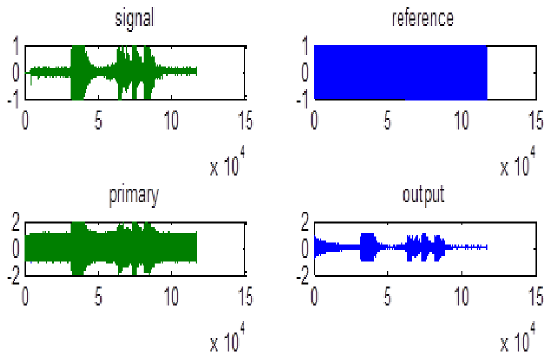

Fig. 3 Results of Noise suppression process

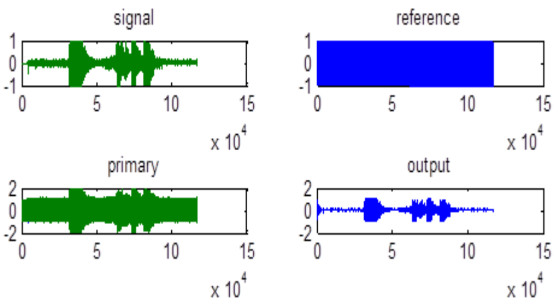

Fig. 4 Results of Noise suppression process

VI. CONCLUSION

Since the advent of the microphone, background noise reduction has garnered the majority of attention in the field of audio processing. There are countless conventional ways to filter audio, but the most, if not all, perform badly with non-static audio and add distortion when the main speaker and the background noise merge. We've been able to train computers to excel at particular jobs thanks to the increase in processing power and our capacity to create deep learning models that can retain complicated patterns over extended periods of time. Computers have gotten incredibly good at eliminating audio noise by training deep learning models with lots of data. The AI approach is far superior to conventional approaches if compute resources and latency are taken into account. This is due to the fact that they are generative, as opposed to subtractive, traditional models. The background noise can be eliminated, and the clear speech can be produced with little to no distortion, using AI methods.

ACKNOWLEDGEMENTS

We express our gratitude to the School of Electronics and Communication Engineering (ECE) at REVA University for conducting this research.

REFERENCES

[1] Ke, Y., Li, A., Zheng, C. et al. Low-complexity artificial noise suppression methods for deep learning-based speech enhancement algorithms. J AUDIO SPEECH MUSIC PROC. 2021, 17 (2021). https://doi.org/10.1186/s13636-021-00204-9

[2] Sai Saranya Thunga & Rajesh Kumar MuthuAdaptive Noise Cancellation Using Improved LMS Algorithm, 2019

[3] Aryan Saini; Sparsh Rana, "Improved LMS Adaptive Noise Cancellation to Enhance SNR", International Conference on Communication and Signal Processing, 2020

[4] Deepanjali Jain, Dr. Poonam Beniwal,"Application of Improved LMS Algorithm for Adaptive Noise Cancellation",Ijraset Journal For Research in Applied Science and Engineering Technology, 2022

[5] Durba Satpathy, Prabodh Kumar Sahoo,"Design of Adaptive Noise Canceller using LMS Algorithm",2015 VOLUME 3 - ISSUE 25

[6] J. Gerardo Avalos, Juan C. Sanchez and Jose Velazquez, "Applications of Adaptive Filtering", Adaptive Filtering Applications,2011

[7] Qianhua Ling, Mohammad Asif Ikbal and P. Kumar, "Optimized LMS algorithm for system identification and noise cancellation", De Gruyter February 23, 2021

[8] Manabu Arikawa, Masaki Sato, and Kazunori Hayashi, "Wide range rate adaptation of QAM-based probabilistic constellation shaping using a fixed FEC with blind adaptive equalization", Optics Express, Vol. 28, Issue 2, pp. 1300-1315 (2020)

[9] Alexandru-George Rusu; Silviu Ciochină; Constantin Paleologu, "On the Step-Size optimization of the LMS Algorithm",International Conference on Telecommunications and Signal Processing (TSP),2019

[10] Dinesh Kumar, Ashutosh Sharma, Rajiv Kumar and Neeru Sharma,"A Holistic Survey on Disaster and Disruption in Optical Communication Network",Volume 13, Issue 2, 2020

[11] Xu Yanhong, Zhang Ze, "Design the adaptive noise canceller based on an improved LMS algorithm and realize it by DSP", 2012 Fifth International Conference on Intelligent Computation Technology and Automation

[12] P. Kanagasapabathy, V. R. Vijaykumar, P. T. Vanathi, "Modified Adaptive Filtering Algorithm for Noise Cancellation in Speech Signals ", Electronics and Electrical Engineering ISSN 1392 – 1215 2007

[13] Li-Rong Dai Yong Xu, Jun Du and Chin-Hui Lee, "An experimental study on speech enhancement based on deep neural networks". 2014

[14] Se Rim Park and Jin Won Lee, "A fully convolutional neural network for speech enhancement". 2016.

[15] Jean-Marc Valin, " A hybrid dsp/deep learning approach to real-time full-band speech enhancement", 2018

[16] Chandan K. A. Reddy Harishchandra Dubey Ross Cutler Ivan Tashev Yangyang Xia, Sebastian Braun. "Weighted speech distortion losses for neural-network based real-time speech enhancement", 2020.

Biographies and Photographs

Dr. Nayana Hegde did her Bachelors in Electronics and Communication from Karnataka University, India in year 2001 and Masters in Digital Communication and Networking from VTU, India in 2014. She did her PhD in Electronics and Communication and Engineering from REVA University Bangalore, India. She has published many research articles in reputed journals like IEEE, IGI Global, Taylor and Francis and Inderscience. Currently she is working as Assistant Professor in REVA University.

Ashwini P.., is pursuing his Ph.D. in the field of Speech Processing in the School of ECE, REVA University, Bengaluru, India. She has obtained in B.E. in ECE from VTU, Belgaum, and M.Tech in Digital Communication from VTU, Belgaum, India. Her research areas of interest are Signal Processing, Image Processing, Video Processing, IoT, AI/ML and Communication Systems. She has published research papers extensively in highly reputed international journals and international conferences, including IEEE.