Real-Time Audio Noise Elimination Using A Recurrent Neural Network

Landon Burleson
Electrical and Computer Engineering
Oklahoma State University
Stillwater, Ok, USA
landon.burleson@okstate.edu

Madhusti Dhasaradhan
Electrical and Computer Engineering
Oklahoma State University
Stillwater, Ok, USA
madhusti.dhasaradhan@okstate.edu

Alex Sensintaffar
Electrical and Computer Engineering.
Oklahoma State University
Stillwater, Ok, USA
alex.sensintaffar@okstate.edu

1. Introduction

Noise in audio is detrimental to a listener in terms of understanding and enjoyment. Filtering the undesirable frequency ranges in real-time has proven to be computationally heavy in terms of processing power required to achieve this goal. Graphics card manufacturer, Nvidia, has implemented such a feature for their former generation of Graphics Processing Units (GPU) and beyond by leveraging the bandwidth available in the aforementioned processors. However, using the real-time filtering available on this hardware requires immense power and proprietary algorithms. The goal for this project is to propose an open-source Neural Network architecture that is efficient enough to satisfy the real-time filtering offered elsewhere on various hardware. This problem is interesting due to the complexity of accounting for speed of the neural network and the approach in which to tackle voice recognition within audio tracks or audio channel(s).

2. Problem Statement

Utilizing an RNN architecture, we will be able to filter environmental audio and background noise in real-time for better clarity to a listener in a virtual meeting / conference setting. The primary dataset that will be used to accomplish this will be the MS-SNSD[1] dataset referenced below in the **Data** section as well as a web link shown in the footnote below. Existing implementations utilized LSTM[2] based RNNs. We propose utilizing a transformer based RNN structure to modernize the architecture and allow our model to more easily adapt to long time-series sequences without hindering performance.

3. Data

The primary data that will be used will stem from the MS-SNSD dataset. This dataset has both clean and noisy voice audio files that can be used to train our proposed neural network architecture to filter noise and focus on speech patterns. Other datasets that may be used is the audio MNIST dataset that has short audio recordings of low-noise voices reading simple sentences. This could be used in conjunction with the MS-SNSD dataset to better recognize voices within our network architecture. See the footnotes for hyperlinks to both of the described datasets.

4. Prior Work

A recent paper utilized a LSTM based RNN architecture to provide recognition of voices and similarity between samples[2] . Other implementations, such as Nvidia's, are proprietary and thus encompassed in an unreachable 'black box'. Utilizing these references, we will innovate with an open-source, low overhead RNN model.

5. Evaluation

Our proposed model will be evaluated using the audible quality. This would be done by directly comparing the input and output manually. The scipy module in python also provides audio-based analysis to examine the frequency domain of the audio tracks. Another metric to measure performance is the delay needed to make a computation between a transmitted stream (speaker) and the received stream of audio (listener). Another quantitative measure is the signal-to-noise ratio that can be measured using the scipy python module.

6. Timeline

The following table represents our tentative timeline.

Time	Landon	Madhusti	Alex
Week 1	Prior Work Investigation	Prior Work Investigation	Prior Work Investigation
Week 2	Data collection / cleanup	Data collection / cleanup	Data collection / cleanup
Week 3	Architecture Design	Architecture Design	Architecture Design
Week 4	Test / Debug	Test / Debug	Test / Debug

[1] https://github.com/microsoft/MS-SNSD

[2] Z. Li and P. Song, "Audio similarity detection algorithm based on Siamese LSTM network," 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP), 2021, pp. 182-186, doi: 10.1109/ICSP51882.2021.9408942.