Speech enhancement methods in hands-free communication with emphasis on optimal SNIR Beamformer

Midathala Harish

This thesis is presented as part of Degree of
Master of Science in Electrical Engineering with Emphasis on Signal Processing

Blekinge Institute of Technology
April 2012

Supervisor: Dr. Nedelko Grbic
Examiner: Dr. Benny Sallberg
Department of Signal Processing
School of Engineering (ING)
Blekinge Institute of Technology
This thesis is submitted to the School of Engineering at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering with Emphasis on Signal Processing.

Contact Information:

Author:
Midathala Harish
E-mail: hami10@student.bth.se

Supervisor:
Dr. Nedelko Grbic
School of Engineering (ING)
E-mail: nedelko.grbic@bth.se
Phone no.: +46 455 38 57 27

Examiner:
Dr. Benny Sallberg
School of Engineering (ING)
E-mail: benny.sallberg@bth.se
Phone no.: +46 455 38 55 87
ABSTRACT

A basic speech enhancement can be achieved by the suppression of background noise and reverberation from the clean speech. The point to be noted is to achieve it with a low computational complexity. The aim is to estimate signal arriving optimally from the desired direction in the presence of reverberant-noisy speech signal. Recent studies show that this can be achieved by designing various kinds of robust fixed and adaptive beamformers. A beamformer does spatial filtering in the sense that it separates two signals with overlapping frequency content originating from distinctive directions. In this contribution, robust beamformers namely Elkos beamformer, Wiener beamforming and optimal signal to noise interference ratio (SNIR) beamformer are designed and analyzed collaboratively in a group under the consideration of hearing aid constraints such as the microphone distance and different real world room dimensions. A fractionally delayed (FD) all pass Thiran filters are designed to get a maximally flat group delay. A virtual room image model is designed to achieve different dimensions of the room and their reverberant speech signals.

The objective of this thesis is to design and implement an optimal SNIR beamformer in anechoic and reverberant environments with different noises, i.e. wind, white, factory and interference. It is implemented and simulated offline in MATLAB. The performance of the optimal SNIR Beamformer is evaluated by considering the objective measures such as SNRI, SD, ND, RR and PESQI under different noisy environments in anechoic and reverberated environments. These parameters are measured by assuming input SNR levels at 0dB, 5dB, 10 dB, 15 dB, 20 dB and 25 dB. In addition to this a new parameter RR is also evaluated in reverberated environment. This parameter is measured by varying the number of microphones. The reverberation power suppression is analyzed by using RR. Speech quality is analyzed based on signal to noise ratio Improvement and speech intelligibility is measured using PESQ for different noisy environments. Results show that optimal SNIR beamformer performs best compared to all other beamformers due to its inherent properties.

Keywords: signal to noise ratio, Beamforming, Reverberated and Anechoic, Speech intelligibility, Reverberation.
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and thanks to my thesis supervisor Dr. Nedelko Grbic for providing me a chance to do my thesis research work under his supervision in the field of Speech Processing. I would like to thank him for his persistent help throughout the thesis work. With his deep knowledge in this field, he helped us to learn new things in order to complete master thesis successfully. His continuous feedback and encouragement helped in finishing this thesis work.

I extend my appreciation and thanks to my fellow students Santhurenu Vuppala, Ramesh Telagareddy and Aditya Sri Teja Palanki for their suggestions and discussions regarding solving different problems in doing this research thesis. They have given continuous support and discussed several issues related to thesis work.

I would like to thank BTH for providing me a good educational environment where we can gain the knowledge and learn about new technologies that help us to move forward with the thesis work.

Finally, I would like to extend my immense gratitude and wholehearted thanks to my parents for their moral support and financial support throughout my educational career. They have motivated and helped me for the successful completion of thesis work. I also thank my pals for their support and encouragement during the thesis work. I take an opportunity to thank all the Signal processing staff at BTH.

I would lastly like to thank all for their support and help in any aspect in the successful completion of the thesis work.
CONTENTS

Abstract .................................................................................................................i
Acknowledgements ............................................................................................ii
List of Figures ........................................................................................................v
List of Tables ..........................................................................................................viii
List of Acronyms ....................................................................................................x

1 INTRODUCTION .........................................................................................1
  1.1 Introduction to speech signal .................................................................1
  1.2 Problem statement .................................................................................1
  1.3 Research Question ..................................................................................2
  1.4 Scope of the Thesis ................................................................................2
  1.5 Thesis Outline ..........................................................................................2

2 BACKGROUND THEORIES .........................................................................3
  2.1 Hearing Aids ............................................................................................3
  2.2 Source localization ..................................................................................4
  2.3 Background noise ...................................................................................5
  2.4 Echo cancellation ....................................................................................5
  2.5 Reverberation ..........................................................................................6
  2.6 Fractional delay filters ...........................................................................6
  2.7 Fractional Delay ......................................................................................7
    2.7.1 Implementation of Fractional Delay ..................................................8
    2.7.2 FIR Method (Using Lagrange Interpolator) .......................................8
    2.7.3 IIR Method .......................................................................................9
    2.7.4 Thiran All pass Filter Design ............................................................9

3 ROOM IMPULSE RESPONSE (RIR) ............................................................11
  3.1 Introduction .............................................................................................11
  3.2 Generation of room impulse response ....................................................13
  3.3 Room Image model ..................................................................................14

4 MICROPHONE ARRAY BEAMFORMING ..................................................17
  4.1 Need of Microphone Array .....................................................................17
  4.2 Common signal modeling .......................................................................18
  4.3 Adaptive Filtering ...................................................................................19
  4.4 Elko’s Beamformer ................................................................................19
5 Optimal Signal-to-Noise Interference Ratio (SNIR) Beamformer. 25
5.1 Eigen Decomposition of Autocorrelation Matrix .......................... 27
5.2 Modification for Optimal SNIR in reverberant case ...................... 28
6 Implementation ............................................................................. 29
6.1 Introduction ................................................................................ 29
6.2 Implementation .......................................................................... 29
6.3 Terms and Expansions ................................................................. 30
   6.3.1 Input Signals ........................................................................ 30
6.4 Objective measures .................................................................... 31
   6.4.1 Signal-to-Noise Ratio (SNR) .................................................... 32
   6.4.2 Speech and Noise Distortion (SD) and (ND) ............................. 32
   6.4.3 PESQ Score ........................................................................... 32
   6.4.4 Reverberation ratio (RR) ....................................................... 34
   6.4.5 Test signals .......................................................................... 35
7 Results .......................................................................................... 36
   7.1 Performance Evaluation of Optimal SNIR Beamformer in Reverberated and Anechoic Environment ........................................... 36
      7.1.1 Anechoic Optimal SNIR Beamformer ................................... 36
      7.1.2 Optimal SNIR Beamformer in reverberated environment ....... 47
      7.1.3 Reverberation ratio (RR) ...................................................... 51
   7.2 Comparison of Optimal SNIR with other beamformers in anechoic environment ................................................................. 53
      7.2.1 Elko’s Beamformer .............................................................. 53
      7.2.2 Wiener Beamformer .......................................................... 54
      7.2.3 Delay and Sum Beamformer ............................................... 54
8 Conclusion and Future Work ......................................................... 56
   8.1 Conclusion ................................................................................ 56
   8.2 Future Work ............................................................................ 57
9 Bibliography .................................................................................. 58
LIST OF FIGURES

FIGURE 1: Typical Hands-Free speech communication environment ........................................4
FIGURE 2: Time difference of arrival between microphones ..................................................5
FIGURE 3: Illustration of mobile to landline system ................................................................5
FIGURE 4: Sampled (dots) impulse response of ideal fractional delay filter when the delay is (a) D=2 sample and (b) D= 2.7 samples .................................................................5
FIGURE 5: The group delay of N=15, Thiran maximally flat fractional delay all pass filter ..........................................................................................................................8
FIGURE 6: Illustration of a desired source, a microphone and interfering source ..............10
FIGURE 7: Illustration of a direct path and a single reflection from the desired source to the microphone ........................................................................................................13
FIGURE 8: Illustration of a direct sound, an early and late reverberation from source to the microphone ................................................................................................................13
FIGURE 9: 3-D illustration of a direct sound, an early and late reverberation from source to the microphone ...........................................................................................................14
FIGURE 10: First reflection path of an image source ..............................................................15
FIGURE 11: Reverberated environment with reflected source images .................................15
FIGURE 12: Illustration of direct sound (red color) and reverberated sound (blue color) in a closed room environment ..................................................................................16
FIGURE 13: Microphone array model .................................................................................18
FIGURE 14: Back-to-back Cardioid Microphones .................................................................20
FIGURE 15: Elko’s differential array formation with spatial center at origin ....................21
FIGURE 16: Back-to-back Cardioid Polar pattern .................................................................21
FIGURE 17: Schematic Implementation of an adaptive first order differential microphone using back-to-back cardioids .................................................................22
FIGURE 18: Power Spectral Density of C_B (blue line) and C_F (green line) speech signals for source angle 45 degrees and noise angle at 135 degrees ..................................22
FIGURE 19: Power Spectral Density of C_B (blue line) and C_F (green line) speech signals for source angle 45 degrees and noise angle at 225 degrees ..................................23
FIGURE 20: Fixed optimal beam former .............................................................................26
FIGURE 21: Illustration of reverberated sound in a closed environment ...........................28
FIGURE 22: The experimental setup for validation of optimal beamformer model ..........29
FIGURE 23: Power Spectral density of wind, white, interference and factory noise signals .................................................................................................................................31
FIGURE 24: calculation of objective measure at the input and output of the implemented system..........................................................................................................................................................31
FIGURE 25: Model of PESQ using distorting system .................................................................................................................................................................................33
FIGURE 26: Illustration of room impulse response of room size [2*2*2] .................................................................................................................................34
FIGURE 27: Illustration of original and reverberated speech signal of room size [2*2*2]...........................................................................................................34
FIGURE 28: Illustration of original and reverberant speech signal of room size [5*10*7.5] ........................................................................................................35
FIGURE 29: Illustration of Source (at zero degrees) and interference (at 75 degrees) positions ..........................................................................................................................................................36
FIGURE 30: Plot of SNRI with input SNR in anechoic environment for 2 Mics in factory, white, wind and interference noise environments ..........................................................................................................................41
FIGURE 31: Plot of SNRI with input SNR in anechoic environment for 4 Mics in factory, white, wind and interference noise environments ..........................................................................................................................41
FIGURE 32: Plot of SNRI with input SNR in anechoic environment for 6 Mics in factory, white, wind and interference noise environments ..........................................................................................................................42
FIGURE 33: Plot of PESQI with input SNR in anechoic environment for 2 Mics in factory, white, wind and interference noise environments ..........................................................................................................................42
FIGURE 34: Plot of PESQI with input SNR in anechoic environment for 4 Mics in factory, white, wind and interference noise environments ..........................................................................................................................43
FIGURE 35: Plot of PESQI with input SNR in anechoic environment for 6 Mics in factory, white, wind and interference noise environments ..........................................................................................................................44
FIGURE 36: Plot of speech distortion (SD) with input in anechoic environment SNR for 2 Mics in factory, white, wind and interference noise environments ..........................................................................................................................44
FIGURE 37: Plot of speech distortion (SD) with input SNR in anechoic environment for 4 Mics in factory, white, wind and interference noise environments ..........................................................................................................................44
FIGURE 38: Plot of speech distortion (SD) with input SNR in anechoic environment for 6 Mics in factory, white, wind and interference noise environments ..........................................................................................................................45
FIGURE 39: Plot of noise distortion (ND) with input SNR in anechoic environment for 2 Mics in factory, white, wind and interference noise environments ..........................................................................................................................45
FIGURE 40: Plot of noise distortion (ND) with input SNR in anechoic environment for 4 Mics in factory, white, wind and interference noise environments ..........................................................................................................................45
FIGURE 41: Plot of noise distortion (ND) with input SNR in anechoic environment for 6 Mics in factory, white, wind and interference noise environments ..........................................................................................................................46
FIGURE 42: Illustration of source (at zero degrees) and interference (at 75 degrees) positions ..........................................................................................................................................................47
FIGURE 43: Plot of PESQI with input SNR in reverberated environment for 2 Mics in factory, white, wind and interference noise environments ..........................................................................................................................................................50
FIGURE 44: Plot of SNRI with input SNR in reverberated environment for 2 Mics in factory, white, wind and interference noise environments..............................50

FIGURE 45: Plot of noise distortion(ND) with input SNR in reverberated environment for 2 Mics in factory, white, wind and interference noise environments.........................51

FIGURE 46: Plot of speech distortion(SD) with input SNR in reverberated environment for 2 Mics in factory, white, wind and interference noise environments..........................51

FIGURE 47: Plot of Reverberation improvement with input SNR in reverberated environment for 2 Mics in factory, white, wind and interference noise environments.........53

FIGURE 48: Estimation of Output PESQ for various beamformers in different noisy environments.................................................................55
LIST OF TABLES

TABLE 1: The details of clean speech signal used for evaluation-------------------------------30
TABLE 2: Listening quality scale-----------------------------------------------------------33
TABLE 3: SNR, SD, ND, PESQ, SNRI and PESQI for clean speech signal with wind noise in anechoic environment-----------------------------------------------37
TABLE 4: SNR, SD, ND, PESQ, SNRI and PESQI for clean speech signal with factory noise in anechoic environment---------------------------------------------38
TABLE 5: SNR, SD, ND, PESQ, SNRI and PESQI for clean speech signal with white noise in anechoic environment-----------------------------------------------39
TABLE 6: SNR, SD, ND, PESQ, SNRI and PESQI for clean speech signal with interference in anechoic environment---------------------------------------------40
TABLE 7: SNR, SD, ND, PESQ, SNRI and PESQI for clean speech signal with wind noise in reverberated environment-------------------------------------------48
TABLE 8: SNR, SD, ND, PESQ, SNRI and PESQI for clean speech signal with white noise in reverberated environment-------------------------------------------48
TABLE 9: SNR, SD, ND, PESQ, SNRI and PESQI for clean speech signal with factory noise in reverberated environment------------------------------------------49
TABLE 10: SNR, SD, ND, PESQ, SNRI and PESQI for clean speech signal with interference in reverberated environment-------------------------------------------50
TABLE 11: RRin and RRout for a reverberated speech signal with interference-----------------52
TABLE 12: RRin and RRout for reverberated speech signal with wind noise---------------------52
TABLE 13: RRin and RRout for reverberated speech signal with factory noise------------------52
TABLE 14: RRin and RRout for reverberated speech signal with white noise--------------------52
TABLE 15: SNRI, PESQ, SD and ND values in anechoic environment for two Mics with various noise signals using Elko’s Beamformer----------------------------------53
TABLE 16: SNRI, PESQ, SD and ND values in anechoic environment for two Mics with various noise signals using Wiener Beamformer----------------------------------54
TABLE 17: SNRI, PESQ, SD and ND values in anechoic environment for two Mics with various noise signals using Delay and Sum Beamformer--------------------------------54
# List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLMS</td>
<td>Normalized Least-Mean Square</td>
</tr>
<tr>
<td>ASR</td>
<td>Automatic Speech Recognition</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>LMS</td>
<td>Least Mean Square</td>
</tr>
<tr>
<td>RLS</td>
<td>Recursive Least Square</td>
</tr>
<tr>
<td>APA</td>
<td>Affine Projection Algorithm</td>
</tr>
<tr>
<td>FIR</td>
<td>Finite Impulse Response</td>
</tr>
<tr>
<td>IIR</td>
<td>Infinite Impulse Response</td>
</tr>
<tr>
<td>WLS</td>
<td>Weighted Least Square</td>
</tr>
<tr>
<td>LS</td>
<td>Least Square</td>
</tr>
<tr>
<td>FD</td>
<td>Fractional Delay</td>
</tr>
<tr>
<td>RIR</td>
<td>Room Impulse Response</td>
</tr>
<tr>
<td>RTF</td>
<td>Room Transfer Function</td>
</tr>
<tr>
<td>ISM</td>
<td>Image Source Model</td>
</tr>
<tr>
<td>RADAR</td>
<td>Radio Detection and Ranging</td>
</tr>
<tr>
<td>DSB</td>
<td>Delay and Sum Beamformer</td>
</tr>
<tr>
<td>SNIR</td>
<td>Signal-to-Noise Interference Ratio</td>
</tr>
<tr>
<td>GSC</td>
<td>Generalized Side-lobe Canceller</td>
</tr>
<tr>
<td>LCMV</td>
<td>Linearly Constrained Minimum Variance</td>
</tr>
<tr>
<td>SD</td>
<td>Speech Distortion</td>
</tr>
<tr>
<td>ND</td>
<td>Noise Distortion</td>
</tr>
<tr>
<td>PESQ</td>
<td>Perceptual Evaluation of Speech Quality</td>
</tr>
<tr>
<td>Max-SNR</td>
<td>Optimal Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>SNRI</td>
<td>Signal-to-Noise ratio Improvement</td>
</tr>
<tr>
<td>AEC</td>
<td>Acoustic Echo Cancellation</td>
</tr>
<tr>
<td>ERLE</td>
<td>Echo Return Loss Enhancement</td>
</tr>
<tr>
<td>SNRI</td>
<td>Signal-to-Noise Ratio Improvement</td>
</tr>
<tr>
<td>PESQI</td>
<td>Perceptual Evaluation of Speech Quality Improvement</td>
</tr>
<tr>
<td>MOS</td>
<td>Mean Opinion Score</td>
</tr>
<tr>
<td>BN</td>
<td>Babble Noise</td>
</tr>
<tr>
<td>FN</td>
<td>Factory Noise</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>DN</td>
<td>Destroyer-engine Noise</td>
</tr>
<tr>
<td>REN</td>
<td>Restaurant Noise</td>
</tr>
<tr>
<td>WHN</td>
<td>White Noise</td>
</tr>
<tr>
<td>DOA</td>
<td>Direction of Arrival</td>
</tr>
<tr>
<td>PSD</td>
<td>Power Spectral Density</td>
</tr>
<tr>
<td>RR</td>
<td>Reverberation ratio</td>
</tr>
<tr>
<td>RRI</td>
<td>Reverberation ratio improvement</td>
</tr>
<tr>
<td>VAD</td>
<td>Voice Active Detector</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

1.1 Introduction to Speech signal

Speech is the fundamental way for humans to communicate. Sound is a sequence of waves of pressure which propagates through compressible media. During their propagation, waves can be reflected, refracted, or attenuated by the medium properties. For centuries, efforts have been made to improve individual’s communication over a great distance [3]. Speech processing techniques have been examined to be effective in improving speech intelligibility in noise for hearing impaired listeners. This technique also has the capability of preventing damage to hearing in high-noise environments such as aircrafts, factories and industries. For hearing impaired people, it is especially difficult to communicate with other persons in noisy environments. Therefore, speech enhancement systems have become an integral component of modern speech communication [9].

With today’s technology, multiple-video- conference calling stand outs as one of the widely used methods for multi-level communications. This is essential due to the fact that video-audio conferencing is convenient and cost efficient. All these applications need a problem-free noiseless scenario to have a good communication.

1.2 Problem statement

Direction of arrival (DOA) estimation can be exploited to effectively suppress the noise and can further be used to track the source signal. This DOA can be optimally measured using a microphone array model with the Fractional delay (FD). Also, the best optimal method has to be chosen so as to achieve a high improvement of speech quality and intelligibility.

In addition to these difficulties, a reverberation scenario in closed spaces has to be considered with thorough examination and then has to be suppressed. Reverberation suppression with a microphone array has to be observed in particular.

1.3 Research question

How good is the chosen optimal beamformer compared to other beamformers in real (Reverberant) and anechoic environments?

1.4 Scope of the Thesis

This collaborative thesis work aims at design and implementation of popular beamformers using a fractional delay filter. All these beamformers are examined in reverberated and anechoic environment. The entire work includes Elko’s beamformer, wiener beamformer, optimal-SNIR beamformer and delay-sum-beamforming, echo cancellation; a room image
model is used to simulation of closed room. This complete work is divided between four people.

This work concentrates on the implementation of optimal SNIR beamformer using Thiran fractional delay and examines it in anechoic and closed room environment.

1.5 Thesis outline

The complete report is organized into ten chapters. The introduction chapter deals with the problem and scope of the evaluation. Basic concepts are explained in chapter 2 as background work in which speech enhancement techniques and problems are discussed. Chapter 3 deals with the room impulse response in which room image model is explained. Chapter 4 deals with need of microphone array in which adaptive beamformers are explained. The optimal (SNIR) beamformer part is explained in chapter 5 in which optimal weight exaction by Eigen decomposition is explained in brief. Chapter 6 and Chapter 7 deal with implementation and results. Finally, the paper is concluded with an insight to the future work in the last chapter.
2 BACKGROUND THEORIES

Speech Enhancement is necessary in hands-free communication devices such as teleconferences and Automatic information systems. For example, Speech signals produced in a room generate reverberations, which are noticed when a hand-free single channel telephone system is used and binaural listening is not possible [9]. Enhancement of normal speech is required for hearing impaired persons to fit into their individual hearing capabilities.

Speech Enhancement in hand-free communication is possible by temporal filtering such as Wiener Filtering, noise cancellation and multi-microphone methods using different array techniques [9]. Room reverberation is handled with various array techniques. Hands-Free speech communication is generally characterized by reduction in speech naturalness and intelligibility resulting from corruption of the speech sound field during data capture by microphones, as well as speech distortion generated by data transmission and reproduction [9].

Hands-free speech enhancement is defined as the ability to improve the discrimination between speech and background noise, reverberation and other types of interferences colliding on microphones [9]. The perceptual aspects such as intelligibility and quality are necessary for speech enhancement in hands-free communication systems. The quality and intelligibility are not correlated. Intelligibility and quality cannot be achieved simultaneously. If intelligibility is improved the other, quality should be sacrificed. Intelligibility can be improved by emphasizing the high frequency content of the noisy speech signal. In other words, quality improvement is linked to the loss of intelligibility in the noisy speech signal. As, human ears have been designed in such a way that they have the capability of discrimination of speech in noisy reverberant environments.

2.1 Hearing Aids

In this universe at about 10-20 percentage of people are suffering from temporary hearing impairment. This hearing impairment basically caused by damage of inner ear hair cells in the process of aging or exposure to loud noise. The hearing aid system amplifies the received speech signal without considering the SNR level. If, in case it consists of noise, it is also amplified along with speech signal as hearing impaired people are incapable of distinguishing the noise and speech signals [2]. The main problem with a hearing aid is acoustic feedback. This is caused due to the small distance between speaker and microphone. To overcome the above problems, microphone arrays are used for speech enhancement and echo cancellation is used in order to remove acoustic feedback caused
between speaker and microphone. The microphone array processing is the best solution to remove noise as it has the feature of spatial selectivity known as Beamforming which has the capability of directional hearing. Beamforming reduces the level of directional and ambient noise signals, while minimizing distortion to the speech from a desired direction [2]. The improvements of the hearing aids are as follows.

![Diagram of Diffuse noise and Interference](image)

**Fig.1**: Typical Hands-free speech communication environment.

### 2.2 Source localization

Source localization of a microphone array gives information about the location of the speaker. The source localization is an inherent property of an ordinary human being. This is possible because, our brain matches these signals to the memory data to identify the speaker. This property of ears called as “sound source localization” [10].

In real world scenarios source localization can be achieved by using in a microphone array. Time difference of arrival method is popular for localization of source. A typical Time difference of arrival method is done in [23]. Recent studies show that a steered response power algorithm with phase transform (SRP-PHAT) is a robust algorithm for sound source localization in reverberant and multi-speaker environments. An extensive study on source localization using steered power response algorithms is done in [24].
2.3 Background Noise

Noise is present anywhere in all environments. Background noise is mostly due to wind, vibrations, tire friction, engines, fan noise, car traffic, background music in public places, vibration noise from high power equipment in heavy industries, the revolution of propellers in aircrafts. Severe background noise reduces the intelligibility of speech and also stress. In hands-free speech communication, background noise degrades the performance of Speech automation systems which is a severe problem for hearing aid users.

It also reduces the intelligibility of speech. Acoustic disturbances arrive from all directions which is assumed to be surrounding noise. Background noise contains higher level of low frequency content when compared to speech therefore; spectral based methods are used to extract speech. Generally, background noise is characterized by a Gaussian distribution whereas speech is characterized by Laplacian distribution. By selecting a certain class of distribution techniques can be developed for extracting speech or background noise.

Background noise also generated acoustically, these speech signals are the ones that are generated by echoes and multiple reflections. Suppression of these acoustic noise signals is the latest area of research. A combination of beamformers and microphone array with spatial processing techniques come into use for this enhancement. In order to achieve this one must know the basic description of echo cancellation and reverberation.

2.5 Echo cancellation
The echoes path is the unintended transmission path between transmitter and receiver in hands-free duplex communication. In full duplex communication, the far-end signal is emitted by the speaker propagates in the environment and is captured by the microphones in the same way as other interfering signals [1]. The acoustic feedback also constitutes the disturbance of the speaker who hears his or her own voice echoed which is a double-talk situation [9]. The echoes can affect the intelligibility of conversation between users in a telephone system. The echoes can be suppressed by using a reference signal at the loudspeaker. Acoustic echo mainly occurs due to the acoustic coupling between speaker and microphone in hands-free phones, mobile phones and teleconference systems as shown in figure 3. The acoustic echo is cancelled using adaptive algorithms such LMS, NLMS, and APA algorithms. In this thesis, main concentration is on cancelling the echo using NLMS, APA algorithm.

2.6 Reverberation

In a closed room environment multiple reflections of a speech signal leads to reverberation as shown in Figure 1. These reflections leads to the disturbance of speech produced from the speaker to microphone. Reverberation time is the time required for the reverberation energy to decay by 60 dB which is the main criteria for room reverberations [9]. The energy of confined reverberation depends on the position of sources and acoustic sensors in the room and their relative distances.

Suppression of reverberations signals is a challenging task in recent research acoustic projects. The reverberation affect can be eliminated by keeping the microphone close to the source of signal of interest which is caused by multiple reflections and diffractions of the sound and objects in a room. The dereverberation also adds an advantage to the hearing impaired people as it reduces speech intelligibility [7].

2.6 Fractional delay filters

Fractional delay filters are designed in the discrete-time domain by assuming that the incoming continuous-time signals are fully band-limited up to the Nyquist frequency. However, this assumption is not realistic because no real analog signals are fully band-limited. Moreover, by their very nature, such filters should reconstruct inter-sample signal values, and sampled-data control theory provides an optimal platform for such a problem [17]. We design a filter which minimizes the norm of the error system consisting of a continuous-time delay and a filter with ideal sampler. We show that this design problem is reducible to a discrete-time one by lifting. We also provide an analytical
solution to the case in which such fractional delays are time-varying, under the assumption that the underlying frequency characteristic of the continuous-time input signal is governed by a low-pass filter of first order [18].

2.7 Fractional Delay

The microphone arrays that are implemented for multi-channel speech enhancement should have property of time aligning the incident signal. This means that the setup should be able to delay the signal received by the microphones. In general the delay will be equal to the multiple of integer sample periods. The delay of fraction sample periods can be achieved by using fractional delay filters. As the analog signal is sampled according to the sampling theorem (i.e. \( f_s > f_{max} \)) fractional delay is originated with higher probability.

\[
delay\ in\ samples = \frac{f_sK}{c}
\]  
Eq.2.1

where \( f_s \) is the sampling rate and \( K \) is product of distance between microphones and the angular distance from reference microphone and \( c \) is velocity of sound in the specific medium. The delay contains sum of integer part (‘D’) and the fractional part (‘d’):

\[
delay = \lfloor D \rfloor + d
\]  
Eq.2.2

An impulse function is used for the integer delay. The integer part delay is implemented using an impulse function. From the following figure the sinc function (blue color line) is delayed by 2 units (which is shown by red line).

![Sampled impulse response of ideal fractional delay filter](image)

Fig. 4: Sampled (dots) impulse response of ideal fractional delay filter when the delay is (a) \( D=2 \) sample and (b) \( D=2.7 \) samples.

The vertical dashed line indicates the midpoint of the continuous time impulse response in each case.
2.7.1 Implementation of fractional delay

There are many methods to implement a fractional delay, the basic strategy is based on the Nyquist-Shannon sampling theorem. A sinc interpolator can be used to exactly evaluate a signal value at any point in time. According to Shannon’s sampling theorem, it is possible to recreate the original signal from sampled data by multiplying each sample by a scaled sinc function. Along as it band limited to an upper frequency \( f_s \). Hence, the delay system must be made band limited by using an ideal low-pass filter in the frequency domain. The only condition is that the original waveform is band limited to have a maximum frequency component of less than half the sampling rate.

The impulse response of an ideal FD filter is shifted using sampled sinc function, that is:

\[
y(delay) = \sum_{n=-\infty}^{n=\infty} y(D) * sinc(n - D)
\]

Eq.2.3

The sinc interpolator can be defined as

\[
sinc(n - D) = \frac{\sin(\pi(n-D))}{(\pi(n-D))}
\]

Eq.2.4

The above figure 4 shows the idle impulse response when \( d = 0.0 \) and \( d=0.7 \) samples. The figure 4 shows when \( D \) is an integer i.e., no fractional delay the signal is sampled at zero crossings and when \( D \) is a non-integer the signal is sampled between zero-crossings in which the impulse response is of infinite length. As impulse response is not absolutely summable, the filter is said to be not stable. Therefore, ideal Fractional Delay filter is non-realizable. To realize a Fractional Delay filter some finite length causal approximation filter must be used for non-realizable sinc function [9].

2.7.2 FIR method (using Lagrange interpolator)

There are five different approaches are designed for causal Fractional Delay FIR filters:

1. Windowed Sinc Function (using asymmetric window function with fractional offset) [25].
2. Maximally-Flat FIR approximation (Lagrange Interpolation) [25].
3. Weighted Least Squares (WLS) Approximation [25].
4. Oetken’s Method (a quasi-equiripple Fractional Delay Approximation) [25].
5. Low pass Fractional Delay Approximation with a smooth transition band obtained using low-order spline function [25].

Among all these fir approximations, Legrange interpolation gives maximally-flat FIR approximation. Low pass FD approximation with a smooth transition band obtained
using a low-order spline function followed by Lagrange interpolation and weighted least square approximation.

\[ h(n) = \prod_{k=0}^{N} \frac{D-k}{n-k} \text{ for } n = 0,1,2, \ldots, N. \]  

Eq.2.5

where ‘D’ is the desired total delay and ‘N’ is the order of the filter ‘n’ is the scaling coefficient. Oelken’s method utilizes an odd-length makes a linear phase FIR filter, which can be designed using Remez algorithm.

2.7.3 IIR Method

All pass filters are particularly well suited to FD approximation; because it has a magnitude response is exactly equal to unity for all frequencies. Unlike FIR approximations, allpass filter can be implemented using lower order which translates into computational speed.

The transfer function of a general decrement time all pass filter is given by

\[ H(z) = \frac{a_n z^{-1} a_{n-1} + \cdots + z^{n-1} a_1 + z^{-N}}{1 + z^{-1} a_1 + \cdots + z^{n-1} a_{n-1} + z^{-N} a_n} \]  

Eq.2.6

Where \( N \) is the order of the filter, the numerator is the denominator polynomial, and the filter coefficients \( a_k \) (k = 1, 2, ..., N) are real. The numerator of the all pass transfer function is a mirrored version of the denominator polynomial. The poles (i.e., roots of the denominator) of a stable all pass filter are located inside the unit circle in the complex plane and as a result its zeros (i.e., roots of the numerator) are located outside the unit circle so that their angle is the same but the radius is the inverse of the corresponding pole.

2.7.4 Thiran All Pass Filter Design

An all pass fractional-delay filter with a maximally-flat phase delay models the non-integer delay, D. While the magnitude response of the filter is unity for each frequency, the phase-delay of the filter approximates the fractional delay over a suitable bandwidth. The transfer of this allpass has shown in the equations 2.6 in which For a maximally flat fractional delay D the real valued filter coefficients \( a_k \) can be designed using the closed formula for Thiran all pass filters is represented as below,

\[ a_k = -1^k \binom{N}{k} \prod_{n=0}^{N} \frac{N-n-D}{N-n-n-D} \]  

Eq.2.7

Where

\[ \binom{N}{k} = \frac{N!}{k!(N-k)!} \]  

Eq.2.8
The maximally flat group delay design is the only known FD all pass filter design technique. Here we discuss this solution. Thiran in the year 1971 proposed an analytic solution for the coefficients of an all-pole low-pass filter with a maximally flat group delay response at the zero frequency. Thiran has explained that when \((D-n) > 0\) the denominator polynomial obtained by the above method will have its roots within the unit circle in the complex plane, i.e., the filter will be stable. A drawback of Thiran’s design technique is that the magnitude response of the all-pole low pass filter cannot be controlled. In the all pass design, however, this kind of problem does not exist. Thus it seems that the result of Thiran is better suited to the design of all pass than all-pole filters. The filter with the given coefficients is guaranteed to be stable, and to have maximally-flat phase delay at low frequencies corresponding to the desired fractional delay, \(D\), at the zero frequency. The Thiran approximation can be used to obtain a fractional-delay filter for \(D>N - 1\).

The Thiran all-pole filter can be used for obtaining small delays in which the low pass magnitude response is uncontrolled. The optimal range of ‘\(D\)’ is taken between \(N-0.5\) to \(N+0.5\) [9]. For example, the group delay response with the order number \(N= 15\) is as shown in Figure 5 The group delays are sampled at \(D = N-0.5\) and stopped at \(D = N+0.5\). Therefore, the group delay response in Figure 5 is made between 14.2 and 15.6 samples.

![The group delay of N=15, Thiran maximally flat fractional delay all pass filter.](image)

Fig.5: The group delay of N=15, Thiran maximally flat fractional delay all pass filter.
3. Room Impulse Response (RIR)

3.1 Introduction

In speech communication systems, such as voice-controlled systems, hands-free mobile telephones, and hearing aids, the received microphone are mainly degraded with reverberation, background noise, and other interferences. This degradation may lead to total unintelligibility of the speech and decreases the performance of automatic speech recognition systems.

In the context of this work reverberation is the process of multi-path propagation of an acoustic sound from its source to one or more microphones. The received microphone signal generally consists of a direct sound, reflections that arrive shortly after the direct sound (commonly called early reverberation), and reflections that arrive after the early reverberation (commonly called late reverberation). Reverberant speech can be described as sounding distantly with a noticeable echo and coloration. These detrimental perceptual effects are primarily caused by late reverberation, and generally increase with increasing distance between the source and microphone. Conversely, early reverberations tend to improve the intelligibility of speech. In combination with the direct sound it is sometimes referred to as the early speech component [19].

Microphone systems only exploit the temporal and spectral diversity of the received signal. Reverberation, of course, also induces spatial diversity. To additionally exploit this diversity, multiple microphones must be used, and their outputs must be combined by a suitable spatial processor such as the so-called delay and sum beamformer. It is not a priori evident whether spectral enhancement is best done before or after the spatial processor. Hands-free systems are often used in a noisy and reverberant environment and so the received microphone signal does not only contain the desired signal but also interferences such as room reverberation that are caused by the desired source, background noise, and a far-end echo signal that results from a sound that is produced by the loudspeaker [19]. Usually an acoustic echo canceller is used to cancel the far-end echo. Additionally a post-processor is used to suppress background noise and residual echo, i.e., echo which could not be cancelled by the echo canceller.
To counteract the degradations caused by reverberation, background noise and other interferences, high-performance acoustic signal processing techniques are required. In the context of this work, reverberation is the process of multi-path propagation of an acoustic sound from its source to one or more microphones. Sound is a disturbance of mechanical energy that propagates through matter, e.g., a gas, as a wave. Under the influence of a sound wave, variations of gas density and pressure occur, both of which are functions of time and position. The difference between the instantaneous pressure and the static pressure is called the sound pressure. In this dissertation a microphone is used to transform the pressure (or pressure gradient) present in the air immediately in front of the microphone into an electrical signal. For simplicity we will assume that the microphone is ideal, i.e., that its electrical output is identical (except for a non-dimensionless scaling factor) to the local sound pressure. For this reason we will not distinguish between them in this dissertation.

A major challenge in acoustic signal processing originates from the degradation of the desired signal by the acoustic channel within an enclosed space, e.g., an office room or living room. Because the microphone cannot always be located near the desired source, the received microphone signals are typically degraded by:

(i) Reverberation introduced by the multi-path propagation of the desired sound to the microphones

(ii) Noise introduced by interfering sources.

The main difference between noise and reverberation is that the degrading component in case of reverberation is dependent on the desired signal, whereas in case of noise it can be assumed to be independent of the desired signal. It should be noted that many, if not all, existing acoustic signal processing techniques fail completely or experience a dramatically reduced performance when reverberation is present, e.g., existing source localization and source separation techniques.
Fig. 7: Illustration of a direct path and a single reflection from the desired Source to the microphone.

3.2 Generation of room impulse response

Fig. 8: Illustration of a direct sound, an early reverberation and late reverberation from source to the microphone.
The basic reflection from Direct sound: The first sound received from a source without any reflections is known as direct sound.

Early and late reverberation: The path carried by a reflected sound wave is defined as reflection. The reflected sound from a sound component is known as early reverberation. The early reverberation should not exceed 80-100ms (reflection time taken from source to microphone) time limit. The reflections which exceed the time limit of early reverberation are defined as the late reverberations. These impulse responses are generated from Allen and Berkley’s Image model method which is shown in the figure 8 and figure 9.

3.3 Room Image model

The image model is simulated in a room; it depends on the position of a microphone in a room. Allen and Berkley is explained an efficient method [24]. In this thesis the room impulse response (RIR) is calculated using Image Source Model (ISM). Allen and Berkley were the first researches to implement ISM. In this we find the unit impulse response for each echo with proper time delay. This is achieved by Thiran all pass filter. By using these fractional delay filters, each image source is effectively represented with exact non-integer time delays and Room Transfer Function obtained in frequency domain and the Inverse Fourier transform in time domain also gives the same result [17].
Fig 10: First reflection path of an image source.

In figure 10 shows the path involving the first reflection. This image source ‘S’ is located near to wall, the destination ‘D’ will receive two reflection one is direct path (SD) and another reflected path (SRD). The direct path length is calculated directly. An virtual image is generated next to the wall (S’). from the triangular geometry the distance SR=S’R therefore SRD=S’D.

Fig.11: Reverberated environment with reflected source images.

The figure 11 shows a sound source (green circle) located in a room at 3-D position. Red plus (+) symbol is considered to be origin point or reference point or mid-point of the room and its co-ordinates are assumed to be (0,0,0).every position is measured with reference to the origin of the room. $X_m$ is distance between the microphone and origin. $X_s$ is distance between microphone and origin. $X_r$ is reflecting wall distance from origin. The source 1 and source 2 are the first reflected image sources generated from reverberating image model.

The whole early and late reverberation source image positions $(x_i,y_j,z_k)$ are calculated by using the following equations
Where 'c' is velocity of sound in meters. The \( t_s \) value is estimated for multiple reflections of reverberation. For every reflection there should be some loss of energy which is estimated by using reflection co-efficient (\( \alpha \)) alpha. Calculation of reflection co-efficient and its effect are explained in [10]. From this different reflections and reflection coefficient, the impulse response is plotted in the figure 8. The effect of reverberation for a signal is shown in figure 12. In this figure the red signal indicates the original speech signal. The reverberant signal (blue colored) is amplified due to the addition of reflection energy at particular unit sample.

\[
x_i = (-1)^i x_s + \left[ i + \frac{1-(-1)^i}{2} \right] x_r - x_m \quad \text{Eq.3.1}
\]
\[
y_j = (-1)^j y_s + \left[ j + \frac{1-(-1)^j}{2} \right] y_r - y_m \quad \text{Eq. 3.2}
\]
\[
z_k = (-1)^k z_s + \left[ k + \frac{1-(-1)^k}{2} \right] z_r - z_m \quad \text{Eq.3.3}
\]
\[
d_{ijk} = \sqrt{x_i^2 + y_j^2 + z_k^2} \quad \text{Eq.3.4}
\]
\[
t_s = \frac{d_{ijk}}{c} \quad \text{Eq.3.5}
\]
4. MICROPHONE ARRAY BEAMFORMING

4.1 Need of Microphone Array

In acoustic speech enhancement, the microphone array processing has the potential to perform spatial sensitivity. This leads to the directional hearing using beamformers, which reduces the level of directional and ambient noise signals while minimizing distortion to speech from desired direction. In real world scenario, generally the desired and interfering signal arrive from different spatial direction. The interfering signal will corrupt the desired signal. So there is a need of spatial filtering. A beamformer acts as spatial filter in order to separate the desired signal from the corrupted signal using microphone array. That is why beamforming technique is most popular method to track a radiant signal spatially from a specific direction and attenuate the remaining spatial signal in other directions. With this operation the beamformer can be called as the spatial filter [19]. A beamforming microphone's process begins with acquiring audio feed from two or more adaptive directional microphones placed spatially apart. This adaptation allows the beamforming microphone to focus on sounds that originate directly from the area to where the microphone needs to be pointed. These adaptive directional microphones tend to record less ambient and room reverberation noise than a microphone with a larger field range. Using different audio feeds that are acquired from the microphones, the feed is sent through different processors to adjust the sound quality. Beamforming microphones use a relatively new sound processing technology, which can be used in hearing aids to better isolate a sound and its location. Beamforming technology can also help filter out excess room noise [19]. This technology is useful when recording professional audio in the studio, but it can also be used for simple applications, such as optimizing the sound in your home theater.

The processor then makes an analysis of the acoustic properties of the audio situation, taking into account conditions such as the amount of echoing in the room. The beamforming processor uses scientific knowledge about the quality of the interaction of sound called phasing. Phasing occurs due to the small difference in the sound feed which is coming from each microphone. Two identical microphones are placed with a smaller distance values to record the same audio signal. The microphone array beamforming technology calculates the differences in phase for each microphone with the knowledge of a reference microphone. This process is called geometric beamforming [1].

After the sound is processed to improve its clarity using beamformer, it is transmitted back to the user as one or more audio signals. In a hearing aid, this sound is transmitted back in the hearing aid. The resulting audio should be clearer and easier to hear, and the audio should be more focused, zeroing in on the sounds in the direct path of the microphones.
Beamforming is a revolution that has started with multi-microphone systems. In this, directional arrays are used for the adaptive and fixed beamforming techniques. The adaptive beamformers are track variations as well as to compensate for model mismatch. These types of beamformers steers the directivity on the changes in environment which leads the maximum gain at the source position. Whereas the fixed beamformers are maximize the output of the array in a fixed direction. Fixed beamformers doesn’t steer’s according to the source position. The examples are as follows:

**Adaptive beamformers**
- adaptive beamforming
- Elko’s beamformer

**Fixed Beamformers**
- Wiener beamformer
- Optimal-SNIR beam former
- Delay-and-sum beamformers

### 4.2 Common signal modeling

The section describes about signal modeling and the procedure to construct a microphone array. The speech signal is fed to ‘D’ linear microphones in anechoic environment. The transformation of ‘D’ linear microphone array elements is given as $W(\omega)$. The signal received by the ‘$i^{th}$’ microphone array as given as

$$y_{i} = \sum_{d=1}^{D} S_{d}(t) * W_{d,i}(t) + V(t)$$  \hspace{1cm} \text{Eq.4.1}$$

Fig.13: Microphone array model.
Where the terms $y_i(t)$ is the output of $i^{th}$ microphone and $V(t)$ is the environmental random noise. With this method the input signal is emphasized based on the variance of output signal. Variance and power are directly proportional to each other [1] from the equation 4.1 it is clear that the output of ‘D’ microphone array contain a combination of speech and background noisy signal.

### 4.3 Adaptive Filtering

The optimum weighted vector of a noisy speech signal ($X(n)$) and the desired speech signal ($S(n)$) is generated by wiener solution. The input signal with length ‘$l$’ is defined as the $X(n, l)$ and the desired signal is defined as $S(n, l)$.

$$W_{opt}(n) = R_{xx}^{-1}(n)r_s(n)$$  \hfill Eq.4.2

Where $R_{xx}(n)$ is auto covariance vector and $r_s(n)$ is cross covariance vector. This is defined as

$$R_{xx}(n) = E[X(n, l) . X(n, l)^*];$$  \hfill Eq.4.3

$$r_s(n) = E[X(n, l) . S(n, l)^*];$$  \hfill Eq.4.4

‘$E[ ]$’ is an expectation operator and ($.$)$^*$ is conjugate operator.

### 4.4 Elko’s Beamformer

The qualitative analysis of a one dimensional acoustic speech signal is estimated based on noise in the channel. The optimal noise estimation is essential and for this the directional microphones are designed. Directional microphone arrays can be selective in combating these problems.

This particularly attractive microphone design for personal communicators and teleconferencing with differential microphone array uses sensors that are spaced very closely compared to the acoustic wavelength. To realize directionality, the elements are combined in an alternating sign fashion and as a result of the close-spacing, can be seen to be a differential array. The resulting output of differential sensor is super directional since the directivity is higher than that of the uniformly summed output of all the sensor elements. Typically, optimal directional microphones developed are designed under the assumption that the acoustic reverberation is isotropic. Unfortunately, real acoustic noise never matches ideal theoretical assumptions.

A better potential solution is to design an Elko’s adaptive microphone system that adjusts its directivity pattern to maximize the signal-to-noise ratio. Much work has been done on adaptive beamforming in the past three decades; however, little attention has been given to adaptive differential arrays. One reason for the lack of attention towards differential
sensors is the well-known fact that super directional arrays are extremely difficult to realize practically. However, if the differential order of the sensor is limited to first or second order, then practical designs do exist.

A differential microphone array is arranged, that minimizes the microphone output power adaptively in the rare half plane. This is constructed with array of 2 back to back cardioid microphones. In an array of microphones, the average center valued microphone is considered as the reference microphone. A typical differential microphone array is shown in figure 14.

The fractional delay is designed for each microphone in such a way that it nullifies the delay of the speaker direction. The outputs of three microphones are equal due to accurate calculation of FD ($T=d/c$); in spectral view the output can be written as

$$Y(\omega, \theta) = S(\omega)(1 - e^{-j(\omega T + k d)})$$

$$Y(\omega, \theta) = S(\omega)(1 - e^{-j(\omega T + (d \cos \theta) / c)})$$

Where $d$ is the distance between microphones. $T$ is equal delay applied to the signal. The $K$ is the directional coefficient and defined as $K = \text{angular distance} (\omega)/c$; ‘c’ is the velocity of sound. $K = d \times \cos(\theta)/c$;

![Diagram of differential microphone array](image)

**Fig.14:** Back-to-back cardioid microphones.
The distance between the microphones in an array is chosen so as to avoid the spatial aliasing. In a set of two microphone array two cardioid microphones are arranged back to back. The polar pattern of these microphones is shown in figure 16.

The sampling period is maintained to $T = \frac{d}{c}$ to get a cardioid output as shown in Figure 16. These back-to-back cardioid first–order differential microphones are arranged as shown in the figure 14. The output of $C_F$ and $C_B$ can be written as following equations:

$$C_F(\omega, \theta) = 2j * S(\omega)e^{\frac{j\omega T}{2}} \frac{kd(1+\cos \theta)}{\sin kd(1+\cos \theta)}$$  \hspace{1cm} \text{Eq.4.5}$$

$$C_B(\omega, \theta) = 2j * S(\omega)e^{\frac{j\omega T}{2}} \frac{kd(1-\cos \theta)}{\sin kd(1-\cos \theta)}$$  \hspace{1cm} \text{Eq.4.6}$$

The major strength of a speech signal in $C_F$ and $C_B$ is varied depending on forward and backward directions which have showed in the form of Power spectral densities in figure 18 and figure 19. The low pass filter shown flowing the output ($y(t)$) is used to compensate the differentiator response of the differential microphone. For more qualitative explanation please refer to [8].

The output of signal given by

$$Y = C_F - \beta C_B$$  \hspace{1cm} \text{Eq 4.7}
The optimal $\beta$ minimizes the mean square error of the sensor output. The optimal $\beta$ estimation leads better results of Elko’s algorithms.

Assume that reference Microphone Position is at the distance of $d/2$ from mike 1 and mike 2.

Fig. 17: Schematic implementation of an adaptive first-order differential microphone using back-to-back cardioids.

Fig. 18: Power spectral density of $C_B$ (blue line) and $C_F$ (green line) speech signals for the source angle 45 degrees and noise angle 135 degrees.
Fig. 19: Power spectral density of $C_B$ (blue line) and $C_P$ (green line) speech signals for the source angle 45 degrees and noise angle 225 degrees.

4.5 Wiener Beamforming (WBF)

Beamforming techniques exploit fundamental properties about the spatial and/or temporal distribution of both the speech and noise sources, in order to enhance perception. Fixed beamformers are fundamentally based on modeled assumptions on the speech signal and noise field. Based on this model, the concept of adaption is originated when there is an unknown or random noise. This is a powerful tool to reduce the different unknown-noise signals in echo-environment. Wiener beamforming is one such fixed beamforming technique. It is also known as Mean Square Error Beamformer.

The optimal Minimum Mean Square Error (MMSE) Beamformer is defined as the beamformer weights which minimize the mean square difference between the beamformer output when all sources are present, and an observation, when only the signal of interest is present [9].

The objective can be formulated as

$$w_{opt} = \arg \min E\{[y[n] - s_i[n]]^2\} \quad i \in \{1, 2, \ldots, I\}$$

Eq. 4.8

Where $y[n]$ is defined as

$$y[n] = \Sigma_{i=0}^{I} \Sigma_{j=0}^{L-1} w_i[j] s_i[n - j]$$

Eq. 4.9

Where $L - 1$ is the order of the FIR filters and $w_i[j]$, $j = 0, 1, \ldots, L - 1$, are the FIR filter taps for channel number ‘i’. The signals, $s_i[n]$, are digitally sampled microphone
observations and the beamformer output signal is denoted \( y[n] \). \( s_i[n] \), is one of the microphone observations or a separate reference microphone, chosen as the reference sensor. In theory the true source signal, \( s[n] \), would be desirable to use instead of a sensor observation, but the source signal is practically not accessible. The optimal weights which minimize the expected square difference between the output and the reference signal is found by,

\[
w_{opt} = [R_{ss} + R_{nn}]^{-1} r_s \tag{4.10}
\]
due to the linear property of the expectation operator. The cross correlation vector, \( r_s \), is defined as:

\[
r_s = [r_1 \ r_2 \ \ldots \ r_L] \tag{4.11}
\]

Where

\[
\begin{align*}
  r_i &= [r_i[0] \ r_i[1] \ \ldots \ r_i[L - 1] \\
  i &= 1, 2, \ldots, L
\end{align*}
\tag{4.12}
\]

With each element as

\[
  r_i[k] = E\{s_i[n]s^*_i[n + k]\} \quad i = 1, 2, \ldots, L \quad r \in \{1, 2, \ldots, L\} \quad k = 0, 1, \ldots, L - 1 \tag{4.13}
\]

And \( R_{ss} \) and \( R_{nn} \) is defined as

\[
\begin{align*}
  R_{ss} &= \begin{bmatrix}
    r_{s_i0} & \cdots & r_{s_iL-1} \\
    \vdots & \ddots & \vdots \\
    r_{s_iL-1} & \cdots & r_{s_i0}
  \end{bmatrix} \tag{4.14} \\
  R_{nn} &= \begin{bmatrix}
    r_{n_i0} & \cdots & r_{n_iL-1} \\
    \vdots & \ddots & \vdots \\
    r_{n_iL-1} & \cdots & r_{n_i0}
  \end{bmatrix} \tag{4.15}
\end{align*}
\]

The cross correlation vector, \( r_s \), is equivalent to one column of the source signal correlation matrix, \( R_{ss} \), if the reference sensor is chosen as one microphone observation. Which column is used, depends on which one of the microphones is chosen as the reference microphone and which lag one wish to have as center lag. The weights, \( w \), are arranged in the same way as

\[
\begin{bmatrix}
  w_1^T \\
  w_2^T \\
  \vdots \\
  w_L^T
\end{bmatrix}^T \tag{4.16}
\]

\[
  w_i^T = [w_i[0] \ w_i[1] \ \ldots \ w_i[L - 1]] \quad i = 1, 2, \ldots, L \tag{4.17}
\]

The complete system is implemented in anechoic environment in [9]. The reverberation environment is explained in [10].
5. Optimal Signal-to–noise Interference Ratio Beamformer (SNIR)

The output signal-to-noise plus interference power ratio (SNIR) is defined as

\[ Q = \frac{\text{Average signal output power}}{\text{Average noise–pulse–interference output power}} \quad \text{Eq.5.1} \]

Q, is the optimal Signal-to-Noise plus Interference Beamformer, (SNIB). We need to express the mean signal output power as a function of the filter weights in the beamformer, and find the optimal weights, which maximizes Q. The involved signals are all assumed to be short time stationary [1]. In this fixed beamformer the optimal weights are calculated by Eigenvalue decomposition which optimally which maximizes the SNIR.

We consider a signal model where one speaker is situated in a fixed position with different input SNRs and the noise environment consist of several sources, both fixed point sources and disturbing sources, which can be modeled as a mixture of both coherent and incoherent noise fields. The output of sensor number ‘i’ consists of a speech signal component \( s_i[n] \), and a sum of fixed point noise sources, \( V_{id}[n], d = 1, \ldots, D \), together with the mixture of coherent and incoherent noise source \( V[n] \), as

\[ x_i[n] = s_i[n] + v_{id}[n] + v_i[n]. \quad \text{Eq.5.2} \]

Where \( s_i[n], v_{id}[n], d = 1, \ldots, D \), and \( v_i[n] \) are the \( i \)th microphone observations of the signals, respectively. We wish to construct the filters, \( w_i[j] \), in such a way that the output of the beamformer, \( y[n] \), resembles the signal component, \( s[n] \), while the disturbing components are attenuated or cancelled.

Fig.20: Fixed optimal SNIR beamformer.
Where the output \((y[n])\) is defined as;
\[
y[n] = \sum_{l=0}^{L-1} \sum_{j=0}^{J-1} w_i[j] x_i[n-j]
\]  \text{Eq.5.3}

So the power of a zero-lag clean speech pulse interference signal \(s[n]\) can be expressed as;
\[
\sigma_{s[n]}^2 = E\{y_s[n]y_s^*[n]\} = \sum_{l=0}^{L-1} \sum_{j=0}^{J-1} w_i[j] r_{s[n]}[k-l] w_j^*[l]
\]  \text{Eq.5.4}

In the equation (5.4) the term “\(r_{s[n]}[k-l]\)” denotes the cross-correlation function between reference microphone (‘i’”) and the delayed microphone (‘j’) for speech signal \(s[n]\). In the matrix notation this equation (5.4) can be rewritten as;
\[
r_{y_s y_s}[0] = w^HR_{ss}w
\]  \text{Eq.5.5}

Where \(w\) denotes the hermitian transpose matrix and \(R_{ss}\) is defined as
\[
R_{ss} = \begin{bmatrix}
R_{s_1s_1} & \cdots & R_{s_1s_J} \\
\vdots & \ddots & \vdots \\
R_{s_Js_1} & \cdots & R_{s_Js_J}
\end{bmatrix}
\]  \text{Eq.5.6}

Where
\[
R_{s_is_j} = \begin{bmatrix}
r_{s_is_j}[0] & \cdots & r_{s_is_j}[L-1] \\
\vdots & \ddots & \vdots \\
r_{s_is_j}[L-1] & \cdots & r_{s_is_j}[0]
\end{bmatrix}
\]  \text{Eq.5.7}

\[
w = \begin{bmatrix}
w_1^T \\
w_2^T \\
\vdots \\
w_J^T
\end{bmatrix}
\]  \text{Eq.5.8}

In the same the power of a zero lag clean speech pulse interference signal \(N[n]\) can be expressed as;
\[
\sigma_{n[n]}^2 = \begin{bmatrix}
r_{n_1n_1}[0] & \cdots & r_{n_1n_J}[L-1] \\
\vdots & \ddots & \vdots \\
r_{n_Jn_1}[L-1] & \cdots & r_{n_Jn_J}[0]
\end{bmatrix}
\]  \text{Eq.5.9}

Where the correlation function \(r_{nij}[k]\), is the cross correlation between microphone observation i and j, when all disturbing noise-and interference-sources are active alone, from equation 5.2
\[
r_{n_1n_j}[k] = \{\sum_{d=1}^{D} v_{id}[n] + v_i[n] + \sum_{m=1}^{M} v_{jm}[n+k] + v_j[n+k]\}
\]  \text{Eq.5.10}

Where ‘k’ in the range of \(\{0……L-1\}\)

The optimum weights are calculated by maximizing the ratio of two quadratic forms.
\[
w_{opt} = \max_{w} \frac{w^HR_{ss}w}{w^HR_{nn}w}
\]  \text{Eq.5.11}

5.1 Eigen-decomposition of autocorrelation matrix

Maximizing a ratio between two quadratric forms of positive definite matrices as
\[
w_{opt} = \max_{w} \frac{w^HR_{ss}w}{w^HR_{nn}w}
\]  \text{Eq.5.12}

Let the eigenvalues (\(R_{ss}\)) are \(\lambda_j \{j=1……P\}\). Which are ordered in decreasing order let the corresponding Eigen vectors are \(V_i \{i=1,……,M\}\). In the absence of noise the
Eigen values are $\lambda_j \{j=1\ldots P\}$ while $\lambda_{p+1}=\lambda_{p+2}=\ldots \lambda_M=0$; furthermore, it follows that the signal correlation matrix can be expressed as:

$$R_{ss} = \sum_{i=1}^{p} \lambda_i V_i V_i^H$$

Eq.5.13

Thus, the Eigen vectors $V_i \{i=1,2,\ldots P\}$ span the signal subspace as do the signal vectors $S_i \{i=1,2,\ldots P\}$. The ‘P’ eigenvectors for the signal subspace are called principal eigenvectors and the corresponding Eigen values are called principal Eigen values.

In the presence of noise the noise auto correlation can be written as:

$$R_{nn} = VV^H$$

Eq.5.14

For real world noises:

$$V = R_{nn}^{1/2} \cdot w$$

Eq.5.15

From the equation (5.12)

$$V_{opt} = MAX \left\{ \frac{(V^H R_{nn}^{-H/2} R_{rr} R_{nn}^{-1/2} V)}{V^H V} \right\}$$

Eq.5.16

Where the solution $V_{opt}$ is the eigenvector (which belongs to the maximum eigenvalue $\lambda_i$) in the numerator. This is equivalent to meet the following relation the final output can be written as:

$$W_{opt} = R_{nn}^{(1/2)} V_{opt}$$

Eq.5.17

The square root of the matrix is easily found from the diagonal form of the matrix. In general the optimal vector can only be found by numerical methods and the time domain formulation is there for in general more numerically sensitive since the dimension of the weight space is $L$ times greater than the dimension of the frequency domain weight space. But during the implementation there will be very poor SNRI improvement with “$W_{opt}$” this problem can be avoided by using the modified version of equation 5.17 as:

$$W_{opt} = V_{opt}$$

Eq.5.18

From Equation 5.18 the SNRI results will be improved and the PESQ will be same in both Equations (Equation 5.17 and 5.18).
5.2 Modifications for optimal SNIR in reverberant case

![Illustration of reverberated sound in a closed environment.](image)

As shown in the figure 21, in a closed environment the position of microphone are to be estimated properly in 3-D(x,y,z). Let we take a reference microphone (Mic1) and the remaining microphones' signal are delayed with fractional delay to resemble the reference microphone.

As we discussed before in section 5.1 the optimal weights are used to filter the corrupted signal by suppressing noise. In this situation the room inherently contains the reverberation effect which we also treated as noise signal, therefore the equation 5.12 can be modified as

$$w_{opt} = \max \alpha f \frac{w^H R_{xx} w}{w^H (R_{xx} + R_{rr}) w}$$  \hspace{1cm} \text{Eq.5.19}

In the equation 5.19 the power of the numerator (signal power) is enhanced while the power of denominator (reverberated noise power) is suppressed due to eigen decomposition of autocorrelation matrix. The amount of reverberation present in a signal can be evaluated by a factor reverberation index (RR), by estimating the input and output reverberation index the suppression of reverberation can be estimated which are explained in implantation and results.
6. IMPLEMENTATION

6.1 Introduction

This thesis is a collaborative work of four members in which there are four sub-topics implemented. These topics are divided to each member in this group. The main idea of this thesis is to eliminate all types of disturbances that occur during the hands-free speech communication. The disturbances are discussed in detail in previous chapters. In order to eliminate the disturbances i.e. to enhance the speech quality various Beamformers such as Elko’s Beamformer, Wiener Beamformer, optimal SNIR Beamformer. Among which, this thesis is mainly focused on design and implementation of optimal SNIR beamformer in anechoic and reverberated environment.

6.2 Implementation

![Diagram of beamformer implementation]

Fig. 22: The experimental setup for validation of Optimum beamformer model

The experimental setup of an reverberated environment is shown in the figure 23. In this figure s(n) is the pure speech signal, R(n) is the reverberated signal and d(n) is the noise signal. These s(n) and d(n) are delayed using Thiran All pass fractional delay (FD) filters. The x(n) is the microphone input to the beamformer. The number of microphones can be varied and given to the beamformer model. The x(n) is the input to the beamformer which is a sum of fractionally delayed s(n) and d(n). In the optimal SNIR beamformer Rss and Rnn matrices are generated and optimum filter weights are generated. These generated optimum
filter weights ‘\(w_{opt}\)’ are passed through a filter and gets multiplied with input signal \(x(n)\). The output of the beamformer is \(y'(n)\). Which is explained in section 5.

### 6.3 Terms and expansions

The output \((y'(n))\) (from figure.23) of an optimal SNIR Quality is estimated by signal to noise ratio (SNR), speech Distortion (SD), noise Distortion (ND) and Perceptual Evaluation Speech Quality (PESQ). The Input SNR, Output SNR are calculated in different noise environments and validated based on the improvement of SNR. The performance is measured based on the SNR improvement in different noisy environments. Also, the Input PESQ, Output PESQ is evaluated in different noisy environments. The PESQ improvement in different noise environments are calculated based on that the performance is measured. PESQ ranges from \(-0.5 < \text{PESQ} < 4.5\).

#### 6.3.1 Input signals

A desired clean speech signal (Speech_all.wav) with sampling frequency 16kHz, duration of 11seconds is used to test optimal SNIR module, in reverberated and anechoic environment with four noise signals namely Wind noise (WN), White Noise (WHN), Factory Noise (FN), Interference Noise (IN). These entire noise signals are tested for 0dB, 5dB, 10dB, 15dB, 20dB and 25dB input SNR levels. The Input (Speech_all.wav) signal contains four sentences with female and male voice alternatively. These four sentences are mentioned in table.1

<table>
<thead>
<tr>
<th>File Name</th>
<th>Duration In sec</th>
<th>Type of Voice</th>
<th>Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech_all.wav</td>
<td>3</td>
<td>Female</td>
<td>“It’s easy to tell the depth of the well.”</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Male</td>
<td>“Kick the ball straight and follow through.”</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Female</td>
<td>“Glue the sheet to the dark blue background.”</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Male</td>
<td>“A part of tea helps to pass the evening.”</td>
</tr>
</tbody>
</table>

Table 1: The details of Clean Speech Signal used for Evaluation

The performance of the system is measured with SNR, speech distortion (SD), noise distortion (ND) and PESQ. The noise files are having sampling rate 16 kHz. These signals are multiplied by a factor alpha \((\alpha)\) to get desired input SNR \((\text{SNR}_d)\).

\[
\alpha = \sqrt{\frac{1}{10^{\frac{\text{SNR}_d - \text{SNR}_{in}}{10}}}}
\]

Eq.6.1
The power spectral density of four noise signals is plotted in the following figure as;

Fig.23: Power spectral density of wind, white, interference and factory noise signals

6.4 Objective measures

Figure describes the calculation of objective measure at the input and output of the SNIR system that is implemented any improve in the objective measure can be calculated by subtracting the output value from the input value. The objective measure can be any one of the many available objective measures. A few of them which are used in this works are described in the following sections.
6.4.1 Signal to noise ratio (SNR)

The Signal-to-Noise Ratio (SNR) is used to measure the power of desired signal level $s(n)$ with the power of noise signal $v(n)$. The ideal method for calculating the SNR is the amount of speech energy over the noise energy after the enhancement method. The input SNR and output SNR are calculated as below;

$$SNR_{in} = \frac{|s(n)|^2}{|v(n)|^2}$$  \hspace{1cm} \text{Eq. 6.2}

$$SNR_{out} = 10 \cdot \log_{10}\left(\frac{|S_{out}|^2}{|v_{out}|^2}\right)$$  \hspace{1cm} \text{Eq. 6.3}

$$SNRI = SNR_{out} - SNR_{in}$$  \hspace{1cm} \text{Eq. 6.4}

Where $S_{out}$ filtered Speech output of SNIR module and $N_{out}$ are the filtered noise outputs from SNIR module. $S(n)$ and $N(n)$ are the input signals.

6.4.2 Speech and Noise Distortion (SD) & (ND)

The SD is defined as the spectral deviation in the power of input clean speech signal and the power of the processed speech signal at the output. The speech distortion (SD) is represented as;

$$Speech\ Distortion\ (SD) = 10 \cdot \log_{10}\left(\int_{-\pi}^{\pi} P_s(w) - P_{\tilde{s}}(w) dw\right)$$  \hspace{1cm} \text{Eq. 6.5}

Where the power of clean speech signal is $P_s(w)$ at input and $P_{\tilde{s}}(w)$ is the power of the processed speech signal at output. The ND is defined as the spectral deviation in the power of input noise signal and the power of processed noise signal at the output. The ND is represented as;

$$Noise\ Distortion\ (ND) = 10 \cdot \log_{10}\left(\int_{-\pi}^{\pi} P_n(w) - P_{\tilde{n}}(w) dw\right)$$  \hspace{1cm} \text{Eq. 6.6}

Where $P_n$ the power of pure noise is signal at input and $P_{\tilde{n}}$ is the power of processed signal at output.

6.4.3 PESQ Score

PESQ return a quality score, known as PESQ score, which conforms to ITU-T P.862. PESQ score lies on a scale from -0.5 to 4.5, through in most cases it is between 1 and 4.5. PESQ score correlates with subjective quality scores. However the PESQ score tends to be optimistic for poor quality speech and pessimistic for good quality speech. Alternative mapping for PESQ score have been developed which do not exhibit a better correlation to subjective test scores [14].
PESQ score are closer to the listening quality subjective opinion scale, which is standard in the industry and is defined in ITU-T P.800. This is reproduced in Table 2. Along with the prompt that is given to subjects. Listening quality scores lie in between 1 and 4.5 in subjective test.

In order to execute the PESQ following command is used

Syntax of PESQ:

```
“PESQ” “PESQ +16k/8k referencespeech.wav testspeech.wav”
```

Where +16k and 8k are sampling rates of the speech signals i.e. reference speech and test speech signal.

<table>
<thead>
<tr>
<th>PESQ Quality Score</th>
<th>4.5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Excellent</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
<td>Bad</td>
</tr>
</tbody>
</table>

Table 2: PESQ Quality scale.
6.4.4 Reverberation index (RR)

This reverberation is an index which can estimate the amount of reverberation present in the signal. From the impulse response diagram it is clear that maximum power accumulated in the early reverberation. So this power estimation helps us to find the amount of reverberation.

So the reverberation index can be formulated as

\[ RR = 10 \log_{10} \left( \frac{(Late\ reverbaration)^2}{(Early\ reverbaration)^2} \right) \]  

Eq. 6.7

RRin, RRout can be estimated using same formulas. The RR improvement indicates the RIR suppression in the given signal.

6.4.5 Test Signals

The test reverberant signals are generated from two reverberant environments. One signal is generated in a closed room with dimensions [2*2*2] (in meters), source and microphone positions are at [2,1,2] (corresponds to (x,y,z) positions in free space), [1.6,1,1.3] (corresponds to (x,y,z) positions in free space) respectively.
The second reverberant signal is generated in a room with dimensions [5*10*7.5] (in meters), source and microphone positions are at [1,0.1,1.6] (corresponds to (x,y,z) positions in free space), [1.5,1,1] (corresponds to (x,y,z) positions in free space) respectively.

Comparing these two cases even if the size of room increases the room impulse response changes only at the late reverberation part which is insignificant. Hence the first room impulse response is used for testing.

Fig.28: Illustration of original and reverberant speech signal of room size [5*10*7.5].
7. RESULTS

7.1 Performance evaluation of Optimal SNIR Beamformer in Anechoic and Reverberated environment

7.1.1 Anechoic Optimal SNIR Beamformer.

The optimal SNIR beamforming method is tested with an input Speech signal (Speech_all.wav) in an anechoic environment. The optimal weights of SNIR Beamformer are estimated using the following equation:

\[ w_{opt} = \max \frac{w^{H}R_{ss}w}{w^{H}R_{nn}w} \]  

Eq. 7.1

The main objective of optimal SNIR algorithm is to improve the numerator power \((w^{H}R_{ss}w)\) and decrease the denominator power \((w^{H}R_{nn}w)\). This can be achieved by Eigen-decomposition of autocorrelation matrix which is more clearly explained in chapter 6.

Results is taken for the fixed direction (0 and 75 degrees) of source and noisy environments respectively which are shown in figure 30. Distance between microphones is assumed to be D=0.02 (In meters)

Fig. 29: Illustration of source (at zero degrees) and interference (at 75 degrees) positions.

In the assumed anechoic room, optimal SNIR Method gives an average of 183.8161 dB Signal-to-Noise Ratio Improvement (SNRI), -27.5431 dB speech distortion (SD), -32.1087 dB noise distortion (ND) and 2.007 PESQ Improvement (PESQI) for Wind noise environments in all situations for the clean speech signal (Speech_all.wav). The SNRI, SD, ND, PESQI values are shown in the Table.3. All these parameter are recorded by improving the input signal-to-noise ratio.
**Observation:** Every value is recorded for a fixed DOA (with source angle at zero (0°) and interference at seventy five (75°)) and fixed microphone distance D=0.02 (In meters). optimal SNIR module is very good in suppression of noise signal in an anechoic environment. PESQ value improves as the number of microphone are increased and PESQ ranges from (4.2 - 4.4), the average SNRI for 2 microphone is around 211 dB, for 4 microphones its around 188 dB and for 6 microphones it is observed to be around 147 dB. From the above observations it can be concluded that as the number of microphones increases SNRI decreases for a given input SNR level. The average PESQI is approximately 2.0 for 2, 4 and 6 microphones. SD ranges between (-26 dB to -30dB) and ND ranges between (-23dB to -50dB). Results are shown in the following table:

<table>
<thead>
<tr>
<th>SNR in Mics</th>
<th>SNRout</th>
<th>SNR in dB</th>
<th>SD (dB)</th>
<th>ND (dB)</th>
<th>PESQ</th>
<th>PESQout</th>
<th>PESQI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 2</td>
<td>220.1256</td>
<td>-26.6660</td>
<td>-26.5399</td>
<td>4.570</td>
<td>2.030</td>
<td>2.633</td>
<td></td>
</tr>
<tr>
<td>0 3</td>
<td>219.8268</td>
<td>-26.7770</td>
<td>-26.6479</td>
<td>4.326</td>
<td>2.756</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 4</td>
<td>181.8678</td>
<td>-27.0457</td>
<td>-26.9083</td>
<td>4.312</td>
<td>2.742</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 5</td>
<td>156.4305</td>
<td>-30.7404</td>
<td>-30.4254</td>
<td>4.285</td>
<td>2.715</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 6</td>
<td>144.9305</td>
<td>-29.3264</td>
<td>-29.0967</td>
<td>4.282</td>
<td>2.712</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 7</td>
<td>221.5220</td>
<td>-27.0141</td>
<td>-29.3994</td>
<td>4.319</td>
<td>2.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 8</td>
<td>211.0938</td>
<td>-29.2765</td>
<td>-31.5169</td>
<td>4.340</td>
<td>2.471</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 9</td>
<td>187.4076</td>
<td>-27.2225</td>
<td>-30.3807</td>
<td>4.327</td>
<td>2.458</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 10</td>
<td>149.1808</td>
<td>-26.5667</td>
<td>-34.9738</td>
<td>4.264</td>
<td>2.395</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 11</td>
<td>149.9375</td>
<td>-26.6363</td>
<td>-35.4805</td>
<td>4.337</td>
<td>2.468</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 12</td>
<td>211.8459</td>
<td>-26.1904</td>
<td>-39.9613</td>
<td>4.317</td>
<td>2.134</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 13</td>
<td>223.5826</td>
<td>-26.2680</td>
<td>-38.4980</td>
<td>4.286</td>
<td>2.103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 14</td>
<td>201.3492</td>
<td>-26.9872</td>
<td>-32.9936</td>
<td>4.321</td>
<td>2.138</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 15</td>
<td>151.1916</td>
<td>-26.1513</td>
<td>-41.0670</td>
<td>4.203</td>
<td>2.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 16</td>
<td>160.6546</td>
<td>-28.0206</td>
<td>-23.8929</td>
<td>4.273</td>
<td>2.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 17</td>
<td>227.9405</td>
<td>-26.9318</td>
<td>-31.7312</td>
<td>4.313</td>
<td>1.848</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 18</td>
<td>225.3203</td>
<td>-27.5137</td>
<td>-30.3299</td>
<td>4.320</td>
<td>1.855</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 19</td>
<td>205.9189</td>
<td>-29.1383</td>
<td>-28.3015</td>
<td>4.308</td>
<td>1.843</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 20</td>
<td>170.5892</td>
<td>-28.1592</td>
<td>-29.3173</td>
<td>4.345</td>
<td>1.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 21</td>
<td>164.1655</td>
<td>-29.4741</td>
<td>-28.0428</td>
<td>4.332</td>
<td>1.867</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 22</td>
<td>234.3597</td>
<td>-27.1987</td>
<td>-30.6869</td>
<td>4.324</td>
<td>1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 23</td>
<td>238.8777</td>
<td>-27.8648</td>
<td>-29.4896</td>
<td>4.300</td>
<td>1.576</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 24</td>
<td>220.4305</td>
<td>-27.2483</td>
<td>-30.5783</td>
<td>4.380</td>
<td>1.656</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 25</td>
<td>192.4875</td>
<td>-30.1450</td>
<td>-27.4653</td>
<td>4.313</td>
<td>1.589</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 26</td>
<td>167.3024</td>
<td>-27.6628</td>
<td>-29.8010</td>
<td>4.316</td>
<td>1.592</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 27</td>
<td>236.2333</td>
<td>-27.3640</td>
<td>-30.2486</td>
<td>4.289</td>
<td>1.289</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 28</td>
<td>236.2333</td>
<td>-26.8118</td>
<td>-31.5766</td>
<td>4.401</td>
<td>1.401</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 29</td>
<td>205.5460</td>
<td>-26.6490</td>
<td>-32.1042</td>
<td>4.281</td>
<td>1.281</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 30</td>
<td>190.8766</td>
<td>-27.0515</td>
<td>-30.9286</td>
<td>4.326</td>
<td>1.326</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 31</td>
<td>170.8536</td>
<td>-26.1862</td>
<td>-34.2859</td>
<td>4.314</td>
<td>1.314</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: SNR, SD, ND, PESQ, SNRI and PESQI for clean Speech signal with wind noise in anechoic environment.

In the assumed anechoic room optimal SNIR Method gives an average of 205.6 dB Signal-to-Noise Ratio Improvement (SNRI), -29.5449 dB speech distortion (SD), -30.3844 dB Noise distortion (ND) and 1.620 PESQ Improvement (PESQI) for factory noise environments in all situations for the clean speech signal (Speech_all.wav). The SNRI, SD, ND, PESQI values are shown in the Table.4. All these parameter are recorded by improving the input signal-to-noise ratio.
Observation: Every value is recorded for a fixed DOA (with source angle at zero (0°) and interference at seventy five (75°)) and fixed microphone distance D=0.02 (In meters). Optimal SNIR module is very good in suppression of noise signal in an anechoic environment. PESQ value improves as the number of microphone are increased and PESQ ranges from (4.2 - 4.4), the average SNRI for 2 microphone is around 248 dB, for 4 microphones its around 209 dB and for 6 microphones it is observed to be around 161 dB. From the above observations it can be concluded that as the number of microphones increases SNRI decreases for a given input SNR level. The average PESQI is approximately 1.6 for 2, 4 and 6 microphones. SD ranges between (-26 dB to -44dB) and ND ranges between (-24dB to -40dB). Results are shown in the following table:

<table>
<thead>
<tr>
<th>SNR in</th>
<th>Mics</th>
<th>SNRout</th>
<th>SNIR (dB)</th>
<th>SD (dB)</th>
<th>ND (dB)</th>
<th>PESQ in</th>
<th>PESQout</th>
<th>PESQI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>242.5777</td>
<td>242.5777</td>
<td>-27.2464</td>
<td>-25.3345</td>
<td>4.348</td>
<td>2.429</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>248.6150</td>
<td>248.6150</td>
<td>-27.7687</td>
<td>-25.6635</td>
<td>4.376</td>
<td>2.457</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>164.1008</td>
<td>164.1008</td>
<td>-34.0526</td>
<td>-28.4285</td>
<td>4.376</td>
<td>2.457</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>162.4193</td>
<td>162.4193</td>
<td>-32.6501</td>
<td>-33.0153</td>
<td>4.383</td>
<td>2.464</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>249.0003</td>
<td>244.0003</td>
<td>-26.6977</td>
<td>-32.3702</td>
<td>4.351</td>
<td>2.079</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>231.5959</td>
<td>226.5959</td>
<td>-28.6977</td>
<td>-35.3151</td>
<td>4.373</td>
<td>2.101</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>155.3117</td>
<td>150.3117</td>
<td>-26.5213</td>
<td>-31.7520</td>
<td>4.370</td>
<td>2.098</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>165.5187</td>
<td>160.5187</td>
<td>-27.8137</td>
<td>-40.2384</td>
<td>4.360</td>
<td>2.088</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>261.7158</td>
<td>251.7158</td>
<td>-27.7895</td>
<td>-31.4324</td>
<td>4.368</td>
<td>1.779</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>266.1164</td>
<td>256.1164</td>
<td>-31.1160</td>
<td>-27.9359</td>
<td>4.379</td>
<td>1.79</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>216.4166</td>
<td>206.4166</td>
<td>-29.3158</td>
<td>-24.9397</td>
<td>4.365</td>
<td>1.776</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>179.4136</td>
<td>169.4136</td>
<td>-31.3148</td>
<td>-27.8413</td>
<td>4.376</td>
<td>1.787</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>267.0969</td>
<td>256.0969</td>
<td>-44.7878</td>
<td>-35.7250</td>
<td>4.379</td>
<td>1.477</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>177.8431</td>
<td>162.8431</td>
<td>-30.1923</td>
<td>-27.7309</td>
<td>4.354</td>
<td>1.452</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>6</td>
<td>169.1832</td>
<td>154.1832</td>
<td>-27.2139</td>
<td>-31.2114</td>
<td>4.356</td>
<td>1.454</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>268.8841</td>
<td>248.8841</td>
<td>-27.2412</td>
<td>-30.7690</td>
<td>4.356</td>
<td>1.15</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>183.1510</td>
<td>163.1510</td>
<td>-28.4919</td>
<td>-28.8273</td>
<td>4.378</td>
<td>1.172</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>271.0542</td>
<td>246.0542</td>
<td>-26.3092</td>
<td>-33.8519</td>
<td>4.380</td>
<td>0.836</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>3</td>
<td>272.9192</td>
<td>247.9192</td>
<td>-27.1252</td>
<td>-32.7251</td>
<td>4.244</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>4</td>
<td>222.5350</td>
<td>197.5350</td>
<td>-28.1354</td>
<td>-29.1543</td>
<td>4.375</td>
<td>0.831</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>5</td>
<td>179.7169</td>
<td>154.7169</td>
<td>-27.0861</td>
<td>-30.9958</td>
<td>4.326</td>
<td>0.782</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>6</td>
<td>180.7493</td>
<td>155.7493</td>
<td>-27.044</td>
<td>-31.1012</td>
<td>4.372</td>
<td>0.828</td>
<td></td>
</tr>
</tbody>
</table>

Table 4 SNR, SD, ND, PESQ, SNRI and PESQI for clean Speech signal with factory noise in anechoic environment.

In the assumed anechoic room optimal SNIR Method gives an average of 222.08 dB Signal-to-Noise Ratio Improvement (SNRI), -29.7649 dB speech distortion (SD), -30.1244 dB noise distortion (ND) and 2.1483 PESQ Improvement (PESQI) for white noise environments in all situations for the clean speech signal (Speech_all.wav). The SNRI, SD, ND, PESQI values are shown in the Table.5. All these parameter are recorded by improving the input signal-to-noise ratio.
Observation: Every value is recorded for a fixed DOA (with source angle at zero (0°) and interference at seventy five (75°)) and fixed microphone distance D=0.02 (In meters). the optimal SNIR module is very good in suppression of noise signal in an anechoic environment. PESQ value improves as the number of microphone are increased and PESQ ranges from (4.2 - 4.4), the average SNRI for 2 microphone is around 275 dB, for 4 microphones its around 222 dB and for 6 microphones it is observed to be around 171 dB. From the above observations it can be concluded that as the number of microphones increases SNRI decreases for a given input SNR level. The average PESQI is approximately 2.1 for 2, 4 and 6 microphones. SD ranges between (-26 dB to -38dB) and ND ranges between (-25dB to -40dB). Results are shown in the following table:

<table>
<thead>
<tr>
<th>SNR in</th>
<th>Mics</th>
<th>SNR (dB)</th>
<th>SD (dB)</th>
<th>ND (dB)</th>
<th>PESQ in</th>
<th>PESQout</th>
<th>PESQI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>275.1559</td>
<td>-27.3136</td>
<td>-26.3711</td>
<td>4.377</td>
<td>3.09</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>294.4386</td>
<td>-27.3176</td>
<td>-30.5722</td>
<td>4.382</td>
<td>1.627</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>299.8907</td>
<td>-38.5771</td>
<td>-25.8264</td>
<td>4.361</td>
<td>1.274</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>300.0389</td>
<td>-28.4981</td>
<td>-28.7291</td>
<td>4.388</td>
<td>1.301</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>319.8401</td>
<td>-27.4313</td>
<td>-30.2973</td>
<td>4.375</td>
<td>1.288</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: SNR, SD, ND, PESQ, SNRI and PESQI for clean Speech Signal with white noise in anechoic environment.

In the assumed anechoic room optimal SNIR Method gives an average of 216.4944 dB

Signal-to-Noise Ratio Improvement (SNRI), -28.5341 dB speech distortion (SD), -29.5491 dB noise distortion (ND) and 1.0072 PESQ Improvement (PESQI) for interference noise environments in all situations for the clean speech signal (Speech_all.wav). The SNRI, SD, ND, PESQI values are shown in the Table 6. All these parameter are recorded by improving the input signal-to-noise ratio.
**Observation:** Every value is recorded for a fixed DOA (with source angle at zero (0°) and interference at seventy five (75°)) and fixed microphone distance D=0.02 (In meters). The SNIR module is very good in suppression of noise signal in an anechoic environment. PESQ value improves as the number of microphone are increased and PESQ ranges from (4.2 - 4.4), the average SNRI for 2 microphone is around 267 dB, for 4 microphones its around 166 dB and for 6 microphones it is observed to be around 164 dB. From the above observations it can be concluded that as the number of microphones increases SNRI decreases for a given input SNR level. The average PESQI is approximately 1.0 for 2, 4 and 6 microphones. SD ranges between (-26 dB to -38dB) and ND ranges between (-26dB to -40dB). Results are shown in the following table:

<table>
<thead>
<tr>
<th>SNR in</th>
<th>Mics</th>
<th>SNRout</th>
<th>SNRI (dB)</th>
<th>SD (dB)</th>
<th>ND (dB)</th>
<th>PESQI in</th>
<th>PESQOut</th>
<th>PESQIQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>173.7871</td>
<td>173.7871</td>
<td>-31.5224</td>
<td>-29.3772</td>
<td>4.375</td>
<td>3.088</td>
<td>3.073</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>279.1172</td>
<td>274.1172</td>
<td>-27.3384</td>
<td>-40.0636</td>
<td>4.366</td>
<td>2.675</td>
<td>2.664</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>194.1345</td>
<td>189.1345</td>
<td>-31.114</td>
<td>-30.1171</td>
<td>4.38</td>
<td>2.681</td>
<td>2.681</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>182.1304</td>
<td>177.1304</td>
<td>-30.9474</td>
<td>-30.2542</td>
<td>4.373</td>
<td>2.674</td>
<td>2.674</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>235.3108</td>
<td>225.3108</td>
<td>-37.8265</td>
<td>-25.8388</td>
<td>4.373</td>
<td>2.296</td>
<td>2.296</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>176.6090</td>
<td>166.6090</td>
<td>-28.3368</td>
<td>-30.1087</td>
<td>4.298</td>
<td>2.221</td>
<td>2.221</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>283.8371</td>
<td>268.8371</td>
<td>-26.3148</td>
<td>-34.8758</td>
<td>4.385</td>
<td>1.964</td>
<td>1.964</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>188.1927</td>
<td>173.1927</td>
<td>-38.6601</td>
<td>-25.9765</td>
<td>4.369</td>
<td>1.948</td>
<td>1.948</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>294.4386</td>
<td>274.4386</td>
<td>-27.3716</td>
<td>-30.5722</td>
<td>4.382</td>
<td>1.627</td>
<td>1.627</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>236.5200</td>
<td>216.5200</td>
<td>-28.0363</td>
<td>-29.3577</td>
<td>4.376</td>
<td>1.621</td>
<td>1.621</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>204.1216</td>
<td>184.1216</td>
<td>-36.2335</td>
<td>-26.0327</td>
<td>4.373</td>
<td>1.618</td>
<td>1.618</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>299.8907</td>
<td>279.8907</td>
<td>-38.5771</td>
<td>-25.8264</td>
<td>4.361</td>
<td>1.274</td>
<td>1.274</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>192.4041</td>
<td>162.4041</td>
<td>-27.4311</td>
<td>-30.2397</td>
<td>4.375</td>
<td>1.288</td>
<td>1.288</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>188.7712</td>
<td>163.7712</td>
<td>-27.0056</td>
<td>-31.1896</td>
<td>4.404</td>
<td>1.317</td>
<td>1.317</td>
</tr>
</tbody>
</table>

Table 6: SNR, SD, ND, PESQ, SNRI and PESQI for clean Speech signal with interference in anechoic environment.

The graphs are plotted by varying the input, output SNR, SD, ND and PESQI values for the even number of microphones as followed:
Fig. 30: Plot of SNRI with Input SNR in anechoic environment for 2 mics in factory, white, wind, interference noise environments.

Fig. 31: Plot of SNRI with Input SNR in anechoic environment for 4 mics in factory, white, wind, interference noise environments.
Fig. 32: Plot of SNRI with Input SNR in anechoic environment for 6 mics in factory, white, wind, interference noise environments.

Fig. 33: Plot of PESQI with Input SNR in anechoic environment for 2 mics in factory, white, wind, interference noise environments.
Fig. 34: Plot of PESQI with Input SNR in anechoic environment for 4 mics in factory, white, wind, interference noise environments.

Fig. 35: Plot of PESQI with Input SNR in anechoic environment for 6 mics in factory, white, wind, interference noise environments.
Fig. 36: Plot of speech distortion (SD) with Input SNR in anechoic environment for 2 mics in factory, white, wind, interference noise environments.

Fig. 37: Plot of speech distortion (SD) with Input SNR in anechoic environment for 4 mics in factory, white, wind, interference noise environments.
Fig.38: Plot of speech distortion (SD) with Input SNR in anechoic environment for 6 mics in factory, white, wind, interference noise environments.

Fig.39: Plot of noise Distortion (ND) with Input SNR in anechoic environment for 2 mics in factory, white, wind, interference noise environments.
Fig. 40: Plot of noise distortion (ND) with Input SNR in anechoic environment for 4 mics in factory, white, wind, interference noise environments.

Fig. 41: Plot of noise distortion (ND) with Input SNR in anechoic environment for 6 mics in factory, white, wind, interference noise environments.
7.1.2 Optimal SNIR Beamformer in Reverberated environment

The optimal SNIR beamforming method is tested with an input Speech signal (Speech_all.wav) with the reverberant speech signal (Which is generated in a room of dimensions [2*2* 2] (in meters) , source and microphone positions are at [2,1,2] (corresponds to (x,y,z) positions in free space) , [1.6,1,1.3] (corresponds to (x,y,z) positions in free space ) respectively. So final equation of optimal SNIR is

$$w_{opt} = \max \{ \frac{w^H R_{ss} w}{w^H (R_{nn} + R_{rr}) w} \}$$

Eq.7.2

The main objective of Reverberated optimal SNIR beamformer is to improve the numerator power ($w^H R_{ss} w$) and decrease the denominator power ($w^H (R_{nn} + R_{rr}) w$). This can be achieved by Eigen-decomposition of autocorrelation matrix which is more clearly explained in chapter 7.1. Results are taken for the fixed direction of arrival (0 and 75 degrees) of source and noisy environments. Distance of microphone is also kept consent at D=0.02 (In meters). SNR_out (Output signal to noise ratio), SD (speech distortion), noise distortion (ND) and PESQI (PESQ improvement) tested for 2 microphone array by varying the input SNR (SNR_in) from (0-25 dB).

In the assumed reverberated room optimal SNIR Method gives an average of 24.4014 dB signal-to-noise ratio improvement (SNRI), -13.77951 dB speech distortion (SD), -32.0674 dB noise distortion (ND) and 0.5935 PESQ Improvement (PESQI) for Wind noise environments in all situations for the reverberated speech signal. The SNRI, SD, ND, PESQI values are shown in the Table.7. All these parameter are recorded by improving the input signal-to-noise ratio.
Observation: Every value is recorded for a fixed DOA (with source angle at zero (0°) and interference at seventy five (75°)) and fixed microphone distance D=0.02 (In meters). SNIR module is good in suppression of noise signal but satisfactory in suppression of reverberated signal. PESQ value is very good at 0dB input SNR (SNRin) and the PESQ score ranges from (2.141– 2.517). Average SNRI for 2 microphones is around 24dB and average PESQI is around 0.6. Results are shown in the following table:

<table>
<thead>
<tr>
<th>SNRin</th>
<th>SNRout</th>
<th>SNRI (dB)</th>
<th>SD (dB)</th>
<th>ND (dB)</th>
<th>PESQ in</th>
<th>PESQ out</th>
<th>PESQI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25.1459</td>
<td>25.1459</td>
<td>-20.693</td>
<td>-20.698</td>
<td>1.453</td>
<td>2.141</td>
<td>0.688</td>
</tr>
<tr>
<td>5</td>
<td>29.5671</td>
<td>24.5671</td>
<td>-20.692</td>
<td>-26.094</td>
<td>1.439</td>
<td>2.275</td>
<td>0.836</td>
</tr>
<tr>
<td>10</td>
<td>34.1316</td>
<td>24.1316</td>
<td>-20.691</td>
<td>-32.654</td>
<td>1.734</td>
<td>2.354</td>
<td>0.62</td>
</tr>
<tr>
<td>15</td>
<td>39.2183</td>
<td>24.2183</td>
<td>-20.691</td>
<td>-42.0017</td>
<td>1.971</td>
<td>2.439</td>
<td>0.468</td>
</tr>
<tr>
<td>20</td>
<td>44.1861</td>
<td>24.1861</td>
<td>-20.691</td>
<td>-35.9413</td>
<td>1.976</td>
<td>2.489</td>
<td>0.513</td>
</tr>
<tr>
<td>25</td>
<td>49.1756</td>
<td>24.1756</td>
<td>-20.691</td>
<td>-35.0145</td>
<td>2.081</td>
<td>2.517</td>
<td>0.436</td>
</tr>
</tbody>
</table>

Table 7: SNR, SD, ND, PESQ, SNRI and PESQI for clean Speech Signal with Wind noise in reverberated environment.

In the assumed reverberated room SNIR Method gives an average of 10.49375 dB Signal-to-Noise Ratio Improvement (SNRI), -21.2767 dB speech distortion (SD), -34.4607 dB noise distortion (ND) and 0.7633 PESQ Improvement (PESQI) for White noise environments in all situations for the reverberated speech signal. The SNRI, SD, ND, PESQI values are shown in the Table 8. All these parameters are recorded by improving the input signal-to-noise ratio.

Observation: Every value is recorded for a fixed DOA (with source angle at zero (0°) and interference at seventy five (75°)) and fixed microphone distance D=0.02 (In meters). SNIR module is good in suppression of noise signal but satisfactory in suppression of a reverberated signal. PESQ value is very good at 0dB input SNR (SNRin) and the PESQ score ranges from (2.4 – 2.7). Average SNRI for 2 microphones is around 20dB and average PESQI is around 0.8. Results are shown in the following table:

<table>
<thead>
<tr>
<th>SNRin</th>
<th>SNRout (dB)</th>
<th>SNRI (dB)</th>
<th>SD (dB)</th>
<th>ND (dB)</th>
<th>PESQ in</th>
<th>PESQ out</th>
<th>PESQI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>16.9774</td>
<td>16.9774</td>
<td>-21.277</td>
<td>-20.6103</td>
<td>1.064</td>
<td>2.408</td>
<td>1.344</td>
</tr>
<tr>
<td>5</td>
<td>22.5853</td>
<td>17.5853</td>
<td>-21.277</td>
<td>-25.7178</td>
<td>1.397</td>
<td>2.492</td>
<td>1.095</td>
</tr>
<tr>
<td>10</td>
<td>29.5384</td>
<td>19.5384</td>
<td>-21.276</td>
<td>-31.0739</td>
<td>1.723</td>
<td>2.421</td>
<td>0.698</td>
</tr>
<tr>
<td>15</td>
<td>34.5494</td>
<td>19.5494</td>
<td>-21.276</td>
<td>-37.4343</td>
<td>1.984</td>
<td>2.528</td>
<td>0.444</td>
</tr>
<tr>
<td>20</td>
<td>42.509</td>
<td>22.509</td>
<td>-21.276</td>
<td>-50.3325</td>
<td>2.143</td>
<td>2.625</td>
<td>0.482</td>
</tr>
<tr>
<td>25</td>
<td>46.803</td>
<td>21.803</td>
<td>-21.276</td>
<td>-41.5953</td>
<td>2.242</td>
<td>2.747</td>
<td>0.705</td>
</tr>
</tbody>
</table>

Table 8: SNR, SD, ND, PESQ, SNRI and PESQI for clean Speech Signal with white noise in reverberated environment.

In the assumed reverberated room SNIR Method gives an average of 17.7729 dB signal-to-noise ratio improvement (SNRI), -21.2763 dB speech distortion (SD), -34.135 dB noise distortion (ND) and 0.6133 PESQ improvement (PESQI) for factory noise environments in all situations for the reverberated speech signal. The SNRI, SD, ND,
PESQI values are shown in the Table 9. All these parameter are recorded by improving the input signal-to-noise ratio.

**Observation:** Every value is recorded for a fixed DOA (with source angle at zero (0°) and interference at seventy five (75°)) and fixed microphone distance D=0.02 (in meters). SNIR module is good in suppression of noise signal but satisfactory in suppression of a reverberated signal. PESQ value is very good at 0dB input SNR (SNRin) and the PESQ score ranges from (2.4 – 2.8). Average SNRI for 2 microphones is around 18 dB and average PESQI is around 0.6. Results are shown in the following table:

<table>
<thead>
<tr>
<th>SNRin</th>
<th>SNRout (dB)</th>
<th>SNRI (dB)</th>
<th>SD (dB)</th>
<th>ND (dB)</th>
<th>PESQin</th>
<th>PESQout</th>
<th>PESQI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>19.9791</td>
<td>19.9791</td>
<td>-21.2763</td>
<td>-19.8667</td>
<td>1.608</td>
<td>2.456</td>
<td>0.848</td>
</tr>
<tr>
<td>5</td>
<td>23.1309</td>
<td>18.1309</td>
<td>-21.2763</td>
<td>-24.9559</td>
<td>1.896</td>
<td>2.51</td>
<td>0.614</td>
</tr>
<tr>
<td>10</td>
<td>27.4145</td>
<td>17.4145</td>
<td>-21.2763</td>
<td>-30.2506</td>
<td>2.088</td>
<td>2.616</td>
<td>0.528</td>
</tr>
<tr>
<td>15</td>
<td>32.1238</td>
<td>17.1238</td>
<td>-21.2763</td>
<td>-36.3412</td>
<td>2.205</td>
<td>2.716</td>
<td>0.511</td>
</tr>
<tr>
<td>20</td>
<td>37.012</td>
<td>17.012</td>
<td>-21.2763</td>
<td>-36.3412</td>
<td>2.205</td>
<td>2.716</td>
<td>0.511</td>
</tr>
<tr>
<td>25</td>
<td>41.9734</td>
<td>16.9734</td>
<td>-21.2763</td>
<td>-41.9365</td>
<td>2.319</td>
<td>2.847</td>
<td>0.528</td>
</tr>
</tbody>
</table>

Table 9: SNR, SD, ND, PESQ, SNRI and PESQI for clean Speech Signal with factory noise in reverberated environment.

In the assumed reverberated room SNIR Method gives an average of 12.6312 dB Signal-to-Noise Ratio Improvement (SNRI), -12.623 dB speech distortion (SD), -33.135 dB noise distortion (ND) and 0.35083 PESQ Improvement (PESQI) for Interference noise (road.wav)environments in all situations for the reverberated speech signal. The SNRI, SD, ND, PESQI values are shown in the Table 10. All these parameter are recorded by improving the input signal-to-noise ratio.

**Observation:** Every value is recorded for a fixed DOA (with source angle at zero (0°) and interference at seventy five (75°)) and fixed microphone distance D=0.02 (in meters). SNIR module is good in suppression of noise signal but satisfactory in suppression of a reverberated signal. PESQ value is very good at 0dB input SNR (SNRin) and the PESQ score ranges from (1.5 – 1.6). Average SNRI for 2 microphones is around 13 dB and average PESQI is around 0.3. Results are shown in the following table:

<table>
<thead>
<tr>
<th>SNRin</th>
<th>SNRout (dB)</th>
<th>SNRI (dB)</th>
<th>SD (dB)</th>
<th>ND (dB)</th>
<th>PESQin</th>
<th>PESQout</th>
<th>PESQI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>12.4624</td>
<td>12.4624</td>
<td>-20.9625</td>
<td>-20.938</td>
<td>1.041</td>
<td>1.591</td>
<td>0.55</td>
</tr>
<tr>
<td>5</td>
<td>20.1372</td>
<td>15.1372</td>
<td>-20.9621</td>
<td>-26.102</td>
<td>1.127</td>
<td>1.569</td>
<td>0.442</td>
</tr>
<tr>
<td>10</td>
<td>23.126</td>
<td>13.126</td>
<td>-20.9621</td>
<td>-31.6659</td>
<td>1.239</td>
<td>1.576</td>
<td>0.337</td>
</tr>
<tr>
<td>15</td>
<td>27.5236</td>
<td>12.5236</td>
<td>-20.9621</td>
<td>-39.1745</td>
<td>1.33</td>
<td>1.615</td>
<td>0.285</td>
</tr>
<tr>
<td>20</td>
<td>32.8982</td>
<td>12.8982</td>
<td>-20.9621</td>
<td>-42.4964</td>
<td>1.405</td>
<td>1.645</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Table 10: SNR, SD, ND, PESQ, SNRI and PESQI for clean Speech Signal with interference in reverberated environment.
The graphs are plotted by varying the input, output SNR, SD, ND and PESQI values for the 2 microphones as followed:

Fig. 43: Plot of PESQI with Input SNR in reverberated environment for 2 Mics in factory, white, wind, interference noise environments.

Fig. 44: Plot of SNRI with Input SNR in reverberated environment for 2 Mics in factory, white, wind, interference noise environments.
The Reverberation ratio of optimal SNIR beamforming is tested with an input Speech signal (Specch_all.wav) with the reverberant speech signal (Which is generated in a room of dimensions [2*2* 2] (in meters) , source and microphone positions are at [2,1,2] (corresponds to (x,y,z) positions in free space) , [16,1,1,3] (corresponds to (x,y,z) positions in free space ) respectively.

This reverberation ratio is calculated for number of microphones instead of input SNR so the ‘Input SNR is fixed to zero dB’ and reading is tabulated as; RRin for 0dB:
<table>
<thead>
<tr>
<th>No. of mics</th>
<th>RRin</th>
<th>RRout</th>
<th>RR improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-67.5512</td>
<td>-57.9781</td>
<td>9.5731</td>
</tr>
<tr>
<td>3</td>
<td>-67.5512</td>
<td>-58.0009</td>
<td>9.5503</td>
</tr>
<tr>
<td>4</td>
<td>-67.5512</td>
<td>-55.7071</td>
<td>11.8441</td>
</tr>
<tr>
<td>5</td>
<td>-67.5512</td>
<td>-54.5440</td>
<td>13.0072</td>
</tr>
<tr>
<td>6</td>
<td>-67.5512</td>
<td>-53.997</td>
<td>13.5542</td>
</tr>
</tbody>
</table>

Table 1: RRin and RRout for reverberated speech signal with interference noise.

<table>
<thead>
<tr>
<th>No. of mics</th>
<th>RRin</th>
<th>RRout</th>
<th>RR improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-64.8408</td>
<td>-58.2511</td>
<td>6.5897</td>
</tr>
<tr>
<td>3</td>
<td>-64.8408</td>
<td>-57.9001</td>
<td>6.9407</td>
</tr>
<tr>
<td>4</td>
<td>-64.8408</td>
<td>-55.671</td>
<td>9.1698</td>
</tr>
<tr>
<td>5</td>
<td>-64.8408</td>
<td>-55.021</td>
<td>9.8198</td>
</tr>
<tr>
<td>6</td>
<td>-64.8408</td>
<td>-53.2971</td>
<td>11.5437</td>
</tr>
</tbody>
</table>

Table 12: RRin and RRout for reverberated speech signal with wind noise.

<table>
<thead>
<tr>
<th>No. of mics</th>
<th>RRin</th>
<th>RRout</th>
<th>RR improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-68.6366</td>
<td>-59.4640</td>
<td>9.1726</td>
</tr>
<tr>
<td>3</td>
<td>-68.6366</td>
<td>-58.9712</td>
<td>9.6654</td>
</tr>
<tr>
<td>4</td>
<td>-68.6366</td>
<td>-55.6484</td>
<td>12.9882</td>
</tr>
<tr>
<td>5</td>
<td>-68.6366</td>
<td>-52.7711</td>
<td>15.8655</td>
</tr>
<tr>
<td>6</td>
<td>-68.6366</td>
<td>-50.4986</td>
<td>18.138</td>
</tr>
</tbody>
</table>

Table 13: RRin and RRout for reverberated speech signal with factory noise.

<table>
<thead>
<tr>
<th>No. of mics</th>
<th>RRin</th>
<th>RRout</th>
<th>RR improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-66.9457</td>
<td>-56.9578</td>
<td>9.9879</td>
</tr>
<tr>
<td>3</td>
<td>-66.9457</td>
<td>-53.6464</td>
<td>13.2993</td>
</tr>
<tr>
<td>4</td>
<td>-66.9457</td>
<td>-54.0067</td>
<td>12.939</td>
</tr>
<tr>
<td>5</td>
<td>-66.9457</td>
<td>-52.9488</td>
<td>13.9969</td>
</tr>
</tbody>
</table>

Table 14: RRin and RRout for reverberated speech signal with white noise.

Every value is recorded for a fixed DOA (with source angle at zero (0°) and interference at seventy five (75°)) and fixed microphone distance D=0.02 (in meters). The reverberation ratio is estimated at 0 dB input SNR level. This ratio is measured for different noises by varying number of microphones (2-6) which are shown in tables from Table 11 to
Table 14. The average reverberation ratio improvement is better for white noise which is around 10 dB for 2 microphones. Similarly, the reverberation ratio improvement is better for factory noise which is around 18 dB for 6 microphones.

Fig.47: Plot of Reverberation improvement for factory, white, wind, interference noise environments.

7.2 Comparision of Optimal SNIR Beamformer with Other beamformers in anechoic environments

7.3.1 Elko’s beamformer:

<table>
<thead>
<tr>
<th>No.</th>
<th>Type of Noise</th>
<th>Average SNRI (dB)</th>
<th>SD (dB)</th>
<th>ND (dB)</th>
<th>PESQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Factory Noise (FN)</td>
<td>19.8405</td>
<td>-38.3806</td>
<td>-28.8779</td>
<td>3.408</td>
</tr>
<tr>
<td>2</td>
<td>Wind Noise (WN)</td>
<td>14.345</td>
<td>-38.7754</td>
<td>-33.6536</td>
<td>2.992</td>
</tr>
<tr>
<td>3</td>
<td>White Noise (WHN)</td>
<td>6.7556</td>
<td>-38.3614</td>
<td>-32.4904</td>
<td>2.826</td>
</tr>
<tr>
<td>4</td>
<td>Interference noise (IN)</td>
<td>5.1354</td>
<td>-38.3808</td>
<td>-29.0912</td>
<td>2.007</td>
</tr>
</tbody>
</table>

Table 15: SNRI, PESQ, SD and ND in anechoic environment for 2 Mics with various noise signals using Elko’s beamformer. Elko’s beamformer is tested for the fixed angles (source position at zero degrees and the interference at 75°), and the fixed Speech signal (Speech_all.wav) is tested in four noisy environment. The whole system is tested with 2-microphone array model. The SNRI is around 20 dB and output PESQ is around 3.4 for factory noise which is best among all noisy environments. The results are shown in Table 15.
7.2.2 Wiener beamformer:

Wiener beamformer is tested for the fixed angles (source position at zero degrees and the interference at 75°), and the fixed Speech signal (Speech_all.wav) is tested in four noisy environment. The whole system is tested with 2-microphone array model. The SNRI is around 6 dB for wind noise and output PESQ is around 3.4 for factory noise which is best among all noisy environments. The tables are as follows: The tables are as follows:

<table>
<thead>
<tr>
<th>No.</th>
<th>Type of Noise</th>
<th>Average SNRI (dB)</th>
<th>SD (dB)</th>
<th>ND (dB)</th>
<th>PESQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Factory Noise (FN)</td>
<td>4.0505</td>
<td>-41.3806</td>
<td>-27.8779</td>
<td>3.470</td>
</tr>
<tr>
<td>2</td>
<td>Wind Noise (WN)</td>
<td>5.9345</td>
<td>-42.7754</td>
<td>-28.6536</td>
<td>2.756</td>
</tr>
<tr>
<td>3</td>
<td>White Noise (WHN)</td>
<td>5.7556</td>
<td>-41.3614</td>
<td>-27.4904</td>
<td>2.826</td>
</tr>
<tr>
<td>4</td>
<td>Interference noise (IN)</td>
<td>2.1354</td>
<td>-40.3808</td>
<td>-29.0912</td>
<td>1.76</td>
</tr>
</tbody>
</table>

Table 16: SNRI, PESQ, SD and ND in anechoic environment for 2 Mics with various noise signals using Wiener beamformer.

7.2.3 Delay and sum Beamformer:

Delay and sum beamformer is tested for the fixed angles (source position at zero degrees and the interference at 75°), and the fixed Speech signal (Speech_all.wav) is tested in four noisy environment. The whole system is tested with 2-microphone array model. The SNRI is around 0.8 dB for white noise and output PESQ is around 1.0 for factory noise which is best among all noisy environments. These results are examined for 2-microphone array model, the results can be improved for 4 to 6 microphone array models. The tables are as follows:

<table>
<thead>
<tr>
<th>No.</th>
<th>Type of Noise</th>
<th>Average SNRI (dB)</th>
<th>SD (dB)</th>
<th>ND (dB)</th>
<th>PESQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Factory Noise (FN)</td>
<td>0.0313</td>
<td>-29.4736</td>
<td>-39.308</td>
<td>1.026</td>
</tr>
<tr>
<td>2</td>
<td>Wind Noise (WN)</td>
<td>0.3047</td>
<td>-35.069</td>
<td>-44.2561</td>
<td>0.59</td>
</tr>
<tr>
<td>3</td>
<td>White Noise (WHN)</td>
<td>0.7514</td>
<td>-33.6485</td>
<td>-36.7051</td>
<td>0.69</td>
</tr>
<tr>
<td>4</td>
<td>Interference noise (IN)</td>
<td>0.7245</td>
<td>-31.8565</td>
<td>-35.7871</td>
<td>0.779</td>
</tr>
</tbody>
</table>

Table 17: SNRI, PESQ, SD and ND in anechoic environment for 2 Mics with various noise signals using Delay and Sum beamformer.

The variation in SNRI is very high For Optimal SNIR beamformer when compare to Delay and sum beam former so it is very hard to illustrate in a figure 47 but as the PESQ value is ranges from (0-5) this is very easy to judge the quality of the signals as shown in the following figure;
Fig. 48: Estimation of output PESQ for various beamformers in different noisy environments
8. Conclusion and future work

8.1 Conclusion

This thesis is a collaborative work done by four members, which includes Elko’s, Wiener, optimal SNIR, Delay and sum beamformers in anechoic environments and also optimal SNIR and wiener Beamformers are implemented and evaluated in reverberant environments. Among which this thesis has focused on speech enhancement of noisy speech signal in hands-free speech communication using optimal SNIR beamformer in anechoic and reverberated environment. The performance of optimal SNIR beamformer is tested in anechoic environment with four different noise environments (White Noise, factory noise, wind noise, Interference speech signal). The performance is measured with signal to noise improvement (SNRI), perceptual evaluation of speech quality Improvement (PESQI), speech distortion (SD), noise distortion (ND). The better SNRI indicates more SD i.e. any system compromises between high SNRI and low SD. As the input SNR level increases partially ND, and SNRI and PESQI remains almost constant from overall results which have been shown in tables and plots. The performance of optimal SNIR beamformer has better results in all noisy environments especially in white, wind and factory. So this beamformer can be used as best noise suppresser in anechoic environments. But in real world scenario, it is very hard to find anechoic environments so this thesis has tested the same beamformer in reverberated room environment. In reverberated environment, the optimum SNIR beamformer shows better performance with white, wind and factory noises compared to interference environments. The amount of reverberation present in a signal can be estimated with reverberation ratio (RR), this is estimated and tabulated. So this thesis has concluded that the reverberation is well suppressed as number of microphone increases, this is proved from the reverberation ratio tables.

From this thesis it concludes that, in anechoic environment optimal SNIR Beamformer has best results among all beamformers that have been implemented in this thesis work. The optimal SNIR beamformer gives the best speech quality and intelligibility which are shown in respective tables and plots. In anechoic environment, it can be concluded that the SNRI is better in case of white noise for 2 microphone array model which is around 275 dB and better PESQ improvement which is around 2.1. Similarly SNRI is better for 6 microphones with white noise which is around 171 dB and better PESQI which is around 2.1. In reverberated environment, it is observed that SNRI is better in case of wind noise which is around 24 dB and PESQI is better in case of white noise which is around 0.8. As the
number of microphones increases for a given input SNR (0 dB) level reverberation ratio improvement increases. The reverberation ratio improvement is better for 6 microphones compared to 2 microphones. From the tables it can be observed that the reverberation ratio improvement is better for 6 microphones in case of factory noise which is around 18 dB. Similarly, for 2 microphones the reverberation ratio improvement is better for white noise which is around 10 dB.

The other speech enhancement methods such as Wiener Beamformer, Elko’ Beamformer, Delay and Sum Beamformer were implemented in MATLAB offline mode. The evaluation parameters of these all beamformers are compared and behavior of each beamformer is observed in all different noise environments for only two microphones. The performance of optimal SNIR Beamformer is best when compared to other beamformers that have been implemented in this thesis. It is observed that SNRI and PESQI are best for optimal SNIR Beamformer and poor in case of DSB. The Elko’s performs at its best in case of two microphones and the Wiener Beamformer performance can be improved as the number of microphones increases. The optimal SNIR beamformer is basically designed for comparison purposes, which gives output PESQ of 4.3 for all noise environments. The detailed view of the comparison results are shown in corresponding tables, plots and graphs.

8.2 Future Work

In anechoic environment the optimal SNIR beamformer method using Eigen value decomposition for speech enhancement is completely implemented in time-domain. This can be implemented in DSP kit in real time applications, as it’s giving the best results in offline mode. In reverberated system probably the filter bank method can improve the reverberation ratio results. A combination of beamformers can improve the reverberation ratio. Every part in this thesis done in time domain this can be extended to frequency domain.
9. Bibliography


