Survey Paper - Deep-Learning Based Sound Recognition To Help Hearing Impaired Individuals

Shreya Babulkar[1], Rohit Chaudhari[2], Pavan kamal Pullabhotla[3], Vini Dubey[4], Prof. Rajesh Tak[5]
[1, 2, 3, 4, 5] Dhole Patil College of Engineering, Wagholi, Pune

Abstract- *Classification Techniques using deep learning have had significant advancements in the past decade while having major and significant amounts of growth in importing these techniques to smaller scales edge devices and less resource driven devices. The most commonly used deep learning algorithms are CNNs (Convolutional Neural Network) which excel in classification problems. But most of these classification techniques have been aimed at image classification, for this reason researchers have been applying these techniques for sound classification problems by converting the sound input - in the form of frequency spectrum- to their visual counterparts for example, spectrograms for better results.*

I. INTRODUCTION

Hearing Disability is a major issue prevailing in India. According to **WHO** (2018) data, the prevalence of hearing impairment (HI) in India is around **6.3% (63 million people suffering from significant auditory loss)**. The estimated prevalence of adult-onset deafness in India is **7.6%** and childhood-onset deafness is **2%**. The Data generated doesn't hold up to the people who haven't approached the doctors due to financial reasons or due to social stigma as the amount to be spent for the diagnosis of this issue is high.

While many people use smartphones, which are available in many different cost categories and are highly capable of complex computations that may help such individuals by using deep learning for having a way to interact with the world in a new and less stigmatized manner. In recent times the field of deep learning has had major improvements in accuracy and efficiency using less resources which makes it easier to use it at smaller scales such as smartphones.

The majority of solutions on portable hearing aid solutions or with projects having similar agenda regarding sound classification have architecture consisting of the first model for **Noise Reduction(NR)** for cleaner data input for the latter stage of the process having a **deep neural network (CNN)** for classification and identification of important sounds[1].

The present research has focused on the development of effective **noise reduction (NR)** techniques that either reduce noisy speech and amplify the core signal as input to the processors of **CI(Cochlear Implant)** devices[2]. These noise reduction approaches can be further classified based on the sound capture setup i.e., multiple and single microphone frameworks. The advantage of multiple microphone framework is of better separation in sources of noise and target, processing of the secondary noise source to have its volume reduced. but this setup becomes less advantageous in reverberant environments, with single microphone framework handling of multiple sources is not required.

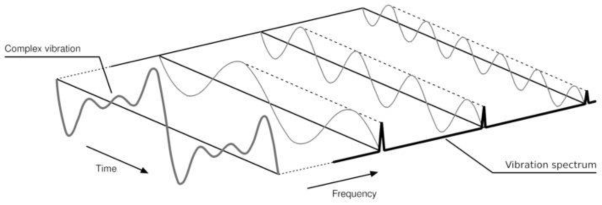
The Captured sound through one of the above frameworks takes a digital form which is less efficient as an input for modern deep learning classification algorithms such as CNN, the sound data is usually visualized as a time variate frequency spectrum which is then converted to more efficient visual representation such as a spectrogram. Before this the input data is processed through stages of clean tones and regular background static and noise including whispering and rustling or with a different approach rather than controlling and searching through different sound types through the data one can use the phenomenon of "destructive interference" to generate a phase shifted sound for the noise part and combine with output from The Noise Reduction module to cancel out the noise which comes with its own downside of optimization dependent and may sometimes destroy parts of data that may be of importance.

II. ARCHITECTURE

Noise Reduction Phase :

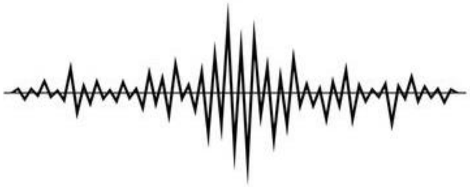
Many researchers use different approaches to solving the noise reduction problem. In case of traditional mathematical approaches these include fourier analysis and destructive interference which are commonly used in signal processing in preference to other acoustic processing techniques. And of the two mentioned fourier Transform are heavily favoured for its advantages over the more resource driven approach of LMS and FXLMS algorithms which better suited for cloud based computation setup.

Fourier Transform uses fourier analysis to find the spectrum of clean tones which make up the susurration(whispering or rustling) spectrum which was picked from the frequency spectrum of the sound. This forms a dactylogram (fingerprint) of the static background in our sound files.then the algorithm (Fourier analysis) finds the frequency spectrum of each short segment from the raw input we give and reduces the volume of these fingerprinted sounds if they are below a certain threshold. That way intentionally hummed or mimicked by a person is picked while the static background noise is lowered in volume. Then there are multiple stages and varieties of sampling and segmentation of the sound in a consecutive manner including **FFT** and hann windowing to smoothen the spectrum.

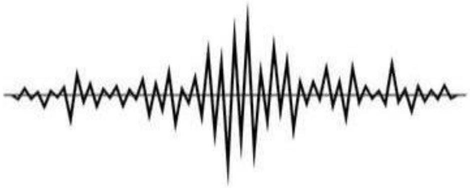


Using “**Destructive Interference**” for noise reduction, generating a phase shifted sound for every unnecessary sound being generated to cause Noise cancellation.

Fig(a) is a sound wave generated. **Fig(b)** is phase shifted sound generated by using the FxLMS algorithm.



Fig(a)



Fig(b)

Every **Peak** generated in Fig(a) has a **Valley** being generated in Fig(b). Playing these sounds together they cancel each other out causing “**Destructive Interference**”.

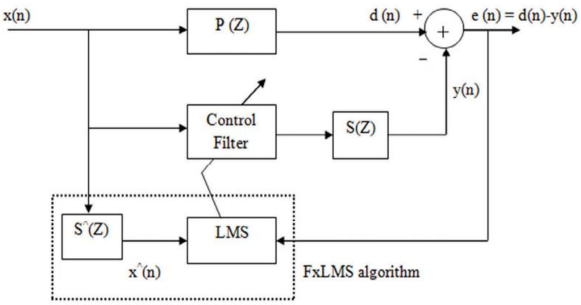
FxLMS Algorithm :

The FxLMS algorithm is a widely used adaptation algorithm used for ANC applications, which is an extension version of the LMS algorithm [7]-[10]. The block diagram for a single-channel feedforward ANC system using the FxLMS algorithm is shown below. Here, **P(z)** is primary acoustic path between the reference noise source and the error microphone and **S(z)** is the secondary path following the ANC filter **w(z)**.The reference signal x(n) is filtered through **S(z)**, and appears as anti- noise signal **y’(n)** at the error microphone. This anti-noise signal combines with the primary noise signal **d(n)** to create a zone of silence in the vicinity of the error microphone. The error microphone measures the residual noise **e(n)** , which is used by ***ω*(z)** , for its adaptation to minimize the sound pressure at the error microphone. Here **S^(z)** accounts for the model of the secondary path **S(z)** between the output of the controller and the output of the error microphone. The filtering of the reference signals **x(n)** through the secondary path model **S^(z)** is demanded by the fact that the output **y(n)** of the adaptive controller ***ω*(z)** is filtered through the secondary path **S(z)**. The expression for the residual error **e(n)** is given as,

$$\mathbf{e(n) = d\ (n)\ \ yc(n)}$$

Where **yc(n)** is the controller output **y (n)** filtered through the secondary path **S(z)** . The **y’(n)** and **y(n)** computed as,

$$\mathbf{y'(n) = S^{T}(n)y(n)}$$



Where, $\boldsymbol{\omega}$**(n) = [**$\boldsymbol{\omega_0}$**(n),**$\boldsymbol{\omega_1}$**(n),...,**$\boldsymbol{\omega_{L\text{-}1}}$**(n)]**$^{\mathbf{T}}$ is tap weight vector, **T x(n) =[x(n),x(n-1),.....,x(n -L+1)]**$^{\mathbf{T}}$ is the reference signal picked by the reference microphone and **S(n)** is the impulse response of secondary path **S(z)** . It is assumed that there is no acoustic feedback from the secondary loudspeaker to the reference microphone. The FxLMS update equation for the coefficients of ***ω*(z)** is given as,

$$\boldsymbol{\omega}\mathbf{(n + 1) =}\ \boldsymbol{\omega}\mathbf{(n) +}\ \boldsymbol{\mu e}\mathbf{(n)x'(n)}$$

Where, $\boldsymbol{\mu}$ is the step size. It determined the convergence rate and **x'(n)** is given by,

$$\mathbf{x'(n) = S\hat{}^{T}\,(z)x(n)}$$

Where **x'(n)** is reference signal **x(n)** filtered through secondary path model **S^(z)** .While using FxLMS with fixed step size, the convergence rate is slow and it affects the stability of the system. In order to overcome this problem, FxLMS with variable step size is preferred.

Fourier Analysis :

It is the method whereby any periodic function can be broken down into a convergent trigonometric series of the form $\boldsymbol{f}(\mathbf{x}) = \boldsymbol{a}_{\mathbf{0}}\mathbf{/2} + \boldsymbol{\Sigma}^{\mathbf{n=1}}{}_{\infty}(\boldsymbol{a}_{\boldsymbol{n}}\mathbf{cos}(\boldsymbol{n}\mathbf{x}) + \boldsymbol{b}_{\boldsymbol{n}}\mathbf{sin}(\boldsymbol{n}\mathbf{x}))$ where $\boldsymbol{a}_{\boldsymbol{n}}$ and $\boldsymbol{b}_{\boldsymbol{n}}$ are constant coefficients. Fourier analysis is the process of determining the frequency domain function from a time function.

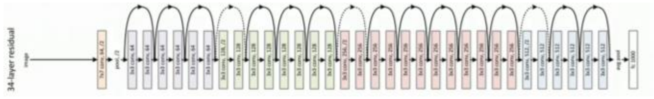
Fast Fourier transform (FFT) :

An algorithm (e.g. the Cooley—Tukey method) which enables the Fourier transformation of digitized wave-forms to be accomplished more rapidly by computer than would be possible using direct evaluation of the **Fourier integral. FFT** usually involves iterative techniques.

Deep Learning Model :

The Classification phase comprises a deep neural network based on Convolutional neural networks which are well suited to classification problems. The convolution in the convolution stage of a single convolution is achieved through kernels that are convolved with the RGB image.In case of multiple kernels all feature maps obtained from distinct kernels are stacked to get the final output of that layer.

To introduce a non linearity to the linear convolution operation, A non-linear Activation function layer is introduced for example, ReLU. Now the Pooling stage of the convolution layer replaces the output of a node at certain locations with a summary statistic of nearby locations and pooling can be of different types : Max, Average, Sum, etc. The max pooling reports the maximum output within a rectangular neighbourhood. Pooling helps to make the output approximately invariant to small translation i.e. it reduces the dimensionality of the spectrogram from the noise reduction module. The model will comprise an Architecture Resembling ResNet with Adam optimizer.



III. PROPOSED SYSTEM

Providing a solution such as an application which can be installed on a mobile device to convert the sound being recognised into transcript to create a hassle free solution for the hearing-impaired.The app converts all the given sounds by noise reduction and classify the sounds such as water running, alarm sounds, dog barking etc. using Fourier analysis and deep learning.

The core processing will be On-Device. The user will be able to refer to the sounds around him/her for a better understanding of the surroundings. This solution will be cost effective in providing help to the hearing-impaired by improving the standard of living.

IV. CONCLUSION

The performance of multi-microphone-based approaches can degrade in reverberant environments, and their applicability is often limited to acoustic conditions in which speech and noise sources differ. Recently, deep learning (Hinton et al. 2006) -based models, constructed with multiple hidden layers, have excelled at a variety of pattern classifications. FFT helps in converting the time domain in frequency domain which makes the calculations easier as we always deal with various frequency bands in communication systems Another very big advantage is that it can convert the discrete data into a continuous data type available at various frequencies.

REFERENCES

[1] Omkar Chavan, Nikhil Kharade, Amol Chaudhari, Nikhil Bhalke, Prof. Pravin Nimbalkar, "Machine Learning and Noise Reduction Techniques for Music Genre Classification" Volume: 06 Issue: 12 | Dec 2019, p-ISSN: 2395-0072, e-ISSN: 2395-0056.

[2] Ying-Hui Lai, Yu Tsao, Xugang Lu, Fei Chen, Yu-Ting Su, Kuang-Chao Chen, Yu-Hsuan Chen, Li-Ching Chen, Lieber Po-Hung Li, Chin-Hui Lee, "Deep Learning–Based Noise Reduction Approach to Improve Speech Intelligibility for Cochlear Implant Recipients." Article in Ear and Hearing · January 2018, Lai et al. / EAR & HEARING, VOL. XX, NO. XX, 00–00.

[3] A. Zorzo, W. D'A. Fonseca , E. Brandão , P. H. Mareze, "Design and analysis of a digital active noise control system for headphones implemented in an Arduino

compatible microcontroller.” ArtigoSIIM-SPS2017 November 2017.

[4] Abhishek Manoj Sharma, “Speaker Recognition Using Machine Learning Techniques”, San Jose State University " (2019). Master's Projects. 685. DOI: https://doi.org/10.31979/etd.fhhr-49pm.

[5] Krishna A/L Ravinchandra, “Active Noise Reduction using LMS and FxLMS Algorithms” Krishna A/L Ravinchandra et al 2019 J. Phys.: Conf. Ser. 1228 012064.