Multi Channel Sub Band Wiener Beamformer

Amerineni Rajesh

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

Blekinge Institute of Technology
October 2012
Contact information:

Author:

Amerineni Rajesh

Email: raam10@student.bth.se, rajesh.amerineni@gmail.com

Supervisor:

Dr. Nedelko Grbic

Department of Electrical Engineering

School of Engineering, BTH

Blekinge Institute of Technology, Sweden

Email: nedelko.grbic@bth.se

Examiner:

Dr. Benny Sällberg

Department of Electrical Engineering

School of Engineering, BTH

Blekinge Institute of Technology, Sweden

Email: benny.sallberg @ bth.se
ABSTRACT

With the recent advance in microphone array speech processing, achieving robustness of speaker localization becomes most significant aspect. At the same time considerable research growth is performed in developing the multiple microphone sensors equipped rooms are developed also called as smart rooms for real time applications.

The accuracy of speaker localization is down casted by acoustic noise and room reverberations. In distributed meeting environment speaker localization is performed by far field microphone arrays with the help of beamforming. But far field Microphone performance is degraded by room reverberations and acoustic noise.

In this master thesis, speaker localization with two adaptive beamforming techniques in distributed meeting application in reverberated environment with the help of far filed microphone arrays is design and implemented. The two beamforming methods examined are multichannel wiener beamformer and multichannel sub band wiener beamformer. These methods use wiener filtering technique for their implementation and they are implemented to capture the human voice using widely separated microphone arrays even when irregular disturbances are present. A smart room is developed with Image source model for generating reverberation in which beamformers are implemented. In sub band beamformer WOLA filter bank is designed. The sub band beamforming is further extended to steered response power with phase transform for speaker localization is achieved with the cross correlation but speech is heavily degraded by the noise which can be further studied to eliminated it.

Finally the quality of the speech is tested using SNR and PESQ (Perceptual Evaluation of Speech Quality) and also the performance of the system with respect to reverberation time is calculated. The results show that the two implementations are acceptable in terms of PESQ score.

Keywords: Reverberation, Linear Microphone array, WOLA Filter Bank, Wiener Beamformer, SRP-PHAT
ACKNOWLEDGEMENT

I would like to express my deepest sense of gratitude to my God who gave me strength to complete this task.

I wish to express my outmost gratitude to my supervisor Dr. Nedelko Grbic for his encouragement, patience, guidance and support. His guide lines have been vital and without them, this work could not have been achieved properly.

I express my grateful thanks to my examiner Dr. Benny Sällberg for his valuable suggestions for increasing the quality of the thesis.

I am grateful to my friends Seshu, Vamsynagh and Hemanth for supporting me to complete this thesis and my roommates for their immense support.

I especially thank my mother and my brother for their continuous support and encouragement.

Amerineni Rajesh

Karlskrona, October 2012
# TABLE OF CONTENT

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF ABREVIATIONS</td>
<td>ix</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Motivation</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Problem Statement</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Hypothesis</td>
<td>2</td>
</tr>
<tr>
<td>1.4. Thesis Overview</td>
<td>3</td>
</tr>
<tr>
<td>2. Room Modelling</td>
<td>4</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>4</td>
</tr>
<tr>
<td>2.2 Acoustic Noise</td>
<td>4</td>
</tr>
<tr>
<td>2.3 Reverberations</td>
<td>5</td>
</tr>
<tr>
<td>2.4. Image Source Model (Room impulse response)</td>
<td>6</td>
</tr>
<tr>
<td>2.4.1 Over view of Image source model and increase its efficiency with Thiran all pass filter</td>
<td>7</td>
</tr>
<tr>
<td>3. Filter Bank Design</td>
<td>12</td>
</tr>
<tr>
<td>3.1 Introduction about time-frequency domain</td>
<td>12</td>
</tr>
<tr>
<td>3.2 Filter bank</td>
<td>13</td>
</tr>
<tr>
<td>3.2.1 FIR filter</td>
<td>13</td>
</tr>
<tr>
<td>3.2.2. Brief description about filter Bank</td>
<td>14</td>
</tr>
<tr>
<td>3.3. Weighted overlapping Add (WOLA)</td>
<td>14</td>
</tr>
<tr>
<td>3.3.1. WOLA analysis filter Bank</td>
<td>15</td>
</tr>
<tr>
<td>3.3.2. WOLA synthesis filter bank</td>
<td>16</td>
</tr>
<tr>
<td>3.3.3. Decimation</td>
<td>16</td>
</tr>
<tr>
<td>3.3.4. FFT/IFFT Transform</td>
<td>16</td>
</tr>
<tr>
<td>4. Multi-Microphone processing</td>
<td>19</td>
</tr>
<tr>
<td>4.1. Array signal processing</td>
<td>19</td>
</tr>
<tr>
<td>4.2. Linear array processing</td>
<td>19</td>
</tr>
<tr>
<td>4.2.1. Basic concepts of Linear array processing</td>
<td>20</td>
</tr>
<tr>
<td>4.3. Microphone array processing</td>
<td>21</td>
</tr>
<tr>
<td>4.3.1. Linear microphone array</td>
<td>21</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Fig. 2.1. Schematic diagram of a typical speaker to receiver room impulse response three parts can be easily distinguished: direct wave, early reflections and late reflections ........................................ 6
Fig. 2.2. Mapping of different virtual sources with mirror method ............................................. 7
Fig. 2.3. Path involving two reflections with two virtual sources .............................................. 8
Fig. 3.1. A bank of eight band pass filters .................................................................................. 14
Fig. 3.2. Block diagram of filter bank ........................................................................................ 15
Fig. 3.3. Block Diagram of WOLA filter bank ........................................................................... 17
Fig. 4.1. Propagating of far field sound wave with microphone array ....................................... 22
Fig. 4.2. Block diagram of Multichannel Sub Band Wiener Beamformer ..................................... 36
Fig. 5.1. Structural diagram of Delay and Sum beamformer ...................................................... 28
Fig. 5.2. Block diagram of Filter and Sum beamformer ............................................................. 28
Fig. 5.3. Block diagram of GSC beamformer ............................................................................. 31
Fig. 6.1. Block diagram of microphone array wiener beamformer ........................................... 34
Fig. 6.2. Block diagram of Multichannel Sub Band Wiener Beamformer ..................................... 36
Fig. 7.1. TDOA between two microphones ................................................................................. 45
Fig. 8.1. Energy decay for reflection coefficient r=0.95 ............................................................. 52
Fig. 8.2. Energy decay curve for reflection coefficient r=0 ....................................................... 52
Fig. 8.3. Plot indicates $RT60$ for the energy curve for r=0.95 ................................................... 53
Fig. 8.4. Plot between $RT60$ and reflection coefficient ‘r’ .......................................................... 54
Fig. 8.5. Impulse response for thiran all pass filter for delay 7.5 ................................................ 55
Fig. 8.6. Group delay response for thiran filter at delay 7.5 for N=1 to N=8 ............................... 55
Fig. 8.7. Graph between reverberated speech signal and original speech signal ..................... 56
Fig. 8.8. Plot between Input SNR and SNR output for two different positions of liner microphone array ........................................................................................................ 58
Fig. 8.9. Plot between Input SNR and SNR improvement for two different positions of liner microphone array ........................................................................................................ 59
Fig. 8.10. Plot between Input SNR and PESQ output for two different positions of liner microphone array ........................................................................................................ 59
Fig. 8.11. Plot among PSD of input speech signal, noise signal, speech+noise signal and output speech signal of wiener beamformer .............................................................................. 60
Fig. 8.12. Magnitude response of WOLA filter bank for 256 sub bands .................................... 61
Fig. 8.13. Plot between Lambda (forgetting factor) and SNR improvement at two different sub band positions ........................................................................................................ 62
Fig. 8.14. Plot between Lambda (forgetting factor) and PESQ score at two different sub band positions ........................................................................................................ 63
Fig. 8.17. Plot of Input speech to mic2 and beamformer speech output for frequency domain ...... 65
Fig. 8.18. Estimating of the $\tau$ form the TDOA observe from the combinations of the mics ...... 66
Fig. 8.19. Graph between the Ideal DOA positions and practical DOA positions ..................... 68
Fig. 8.20. Graph between the Input position and absolute error .............................................. 68
LIST OF TABLES

Table 8.1. Evaluation of wiener beamformer at two different position of linear microphone array……………………………………………………………………………….58

Table 8.2. Evaluation of Multi channel sub band wiener beamformer at two different sub bands of WOLA filter Bank………………………………………………………………………………………………62

Table 8.3. Evaluation of SRP-PHAT on the basis of actual DOA and practical DOA………………………………………………………………………………………………………67
## LIST OF ABREVIATIONS

<table>
<thead>
<tr>
<th>Abreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>RIRS</td>
<td>Room Impulse Response</td>
</tr>
<tr>
<td>STFT</td>
<td>Short Term Fourier Transformer</td>
</tr>
<tr>
<td>FIR</td>
<td>Finite Impulse Response</td>
</tr>
<tr>
<td>IIR</td>
<td>Infinite Impulse Response</td>
</tr>
<tr>
<td>WOLA</td>
<td>Weighted Over Lapping Add</td>
</tr>
<tr>
<td>FIFO</td>
<td>First Input First Output</td>
</tr>
<tr>
<td>DFT</td>
<td>Discrete Fourier Transform</td>
</tr>
<tr>
<td>IDFT</td>
<td>Inverse Discrete Fourier Transform</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>IFFT</td>
<td>Inverse Fourier Transform</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal To Noise Ratio</td>
</tr>
<tr>
<td>LCMV</td>
<td>Linearly Constrained Minimum Variance Method</td>
</tr>
<tr>
<td>GSC</td>
<td>Generalized Side Lobe Canceler</td>
</tr>
<tr>
<td>SRP-PHAT</td>
<td>Steered Response Power Phase Alignment Transform</td>
</tr>
<tr>
<td>DSP</td>
<td>Digital Signal Processor</td>
</tr>
<tr>
<td>OS</td>
<td>Oversampling Ratio</td>
</tr>
<tr>
<td>PESQ</td>
<td>Perceptual Evaluation of Speech Quality</td>
</tr>
<tr>
<td>dB</td>
<td>Decibel</td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunication Union</td>
</tr>
</tbody>
</table>
1. Introduction

This chapter introduce project work carried out and overview of the main motivation and problem faced in development of project. The chapter 1.1 shows the motivation for the thesis. The following chapter 1.2 problem statement last session chapter 1.3 shows the organisation of the thesis.

1.1 Motivation

Human being is sensitive to a large number of signals. Speech is also one of the most important signals that can motivate a person. The process of transferring speech signal from one source to other source is called speech processing. Speech processing techniques are implementing in most of the present technology. This leads to increase in demand for the speech processing technologies in the present world. With the invention of the telephone speech processing technique got a sudden boost up. Telephone is a close talking microphone approach. Close talking approach can easily isolate the noise from the speech signal [1].

But the technology now moves from telephone to video conference through the revolution of the internet. Video conference increases the conference room environment where a group of people perform a video conference with an individual or a group of people. Here speech recognition and enhancement of the speech signal place a significant role. In general conference room environment provides close talk microphone to each and every individual it leads to increase in expensive and complexity in such case of application. This can be overcome by implementing far field microphone array for speech enhancement and speaker localization in conference room [2, 3].

Speech enhancement can be performed by the beamformer [4]. Video conference employ on the basic concept of amplify all the sources in the room so both noise and speech signals are amplified simultaneously. In such situation it is difficult to identify the primary source in the room. In those cases beamforming algorithm become a great tool for noise reductions. As the number of microphone array increases leads to increase the aperture size of the beamformer this final leads to increase the ability of beamformer to identify the primary source using spatial information improves[6, 7].

Various Beamforming Methods are exited for speech enhancement and recognition in closed room acoustic environment. Closed room has the capability of generating the reverberations for the speech and noise signals which leads to distortion of the speech signal.
Dereverberation of speech signal in a closed room can be performed by Beamformers. Wiener filter has the capability of generating the beamformer [8].

The implementation of the Microphone array with beamformer is further extending to find out the source localization in the conference room. That means steering the beamformer response towards the speaker. Localization can be performed by algorithm called SRP-PHAT with the help of Microphone arrays. This algorithm performs source localization on bases of TDOA technique [9]. Source localization also implement in Human to Machine Interface. SRP-PHAT algorithm is little bit complex and expensive when compare to computations. SRP-PHAT localizes the source by estimating the TDOA between pair of microphones. In a closed room environment reverberations are present with the speech and noise signal in such cases it is difficult to estimate the localization of the speech source with SRP-PHAT. But it is not a complex if beamformer is present as it can reduce the reverberations.

1.2 Problem Statement

Conference rooms are the far field environment which implements microphone sensors for tracking of the speech signal. But the microphone sensors receive both statistical data of the speaker and the undesirable signals (acoustic, reverberated and unwanted human signals).

This thesis is aim to illustrate the problem of Speech enhancement, dereverberation and localization in far field environment. The main focus of the thesis is to bring into possibility a sophisticate algorithm that can perform speech enhancement in the far field environment.

1.3 Hypothesis

The goal of the thesis is to provide a reliable Beamformer technique that can provide a satisfactory solution for the problem statement. This thesis presents a series of assumptions and theories about the generation of the reverberation for attaining particular room environment, signal and sensor models, basic information about various beamformers and speaker localization technique.

In this thesis two beamformer techniques are implemented one is wiener filtering technique and another one is sub band wiener technique for speech enhancement in linear microphone environment for far filed assumption. This thesis also deals with the speaker localization in the meeting room so a reliable SRP-PHAT algorithm is implemented for this purpose.
1.4. Thesis Overview

The thesis is divided into nine chapters. The chapter 1 gives a brief introduction and motivation for the thesis.

Chapter 2 discuss about the image source model for creating a closed room environment and generating reverberations and generating of multi channel environment.

Chapter 3 provides the necessary information for the generating of the WOLA filter bank and their implementations in the speech processing.

Chapter 4 deals with the providing of the information about the microphone arrays and implementation of array processing in speech processing algorithms.

In chapter 5 speech enhancement algorithms are provided. The objective of this chapter is to provide the basic information about the different beamforming techniques and implementation of beamformer in the application of speech enhancement and noise cancelation.

Chapter 6 provides the experimental implementation of the two beamformer techniques proposed. The two beamformers are linear microphone array wiener beamforming and multichannel sub band wiener beamformer. The flow of signalling in both of the beamformers is clearly mentioned.

Chapter 7 provides the information about the speaker localization. Here how SRP-PHAT provides the solution for the speaker localization is clearly discussed.

Chapter 8 provides the simulation results for the chapter 6 and chapter 7 and the performance of the algorithms in various noisy fields.

Chapter 9 finally provides the concluded information of the thesis and need for extension in future.
2. Room Modelling

This chapter describes the problem with the acoustic noises and reverberations in the far field environment. Here discuss the implementation of Image source model for generation of reverberation for particle far field room environment case.

2.1 Introduction

In order to achieve successful communication system it is compulsory that the emitter and receiver use same conditions. In other words the channel between the emitter and transmitter must be suitable for transmitting message.

In any communication system the channel is affected by several sources of distortion that can affect the success of the communication. Some of the distortion affects are acoustic noise, reverberations and interference speech of the other speaker instead of active speaker.

In far field microphone environments like meeting rooms where sensors are placed far apart from the source leads to the decay in the speech signal. Decay in the speech signal in far field environment is caused by two factors one is acoustic noise and the second one is reverberation.

2.2 Acoustic Noise

Acoustic noise refers to the undesirable sound effect that is summation to the speech signal when it is received by the microphone. Acoustic noise does not refer to statistical, frequency, spatial or propagation characteristics.

If it is a conference room except the active speaker all other speaker and any undesirable sound effects like (fan sound, door sound) are taken into account as noise. In this case noise is considered as speech signal too.

In reality case noises are considered into two categories, one is directional noise and another one is non directional noise.

Non directional noise is generally considered as back ground noise to the microphone which is considered as noise coming from everywhere. Here in the thesis white noise is considered as the non directional noise ass the white noise is uncorrelated with the speech signal.
2.3 Reverberations

The acoustic signals are propagated in multi path in a closed room that means sound signal generated from a speaker will flow in direct path and multiple paths to the receiver (human or mic). The mic at the receiver will record the data of the direct path sound wave and multiple path waves those are generated by the reflection of the walls in the closed room. These multiple paths will depend on various constrains they are reflection coefficient, absorption coefficient, surface of the room and size of the room. The multiple paths sound waves are different when compared to distant echo that is generated after 50 to 100ms after initial sound wave [15]. These multiple paths sound waves are commonly referred as reverberations.

In real world the existence of reverberation after the sound wave is observed is clearly observe in large closed rooms. This phenomenon is observed by reverberation time in a room. In technical format reverberation time is the time taken for soundwave in a room to decay by 60dB.

The reverberation time is regulated by two factors those are size of the room and surface of the room, there are so many other factors like objects in the room, temperature etc which are also effect the reverberation but these are two most important. Surface of the room plays a crucial role because if the surface is reflective material then the number of reflections increase or else it is an absorption material then the number of reflections decreases. The Sabine’s equation illustrate the relation among the volume of the room ($V$ ($m^3$)), the area of each surface($S_i$ ($m^2$)) and its absorption coefficient $a_i$ is shown in equation (2.1)

$$T_{60} = \frac{0.161 v}{\sum i a_i S_i}$$ (2.1)

The nature of reverberation in a closed room is characterised by the speaker to receiver room impulse response [15]. The figure (2.1) display indicates the room impulse response of a closed room in which x axis is time taken for the impulses and the y axis is the amplitude of the impulses. RIR is union of three things direct path, early reflection and late reflections. Early reflections are just some delayed and attenuated from direct sound wave. They are generated from the near by surface and they are used for increasing the strong audibility of sound and special impression [20]. Early reflections are essential in most of the applications.
In the case of late reflections the rate of reflections increases and the interval time between the reflections decreases. These are dense pack of echoes and travel in all directions. In most of the cases late reflections causes spreading the speech spectra which leads to lack of audibility in the speech. In some of the application the presence of reverberation increases the quality of the hearing experience so the effect of reverberation depends on the applicable field. In the field of microphone array reverberation causes inaudibility and causes problem in identifying the speaker position.

![Graph showing early and late reflections](image)

Fig. 2.1. Schematic diagram of a typical speaker to receiver room impulse response three parts can be easily distinguished: direct wave, early reflections and late reflections

So in experimental set of speech processing the reverberation makes an crucial role for their existence whether in the form of undesirable signal or usefulness. In experimental case reverberation of a room are generated from different methods like ray and beam tracing methods, right now we are using the image source method for their generation.

### 2.4. Image Source Model (Room impulse response)

Image source method becomes a pre dominant tool in most of the research fields of speech processing [28]. Because of its simplicity and flexibility for achieving results it is widely implemented in acoustic and signal processing [22]. ISM is implemented to generate impulse responses for a virtual source in a room. ISM’s one of the most important implementation is, performance assistant of various signal processing algorithms in reverberation environment. Few implementations of ISM are validation the blind source separation, channel identification, equalisation, acoustic source localization and many others. In such cases ISM
is participate to test the specific algorithm in order to decide its endurance in different environmental reverberation [28].

2.4.1 Over view of Image source model and increase its efficiency with Thiran all pass filter

Image source model approximate the room impulse response of multiple reflections from a single source in a room by transferring them into direct paths from multiple virtual sources.

The virtual sources are the mirror images of the original source in the room [26, 27]. The figure shown below is a two dimensional structure of an original source and their mirror images of it. Although mirroring is performed in three dimensional structural but for illustration purpose two dimensional structural is shown

In the figure (2.2) original room is the one in which black star and green circle is actual position of the source and all black circles are the virtual sources in mirror images of the original room. Black star is mic position, green circle is source and black circles are virtual sources. For simple illustrate the direct path and reverberated path here simple figure (2.3)
In the figure (2.3) path between the source and mic with out any deflection is direct path. Remaining two paths are reverberated paths as both are having deflections. Black path indicates the original sound waves and the blue path indicated the reverberation of the sound wave. The perceived path and the original path are same with respect to distance, absorption coefficient when the sound wave hits the wall the only difference is transferring multiple reflections of a single source into direct paths of multiple virtual sources.

2.4.1.1 Generation of discrete time domain room impulse response

The equation (2.2) is the time domain representation of room impulse response

\[ h[t] = \sum_{i,j,k=-n}^{n} \alpha_{i,j,k} \beta_{i,j,k} \gamma_{i,j,k} \]  

(2.2)

Variables in the equation (2.2) are described as follow \( \alpha \) is the reflection coefficient, \( \beta \) is the propagation attenuation and \( \gamma \) is the unit impulse response function. \( \alpha \) and \( \beta \) have the capacity to alter the magnitude of the each impulse of the room impulses response. Detail description of these two variables are described below.

Locating of virtual sources

ISM is a well established model for simulating RIRs in a given room. Here Cartesian coordinates system\((x, y, z)\) are assumed for a enclosed room [28]. \( i, j \) and \( k \) are the reflection indexes. The locating of the virtual source is formulated with respect to one of the coordinates of the room is

\[ x_i = (-1)^i x_s + \left[ i + \frac{1 - (-1)^i}{2} \right] x_r \]  

(2.3)
Where $x_i$ is x coordinate of the virtual source, $x_s$ is the x coordinate of the sound source and $x_r$ is the length of the room in x dimension [14]. Similarly we can calculate for y and z coordinates of the virtual source. Reflection indices i, j, k are along the spatial axes of the room. As a part these indices can be positive or negative, if the indices i, j, k equals to [0, 0, 0] then the virtual source is original sound source. Virtual source with reflection indices i, j, k is shown is equation (2.4)

$$V_{i,j,k}=(x, y, z)_{i,j,k} \quad (2.4)$$

**Unit impulse response and Propagation Attenuation**

The direct path propagation vector between the virtual source and microphone is expressed in equation (2.5)

$$d_{i,j,k} = (x_m, y_m, z_m) - V_{i,j,k} \quad (2.5)$$

$(x_m, y_m, z_m)$ is the microphone coordinates and $V_{i,j,k}$ is virtual source location. Before generating unit impulse function there has a principle importance for find out the time delay for each echo let commence the time delay with respect to a function for generating unit impulse response. This time delay is also called as propagation delay is

$$\delta_{i,j,k}(t) = t - \frac{d_{i,j,k}}{c} \quad (2.6)$$

In the equation (2.6) $t$ is the time, $d_{i,j,k}$ is the propagation distance and $c$ is the speed of the sound. $\frac{d_{i,j,k}}{c}$ is the effective time delay for each echo. Now it is time to generate the unit impulse response function. It is formulated as

$$y_{i,j,k}(\delta_{i,j,k}(t)) = \begin{cases} 1 & \text{if } \delta_{i,j,k}(t) = 0 \\ 0 & \text{otherwise} \end{cases} \quad (2.7)$$

**Thiran all pass filter**

In the simulation environments generation of fraction values is not as easy as in the case of theoretical expressions. In generating impulse response for each echo, time delay plays an important role. This time delay may be round off value or fraction value. For obtaining fractional time delays, thiran all pass filter is implemented. Thiran is popular because of it flat magnitude response and more concentration on the phase response. Transfer function for digital IIR all pass filter is formulated in equation (2.8)
Where \( N \) is the order of the filter and \( D(Z) = 1 + a_1 Z^{-1} + a_2 Z^{-2} + \cdots + a_N Z^{-N} \) is the denominator polynomials with real value coefficient \( a_k \) and the numerator polynomials are the reverse version of the denominator ones [30]. Thiran formula for all pass filters is

\[
a_k = (-1)^k \binom{N}{k} \prod_{n=0}^{N} \frac{d+n}{d+k+n}
\]

(2.9)

d is the delay parameter, \( k=1, 2, 3, 4 \ldots \ldots N \), \( a_k \) will generate the coefficients for the IIR all pass filter in equation (2.9).

**Propagation attenuation**

Propagation is one of the factors that reduce the magnitude of the echoes. Not all the echoes have same magnitude; their magnitude will depend on a factor called as propagation attenuation. Propagation attenuation is expressed in equation (2.10)

\[
\beta_{i,j,k} \propto \frac{1}{\|a_{i,j,k}\|}
\]

(2.10)

**Reflection Coefficient**

Reflection coefficient is another factor which will directly affect the magnitude of the echoes. Sound wave experiences partial reflection and partial transmission when it hit to walls of a room. Reflection coefficient is the amount of sound wave reflects when it hits a surface (walls, objects in the room). In the thesis concern coefficient is not the ratio between reflection and transmission sound wave, here coefficient means amount of sound wave reflection after sound wave absorb by the surface. The most important factor in the ISM is found out the total number of the reflections that engage in the room. Indexing schema \( |i|, |j| \) and \( |k| \) are important for find out the total number of reflections. Let’s take wall reflection coefficient as \( \alpha \) and raise exponent \( n \) where \( n = |i| + |j| + |k| \) will give the total reflection of the sound wave is given in equation (2.11)

\[
\alpha_{i,j,k} = \alpha^{|i| + |j| + |k|}
\]

(2.11)

Here \( \alpha < 1 \), reflection coefficient never greater than 1. If each wall has different reflection coefficient then it will lead to some more complex equations. If \( \alpha_{x=0} \) is the reflection coefficient of the wall perpendicular to the x-axis near the origin and \( \alpha_{x=xr} \) is reflection of
the wall opposite that, then the combined reflection coefficient by the \(i^{th}\) virtual source is below [14].

\[
\alpha_{xi} = \alpha_{x=0} \left| \frac{1}{2} - \frac{1}{4} + \frac{1}{4}(-1)^i \right| \alpha_{x=xr} \left| \frac{1}{2} + \frac{1}{4} - \frac{1}{4}(-1)^i \right|
\]  
(2.12)

Similarly we can calculate the \(\alpha_{yj}\) and \(\alpha_{zk}\) with the indices \(j, k\) in order to generate the total reflection coefficient in equation (2.13)

\[
\alpha_{i,j,k} = \alpha_{xi} \cdot \alpha_{yj} \cdot \alpha_{zk}
\]  
(2.13)

The propagation attenuation and the reflection coefficient are directly propositional to the total number of reflections. As the total number of reflections increases then the both attenuation factors increases but this should be limited with the index schema to below reference order i.e.

\[
|i|+|j|+|k|<N_{ref}
\]  
(2.14)
3. Filter Bank Design

This chapter provides basic information about time-frequency domain.
Overview and brief information about filter bank architecture. Designing procedure of the WOLA analysis and synthesis filter bank.

3.1 Introduction about time-frequency domain

Time frequency domain examines the techniques that study the signal in both time and frequency responses simultaneously. Discussing about time and frequency domain separately, time domain graph shows how a signal varies with time where as in frequency domain graph shows how much signal lies in a given frequency band for a given frequency range. These frequency components in frequency domain are combined by applying phase shift to each sinusoid to recover the time response. The signal characteristic changes from time domain to frequency by Fourier Transform. Inverse can be achieved by inverse Fourier transform. The prime motive for developing the time-frequency domain is, it is implemented for short time interval signals where Fourier transform is inefficient for short interval signals. Fourier transform assume signals are infinite duration.

The most basic form of time-frequency domain analysis is short term Fourier transforms. This time-frequency domain is also called as hybrid domain. Some other examples of time-frequency domain analysis are wavelet transforms, Wigner distribution function, Gabor-Wigner function. To harness the power of a frequency representation without the need of a complete characterization in the time domain, one first obtains a time–frequency distribution of the signal, which represents the signal in both time and frequency domains simultaneously. In such a representation the frequency domain will only reflect the behaviour of a temporally localized version of the signal. This enables one to have a brief idea about signals whose component frequencies vary in time.

Most of the time-frequency domain analysis like STFT will have window function for materialising the small part of the signal. This window function can be compared to a filtering technique. Time-frequency domain can be achieving by filtering the signal over a series of filter banks. In most of the cases these filter banks will be band pass filters. These filter banks decompose the signal into sub band signals when signal pass through them.
3.2 Filter bank

In signal processing filter banks are used to perform short time spectrum analysis [25]. The most important one is the sum of individual frequency of band pass filters should be flat with linear phase. Filter banks is an array of band pass filter that slices the total band width of the input signal into sub bands, each with a certain bandwidth component of the input signal. Filter bank increase the efficiency of the process as the complex solution is sub divided into smaller ones. Simple plot of eight sub band filter bank is show in the figure(3.1) [15].most case implemented filters are Finite impulse (FIR) response filter and Infinite impulse response (IIR) filter. Both are having their own advantages and disadvantages in various segments. FIR filter is most common term in literature where as IIR filter is less popular because of its designing is more complicated. Although IIR filters have many advantages like lower complexity, shorter delay and better frequency selectivity [21].

3.2.1 FIR filter

Digital filter gain significant importance during the design filter banks. These filters in filter banks are constructed by Fir filter or IIR filter. Both the filter have there own gain and loss, FIR filter is easy to implement. General equation of a output signal $y(n)$ at a particular time when a FIR filter by a set of $N$ coefficients ($h(k)$) is implemented to input signal $x(n)$ is given in equation (3.1).

$$y(n) = \sum_{k=0}^{N-1} x(n - k) * h(k)$$  \hspace{1cm} (3.1)

$x$ is the input signal, $y$ is the output signal, $h$ is the filter and $*$ is the convolution operation

Filter $h$ coefficients will significantly varies depend on the implemented filter type. In general cases of the filter bank, band pass filters are implemented because they are completely flexible in choosing the center frequencies. FIR filter equation for construction band pass filter is implemented in equation (3.2)

$$h_k(n) = \begin{cases} 
    h_{lk}(n) \cos(\omega_c n T) & 0 \leq n \leq N - 1 \\
    0 & \text{otherwise}
\end{cases}$$  \hspace{1cm} (3.2)

$h_{lk}$ is the impulse response of a $k^{th}$ order linear phase low pass filter $\omega_c$ is the cut off frequency [21]. The performance of the fir filter is enhanced by window function.
3.2.2. Brief description about filter Bank

Filter bank will be segregated into Analysis and synthesis filter bank. The general structure of the filter bank is equally divided between them. The raw view of the analysis filter bank starts with the band pass analysis filters. These analysis filters engage a bank of $K$ band pass filters to bind the bandwidth of the input signal to a frequency band of that band pass filter. From here the signal is received by a decimator factor $D$ in order to reduce the number of samples of the sub band signal. $h_k$ is the impulse response of each sub band filter. Where $k=0$ corresponds to DC frequency and $k=K/2$ is the Nyquist rate [15]. Processing block consist of wiener beamforming technique for reducing noise and localization of sound source. Processing block is followed by synthesis block for reconstruction of original speech signal. The synthesis part will employ interpolator factor $D$ in order to increase the sampling rate. This is followed by a synthesis filter bank in order to reconstruct the original signal from low rate sub band filters. Figure 3.2 is the simple block diagram of a filter bank.

3.3. Weighted overlapping Add (WOLA)

WOLA method is implemented in analysis and synthesis filter bank for overlapping the adjacent windows. The explanation of WOLA filter bank is performed with respect to the figure (3.3). However this will introduce time domain aliasing in the analysis part but this can be overcome in the synthesis for perfect reconstruction of the signal. WOLA is an efficient method for discrete convolution of the large signal with FIR filter. WOLA filter bank is extensively implemented in low power consumption technology like deep Submicron
technology. WOLA design provides high degree of flexibility in sub band coding, sub adaptive algorithms [18]. Analysis window $w[n]$, $L$ length of the analysis window, decimation ratio and synthesis window are the four variables comprises to form a efficient WOLA filter bank.

![Fig. 3. 2. Block diagram of filter bank](image)

### 3.3.1. WOLA analysis filter Bank

In the analysis filter bank, the input signal is slice into frames (block). Each frame of size $D$ is stored by the input FIFO buffer of length $L$. The data in the FIFO buffer is windowed by the analysis window $w[n]$. This processed data in stored in the temporary buffer $t_1[n] = u[n].w[n]$, of length $L$ samples [15]. The resulting vector is added modulo $K$ (i.e. folded) and stored in another temporary vector $t_2[n]$ [18], which can be expressed in equation (3.3)

$$t_2[n] = \sum_{m=0}^{L-1} t_1[n + mK]$$  \hspace{1cm} (3.3)

Circular shift is performed for the temporary vector $t_2[n]$ by $K/2$ samples in order to provide the zero phase signals for FFT. Then the FFT of the resulted window is time segmented. The output of the analysis filter is expressed with respect to magnitude and phase response [18]. This can be clearly shown in the block diagram (3.2) of WOLA filter bank.
3.3.2. WOLA synthesis filter bank

Synthesis filter bank is responsible for perfect reconstruction of the original speech signal. The input signal to the synthesis filter bank is followed from the wiener beamformer. The data collect from the beamformer is received by the synthesis filter bank for further processing to generate a modified time domain signal. The synthesis stage initiate with the implement of size K IFFT to process sub band signals from the beamformer. Circular shifting of K/2 samples is performed on the IFFT output signal in Synthesis stage for counterpart the circular shift that is performed in the analysis stage. Temporary buffer \( t_3[n] \) is generated for storing the data from the circular shifting. These vectors in buffer \( t_3[n] \) are again transferred to \( t_4[n] \) of length \( L/D_F \) where \( L \) is the length of the synthesis window and \( D_F \) is the synthesis decimation factor. The resulting \( t_4[n] \) is then windowed with the synthesis window and the result is accumulated in FIFO buffer. This generated samples are then shift out of the output FIFO and finally D zeros are shifted into the output FIFO, this is performed to balance the decimation perform in the Analysis filter bank. Total process again repeats for the next block [15].

3.3.3. Decimation

In Digital signal processing decimation is a technique for reducing the sampling rate of a signal in analysis filter bank. But this can be compensated by an interpolation procedure in synthesis filter bank. There are K sub bands for representing the signal but they can represent only 1/K part of the frequencies of a signal. So decimation is implemented in analysis filter bank which is represent by a factor D i.e. decimation rate, then the analysis filter bank output will be \( x_k[nD] = (h_k * x)[nD] \)[23]. Decimation factor is represented by the no of sub bands by oversampling ratio. Oversampling ratio is represented by O. O=K/D if O=1 then number of sub bands is equal to decimation rate which means filter bank is critically sampled. In general cases O=2 which means factor two oversampled [23]. Naturally decimation procedure in analysis filter bank is compensated by interpolation procedure in synthesis filter bank. If a signal is down sample by a factor D then it new sampling frequency should be \( f_s/ D \). Where \( f_s \) is the previous sampling frequency before decimation is performed.

3.3.4. FFT/IFFT Transform

Frequency domain increases its popularity in digital signal processing over time domain, this leads to computation of most of the algorithms in frequency domain. Fourier transform will convert the time domain into frequency domain. Signal will be continuous or
discrete but in the digital signal processing signal are discrete in form so this lead to increase the implementation of Discrete Fourier Transform (DFT). DFT decomposes a sequence of values into components of various frequencies. The main drawback of DFT is its computational complexity. DFT has more floating point multiplications than additions which lead to increase its computational complexity.

An alternative to DFT is FFT which compute the same results with less number of computations. The total number of computational operations for a N point FFT is $O(N \log N)$.
where as for DFT is $O(N^2)$. In most of the case FFT is depend on the twiddle factore $\frac{2\pi}{N}$. There are so many methods for computations FFT one of the efficient method is Decimation In Time radix N point FFT [17]. Inverse DFT is similar to DFT but a positive in exponential and 1/N factor. Any FFT algorithm can accommodate to it.
4. Multi-Microphone processing

This chapter shows the main contribution of the thesis. This chapter discusses about the speech enhancement with the implementation of microphone arrays to the drastically degraded speech signal. Brief information is provided on the microphone array signal processing and beamforming techniques.

4.1. Array signal processing

Array signal processing is derived from the simple concept of sharing the load among various sources. The information provided by the multiple sensors is more reliable than the information provided by the single sensor. These arrays of sensors receive samples of signal from different spatial position this will create diversity in space domain in terms of time domain. Space diversity is the term used in many applications like wireless communication, speech processing, radars, sonar and many other applications. It is possible to make correspondence between signal processing based on time diversity and array signal processing on special diversity [15]. The information provided by the each sensor will depend on the direction of source located. The information received by each sensor at a time instant depends on the delay of arrival between each sensor (relative positions of sensors) and also on the temporal frequency of the signal. The basis of the array signal processing counts on the information of the source position that is involved in the phase signals received by the each sensors and in the correct alliance of these signals. In most of the cases in array signal processing source is assumed to be far field and the signal receives from the source is a plane wave.

There are different types of array processing with respect to the geometrical arrangement of the sensors. Some of them are linear array, rectangular arrays, perimeter array, Random ceiling array, Endfire cluster. In this thesis linear array is implemented. The advantage of linear array towards remain other is geometric flexibly, less computational, easy to implement.

4.2. Linear array processing

In linear Array, the multiple pieces of information of a signal are tracked by sensors that are arranged in linear array. Each sensor is placed in linear order with an unified separable distance. Linear array processing is the simplest array processing algorithm for generating effective beamforming algorithm and find out the time directional of arrival to
identify the source localization. The distance between the linearly arranged sensors should be fixed carefully in order to reduce the aliasing problem.

4.2.1. Basic concepts of Linear array processing

Wave propagation

Wave propagation is the direction in which wave travels. The information about the speech is carried by the waves [15]. The physical properties of a wave is governed by the wave equation

Wave equation describes the medium of propagation with some boundary condition is shown below equation (4.1)

\[ \nabla^2 S(r,t) = \frac{1}{c^2} \frac{\partial^2 S(r,t)}{\partial t^2} \]  

(4.1)

\( S(r,t) \) is a general scalar field which represents the electric field \( E \). \( \nabla^2 \) is a laplacian operator. \( C \) is the speed of the wave. \( \partial^2 \) is a derivative operator. \( S(r,t) \) represents the wave equation which will vary according to the nature of the wave.

Time delay of arrival

The information from a source is received by the sensors at different instance of time. The time delay taken for each sensor to receive information can be formulated as

\[ x_q(t) = x_0(t - t_0 - \tau_q) \]  

(4.2)

\( t_0 \) is the propagation delay to the each sensor. \( \tau_q \) is the delay with respect to reference point that will show each sensor with respect to its position and direction of arrival of sound. \( \tau_q \) can be calculated from a phase alignment transform which can be discussed further in the next chapter. No sensor will have the same time delay of arrival.

Spatial aliasing

Due to the coupling of the temporal and the spatial domains there arises so many problems, one of the most important problem is the spatial aliasing. It is clearly illustrated that spatial domain of the signal depends on the direction of arrival of the sound wave, distance between the sensors, and carrier frequency of the source signal [15]. If a sound wave is generated from one or more directions then a beamformer will enhance one as desirable and other sound wave as undesirable. This is call spatial aliasing.

Spatial aliasing is also caused if the distance between the sensors are more than the half the wavelength of the sound wave. This is caused if the sensors are placed too far from
each other. The condition for eliminating the spatial aliasing in the case of distance is given as \( d \leq \frac{\lambda}{2} \). Where \( d \) is the distance between the sensors, \( \lambda \) is the wavelength of the receive signal at the sensor. The critical case for spatial aliasing is \( d = \frac{\lambda}{2} \) if \( d > \frac{\lambda}{2} \) then spatial aliasing will arise in the array processing.

4.3. Microphone array processing

The performance of the single channel system has limitation for improving the speech quality when compare to multiple channels system in speech processing algorithms. Due to the increase in demand of multiple channels application lead to excessive implementation of the microphone arrays. Microphone array processing is the pre processing stage in order to enhance the speech signal. Microphone array processing basically consist in the designing of the spatial filter capability for selecting the particular space direction and eliminating the all other [15]. Before generating and implementing the beamforming algorithm with the microphone arrays, the geometry of the array must be discussed [32]. The characteristics like number of microphones, spacing, and array arrangement must be established. There are too many array architecture arrangements like linear, square, circular, logarithmic and many more. Depend on providing the optimistic solution array architecture is determined. In the present case of speech enhancement and localization linear and square array are best suit and they are the most common practise in speech processing. Linear array have more advantage like less computational time, less complexity than square array. One of the draw back of the linear array is it is operated in two dimensional space when compare to square is implemented in three dimensional space. In an environment like multiple speakers two dimensional linear array architecture is suitable. So this research utilises linear array architecture.

4.3.1. Linear microphone array

Linear array processing is the basic concept in many applications like communication system, radar, speech processing. Depend on the application; sensor’s physical characteristics will change. In the case of speech enhancement microphone is used as a sensor for tracking of the speech and localization of the source. Implementation of linear microphone array for automatic speech reorganisation is investigated in this session. Most of the work of the beamformer conducts on the linearly equispaced microphone arrays. Linear microphone array processing is suitable
for narrow frequency ranges that depend on inter microphone spacing. The time taken for each microphone to receive speech is described as

$$\Delta t = \frac{(m-1)dsin\theta}{v}$$  \hspace{1cm} (4.3)

Where m is the microphone number, d is the distance between microphones, $\theta$ is the direction of arrival of speech signal. The pictorial representation of the linear microphone array is described in figure (4.1)

![Diagram of microphone array](image)

**Fig.4.1.** propagating of far field sound wave with microphone array

Number of microphone plays a crucial role with respect to the shape of the microphone array for speech enhancement. Microphone array aperture depends on the number of microphones use in the array. Increase in microphones will lead to increase the size of aperture of the array. This means number of microphone is directly propositional to the size of the array’s aperture. With the increase in the aperture, then the resolution is also increase because of the increase in the spatial filter of the array. Speaker localization can be performed effectively with the high resolution. But infinite long aperture is not the solution for getting the precise location of the speech signal. In practical case infinite aperture is not possible so size of the aperture depends on application.

Microphone placing plays a crucial role in the performance of the beamforming. As discussed in the above, the concept of spatial aliasing in array processing. The distance
between the microphones is governed by the condition in the spatial aliasing. Spatial filtering is the primary tool in beamforming that is utilised when the microphone array extract the information from specific location. It is similar like nyquist criteria in frequency sampling. Further the condition for the spatial filtering is discussed clearly in the spatial aliasing in the array signal processing.

4.3.2. Source localization with microphone arrays

A major field of microphone array processing is detection of source and its localization. The location of the speaker is detected by the microphone arrays [32]. Microphone arrays can also detect the angle of direction of arrival of speech, number of speakers and to track the position of the speaker [32]. Source localisation is the most important in some of the application like gaming, videoconference, and teleconference. Spatial aliasing place an important role for source localization, source localization is determined by the time difference of arrival between the microphones. Source localization benefits from the large spacing between the microphones where as source extraction benefits form the less spacing between the microphones. Therefore both source localization and source extraction is difficult to implement in array.

4.4. Alternative to microphone approach

Microphone approach based on thee array signal processing for speech enhancement can be compensated with alternative approaches. Some of the techniques that are implemented as an alternative to microphone array approach are binaural processing, blind source separation and multichannel dereverberation.

4.4.1. Binaural processing

Binaural processing is based on the binaural signals. Binaural signals are the signals that are implemented in the human system for identification, recognise and focus on different sound source. The information about the speech is received by the binaural signal with the help of ears.

Based on the binaural signal one can indentify the source localization by implement these signals in real time application. Most of the binaural application depends on two fundamental cues; those are Interaural Level differences (ILD) and Interaural time differences (ITD) [19]. That is amplitude and time difference of two observed signals by the ears.
4.4.2. Blind source separation

Blind source separation (BSS) depends on the demixing matrix. This estimated from the independent compound analysis. BSS technique has no prior knowledge of the microphones sensors or the source physical characteristics. BSS estimates the demixing matrix by independent compound analysis (ICA). ICA does not have any information about the number of sensors or source characteristics. ICA estimates the values by maximisation of appropriate independence criteria by measuring the degree of independence of the demixing output signals.

4.4.3. Multichannel Dereverberation techniques

It is the best for the speech enhancement algorithms. The name it self describes that it will remove reverberation of the signal that is received by the sensors by developing the inverse filter. Inverse filter will remove the reverberations and equalise the speaker to receiver impulse response. It is hard to develop the inverse filter but multiple channels can create the inverse filters with the help of match filters by estimating the transfer function between the source and receiver.

4.5. Other microphone array considerations

In most of the Microphone array applications it is assumed that, the plane wave propagation and narrow band. These assumptions do not affect the result in most of the cases but in some cases they can seriously affect the validation results. Some of the most important affects that deserve to be mentioned are ideal propagation channel, punctual emitting sources, calibrated and isotropic sensors.

    Ideal propagation channel is assumed theoretically to be linear and no distorting. In real-time channels show some typical characteristics like dispersion, attenuation, refraction and diffraction. These characteristics do not affect the performance of the channel at all the time but in some cases results will be hazard by them.

    In most of the case it is assumed that the source is punctual and emitting from single point from the space but in most of the case the source is distributed and this assumption can not be consider.

    The last one is characteristic of a sensor; these characteristics change from sensor to sensor. Different sensors show different gains and simultaneously each sensor has direction and frequency dependent. If the sensor is punctual then there is a need of considering the calibration of sensor into account.
5. Speech Enhancement with Beamforming

This chapter shows the main contribution of beamforming for speech enhancement with the present state of art in beamforming. Significant work on the beamforming for speech enhancement is briefly describes.

5.1. Introduction about Beamforming

Beamforming is a spatial filter technique implements to insulate the speech signal based on its position in the room. The technique was first introduce in the radio technology by a way of collecting the antenna information from an array antenna dishes it was introduce in 1950’s. Then this technique is implemented lately in speech processing for enhancement of the speech signal in mid 1970’s. Beamforming exploring as a general signal processing algorithm implemented in most of the applications where a cluster of sensors are used. It is used in sonar technology in submarines for detection of enemy vessels using hydrophones in order to enchase the ability of ground sensors for detection of vessels.

1970’s is the era for implementing the microphone array beamforming in audio signal processing and it is the time for microphone array beamforming for active area of research [24]. The implementation of audio beamforming gain importance in the applications like hand free environment, speaker tracking in conferences, human to machine interface, text to speech conversion and a lots of other applications.

The present quality of thesis shows significant improvement in the SNR (Signal to Noise ratio) with the implementation of microphone array in audio beamforming. With the implementation of the wiener filter beamforming in reverberated environment in order to reduce the reverberation and non directional noise is studied. Here wiener beamforming technique is examined in two platforms; time and frequency domain. Most of the thesis is concentrated on frequency domain as it is best platform to perform both speech enhancement and speaker localization.

5.2. Beamforming for Speech Enhancement

The spatial filter operation (Beamforming) can be represent with the input vector equation and the weight \( \mathbf{w} \) vector applied to each sensor of the array can be represented in equation (5.1).

\[
y(n) = \mathbf{w}^H \mathbf{x}_n
\]  

(5.1)
\( \mathbf{X}_n \) is a vector which carries information about speech signal noise to the sensor. It can be represented in the following steps

\[
x_t = a(t) \mathbf{s} \exp(j2\pi f_0 t)
\]  

(5.2)

The input signal \( x_t \) is represented at a sample time instance \( t \). Where \( a(t) \) is the complex envelope constant or amplitude envelope of a monochromatic plane wave that will vary at every instant of time. \( a(t) = |s(r, t)| \), where \( s(r, t) \) is the monochromatic plane wave. \( f_0 \) is the carrier frequency. In \( \exp(j2\pi f_0 t) \), \( 2\pi f_0 t \) is the argument function of the wave and \( \mathbf{s} \) is the steering vector which has the information about the phase delays at each sensor.

\[
\mathbf{s} = [\exp(-j2\pi f_0 \tau_1), \ldots \exp(-j2\pi f_0 \tau_q), \ldots \exp(-j2\pi f_0 \tau_q)]^T
\]  

(5.3)

For the case of simplicity from now onwards it is going to omit the carrier equation in the speech signal. Now it is going to consider that the non directional uncorrelated noise is added to the speech signal represent as \( \mathbf{n}_t \) and the equation (5.2) can be modified as equation (5.4)

\[
x_t = a(t) \mathbf{s} + \mathbf{n}_t
\]  

(5.4)

This speech signal receives at the sensor is represented in digital notation as

\[
\mathbf{x}_n = a(n) \mathbf{s} + \mathbf{n}_n
\]  

(5.5)

Substituting the input signal \( \mathbf{x}_n \) is the beamforming equation to get more simplified equation

\[
y(n) = \mathbf{w}^H a(n) \mathbf{s} + \mathbf{w}^H \mathbf{n}_n
\]  

(5.6)

The beamforming response at particular angle and for a concern frequency can be expressed as the product of the weights of the beamformer and the steering vector of particular angle at that concern frequency

\[
|H(f, \varphi)| = |\mathbf{w}^H \mathbf{s}_\varphi|^2
\]  

(5.7)

It is preliminary discuss that the beamformer work is to estimate the weights in order to enhance the speech signal or desire signal and eliminate the non desired components or noise. Basically beamforming design method is divided into two, one is data independents and other one is statistical optimum method.
In data independent method, the functionality of beamformer is illustrated by the name itself. Here fixed beamformer is design which does not depend on the input data even though the data is bounded to spatial restrictions or any other boundary conditions.

In statistical optimum methods a beamformer is designed based on the statistic of the receiving data to design a optimise function that makes beamformer optimum. Basically optimum functions are designed in order to minimize the output noise power of the beamformer. This kind of beamformer has some problems with spatial restrictions resulting in conditional optimal problems which may leads to cancelling the desire data. The problem with the spatial restrictions can be over by forcibly injecting the steering vector to the desire signal so the beamformer will have unity response particular direction and eliminates the all other directions. In the same way other restrictions can also inject in additions to get better response.

5.2.1. Fixed Beamforming

Beamforming in which the direction of response is fixed to particular azimuth and elevation is known as fixed beamforming. Beamforming is fixed to particular direction because of the fixing array strategies to particular direction. Most conventional implementations of fixed beamformers are delay and sum beamformer, filter and sum beamformer and super directive beamformer.

Delay and sum beamformer (DSB)

It is the simplest beamformer architecture in the fixed beamforming technique. DSB is a combination of different microphone signals for guiding the different path length from source to microphones. Output is the combination of these signals which can be expressed as equation (5.8).

\[ y(n) = \sum_{k=1}^{K} \alpha_k x_k(n - \tau_k) \quad (5.8) \]

\( \alpha_k \) is the weight given to each microphone, \( \tau_k \) delay for balancing the propagation delay. In most of the cases \( \alpha_k \) is equal to \( 1/K \) which means average of the aligned signals. Depend on the propagation model (far field or near field) one can derive weights. Obtaining \( \tau_k \) is the most important and problematic in DSB. The main positive sign of DSB is its simplicty of implementation and in most of the cases the result are convincible. The main draw back of delay and sum beamformer is its inefficiency in directive noise presence. Simple block diagram of DSB is illustrated figure (5.1).
Filter and sum beamforming (FSB)

It is the generalise version of the delay and sum beamformer. The difference between two beamforming techniques is implementation of filter bank. Here simple summation of the microphone array signals is performed. Each microphone signal is filtered with the corresponding filter depends on the channel before summation. FSB can be expressed as

$$y(n) = \sum_{k=1}^{K} h_k(n) * x_k(n - \tau_k) = \sum_{k=1}^{K} \sum_{l=0}^{L-1} h_k(l) x_k(n - l - \tau_k)$$

(5.9)
From the equation (5.9) $h_k(n)$ is the filter for the microphone $l$ of length $L$. When compare to DSB, FSB can equip more sophisticated and accurate results. FSB has more specified array response that is not possible in DSB. The main drawback of the FSB is complexity of designing the filter bank and increase of computational complexity. The output of the FSB with respect to frequency response can be expressed in equation (5.10)

\[
Y(t, f) = \sum_{k=1}^{K} W(t, f)_k X(t, f)_k
\]  

(5.10)

FSB can permit of applying spatial restrictions as it is expressed in frequency domain. The equation (5.10) can be pictorial representation as figure (5.2).

**Super Directive Beamformer**

Super directive beamformer is more conditional form of FSB. It will increase the flexibility of FSB by estimating the directivity of the FSB towards the source direction. This character of super directive beamformer has increased its implementation in microphone arrays. The frequency domain of super directive beamformer is derived from the time domain of the Minimum Variance Distortion less Response Beamformer (MVBR). The weights for the super directive beamformer is derived from the equation below

\[
W(f) = \Gamma^{-1}(f)s_d(s_d^H\Gamma^{-1}(f)s_d)^{-1}
\]  

(5.11)

$\Gamma(f)$ is the cross correlation of the diffuse noise between the sensors and the $s_d$ is the steering vector of the desire direction at the frequency $f$.

**5.2.2. Adaptive Beamformer**

An adaptive beamforming is a beamforming technique which carries out adaptive spatial signal processing on the microphone array. Adaptive beamforming is a data dependent type which optimises the filter coefficient according to the receiving data. Adaptive beamforming techniques are implementing in order to adaptively filter the receiving signal. Data driven beamformer will always have more importance if those beamformers are deals with the adaptive estimation of the noise signals. The adaptive methods deals with the construction of weights with respect to noise constrain in order to estimate the minimum least square solution for eliminating noise. Some of the adaptive beamforming techniques are Frost’s method, GSC method and post filtering techniques. Adaptive beamformers provide more efficient results when compare to fixed beamformer. The draw back of adaptive beamformer is sensitive to the desire signal direction and may suffer signal leakage.
**Frost’s method**

Frost’s method is the first and foremost adaptive beamforming method. Frost’s algorithm estimates the filter coefficients in order to estimate the minimising the mean square error at the same time maintaining a constant response for the desire speech signal. Frost’s method classify under LCMV beamformer. Frost’s algorithm implement LMS adaptive algorithm for estimating the weights.

**GSC Beamforming**

Generalized Side lobe Canceller beamforming is the most commonly used LCMV beamformer. The simple diagram of GSC beamformer is illustrated in figure (5.3). GSC beamformer is a collaboration of both fixed and adaptive beamformer. GSC deals with the same problem that is handled by the Frost algorithm. GSC beamformer is a two structure procedure. First structural procedure is fixing a standard beamformer with constrain on the desire signal and the second structure deals by designing the adaptive procedure which is to construct a set of filters for minimizing output power. In second procedure desire signal information is eliminated as in this case it is noise power that is to minimize by designing a blocking matrix. In this noise reduction stage a set of adaptive filters are designed on the bases of LMS algorithm and the blocking matrix will block the information about the desire signal.

\[
y_c[n] = \frac{1}{L} \sum_{k=1}^{L} x_k[n]
\]  
(5.12)

The second most importance stage blocking stage is achieved by arranging the information of each sensor in column wise in blocking matrix \(w_x\). The blocking matrix can be expressed as the product of the blocking matrix and the matrix of the current input

\[
z[n] = w_x x[n]
\]  
(5.13)

The final output of the GSC beamformer with the minimising the noise factor is given as

\[
y[n] = y_c[n] - \sum_{k=1}^{L} w_k[n] z_k[n]
\]  
(5.14)

The weighted matrix is updated according to the LMS algorithm. There are many other adaptive algorithms for implementing adaption procedure.
GSC beamformer’s adaptive filter stage is worth in the case of coherence noise if it is non-coherence noise it is unworthy. Designing of blocking matrix is another complex problem. Because it deals with various factors like estimating of source position, microphone array arrangement and many more constrains.

5.2.3. Post filter techniques

As the time increases need for eliminating the noise is also increase. The data independent beamformer are not providing the sufficient results during the present of noise so this leads to increase of post filtering technique in front of beamformer in order to increase its performance.

Post wiener filtering

Wiener filter is designed by Norbert Wiener in order to eliminate the noise that is present in a signal by comparing the estimating with the desire noiseless signal. In normal case wiener filter is consider for stationary signal. The wiener filter can be converted into
adaptive by implementing adaptive techniques. Implementing the adaptive wiener filter in front of the conventional beamformer is one of the posts filtering technique. This will leads to enhancing the efficiency of the both spatial and frequency domains. Wiener post filtering can be expressed in terms of cross spectrum densities of the noise and desire signal at the beamforming output is formulated below.

\[ H(f) = \frac{\Phi_{sd}(f)}{\Phi_{ss}(f)+\Phi_{nn}(f)} \]  

(5.15)

The computation of the post filtering technique is based on the power spectrum density of the target signal, power spectrum density of the noise signal and power spectrum density among the microphone arrays. More information about the wiener post filtering implementation for delay and sum beamformer can be provided in [15].

5.3. Brief information about Speech Enhancement and Recognition

The main prospective of speech enhancement is improving the quality of the speech signal. Speech enhancement can be performed in an efficient way with the introduction of the microphone arrays. The main objective of the speech enhancement is improving the perceptual of the degrading signal using audio signalling algorithms. Speech enhancement in the case of hands free environment is always a challenging task. Algorithms for speech enhancements are classified into three categories. Those three categories are Filtering techniques, spectral restoration and speech model base techniques. Beamforming technique is implemented in order to identify the speaker and enhancing the speech information. Both the works can be done simultaneously by the beamformer at the same time.

In this thesis it is clearly illustrate the use of beamforming for speech enhancement in microphone array. The speech data from the reverberated environment is received by the microphone sensors and pass through filter bank for spatial information is given to beamformer. Beamformer eliminates the noise and enhance the speech signal. The next chapter follows the implementation of wiener beamformer with microphone arrays for speech enhancement. In the preceding chapter discuss the implementation of head tracking with the srp phat algorithm which is also part of thesis.
6. Experimental modelling of Multichannel Beamformers

This chapter deals with the experimental construction of the two beamformers and their implementation procedure in step by step.

6.1. Implementation of Multi channel sub band wiener beamforming

Till now in the previous chapter discuss about the various modules like beamforming algorithms, filter banks, reverberation techniques and microphone arrays. Till now there no actual discuss about the experiment that will implement the speech enhancement algorithm. This chapter present contribution. One is Multichannel wiener beamforming. Another one is Multichannel sub band beamformer with SRP-PHAT. The series of experiments is conduct on the two algorithms on the bases of difference noise presence at different reverberation environments. Here source is considered as the stationary source and noise is non directional noise.

Each of the experiment is conduct on the Matlab programming language. First case of multichannel wiener beamformer is conduct to gain knowledge about the speech enhancement. This is conduct on the time domain bases for easy to understandable about the speech signal in reverberation environment, characteristics of noise and performance of the beamforming. Second experiment multichannel sub band beamforming is conduct about the spatial representation of the source, aliasing effect, steering response, filter bank working.

Microphone array speech enhancement has gain popularity and considerable important for the research in audio signal processing. It is very important in the speech recognition field as the speech is enhancing before it is recognition. Complete description about the microphone array processing is explained in chapter 4. Detail description about reverberation chapter 2 and Filter bank is shown in chapter 3.

Till now the problem with the enhancing the far field speech signal is a challenge task in the audio signal processing. New algorithms and methods are developed but the problem is not completely rectified, this increase the research work on far field speech signals. Microphone array speech processing in far field environment like conference room is the main objective of the thesis. Most of the thesis work is illustrated in this chapter.

Concentrated first on this chapter experiment develop of the microphone array wiener filtering for speech enhancement. Most of the work on the previous chapter describes about
the array processing and beamforming so there is no need for detail description about the microphone array wiener beamforming.

6.2 Experimental construction of microphone array wiener beamforming

Microphone array wiener beamformer is the pilot experiment for the multichannel frequency domain wiener beamformer. This is performed in order to get a brief idea about the microphone array and beamforming algorithm. This is simplest beamforming techniques to implement and can achieve better results. Wiener filtering is employed for the generation the coefficients for the beamformer. Wiener filter can easily distinguish between the noise and the speech signal. The simple diagram of block diagram of the microphone array beamformer is shown in figure (6.1).

\[
\begin{align*}
\mathbf{w} &= [w_1(t), w_K(t)] \\
\mathbf{x} &= [x_1(t), \ldots, x_K(t)]
\end{align*}
\]

Fig.6.1. Block diagram of microphone array wiener beamformer

Wiener beamformer is similarly like filter and sum beamformer in the fixed beamformer categories. Here it is implemented in time domain instead of frequency domain. \(x(t)\) is the speech in the collaboration of the noise signal in reverberated environment. Weights are calculated for this speech signal according to the wiener filtering technique and applied to them in order to get the noise free speech signal. For the discrete time signal, \(x(t)\) is the combination of all speech signal at the time instance form the microphone array. Weighted vector and speech signal vector \(x\) is given below as

\[
\mathbf{w} = [w_1(t) \ldots w_K(t)] \\
\mathbf{x} = [x_1(t) \ldots x_K(t)]
\]
6.2.1. Wiener solution for time domain beamformer

\[ x(t) = s(t) + n(t) \]  \hspace{1cm} (6.1)

Cross correlation of the input signal \( x(t) \) is expressed as below

\[ r_{xx}(t) = r_{ss}(t) + r_{tt}(t) + r_{nn}(t) \]  \hspace{1cm} (6.2)

From the equation (6.2) \( r_{ss}(t) \) is the cross correlation of the speech signal and \( r_{nn}(t) \) is the cross correlation of the noise signal. In general case it is considers that the noise and speech signal are uncorrelated but there may be some interference between them so \( r_{tt}(t) \) is the cross correlation of the interference signal.

Optimal weights can be calculate from the wiener filter by using the cross correlation matrix is elaborated with the equations below

\[ w_{opt} = [r_{xx}]^{-1}r_{sx} \]  \hspace{1cm} (6.3)

\([r_{xx}]^{-1}\) is the pseudo inverse matrix and \( r_{sx} \) is the desire speech signal that can be expressed as the expectation of the \( x(t) \) and the speech signal \( s(t) \)

\[ r_{sx} = E[x_s(t)s^H(t)] \]  \hspace{1cm} (6.4)

The optimal weighted coefficient is the vector is the combination of individual weights of the each microphone of the array that is expressed as

\[ w_{opt} = [w_1(t)\ldots w_k(t)\ldots w_k(t)] \]  \hspace{1cm} (6.5)

The output of the wiener beamformer with respect to the optimal weights and the speech signal at the output is given as

\[ y(t) = w_{opt}^H x(t) \]  \hspace{1cm} (6.6)

The main draw back of this algorithm is identification of the speaker location.

6.3. Experimental construction of Multichannel Sub Band Beamformer

This is the extend version of the wiener filter beamformer. Generally this sub band wiener beamformer is the best option when compare to fixed wiener beamformer as it prevents spectral leakage. The wiener filter makes an optimal tradeoff between speech signal
preservation and noise cancellation. The block diagram of Multichannel wiener sub band beamformer with steering response is shown in figure (6.2).

This is the combination of the filter sum beamforming with the wiener filter coefficients for making this as optimal beamforming. Steering is done by the srp phat algorithm for making it more realistic. Sub Band itself shows that it is implemented in frequency domain. This is most complex beamforming technique when compare to the pre wiener beamformer discussed before. The block diagram in figure (6.2) shows the different modules for the implementation of this multichannel sub band wiener beamformer and discussing each of them in detail. The different modules of multichannel sub band wiener beamformer are
• Reverberated speech signal
• Microphone arrays
• WOLA Analysis /Synthesis
• Wiener Beamformer

**Reverberated Speech Signal**

The speech signal in far field environment will generate reverberations; this is one of the main hurdles in the far field environment. Image source model is implemented in order to generate reverberations for the speech signal in the experiment. The more information about the image source model and reverberation is presented in chapter 2. According to the chapter it is clearly present that image source model can generate reverberation to the speech signal so the output from this block is the speech signal and the reverberations create by the speaker in the far field environment. This reverberated speech signal is given to the Microphone array.

**Linear Microphone Array**

The reverberated speech signal is receive by the linear microphone arrays. These microphone arrays are the sensors for tracking of the speech and localization of the source. The complexity of the noise is degraded by the microphone arrays. Most of the information about the spatial aliasing, localization of the source, alternative approaches for the microphone arrays is discussed in the chapter 4. Microphone array decrease the work of the beamformer as it can track the speech signal and some information about the non directional noise. The summation of speech and noise signal in reverberated environment is received by the beamformer from the linear microphone array sensors.

**WOLA Analysis /Synthesis Filter Bank**

Filter bank filters the linear microphone array data in several sub bands those are individual consider as the band pass filters and the processed data is received by the beamformer for further estimation. The filtering array data is applied to the fast fourier transform which convert the time domain data into frequency domain. Each of the Band pass signal has particular central frequency corresponding to the frequency bin of the fourier transform. All of this work comes under the analysis stage of the filter bank. After implementing of the beamformer to these Sub band data, this frequency domain processing results of beamformer sub bands is re synthesis to time domain by means of inverse filtering.
which is also termed as synthesis filter bank. More information about the WOLA filter bank about analysis and synthesis stage is given chapter 3.

**Wiener filtering**

The working of this module is similar to the time domain wiener beamformer but the difference is implementation of the matrix inverse lemma for identification of the inverse cross correlation function. The implementation of this is shown in the below steps

**Implementation wiener filtering with the combination of matrix inversion lemma algorithm for sub bands**

It shows the implementation of the wiener filtering algorithm to the sub band speech signals for finding out the weights for the beamformer in order to reduce the noise as follows. Let there be $K$ sub bands and each is having a corresponding frequency at $k^{th}$ among 0 to $K-1$ sub bands is $\frac{2\pi k}{K}$. Let the observed microphone signal in $k^{th}$ is $x^k_i(n)$, $i=1,2,3,...I$ is the number of the microphones and $n$ is the sample among the total number $N$ samples of each microphone. $x_r$ is the reference microphone. Then the correlation matrix of the speech and noise is

$$X^k(n) = [x^k_1(n)x^k_2(n)...x^k_I(n)]$$

(6.7)

$$\mathbf{F}^k_s = \frac{1}{N}\sum_{n=1}^{N} x^k(n)x_r^k(n)^*$$

(6.8)

The above expression is the cross correlation of the speech signals between the reference microphone and another microphone.

Now consider the another equation of autocorrelation between the reference microphone

$$\mathbf{R}_{ss}^k = \frac{1}{N}\sum_{n=1}^{N} x_r^k(n)x_r^{kH}(n)$$

(6.9)

The above are the equations with respect to the only speech signals

Now consider the signal is the noise one

$$\mathbf{R}_{nn}^k = \frac{1}{N}\sum_{n=1}^{N} x^k(n)x^k(n)^*$$

(6.10)

Here $X$ will have the information about the noise
In general case both speech and noise are present simultaneously and assume that they are uncorrelated so at a instant the microphone signal cross correlation function will be addition of the two cross correlation function of speech and noise which can be expressed in the equation (6.11). The inverse cross correlation is calculated according to the matrix inversion lemma [37] [38].

\[
(\mathbf{R}_{ss}^{-k} + \mathbf{R}_{nn}^{-k}) = Q^{kH} \Gamma^k Q^k
\]  

(6.11)

Where \(Q^k\) is the set of eigenvector represent as follows

\[
Q^k = [q_1^k, q_2^k, ..., q_I^k]
\]  

(6.12)

Where \(\Gamma^k\) is the set of Eigen values denoted as

\[
\Gamma^k = diag([\gamma_1^k, \gamma_2^k, ..., \gamma_I^k])
\]  

(6.13)

These Eigenvector, Eigen values and cross correlation functions are implemented in further equations for finding the weights.

For a \(k^{th}\) sub band there are total of \(I\) number of weights those are expressed as

\[
\mathbf{W}_n^k = [W_1^k, W_2^k, ..., W_I^k]
\]  

(6.14)

The inverse of the total cross correlation matrix which includes both speech and noise is represent by a variable \(p_n^k\) where \(n\) is the \(n\) instance of total \(N\) variables and \(k\) is the \(k^{th}\) sub band.

\[
p_0^k = Q^{kH} \Gamma^{-1} Q^k
\]  

(6.15)

For the weighted RLS algorithm there present \(\lambda\) the forgetting factor and \(\alpha\) the smoothing factor. Both the factors remain constant for all frequency.

By including these entire assumptions Weighted wiener algorithm final equations will be represent by

The input vector for the microphone array for the \(k^{th}\) sub band is

\[
\mathbf{X}_k(n) = [X_1^k(n)X_2^k(n)...X_I^k(n)]
\]  

(6.16)

\[
p_k = \lambda^{-1} p_{n-1}^k - \frac{\lambda^{-1} p_{n-1}^k X_n^k \mathbf{p}_{n-1}^k}{1 + \lambda^{-1} X_n^k \mathbf{p}_{n-1}^k}
\]  

(6.17)
\[ p^k_n = p^k - \frac{\gamma_p(1-\lambda)p^k q^k p^k H_p}{1+\gamma_p(1-\lambda)q^k H_p q^k} \]  \hspace{1cm} (6.18)

Where p is \((n \text{ mod } I) + 1\)

The weight vector equation is given as

\[ W^k_n = p^k r^k_s \]  \hspace{1cm} (6.19)

Filter equation of the weights and input speech signal given the result as output noiseless signal which can be expressed in equation as

\[ \tilde{y}_n(n) = W^T(n)X(n) \]  \hspace{1cm} (6.20)

**6.5. Brief Information on Multichannel Sub band Wiener Beamformer**

The functionality of multichannel sub band wiener beamformer start with the calculation of the reverberations for the speech signal as the beamformer is implemented in far field environments like conferences. The reverberated speech and noise is given to the sub band analysis stage WOLA filter bank for estimating the frequency domain of the signal. This is succussed by the wiener beamformer in order to eliminate the noise and dereverberate the speech signal as the beamformer is tuned to the source direction. For these frequency compounds synthesis WOLA filter bank is implemented in order to estimate the time domain for the speech signal. The output is the original speech signal which is absence of reverberations and noise.
7. Source localization by implementing SRP-PHAT

The main contribution of this chapter to present the art of source localization and some of the resent work that is performed on the audio source localization and tracking of the source.

7.1. Introduction

Previous chapters present the working of linear microphone array speech enhancement in far field environment. It is also shown the working of the different beamformers for noise reduction in noisy far field environment with experimental application of constant wiener beamformers. Most of the beamformer technique in real time is to be conditioned on source localization fundamentals which are unknown. This leads to perform considerable research in the field of acoustic source localization for estimating the source localization parameters. The information provided by the source localization parameters can be implemented to identify the active speaker in a meeting environment and can also employ for knowing the group behaviour. Source localization parameters can also provide information for automatic steering camera systems in presentation environment.

But here a complete postulate of source localization is not studied. In this chapter how to perform steering the beamformer response towards the direction of the source is investigated. The solution for the problem of steering the beamformer response is provided by identifying of Time Difference of Arrival (TDOA) and direction of arrival (DOA). Directional of arrival deals with the direction of the arrival of the speech signal towards the linear microphone arrays and the Time difference of arrival deals with the difference between the times of arrival of a speech signal between microphone pair.

The source localization technique depends on various parameters like intensity of noise, reverberations of the room, presence of single or multiple sound sources, physical characteristic of the speech signal and geometrical representation of the microphone array towards the sound source. In some cases representation of these postulates are simpler when compare to some case it will be more complicated. These parameters can even detect the acoustic events like higher order door slams that highly differ from the speech characteristics.
Let’s consider the person tracking application which is higher order to the source localization technique. Now it is concern area of research in present days. If one knows the information about the person tracking it is easy to find out the physical analysis information what is happening in the meeting. It will lead to provide better noise reduction technique towards the particular direction of the person and it is helpful to know the behaviour of the person in the meeting room. These all are the extended to the person tracking application.

Chapter gives the complete overview of the source localization and tries to achieve the steering beamformer response towards the direction of the localised source.

There are different source localization strategies which employs microphone arrays. They are Steering beamformer base locator, High resolution spectral estimation base locator, and Time difference of arrival estimator. Among three localization estimators Time difference of arrival locator is the beast and simplest one to implement. This makes the locator to implement in most of the application.

7.2. TDOA and DOA estimation approach

The signal arrival towards the linear microphone arrays can be viewed with respect to two acoustic signal characteristics. One is Direction of arrival (DOA) of the acoustic speech signal towards the linear microphone arrays. Second is Time Difference of Arrival (TDOA) which deals with the difference in time of arrival of a speech signal between microphone pair.

The solution for the DOA is estimated by the array signal processing discussed in chapter 4 that mostly developed for narrowband signals located in far field environment. At the same time the solution for the TDOA is furnished by the Time Delay Estimation (TDE) technique which arrives as the best technique for TDOA for acoustic signal processing.

In practical case DOA technique are not as popular as TDOA technique because of their complexity and geometric problems that is caused by the array signal processing. The geometric problems are caused because of implementing the particular characteristics of the speech signal to the theoretical array signal processing that is assumed. How ever with implementation of array signal processing in other fields like radar and sonar [32] influence the research area of array processing for multiple sources localization makes these techniques an interesting trend to research. With the recent progress in audio signal makes significant effort in order to combine both TDOA and DOA techniques to achieve better source localization systems.
In this thesis TDOA for source location is investigated but before going into TDOA concept a brief discussion on DOA is provided as basic information.

7.3. Direction of Arrival estimation

Direction of arrival estimation is depending on the two fundamental compounds in arrays in microphone array processing, they are far field and narrow band condition. Most popular direction of arrival technique can be seen in two categories one is beamforming method which is called as steered power response technique and the other category is sub band base method.

Steered Response Beamformer

In Steered power response category a narrow band beamformer is implemented but the narrow band beamformer may raise the problem of spatial aliasing so in this case a sub band beamformer is implemented for broad band signals as shown in chapter 2. DOA technique is implemented to each sub bands. Steering beamformer explore the all positions in the spatial room by steering the beamformer and indentify the direction where maximum energy in present. The main effort of the steering beamformer is performed by the weights of the beamformer. Weights are adjusted according to the potential positions of the source explored.

The simple example of steering beamformer is delay and sum beamformer. In delay and sum beamformer the speech signal capture between the two microphones are delay and sum at different time delays to form a function with respect to time delays. This function is the DOA.

The steering problem can be solved by finding the maximum position in spatial likelihood function \( F(x) \) that can be expressed in equation below.

\[
\hat{x} = arg\max(F(x))
\]

More information about the steering beamformer is provided in [15].

Sub Band base method

Sub band base method is also term as High Resolution Spectral Estimation techniques. This technique based on the decomposing and exploitation of the properties of the covariance matrix [15]. The main representation of this group is Root Music algorithm.
Multi source localization in reverberation environment is addressed in [31] bases on the DOA estimation with the help of root music algorithm [31] to obtain the speaker localization.

There is certain limitation to the sub band base method that is this Method is less sensitive to the sensor and source modelling errors.

**Time Difference of Arrival (TDOA)**

This is one of the oldest and most sophisticated algorithms implemented for source localization. TDOA estimates the time difference between the pair of microphone arrays. Early work of TDOA is implemented in sonar technology by LMS algorithm by estimating the time difference between the pair of signals.

This TDOA technique attain popular state base on the computation of cross correlation for the pair of signals. A detail effort is done on the cross correlation function for estimating TDOA in this thesis. A multichannel cross correlation function with cross correlation function is performed in [12] between pair of signals. A multi channel correlation matrix can be deduced by implementing the notation of spatial prediction and interpolation which is later employ to estimate the TDOA. This redundancy matrix can be implemented for more than two microphone application. A blind modelling technique on the bases on correctly model reverberation environment where the peak samples of the model for blindly identifying the TDOA same technique is implemented for multichannel microphone arrays.

A realistic review and comparison of the most successful TDOA techniques can be found in [15]

In noisy and reverberation environment with the statistical available information of the noise maximum likelihood time delay estimation can be performed with the help of estimations obtain from the SNR weighted version of the Generalized cross correlation function [16]. An enhance version is provided by implementing Phase transform to the GCC which intern written as GCC-PHAT.

The main draw back of the GCC-PHAT is its poor functioning in the source localization in multiple source environments. At the same time GCC-PHAT doesn’t perform well in highly noisy and reverberated environment.

These limitations and draw backs are overcome by implementing a more realistic source localization technique is provided which is term as SRP-PHAT (Steered Response
Power Phase Transform). This is more robust source localization technique. In this thesis experimental work is done on the SRP-PHAT for estimating the source localization in the sub band beamformer. But before progressing into SRP-PHAT algorithm let go for a brief description for GCC-PHAT algorithm whose information uses for SRP-PHAT.

### 7.4. GCC-PHAT

GCC-PHAT algorithm is a pertaining to most famous algorithm for estimating the TDOA between pair of microphones [10, 11, 12]. Source location can be calculated from multiple TDOA values. Figure (7.1) is an example for theoretical calculation of TDOA.

![Fig.7.1. TDOA between two microphones](image)

The distance from source to microphone m is $r_m$ then the travelling time for speech to reach microphone m from source is

$$\tau_m = \frac{r_m}{c}$$  \hspace{1cm} (7.1)

Where $c$ is the speed of sound similarly travelling time for microphone n is

$$\tau_n = \frac{r_n}{c}$$  \hspace{1cm} (7.2)

Then the time difference of arrival (TDOA) between microphone pair (m, n) is

$$\tau_{mn} = \frac{r_m - r_n}{c}$$  \hspace{1cm} (7.3)

Generalised Cross correlation using phase transform for two microphone m and n is as shown in equation(7.4)
Most of the limitations and drawbacks of the GCC-PHAT are overcome by the SRP-PHAT. The problem with the maximum likelihood map for every possible exploration position leads to develop research for fixing flexible manner of likelihood functions. It is not only necessary to find out the constrains for the energy receive by the beamformer, it is also necessary to find out the time delay between the microphone pair of the maximum cross correlation value. The main advantage of the srp phat technique is it is speaker independent.

Wiener beamformer output can be described in the equation (7.5)

\[ Y(\omega, q) = \sum_{n=1}^{M} G_n(\omega)X_n(\omega)e^{-j\omega\delta_n} \]  

(7.5)

Where \( G_n(\omega) \) is the Fourier transform of the wiener filter of microphone \( n \) and \( X_n(\omega) \) is the Fourier transform of the speech signal \( x_n(t) \). Wiener technique deals with the eliminating of the noise so wiener filter remove reverberations and noises.

### 7.5.1. Steered response Power

A conventional SRP is the response power achieved by steering the beamforming coefficients at some particular location. It can be expressed in frequency domain as

\[ P(q) = \int_{-\infty}^{\infty} Y(\omega, q)Y^*(\omega, q) \, d\omega \]  

(7.6)

Inserting equation (7.4) here \( l \) and \( k \) are the two microphones pair in equation (7.5) then the result will be

\[ P(q) = \int_{-\infty}^{\infty} \left( \sum_{l=1}^{M} G_l(\omega)X_l(\omega)e^{-j\omega\delta_l} \right) \left( \sum_{k=1}^{M} G_k^*(\omega)X_k^*(\omega)e^{-j\omega\delta_k} \right) \, d\omega \]  

(7.7)

To place in proper the above equation

\[ P(q) = \int_{-\infty}^{\infty} \left( \sum_{l=1}^{M} \sum_{k=1}^{M} (G_l(\omega)G_k^*(\omega))(X_l(\omega)X_k^*(\omega))e^{-j\omega\delta_k - \delta_l} \right) \, d\omega \]  

(7.8)

The steering delays \( \delta_k \) and \( \delta_l \) will be estimated using TDOA of each microphone pair, which can be written as

\[ \tau_{kl} = \delta_k - \delta_l \]  

(7.9)

Now the power of beamformer in frequency domain is expressed as
\[ P(q) = \int_{-\infty}^{\infty} \left( \sum_{l=1}^{M} \sum_{k=1}^{M} \left( G_l(\omega) G_k^*(\omega) \right) X_l(\omega) X_k^*(\omega) \right) e^{i\omega \tau_{kl}} \, d\omega \] \quad (7.10)

Weighting function for the filter is defined as

\[ \Psi_{lk}(\omega) = G_l(\omega) G_k^*(\omega) \] \quad (7.11)

Rearranging the integration and summations of the equation (7.9) for a finite length

\[ P(q) = \sum_{l=1}^{M} \sum_{k=1}^{M} \int_{-\infty}^{\infty} \Psi_{lk}(\omega) X_l(\omega) X_k^*(\omega) e^{i\omega \tau_{kl}} \, d\omega \] \quad (7.12)

Rearranging the above equation in order to find out the generalized SRP-PHAT algorithm for minimum computations is rewritten as

\[ P(q) = \sum_{l=1}^{M} \sum_{k=l+1}^{M} \int_{-\infty}^{\infty} \Psi_{lk}(\omega) X_l(\omega) X_k^*(\omega) e^{i\omega \tau_{kl}} \, d\omega \] \quad (7.13)

PHAT weighting function can be expressed as

\[ \Psi_{lk}(\omega) = \frac{1}{|X_l(\omega) X_k^*(\omega)|} \] \quad (7.14)

\[ \Psi_{lk}(\omega) \] is the desire PHAT filter for the speech signal between microphone pair(l, k) similarly the PHAT weight for the adaptive filter is expressed in the equation (7.15)

\[ G_l(\omega) G_k^*(\omega) = \frac{1}{|X_l(\omega) X_k^*(\omega)|} \] \quad (7.15)

To place in proper the equation (7.14) in equation (7.12)

\[ P(q) = \sum_{l=1}^{M} \sum_{k=l+1}^{M} \int_{-\infty}^{\infty} \frac{1}{|X_l(\omega) X_k^*(\omega)|} X_l(\omega) X_k^*(\omega) \, d\omega \] \quad (7.16)

\[ \tau_{kl} \] is the time difference of arrival between the microphone pair(k, l). The far field assumption of the microphone arrays and a plane of sound wave are expressed below

### 7.6. TDOA Estimation using SRP PHAT

To estimate the speaker location, TDOA \( \tau_s \) should be first estimated. The GCC-PHAT algorithm used in \([34]\) is defined as

\[ \tau_s = \arg\max_{\tau_{kl}} P_{rl}(\tau_{kl}) = \arg\max \left( \frac{1}{2\pi \int_{-\infty}^{\infty} \frac{1}{|X_l(\omega) X_k^*(\omega)|} X_l(\omega) X_k^*(\omega) e^{i\omega \tau_{kl}} \, d\omega} \right) \] \quad (7.17)
By inserting equation 7.16 in equation 7.17 we get equation (7.18)

$$\tau_s = \arg \max \{ \tau \} = \arg \max \left( \sum_{i=1}^{M} \sum_{k=1}^{M} \int_{-\infty}^{\infty} \frac{1}{X_i(\omega)X_k^*(\omega)} X_i(\omega)X_k^*(\omega) e^{j\omega \tau(k-l)} \ d\omega \right)$$

(7.18)

The TDOA $\tau_s$ will be the value which will give the maximum output power of SRP-PHAT. This SRP-PHAT algorithm has only one parameter output $\alpha$, which indicates the DOA of sound source as expressed below

$$\alpha = \sin^{-1} \left( \frac{v^*\tau_s}{d + fs} \right)$$

(7.19)
8. Simulation results and analysis

The evaluation of wiener beamformer and sub band beamformer with linear microphone arrays in closed meeting room environment is performed with the help of Matlab simulator. The simulation process implements a speech signal and wind noise signal in the two beamformers. The speech signal is sampled at 8000 kHz and it is a male voice. The speech and noise signals both are simulated with image source model in order to generate the reverberations for each signal. Linear array consist of four microphones and each microphone receives the summation of reverberated speech and noise signals. The test is performed for both the algorithms at different SNR input by varying the noise power. The simulation is conducted at two different microphone array positions in the case of time domain wiener beamformer. In Multichannel sub band wiener beamformer performance is performed in various cases by changing the forgetting factor and distance between the mics for two noise signal one is Wind noise and another one is Additive White noise.

Frequency domain analysis is performed at certain conditions (conditions like input snr of 13.68 db and number of sub bands are 128,256) as if the beamformer extend from these conditions leads to degradation of SNR improvement and PSEQ values.

In time domain the SNR input is varied according to the equation (8.3) and (8.4) as follows but this type approach is not performed in frequency as there has one limitation that is weights are sensitive with respect to variation in the input SNR. That means weights are stable if the input snr is 13.68dB.

SNR improvement is calculated as

\[
SNR_{db} = 10 \log_{10} \left( \frac{\sigma^2_{speech}}{\sigma^2_{noise}} \right) \quad (8.1)
\]

\[
SNR_{IMPROV} = 10 \log_{10} \left( \frac{\sigma^2_{speech_{out}}}{\sigma^2_{noise_{out}}} \right) - 10 \log_{10} \left( \frac{\sigma^2_{speech_{input}}}{\sigma^2_{noise_{input}}} \right) \quad (8.2)
\]

The input signal at the microphone mic_i can be expressed in equation (8.3)

\[
mic_i = s_i + n_i \quad (8.3)
\]

Where i=1:4 (i representing mic positions of four mic)

Where \( \alpha \) can be calculated from the formula expressed in equation (8.4)
\[ \alpha = \frac{1}{\sqrt{10^{\frac{SNR_{input} - SNR}{10}}} } \]  

(8.4)

\(SNR_{input}\) is the desire to noise ratio and it is varies as 0dB, 5dB, 10dB, 15dB, 20dB. The signal to noise ratio is calculated according to the equation.

The performance of the system is calculated with respect to the SNR improvement and PESQ. A small description about PESQ is provided in the below paragraph

PESQ (Perceptual Evaluation of Speech Quality) is another objective measurement tool that predicts the results of subjective listening tests on telephony systems. PESQ uses a sensory model to compare the original, unprocessed signal with the degraded signal from the network or network element. The resulting quality score is analogous to the subjective MOS measured using panel tests according to ITU-T P.800. The PESQ scores are calibrated using a large database of subjective tests. The most eminent result of PESQ is the MOS. It directly expresses the voice quality. The PESQ MOS as defined by the ITU recommendation P.862 ranges from 1.0 (worst) up to 4.5 (best) [35].

PESQ takes into account coding distortions, errors, packet loss, variable delay, and filtering in analogue network components. The user interfaces have been designed to provide a simple access to this powerful algorithm, either directly from the analogue connection or from speech files recorded elsewhere.

This powerful test tool can be deployed in many different areas of a business, on any speech carrier technology:

- In the research laboratory; providing rapid feedback on promising areas of signal processing development, validation of design implementation, ranking alternative design solutions, providing a higher degree of confidence before submission to subjective testing
- In network equipment evaluation; comparing different vendor offerings and determining their impact on network performance
- In sales and marketing; demonstrating the excellence of a new product, assuring the customer of the system performance
The results are arranged in systematic procedure. The result session starts with the performance analysis of Room impulse response and then follows the polar coordinated of the microphone array at different frequency. After this response of WOLA filter bank and later follows the simulation of beamformers and srp phat.

8.1. Simulation of Room impulse Response

Room impulse response generated the reverberations for the speech and noise signal in order to meet the limitations of closed room environment. RIR is the filter function that is generated according to the limitations of the image source model that is described in second chapter. Thiran filter is implemented in order to find out the fraction delays of each and every reverberation from source to mic.

Here Room impulse response is design as a filter in the Matlab simulation tool bar as user defined function and it is convoluted with the speech and noise signal in order to generate the reverberations for those signals. More theoretical description about the image source model to generate reverberations in closed room is available in chapter 2

\[ x(t) = s(t) * h(t) \]  

(8.6)

In equation (8.6) s(t) is the original speech signal and h(t) is the room impulse function and * is the convolution operator. Noise signal is also shown in similar passion by varying the variables in the equation.

List of variables in the image source model are room coordinates, source coordinates, reflection coefficient, sampling frequency and mic coordinate. Room impulse response function is generated by Image source model with the following variables are taken into consideration.

Here a room of [6 4 2.8] coordinates is taken with respect to x axis, y axis and z axis. Source coordinates are taken as [5 2 1] and mic coordinates are [4 2 1]. Sampling frequency is taken as 32 kHz; n is the number of reflections that is required and reflection coefficient coordinate are various according to the required criteria.

The figure (8.1) is the room impulse response for the reflection coefficient 0.95 which means 95 % of the energy of the wave is reflected back when hits surrounding and only 5 % is absorbed by the surroundings. Y axis is the amplitude of the impulses and x axis the time taken for the impulse response.
Fig. 8.1. Energy decay for reflection coefficient $r=0.95$

Fig. 8.2. Energy decay curve for reflection coefficient $r=0$
Now considering the another case if total reverberations are absorbed by the surroundings that means reflection coefficient is zero and only the direct sound wave between the source and the mic present it is illustrated in figure(8.2).

**Reverberation Time (\(RT_{60}\))**

Reverberation time is the time required for the reflection of the sound wave to decay by 60dB. Reverberation is the single value however it is measured for the wide range of frequencies. The figure (8.3) shows the calculation of reverberation time in the energy decay curve. In the energy decay curve reverberation time is the time taken for the curve to reach the 0.001 magnitude value. It is 0.001 because maximum amplitude value is 1 then 0.001 is the 60dB of the maximum amplitude. In the figure (8.3) the green indicates the threshold value \(RT_{60}\) that is 0.001 and the blue is the energy decay curve. The \(RT_{60}\) for the below graph is 0.2691 sec, X=0.2691 is reverberation time and Y=0.001 is the threshold value to identify reverberation time.

**Room impulse response**

![Room impulse response graph](image)

Fig.8.3.plot indicates \(RT_{60}\) for the energy curve for r=0.95
Reverberation time is directly proportional to absorption coefficient and dimensions of the rooms. The plot in figure (8.4) is the reflection coefficient and reverberation time. X axis is the reflection coefficient ‘r’ and y axis is the rt60. As the reflection coefficient increases reverberation time also increases. This is clearly shown in graph for different reflection coefficients reverberation is calculated and plotted in the graph.

Fig.8.4. Plot between RT60 and reflection coefficient ‘r’

**Thiran All Pass Filter**

A novel fraction delay all pass filter is implemented in Image Source Method for generation of reverberations to estimate the delay of each reverberate signal from source to mic. A simple impulse response for a thiran filter for delay d=7.5 and filter order N=8 is shown in figure (8.5). X axis is the time taken for impulse and y axis is the amplitude of the impulse. According to the specification the impulse response shows a sudden increase in the peak at the time instant of 7.5 in order to generate a delay at that time.

**Performance evaluation of Thiran filter**

In figure (8.6) the performance of the thiran filter is evaluated with group delay response. According to the condition N-1<D<N, D should be always in between the boundary condition. As D=7.5 then N should be 8. From the figure it shown the performance of thiran
for filter order N=1 to 8. For filter order 8 it is providing the better response at the group delay is constant for 7.5 there has a sudden peak at the final value of normalised frequency.

Fig.8.5. Impulse response for thiran all pass filter for delay 7.5

Fig.8.6. Group delay response for thiran filter at delay 7.5 for N=1 to 8
because of non linearity of IIR filter. For remaining N values the group delay response it not satisfactory as it is outside the boundary condition.

Fig. 8.7. Graph between reverberated speech signal and original speech signal

The figure (8.7) is the graph the original speech and the reverberated signal of the speech signal in the closed room condition here room coordinates and all variables are similar to the figure(8.1) except reflection coefficient is 0.95 is varied. X axis is the time period of the sound wave and y axis is the amplitude of the sound wave. Reverberated sound wave will have more impulses and increase in amplitude of the impulses than the ordinary sound wave.
8.2. Simulation of Wiener beamformer in time domain

For the Wiener beamformer in time domain evaluation, the optimum weights are calculated by using the formula in the equation (8.3).

\[ W_{opt} = \arg\min_{r} E\{[y[n] - S_r[n]]^2\} \quad r \in [1,2,\ldots,l] \quad (8.3) \]

Where \( S_r[n] \) is considered as the reference speech signal. Practically as the reference signal is not accessible the optimal weights are found by calculating the mean square difference between the beamformer output \( y[n] \) and the reference signal \( S_r[n] \) and is generalized as follows due to the linearity property of expectation operator.

\[ W_{opt} = [R_{ss} + R_{nm}]^{-1}r_s \quad (8.4) \]

The cross-correlation vector \( r_s \) is defined as

\[ r_s = [r_1r_2 \ldots \ldots r_l] \quad (8.5) \]

In equation (8.4) \( R_{ss} \) and \( R_{nm} \) are the source and noise auto correlation matrices respectively.

Time domain beamformer is implemented in with the room coordinates [6 4 2.8]. Source coordinate position at [5 2 1], noise coordinates at [3 1 0.5], frequency at 8000 kHz and reflection coefficient is at 0.3. Wind noise is taken as the noise in the closed room environment. Two microphone positions are taken. Initial conditions mic are placed at a distance of 0.02 then the coordinates of all four mics are mic1= [4 2 1], mic2= [4 2.02 1], mic3= [4 2.04 1] and mic4= [4 2.06 1]. Second conditions mics are placed at 0.01 then the all four mics coordinates will be mic1= [4 2 1], mic2= [4 2.01 1], mic3= [4 2.02 1] and mic4= [4 2.03 1]. For each of the position input SNR is varied for values 0dB, 5dB, 10dB, 15dB and 20dB corresponding output SNR, SNR improvement and PESQ values are calculated. SNR is calculated from the formula (8.1). Corresponding \( \alpha \) value for multiplying to noise to generate a variable SNR input is calculated form the equation (8.4). Below table (8.1) is corresponding of various SNR input for two different linear microphone positions, their corresponding SNR and PESQ for the output speech signals after implementing wiener beamformer are placed in the table columns. The system provide better results when the SNR input is 20dB but there has a slight decrease in the SNR improvement. The average PESQ of the total system is two microphone conditions is 2.74 and the best PESQ value is 3.384, this shows this is the better algorithm for speech enhancement and noise cancelation. The tabular form is further.
synthesis into plots among input SNR, Output SNR, SNR improvement and PESQ output. These are shown in figure (8.8), figure (8.9) and figure (8.10)

### Table 8.1: Evaluation of wiener beamformer at two different position of linear microphone array

<table>
<thead>
<tr>
<th>Two different position of mics</th>
<th>SNR Input (dB)</th>
<th>SNR Output (dB)</th>
<th>SNR Improvement (dB)</th>
<th>PESQ Input</th>
<th>PESQ Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>mic1=[4 2 1]</td>
<td>0</td>
<td>18.2501</td>
<td>18.2501</td>
<td>1.302</td>
<td>1.911</td>
</tr>
<tr>
<td>mic2=[4 2.02 1]</td>
<td>5</td>
<td>21.4342</td>
<td>16.4342</td>
<td>1.520</td>
<td>2.194</td>
</tr>
<tr>
<td>mic3=[4 2.04 1]</td>
<td>10</td>
<td>24.7517</td>
<td>15.1264</td>
<td>1.842</td>
<td>2.571</td>
</tr>
<tr>
<td>mic4=[4 2.06 1]</td>
<td>15</td>
<td>28.4110</td>
<td>13.7865</td>
<td>2.250</td>
<td>2.960</td>
</tr>
<tr>
<td>[position 1]</td>
<td>20</td>
<td>32.2699</td>
<td>12.6446</td>
<td>2.682</td>
<td>3.336</td>
</tr>
<tr>
<td>mic1=[4 2 1]</td>
<td>0</td>
<td>19.4883</td>
<td>19.8626</td>
<td>1.302</td>
<td>2.843</td>
</tr>
<tr>
<td>mic2=[4 2.01 1]</td>
<td>5</td>
<td>22.7025</td>
<td>18.0771</td>
<td>1.520</td>
<td>2.914</td>
</tr>
<tr>
<td>mic3=[4 2.02 1]</td>
<td>10</td>
<td>23.3066</td>
<td>15.6813</td>
<td>1.842</td>
<td>2.906</td>
</tr>
<tr>
<td>mic4=[4 2.03 1]</td>
<td>15</td>
<td>27.8462</td>
<td>13.2208</td>
<td>2.250</td>
<td>2.944</td>
</tr>
<tr>
<td>[position 2]</td>
<td>20</td>
<td>30.9128</td>
<td>11.2874</td>
<td>2.682</td>
<td>3.384</td>
</tr>
</tbody>
</table>

**Fig. 8.8.** Plot between Input SNR and SNR output for two different positions of linear microphone array
Fig. 8.9. Plot between Input SNR and SNR improvement for two different positions of liner microphone array.

Fig. 8.10. Plot between Input SNR and PESQ output for two different positions of liner microphone array.
In the figure (8.11) shows the power spectrum densities of various signals in time domain. Wiener beamformer x axis is the normalized frequency and y axis is the power (dB). It is performed at when input SNR is -7.443 (It means nose power is more than the signal power) and SNR output is 17.1540 and SNR improvement is 24.5983. In the graph plot power spectrum of input speech and output is likely to be same when compare to the power spectrum of the speech+noise.

8.3. Simulation of Multichannel Sub Band wiener Beamformer

The simulation is conducted in a closed meeting room in reverberated environment where Multichannel sub band beamformer is implemented in order to estimate the direction of arrival of speech signal and removing of the noise by implementing linear microphone array. This is the extended version of the wiener beamformer. It is implemented in the frequency domain. The main work of these beamformer is to steering the beamformer response. This is conducted in boundary conditions. As it is implemented in frequency domain it has made an immense work to stop the spectral leakage but if the beamformer extend the boundary condition then the PESQ and SNR improvement decreases. Test of multichannel sub band wiener beamformer is performed on the bases of Additive white noises. In this beamformer
forgetting factor $\lambda$ is crucial. The performance of the beamformer is mostly depending on this as the inverse autocorrelation is calculated on these bases.

Before the implementation of this beamformer a WOLA filter bank is designed and tested. The figure (8.12) is the magnitude response of the WOLA filter bank. Where x axis is the normalized frequency and y axis is the amplitude. It is plotter for 128 sub bands with the help of tfestimate tool in Matlab.

In the table (8.2) are the results for the implementation of Multi channel sub band wiener beamformer. Here room coordinates are taken as [8 8 8], source coordinates are [5.8 2 1.5], noise coordinates as [2 1 0.5], mic1 is located at [4 2 1], all other mics are linearly equidistance by 0.02 and oversampling ratio is 64. The table (8.2) has the parameters that can evaluate the performance of the conventional beamformer with two different number of sub bands as the reference points hold. The beast SNR improvement in the case of frequency domain is 16.3777 and its corresponding PESQ is 3.432. Even though time domain beamformer is implemented with wind noise with various SNR input relations but frequency domain beamformer is implemented with Additive white noise because frequency domain beamformer is effect with spectral leakage when wind noise is implemented. The Multichannel sub band beamformer is facing problem when the boundary conditions exceed the provided values in the tabular column which finally leads to decrease in the SNR.
improvement and PESQ values. Corresponding plots are shown in the figure (8.13) and (8.14).

Table 8.2. Evaluation of Multi-channel sub band Wiener beamformer at two different sub bands of WOLA filter Bank

<table>
<thead>
<tr>
<th>Number of Sub bands 128 [Position 1]</th>
<th>( \lambda ) forgetting factor (lambda)</th>
<th>SNR improvement</th>
<th>PESQ output</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>12.1904</td>
<td>2.984</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>14.2844</td>
<td>1.963</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>14.2677</td>
<td>1.134</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>16.3773</td>
<td>3.432</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>11.9247</td>
<td>1.106</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of sub bands 256 [position 2]</th>
<th>( \lambda ) forgetting factor (lambda)</th>
<th>SNR improvement</th>
<th>PESQ output</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>6.2159</td>
<td>1.369</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>7.1684</td>
<td>1.729</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>11.8998</td>
<td>1.818</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>12.8465</td>
<td>1.405</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>14.1897</td>
<td>1.946</td>
<td></td>
</tr>
</tbody>
</table>

Figure (8.13) is the bar graph plotted between the lambda(x axis) and SNR improvement(y axis). Here lambda is a variable at two different sub band positions. When lambda is 0.7 at 128 bands gives maximum SNR improvement of 16.3773.
Fig. 8.14. Plot between Lambda (forgetting factor) and PESQ score at two different sub band positions

Figure (8.14) is the bar graph plotted between the lambda (x axis) at two different sub bands and PSEQ value (y axis). From the graph when lambda is 0.7 at 128 sub band provide best PESQ score of 3.432.

Fig. 8.15. Plot between Lambda (forgetting factor), SNR improvement for adaptive white noise and wind noise

SNR in dB

0.45 0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95

SNR imp for rand noise
SNR imp for wind noise
Figure (8.15) shows the relation between the lambda and the SNR improvement. The SNR improvement is calculated in two cases one is in wind noise and another in additive white noise. The plot shows the performance of the frequency domain beamformer is better in the case of additive white noise when compare to wind noise. Wind noise will get better speech enhancement with the forgetting factor is more than 0.9. For this reason here most of the stress is done for the beamformer that is implemented in additive white noise environment.

In the figure (8.16) shows the power spectrum densities of various signals in Frequency domain wiener beamformer x axis is the normalized frequency and y axis is the power(dB). In the graph plotted power spectrum of the output speech follows mostly power spectrum of the input speech. Between the normalized frequencies 0.4 to 0.6 the output speech signal is having power (dB) maximum amplitude than the original speech signal.
The figure (8.17) is the graph the original speech and the reverberated signal of the speech signal to mic 2 in the closed room condition and the output speech wave. X axis is the time period of the sound wave and y axis is the amplitude of the sound wave. The output speech wave clearly showed the dereverberation of the speech wave to the mic2.

**Performance of the Time domain and frequency domain beamformers**

The performance of each beamformer is illustrated in the form of tabular and plots in order to have better understanding of the each beamformer. In general case both of the beamformers need to be compared in order to better understand the performance. That attempt is not performed as the both the beamformers are showing better performance at different cases not the same case. Time domain beamformer is performing better in the case of wind noise where as frequency domain beamformer is perform better in the case of additive white noise. The figure (8.15) shows that time domain beamformer has better performance when compare to frequency domain beamformer in the case of wind noise. But the frequency domain has better performance in the case of additive white noise and the important one is source localization in the frequency domain beamformer. Initially it is presented that desire signal constrains in wiener filter are calculated from the SRP PHAT but this cases huge amounts of noise in the beamformer so SRP-PHAT is limited to source localization in the frequency domain. The main advantage of frequency domain beamformer when compare to time domain beamformer
is the speaker localization that is frequency domain can not be perform in time domain
beamformer.

8.4. Estimation of angle of arrival in close room environment wit SRP-PHAT

The utilisation of Multichannel sub band wiener beamformer is enhanced more with the
implementation of SRP-PHAT algorithm. The main problem in closed room environment is
estimation of the speaker. Steering the beamformer in order to estimate the maximum
response power for identification of the active speaker is performed by the SR-PHAT on the
bases of calculation of TDOA and DOA. More information of the SRP-PHAT algorithm is
provided in chapter 7.

SRP-phat test is performed in Matlab simulation environment with linear microphone array
and WOLA filter bank. Here a room of [6 4 2.8], source coordinates is vary by the phase
array system tool box. Linear microphone array of four mics are at coordinates [4 2 1, 4 2.02
1, 4 2.04 1, 4 2.06 1]. Speed of sound is 343 m/sec, noise coordinates of [2 1 0.5] are hold.
The figure (8.18) is the estimation of the DOA of the signal by calculating the final TDOA of
the linear array. In the figure x axis is the variation of $\tau$ and y axis is the amplitude of the
$\tau$.The final $\tau$ is calculated by taking the mean of the all three TDOA between the mic1 and
rest of the mics. For this case initial DOA of the speech is $30^0$ and the estimated DOA is
$29.0975^0$. 

![Fig.8.18 Estimating of the $\tau$ form the TDOA observe form the combinations of the mics](image-url)
Table 8.3: Evaluation of SRP-PHAT on the basis of actual DOA and practical DOA

<table>
<thead>
<tr>
<th>Actual DOA (degrees)</th>
<th>Practical DOA (degrees)</th>
<th>Absolute error (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-90</td>
<td>-72.3243</td>
<td>17.6757</td>
</tr>
<tr>
<td>-80</td>
<td>-69.7828</td>
<td>10.2175</td>
</tr>
<tr>
<td>-70</td>
<td>-63.6861</td>
<td>6.3139</td>
</tr>
<tr>
<td>-60</td>
<td>-55.8498</td>
<td>4.1505</td>
</tr>
<tr>
<td>-50</td>
<td>-47.2521</td>
<td>2.7479</td>
</tr>
<tr>
<td>-40</td>
<td>-38.2522</td>
<td>1.7478</td>
</tr>
<tr>
<td>-30</td>
<td>-28.9926</td>
<td>1.0074</td>
</tr>
<tr>
<td>-20</td>
<td>-19.5036</td>
<td>0.4964</td>
</tr>
<tr>
<td>-10</td>
<td>-9.5456</td>
<td>0.4544</td>
</tr>
<tr>
<td>0</td>
<td>-0.0115</td>
<td>0.0115</td>
</tr>
<tr>
<td>10</td>
<td>9.5456</td>
<td>0.4544</td>
</tr>
<tr>
<td>20</td>
<td>19.5523</td>
<td>0.4477</td>
</tr>
<tr>
<td>30</td>
<td>29.0975</td>
<td>0.9025</td>
</tr>
<tr>
<td>40</td>
<td>38.3690</td>
<td>1.6310</td>
</tr>
<tr>
<td>50</td>
<td>47.3873</td>
<td>2.6127</td>
</tr>
<tr>
<td>60</td>
<td>55.9722</td>
<td>4.0278</td>
</tr>
<tr>
<td>70</td>
<td>63.7898</td>
<td>6.2102</td>
</tr>
<tr>
<td>80</td>
<td>69.8490</td>
<td>10.1510</td>
</tr>
<tr>
<td>90</td>
<td>72.4000</td>
<td>17.6000</td>
</tr>
</tbody>
</table>

Table (8.3) shows a set of original direction of arrival generated by the phase array tool box and the practical values calculated by the SRP-PHAT algorithm from the estimated TDOA between the linear microphone arrays. Reference mic is mic1, actual DOA is estimated from $-90^\circ$ to $90^\circ$. The SRP-PHAT calculates particle values for the actual DOA. The absolute error between the actual DOA and practical DOA is high when the source DOA is below $-70^\circ$ and above $70^\circ$. The average of the absolute error is $4.3^\circ$ which is acceptable for source localization.

The tabular form (8.3) is further illustrated into figure (8.19) and figure (8.20). Figure (8.19) is the graph between the ideal DOA positions which is indicated with the read line and the blue bots are the practical DOA calculations. X axis is the input source position in degree and
Fig. 8.19. Graph between the Ideal DOA positions and practical DOA positions

Fig. 8.20. Graph between the Input position and absolute error
y axis is the estimated position with the SRP-PHAT in degree. Comparing the both the line and dotes it is shown that for the values of $-50^0$ to $50^0$ input source positions exist small error of from 2.7 to 0.0115 for the calculated position from the SRP-PHAT.

Figure (8.20) is the graph between the root mean square error and input positions. Here it is shown that the graph is curve which indicated that at initial DOA $-90^0$ root mean square is more and it goes on decreasing until $0^0$ and again root mean square error value goes on increasing form DOA$0^0$ to $90^0$.

These two figures (8.19) and (8.20) are graphed in order to obtain for better understanding of the results of the SRP-PHAT.
9. Conclusion and Future work

The final chapter is aimed to summarized the contribution and major results of this thesis in addition to highlight some directions for the future work.

The study of Linear array microphone with beamformer techniques for noise cancellation and speech enhancement has been sought in this thesis. More concretely two different but related research lines have been followed: Speech enhancement with microphone arrays and speaker localization with SRP-PHAT algorithm.

Microphone array processing is introduce as a possible solution to the problems that present in the distant talking speech applications which are mainly of noise and reverberations. By the means of beamformer one can train the beamformer spatial filtering to the particular direction of the source in order to get the enhance version of the speech by the array of sensors.

The thesis consist of implementation and analysis of the microphone array wiener beamformer, multichannel sub band wiener beamformer and source localization with the implementation of SR-PHAT algorithm in multichannel sub band wiener beamformer. The simulation work is carried in Matlab programming. A linear microphone array of four microphones is implemented in two beamformers with wind noise in wiener beamformer and Additive white noise in multichannel sub band wiener beamformer. The performance of the each of the beamformer is analysis with the simulation result presented in the chapter 8.

Distance between the microphone and the source, noise coordinates plays a crucial role in the case of noise attenuation and speech enhancement. Microphone array wiener beamformer provide a better result with the SNR improvement of 19dB and PESQ score of 3.386 in the case of wind noise considering boundary conditions same as closed meeting room environment. In the case of Multichannel sub band beamformer is the room environment but with additive white noise with the help of WOLA filter bank provides satisfactory results of SNR improvement of 16.377dB and PESQ score of 3.4. However additional advantage of the Multichannel sub band wiener beamformer is implementation of SRP-PHAT algorithm in it in order to provide source localization. SRP-PHAT is providing better DOA. The minimum error occurs when the source is in front of the microphone array and maximum error occurs when the source is same axis of the array. This thesis is finally
concluded with the satisfactory results of the multichannel sub band wiener beamformer for speech enhancement and source localization.

**Future work**

The solution provided by the multichannel sub band wiener beamformer is appropriate one but not the most sophisticated one. There has a need to improve the performance of the beamformer. In the future if the desire signal characteristics in the wiener filter are observed form the SRP-PHAT then it can improve the performance of the beamformer. There is a need of implementing the thesis in real time environment instead of Matlab offline.
10. References


