A NOVEL APPROACH OF NOISE SUPPRESSION IN AUDIO

P SAI SANDEEP (17SS1A0438)

ALOKE VISHWAKARMA (17SS1A0403)

SEGGARI KOUSHIK (17SS1A0449)

Department of Electronics and Communication Engineering

Jawaharlal Nehru Technological University Hyderabad College of Engineering Sultanpur.

Sultanpur(v), Pulkal(M), Sangareddy-502273, Telangana

2020-2021

A NOVEL APPROACH OF NOISE SUPPRESSION IN AUDIO

A PROJECT REPORT

SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

BACHELOR OF TECHNOLOGY IN

ELECTRONICS AND COMMUNICATION ENGINEERING

BY

P SAI SANDEEP (17SS1A0438)

ALOKE VISHWAKARMA (17SS1A0403)

SEGGARI KOUSHIK (17SS1A0449)

Under the Guidance of

Dr. Y RAGHAVENDER RAO

Department of Electronics and Communication Engineering

Jawaharlal Nehru Technological University Hyderabad
College of Engineering Sultanpur.
Sultanpur(V), pulkal(M), Sangareddy(Dist) – 502273, Telangana
2020-2021

Jawaharlal Nehru Technological University Hyderabad
College of Engineering Sultanpur.
Sultanpur(V), pulkal(M), Sangareddy(Dist) – 502273, Telangana
2020-2021

Department of Electronics and Communication Engineering

CERTIFICATE

Date:

This is to certify that the project report entitled **A NOVEL APPROACH OF NOISE SUPPRESSION IN AUDIO** *is a bonafide work carried out by* **P SAI SANDEEP, ALOKE VISHWAKARMA, SEGGARI KOUSHIK** *bearing Roll Numbers* **17SS1A0438, 17SS1A0403, 17SS1A0449** *in partial fulfilment of the requirements for the degree of* ***BACHELOR OF TECHNOLOGY*** *in* ***ELECTRONICS & COMMUNICATION ENGINEERING*** *by the Jawaharlal Nehru Technological University, Hyderabad during the academic year 2020-2021.*

The results embodied in this report have not been submitted to any other University or Institution for the award of any degree or diploma.

DR.Y. RAGHAVENDER RAO	**DR.Y. RAGHAVENDER RAO**
Head of the Department	Project Guide

ACKNOWLEDGMENT

We wish to take this opportunity to express our deep gratitude to all those who helped, encouraged, motivated and have extended their cooperation in various ways during our Major project work. It is our pleasure to acknowledge the help of all those individuals who were responsible for foreseeing the successful completion of our project.

We express sincere gratitude to **Dr. B. BALU NAIK** Principal of JNTUHCES for his support during the course period.

We sincerely thank **Dr. V. VENKATESWARA REDDY** the Vice Principal of JNTUHCES for his kind help and cooperation.

We are thankful to **Dr. Y. RAGHAVENDER RAO**, Professor and Head of the Department of Electronics and Communication Engineering of JNTUHCES for his effective suggestions during the course period.

We would like to express our gratitude to **Sri V. RAJANESH**, Assistant Professor, for his guidance and constant supervision.

Finally, we would like to express our gratitude to **Mr T. MOHAN DAS**, for spending his valuable time and providing their guidance throughout the course period.

BY

P SAI SANDEEP (17SS1A0438)

ALOKE VISHWAKARMA (17SS1A0403)

SEGGARI KOUSHIK (17SS1A0449)

ABSTRACT

Noise is ubiquitous in almost all acoustic environments. The speech signal, which is recorded by a microphone, is generally infected by noise originating from various sources. Such contamination can change the characteristics of the speech signals and degrade the quality of speech and intelligibility, thereby causing significant harm to human-to-machine communication systems. Noise reduction for speech applications is often formulated as a digital barrier, where the clean speech estimation is obtained by passing the noisy speech through a linear filter. With such a formulation, the core issue of noise reduction becomes how to design an optimal filter that can significantly suppress noise without noticeable speech distortion.

We present a novel approach to denoise the unwanted audio signal so as to process the useful signal and provide a clear sound out of random noises. The sound denoising algorithm is based on the popular spectral subtraction technique. Based on the spectrum of the sound, this technique simply computes an attenuation map in the time-frequency domain. Then, the audio signal is restored by computing the inverse STFT. This is a method for restoration of the power, or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum. The **noise spectrum** is **estimated**, and **updated**, from the periods **when the signal is absent and only the noise is present.**

This project can be used at places where an audio signal is filled with unwanted noise, but the speech or the useful signal needs to be extracted with high precision, and this project does that work effectively and efficiently.

Keywords: Digital signal processing, Noise reduction, Spectrogram analysis, Spectral subtraction technique.

Software: MATLAB

P. SAI SANDEEP (17SS1A0438)

ALOKE VISHWAKARMA (17SS1A0403)

SEGGARI KOUSHIK (17SS1A0449)

CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Speech has always been one of the most important carriers of information for people and it becomes a challenge to maintain its high quality. In many applications of noise cancellation, the changes in signal characteristics could be quite fast. Speech signal is often degraded by additive background noises like babble noise, train noise, restaurant noise, car noise, white noise etc. In such noisy environment listening task is very difficult at the end user.

communication via speech is one of the essential functions of human beings. Humans possess varied ways to retrieve information from the outside world or to communicate with each other, and the three most important sources of information are speech, images and written text. For many purposes, speech stands out as the most efficient and convenient one. Speech not only conveys linguistic contents, but also communicates other useful information like the mood of the speaker. Language communication through speech is closely intertwined with the evolution of human civilization. Noise presented in speech signal Recorded under the real conditions can impair the quality of the signal, reduce intelligibility, and Increase listener fatigue. Since in practice many of noise is presented in recording speech, the problem noise reduction is essential in the world of Telecommunications and has gained much attention in recent years. Noise reduction algorithms in general, attempt to improve the performance of communication systems when their input or output signals are corrupted by noise. The main objective of speech is to improve the speech quality or intelligibility. A hearing aid is a battery-powered, electronic device that makes listening easier for people with a hearing loss. A hearing aid consists of a microphone, an amplifier and a receiver. The microphone picks up sounds in your acoustic environment and turns them into electronic signals.

Fig 1.1: Human – machine communication

Speech communication is the exchange of information via speech either between humans or between human to machine in the various fields, for instance automatic speech recognition and speaker identification.

Generally, we come across many situations, where speech signals are degraded by the ambient noises that limit their effectiveness of communication. Therefore, enhancement of speech is normally required to reduce annoyance due to noise.

1.2 AIM OF THE PROJECT

Speech communication is the exchange of information via speech either between humans or between human to machine in the various fields,

Generally, we come across many situations, where speech signals are degraded by the ambient noises that limit their effectiveness of communication. Therefore, enhancement of speech is normally required to reduce annoyance due to noise. The main purpose of speech enhancement is to decrease the distortion of the desired speech signal and to improve one or more perceptual aspects of speech, such as the quality and/or intelligibility. But the clean speech signal is often corrupted with additive noise every time we record an audio which relate to the situation like the sound of vuvuzela Which sometimes do not let us listen to the commentary while watching soccer matches, and situations where the random

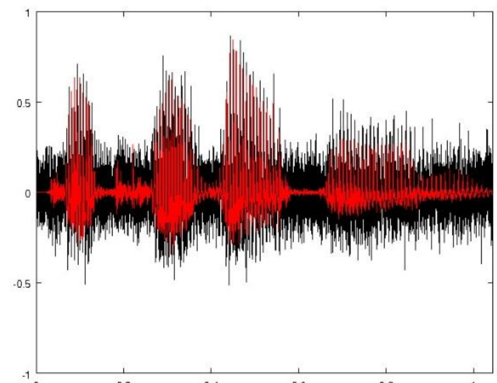

Fig 1.2: Noise signal

noises are more dominant over the speech in hearing aids which is intended to amplify the surrounding sounds of a deaf person.

So, here arises the need of a system which could efficiently suppress the unwanted noise and pull-out only the clean speech out of an audio full of noisy speech signals. There two measures are not necessarily correlated which are quality and intelligibility and therefore, an increase in speech quality does not necessarily lead to an improvement in intelligibility.

The solution to the above problem statement could be a system which can extract the clean speech signal from the inputted audio containing a blend of noise and speech.

Therefore, the AIM of this project is to develop a system or a software which will extract clear speech signal from the provided audio by making use of various methods, filters and algorithms to differentiate and suppress the noise from speech.

1.3 METHODOLOGY

Speech enhancement techniques can be classified into, single channel, dual channel, or multi-channel enhancement. Although the performance of multi-channel speech enhancement is better than that of single channel enhancement, the single channel speech enhancement is still a significant field of research interest because of its simple implementation and ease of computation. In single channel applications, only a single microphone is used, and the characterization of noise statistics is extracted during the periods of pauses, which requires a stationary assumption of the background noise.

The estimation of the spectral amplitude of the noise data is easier than estimation of both the amplitude and phase. It is revealed that the short-time spectral amplitude (STSA) is more important than the phase information for the quality and intelligibility of speech.

Based on the STSA estimation, the single channel enhancement technique can be divided into two classes.

- The first class attempts to estimate the short-time spectral magnitude of the speech by subtracting a noise estimate. The noise is estimated during speech pauses of the noisy speech.
- The second class applies a spectral subtraction filter (SSF) to the noisy speech, so that the spectral amplitude of enhanced speech can be obtained.

The design principle is to select appropriate parameters of the filter to minimize the difference between the enhanced speech and the clean speech. These two classes belong to the family of spectral-subtractive type algorithms. The spectral subtraction method of single channel speech enhancement is the most widely used conventional method for reducing additive noise. Many improvements are proposed to deal with the problems typically associated to spectral subtraction such as remnant broadband noise and narrow band tonal noise referred as musical noise. In this document, a simulation study of different forms of spectral subtractive-type algorithms is described. Other variants of spectral subtraction include spectral over-subtraction, multi-band spectral subtraction, Wiener filtering, iterative spectral subtraction, and spectral subtraction based on perceptual properties.

1.4 SIGNIFICANCE OF WORK

Audio and digital Signal processing has grown dramatically in recent times, in parallel with the growth of powerful and low - cost processing circuits, and the reduction in price of computer memory. This has led, in turn, to many new applications, including multimedia delivery and handheld communication devices, with the convergence of computer and telecommunications technologies.

However, many important applications of Noise suppression are as follows:

- **Speech enhancement** is required for Enhancing the speech degraded by noise and improvement of audio quality, particularly in “acoustically difficult” environments .
- **Hearing aids** for the impaired to improve the speech quality recorded by their hearing aids in noisy environments.

- **Speech recognition** provides a more natural interface to computer systems and information retrieval systems (such as telephone voice - response systems).
- **Teleconferencing systems** for communicating from remote places.
- **Biomedical applications** such as patient monitoring are indispensable in modern medical practice. Audio processing and storage continues to attract much research attention.
- **Automated systems** such as VCD which allows the users to issue voice commands to the smart systems.

1.5 ORGANIZATION OF THESIS

The remainder of this project is structured as follows:

- **Chapter 2: "Literature review"** gives a comprehensive overview of the speech enhancement system and its applications. Preliminary studies on speech enhancement techniques are presented. In addition, outlines the important design considerations that have to be included in the noisy environment which can affect the overall performance.
- **Chapter 3: "System analysis and design"** proposes a model of speech enhancement by using single channel spectral subtraction algorithm for noise suppression and its detailed overview required for the project are discussed.
- **Chapter 4: "System development and implementation"** gives an overview of the system required to implement the algorithm and basic functionalities are discussed very keenly and intently.
- **Chapter 5: "Results and discussion"** concludes the findings and outlines the suggestions for further research.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

In this chapter, a brief review of literature on speech signal characteristics, functioning of human ear with respect to sound perception, additive noise interference, adaptive signal processing algorithms and comparison of different speech enhancement algorithms is done extensively by referring different articles and wide range of research papers and a pile of books to understand the need and use of noise suppression in audio.

For many speech communication applications, such as hands-free communications, hearing aids and teleconferencing systems, speech quality and intelligibility have a direct effect on the ease and accuracy of information exchange. In acoustically adverse environments, due to the presence of different sources of disturbances such as background noise, reverberation or interference, the desired speech signal captured by the microphones is contaminated as shown in Figure 2.1, where a person (desired source) is talking in a busy conference room with different sources of noise. Consequently, this leads to a significant degradation in the quality of the picked-up speech. Figure 2.2 shows the corrupted speech signal received by the microphone array (bottom figure) and the clean speech signal (top figure). It can be clearly noticed how the corrupted speech signal is different from the clean speech, in which great portions of the speech spectra are masked and less distinct. This can significantly impair the speech quality and intelligibility. Moreover, it can increase listener’s fatigue and lower information exchange ability. Therefore, speech enhancement systems are useful to clean up the desired speech signal and mitigate the effect of the corruption in order to improve the performance of the aforementioned applications [1].

Noise suppression or speech enhancement has attracted considerable research effort in the last decades due to its uses in widely spread devices like for instance, mobile phones, hearing aids, assistive listening devices and voice communication. Particularly, hearable devices have been poised to assist people with hearing difficulties in social environments. For noise suppression and speech enhancement to work in those environments where acoustic noise becomes more intrusive, it is important to preserve weak speech components while still balance the amount of noise reduction. Accordingly, techniques that can enhance speech signals while preserving weak speech components under a large variety of acoustic scenarios are underpinning the success of different applications such as speech coding, speech recognition and hand free telephony [2, 3, 4, 5].

The type of speech enhancement or noise reduction algorithm to be selected depends on many factors that need to be considered such as, the noise type and its characteristics, applications at hand and the number of microphones available. These factors have significant impact on the quality of the estimated speech and the overall performance of the speech processing systems.

Generally speaking, it is important to consider the characteristics of the noise to achieve a speech enhancement/noise reduction method that works effectively 10in different conditions. Different types of noise based on characteristics of the noise source, such as stationary noise (pink noise) which statistically does not change over time, or non-stationary noise with high variability and similar characteristics as a speech signal (babble noise) [6]. Moreover, based on the way the noise contaminates the speech signal, the noise can be categorized into additive noise originating from different noise sources such as fans, air-conditions, traffic, babble etc, and a noise caused by multi path propagation due to the room acoustics (reverberation). Hence, according to the type of noise, different signal processing techniques can be implemented, such as noise reduction and speech dereverberation or a combination of both [7].

The goal of the speech enhancement algorithm is to remove noise and recover the original signal with as little distortion and residual noise as possible. The procedure of denoising is being complicated by the fact that, no transform domain, e.g., time, frequency, or others, exists where signal and noise have non-overlapping supports; hence aggressive removal of noise is always accompanied by signal distortion and efforts to reduce distortion conflict with the amount of reduced noise [8].

For a computationally efficient implementation, most of the speech enhancement techniques are utilizing the short time Fourier transform (STFT), where the desired signal is estimated from the degraded speech by applying noise reduction algorithm to the complex STFT coefficients [3, 5, 9, 10]. The main advantage of using STFT is the flexibility in exploiting the noise statistics to optimize the noise reduction performance since this type of decomposition helps to handle different frequencies independently. However, deploying STFT results in uniform resolution for the whole band of frequencies, which is not the case for the human auditory system with a non uniform resolution. This fact has motivated many researchers to propose alternative speech enhancement methods based on the human auditory system in order to improve the speech quality and reduce the annoying residual musical artifacts that known as musical noise [11, 12, 13, 14].

Human auditory spectrum model consists of a bank of bandpass filters which follows a spectral bark scale or so-called critical bands [14, 15]. In [14], a standard subtractive speech enhancement method is presented to eliminate the musical artifacts in very noisy situations. The masking properties of the auditory system are utilized to compute the subtraction parameter. In [16], a spectral subtraction noise reduction method was proposed using a spatial weighting technique based on the inhibitory property of the auditory system, which results in improving the estimated speech while reducing the musical noise.

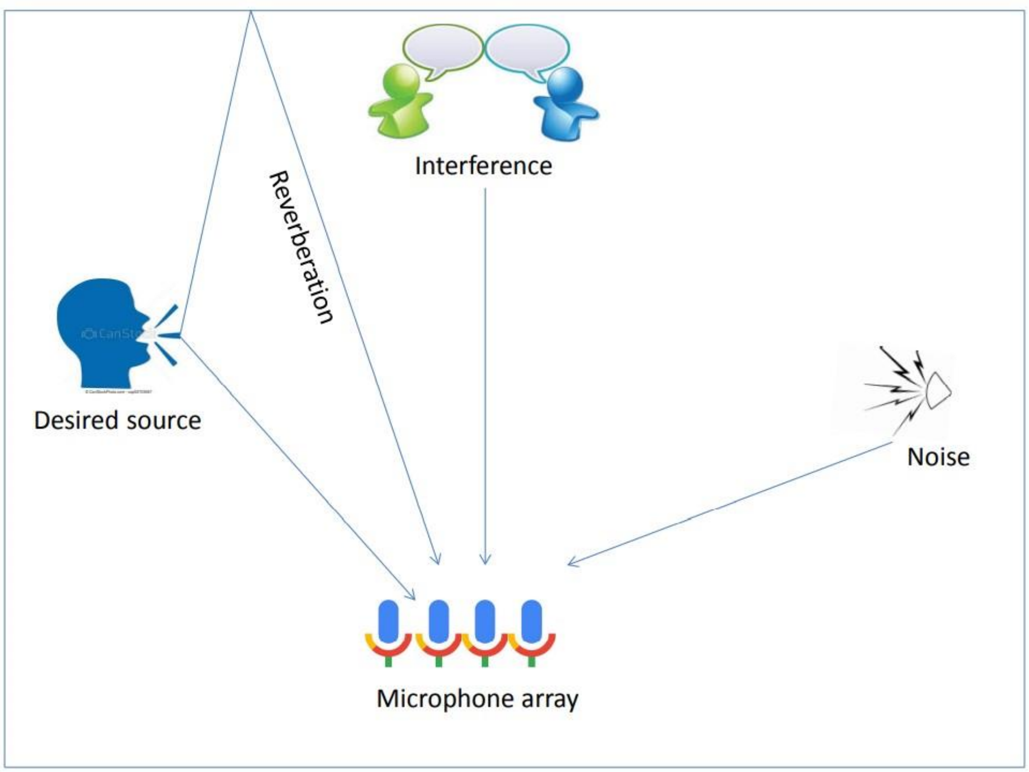

Fig 2.1: Speech degradation scenario in noisy environment.

Besides noise reduction algorithm, dereverberation algorithm in adverse environments has been discussed in this thesis. In such environment, the microphone picked up the direct path signal as well as a multiple attenuated and delayed replicas of the signal due to the room acoustics. This results in a severe degradation in the observed signal. As such for more than four decades, researchers were trying to improve the performance of speech applications in adverse environments, and recent research is interested in a robust combination of noise reduction and dereverberation technique under different noise conditions.

This chapter presents the main problem of speech communication in noisy environments and reviews recent speech enhancement techniques. First, some applications where speech enhancement is used are presented. Consequently, different speech enhancement techniques for additive noise are described for single microphone techniques. Thereafter, convolutive distortion is discussed and dereverberation processing is described using some recent efficient methods.

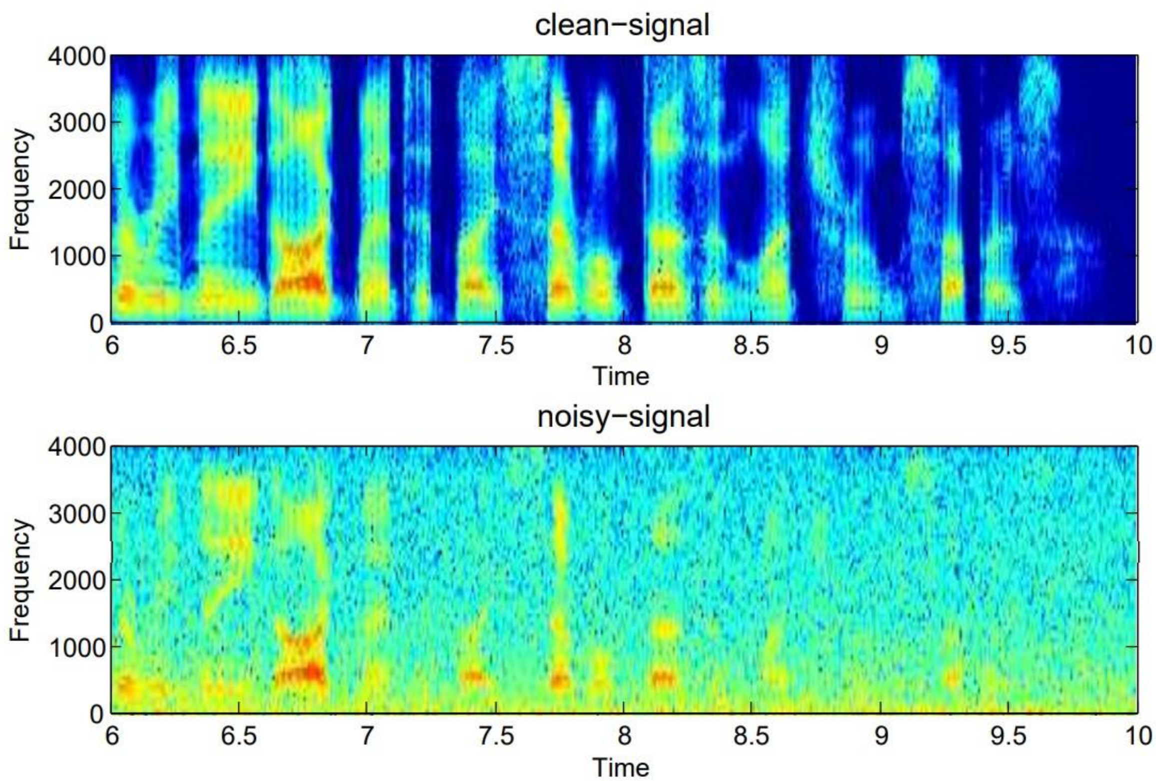

Fig 2.2: Spectrograms of clean speech signal and noisy speech signal consisting of clean speech and pink noise at 0 dB SNR.

The literature review of Spectral subtraction for enhancement of speech corrupted by additive noise involves estimating the magnitude spectrum of the noise and using it for estimating the magnitude spectrum of the speech signal. The enhanced magnitude spectrum along with the phase spectrum of the noisy speech is used to resynthesize the enhanced speech.

The following speculations are made from a study of various articles, research papers and a pile of unique books:

- **S F Boll** proposed a technique for reducing background noise by using the spectral subtraction algorithm in combination with gaussian window as an effective method for reducing background noise from the noisy speech signal. While designing this algorithm, some of the hypotheses considered are, the background noise is added to speech acoustically or electronically. The background noise condition stays constant locally to the degree that its expected value of spectral magnitude is equal to its expected value after speech activity just before speech activity. Through subtracting the noise amplitude scale from the noisy speech spectrum the technique was used to approximate the magnitude frequency spectrum of the corresponding clean speech The measurement of noise amplitude range is calculated by combining the first few frames of non-speech activity from the noisy speech signal and the secondary procedures applied to reduce the effect of background noise by half-wave rectification, residual noise reduction, and additional signal attenuation during non-speech activity.

- **M.Beroutiet** suggested a strategy for optimizing speech distorted by broadband interference by taking into account two considerations in its implementation: first, the Author excludes an α factor (over-subtraction factor) from the noise spectrum where α 1 differs from frame to frame. Second, we prohibit the filtered signal's spectral components from going below a certain minimum β (spectral floor) pre-set level. The method involves subtracting from the speech power spectrum an estimate of the noise power spectrum, setting negative discrepancies to zero, recombining the new power spectrum with the initial phase, and then reconstructing the time waveform. While this approach eliminates broadband interference, distracting "musical distortion" is usually added as well. The researcher suggested another approach to remove this "musical noise" accordingly.

- **Yariv Ephraim** and **David Malah** suggested a voice amplification algorithm corrupted by irrelevant additive noise when there is only the distorted speech signal. The basic approach is to optimally approximate (under the criteria of MMSE (Minimum Mean Square Error) and an expected mathematical method, the short-term spectral amplitude (STSA) and the complex exponential of the speech signal process. Author uses this method to measure the two elements of the short-term Fourier transformation (STFT) independently as optimally as possible rather than estimating the STFT itself optimally. The researcher has shown that the STSA and the exponential complex cannot be optimally calculated simultaneously. The researcher then uses an optimized MMSE STSA estimator and blends it with an optimal MMSE estimator of the complex phase exponential that does not impact the calculation of STSA. The researcher found that at high SNR, the estimator of MMSE STSA and the estimator of Wiener STSA resulting from the optimum estimator of MMSE STFT are almost equal.The aim of this paper is to incorporate recent research on model-based speech enhancement and provide a coherent quantitative context for specific speech enhancement issues.

- **S. D Kamath** and **P. Loizou** suggested a multi-band spectral subtraction method that takes different frequencies into account the idea that colorful noise affects the speech spectrum. Based on the frequencies the voice spectrum is separated into 'N' non-overlapping bands, and spectral subtraction is done in each band separately.

- The technical article on '**A comparison of noise reduction techniques for speech recognition in telecommunication environments**' shows how human speech carries different two types of sounds like voiced sounds and unvoiced sound. The production and characteristics of the acoustic speech signal upon which recognition systems operate is well understood and can be modeled quite accurately. paper also shows the Physiology of the human voice tract imposes articulatory constraints on the range of sounds which may be generated. the sounds may be classified according to their excitation.

- A technical article on '**Noise reduction in hearing aids; a review**' shows how speech intelligibility for people with hearing loss is affected by background noise. The problem of reducing noise is in hearing aids is one of the great importance and great difficulty. paper also describes the problem has been addressed in many different ways over the years and the techniques used range from relatively simple forms of filtering to advance signal processing methods.

S Rangachari and **philipos C Loizou** suggested a noise estimating algorithm. The noise approximation is modified by using time and frequency dependent smoothing factors to average the noisy speech power spectrum, which are aligned in individual frequency bins based on the probability of signal-presence. Signal intensity is calculated by measuring the ratio of the noisy sound power spectrum to its local average, which is continuously updated with a look-ahead parameter by comparing previous values of the noisy speech power spectrum. The local minimum algorithm for approximation adapts to extremely non-stationary noise conditions very easily.

Xiong proposed a new model for multi-stream speech enhancement that is used by multi-stream information even when some of the data streams are not directly concerned with speech waveform. The approach proposed is based on the single-channel model-based speech enhancement technique, with the exception of using a multi-stream system to distinguish the noisy speech frame. Based on the results of identification, a class-dependent filter improves the noisy speech as the traditional model-based methods of enhancement. In this way, multi-stream information does not need to be used to directly retrieve the voice waveform, but to improve the frame classier's robustness, which can outperform conventional model-based enhancement methods by extracting single data stream from noisy acoustic voice signals.

Betty Kurian suggested a PNCC (Power Normalized Cepstral Coefficients) which is a methodology for removing characteristics using nonlinearity of power-law. Time analysis is used in the PNCC processing method to estimate the degradation of the environment. It uses nonlinearity of power-law which approximates the nonlinear relation between the frequency of the signal and the rate of auditory nerve firing. It uses temporary masking to suppress noise and to support real-time online processing. Speech signal has a high spectrum of modulation and attempting to speak power varies rapidly from components of noise. Thus, the parameters characterizing environmental degradation are analyzed using a longer window of 50-120ms duration. The processing of PNCC has three levels of preliminary processing, environment compensation and final processing. Initial processing is similar to conventional processing of MFCC except that the frequency analysis is carried out using a bank of gamma tone filters. Gamma tone filters are used to design auditory filters and gamma tone weighting improves accuracy. Speech enhancement is done in the second stage. Using longer-term temporal analysis, nonlinear time-varying operations are conducted to improve robustness.

2.2 APPLICATIONS OF SPEECH ENHANCEMENT:

Prior to design speech enhancement algorithm, it is important to consider the target application. What the industry is looking for is a way to provide a clean voice signal (free of all ambient noise) to the desired application such as Automatic Speech Recognition (ASR) engine. The key concept here is extracting a clean voice signal in very noisy environments while mitigating the impact of unwanted signals such as background noise, echo, other speech sources and reverberation by using speech enhancement techniques. However, different speech communication applications have different preferences and approaches that need to be tailored to the specified application.

In many applications, distance from the speaker, noise characteristics and levels are important properties to consider not only in single channel techniques but also in multi-channel techniques.

Aviation and military applications

Speech enhancement algorithms play an important role for effective aviation and military operations. They can be used either as a standalone component to improve the speech quality or as a pre-processor in a larger speech communication system to enhance the input signal prior to further processing. Many research works have been presented in the literature that develop speech enhancement technologies. These technologies have been employed in different military and civic applications, such as for communication between pilot and civil air traffic controller (ATC) systems [10] or advanced air traffic control training system in military applications.

The ability to capture the information from the verbal messages which are used to predict the current state of the airspace or providing an early warning to avoid hazard helps to ensure passenger safety. The main problem for ATC is the presence of background noise, which can dramatically degrade the intelligibility of the verbal communications since low quality messages delivered to the ATC may have fatal effects. Thus, speech enhancement algorithm can be used as a pre-processor of the noisy speech signal before being fed to an automatic speech recognizer (ASR) in order to increase the robustness to background noise. In most of these applications, improving quality and intelligibility of the desired speech signal that has been degraded and interfered due to the harsh noisy environments are highly desirable. In [11], an advanced headphone technology was presented for speech communication application in military aircraft environment such as an air fighter cockpit. The main task of such a system is to reduce the noise in the listener's ears in very noise environments.

Biomedical applications

In industrial and heavy manufacturing workplaces such as mining, and construction sites, the workers are exposed to high level of noise for prolonged time which may result in a temporary or permanent hearing impairment. Hearing protection device (HPD) is a good solution to avoid the hearing impairment in very loud noise environments. Noise control is the most used noise reduction technique for HPD.

It can be classified into passive and active techniques. Passive noise control is used to reduce the ambient noise in noisy environment such as areas surrounding airports in order to sleep conveniently or listen to music without disturbance. This kind of protection devices is sound reduction by noise-isolating materials such as insulation, sound-absorbing tiles, or a muffler. The main drawback of such a technique is the reduction of the background noise as well as the wearer ability to hear speech. This affects the verbally communication between workers and reduce the awareness of their surroundings and any kind of hazard . In contrast, active noise control (ANC) reduces offensive (especially low frequency portions) of the noise by using cancellation techniques. ANC systems use microphones, speakers and digital signal processor. The main advantage of such devices is their ability to increase the efficiency of verbal communication and reducing the background noise while achieving a hearing protection [12].

Hence, Assistive Listening Devices (ALD) such as Hearing Aids (HA) and Choclear Implant (CI) have been widely employed to provide assistance for people with hearing loss. The main aim of ALD is to reduce the background noise while enhancing the useful signals. The efficiency of such devices is strongly dependent on the performance of the signal processing for the speech enhancement. Current research has focused on the speech enhancement algorithms in order to improve the speech quality and intelligibility of the hearing aid devices. This is achieved by contrasting speech and noise through a masking function or a gain function which localizes and preserves the speech components while attenuates the undesired noise [13]. This assist people with hearing loss to have a better verbal communication in noisy environments.

Commercial applications

Apart from the biomedical or civic and military applications, speech enhancement plays an important role to most of the commercial voice applications. One of the most popular applications of voice communication systems is the use of voice-controlled devices (VCD). It can be defined as a device that is controlled by the human voice which can be found in mobile phones, cars, internet search engine and home appliances. In the automotive industry, an increasing number of new models feature voice-activated controlled system. Hyundai Motors has been one of the major manufacturers at the forefront of bringing this new technology to the market. The VCD allows the driver to issue voice commands in order to control the mobile phone, play music, send messages, give GPS navigation addresses or coordinates all the above via the cars inbuilt microphone and without being distracted. Automatic Speech recognition (ASR) system is the main speech processing technique employed to achieve this task. For a robust ASR system in highly noisy environments, the speech enhancement technique is utilized as a pre-processing operation which helps to mitigate unwanted signals such as background noise, while preserving the desired speech signal for further processing. This improves the robustness of ASR system and enables the human-machine communication.

Another important commercial application for speech enhancement system is the teleconferencing application. The ultimate intention of any conference system is to facilitate communication between remote participants with high speech quality and low latency technique. High quality teleconferencing also saves the environment by providing the means to have effective remote meetings without the need to meet face to face. Speech enhancement is an enabling technology for this application. Depending on the size of the conference room, or type of the activities such as, formal presentation or distance learning, different speech enhancement techniques with 16different number of microphones and configuration need to be used [14].

2.3 MOTIVATION

Speech enhancement aims to improve speech quality by using various algorithms. The objective of enhancement is improvement in intelligibility and/or overall perceptual quality of degraded speech signal using audio signal processing techniques.

Enhancing of speech degraded by noise, or noise reduction, is the most important field of speech enhancement, and used for many applications such as mobile phones, VoIP, teleconferencing systems, speech recognition, and hearing aids .

The speech enhancement technique should have low algorithmic delay. Its computational complexity should be low to permit its implementation on a low-power processor. Spectral subtraction is a single-input speech enhancement technique developed for use in audio codecs and speech recognition. It involves estimating the noise spectrum, subtracting it from the noisy speech spectrum, and resynthesizing the enhanced speech signal.

2.4 CONCLUSION

In this chapter, we have discussed the overall literature survey of the methods available for speech enhancement and posed an insight of basic single channel speech enhancement technique as the state of the art. The technique for speech enhancement which is discussed here is the most popular technique used in the hearing aids and mobile phones. The main advantages of single channel speech enhancement techniques consist of being simple, easy to implement in hardware and cost effective in practice. A detailed framework of single channel speech enhancement is presented to discuss the design decision considerations that have to be taken into account in the design procedure in order to preserve the speech components while reducing the musical noise.

CHAPTER 3

SYSTEM ANALYSIS AND DESIGN

Speech is the most sophisticated signal naturally produced by humans. A speech signal carries linguistic information for the sharing of information and ideas. It allows people to express emotions and verbally share feelings. It is the most fundamental form of communication among humans. The aim of digital speech processing is to take advantage of digital computing techniques to process the speech signal for increased understanding, improved communication, and increased efficiency and productivity associated with speech activities.

The aim of a speech enhancement system is to suppress the noise in a noisy speech signal. For robust speech recognition, such a system is used as a pre-processor to a speech recognizer. Since it produces a clean speech signal, no changes in the recognition system are necessary to make it robust.

3.1 SPEECH ENHANCEMENT TECHNIQUES

The speech enhancement system based on the number of microphones can be classified into single and multi-channel speech enhancement techniques. Multichannel techniques can provide an improved dereverberation, strong noise suppression and interference rejection as compared to the single channel techniques. Single channel speech enhancement technique, however, is still useful because of its simple implementation. Accordingly, we provide an overview of single and multi-channel speech enhancement techniques. Hence speech enhancement techniques are basically divided into two main types.

1. **Single channel space enhancement technique:** In real-time applications like mobile communications and hearing aids, single channel speech enhancement is usually preferred. The main formulation of such technique is to estimate the speech signal that is degraded by uncorrelated additive noise. It often consists of one microphone used to estimate the clean speech using the temporal and spectral information of the degraded speech signal [15].

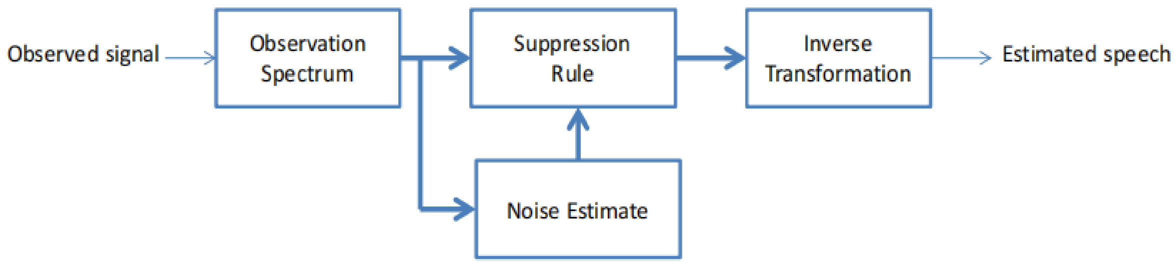

Fig 3.1 : Framework of single channel speech enhancement.

2. **Multi-channel space enhancement technique:** multi-channel speech enhancement technique consists of microphone array with elements located at diverse spatial positions. The main advantage of using a microphone array is its ability to exploit the spatial information of the received signal in addition to the temporal and spectral information. Since the speech signal and noise are located in different positions in the room, the desired signal can be spatially separated from the noise. This provides extra information about the desired signal characteristics and the noise properties [16].

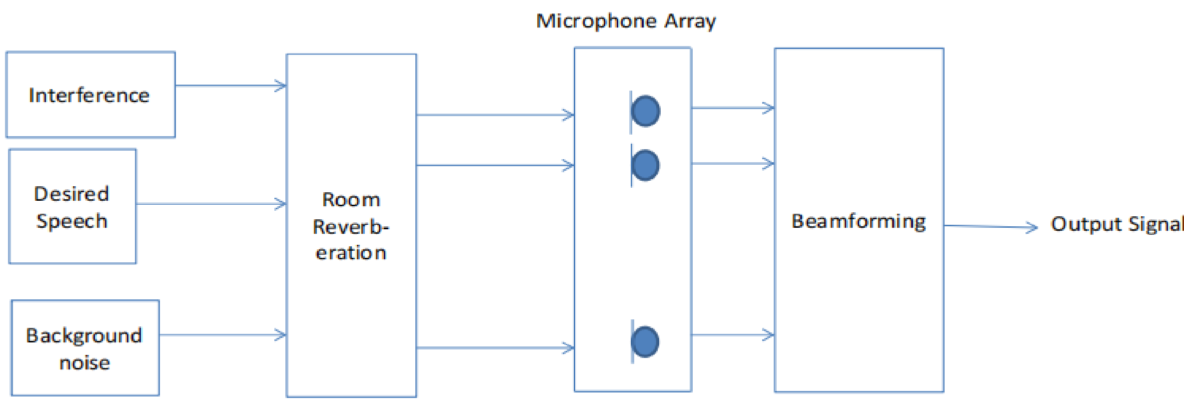

Fig 3.2 : Framework of multiple channel speech enhancement.

3.2 SPEECH ENHANCEMENT ALGORITHMS

The algorithms of speech enhancement for noise reduction can be categorized into three fundamental classes: filtering techniques, spectral restoration, and model-based methods [17].

- **Filtering :**

  1. **Spectral Subtraction Method:** Spectral subtraction is a method for restoration of the power, or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum.

  2. **Wiener Filtering Method:** The Wiener filter can be used to filter out the noise from the corrupted signal to provide an estimate of the underlying signal of interest. The Wiener filter is based on a statistical approach, and a more statistical account of the theory is given in the minimum mean square error (MMSE) estimator article.

  3. **Signal Subspace Approach:** Subspace algorithms decompose the vector space of the noisy signal into noise subspace having noise signals and speech subspace having speech signals. SSA provides dimensionality reduction and a better compromise between signal distortion and residual noise over other speech enhancement methods.

- **Spectral Restoration:**

  1. **Minimum Mean-Square-Error Short-Time Spectral Amplitude Estimator:** A system which utilizes a minimum mean-square error (MMSE) STSA estimator is proposed and then compared with other widely used systems which are based on Wiener filtering and the "spectral subtraction" algorithm.

- **Speech Model Based:** The idea of model-based speech enhancement is first to detect those time-frequency areas that seem to be appropriate for reconstruction. In order to achieve a successful reconstruction, it is necessary that at least a few time-frequency areas have a sufficiently high SNR. These signal parts are then used to reconstruct those parts with lower SNR. For reconstruction, several speech signal properties such as pitch frequency or the degree of voicing need to be estimated in a reliable manner.

3.3 SPECTRAL SUBTRACTION ALGORITHM

Spectral subtraction is a method for restoration of the power spectrum, or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum. The noise spectrum is usually estimated, and updated, from the periods when the signal is absent and only the noise is present. The assumption is that the noise is a stationary or a slowly varying process, and that the noise spectrum does not change significantly in between the update periods.

For restoration of time-domain signals, an estimate of the instantaneous magnitude spectrum is combined with the phase of the noisy signal, and then transformed via an inverse discrete Fourier transform to the time domain. In terms of computational complexity, spectral subtraction is relatively inexpensive. However, owing to random variations of noise, spectral subtraction can result in negative estimates of the short-time magnitude or power spectrum. The magnitude and power spectrum are non-negative variables, and any negative estimates of these variables should be mapped into non-negative values. This nonlinear rectification process distorts the distribution of the restored signal. The processing distortion becomes more noticeable as the signal-to-noise ratio decreases. In this chapter, we study spectral subtraction, and the different methods of reducing and removing the processing distortions [18].

3.3.1 SPECTRAL SUBTRACTION

In applications where, in addition to the noisy signal, the noise is accessible on a separate channel, it may be possible to retrieve the signal by subtracting an estimate of the noise from the noisy signal. For example, the adaptive noise canceller takes as the inputs the noise and the noisy signal, and outputs an estimate of the clean signal. However, in many applications, such as at the receiver of a noisy communication channel, the only signal that is available is the noisy signal. In these situations, it is not possible to cancel out the random noise, but it may be possible to reduce the average effects of the noise on the signal spectrum. The effect of additive noise on the magnitude spectrum of a signal is to increase the mean and the variance of the spectrum as illustrated in Figur. The increase in the variance of the signal spectrum results from the random fluctuations of the noise, and cannot be cancelled out [19]. The increase in the mean of the signal spectrum can be removed by subtraction of an estimate of the mean of the noise spectrum from the noisy signal spectrum.

The noisy signal model in the time domain is given by

$$y(m) = x(m) + n(m)$$

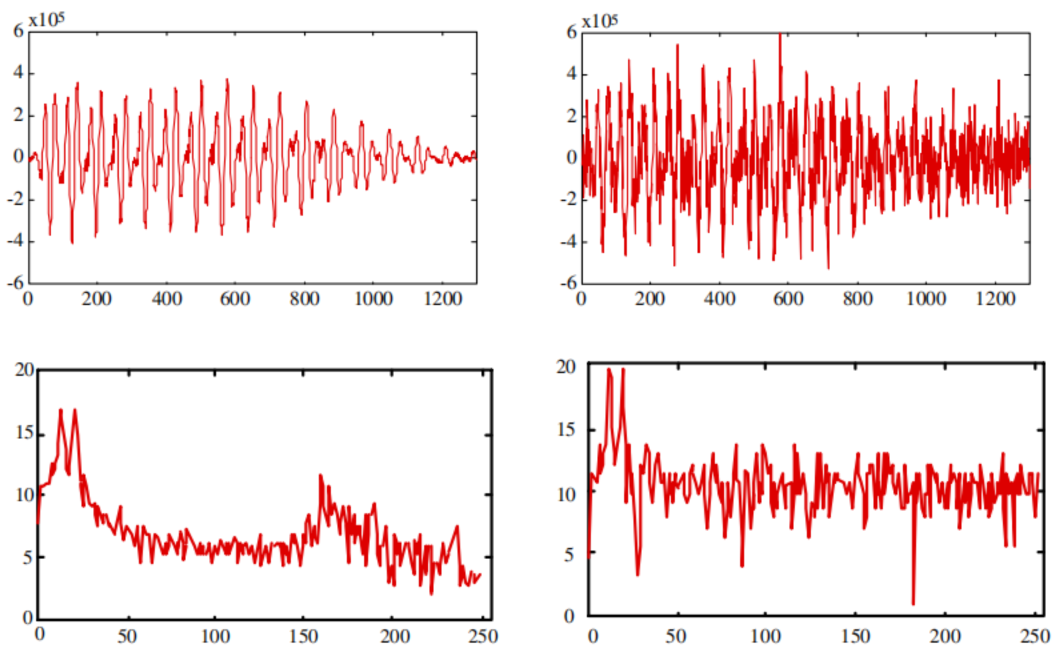

Fig 3.3 : Effect of noise on a signal in the time and the frequency domains.

Where y(m), x(m) and n(m) are the signal, the additive noise and the noisy signal respectively, and m is the discrete time index. In the frequency domain, the noisy signal model is expressed as

$$Y(f) = X(f) + N(f)$$

where Y(f), X(f) and N(f) are the Fourier transforms of the noisy signal y(m), the original signal x(m) and the noise n(m) respectively, and f is the frequency variable. In spectral subtraction, the incoming signal x(m) is buffered and divided into segments of N samples length. Each segment is windowed, using a Hanning or a Hamming window, and then transformed via discrete Fourier transform (DFT) to N spectral samples. The windows alleviate the effects of the discontinuities at the endpoints of each segment. The windowed signal is given by

$$\begin{aligned} y_w(m) &= w(m)y(m) \\ &= w(m)[x(m)+n(m)] \\ &= x_w(m)+n_w(m) \end{aligned}$$

The windowing operation can be expressed in the frequency domain as

$$\begin{aligned} Y_w(f) &= W(f) * Y(f) \\ &= X_w(f)+N_w(f) \end{aligned}$$

where the operator * denotes convolution. Throughout this chapter, it is assumed that the signals are windowed, and hence for simplicity we drop the use of the subscript w for windowed signals. Figure illustrates a block diagram configuration of the spectral subtraction method.. The equation describing spectral subtraction may be expressed as

$$\left|\hat{X}(f)\right|^b = \left|Y(f)\right|^b - \alpha\overline{\left|N(f)\right|^b}$$

where $|\hat{X}(f)|^b$ is an estimate of the original signal spectrum $|X(f)|^b$ and $\overline{|N(f)|^b}$ is the time-averaged noise spectra. It is assumed that the noise is a wide-sense stationary random process. For magnitude spectral subtraction, the exponent b=1, and for power spectral subtraction, b=2. The parameter α controls the amount of noise subtracted from the noisy signal. For full noise subtraction, α=1 and for over-subtraction α>1. The time-averaged noise spectrum is obtained from the periods when the signal is absent and only the noise is present as

$$\overline{|N(f)|^b} = \frac{1}{K}\sum_{i=0}^{K-1} |N_i(f)|^b$$

$|N_i(f)|$ is the spectrum of the ith noise frame, and it is assumed that there are K frames in a noise-only period, where K is a variable. Alternatively, the averaged noise spectrum can be obtained as the output of a first order digital low-pass filter as

$$\overline{|N_i(f)|^b} = \rho \overline{|N_{i-1}(f)|^b} + (1-\rho)|N_i(f)|^b$$

where the low-pass filter coefficient ρ is typically set between 0.85 and 0.99. For restoration of a time-domain signal, the magnitude spectrum estimate is combined with the phase of the noisy signal, and then transformed into the time domain via the inverse discrete Fourier transform as

$$\hat{x}(m) = \sum_{k=0}^{N-1} |\hat{X}(k)| e^{j\theta_Y(k)} e^{-j\frac{2\pi}{N}km}$$

Where $\theta_Y(k)$ is the phase of the noisy signal frequency Y(k). The signal restoration equation is based on the assumption that the audible noise is mainly due to the distortion of the magnitude spectrum, and that the phase distortion is largely inaudible. Evaluations of the perceptual effects of simulated phase distortions validate this assumption.

Owing to the variations of the noise spectrum, spectral subtraction may result in negative estimates of the power or the magnitude spectrum. This outcome is more probable as the signal-to-noise ratio (SNR) decreases. To avoid negative magnitude estimates the spectral subtraction output is post processed using a mapping function T[·] of the form

$$T[|\hat{X}(f)|] = \begin{cases} |\hat{X}(f)| & if\ |\hat{X}(f)| > \beta |Y(f)| \\ \text{fn}[|Y(f)|] & \text{otherwise} \end{cases}$$

For example, we may choose a rule such that if the estimate $|\hat{X}(f)| > 0.01|Y(f)|$ (in magnitude spectrum 0.01 is equivalent to –40 dB) then the noise estimate should be set to some function of the noisy signal fn[Y(f)]. In its simplest form, fn[Y(f)] =noise floor, where the noise floor is a positive constant. An alternative is fn[|Y(f)|] =β |Y(f)|. In this case,

$$T[|\hat{X}(f)|]=\begin{cases}|\hat{X}(f)| & \text{if } |\hat{X}(f)|>\beta\,|Y(f)| \\ \beta|Y(f)| & \text{otherwise}\end{cases}$$

Spectral subtraction may be implemented in the power or the magnitude spectral domains. The two methods are similar, although theoretically they result in somewhat different expected performance.

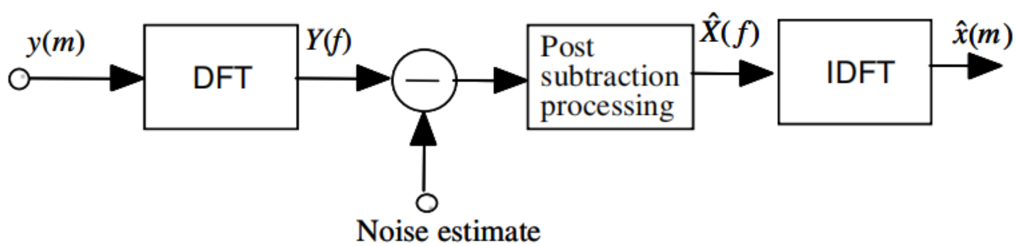

Fig 3.4: A block diagram illustration of spectral subtraction.

3.3.2 POWER SPECTRUM SUBTRACTION

The power spectrum subtraction, or squared-magnitude spectrum subtraction, is defined by the following equation:

$$|\hat{X}(f)|^2=|Y(f)|^2-\overline{|N(f)|^2}$$

where it is assumed that α, the subtraction factor. We denote the power spectrum by E{ $|X(f)|^2$ } the time-averaged power spectrum by $|\hat{X}(f)|^2$ and the instantaneous power spectrum by $|X(f)|^2$. .By expanding the instantaneous power spectrum of the noisy signal $|Y(f)|^2$ and grouping the appropriate terms, equation may be rewritten as

$$|\hat{X}(f)|^2 = |X(f)|^2 + \underbrace{\left(|N(f)|^2 - \overline{|N(f)|^2}\right)}_{\text{Noise variations}} + \underbrace{X^*(f)N(f) + X(f)N^*(f)}_{\text{Cross products}}$$

Taking the expectations of both sides and assuming that the signal and the noise are uncorrelated ergodic processes, we have

$$\mathcal{E}[|\hat{X}(f)|^2] = \mathcal{E}[|X(f)|^2]$$

The average of the estimate of the instantaneous power spectrum converges to the power spectrum of the noise-free signal. However, it must be noted that for non-stationary signals, such as speech, the objective is to recover the instantaneous or the short-time spectrum, and only a relatively small amount of averaging can be applied. Too much averaging will smear and obscure the temporal evolution of the spectral events [19].

3.3.3 MAGNITUDE SPECTRUM SUBTRACTION

The magnitude spectrum subtraction is defined as

$$|\hat{X}(f)| = |Y(f)| - \overline{|N(f)|}$$

where $\overline{|N(f)|}$ is the time-averaged magnitude spectrum of the noise. Taking the expectation, we have

$$\begin{aligned}
\mathcal{E}[|\hat{X}(f)|] &= \mathcal{E}[|Y(f)|] - \mathcal{E}[\overline{|N(f)|}] \\
&= \mathcal{E}[|X(f) + N(f)|] - \mathcal{E}[\overline{|N(f)|}] \\
&\approx \mathcal{E}[|X(f)|]
\end{aligned}$$

For signal restoration the magnitude estimate is combined with the phase of the noisy signal and then transformed into the time domain [19].

3.3.4 SPECTRAL SUBTRACTION FILTER

The spectral subtraction equation can be expressed as the product of the noisy signal spectrum and the frequency response of a spectral subtraction filter as

$$\begin{aligned} |\hat{X}(f)|^2 &= |Y(f)|^2 - \overline{|N(f)|^2} \\ &= H(f)|Y(f)|^2 \end{aligned}$$

where H(f), the frequency response of the spectral subtraction filter, is defined as

$$\begin{aligned} H(f) &= 1 - \frac{\overline{|N(f)|^2}}{|Y(f)|^2} \\ &= \frac{|Y(f)|^2 - \overline{|N(f)|^2}}{|Y(f)|^2} \end{aligned}$$

The spectral subtraction filter H(f) is a zero-phase filter, with its magnitude response in the range $0 \geq H(f) \geq 1$. The filter acts as a SNR-dependent attenuator. The attenuation at each frequency increases with the decreasing SNR, and conversely decreases with the increasing SNR. The least mean square error linear filter for noise removal is the Wiener filter covered in chapter 6. Implementation of a Wiener filter requires the power spectra (or equivalently the correlation functions) of the signal and the noise process. Spectral subtraction is used as a substitute for the Wiener filter when the signal power spectrum is not available. In this section, we discuss the close relation between the Wiener filter and spectral subtraction. For restoration of a signal observed in uncorrelated additive noise, the equation describing the frequency response of the Wiener filter is given as

$$W(f) = \frac{\mathcal{E}[|Y(f)|^2] - \mathcal{E}[|N(f)|^2]}{\mathcal{E}[|Y(f)|^2]}$$

A comparison of W(f) and H(f) and the above equations shows that the Wiener filter is based on the ensemble-average spectra of the signal and the noise, whereas the spectral subtraction filter uses the instantaneous spectra of the noisy signal and the time-averaged spectra of the noise. In spectral subtraction, we only have access to a single realisation of the process. However, assuming that the signal and noise are wide-sense stationary ergodic processes, we may replace the instantaneous noisy signal spectrum in the spectral subtraction equation with the time averaged spectrum $\overline{|Y(f)|^2}$, to obtain

$$H(f) = \frac{\overline{|Y(f)|^2} - \overline{|N(f)|^2}}{\overline{|Y(f)|^2}}$$

For an ergodic process, as the length of the time over which the signals are averaged increases, the time-averaged spectrum approaches the ensemble averaged spectrum, and in the limit, the spectral subtraction filter of equation approaches the Wiener filter equation. In practice, many signals, such as speech and music, are non-stationary, and only a limited degree of beneficial time-averaging of the spectral parameters can be expected [19].

3.3.5 NON-LINEAR SPECTRAL SUBTRACTION

The use of spectral subtraction in its basic form of Equation may cause deterioration in the quality and the information content of a signal. For example, in audio signal restoration, the musical noise can cause degradation in the perceived quality of the signal, and in speech recognition the basic spectral subtraction can result in deterioration of the recognition accuracy. In the literature, there are a number of variants of spectral subtraction that aim to provide consistent performance improvement across a range of SNRs. These methods differ in their approach to estimation of the noise spectrum, in their method of averaging the noisy signal spectrum, and in their post processing method for the removal of processing distortions. Non-linear spectral subtraction methods are heuristic methods that utilise estimates of the local SNR, and the observation that at a low SNR over-subtraction can produce improved results. For an explanation of the improvement that can result from over-subtraction, consider the following expression of the basic spectral subtraction equation:

$$\begin{aligned} |\hat{X}(f)| &= |Y(f)| - \overline{|N(f)|} \\ &\approx |X(f)| + |N(f)| - \overline{|N(f)|} \\ &\approx |X(f)| + V_N(f) \end{aligned}$$

where $V_N(f)$ is the zero-mean random component of the noise spectrum. If $V_N(f)$ is well above the signal X(f) then the signal may be considered as lost to noise. In this case, over-subtraction, followed by non-linear processing of the negative estimates, results in a higher overall attenuation of the noise. This argument explains why subtracting more than the noise average can sometimes produce better results. The non-linear variants of spectral subtraction may be described by the following equation:

$$|\hat{X}(f)| = Y(f)| - \alpha\big(SNR(f)\big)\overline{|N(f)|}_{NL}$$

where α(SNR(f)) is an SNR-dependent subtraction factor and $\overline{|N(f)|_{NL}}$ is a non-linear estimate of the noise spectrum. The spectral estimate is further processed to avoid negative estimates as Spectral Subtraction

$$|\hat{X}(f)| = \begin{cases} |\hat{X}(f)| & \text{if } \ |\hat{X}(f)| > |\beta Y(f)| \\ |\beta Y(f)| & \text{otherwise} \end{cases}$$

One form of an SNR-dependent subtraction factor for Equation is given by

$$\alpha\big(SNR(f)\big) = 1 + \frac{sd\big(|N(f)|\big)}{\overline{|N(f)|}}$$

where the function sd(|N(f)| is the standard deviation of the noise at frequency f. For white noise, *sd*(|N(f)|=σ_n, where σ_n^2 is the noise variance. Substitution of Equation in Equation yields

$$|\hat{X}(f)| = |Y(f)| - \left[1 + \frac{sd\left(|N(f)|\right)}{|N(f)|}\right] \overline{|N(f)|}$$

In Equation (11.27) the subtraction factor depends on the mean and the variance of the noise. Note that the amount over-subtracted is the standard deviation of the noise. This heuristic formula is appealing because at one extreme for deterministic noise with a zero variance, such as a sine wave, α(SNR(f)) =1, and at the other extreme for white noise α(SNR(f))=2. In application of spectral subtraction to speech recognition, it is found that the best subtraction factor is usually between 1 and 2.

3.3.6 VUVUZELA DENOISING ALGORITHM

The Vuvuzela denoising algorithm follows spectral subtraction method i.e. all the above mentioned steps in a more efficient manner in order to reduce the noise level of a speech signal and enhance it. Hence Vuvuzela denoising algorithm is used in this project.

The **vuvuzelas** have the potential to cause hearing loss. Vuvuzelas can have a negative effect when a listener's eardrums are exposed to the instrument's high-intensity sound. The Vuvuzela is like a straightened trumpet and is played by blowing a raspberry into the mouthpiece. The player's lips open and close about 235 times second, sending puffs of air down the tube, which excite resonance of the air in the conical bore. A single Vuvuzela played by a decent trumpeter is reminiscent of a hunting horn – but the sound is less pleasing when played by the average football fan, as the note is imperfect and fluctuates in frequency. It sounds more like an trumpeting. This happens because the player does not keep the airflow and motion of the lips consistent. A flared instrument has louder higher-frequency harmonics than a cylindrical one. The flared instrument is perceived as louder because the higher harmonics are at frequencies where our hearing is most sensitive. This is partly why the conical saxophone sounds louder than the cylindrical clarinet. Since it produces 116 decibels at 1 metre. A whole crowd produces even higher levels, and measurements at a training match have shown temporary hearing loss among spectators. Experiments on other noise sources show that louder sounds are more annoying. Our hearing is an early-warning system: we listen out for sudden changes in the sounds around us which might indicate threats, and ignore benign, persistent noise.

The Vuvuzela Denoising algorithm is given by the following flowchart. The vuvuzela Denoising algorithm is done by loading the .wav file and then estimating the noise is carried out by taking the short time Fourier transform, the windowing function used here is hamming window, and then noise is estimated and inverse short time Fourier transform is performed, and then the signal is retained by using overlap and add method [21].

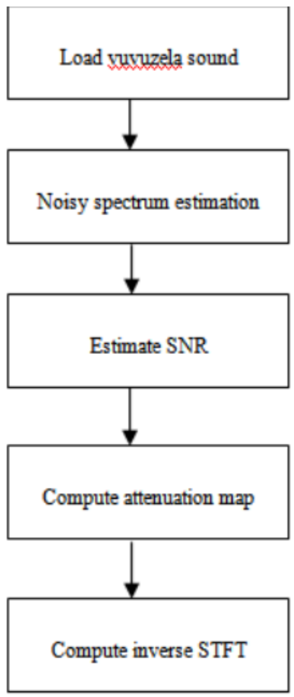

Fig 3.7 : Flowchart of the Vuvuzela Denoising algorithm.

The output of this algorithm is given by the following figures which are as follows. The picture describes the original input signal which is given when the Vuvuzela sound is introduced. This figure consists of speech and the noise signal which is differentiated by various colours. It also describes the spectrogram which corresponds to both speech and the noise. Here the signal is accompanied along with the noise. This noise is reduced by using the Vuvuzela denoising method.

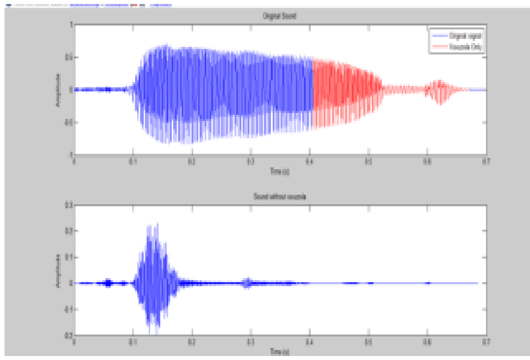

Fig 3.8 : Original sound and sound without vuvuzela.

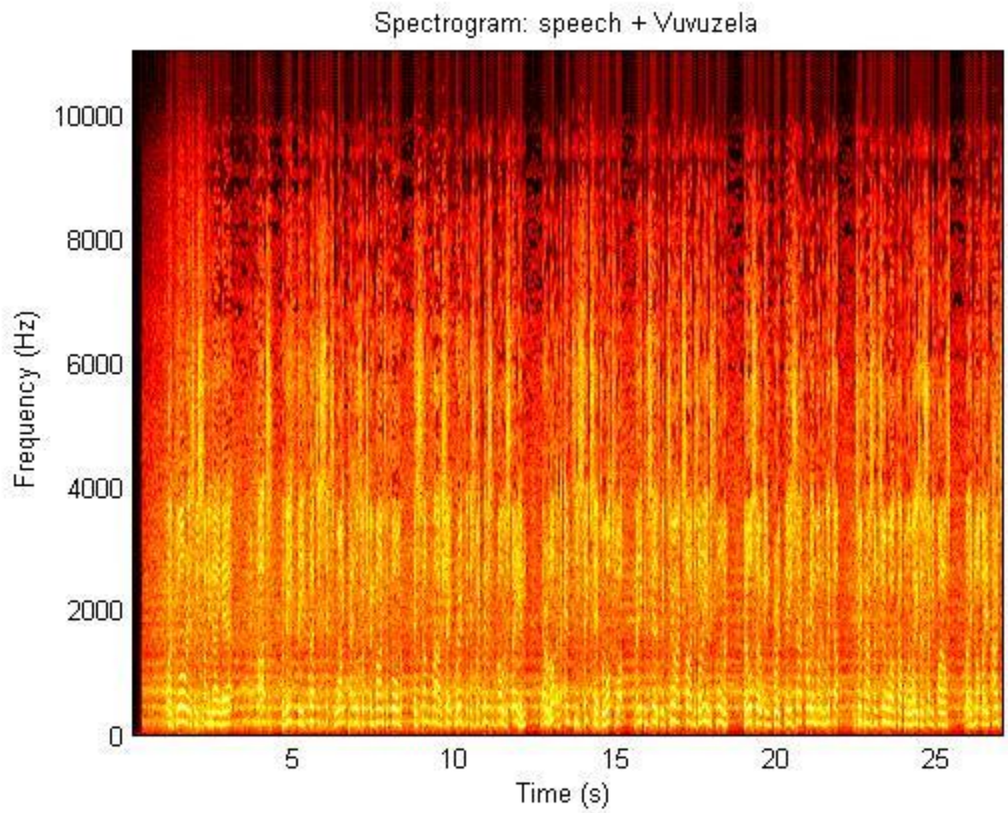

Fig 3.9 : Spectrogram of speech+vuvuzela signal.

The figure shows the signal without noise and this is achieved by following the steps of the algorithm which is given in the flowchart..

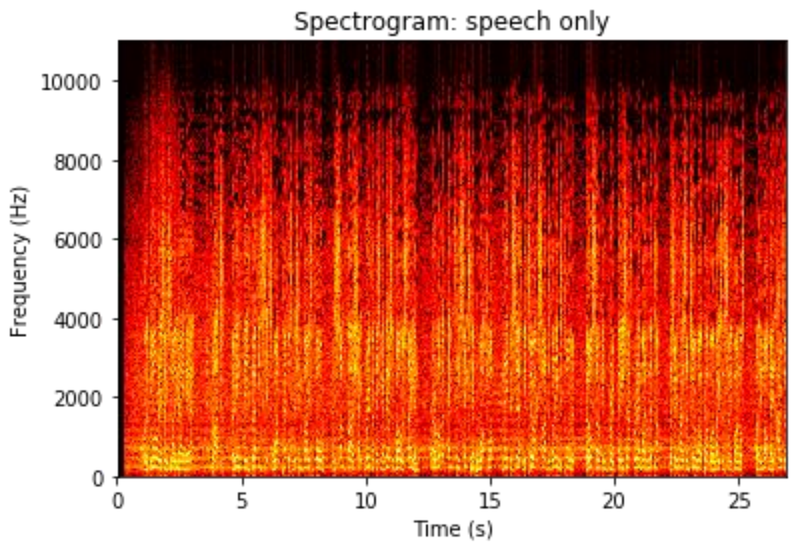

Fig 3.10 : Spectrogram of speech only signal.

For effective noise reduction, this algorithm takes in account to perceptual aspects of human ear. It can be seen from the experimental results that it effectively reduces background noise in comparison with commonly used other types of algorithms. This method results in greater improvement of noise reduction and considerably improvement of perceptual speech quality in compassion to conventional method [21,22].

3.4 CONCLUSION

In this chapter, we have discussed about the core component of our project which is the spectral subtraction algorithm and its dependencies. The noise can be clearly removed to the maximum possible extent using this algorithm. The parameters required for better functioning of the algorithm are mentioned and the variation of these parameters could provide better performance in terms of speech quality and intelligibility.

CHAPTER 4

SYSTEM DEVELOPMENT AND IMPLEMENTATION

TOOLS USED

MATLAB

4.1 MATLAB

4.1.1 WHAT IS MATLAB?

MATLAB (an abbreviation of “matrix laboratory”) is a proprietary multi-paradigm programming language and numeric computing environment developed by MathWorks (by the LINPACK & EISPACK). MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages.

It is a programming platform designed specifically for engineers and scientists. The heart of MATLAB is the MATLAB language, a matrix-based language allowing the most natural expression of computational mathematics.

As of 2020, MATLAB has more than 4 million users worldwide. MATLAB users come from various backgrounds of engineering, science, and economics MATLAB is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation.

Fig 4.1 – MATLAB logo

Typical uses include:

- Math and computation
- Algorithm development
- Modelling, simulation, and prototyping
- Data analysis, exploration, and visualization
- Scientific and engineering graphics
- Application development, including GUI build.

MATLAB is an interactive system whose basic data element is an array that does not require dimensioning. This allows you to solve many technical computing problems, especially those with matrix and vector formulations, in a fraction of the time it would take to write a program in a scalar noninteractive language such as C or Fortran [23].

MATLAB features a family of application-specific solutions called toolboxes. Very important to most users of MATLAB, toolboxes allow you to *learn* and *apply* specialized technology. Toolboxes are comprehensive collections of MATLAB functions (M-files) that extend the MATLAB environment to solve particular classes of problems. Areas in which toolboxes are available include signal processing, control systems, neural networks, fuzzy logic, wavelets, simulation, and many others [23]. Millions of engineers and scientists in industry and academia use MATLAB. You can use MATLAB for a range of applications, including deep learning and machine learning, signal processing and communications, image and video processing, control systems, test and measurement, computational finance, and computational biology.

MATLAB features a family of application-specific solutions called toolboxes. Very important to most users of MATLAB, toolboxes allow you to *learn* and *apply* specialized technology. Toolboxes are comprehensive collections of MATLAB functions (M-files) that extend the MATLAB environment to solve particular classes of problems. Areas in which toolboxes are available include signal processing, control systems and others.

4.1.2 IMAGE PROCESSING: -

In computer science, **digital image processing** is the use of computer algorithms to perform image processing on digital images. As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distortion during processing. Since images are defined over two dimensions (perhaps more) digital image processing may be modelled in the form of multidimensional systems.

Digital image processing allows the use of much more complex algorithms, and hence, can offer both more sophisticated performance at simple tasks, and the implementation of methods which would be impossible by analog means.

In particular, digital image processing is the only practical technology for:

1. Classification
2. Feature extraction
3. Multi-scale signal analysis
4. Pattern recognition
5. Projection

4.1.3 UNDERSTANDING THE IMAGE

It is defined as a two dimensional function **f(x,y)**, where x and y are spatial coordinates. Amplitude of **f** at any pair of coordinates **(x,y)** is called intensity or gray level of image at that point. When **(x,y)** and amplitudes values of f are all finite , discrete quantities , image is said to be digital image.

Hence digital image processing refers to processing digital images by means of a digital computer.

A digital image is composed of a finite number of elements are referred to as picture elements, pels and pixels. Pixel is the most widely used term to denote the elements of an image.

Moreover, digital image processing encompasses processes whose inputs and outputs are images and in addition, encompasses processes hat extract attributes from images up to and including the recognition of individual objects.

4.1.4 AUDIO PROCESSING

Audio signal processing is a subfield of signal processing that is concerned with the electronic manipulation of audio signals. Audio signals are electronic representations of sound waves-longitudinal waves which travel through air, consisting of compressions and rarefactions. The energy contained in audio signals is typically measured in decibels. As audio signals may be represented in either digital or analog format, processing may occur in either domain. Analog processors operate directly on the electrical signal, while digital processors operate mathematically on its digital representation.

- An analog audio signal is a continuous signal represented by an electrical voltage or current that is “analogous” to the sound waves in the air. Analog signal processing then involves physically altering the continuous signal by changing the voltage or current or charge via electrical circuits.
- Historically, before the advent of widespread digital technology, analog was the only method by which to manipulate a signal. Since that time, as computers and software have become more capable and affordable, digital signal processing has become the method of choice. However, in music applications, analog technology is often still desirable as it often produces nonlinear responses that are difficult to replicate with digital filters [24].
- Audio signal processing is used when broadcasting audio signals in order to enhance their fidelity or optimize for bandwidth or latency. In this domain, the most important audio processing takes place just before the transmitter. The audio processor here must prevent or minimize overmodulation, compensate for non-linear transmitters (a potential issue with medium wave and shortwave broadcasting), and adjust overall loudness to desired level.

4.2 THE MATLAB SYSTEM:

The MATLAB system consists of five main parts:

The MATLAB language:

This is a high-level matrix/array language with control flow statements, functions, data structures, input/output, and object-oriented programming features. It allows both "programming in the small" to rapidly create quick and dirty throw-away programs, and "programming in the large" to create complete large and complex application programs.

The MATLAB working environment:

This is the set of tools and facilities that you work with as the MATLAB user or programmer. It includes facilities for managing the variables in your workspace and importing and exporting data. It also includes tools for developing, managing, debugging, and profiling M-files, MATLAB's applications.

Handle Graphics:

This is the MATLAB graphics system. It includes high-level commands for two-dimensional and three-dimensional data visualization, image processing, animation, and presentation graphics. It also includes low-level commands that allow you to fully customize the appearance of graphics as well as to build complete Graphical User Interfaces on your MATLAB applications.

The MATLAB mathematical function library:

This is a vast collection of computational algorithms ranging from elementary functions like sum, sine, cosine, and complex arithmetic, to more sophisticated functions like matrix inverse, matrix eigenvalues, Bessel functions, and fast Fourier transforms.

The MATLAB Application Program Interface (API):

This is a library that allows you to write C and Fortran programs that interact with MATLAB. It includes facilities for calling routines from MATLAB (dynamic linking), calling MATLAB as a computational engine, and for reading and writing MAT-files [23].

4.3 MATLAB Installation

Step 1: Double click on the MATLAB icon (the binary file which we need to download from official MATLAB website https://in.mathworks.com/downloads). After clicking the icon, a pop-up will ask for the installer to run, click on the *Run*. A MathWorks Installer window will pop-up on the screen [25].

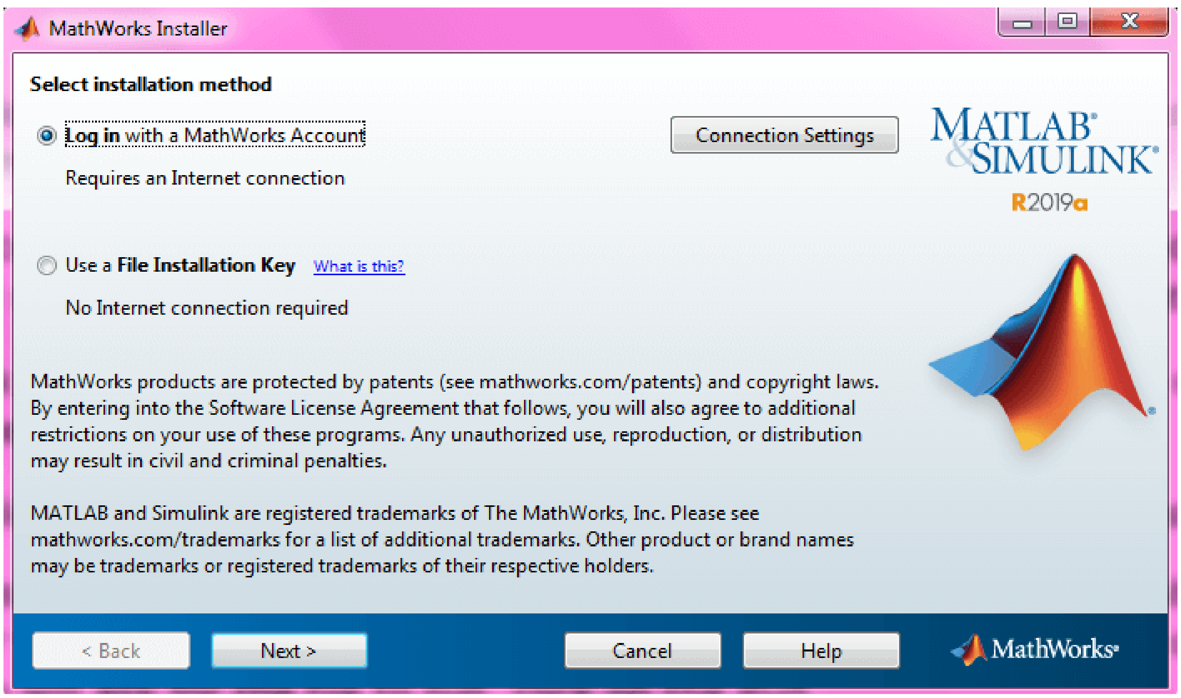

Fig 4.2 – MATLAB installation prompt in windows

- Click on **Log in with a MathWorks Account** and click next.

Step 2: A license Selection window will appear, a preselected license id will be highlighted with a blue background. Here you have to select your license id; this is the id which we have saved during STEP 9 of downloading of the installer (we urged to note down that id during that time) and again click on Next.

- A new Folder Selection window appears, no need to change the folder location for installation of MATLAB, click on Next.

Fig 4.3 – Choose installation folder prompt

Step 3: Next is Product Selection window, the first product is MATLAB 9.6, this is mandatory to select because it is the MATLAB environment, and from other products, you can choose as many of your choices and click on Next.

- Next is the *Installation Options* window, select options as per your choice. Any time you feel something to change, you can go back to the previous step by clicking on the Back button.
- Next is the Confirmation window, here you no need to do anything, confirm what you are going to download in the process of the installation of MATLAB, its other Add-on products, and what is the size of the downloads; and click on Install.
- By clicking on Install, downloading of all the products will be started. It's a massive download, so you have to wait for some time to complete the download.

Step 4: After downloading of all products and completion of the installation, a window appears that says to Activate the MATLAB, no need to do anything, click on the Next button.

- After clicking on Next, a new window appears that says about what is the meaning of activation. Proceed by clicking on *Next*.

- Again a new window appears displaying your email id and your products' license id, proceed by clicking on *Confirm button*.
- Congratulations, you have completed the installation process and successfully installed the MATLAB and its other products. Now click on the *Finish button*.

Step 5: A MATLAB shortcut will be created on the desktop as per our choice during the installation process. Now we can work with MATLAB by clicking on the icon placed there on the desktop.

MATLAB provides a Desktop GUI based environment. The Home tab of the environment contains three panels there:

1. Current folder: It is located on the left side; here, you can access your files.
2. Command Window: It is located on the right side; it is the command prompt, and here you can enter commands to operate the functions, to assign variables, and for calculations.
3. Workspace: It is located on the left side right below the Current Folder; here, all variables that you create are stored, and data from other files can also be imported here.

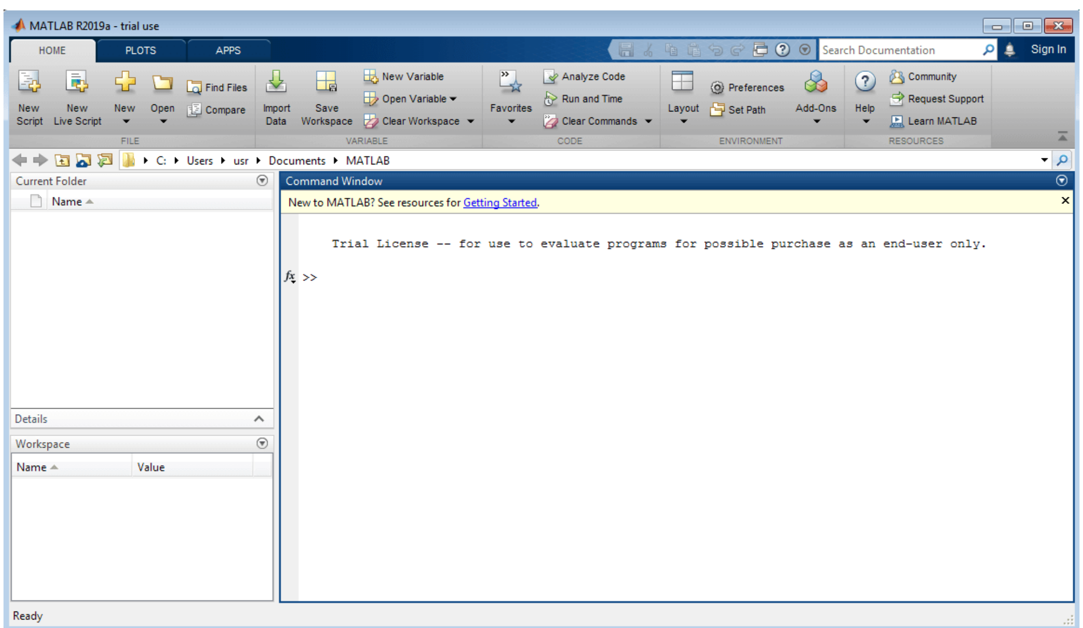

Fig 4.4 – MATLAB Homepage UI

4.4 MATLAB IMAGES:

4.4.1 READING IMAGE IN MATLAB:

First, clear the workspace of any variables and close open figure windows.

Clear all; close all;

Read an image into the workspace, using the **imread** command. The example reads one of the sample images included with the toolbox, an image of a cameraman in a file named cameraman.tif , and stores it in an array named I . imread infers from the file that the graphics file format is Tagged Image File Format (TIFF).

I = imread('cam.jpg');

4.4.2 DISPLAYING IMAGE:

Display the image using imshow function. You can also view an image in the Viewer app.The imtool function opens the Image Viewer app which presents an integrated environment for displaying images and performing some common image processing tasks. The Image Viewer app provides all the image display capabilities of imshow but also provides access to several other tools for navigating and exploring images, such as scroll bars, the Pixel Region tool, Image Information tool, and the Contrast Adjustment tool.

Z = imread('cam.jpg');

imwrite(z,'cameraman.png');

img = imread('cameraman.png');

imshow(img);

Fig 4.5 – Image displayed using imshow() built-in function

4.4.3 WRITING IMAGES:

imwrite(A,filename) writes image data A to the file specified by **filename**, inferring the file format from the extension. **Imwrite** creates the new file in your current folder. The bit depth of the output image depends on the data type of A and the file format. For most formats:

If A is of data type **uint8**, then imwrite outputs 8-bit values.

If A is of data type **uint16** and the output file format supports 16-bit data (JPEG, PNG, and TIFF), then imwrite outputs 16-bit values. If the output file format does not support 16-bit data, then imwrite returns an error.

If A is a grayscale or RGB color image of data type double or single, then imwrite assumes that the dynamic range is [0,1] and automatically scales the data by 255 before writing it to the file as 8-bit values. If the data in A is single, convert A to double before writing to a GIF or TIFF file.

If A is of data type logical, then imwrite assumes that the data is a binary image and writes it to the file with a bit depth of 1, if the format allows it. BMP, PNG, or TIFF formats accept binary images as input arrays.

4.5 MATLAB AUDIO:

4.5.1 READING AUDIO FILE

audioread() is a special function that provides a single, unified Matlab property for reading audio files in a range of different file formats, including wav, mp3, aac, flac, AIFF, etc. In most cases, access is actually provided by external binaries, but this is hidden within audioread (and its subfunctions). The idea is to make your code independent of the kind of audio files used as input.

In addition, audioread provides support for common modifications performed during sound input, namely resampling (changing the sampling rate), casting multi-channel signals to mono, and loading only a limited time range of the sound. Care has been taken to ensure that short subregions of very large audio files can be read with the minimum of memory overhead (for instance, without having to load in the entire audio file at any time) [26]

FORMAT:

[y,Fs] = audioread(filename)

[y,Fs] = audioread(filename,samples)

DESCRIPTION

[y,Fs] = audioread(filename) reads data from the file named filename, and returns sampled data, y, and a sample rate for that data, Fs.

[y,Fs] = audioread(filename,samples) reads the selected range of audio samples in the file, where samples is a vector of the form [start,finish].

EXAMPLE:

[y,Fs]=audioread('audio.wav');

4.5.2 WRITING AUDIO FILE

audiowrite(), a function to write out soundfiles in different formats, where the format is guessed from the provided filename extension. This is only a quick attempt, however, since there is no provision for controlling the encoding performed by lossy formats (mp3 etc.) [26].

FORMAT:

audiowrite(filename,y,Fs)

audiowrite (filename,y,Fs,Name,Value)

DESCRIPTION

audiowrite(filename,y,Fs) writes a matrix of audio data, y, with sample rate Fs to a file called filename. The filename input also specifies the output file format. The output data type depends on the output file format and the data type of the audio data, y.

audiowrite(filename,y,Fs,Name,Value) uses additional options specified by one or more Name, Value pair arguments.

EXAMPLE:

audiowrite(filename,y,Fs);

4.5.2 PLAYING AUDIO IN MATLAB

y = audioplayer(x,Fs) returns a handle to an audio player object y using input audio signal x. The audio player object supports methods and properties that you can use to play audio data.

The input signal '**x**' can be a vector or two-dimensional array containing single, double, int8, uint8, or int16 MATLAB data types. The input sample value range depends on the MATLAB data type.

Fs is the sampling rate in Hz to use for playback. Valid values for Fs depend on the specific audio hardware installed. Typical values supported by most sound cards are 8000, 11025, 22050, and 44100 Hz.

Y = audioplayer(x,Fs,nbits) returns a handle to an audio player object where nbits is the bit quantization to use for single or double data types. This is an optional parameter with a default value of 16. Valid values for nbits are 8 and 16 (and 24, if a 24-bit device is installed). You do not need to specify nbits for int8, uint8, or int16 data because the quantization is set automatically to 8 or 16, respectively.

Y = audioplayer® returns a handle to an audio player object from an audiorecorder object r.

y = audioplayer(r,id) returns a handle to an audio player object from an audiorecorder object r, using the audio device specified by id for output. This option is only available on systems running Windows

After the audioplayer function returns an object, the object needs to be called to play the audio

play(y);

Starts playback from the beginning and plays to the end, or from start sample to the end, or from start sample to stop sample. The values of start and stop can be specified in a two-element vector range.

Playblocking(y)
playblocking(y,start)
playblocking(y,[start stop])

Same as play but does not return control until playback completes.

Stop(y)	Stops playback.
Pause(y)	Pauses playback.
Resume(y)	Restarts playback from where playback was paused.
Isplaying(y)	Indicates whether playback is in progress. If 0, playback is not in progress. If 1, playback is in progress.
Display(y) disp(y) get(y)	Displays all property information about audio player y.

4.6 MATLAB GUI

GUIs (also known as graphical user interfaces or Uis) provide point-and-click control of software applications, eliminating the need to learn a language or type commands in order to run the application.

MATLAB apps are self-contained MATLAB programs with GUI front ends that automate a task or calculation. The GUI typically contains controls such as menus, toolbars, buttons, and sliders. Many MATLAB products, such as Curve Fitting Toolbox, Signal Processing Toolbox™, and Control System Toolbox™ include apps with custom user interfaces. You can also create your own custom apps, including their corresponding Uis, for others to use. Graphical user interfaces (GUIs), also known as apps, provide point-and-click control of your software applications, eliminating the need for others to learn a language or type commands in order to run the application. You can share apps both for use within MATLAB and also as standalone desktop or web apps.

Step 1 : Start GUIDE by typing `guide` at the MATLAB prompt.

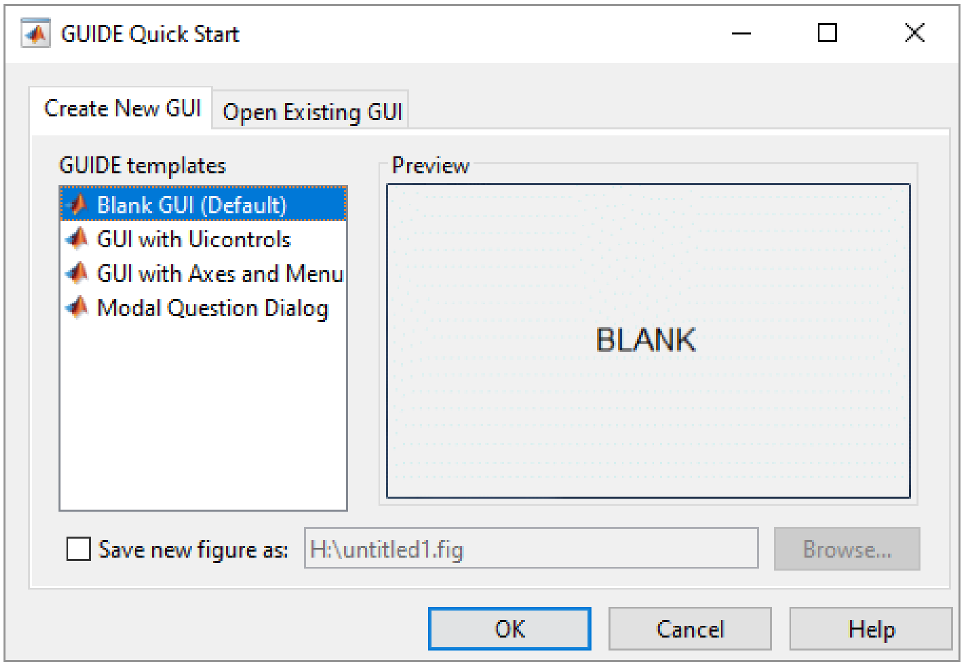

Fig 4.6 – MATLAB Guide prompt

Step 2 :In the GUIDE Quick Start dialog box, select the **Blank GUI (Default)** template, and then click **OK**.

a. Select **File** > **Preferences** > **GUIDE.**

b. Select **Show names in component palette**.

c. Click **OK**.

Step 3 : Set the size of the window by resizing the grid area in the Layout Editor. Click the lower-right corner and drag it until the canvas is approximately 3 inches high and 4 inches wide. If necessary, make the canvas larger.

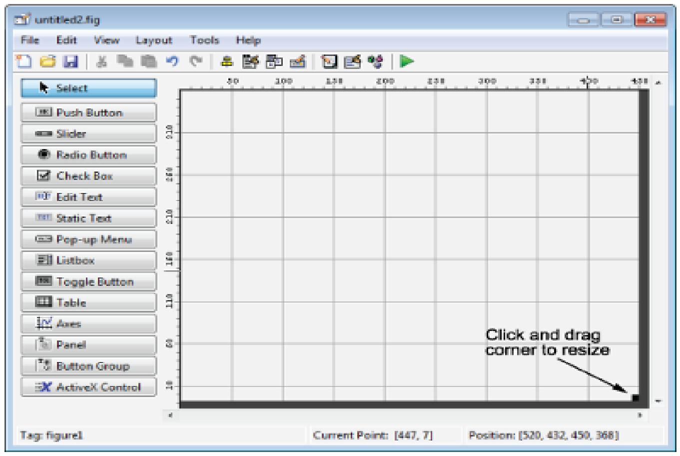

Fig 4.7 – Empty GUI

Step 5: Creating the layout arranging buttons, axes, static text, edit text in order

- **2 Axes** to display the original picture and encrypted picture
- **3 Buttons to** display the original image to encrypt the text and other to decrypt the text
- **1 Edit text** to enter the text which is to be encrypted
- **1 Static text** to display the decrypted text

After arranging all the components, the layout looks like:

After clicking the run button, the following execution window appears

Step 6 : Save the Layout

When you save a layout, GUIDE creates two files, a FIG-file and a code file. The FIG-file, with extension. Fig, is a binary file that contains a description of the layout. The code file, with extension .m, contains MATLAB functions that control the app's behavior.

1. Save and run your program by selecting **Tools** > **Run**.

2. GUIDE saves the files simple_gui.fig and simple_gui.m, and then runs the program. It also opens the code file in your default editor.

The app opens in a new window. Notice that the window lacks the standard menu bar and toolbar that MATLAB figure windows display. You can add your own menus and toolbar buttons with GUIDE, but by default a GUIDE app includes none of these components. When you run simple GUI, you can select a data set in the pop-up menu and click the push buttons, but nothing happens. This is because the code file contains no statements to service the pop-up menu and the buttons..

4.7 CONCLUSION

In this chapter, we have reviewed the system requirements for the implementation of the project and the various components of MATLAB are discussed. The installation procedure and setup of the MATLAB software is mentioned in the chapter. A study of various functions required for the project for successful execution of the spectral subtraction algorithm is discussed.

CHAPTER 5

RESULT AND DISCUSSION

In noisy environments, speech communications would cease and deteriorate due to the noise contamination. Thus, speech enhancement systems are used as means to provide adequate and effective noise suppression to enhance the quality of speech in such adverse environments. The main goal of this project is to develop speech enhancement techniques that provide the ability to suppress the background noise, and other interference while preserving the original speech signal without too much distortion. In order to achieve this goal, different speech enhancement techniques have been investigated. In this chapter, we summarize the main conclusions drawn from this project, as well as highlight some suggestions for further works.

We have developed a speech enhancement technique that is applicable in many speech communication systems such as hands-free mobile phones and hearing aids. There are many considerations that have to be taken into account when designing such systems such as good noise reduction performance without too much speech distortion, quick response to abrupt changes in the observed noisy signal and low computational complexity for less power usage.

Moreover, in the proposed techniques we focus on overcoming the drawbacks of the conventional speech enhancement systems. In that case, reducing the musical noise and improving the tracking speed of the a priori SNR estimation are the main problems we aim to reduce.

A stand-alone noise suppression algorithm is presented for reducing the spectral effects of acoustically added noise in speech. Effective performance of digital speech processors operating in practical environments may require suppression of noise from the digital waveform. Spectral subtraction offers a computationally efficient, processor-independent approach to effective digital speech analysis. The method, requiring about the same computation as high-speed convolution, suppresses stationary noise from speech by subtracting the spectral noise bias calculated during nonspeech activity. Secondary procedures are then applied to attenuate the residual noise left after subtraction. Since the algorithm resynthesizes a speech waveform, it can be used as a pre-processor to narrow-band voice communications systems, speech recognition systems, or speaker authentication systems.

5.1 MATLAB GUI

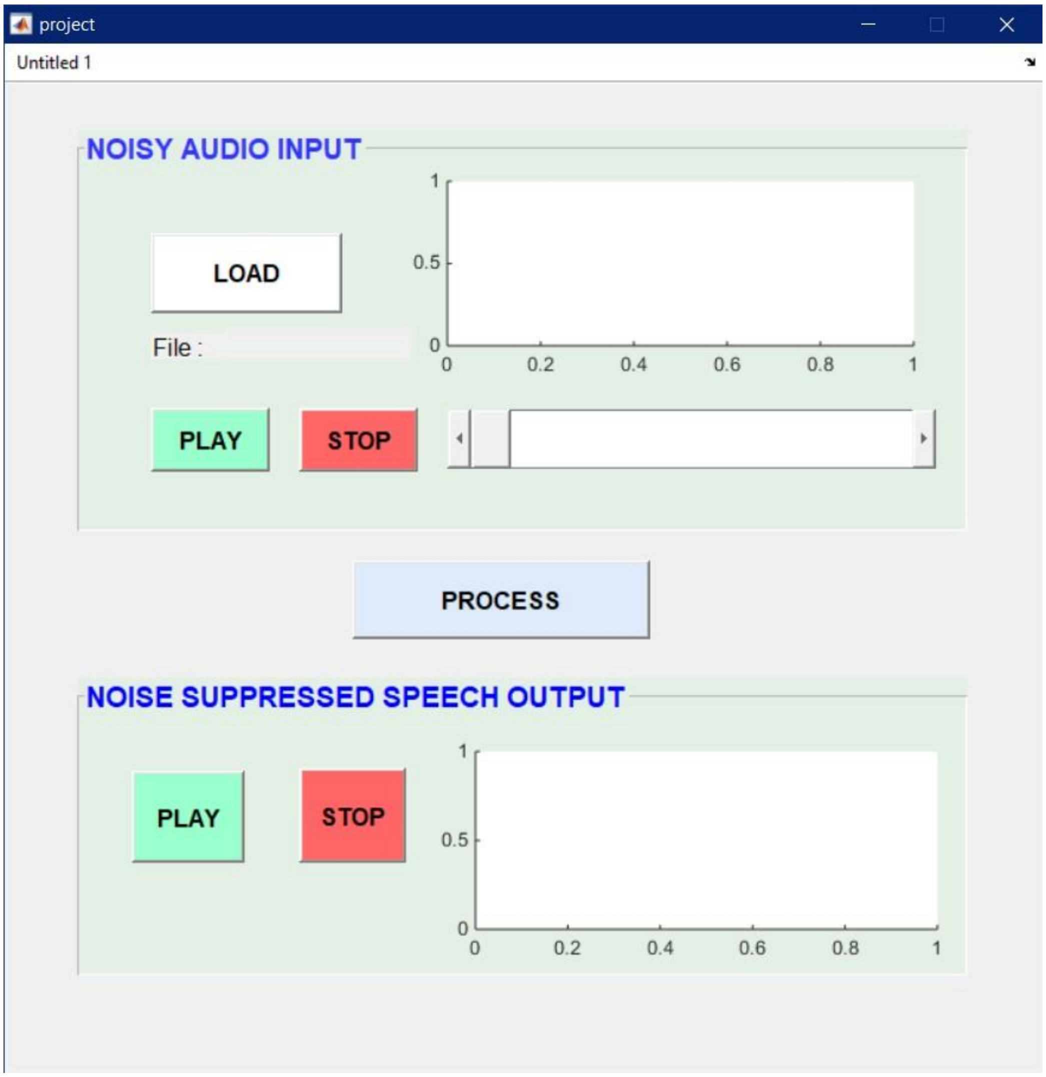

Fig 5.1 MATLAB GUI

A graphical user interface is created using in-built MATLAB GUI function to control both the input and output audio, where we can see three buttons which are used for loading, playing and pausing as well as completely stopping the audio file. We also have an axes to show the time vs amplitude plot of the input audio signal.

After loading the audio file into MATLAB platform, the loaded noisy speech is inputted to the spectral subtraction algorithm for noise suppression mechanism. On clicking the process button, The audio signal then undergoes pre-processing where all the parameters are defined and sent for the digital signal processing.

The audio goes through five steps of processing where all the features are extracted, and average of the noisy signal is calculated to get the noise spectrum by calculating the signal to noise ratio of the noisy signal. The signal is next taken through spectral subtraction algorithm to remove the relative background noise and retain the speech signal.

5.2 MATLAB OUTPUT

The algorithm pertaining to the noise suppression in audio is written in MATLAB and the following results are obtained. The following picture shows step by step execution of the complete process MATLAB.

```
Command Window
>> project
<------------------- Noise suppression project -------------->


-> Step 1: Loading Audio file -----------> Done

-> Step 2: Extracting noise spectrum ----> Done

-> Step 3: Estimating SNR ---------------> Done

-> Step 4: Compute TF attenuation map ---> Done

-> Step 5: Compute Inverse STFT ---------> Done
```

Fig 5.2 MATLAB Command line output

In the first step of the process, the audio file is read onto the MATLAB platform using audioread() in-built function, then the noise spectrum is extracted from the loaded audio signal and then the estimated SNR is calculated. In the next steps attenuation map is computed from the estimated SNR and then moving ahead the signal is transformed back into time domain by using Inverse STFT. The algorithm shows “Done” as the execution message at every step to know everything runs without an error at every step.

Then five seconds of original noisy audio is played which is compared with that of five seconds off output noise-less audio. The difference between both the audio signals is clearly audible and the difference is noted as well.

5.3 SPECTROGRAM OUTPUT

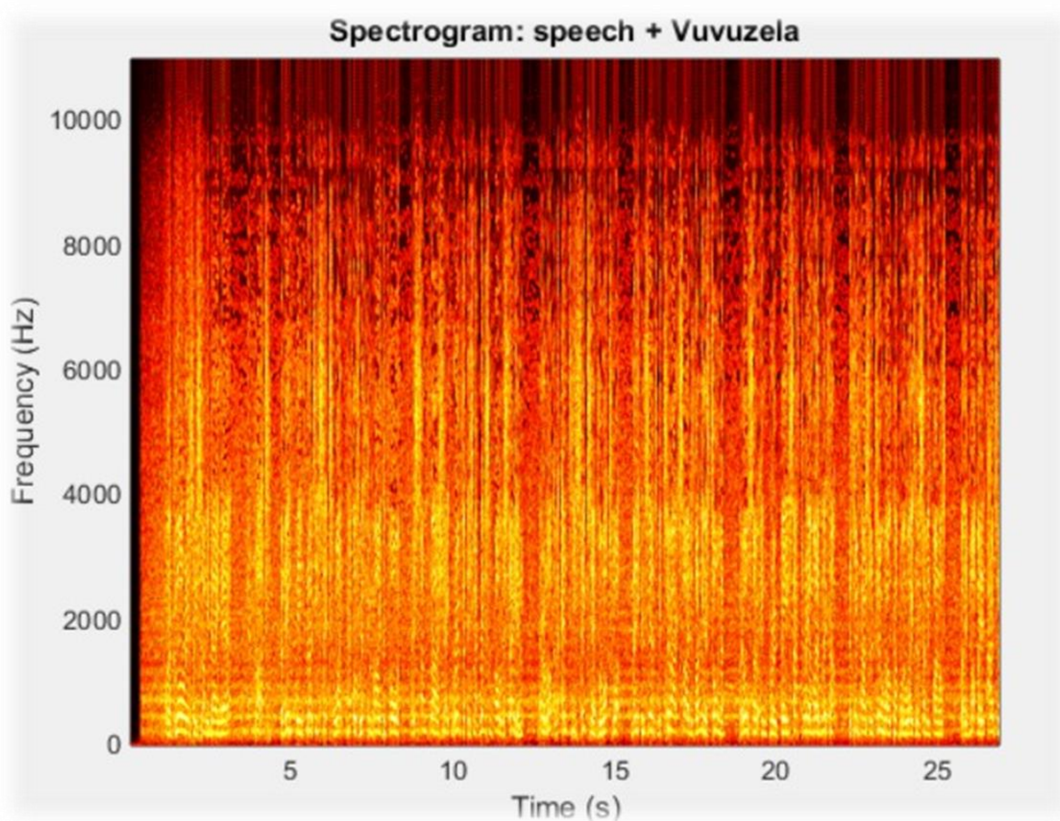

Fig 5.3 Spectrogram of noisy speech signal

A spectrogram of the noisy speech signal consisting of Speech and noise is plotted in MATLAB figure using spectrogram( ) in-built function. The spectrogram of the signal looks like heat-map due to heavy interference of noise in speech while recording.

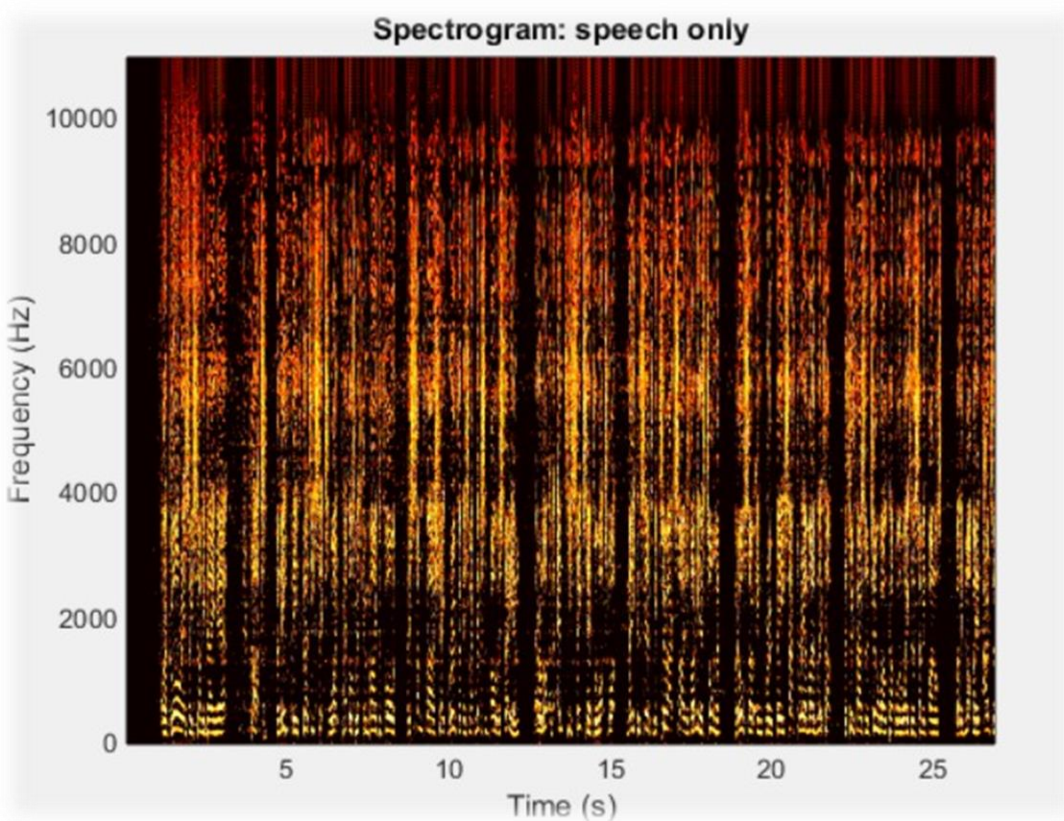

Fig 5.4 Spectrogram of output signal

The spectrogram of output signal which is clear speech signal is displayed, where the speech only signal is visible in the spectrogram map.

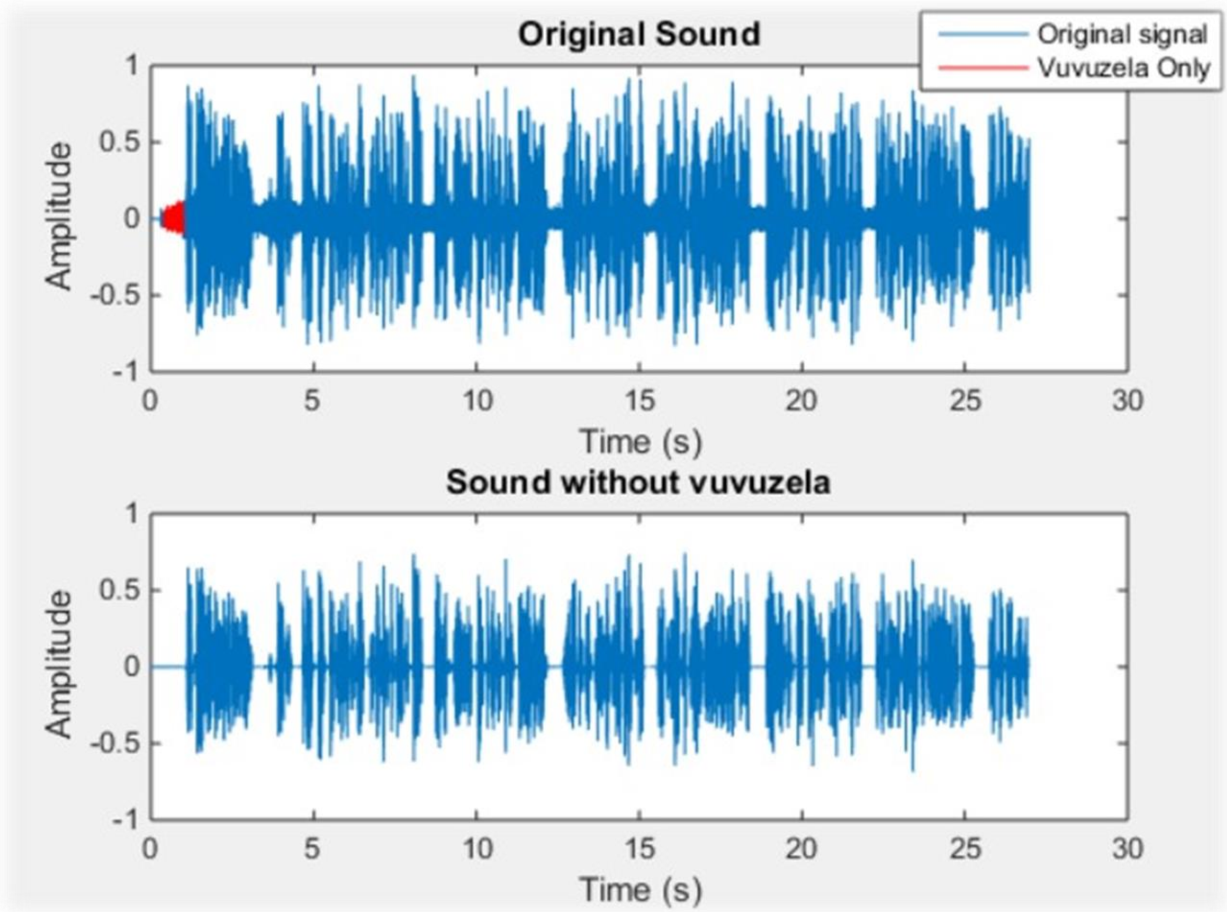

Fig 5.5 Comparison between original signal and output signal

A comparison of input and output sounds is shown in the above figure where we can see visible changes in the amplitude vs time graph of both the signals. The amplitude of output speech signal goes low when the user's speech is completely zero and rest the times the amplitude varies in accordance with time.

5.4 FUTURE WORK

Throughout this project, we have developed a speech enhancement techniques that is applicable in different places such as hands-free mobile phones, hearing aids and teleconferencing systems. The main goals of the proposed technique are to extract the desired speech signal while mitigate the unwanted signals. Different objective measurements are used to evaluate the performance of the developed algorithms in improving the speech quality and suppressing the noise signals. In the following subsections, several issues that may be addressed for further research are discussed.

Speech enhancement has substantial interest in the utilization of speaker identification, videoconference, speech transmission through communication channels, speech-based biometric system, mobile phones, hearing aids, microphones, voice conversion etc. Pattern mining methods have a vital step in the growth of speech enhancement schemes. To design a successful speech enhancement system consideration to the background noise processing is needed. A substantial number of methods from traditional techniques and machine learning have been utilized to process and remove the additive noise from a speech signal. Methods of speech enhancement consist of different stages, such as feature extraction of the input speech signal, feature selection. Deep learning techniques are also an emerging field in the classification domain, which is discussed in this review. The intention of this project is to provide a state-of-the-art summary and present approaches for using the widely used machine learning and deep learning methods to detect the challenges along with future research directions of speech enhancement systems.

Further investigation of the efficiency of the perceptual based filter bank in improving the speech intelligibility is of great importance. This might improve the applicability of the proposed technique in different signal processing systems such as speech recognition. Moreover, future research could focus on different choices of filter banks to further improve the performance of the speech enhancement techniques.

REFERENCES

[1] L. R. Rabiner and R. W. Schafer, Theory and applications of digital speech processing. Pearson Upper Saddle River, 2011, vol. 64.

[2] J. Benesty, S. Makino, and J. Chen, Speech enhancement. Springer Science & Business Media, 2005.

[3] P. C. Loizou, Speech enhancement: theory and practice. CRC press, 2013.

[4] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft decision noise suppression filter," IEEE Transactions on Acoustics Speech and Signal Processing, vol. 28, no. 2, pp. 137–145, 1980.

[5] H. Gustafsson, S. E. Nordholm, and I. Claesson, "Spectral subtraction using reduced delay convolution and adaptive averaging," IEEE Transactions on Speech and Audio Processing, vol. 9, no. 8, pp. 799–807, 2001.

[6] E. Referent, I. T. Fingscheidt, and I. R. Martin, "Speech enhancement using data-driven concepts," Braunschweig University, 2012.

[7] W. Charoenruengkit, Spectral refinements to speech enhancement. Florida Atlantic University, 2009.

[8] J. Benesty, S. Makino, and J. Chen, Speech enhancement. Springer Science & Business Media, 2005.

[9] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," Signal processing, vol. 81, no. 11, pp. 2403–2418, 2001.

[10] M. Abbott, "The use of speech technology to enhance the handling of electronic flight progress strips in an air traffic control environment," Proceedings of Voice Systems Worldwide, pp. 126–134, 1990.

[11] C. J. Weinstein, "Opportunities for advanced speech processing in military computer-based systems," in Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27, 1990, 1990.

[12] **]** A. J. Brammer, "Hearing protection and communication in an age of digital signal processing: Progress and prospects," in Proceedings on Noise as a Public Health Problem, 2008, pp. 21–25.

[13] N. Madhu, A. Spriet, S. Jansen, "The potential for speech intelligibility improvement using the ideal wiener filter in single channel noise reduction systems:" IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 1, pp. 63–72, 2013.

[14] A. Johansson, S. Nordholm, and E. Ostlin, "Microphone array designed ¨ for a conference room," Report, Australian Telecommunications Research Institute, Bentley, Australia, Tech. Rep., 2001.

[15] P.C.Yong, "Speech enhancement binaural hearing protection devices," Ph.D. dissertion, Curtan University, 2013.

[16] S.Doclo, "Multi-microphone noise reduction and deverberation techniques for speech applications ," Ph.D.thesis, University of Leuven, Leuven, Belgium, 2003.

[17] J. Benesty, M. M. Sondhi, Y. Huang (ed). Springer Handbook of Speech Processing. pp.843-869. Springer, 2007.

[18] Boll S.F (1979) Suppression of Acoustic Noise I Speech Using Spectral Subtraction .IEEE Tran. On Acoustics, Speech and Signal Processing ASSP-27,2pp. 113-120.

[19] Lim J.S (1978) Evaluations of Correltion Subtraction Method for Enhancing Speech degraded by Additive White Noise . IEEE Trans. Acoustics, Speech and Signal Processing, ASSP-26, 5, pp.471-472.

[20] I. Y. Soon and S. N. Koh, ―Low distortion speech enhancement,‖ Proc.Inst. Elect. Eng., Vis. Image Signal Process., vol. 147, pp. 247–253. 2000.

[21] N.Virag, ―Single channel speech enhancement system based on maskingproperties of the human auditory system,‖ IEEE Trans. Speech AudioProcessing, vol. 7, pp. 126–137, Mar. 1999.

[22] J. H. L. Hansen and B. L. Pellom, ―An effective evaluation protocolfor speech enhancement algorithms,‖ in Proc. ICSLP, vol. 7, Sydney,Australia, 1998, pp. 2819–2822.

[23] What is MATLAB? https://cimss.ssec.wisc.edu/wxwise/class/aos340/spr00/whatismatlab.htm

[24] Audio processing https://en.wikipedia.org/wiki/Audio_signal_processing

[25] MATLAB installation https://www.javatpoint.com/how-to-install-matlab

[26] AUDIO in MATLAB https://in.mathworks.com/help/matlab/ref/audioread.html

APPENDIX

PROJECT SOURCE CODE:

```
clc;

close all;

fprintf('--- Vuvuzela Cancelation Program ---\n\n');


%load vuvuzela sound example

fprintf('-> Step 1/5: Load vuvuzela.wav:');

[y,Fe]=audioread('Vuvuzela.wav');

x=y(100000:end,1).'; %remove the beginning of the sample

Nx=length(x);

fprintf(' OK\n');


%algorithm parameters

apriori_SNR=1; %select 0 for aposteriori SNR estimation and 1 for apriori

alpha=0.05;     %only used if apriori_SNR=1

beta1=0.5;

beta2=1;

lambda=3;
```

```
%STFT parameters
NFFT=1024;
window_length=round(0.031*Fe);
window=hamming(window_length);
window = window(:);
overlap=floor(0.45*window_length); %number of windows samples without overlapping

%Signal parameters
t_min=0.4;    %interval for learning the noise
t_max=1.00;   %spectrum (in second)

%construct spectrogram
[S,F,T] = spectrogram(x+i*eps,window,window_length-overlap,NFFT,Fe);
%put a short imaginary part to obtain two-sided spectrogram
[Nf,Nw]=size(S);

%       noisy spectrum extraction  %

fprintf('-> Step 2/5: Extract noise spectrum -');
t_index=find(T>t_min & T<t_max);
absS_vuvuzela=abs(S(:,t_index)).^2;
```

vuvuzela_spectrum=mean(absS_vuvuzela,2); **%average spectrum of the vuvuzela (assumed to be ergodic))**

vuvuzela_specgram=repmat(vuvuzela_spectrum,1,Nw);

fprintf(' OK\n');

% Estimate SNR %

fprintf('-> Step 3/5: Estimate SNR -');

absS=abs(S).^2;

SNR_est=max((absS./vuvuzela_specgram)-1,0); **% a posteriori SNR**

if apriori_SNR==1

SNR_est=filter((1-alpha),[1 -alpha],SNR_est); **%a priori SNR:**

end

fprintf(' OK\n');

% Compute attenuation map %

fprintf('-> Step 4/5: Compute TF attenuation map -');

an_lk=max((1-lambda*((1./(SNR_est+1)).^beta1)).^beta2,0);

STFT=an_lk.*S;

```
fprintf(' OK\n');


%   Compute Inverse STFT   %


fprintf('-> Step 5/5: Compute Inverse STFT:');
ind=mod((1:window_length)-1,Nf)+1;
output_signal=zeros((Nw-1)*overlap+window_length,1);


for indice=1:Nw  %Overlapp add technique
    left_index=((indice-1)*overlap) ;
    index=left_index+[1:window_length];
    temp_ifft=real(ifft(STFT(:,indice),NFFT));
    output_signal(index)= output_signal(index)+temp_ifft(ind).*window;
end
fprintf(' OK\n');




%-----------------   Display Figure   -----------------------------------


%show temporal signals
figure
```

```
subplot(2,1,1);

t_index=find(T>t_min & T<t_max);

plot([1:length(x)]/Fe,x);

xlabel('Time (s)');

ylabel('Amplitude');

hold on;

noise_interval=floor([T(t_index(1))*Fe:T(t_index(end))*Fe]);

plot(noise_interval/Fe,x(noise_interval),'r');

hold off;

legend('Original signal','Vuvuzela Only');

title('Original Sound');

%show denoised signal

subplot(2,1,2);

plot([1:length(output_signal)]/Fe,output_signal );

xlabel('Time (s)');

ylabel('Amplitude');

title('Sound without vuvuzela');
```

%show spectrogram

```
t_epsilon=0.001;

figure
```

```
S_one_sided=max(S(1:length(F)/2,:),t_epsilon); %keep only the positive frequency

pcolor(T,F(1:end/2),10*log10(abs(S_one_sided)));

shading interp;

colormap('hot');

title('Spectrogram: speech + Vuvuzela');

xlabel('Time (s)');

ylabel('Frequency (Hz)');


figure

S_one_sided=max(STFT(1:length(F)/2,:),t_epsilon); %keep only the positive frequency

pcolor(T,F(1:end/2),10*log10(abs(S_one_sided)));

shading interp;

colormap('hot');

title('Spectrogram: speech only');

xlabel('Time (s)');

ylabel('Frequency (Hz)');



%-----------------   Listen results   ------------------------------------


fprintf('\nPlay 5 seconds of the Original Sound:');
```

```
gong = audioplayer(x, Fe);

play(gong);


fprintf(' OK\n');

fprintf('Play 5 seconds of the new Sound: ');

out=audioplayer(output_signal,Fe);

%play(out)

fprintf('OK\n');

fprintf('Write anti_vuvuzela.wa:');

audiowrite('anti_vuvuzela.wav',output_signal,Fe*5);

fprintf('OK\n');
```