Machine Learning Approach for Source Separation in Hearing Aids

A Nirma University Funded
Minor Research Project Report

Submitted By

Dr. Manish I. Patel
(Asst. Prof, EC, IT-NU)

Department of Electronics and Communication Engineering
Institute of Technology
Nirma University
Ahmedabad 382 481

July 2021

Acknowledgement

I extend my sincere gratitude to management of Nirma University and Institute of Technology to provide me an opportunity to carry out a minor research project. I thank the Director, Institute of Technology, for showing trust and granting permission to carry out the Nirma University funded minor research project.

With immense pleasure I express my sincere gratitude, regards and thanks to my project mentor Dr. N. P. Gajjar for his guidance, invaluable suggestions and continuous encouragement at all the stages of my research work.

I would like to express my thanks to Dr. Dhaval Pujara, HOD – EC Department for his valuable support. I would like to thanks the members of the review committee of this project Dr. D. K. Kothari and Dr. Manisha Upadhyay for their valuable inputs and time. I would like to thank faculty members of the department who have helped me directly or indirectly in the project work. I would also like to thank Mrs. Vaibhavi Mehta for carrying out purchase related work for the project.

Dr. Manish I. Patel

Abstract

Hearing is very important for effective communication and so for quality of life for every individual. But noisy environment is a hurdle especially for hearing impaired listeners. As per the reports by National Institute on Deafness and other Communication Disorders (NIDCD), more than 360 million people across the world suffer for some kind of hearing loss.

Consider the scenario of bus station, train station or market. While people are talking, they have background noise. If they can't isolate the main speech signal from noise they can't communicate properly. In Hearing Aids (HA) (which should not be simply an amplifier but more than that), suitable speech processing and more importantly Speech Enhancement (SE) techniques are used which can solve the problem of source separation from background noise. In this research project, machine learning approach is explored for speech signal separation from the mixed/noisy background which is suitable for the hearing aid application.

Short Time Fourier Transform (STFT) is used as a feature to represent the speech in transform domain. The STFT is used to find the binary mask for the target speech. The binary mask is then used to recover the targeted speech by taking the inverse DTFT. The simulation is carried out first in MATLAB and then in Python also to get its good hardware support. The technique is deployed on hardware like Raspberry pi in which good quality of separated signal is observed from the separated speech signal.

INDEX

Chapter 1 Introduction

1.1 Introduction

Hearing is very important for effective communication and so for quality of life for every individual. But noisy environment is a hurdle especially for hearing impaired listeners. Hearing aid can help people to improve the listening capability. The simplest hearing aid can amplify the speech signal. In the modern hearing aid, they also remove some background noise using some frequency selective filters. In those hearing aid the techniques used are enhancing the signal in the frequency of interest while attenuates the frequency of noise or background speech. Such decisions are based on any well-established theories and mathematical methods. However as reported in recent years machine learning can also help for such decision or classification.

1.2 Motivation

The reports and statistics reviewed by National Institute on Deafness and other Communication Disorders (NIDCD), showing seriousness of hearing loss. Overall, 15% of adults aged 18 years and over had more or less hearing trouble without a hearing aid [1]. Age is the strongest predictor of hearing loss among adults aged 20-69 years, with the greatest amount of hearing loss in the 60 to 69 age group [2]. The numbers are increasing in recent years. Many of those people are avoiding use of hearing aid may be because of its cost, unavailability or poor performance. This ignorance affect their communication and ultimately quality of their life.

In recent years due to the advancement in computational resources, datasets and machine learning algorithms, people have started use of machine learning for certain decisions may be in terms of classification, clustering or regression. In many areas machine learning has already given prominent improvements and people are exploring many more areas. Machine learning is also being used for speech processing. Specifically speech separation for hearing aid using machine learning can improve the performance. Such speech or its source separation also increases the effectiveness of Automatic Speech Recognition (ASR) by separating main speech from background speech or noise.

1.3 Objectives

The objective of the project is to investigate machine learning based speech or its source separation techniques for hearing aids. Various signal processing techniques, speech signal features and machine learning models investigated for separation of speech to be used in hearing aid. So the main objectives of the project are:

- machine learning approach for hearing aid
- Investigate speech features for machine learning model
- Speech separation for improved speech enhancement in hearing aid

1.4 Organization of the Report

The report consists of six subsequent chapters as mentioned below.

- Chapter-1: The first chapter introduces this work and it also covers motivation and objectives.
- Chapter-2: In the second chapter, the literature review of speech processing techniques and approaches available in literature is presented.
- Chapter-3: The third chapter describes the proposed methodology for speech separation for its low-cost hardware implementation.
- Chapter-4: In the fourth chapter, the hardware and software used are presented with the discussion on why they have been selected.
- Chapter-5 Finally, in the fifth chapter the work carried out on software platforms and its hardware set-up is presented.

Chapter 2 Literature Survey

Many of the existing algorithms for speech separation are based on some background knowledge of speech, its characteristics and nature of underlying noise and so they may not performing well in complex real world scenario [3]. Machine learning takes some derived features as input and based on it takes some decision may be in terms of classification, regression or clustering.

In addition to the machine learning model the other important part is features. Some of recent methods [4]-[6] which are based on deep neural networks, don't need such background knowledge. In these methods a regression model is trained to predict clean speech from the noisy speech. In [7], transfer learning is used along with neural network, which improved the performance. In that transfer learning a pre-trained network is used which is improved further. In [8] a prototype model is developed for implementing some of the speech processing techniques such as noise reduction filter, frequency shaper and compression for hearing aids application, but machine learning is not incorporated in their work.

In addition to these a mask based methods some other methods are also reported. The authors in [9] used two dimensional Recurrent Neural Network (RNN) and skip connections to separate a single speech signal from mixtures of multiple speeches.

Different models investigated for speech enhancement in [10]. Google's WaveNet architecture has been modified for noise reduction in a model called WaveNet de-noising and has proven to be state-of-the-art. They also extended their work for Speech Enhancement Generative Adversarial Network (SEGAN) which adapts the GAN architecture into applications on speech. This also gave the idea to investigate possibilities of such different models for speech separation as well.

Chapter 3 Proposed Design

3.1 Proposed System Overview

The research project is mainly on application of machine learning for speech enhancement for hearing aids by separating the target speech from mixtures of speech. The general system level block diagram is shown in figure 1.

On the selected Single Board Computer (SBC) along with the main processor other required interfaces are available or can be made available on the board such as mic, loudspeaker, ADC (to convert analog audio signal into digital signal) and DAC (to convert processed/enhanced/separated digital signal into analog signal for loudspeaker) as shown in the figure 1. But the focus of work is on digital speech signal processing technique using machine learning approach to enhance its quality by separating the targeted speech signal from the mixtures of two speech signals or speech signal with background noise. For that some available options as single board computer are Raspberry pi, NVIDIA Jetson nano and likewise. Such boards are like mini-computer with different level of complexity, resources (computational and memory) and features.

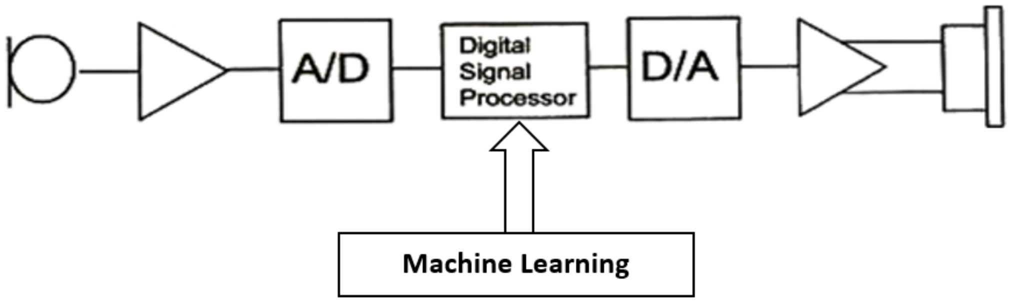

Figure 1 System Level Block Diagram

Selection of the board depends on some criteria of application. For complex algorithm and machine learning techniques the NVIDIA Jetson nano is preferred on which good computational and memory resources are available. For a light algorithm small sized low-cost board such as Raspberry pi can be preferred. Nowadays their operation system also support Python environment so that deployment of algorithm development made in Python can be made easy, just it is required to install the respective libraries.

3.2 Software Implementation Overview

In the conventional algorithm speech signal can be separated using some filters or other such analytical methods. In this project machine learning approach is to be investigated. In machine learning the data contains certain important attributes which are called features. This features are to be derived or decided based on the application. In machine learning the selected or derived features are given as input to classification algorithm. In our approach short Time Fourier Transform (STFT) is used as features because it contains both time (i.e. spatial) domain and frequency domain information. Specifically for a speech signal a value at some time instance only can't explain the speech signal well instead additional temporal information needed to understand the meaning of speech. Use of STFT in a machine learning framework is presented in figure 2. First say from the mixture of two speech signals, its STFT is calculated and then based on it binary map is created from the data or using a machine learning model like deep neural network. Then for the actual testing signal, after its STFT it is passed from the binary mask then inverse DTFT to recover the targeted speech. This process can be understood from figure 2.

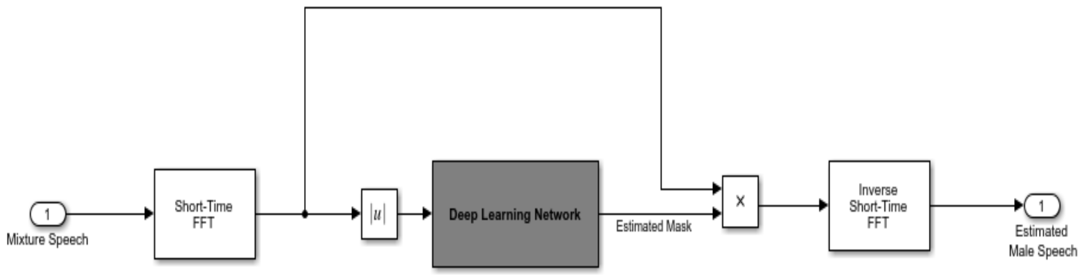

Figure 2 Source separation using binary mask

Chapter 4 Software and Hardware Description

For the proposed approach first the required hardware and software are presented in this chapter.

4.1 Hardware

To deploy the speech separation technique some of the options are studies and then Raspberry pi board and NVIDIA Jetson nano board are used. Selection of the board depends on some criteria of application. For complex algorithm and machine learning algorithm the NVIDIA Jetson nano is preferred. For a light algorithm small sized low-cost board such as Raspberry pi can be preferred. Nowadays their operation system also support Python environment so that deployment of algorithm development done in Python can be made easy.

Figure 3 Raspberry pi 4 Model B

The Raspberry Pi is a credit card-sized printed circuit board that may be used for a variety of tasks on a computer, including games, word processing, spreadsheets, and even for playing High Definition videos. The Raspberry Pi 4 Model B is the most recent version of the popular Raspberry Pi computer line. It delivers substantial improvements in processing speed, multimedia performance, memory, and connection over the previous generation Raspberry Pi 3 Model B+ while preserving backward compatibility and similar power consumption. The Raspberry Pi 4 Model B used in the work is shown in figure 3.

Figure 4 NVIDIA Jetson Nano board

The NVIDIA Jetson Nano is a compact, powerful single-board computer that allows many neural networks to operate in parallel for applications including image classification, object identification, segmentation and speech processing. It includes a full development environment (JetPack SDK) as well as libraries for embedded applications, deep learning, IoT, computer vision, graphics, multimedia and other topics. The combination of a Jetson Nano and a GeForce-enabled graphics processor with the same CUDA cores offers a very powerful application development environment. In addition, the Jetson Nano includes a CPU-GPU hybrid architecture, which allows the operating system to be booted by the CPU and designed to accelerate the CUDA capable GPU's sophisticated machine learning tasks. Artificial intelligence demonstrates that running algorithms use very little electricity. Jetson nano, the Nvidia Jetson ecosystem's medium-sized board, it comes in handy when it comes to computer vision and deep learning.

The technical characteristics of the Raspberry Pi4, Jetson Nano boards are listed in Table 1.

Table 1: Comparison of specifications of available two single computer boards Raspberry Pi4 and Jetson Nano

Parameter	Raspberry PI 4 B	Jetson Nano
Performance	13.5 GFLOPS	472 GFLOPS
CPU	Quad-core ARM Cortex A72 64-bit @ 1.5 GHz	Quad-Core ARM Cortex-A57 64-bit @ 1.42 GHz
GPU	Broadcom Video Core VI (32-bit)	NVIDIA Maxwell w/ 128 CUDA cores @ 921 MHz
Memory	8 GB LPDDR4	4 GB LPDDR4 @ 1600MHz, 25.6 GB/s
Networking	Gigabit Ethernet / Wi Fi 802.11ac	Gigabit Ethernet / M.2 Key E
Display	2x micro HDMI (up to 4Kp60)	HDMI 2.0 and eDP 1.4
USB	2x USB 3.0, 2x USB 2.0	4x USB 3.0, USB 2.0 Micro-B
Other	40-pin GPIO	40-pin GPIO
Video Encode	H264(1080p30)	H.264/H.265 (4Kp30)
Video Decode	H.265(4Kp60), H.264 (1080p60)	H.264/H.265 (4Kp60, 2x 4Kp30)
Camera	MIPI CSI port	MIPI CSI port
Storage	Micro-SD	16 GB eMMC
Power under load	2.56W-7.30W	5W-10W
Price	$35	$89

4.2 Software

Initial simulation is carried out in MATLAB. MATLAB is good for detailed simulation and analysis. But as it not open source, research community in the domain is preferring the open source environment like Python. It gives large number of libraries many of which are for speech processing and hardware implementation also. In this work the main Python libraries used are scipy, matplotlib, numpy and libreso. On the computer system Spyder Python environment is used for simulation while on raspberry pi board the Thony Python environment is used for python 3 programming which is available in its operating system.

Chapter 5 Experimental Results

For the proposed approach, first the simulation is carried out in MATLAB presented in section 5.1. Then for hardware support, the simulation in also done in Python environment which is covered in section 5.2. The experimental set-up is shown in section 5.3.

5.1 Binary Mask using STFT for Speech Separation in MATLAB

STFT is used as a feature which in turn used for speech separation. From the mixtures of male and female speech, one is to be separated. For a given sample speech data-1 having mixture of a male speech and a female speech is shown in figure 5, corresponding STFT features are shown in figure 6. First the purpose is to create a binary mask as shown in figure 7. Then in testing phase this mask is used to separate the targeted speech from mixtures of speech. The original and reconstructed speech signals are shown in figure 8 and 9 for targeted male speech and female speech signals respectively. Their STFT graphs are also shown in figure 10 for comparison purpose.

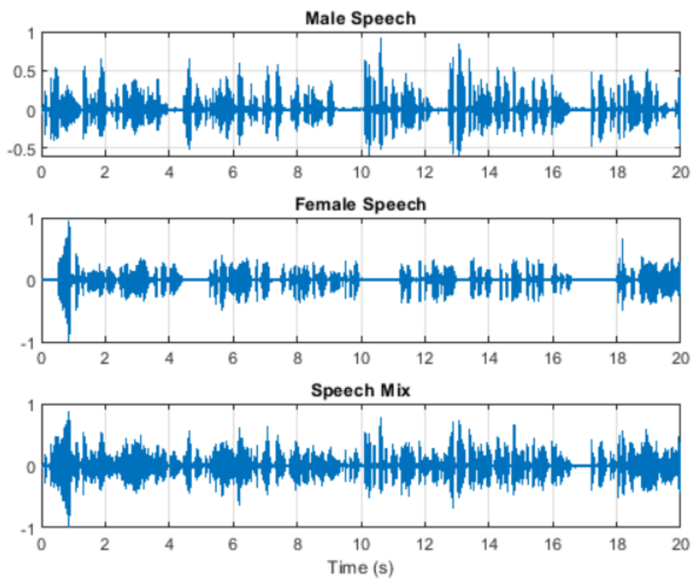

Figure 5 Speech signals – male, female and combined

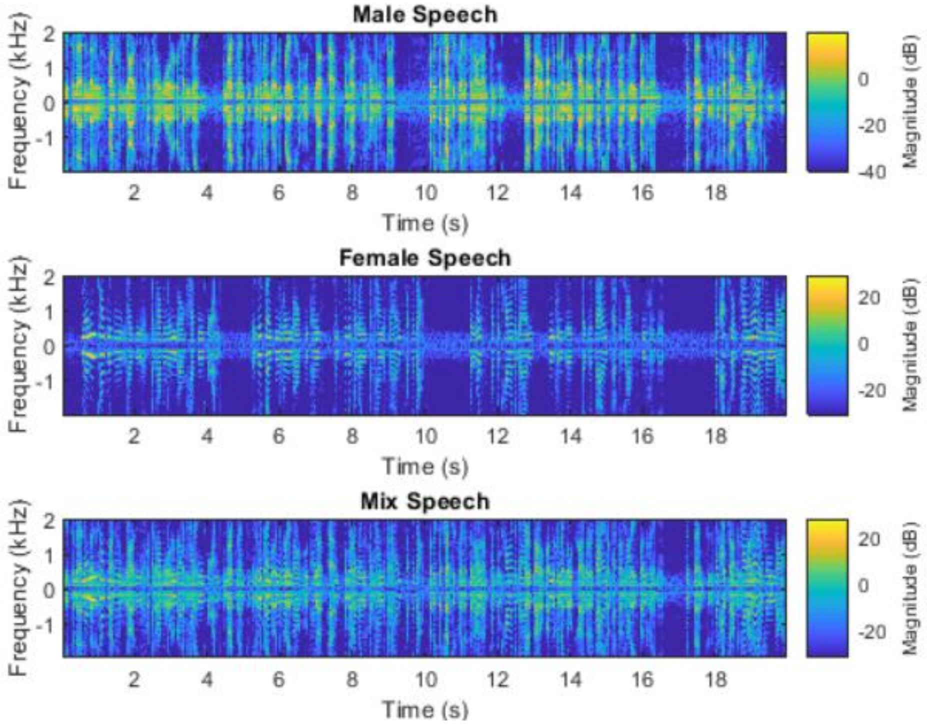

Figure 6 STFT of the three speech signals - male, female and mix

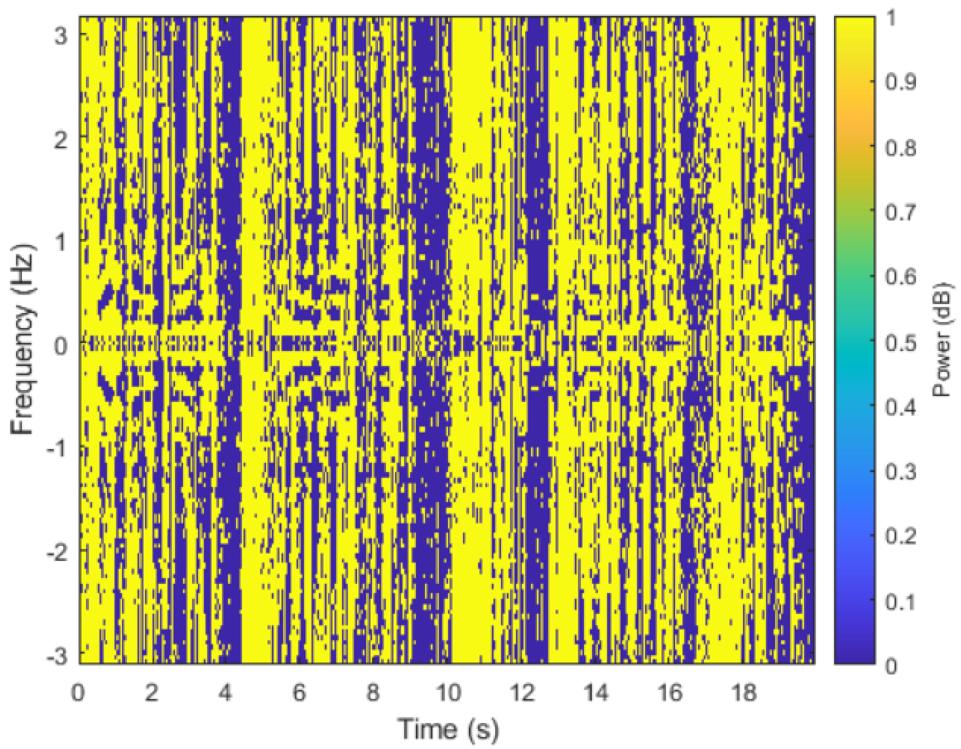

Figure 7 Binary Mask based on STFT

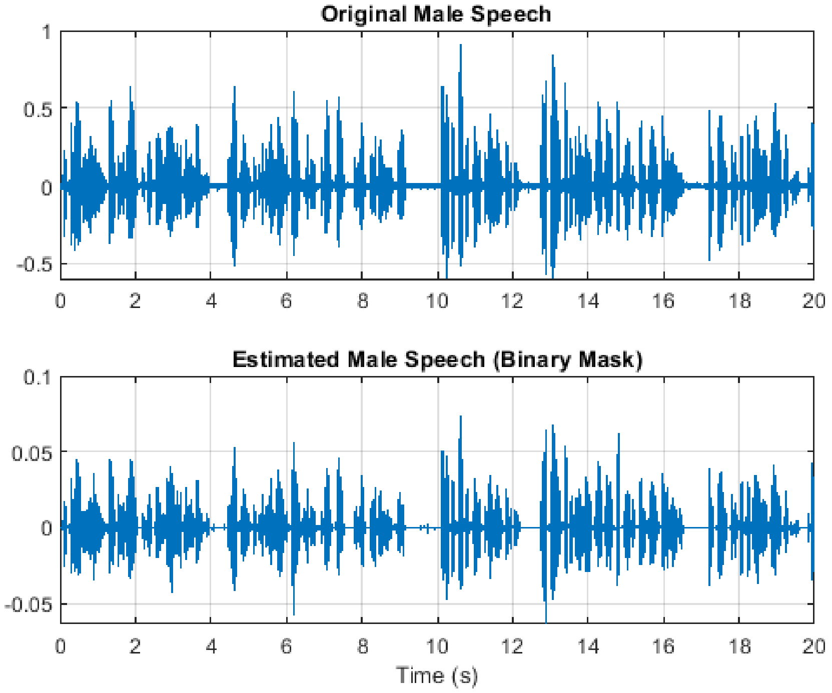

Figure 8 Original and reconstructed male speech signals

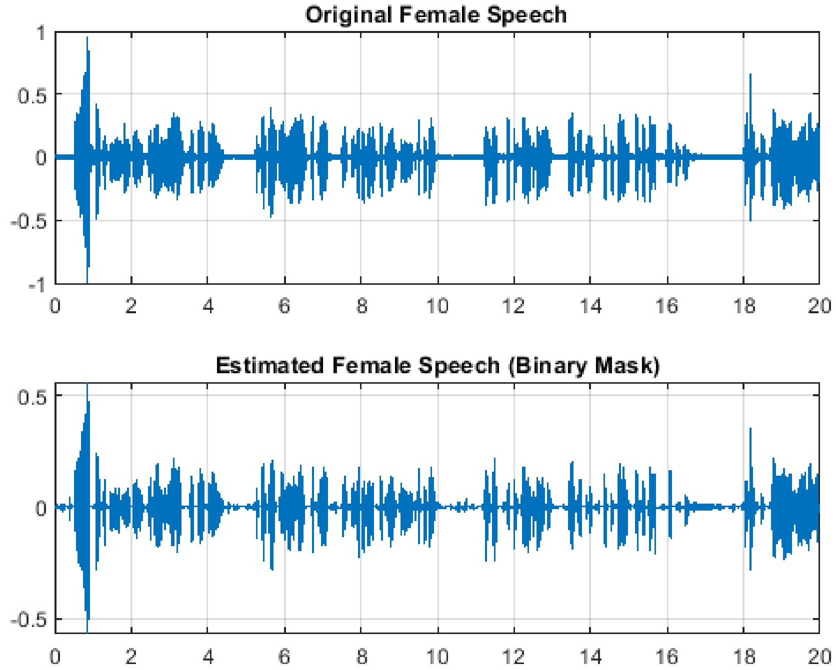

Figure 9 Original and reconstructed female speech signals

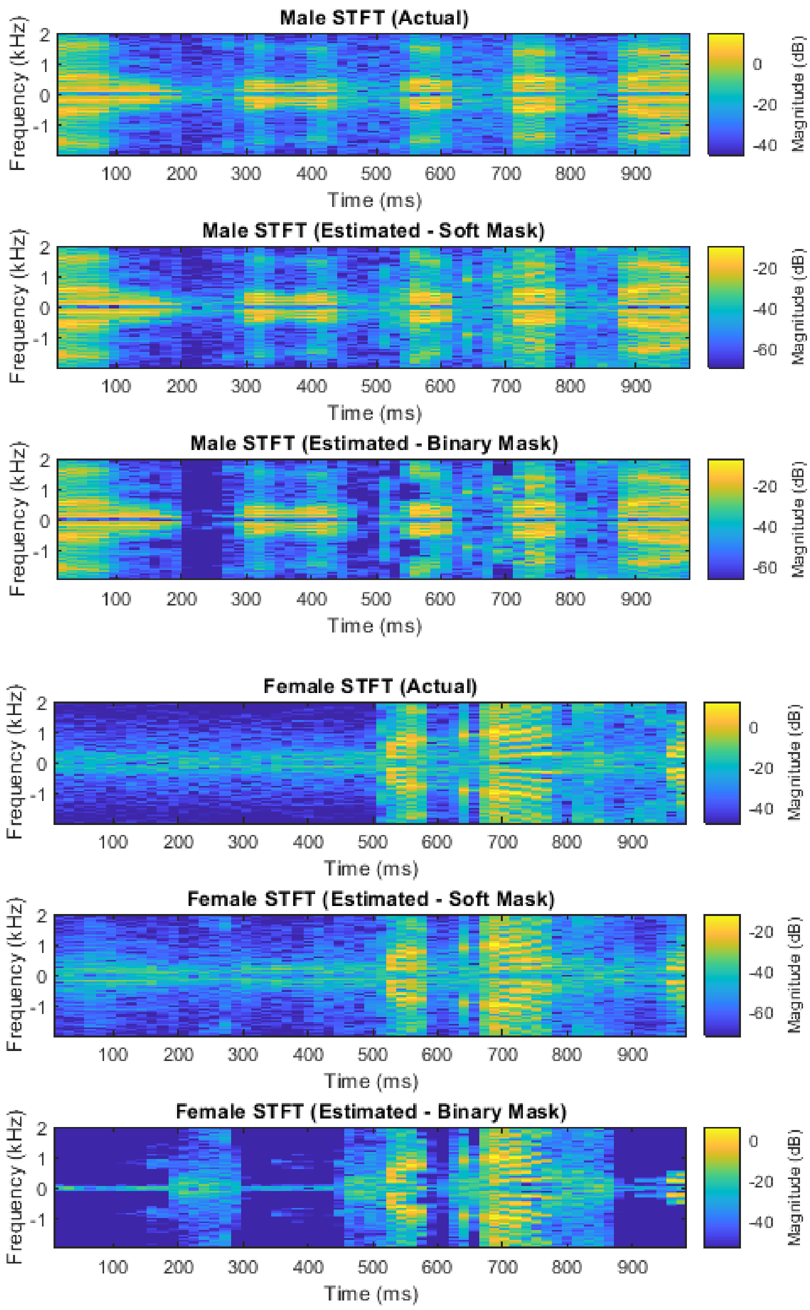

Figure 10 STFT of original and reconstructed speech signals

5.2 Binary Mask using STFT for Speech Separation in Python

MATLAB is good for detailed simulation and analysis. But as it is not open source, research community in the domain is preferring the open source environment like python. It gives large number of libraries many of which are for hardware implementation support also. So in the Spyder IDE (Integrated Development

Environment) for python, the speech separation based on binary mask is obtained using STFT as a feature is presented here on a different sample speech data-2.

Figure 11 shows the combined signals which is basically combination of a male and female speech signal shown in different colour. It is required to separate one of the signal. Similar to the previous section here also the binary map targeting the signal 2 is created based STFT which is presented in figure 12. The reconstructed signal 2 using the binary map and inverse DTFT is shown in figure 13 along with the original signal 2. This indicates the reconstructed signal 2 is very close to the original signal-2. Further same steps were repeated by targeting the signal 1, corresponding binary mask is shown in figure 14. From figure 12 and figure 14, it can be observed that the binary mask for both the experimentations are inverse to each other. The reconstructed and the original signal 1 shown in figure 15.

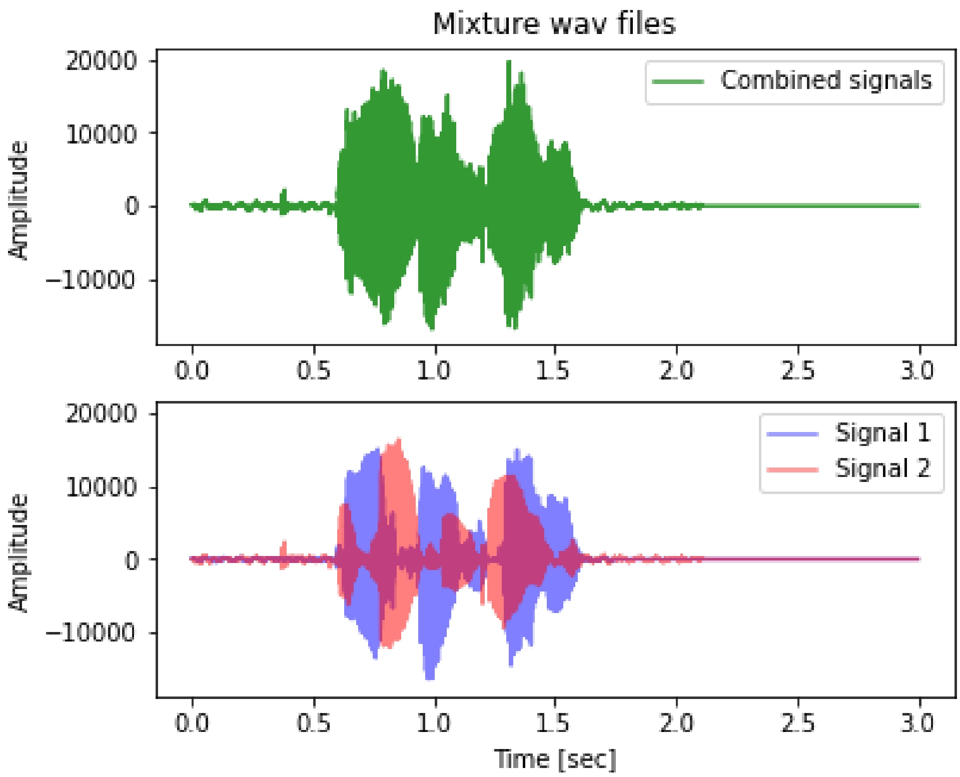

Figure 11 Combined signal and both signals in different color

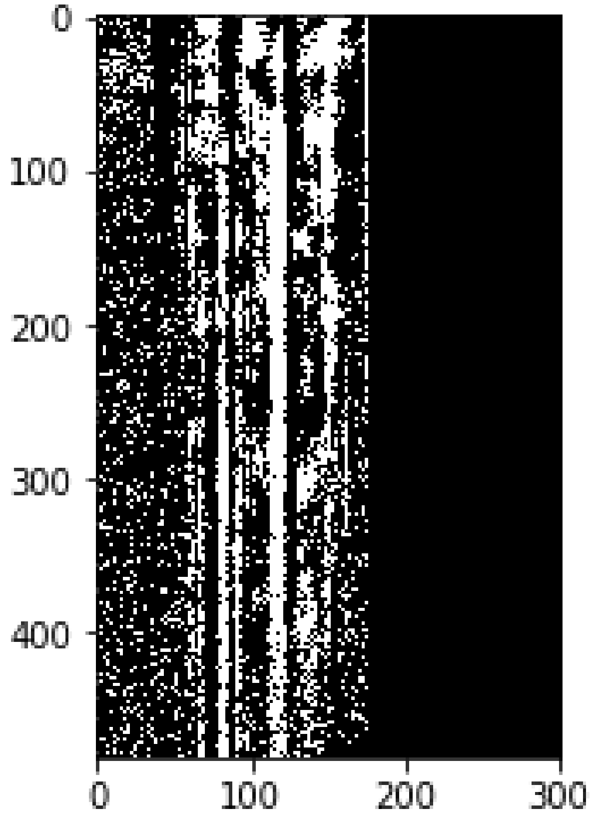

Figure 12 Binary mask targeting the signal2

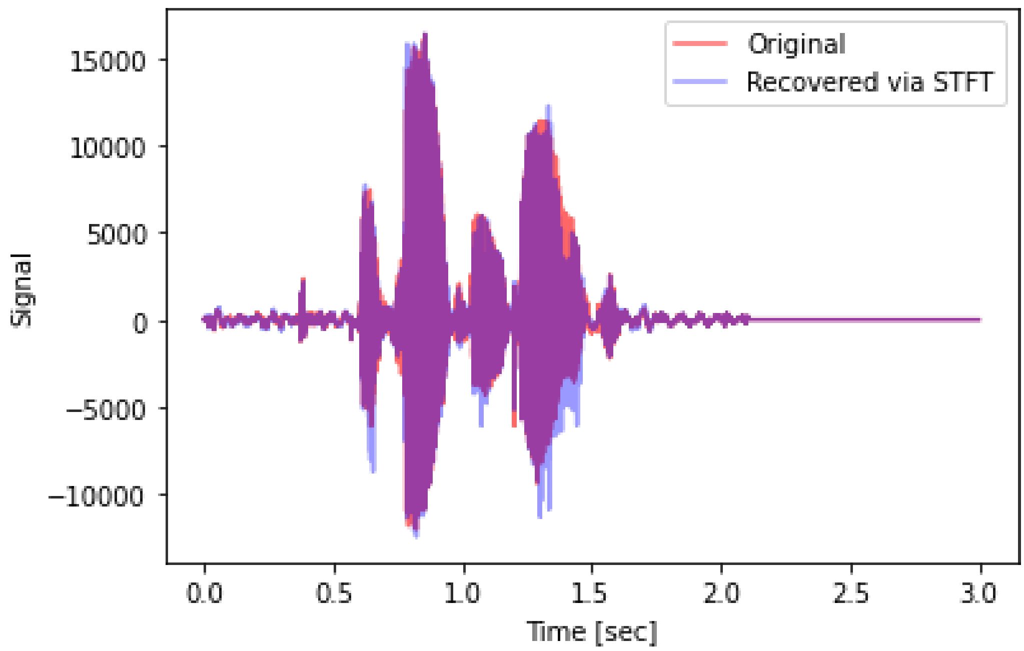

Figure 13 original and reconstructed signal 2

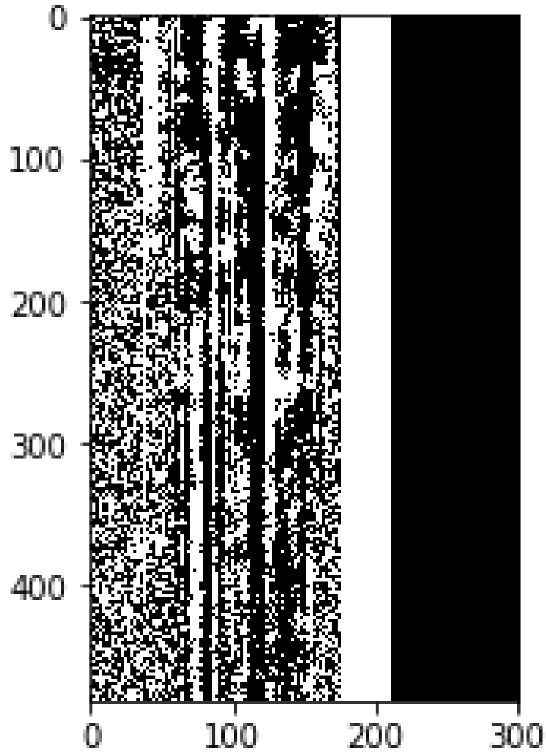

Figure 14 Binary mask targeting the signal 1

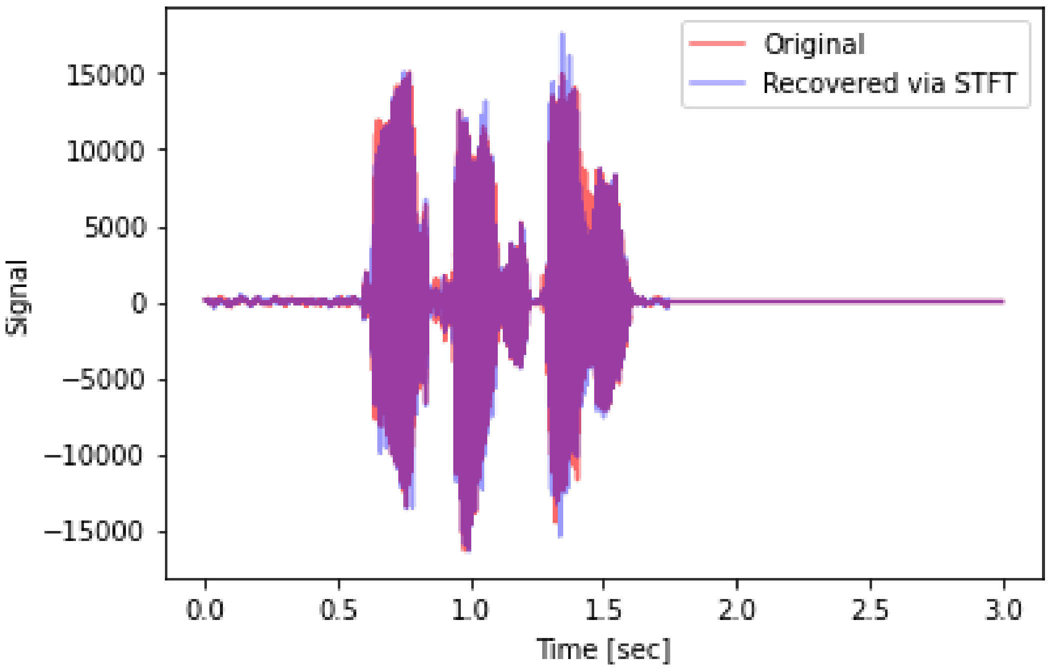

Figure 15 original and reconstructed signal 1

5.3 Hardware Set-up with Raspberry pi

The Python implementation of the proposed approach is deployed on the comparatively light board that is raspberry pi. One such complete hardware set-up is shown in figure 16. The mixture of male and female speech signal is given as an input. Based on the binary mask the targeted speech signal is reconstructed which we can hear from the speaker. This is not image processing in which we visualize the processed image but speech signal we need to listen.

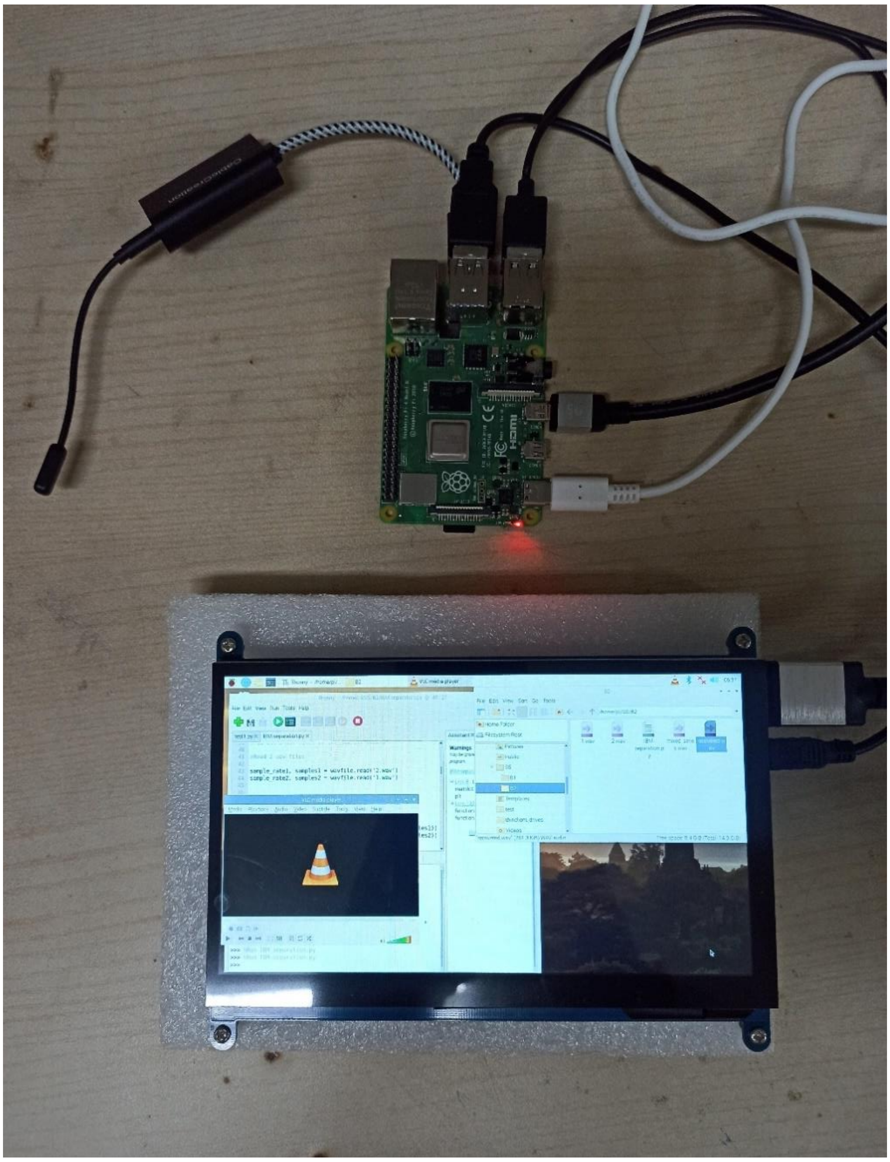

Figure 16 Hardware set-up using raspberry pi board

Chapter 6 Conclusion and Future Scope

6.1 Conclusion

Speech enhancement is achieved by speech separation. From the mixtures of speech using STFT as a feature representation, the binary map is created. Using the binary map the inverse STFT can recover the targeted speech signal. The steps are observed in simulation results carried out in MATLAB first and then also in Python environment to get its advantages of comparatively better hardware support. The software is then deployed on hardware i.e. on raspberry pi. The good quality of separated speech can be realised by hearing it through headphones or earphones connected to the raspberry pi board. Deployment of the algorithm in such a small lightweight board can be considered for the prototype of low-cost small size application such as hearing aid. In addition of the amplification such approach will lead to enhance the performance of speech hearing on hearing aids devices.

6.2 Future Scope

In the future, the proposed approach can be extended for comparative analysis of different speech features. In literature some audio features are available like MFCC, spectrogram model. Further, in addition to the binary mask, phase mask and ratio mask can also be investigated for possible improvements as suggested in literature.

References:

[1] Blackwell DL, Lucas JW, Clarke TC. “Summary health statistics for U.S. adults: National Health Interview Survey”, 2012. National Center for Health Statistics. Vital Health Stat 10(260). 2014.

[2] Hoffman HJ, Dobie RA, Losonczy KG, Themann CL, Flamme GA. “Declining Prevalence of Hearing Loss in US Adults Aged 20 to 69 Years”. JAMA Otolaryngology – Head & Neck Surgery, 2017.

[3] 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

[4] Y. Xu, J. Du, L. R. Dai, and C. H. Lee, “A regression approach to speech enhancement based on deep neural networks,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 1, pp. 7_19, Jan. 2015.

[5] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, “An experimental study on speech enhancement based on deep neural networks,” *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 65_68, Jan. 2014.

[6] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, “Dynamic noise aware training for speech enhancement based on deep neural networks,” in *Proc. 15th Annu.Conf. Int. Speech Commun. Assoc.*, 2014, pp. 1-5

[7] Mondol, S.I.M.M.R.; Lee, S. A Machine Learning Approach to Fitting Prescription for Hearing Aids. *Electronics* 2019, *8*, 736

[8] R. Arsinte, E. Lupu and T. Sumalan, “A Rapid Prototyping model concept for a DSP based hearing aid”, *2017 E-Health and Bioengineering Conference (EHB)*, Sinaia, 2017, pp.337-340

[9] Voice Separation with an Unknown Number of Multiple Speakers Eliya Nachmani 1 2 Yossi Adi 1 Lior Wolf 1

[10] Deep Learning for Speech Enhancement - A Study on WaveNet, GANs and General CNN-RNN Architectures OSCAR XING LUO Master in Computer Science Date: July 1, 2019

Speech Separation

by Manish Patel

Submission date: 29-Jul-2021 04:36PM (UTC+0530)
Submission ID: 1625406229
File name: Chapters.pdf (986.95K)
Word count: 3006
Character count: 15703

Speech Separation

ORIGINALITY REPORT

8%	6%	5%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Zuraida Muhammad, Muhammad Azri Asyraf Mohd Hafez, Nor Adni Mat Leh, Zakiah Mohd Yusoff, Shabinar Abd Hamid. "Smart Agriculture Using Internet of Things with Raspberry Pi", 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 2020 Publication	1%
2	blog.cavedu.com Internet Source	1%
3	www.hearingloss.org Internet Source	1%
4	www.nvidia.com Internet Source	1%
5	www.alamy.com Internet Source	1%
6	Ümit Güz, Hakan Gürkan, Binboga Sıddık Yarman. "A New Method to Represent Speech Signals Via Predefined Signature and	<1%

Envelope Sequences", EURASIP Journal on Advances in Signal Processing, 2006
Publication

7	Serkan Tokgoz, Stephanie Tittle, Linda Thibodeau, Issa Panahi. "A Hearing Test Simulator GUI for clinical testing of speech recognition", Acoustical Society of America (ASA), 2018 Publication	<1%
8	ethesis.nitrkl.ac.in Internet Source	<1%
9	Agus Kurniawan. "IoT Projects with NVIDIA Jetson Nano", Springer Science and Business Media LLC, 2021 Publication	<1%
10	Gautam S. Bhat, Nikhil Shankar, Chandan K. A. Reddy, Issa M. S. Panahi. "A Real-Time Convolutional Neural Network Based Speech Enhancement for Hearing Impaired Listeners Using Smartphone", IEEE Access, 2019 Publication	<1%
11	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1%
12	Michael Syskind Pedersen. "Two-Microphone Separation of Speech Mixtures", IEEE Transactions on Neural Networks, 03/2008 Publication	<1%

13	dissertations.umi.com Internet Source	<1 %
14	www.liebertpub.com Internet Source	<1 %
15	www.seeedstudio.com Internet Source	<1 %
16	"IoT as a Service", Springer Science and Business Media LLC, 2021 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		