International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 07 | July 2022 www.irjet.net p-ISSN: 2395-0072

A review of Noise Suppression Technology for Real-Time Speech Enhancement

Keshav Patta[1], Hitesh Tiwari[2], Mohit Kumar Tiwari[3], Prof. Vaishali Gatty[4]

[1,2,3] PG Student, Department of M.C.A. VESIT, Mumbai, Maharashtra, India.
[4]Professor Vaishali Gatty, Department of M.C.A, VESIT, Mumbai, Maharashtra, India.

Abstract - *Despite noise suppression being a mature space in the signal process, it remains hugely captivated by the fine calibration of reckoner algorithms and parameters. In this paper, we demonstrate a Real-Time learning approach to noise suppression. We tend to focus powerfully on keeping the quality as low as potential, while still achieving high-quality increased speech. A huge number of communication programs or systems have introduced next-generation noise-canceling AI as an alternative to reduce background sounds/noise in their online meetings or work and made this process well organized. Yet, noise suppression is far more prominent than active noise canceling systems out there in existing systems and provides a better result. Currently leading tech companies to choose noise suppression for communication applications to offer better result, even though these system does not yield 100% accuracy it is still far superior than the other conventional systems.*

Key Words: Noise Detection, Noise Suppression, Deep Learning, Voice Detector, Microphone,

1. INTRODUCTION

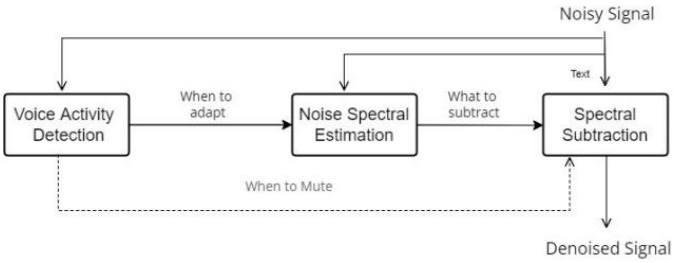
For decades we hear that AI is the future and with its help, Noise suppression has gained much interest in the field. Despite important enhancements in quality, the high-level structure has remained principally equivalent. The noise spectrum estimation technique is the backbone of noise suppression AI, it is derived by a voice activity detector (VAD). It has three component which needs correct estimators and is tough to tune. for instance, the crude initial noise estimators and also spectral estimators that supported spectral subtraction are replaced by a lot of correct noise estimators and spectral amplitude estimators. Despite the enhancements, these estimators have remained tough to style and have needed important manual calibration effort. that's why recent advances in deep learning techniques are appealing for noise suppression.

Active noise-canceling technology has been the focus of the world for many years because of its easy implementation and the technology sector see it as a compelling option, but since it has drawback on hazardous level (security of user) noise suppression technology has been seen as an only solution for that problem. Noise suppression in devices is achieved with twin microphones or quad (Omni-directional) microphones, that square measure accustomed pull in audio signals (noise, speech, alarms, etc.) from the encompassing setting. These audio signals square measure sent to a digital signal processor (DSP) with algorithms to assist separate and suppressing background signals. Also, with advanced algorithms, this could embody uninflected and enhancing speech thus you'll hear and be detected clearly in spite of the noise around you. Noise suppression AI also achieved prominent results through deep learning which can detect human voice between different noises given as an input, these results show the advancement in Artificial Intelligence in today's world.

2. Voice Activity Detector (VAD)

Voice activity detection (VAD) can be defined as a technique during which the presence or absence of a human voice is detected. The detection is accustomed to triggering a method. VAD has been applied in speech-controlled applications and devices like smartphones, which can be operated by voice commands. Most common VAD algorithms are based on using amplitude and are very efficient namely Short-Time average Energy (STE) and Zero-crossing Rate (ZCR) but these techniques do not work in loud or noisy environments. Using spectral energy for this process is the best approach since it works with high accuracy in high speech noisy environments, some of the spectral energy techniques are long-term spectral envelope, Mel-frequency cepstral coefficients (MFCCs), linear predictive coding coefficients (LPCC).

The detection process has been carried out by the technique chosen by the engineer, in the spectral energy base technique when input comes with speech + noise, it sets the idle frequency range in which speech can be recognized and set the condition to remove everything else when spectral energy is less than zero. In the next part, sum up all the parts and set a threshold for energy to recognize active and non-active parts in the detection, and create a length filter for removing non-active segments, and in the last part, it creates a buffer on both sides (e.g. 0.5 seconds) to complete the process.



High-level Structure Of Most Noise Suppression Algorithm

3. Noise Estimation Algorithm

The need for noise Estimation is to calculate the noise in the speech and subtract the estimated value to get the desired speech. With this basic idea noise estimation algorithm has been created, the working of the algorithm is very simple in the initial part it divides the speech signal into small parts and starts at the beginning. The first part of the signal starts being estimated for noise and compared with the next part and so on when the noise signal gets estimated in a few parts it starts subtracting that signal from the rest of the speech signal to get the desired output.

There are three types of noise estimation algorithms, first is a histogram-based algorithm in which the speech signal is divided into bins, and each bin contains a speech signal with noise. The spectral energy in each bin has been calculated as the density of the noise and once it's been calculated the algorithm subtracts it from the speech signal and a clean signal can be achieved. The second algorithm is the Minimal tracking Algorithm in this algorithm we track the minimum band frequency of the noise in the speech spectrum and calculate the spectral energy similar to the previous algorithm, this algorithm has two types 1 - Minimum statistics (MS) Noise Estimation and 2 - Continuous Spectral Minimum. Time Recursive Algorithm is the third algorithm that uses the knowledge that speech signal has a non-constant effect on the spectrum signal in some parts, those parts then been calculated for signal-noise ratio (SNR) and each part will give different SNRs to calculate with which will give a high value and a low value. The noise considers being absent at high and low values by which we can get the average spectrum signal to subtract from the speech signal,

Time recursive algorithm is further derived in minima controlled recursive averaging (MCRA) algorithm.

4. Conventional use of Noise Suppression Technology

Current generation devices like computers, tablets, smartphones, and many more are the prime example of the Conventional use of Noise Suppression Technology. Conventionally devise takes the noise as input and processes it to produce clear sound for the end-user.

For past decades devices have been improving vigorously and competing with each other to improve more and more, old devices like mobile phone have bad reception and user have to suffer from background noises a lot. Later smartphones introduce Active Noise-canceling technology which can cancel the background noises and provide better reception still the results were not promising since all the devices have to rely on the hardware of that device.

Smartphones are equipped with multiple microphones to process the incoming sound at different positions (e.g., the first mic at bottom of the device, the second at top of the device) to capture user sound at a different angle, after processing input data (sound) from all the mics background noise can be detected by matching the data and removing it in the process and in the end, the output will be a clean audible sound.

All this process is done by a combination of simple hardware and software but it has some major flaws, imagine if the device is only receiving data of background sound or the device has been put far apart then it won't get the desired input to the process. This type of flaw compels the tech industry to improve its hardware and software even more but doing so also increases the cost of devices and also has to upgrade them with the latest technology.

Conventional smartphone flaws in using the mic for ANC is a huge problem, especially in the condition where the user is doing exercise away from the device or walking in the market, but to some extent, these problems are solved by new devices like a smartwatch, Bluetooth earphone, etc. but still, the accuracy is not up to the mark and in some extreme condition noise suppression simple fail to achieve its desire output.

5. Background Noise Separation using Deep Learning

Using deep learning to separate noise from the sound is most effective and revolutionized way. A study in 2015 says that using a regression method we can create a model which learns to produce a signal-noise ratio (SNR) for every single spectrum and the produced signal will only give a human voice leaving extra noise. But still, it was not perfect since it was the initial model.

With time many more models have been introduced but the conventional approach is a three-part process that consists of Data gathering, training the data, and Deduction. In data gathering, users gather a huge dataset of loud noises, with different variations of human-animal background noise, etc. training the data is the process in which we train the model according to the dataset and achieve the highest accuracy possible. Lastly, come deduction where final trained weights are created to detect human voice with high accuracy and mask out all the noises.

By working on different models and creating a distinctive model we have reached a time where we can get exceptional results with the current-generation model and also with high accuracy of 99.0%. with such a model, next-generation devices can suppress sound using a single microphone design.

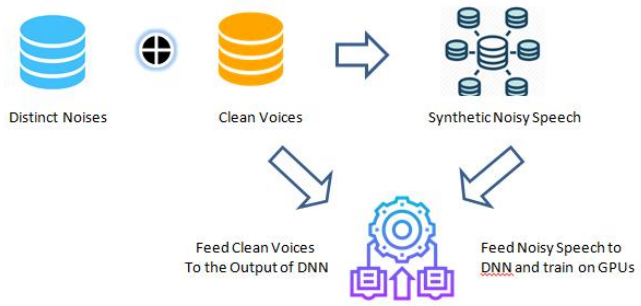


Fig- Prosess of Deep Learning Algorithm.

6. Single Microphone Design over Multi Microphone Design

Multi-microphone styles have a couple of vital shortcomings.

1 – This design is more centric on smartphones which have a conventional design to be used in close proximity to the user (near the mouth)

2 – This conventional design put the tech industry at a disadvantage by creating an expensive design with costly components and software.

3 – The sound input can only be processed at the device end so the software cannot be too precise or the result cannot be too accurate to achieve cheap cost

If all the process can be done on the software side of a device with the help of a single microphone in a smartphone it will put the hardware cost as cheaper side and also will be cost-effective. Currently extracting a person's voice from a speech signal is a very difficult task even for an algorithm, no perfectly accurate model is available yet.

The conventional model of digital signal tries to learn continuously from the human voice by processing it bit by bit, these models work well in some cases but fail to achieve high accuracy in real-world trials.

Nonstationary sound can be extremely tough to recognize when compared to a person's speech since the signal is very unpredictable and very thin (e.g. whistle or short ring) whereas, Stationary sound has the same regressive characteristic as a person's voice with the difference in the signal, conventional algorithms can be highly successful while working on such sound. To achieve high accuracy regardless of the type of sound we need to improve the conventional algorithm models and it's not too far when using deep learning can achieve it.

7. Leading names in Noise Suppression Technology

Microsoft Teams

Noise Suppression Technology has come a long way since it has been introduced especially after the effects of covid-19.

Microsoft has introduced a superior suppression technology for its meeting application (Teams) using artificial intelligence to recognize and separate unwanted noise from calls.

Achieving it was not a simple task Microsoft have to collect over a thousand hours of sound data and categorized it into human voice or just a noise. Microsoft has collected sound data over dozens of languages and hundreds of other noises to achieve high accuracy. They trained their artificial intelligence model with this data to recognize the person's voice communication over the call.

Nvidia

A new way of the noise suppression model is being developed by Nvidia using DL. Nvidia has created RTX-voice which is very identical to its competitor Krisp.ai, which also used a virtual AI base system to suppress noises. RTX-voice currently runs on the latest windows only (10 - 11) and needs a high-end GPU to process.

This extraordinary growth in this industry is a result of the high need for clear communication over distance meetings on online calls.

Krisp.ai

When main competitors were creating a single platform-centric AI-based model, an AI-based company created a platform-independent AI solution to work with many connective applications used for meeting out in the market.

Krisp.ai give its solution to a single user or an organization on call platform from each end (outgoing-incoming).

The most fascinating feature of krisp.ai is that it not only blocks the outgoing extra sound from the surroundings but also blocks the incoming surrounding sound and only gives a filtered clean voice even when the opposite user has not krisp.ai. it has trained to suppress unwanted sounds immediately on the call while it's passing

8. Moving to the Cloud

Considering all the model is completely software-centric, the question arises whether they can be transferred over a cloud, it can be simply answered as 'yes'. N number of devices support the cloud and hence the model can work on

them, it is also very convenient to use on both incoming and outgoing calls over a connectivity application. Today it is possible to migrate over the cloud but it was still a dream decade ago due to the heavy hardware requirements of the suppression system. An original equipment manufacturer must meet the standard set by mobile companies to achieve high-quality communication, sadly even today multi-microphone design is the only solution we get. Still, we can get a high level of noise suppression thanks to DL in the cloud which can support a singular microphone design

9. CONCLUSIONS

In this paper we have tried to demonstrate how the noise suppression technology has been replacing Conventional Active Noise Canceling Technology and get the summarized knowledge of voice activity detection which help to detect voice in the transmission we also learned about different noise estimation algorithm which is important to estimate the noise present in a signal. Learning about the leading company and technology was the important part of this study and we learn how conventional method are being used but still yield mediocre result. Next generation need to be shifted over single microphone design from multi microphone design if industry need to be innovative as well as cost effective, it can be achieved with the help of cloud technology and we are sure noise suppression technology will yield high accuracy then any current available technologies.

A superior noise suppression technology can help user experience grow like never before, it can help improve customer service (prevent background noise disturbance on important calls), it can be used in restaurant's for new gen ordering system over voice commands (such technology is being used in today's date with some flows), personal smartphone assistant is common nowadays just imagine it with more higher accuracy to command such system with any hassle of repeating our self over and over again. The more dynamic these model gets the more possibility can be open with such system, deep learning has a potential to create flawless suppression system.

REFERENCES

[1] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Transactions on acoustics, speech, and signal processing, vol. 27, no. 2, pp. 113–120, 1979.

[2] H.-G. Hirsch and C. Ehrlicher, "Noise estimation techniques for robust speech recognition," in Proc. ICASSP, 1995, vol.1, pp. 153–156.

[3] T. Gerkmann and R.C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 4, pp. 1383–1393, 2012.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 33, no. 2, pp. 443–445, 1985.

[5] A. Maas, Q.V. Le, T.M. O'Neil, O. Vinyals, P. Nguyen, and A.Y. Ng, "Recurrent neural networks for noise reduction in robust ASR," in Proc.INTERSPEECH, 2012.

[6] D. Liu, P. Smaragdis, and M. Kim, "Experiments on deep learning for speech denoising," in Proc. Fifteenth Annual Conference of the International Speech Communication Association, 2014.

[7] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "A regression approach to speech enhancement based on deep neural networks," IEEE Transactions on Audio, Speech and Language Processing, vol. 23, no. 1, pp. 7–19, 2015.

[8] A. Narayanan and D. Wang, "Ideal ratio mask estimation using deep neural networks for robust speech recognition," in Proc. ICASSP, 2013, pp. 7092–7096.

[9] S. Mirsamadi and I. Tashev, "Causal speech enhancement combining data-driven learning and suppression rule estimation.," in Proc. INTERSPEECH, 2016, pp. 2870–2874.

[10] https://grubbrr.com/the-importance-of-background-noise-suppression-in-ai/

[11] https://developer.nvidia.com/blog/nvidia-real-time-noise-suppression-deep-learning/

[12] Jean-Marc Valin, " A Hybrid DSP / Deep Learning Approach to Real-Time Full-Band Speech Enhancement," arXiv:1709.08243v3 [cs.SD] 31 May 2018

[13] Urmila Shrawankar and Vilas Thakare, "Noise Estimation and Noise Removal Techniques for Speech Recognition in Adverse Environment,".