

*A Major Project report*

*on*

**“Sound Sound Localization for Spoken Conversation  
Humanoid”**

*Submitted By*

**Jennifer Yennam (18BEC018)**

**Kakulavarapu Lakshmi Gayathri (18BEC019)**

*Under the guidance of*

**Dr. K.T. Deepak**

**HoD/Asst.Professor, Department of ECE**



**INDIAN INSTITUTE OF INFORMATION TECHNOLOGY  
DHARWAD**

**Academic Year : 2021-2022**

# CERTIFICATE

This to certify that the report of the major project submitted is the outcome of the project work entitled '**Sound Source Localisation for Spoken Conversation Humanoid**' carried out by

Jennifer Yennam (18BEC018)  
Kakulavarapu Lakshmi Gayathri (18BEC019)

under the guidance and supervision of **Dr. K.T. Deepak** (Assistant Professor, Electronics and Communication Engineering) during the final Semester and that this work has not been submitted elsewhere for a degree.

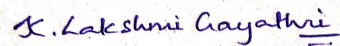
Dr. K.T. Deepak  
Assistant Professor  
HoD, ECE Department  
IIIT Dharwad  
May, 2022

# DECLARATION

We declare that this written submission represents our ideas in our own words and where other ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We moreover understand that any consequences of falsifying any information or misinterpretation of data and figures, hence we tried to come up and present with best possible results.



Jennifer Yennam  
18BEC018



Kakulavarapu Lakshmi Gayathri  
18BEC019

# APPROVAL SHEET

The project entitled “Sound Source Localisation for Spoken Conversation  
Humanoid” by

Jennifer Yennam      18BEC018

Kakulavarapu Lakshmi Gayathri      18BEC019

is approved for the degree in Bachelor of technology in Electronics and  
Communication Engineering.

## **Supervisor**

Dr. K. T. Deepak  
Assistant Professor, HoD  
Department of ECE  
IIIT-Dharwad

## **Head of Department**

Dr. K. T. Deepak  
Assistant Professor, HoD  
Department of ECE  
IIIT-Dharwad

## **Examiners**

Dr. K.T. Deepak,  
Assistant Professor, HoD  
Department of ECE  
IIIT-Dharwad

Dr. Prakash Pawar  
Assistant Professor  
Department of ECE  
IIIT-Dharwad

# Acknowledgment

It gives us huge delight and fulfillment in introducing this Major Project “**Sound Source Localization Module for SPOKEN CONVERSATION HUMANOID**”

We are extremely grateful and would like to express our special thanks to Dr. K. T. Deepak, HOD/Asst.Professor, Department of Electronics and Communication (ECE) - IIIT Dharwad for his continuous guidance and constant support throughout the course of this major project.

We would also like to thank all the faculty, research scholars and administration of the institute who ensured the needs fulfilled for the completion of this project.

Last but not the least, we would like to thank every one of those, who have straightforwardly or by implication made a difference for the successful completion of the work.

Date : May 2022

Place: Indian Institute of Information Technology, Dharwad

Jennifer Yennam (18BEC018)

Kakulavarapu Lakshmi Gayathri (18BEC019)

# Preamble

The main aim of the project proposal is to develop different modules involved in a humanoid that can navigate autonomously in a specific environment. Furthermore, the humanoid should be able to converse with patients/elderly in their native language and carry out limited electro-mechanical activities. The main idea is to attain an off-the-shelf (ready-made) humanoid and develop independent modules such as Automatic Speech Recognition (ASR) interfaced to Chabot, Text to Speech synthesis (TTS), Sound source localization (SSL), computer vision, artificial intelligence (AI), and necessary computer networking, and security applications.

Though the proposed application targets health care and the aging population, it can easily extend the applications to other areas. The proposed work can be broadly categorized as hardware and software stack. The idea is to maintain common hardware and make the necessary changes to the software stack to meet different objectives. This type of architecture enables the building of different applications with easy customizations rapidly.

For Spoken Conversation humanoids who can guide the elderly, assist and look after them, assistive robots can be perceived in two main ways: tools or partners. Considering the past and latest research, assistive humanoids that are invented to provide physical assistance for the purpose of the elderly people are often invented in the context of a tool analogy. The orientation of conversational robots to hide their interlocutors is essential for natural and efficient Human-Robot Interaction (HRI). Knowing the origin of the sound source is a very important skill for a robot because this skill plays an important role during the interaction, for instance, in calling a person over, or assisting them while they do their work, etc. This assistive humanoid can detect the direction of a user, and orient itself towards him/her, in a complex auditive environment, using only voice and a 4-microphone system. This functionality is integrated within Spoken Human Robot Interaction using dialogue modules and theoretical architecture.

# Contents

<b>Chapter 1: Introduction</b>	10
<b>Chapter 2: ReSpeaker and it's Components</b>	13
2.1 XMOS XVF-3000	14
2.2 MP34DT01-M (Digital Microphone)	16
2.3 Algorithms involved in the Respeaker USB Mic Array	16
2.3.1 Voice Activity Detection (VAD)	16
2.3.2 Direction of Arrival (DOA)	17
2.3.3 Noise Suppression	17
2.3.4 DeReverberation	18
2.3.5 Acoustic Echo Cancellation	18
2.3.6 Beamforming	20
<b>Chapter 3: Beamforming</b>	21
3.1 Fixed Beamforming	25
3.1.1 Butler Matrix	26
3.2 Adaptive Beamforming	27
3.2.1 Minimum Variance Distortionless Response (MVDR)	28
3.2.2 Linearly Constrained Minimum Variance (LCMV)	29
<b>Chapter 4: AIRA Corpus</b>	31
4.1 Triangular Array	32
4.2 3D Array	32
<b>Chapter 5: GCC PHAT</b>	34
5.1 GCC-PHAT method	34
<b>Chapter 6: MUSIC Algorithm</b>	36
6.1 MUSIC Algorithm method	37
<b>Chapter 7: Result and Discussions</b>	42
7.1 GCC-PHAT	42
7.2 MUSIC	42
7.2.1 Data Covariance matrix of the input signal $x_1$ and $x_2$	42
7.2.2 Autocorrelation matrix of the Data Covariance matrix $x_{11}, x_{12}, x_{21}, x_{22}$	44
7.2.3 Outputs for the given data	45
7.2.4 Steering matrix for the Data Covariance matrix	48
7.2.5 MUSIC Spectrum ( $P_{music}$ )	48
7.2.6 DOA Estimation based on MUSIC algorithm Plotting Diagram	49

<b>Chapter 8: Summary and Conclusion</b>	50
First Method : GCC-PHAT	50
Second Method : MUSIC Algorithm	51
<b>Chapter 9: Appendices</b>	52
Appendix I : Code for GCC-PHAT	52
Appendix II : Code for MUSIC Algorithm	54
<b>Chapter 10: Future Scope</b>	56
<b>Chapter 11: Workflow</b>	57
<b>Chapter 12: References</b>	58



# List of Figures

<b>Figure 2.1 : Hardware Overview of ReSpeaker USB Mic Array</b>	<b>13</b>
<b>Figure 2.2 : XMOS XVF-3000</b>	<b>14</b>
<b>Figure 2.3 : Internal chip design of XMOS XVF-3000</b>	<b>15</b>
<b>Figure 2.4 : Block diagram of Acoustic Echo Cancellation</b>	<b>19</b>
<b>Figure 2.5 : Functional Block diagram of ReSpeaker USB Mic Array</b>	<b>20</b>
<b>Figure 3.1 : Classification of Beamforming</b>	<b>22</b>
<b>Figure 3.2 : Fixed and Adaptive Beamforming</b>	<b>25</b>
<b>Figure 3.3 : Butler Matrix for both (i)2x2 and (ii)4x4 matrix</b>	<b>27</b>
<b>Figure 3.4 : Classification of Adaptive Beamforming</b>	<b>28</b>
<b>Figure 4.1 : Triangular Array Configuration</b>	<b>32</b>
<b>Figure 4.2 : 3D Array Configuration (i)Side view (ii)Top view</b>	<b>33</b>
<b>Figure 6.1 : Array of Antennas of M elements with <math>\theta</math> arriving angles</b>	<b>37</b>
<b>Figure 6.2 : Flowchart of Multiple Signal Classification algorithm</b>	<b>39</b>
<b>Figure 7.1 : Table for DOA Estimation using GCC-PHAT for each input signal</b>	<b>42</b>
<b>Figure 7.2 : Data Covariance matrix for x11</b>	<b>42</b>
<b>Figure 7.3 : Data Covariance matrix for x12</b>	<b>43</b>
<b>Figure 7.4 : Data Covariance matrix for x21</b>	<b>43</b>
<b>Figure 7.5 : Data Covariance matrix for x22</b>	<b>43</b>
<b>Figure 7.6 : Autocorrelation matrix for the Data Covariance matrix x11</b>	<b>44</b>
<b>Figure 7.7 : Autocorrelation matrix for the Data Covariance matrix x12`</b>	<b>44</b>
<b>Figure 7.8 : Autocorrelation matrix for the Data Covariance matrix x21</b>	<b>44</b>
<b>Figure 7.9 : Autocorrelation matrix for the Data Covariance matrix x22</b>	<b>45</b>
<b>Figure 7.10 : Outputs for the inputs given in the code (Part I)</b>	<b>45</b>
<b>Figure 7.11 : Outputs for the inputs given in the code (Part II)</b>	<b>46</b>
<b>Figure 7.12 : Outputs for the inputs given in the code (Part III)</b>	<b>47</b>
<b>Figure 7.13 : Outputs for the inputs given in the code (Part IV)</b>	<b>47</b>
<b>Figure 7.14 : Outputs for the inputs given in the code (Part V)</b>	<b>48</b>
<b>Figure 7.15 : Steering Matrix for the Data Covariance matrix x11</b>	<b>48</b>

<b>Figure 7.16 : MUSIC Spectrum (<math>P_{\text{music}}</math>) for <math>x_{11}</math></b>	<b>48</b>
<b>Figure 7.17 : DOA Estimation using MUSIC algorithm Plotting Diagram</b>	<b>49</b>
<b>Figure 11.1 : Sound Source Localisation Workflow Block Diagram</b>	<b>57</b>

# Chapter 1

## Introduction

The aging population is one of the major issues in Asian as well as European countries. Wherein soon, Asian countries, including India, are about to face similar challenges. As the younger population gets educated, there will be a mass migration towards urban cities and abroad seeking job opportunities. Mostly this leaves elderly parents to mend their own lives. Already such patterns are evident in urban Indian cities, where plenty of elderly care homes accommodate such people. However, this puts older adults under a lot of psychological and physical stress and may lead to health deterioration.

This project mainly focuses on the health care of the elderly, wherein many health ailments which occur in old age including Dementia, Alzheimer's, Parkinson's, Heart attack, Diabetes, High Blood Pressure, etc which need special attention and social care. Doctors, nurses, midwives, etc have to take care of the well being of the elderly people and assist them in mundane activities which includes taking the measurements like temperature, checking pulse rate, reminders of tablets and medicine, supporting patients & elderly to the lavatory, and providing food & water but not limited and other tasks which includes taking care of their psychological well-being of the patients (elderly).

Though such tasks are natural for humans; however they can be quite complex to attain through technology. One such possible option is to use robots. Understandably, most of the civil infrastructure is built keeping human ergonomics in consideration. To psychologically engage patients and the elderly, robots similar to humans' look and feel are desirable. Specifically, such humanoids for humankind services are called Socially Assistive Robots (SARs) in the conventional literature.

### **Humanoid Functionality:**

1. Autonomous navigation with obstacle avoidance.
2. Person identification using multi-modal means (mainly using vision and speech).

3. Sound source localization features using an array of microphones.
4. Voice bot consists of automatic speech recognition using multi-microphone, natural language processing, and text to speech synthesis modules in Hindi language. The prototype development involves limited vocabulary.

Internet connection facility using wired and Wi-Fi facilities along with network security features. Depending on the OS system-level support from the industrial partner, network security features may be added.

### **Sound Source Localization :**

One of the fundamental and most important features of sound source detection (localisation) is the ability of a living thing (humans or animals) to estimate source localisation as a first step in behaving accordingly in response to the sound. The detection of sound in three-dimensional space is also related to the capacity to evaluate the distance of the sound source (Moore 1997).

The technology of sound source localization for robots could be a simulation of the human auditory system. It receives sound signals employing a microphone array and other sensors which are placed on the robot. These signals are processed so as to realize sound source position detection, speech recognition and so on.

The technologies for sound source localization supported on a microphone array may be categorized into three classes:

- (1) Directional technology supported on high resolution spectral estimation
- (2) Controllable beamforming technology supported on the largest output power
- (3) Technology supported on time delay of arrival (TDOA)

### **Directional technology supported on high resolution spectral estimation :**

This method aims at narrowband signals, but voice signals are broadband signals which are required to boost positioning accuracy with high computational complexity.

### **Controllable beamforming technology based on the highest output power :**

This method requires a priori knowledge of sound source and environmental noise, and also the computational complexity is additionally high.

### **Technology based on Time Delay of Arrival (TDOA) :**

The TDOA method has powerful real-time applicability and is convenient for single speech sound source localization. Through appropriate improvements to beat noise and reverberation, it can do better positioning accuracy.

# Chapter 2

## ReSpeaker and Components

We used the ReSpeaker USB mic array for this project. Here is a very brief description on the ReSpeaker USB mic array.

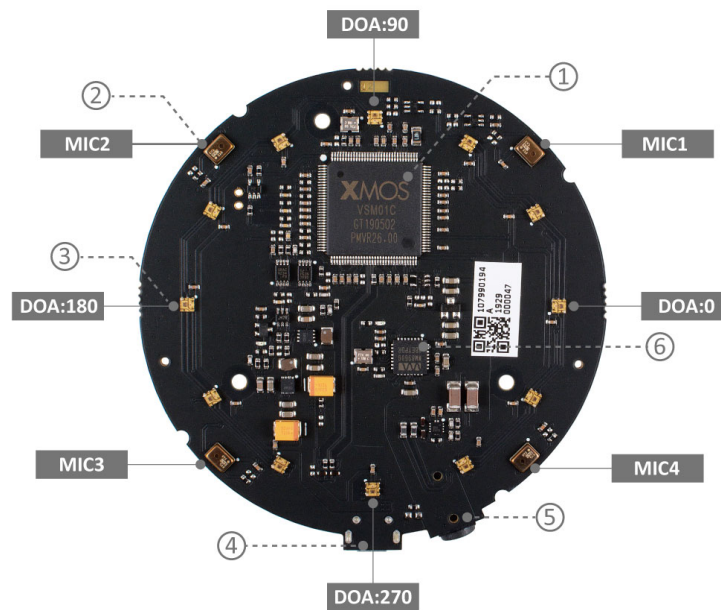


Fig 2.1 Hardware Overview of ReSpeaker USB mic array

The components labeled according to the numbers given in the hardware overview of the ReSpeaker USB mic array are :

1. XMOS XVF-3000
2. Digital microphone
3. RGB (Red Green Blue) LEDs
4. USB Port
5. 3.5mm headphone jack
6. WM8960

## 2.1 XMOS XVF-3000 :

This is a 32 bit multicore microcontroller which is the main part of the ReSpeaker USB mic Array that brings a low latency (time taken for a data to be transferred between its original source and its destination, measured in milliseconds) and timing determinism (provides a measure of reliability that an output will not only be correct but will happen in a specific time).



Fig 2.2 XMOS XVF-3000

- It can execute multiple RealTime tasks simultaneously and communicate between tasks using a high speed network.
- It has an architecture to voice interface with applications.
- It can execute multiple RealTime tasks simultaneously and communicate between tasks using a high speed network.

Key features include:

1. Logical cores: can execute tasks such as computational code, DSP(digital signal processing) code, control software.
2. xTime Scheduler: performs functions similar to a RTOS(real time operating system) in hardware.
3. Ports: The I/O pins are connected to the processing cores by Hardware response ports. These ports have something called port logic which might drive high and low, it can sample the value on its pins optionally awaiting particular conditions.

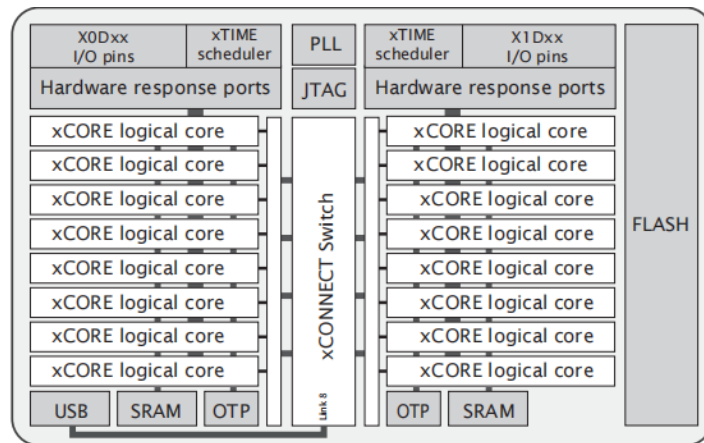


Fig 2.3 Internal chip design of XMOS SVF-3000

4. Memory: Each tile integrates a bank of SRAM for instructions and data, and a block of one time programmable memory will be configured for system wide safety features.
5. PLL (phased locked loop): It is used to create a high-speed processor clock given an occasional speed external oscillator.
6. USB: It provides high speed and full speed device, host, on the go functionality. Data is communicated through ports on the digital node.
7. Software: This chip allows you to program using C, C++ where it provides tested and proven software libraries, which also allows you to quickly add interface and processor functionality like USB, Ethernet graphics driver and audio to your applications.
8. Flash: The device encompasses a built-in 2MB flash.

XVF3000 devices integrate advanced DSP algorithms that include Acoustic Echo Cancellation(AEC), Beamforming, Dereverberation, Noise separation and Gain control. They will be accustomed to deliver superior far-field voice interface solutions for home and conferencing applications.

Other features include:

1. Ambient Temperature Range: 0 °C to 70 °C
2. Power Consumption: 580 mA (typical)



## **2.2 MP34DT01-M (Digital Microphone):**

It is an ultra-compact, low power, omnidirectional. It is a digital MEMS (micro-electro-mechanical system) microphone built with a capacitive detector and an IC interface.

The MP34DT01-M has an acoustic overload point of 120dB SPL with 61dB signal to noise ratio (SNR) and -26dBFS sensitivity.

## **2.3 Algorithms involved in Respeaker USB Mic Array:**

The respeaker mic array has these following built-in algorithms.:

1. Voice Activity Detection
2. Direction of Arrival
3. Beamforming
4. Noise Suppression
5. De reverberation
6. Acoustic Echo Cancellation

### **2.3.1 Voice Activity Detection (VAD):**

Voice Activity Detection may be a technique during which the presence or absence of human speech is detected. The detection is often further used to trigger a process.

VAD, also referred to as speech detection, aims to detect presence or absence of speech and differentiates speech from non speech sections. Some VAD algorithms also provide further analysis, for instance, whether the speech is voiced, unvoiced or sustained. Voice Activity Detection is typically independent of language.

The typical design of a VAD algorithm as follows:

1. There may first be a noise reduction stage, e.g. i.e., spectral subtraction.
2. Then some features or qualities are calculated from a section of the input signal.
3. A classification rule is applied to classify the section as speech or non speech - often this classification rule finds when a value exceeds a particular threshold.

There is also some feedback during this sequence, during which the VAD decision is employed to boost the noise estimate within the noise reduction stage, or to adaptively vary the threshold(s). These feedback operations improve the VAD performance in non-stationary noise (i.e., when the noise varies a lot.)

### **2.3.2 Direction of Arrival (DOA):**

In signal processing, direction of arrival (DOA) denotes the direction from which usually a propagating wave arrives at a degree, where usually a group of sensors are located. These sets of sensors form a sensor array, often there's the associated technique of beamforming which is estimating the signal from a given direction.

### **2.3.3 Noise Suppression:**

The suppression of noise, especially so as to boost the standard of audio signals by selectively reducing the noise in one or more frequency bands.

As technology addressed the difficulty of hearing protection, active noise canceling was developed. which provided a more practical solution. This analog technology functions by detecting the sound coming into the headset, and generating signals that are out of phase with the offending signals, canceling them out. This enables any sounds generated within the headset to be understood more clearly (music, radio communications, etc.).

Unfortunately, these very noise canceling attributes also isolate the wearer from sounds that will make them responsive to hazardous conditions in their surroundings. A serious concern with a noise cancellation approach is that each sound external to the headset is subjected to cancellation.

This can mean that the wearer is less likely to be aware of vehicle movement hazards, heavy equipment motion, alarms or maybe warnings shouted by other workers. With safety being a significant concern for any worksite, this presents a serious risk to the well being of workers. This could also impact productivity and effective communications between workers, who are unable to obviously understand face to face discussions.

Implementing the digital technology provided by noise suppression becomes a lucid preference for noisy work environments. Effective applications can include the oil and gas industry, environments in proximity to aircraft, mining and data centers, among others. Each environment poses unique challenges to hearing protection requirements and can get pleasure from the employment of noise suppression devices.

#### **2.3.4 De Reverberation:**

Dereverberation is the process by which the results of reverberation are eliminated from sound, after such reverberant sound has been picked up by microphones. Dereverberation may be a subtopic of acoustic digital signal processing and is most typically applied to speech but also has relevance in some aspects of music processing.

Dereverberation of audio (speech or music) may be a corresponding function to blind deconvolution of images, although the techniques used are usually very different.

Reverberation itself is caused by sound reflections in an exceedingly large room (or other enclosed space) and is quantified by the area reverberation time and therefore the direct to reverberant ratio. The effect of dereverberation is to extend the direct to reverberant ratio so that the sound is perceived as closer and clearer.

A main application of dereverberation is in hands free phones and desktop conferencing terminals because, in these cases, the microphones aren't near to the source of sound - the talker's mouth - but at arm's length or further distance. Still as telecommunications, dereverberation is important in automatic speech recognition because speech recognizers are usually error prone in reverberant scenarios.

#### **2.3.5 Acoustic Echo Cancellation:**

The Acoustic Echo Cancellation (AEC) block is meant to get rid of echoes, reverberation and unwanted added sounds from a proof that passes through an acoustic space.

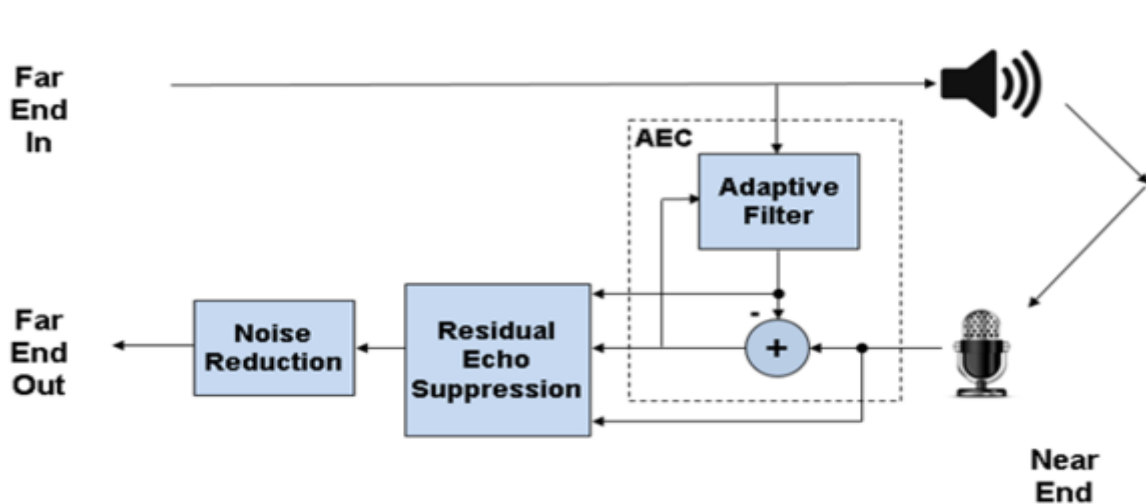


Fig 2.4 Block Diagram of Acoustic Echo Cancellation

As shown within the diagram above, the sound coming from the remote person speaking, referred to as the Far End In, is shipped in parallel to a DSP path and to an acoustic path. The acoustic path consists of an amplifier/loudspeaker, an acoustic environment and a microphone returning the signal to the DSP. The AEC block relies on an adaptive FIR filter.

The algorithm continuously adapts this filter to model the acoustic path. The output of the filter is then subtracted from the acoustic path signal to supply a 'clean' signal output with the linear portion of acoustic echoes largely removed. The AEC block also calculates a residual signal containing non linear acoustic artifacts. This signal is shipped to a Residual Echo Cancellation block (RES) that further recovers the input signal.

The signal is then (optionally) capable of a noise reduction function to supply the output, which is understood because the 'Far End Out'. AEC is required when a far end signal (voice originating at the opposite end of a line of communication) is played over a loudspeaker into a reverberant acoustic space and is picked up by a microphone.

If the AEC algorithm weren't implemented, an echo such as the delay for the sound to travel from the speaker to the microphone, additionally as any reverberation, would be returned to the far end. In addition to sounding unnatural and being unpleasant to concentrate on, the artifacts substantially reduce speech intelligibility.

### 2.3.6 Beamforming:

Beamforming or spatial filtering may be a signal processing technique employed in sensor arrays for directional signal transmission or reception. This can be achieved by combining elements in an antenna array in such a way that signals at particular angles experience constructive interference while others experience destructive interference. Beamforming will be used at both the transmitting and receiving ends so as to realize spatial selectivity. The development compared with omnidirectional reception/transmission is understood because of the directivity of the array.

Beamforming will be used for radio or sound waves. It found numerous applications in radar, sonar, seismology, wireless communications, radio astronomy, acoustics and biomedicine. Adaptive beamforming is employed to detect and estimate the signal of interest at the output of a sensor array by means of optimal (e.g. least squares) spatial filtering and interference rejection.

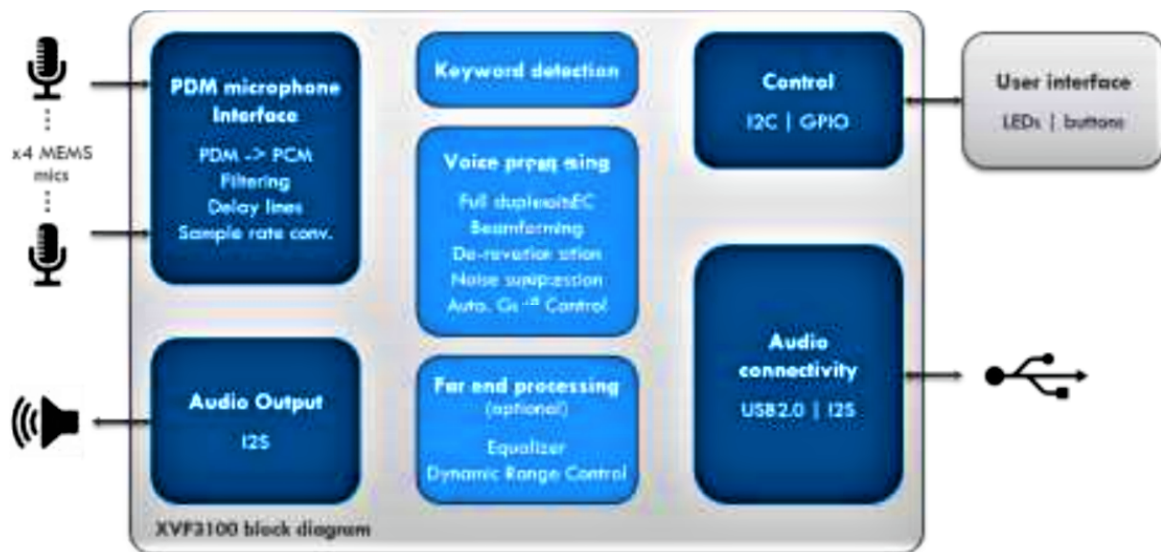


Fig 2.5 Functional Block Diagram of ReSpeaker USB Mic Array

# Chapter 3

## Beamforming

Beamforming can be utilized to attempt to separate and draw out the sound sources in a room, for example, various speakers in a party. This requires the location of the speakers to be known ahead of time, for instance by using the time of arrival from the sources to mics in the array, and deriving the location from the distances.

Contrasted with carrier wave telecommunications, normal sound contains an assortment of frequencies. It is profitable to separate frequency bands preceding the beamforming considering that various frequencies have different optimal beamforming filters (and subsequently can be treated as separate issues, in parallel, and afterward recombined a short time later).

Appropriate disconnecting these bands includes particular non standard filter banks. Conversely, for instance, the standard fast fourier transform (FFT) band filters certainly expect that the main frequencies present in the signal are precise harmonics; frequencies which lie between these harmonics will regularly enact all of the FFT channels (which isn't what is needed in a beamform analysis).

All things considered, filters can be planned in which just nearby frequencies are recognized by each channel (while holding the recombination property to be able to recreate the first signal), and these are regularly non-symmetrical unlike the FFT basis.

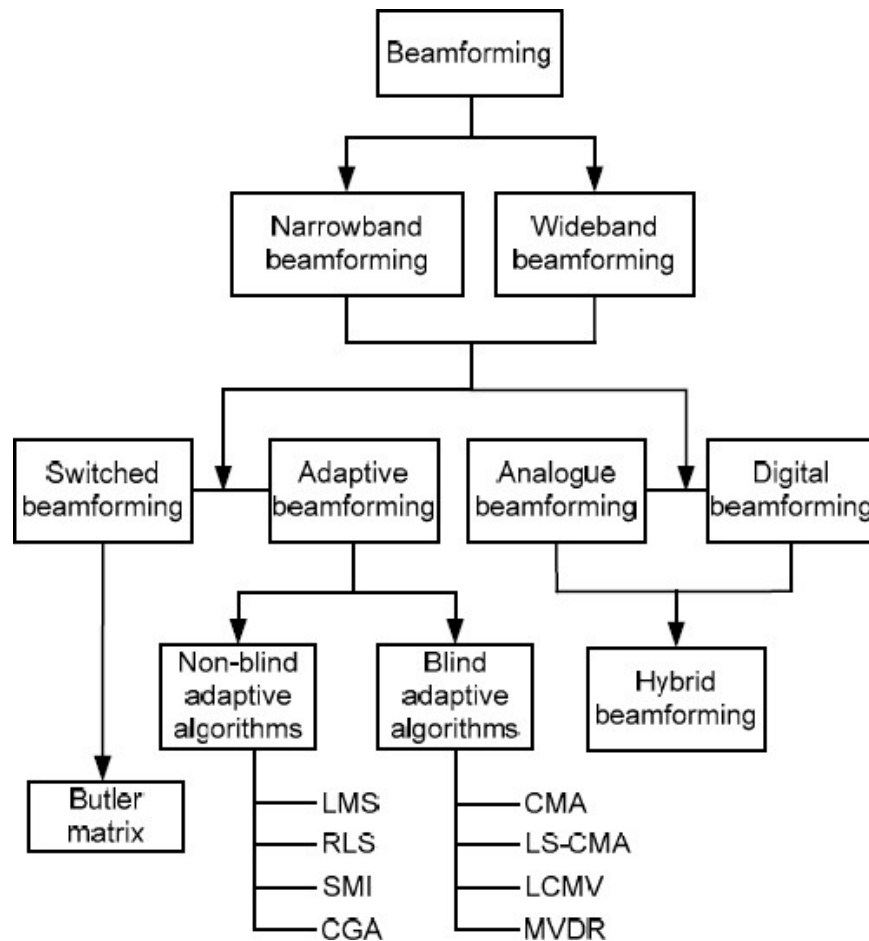


Fig 3.1 Classification of Beamforming

Beamforming can be divided into two classes relying upon the signal bandwidth:

1. Narrowband Beamforming and
2. Wideband Beamforming.

Narrowband beamforming is accomplished by an instantaneous linear combination of the received array signals. In any case, when the involved signals are wideband, an additional processing dimension should be utilized for effective operation, for example, tapped delay lines (or FIR/IIR channels) or the newly proposed sensor delay lines, which lead to a wideband beamforming framework.

Most of current wireless communication applications are still centered around Narrowband Beamforming; nonetheless, Wideband Beamforming becomes a significant subject for future wireless communication applications attributable to 5G requirements concerning high-frequency band signals for accomplishing a very high data rate. The best illustration of wideband beamforming that can be carried out for 5G for laying out very high speeds and high potentials is mm-wave beamforming.

Narrowband implies to radio communications whose signal bandwidth is within the bounds of the coherence band of a frequency channel. This means that in narrowband communications, bandwidth of the signal doesn't fundamentally surpass the coherent bandwidth of the frequency bandwidth of the frequency channel. Narrowband beamformers are regularly used in numerous applications that vigorously rely upon accomplishing solid connections in various working conditions, for example, in handheld and manpack military radios, as well as ISR. They are likewise used for other shorter range, fixed area non military personnel applications, like radio-recurrence recognizable proof (RFID) and business vehicle remote keyless entry (RKE) gadgets.

Wideband refers to broadband communications that use a corresponding wide range of frequencies. Wideband radio channels' functional bandwidth may altogether surpass the coherence bandwidth of the channel. In general, the bandwidths over which wideband beamformers work over bandwidths lower than 1 octave. Additionally ISR, wideband beamformers are appropriate for SIGINT and EW.

Here are a few key contrasts between Wideband frameworks and Narrowband frameworks :

- Overall Complexity: Narrowband systems are similarly less intricate than wideband systems, which require a more different network of circuits and stages.
- Frequency Spectrum: The frequency range of narrowband systems is separated into numerous channels as the frequency permits. Interestingly, for wideband systems, either a significantly high portion or the whole of the frequency range is accessible to its users.



- **Channel to Channel Isolation:** For narrowband systems, transmitted energy can be focused on a smaller portion of the range. Subsequently, channel to channel separation is higher for narrowband systems, compared to wideband systems.
- **Signal Strength:** In narrowband systems, signals blur consistently across frequencies. That being said, adding more frequencies won't reinforce the signal. On the contrary, various parts of wideband signals will be impacted by the varying frequencies. Basically, the signal diminishes as the frequency band broadens, making it harder to send and identify wideband signals.
- **Signal Interference:** One of the primary advantages of having a smaller signal bandwidth (narrowband) is that the likelihood of cross-over with interfering signals is somewhat lower. The bigger the bandwidth (wideband), the higher the likelihood of interference. This implies that wideband communications require more filters to accomplish a higher signal to interference plus noise ratio (SINR). A class of signal processing, a filter is a device or process that eliminates some undesirable 'noise' from a signal.
- **Operational Power:** Narrowband channels have lower operating power requirements, making them ideal for applications that require transmission of restricted (limited) data over generally short distances. The tradeoff for wideband channels is that being able to carry more data over additional distances is that they require essentially higher operating power. Likewise, a wideband channel's higher operating power overcomes more significant levels of signal interference as recently mentioned.
- **Data Rate:** Narrowband systems commonly have lower data rate transmissions, while wideband systems support generally higher data rate transmissions. To lay out plainly, wideband systems consider quicker communication.

In summation, narrowband and wideband systems are appropriate for various applications because of their strengths and weaknesses.

Mainly, Beamforming techniques are said to be classified into 2 types, namely:

1. Fixed Beamforming
2. Adaptive Beamforming

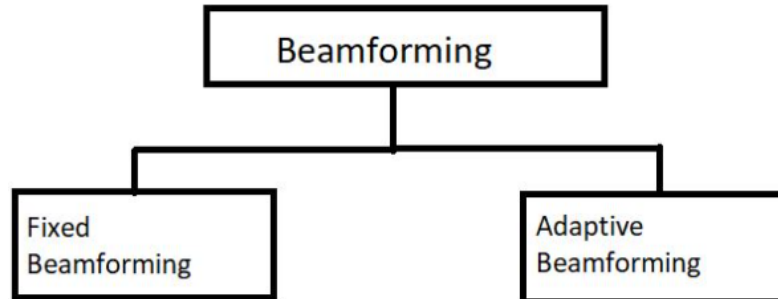


Fig 3.2 Fixed and Adaptive Beamforming

### 3.1 Fixed Beamforming:

A Fixed Beamformer is a spatial filter that has the capability to form a main beam in the direction of the desired signal and, perhaps, place nulls in the direction of interferences. The coefficients of this filter are fixed and don't rely upon the acoustic environment where the array performs.

An switched beam system or fixed bar framework requires an exchanging organization, with the target of choosing a suitable beam to get the ideal signal from a particular terminal. Most of the chosen emitted beam probably won't highlight the desired direction.

Fixed beamforming uses data about the location of the sensors in space and the directions of the desired and interference sources through the steering vectors.

Fixed Array strategies improve the microphone filtering for a specific direction and don't adjust with changing incident source direction. In this way the directional response of the array is fixed to a specific azimuth and elevation. Nevertheless, in the event that the target source is non fixed, the signal enhancement performance is decreased as the source moves away from the steering direction. Spatial beam width constraints might be added to the fixed array to such an extent that the directionality of the response is exchanged for beam width to

make up for small movements in the source. As the beam width increments, so too does the level of ambient noise pickup.

### **3.1.1 Butler Matrix:**

A butler matrix is a beamforming network used to feed a phased array of antenna/microphone elements. Its objective is to control the direction of a beam of radio transmission. It comprises of a  $n \times n$  matrix of hybrid couplers and fixed value phase shifters where  $n$  is some power of 2. The device has  $n$  input ports to which power is applied, and  $n$  output ports (the component ports) to which  $n$  antenna/microphone components are associated.

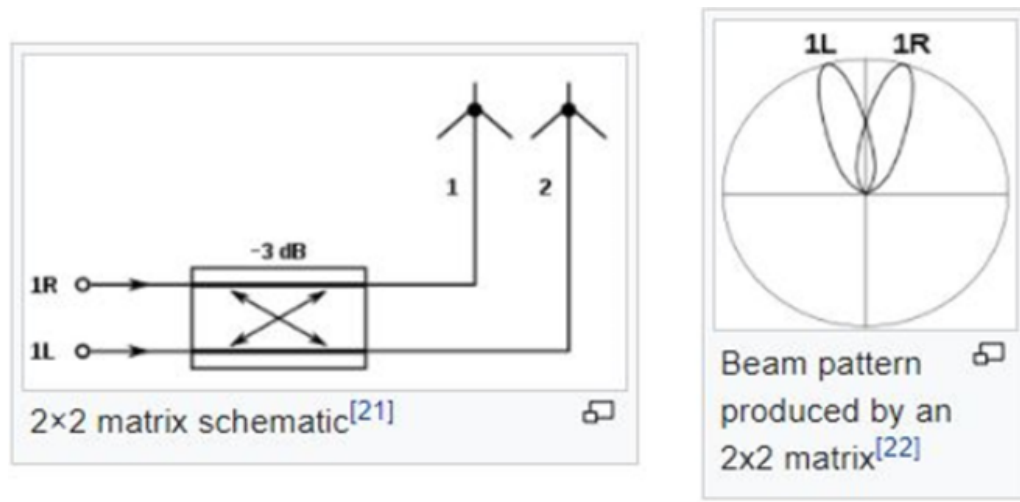
The butler matrix takes care of the capacity to the elements with a progressive phase difference between elements to such an extent that the beam of radio transmission is within the desired direction. The beam direction is controlled by switching power to the specified beam port. More than one beam, or even all  $n$  of them can be activated at the same time.

Butler matrices can be utilized with the two transmitters and receivers. Since they are passive and complementary, a similar matrix can do both - in a transceiver for example. They have the beneficial property that in transmit mode they deliver the full force of the transmitter to the beam, and in receive mode they collect signals from every one of the beam directions with the full gain of the antenna array.

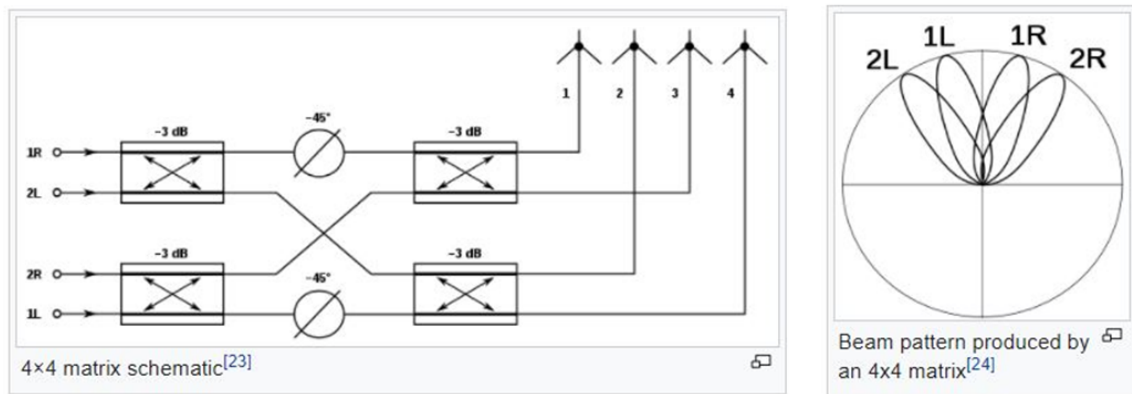
A normal use of Butler matrices is in the base stations of mobile networks to keep the beams pointing towards the mobile users.

The primary characteristics of the Butler matrix are:

1.  $N$  inputs and  $N$  outputs, with  $N$  usually 4, 8 or 16.
2. Inputs are segregated from each other.
3. Phases of  $N$  outputs are linear with respect to position, so the beam is shifted off the main axis.
4. None of the inputs produces a broadside beam.
5. The phase increment between the outputs depends on which input you use.



(i)



(ii)

Fig 3.3 Butler Matrix for both (i) 2x2 and (ii) 4x4 matrix

### 3.2 Adaptive Beamforming:

An adaptive beamformer is a system that performs adaptive spatial signal processing with a variety of array transmitters or receivers. The signals are joined in a way which increases the signal strength to/from a desired direction. Signals to/from different directions are joined in a favorable or detrimental manner, bringing about degeneration of the signal to/from the undesired direction. This technique is utilized in both radio frequency and acoustic arrays, and accommodates directional sensitivity without substantially moving an array of receivers or transmitters.

An adaptive beamforming system depends on principles of wave propagation and phase relationships. Utilizing the principles of superimposing waves, a sequentially higher and lower amplitude wave is produced (for example by delaying and weighting the signal received). The adaptive beamforming system progressively adjusts to maximize or minimize an ideal parameter, like Signal to Interference plus noise ratio.

Adaptive beamforming is majorly classified into 2 techniques, namely:

1. Minimum Variance Distortionless Response (MVDR)
2. Linearly Constrained Minimum Variance (LCMV)

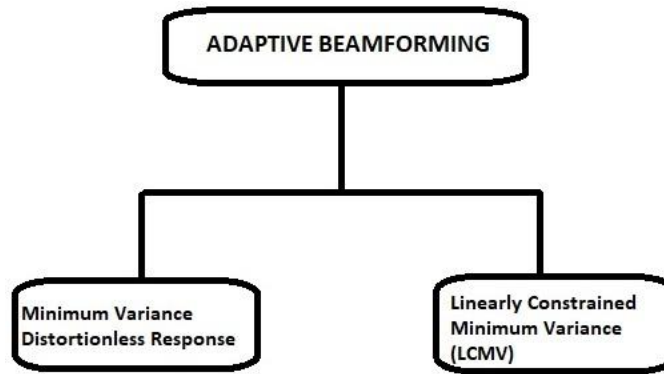


Fig 3.4 Classification of Adaptive Beamforming

### 3.2.1 Minimum Variance Distortionless Response (MVDR):

Minimum Variance Distortionless Response (MVDR) is a typical beamforming algorithm that makes the resulting power with least interference and noise in the desired direction through changing a weight factor.

The minimum distortionless response (MVDR) beamformer is exceptionally sensitive to errors, like the finite snapshots and the steering vector errors. A small mismatch can prompt serious deterioration to its performance, particularly at high sign to noise ratio (SNR).

The resulting power of interference signal adding noise can be expressed as:

$$E \left\{ \left| \omega^H \mathbf{X}_i(t) + \omega^H \mathbf{x}_n(t) \right|^2 \right\} = E \left\{ \left| \omega^H \mathbf{X}_{i+n}(t) \right|^2 \right\} = \omega^H E \left\{ \mathbf{X}_{i+n}(t) \mathbf{X}_{i+n}^H(t) \right\} \omega = \omega^H \mathbf{R}_{i+n} \omega.$$

For convenience,  $\mathbf{X}_{i+n}(t) = \mathbf{X}_i(t) + \mathbf{x}_n(t)$ ,  $\mathbf{R}_{i+n} = E\{\mathbf{X}_{i+n}(t)\mathbf{X}_{i+n}^H(t)\}$

where,  $\mathbf{x}_n(t)$  is the interference signal adding noise. Mathematically,  $\mathbf{R}_{i+n}$  is equal to the covariance matrix of  $\mathbf{X}_{i+n}(t)$ .

To ensure the non distortion output of the desired signals, we can deduce the necessary and sufficient condition:

$$\boldsymbol{\omega}^H \mathbf{s} = 1$$

where  $\mathbf{s}$  is the directional vector with normalizing factor

$$\mathbf{s} = \mathbf{a}(\theta) / \sqrt{\mathbf{a}^H(\theta)\mathbf{a}(\theta)}.$$

Thus, beamforming MVDR algorithm can be expressed as the following minimization problem:

$$\begin{aligned} \min \quad & \boldsymbol{\omega}^H \mathbf{R}_{i+n} \boldsymbol{\omega} \\ \text{such that, } & \boldsymbol{\omega}^H \mathbf{s} = 1. \end{aligned}$$

### 3.2.2 Linearly Constrained Minimum Variance (LCMV):

The linearly constrained minimum variance (LCMV) beamformer has been broadly utilized to extract multiple desired speech signals from a collection of microphone signals, which are likewise polluted by other interfering speech signals and noise components.

LCMV Beamformer is presently utilized widely in EEG/MEG source estimation for which the goal is to select weights such that the output variance of the beamformer is limited, under the constraint that only the brain signal from the desired location is passed.

A volumetric picture of source image can be produced by successively applying beamforming to various predetermined locations or every location on the source space (i.e., the brain). The areas with high variance in source power can be explained by contributing to most of the EEG/MEG estimations, while areas with little variance can be considered as dormant.

The LCMV beamformer makes two primary suppositions in EEG/MEG methods:

1. Brain sources can be adequately demonstrated as dipoles
2. The time-courses of the brain sources to be assessed are uncorrelated.

# Chapter 4

## AIRA Corpus

The aim of the Acoustic Interactions for Robot Audition (AIRA) corpus is to be used for research on sound source localization and separation, likewise as for multi-user speech recognition, in circumstances where the sound source is outside of the microphone array. This provides great potential for Robot Audition applications, additionally as for Auditory Scene Analysis generally, within the aspects of evaluation and model training.

To the Robot Audition community, also on the Auditory Scene Analysis field, it's of great interest to own a corpus that comes with the advantages of the afore-mentioned corpora, such as: a varying amount of microphones, with different array configurations, recorded in real scenarios, etc.

To this effect, this dataset presents an intensive description of the Acoustic Interactions for Robot Audition (AIRA) corpus which we believe covers these benefits, since it's the subsequent characteristics:

1. It uses two array configurations: a triangular array and a 16 microphone three dimensional array.
2. It was recorded in 6 different daily life scenarios, including an anechoic chamber as a reference point.
3. There is a substantial amount of variations between the scenarios in terms of noise presence and reverberation time.
4. Static speech sources were simulated by high end flat-response studio monitors reproducing the recordings from another cleanly recorded corpus in Mexican Spanish: the DIMEx100 corpus (Pineda et al., 2010). All clean speech data from these static speech sources is provided together with the real life recordings.



5. Mobile speech sources were doled out by human volunteers, and their position through time are either provided by a laser based tracking system or by an estimation from their start and end position (to stimulate noisy localization results).

#### 4.1 Triangular Array:

This configuration employs an array of microphones set in an equilateral triangle. The objectives of this configuration are:

1. For algorithms that only require a tiny amount of microphones.
2. For circumstances where there are more sources than microphones.

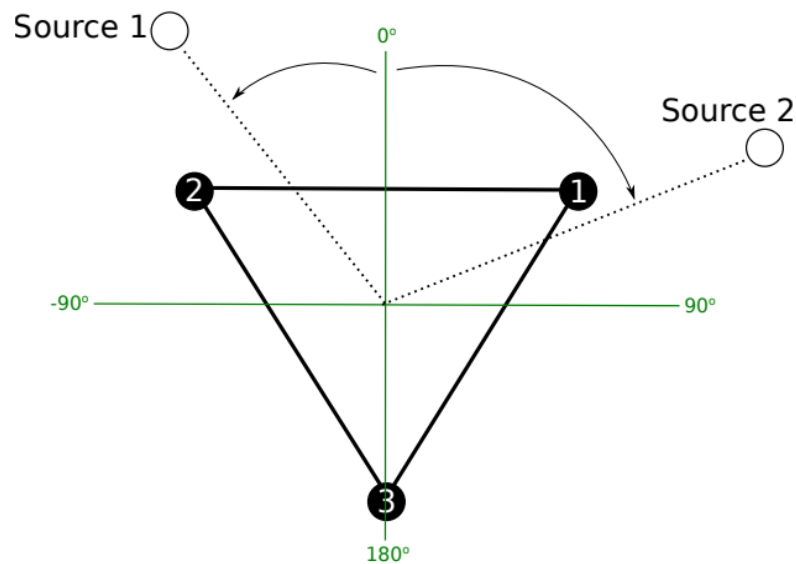


Fig 4.1 Triangular array configuration

#### 4.2 3D Array:

This configuration employs a three dimensional array of 16 microphones, all set over a hollow plastic body.

The objectives of this configuration are:

1. For algorithms that require a high amount of microphones.
2. For circumstances that break the inter microphone free field assumption.

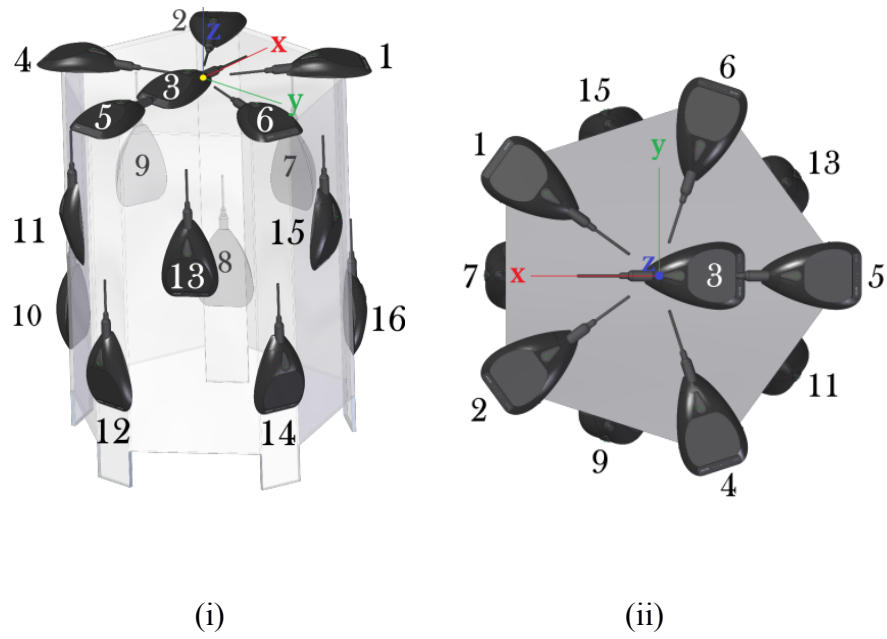


Fig 4.2 3D Array Configuration

(i) Side view, (ii) Top view

# Chapter 5

## GCC PHAT

The computation of the time delay of arrival (TDOA) between each of the considered channels and the reference channel is repeated along the recording in order for the beamforming to respond to changes in the speaker.

In this implementation, it is computed every 250ms (called segment size or analysis scroll) over a window of 500ms (called the analysis window) which covers the current analysis segment and the next. The size of the analysis window and the segment size constitute a trade off. A big analysis window or segment window leads to a reduction in the resolution of changes in the TDOA.

On the other hand, using a very small analysis window reduces the robustness of the cross correlation estimation, as less acoustic frames are used to compute it. The reduction of segment size also increases the computational cost of the system, while not increasing the quality of the output signal.

In order to compute the TDOA between the reference channel and any other channel for any other channel for any given segment it is usual to estimate it as the delay that causes the cross correlation between the two signals segments to be maximum. In order to improve robustness against reverberation it is normal practice to use the Generalized Cross Correlation Phase Transform (GCC PHAT) as presented by Knapp and Carter (1976) and Brandstein and Silverman (1997).

### 5.1 GCC-PHAT method :

Given two signals  $x_i(n)$  and  $x_j(n)$  the GCC-PHAT is defined as:

$$\hat{G}_{PHAT}(f) = \frac{X_i(f)[X_j(f)]^*}{|X_i(f)[X_j(f)]^*|} \quad (1)$$

Where,  $X_i(f)$  and  $X_j(f)$  are the Fourier transforms of the two signals and  $[\ ]^*$  denotes the complex conjugate. The TDOA for these two microphones is estimated as:

$$\hat{d}_{PHAT}(i, j) = \underset{d}{argmax} (\hat{R}_{PHAT}(d)) \quad (2)$$

Where,  $\hat{R}_{PHAT}(d)$  is the inverse Fourier transform of Eq (1).

Although the maximum value corresponds to the estimated TDOA for that specific segment, there are three particular cases that it has been considered not appropriate to use absolutely the maximum from the cross correlation function.

On the other hand, the utmost are often because of a spurious noise or event not associated with the speaker active at that point within the surrounding acoustic region, being the speaker of interest represented by another local maximum of the cross correlation.

On the other hand, when two or more speakers are overlapping with one another, each speaker is represented by a maximum of the cross correlation function, but absolutely the maximum won't be constantly assigned to the identical speaker, leading to artificial speaker switching. So as to effectively enhance the signal it would be optimum to first detect when quite more than one speaker is speaking at the same time and so to obtain a filter and sum signal for each one, stabilizing the chosen delays and avoiding them from constant speaker switching.

# Chapter 6

## MUSIC Algorithm

MUltiple SIgnal Classification (MUSIC) is the widely used technique employed in Direction of arrival estimation. This method is additionally referred to as spectral MUSIC, which estimates the noise subspace from available samples.

The same may be done by either Eigenvalue decomposition of the estimated antenna array correlation matrix or singular value decomposition of the data matrix, with its  $N$  columns adequate to  $N$  snapshots of the array signal vectors.

When the noise subspace has been estimated, a search for an angle pair within the range is formed by searching for steering vectors orthogonal to the noise subspace as possible. Normally this can be accomplished to go looking for peaks within the MUSIC spectrum.

MUSIC belongs to the family of subspace based direction finding algorithms. The MUSIC algorithm includes a high resolution, accuracy and stability under certain conditions for full research and analyses.

In general, it has the following advantages when it is used to estimate a signal's DOA :

1. The ability to simultaneously measure multiple signals.
2. High precision measurement.
3. High resolution for antenna beam signals.
4. Applicable to short data circumstances.
5. It can achieve real time processing after using high speed processing technology.

Most of the DOA estimation methods which are based on signal processing rely on certain assumptions made on the received antenna array signals.

## 6.1 MUSIC Algorithm method :

Let's consider a 2D pattern to briefly describe the signal received model of DOA.

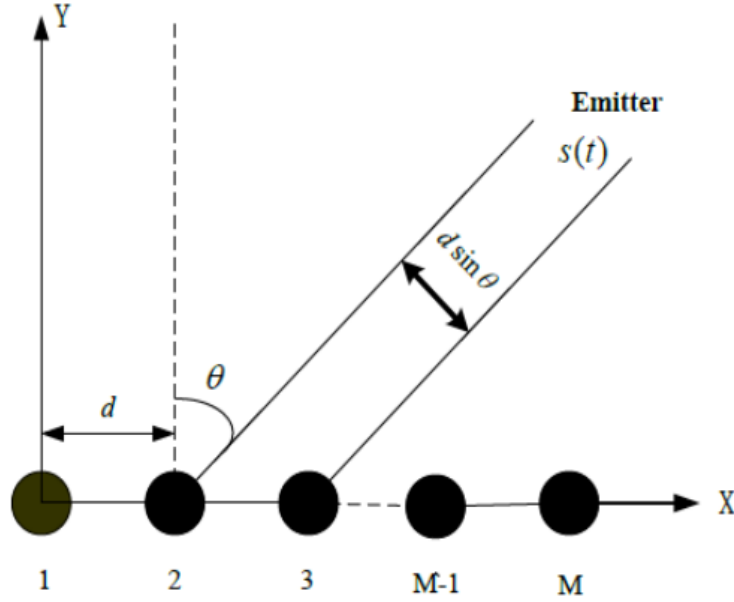


Fig 6.1 Array of Antennas of M elements with  $\theta$  arriving angles

We have considered a scenario with emitting sources from various directions with narrowband property. We have a linear antenna array with M elements to receive each emitting signal separated by distance d as shown in the above figure.

The array is receiving a signal impinging on the array axis at an angle  $\theta$ . Therefore the signal of direction whose element is expressed as and the output signals of received antenna array at the sampling time can be shown as,

$$x(k) = \sum_{i=1}^D a_i s_i(k) + n(k) = A \cdot s(k) + n(k)$$

Where  $A = [a_1, a_2, \dots, a_D]^T$ ,  $s(k) = [s_1(k), s_2(k), \dots, s_D(k)]^T$ , and the signal wavelength. Moreover  $n(k) = [n_1(k), n_2(k), \dots, n_M(k)]$  is assumed to be spatially white Gaussian noise with variance. Generally, the designed antenna arrays require that  $M > D$ .

MUSIC relies on the array correlation matrix. Assuming ergodicity, the time averaged array correlation matrix of is given by,

$$\widehat{R_{xx}} = \frac{1}{k} \sum_{k=0}^{k-1} x(k)x^H(k)$$

where, k is the total number of samples.

MUSIC is a subspace method that can potentially provide high resolution by exploiting the structure of the input data model. The MUSIC spectrum is computed by:

$$P_{MU}(\theta, \phi) = \frac{1}{\alpha^H(H_N)E_N^H\alpha(\phi)}$$

Where  $E_N$  represents the noise eigenvectors of  $\widehat{R_{xx}}$ . In order to determine  $E_N$ , it is necessary to separate the signal subspace from the noise subspace. This separation can be achieved either by using a threshold or by more advanced techniques such as more robust performance.

### Flow chart of MUSIC Algorithm:

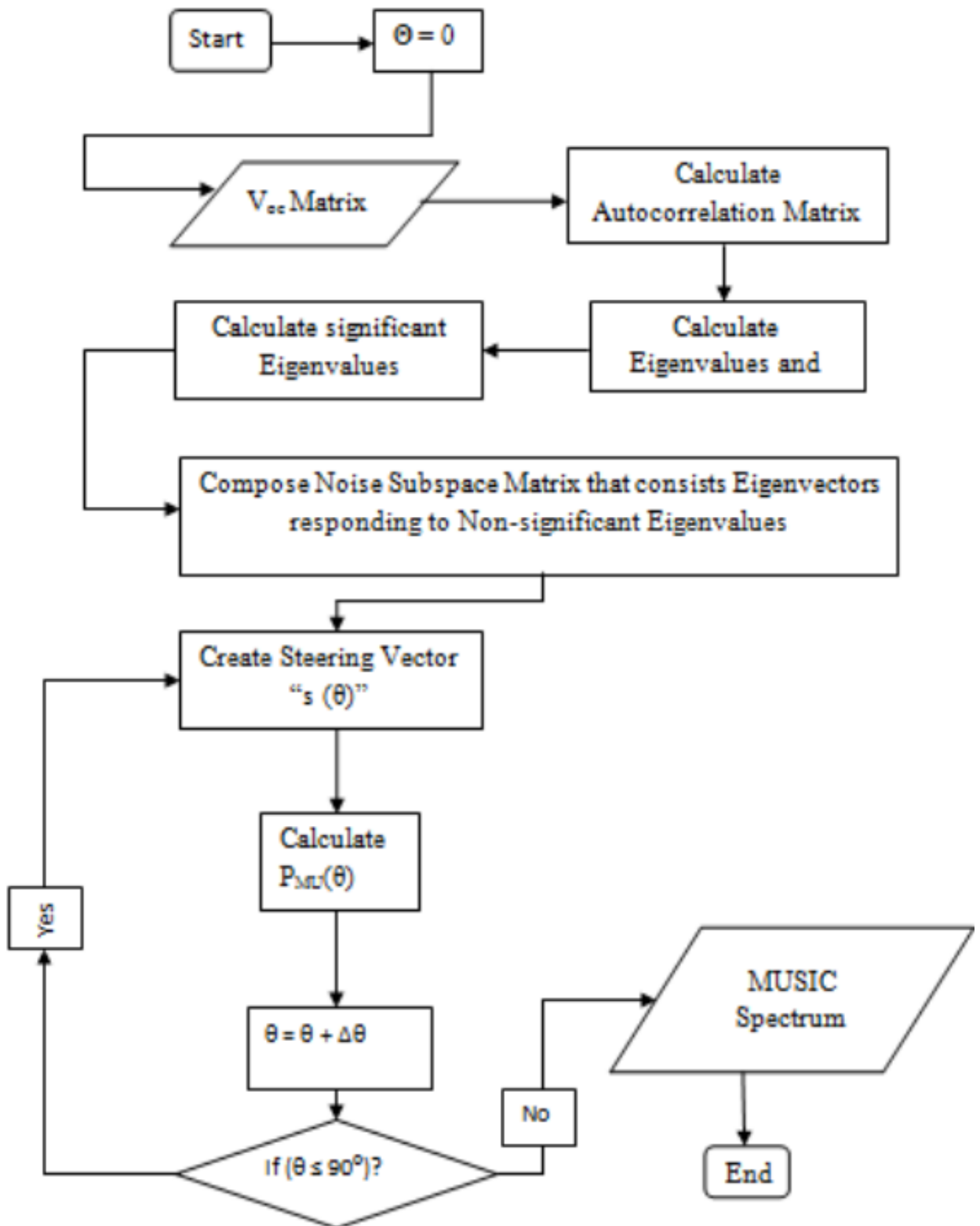


Fig 6.2 Flowchart of Multiple Signal Classification Algorithm



1. Sensor array matrix  $X$  is a  $l$  by  $n$  matrix where  $l$  is the number of sensor antennas and  $n$  is the number of snapshots taken.

$$X^T = [x_1 \dots x_l] \quad (1)$$

2. Total signal induced on the  $l^{\text{th}}$  element of the receiver array can be formulated in the below equation.

$$x_l = \sum_{k=1}^K m_k(t) e^{j2\pi f_0 \tau_l(\theta_k)} + n_l(t) \quad (2)$$

3. Time taken (used as time delay) by the signal to reach the  $l^{\text{th}}$  element of the receiver array from the reference array element by the  $k^{\text{th}}$  signal coming from  $(\theta_k)$  can be calculated in the equation given below.

$$\tau_l(\theta_k) = \frac{\bar{r}_l \cdot \tilde{v}(\theta_k)}{c} \quad (3)$$

where  $\bar{r}_l$  denotes the position vector of  $l^{\text{th}}$  antenna  $\tilde{v}(\theta_k)$  denotes the unit vector directed to  $k^{\text{th}}$  incoming signal.

4. The autocorrelation matrix of the sensor array can be obtained in the equation given below.

$$R = E\{XX^T\} \quad (4)$$

5. Find the correlation matrix of the receiver antenna array elements by using the formula given in the equation below.

$$R = \frac{1}{K} \sum_{n=1}^N x_n x_n^H \quad (5)$$

6. Calculate eigenvalues and eigenvectors of the correlation matrix using equation 5. Compose a noise subspace matrix which is eigenvectors that correspond to smallest eigenvalues of the correlation matrix. For both of all the theta and phi angles, create a steering vector by using the formula given in the equation below.

$$S(\theta) = [e^{2\pi f_0 \tau_2(\theta)} \dots e^{2\pi f_0 \tau_1(\theta)}] \quad (6)$$

7. Calculate Pmu for all angle values by using 10 Peaks of the MUSIC spectrum are the estimated DOA angles.

$$P_{MU}(\theta) = \frac{1}{|S^H(\theta,.)U_L|^2} \quad (7)$$

Where  $U_L$  denotes an  $l$  by  $l - m$  dimensional matrix with its  $l - m$  columns being eigenvectors corresponding to the  $l - m$  smallest eigenvalues of the array correlation matrix.  $S^H(\theta)$  is the hermitian (transpose of complex conjugate) of the steering vector that is used for scanning the range of meaningful angles for the user.

# Chapter 7

## Results And Discussions

### 7.1 GCC-PHAT :

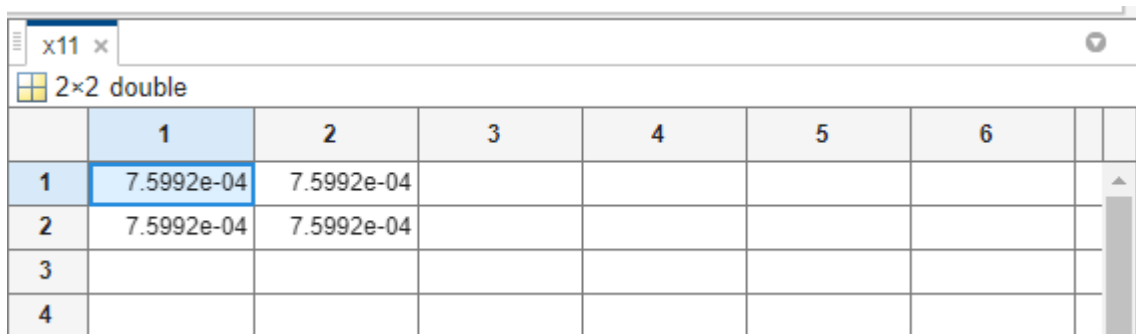
Signal Name	Direction of Arrival
wav_mic1.wav	-19.8633
wav_mic2.wav	0
wav_mic3.wav	90
Audio Track-3.wav	-72.1331
Audio Track-4.wav	90

Fig 7.1: Table for DOA Estimation using GCC-PHAT for each input signal

### 7.2 MUSIC :

#### 7.2.1 Data Covariance matrix of the input signal x1 and x2 :

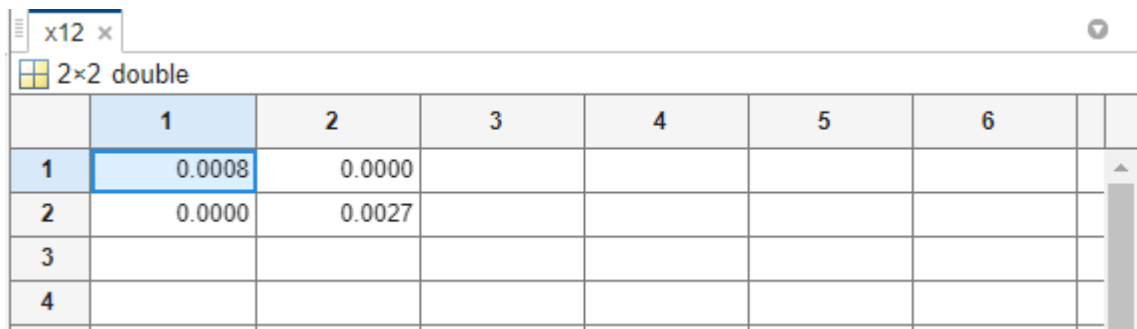
x11:



	1	2	3	4	5	6
1	7.5992e-04	7.5992e-04				
2	7.5992e-04	7.5992e-04				
3						
4						

Fig 7.2: Data Covariance matrix for x11

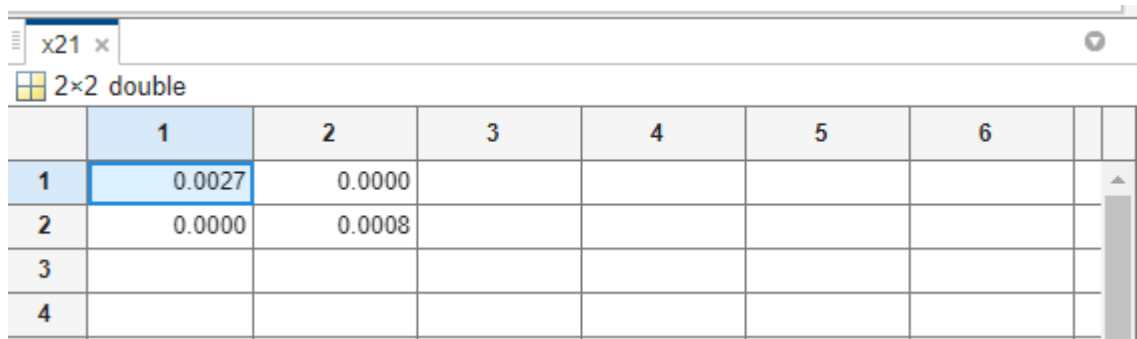
x12:



	1	2	3	4	5	6
1	0.0008	0.0000				
2	0.0000	0.0027				
3						
4						

Fig 7.3: Data Covariance matrix for x12

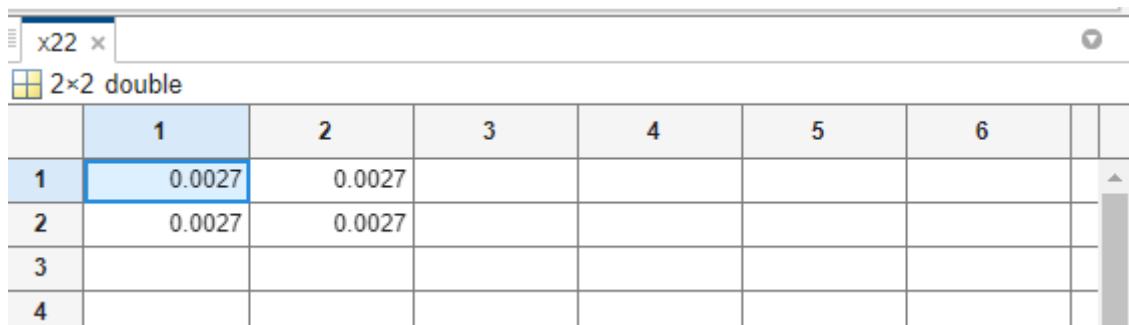
x21:



	1	2	3	4	5	6
1	0.0027	0.0000				
2	0.0000	0.0008				
3						
4						

Fig 7.4: Data Covariance matrix for x21

x22:



	1	2	3	4	5	6
1	0.0027	0.0027				
2	0.0027	0.0027				
3						
4						

Fig 7.5: Data Covariance matrix for x22

## 7.2.2 Autocorrelation Matrix of the Data covariance matrix x11, x12, x21 and x22 :

xa :

	1	2	3	4	5	6
1	5.7748e-07	5.7748e-07	5.7748e-07	5.7748e-07		
2	1.1550e-06	1.1550e-06	1.1550e-06	1.1550e-06		
3	5.7748e-07	5.7748e-07	5.7748e-07	5.7748e-07		
4						

Fig 7.6: Autocorrelation matrix for the Data Covariance matrix x11

xb :

	1	2	3	4	5	6
1	1.2893e-08	2.0208e-06	2.8784e-10	4.5115e-08		
2	5.7777e-07	5.8008e-08	5.8008e-08	7.0715e-06		
3	1.2893e-08	2.8784e-10	2.0208e-06	4.5115e-08		
4						

Fig 7.7: Autocorrelation matrix for the Data Covariance matrix x12

xc :

	1	2	3	4	5	6
1	4.5115e-08	2.0208e-06	2.8784e-10	1.2893e-08		
2	7.0715e-06	5.8008e-08	5.8008e-08	5.7777e-07		
3	4.5115e-08	2.8784e-10	2.0208e-06	1.2893e-08		
4						

Fig 7.8: Autocorrelation matrix for the Data Covariance matrix x21

xd :

	1	2	3	4	5	6
1	7.0712e-06	7.0712e-06	7.0712e-06	7.0712e-06		
2	1.4142e-05	1.4142e-05	1.4142e-05	1.4142e-05		
3	7.0712e-06	7.0712e-06	7.0712e-06	7.0712e-06		
4						

Fig 7.9: Autocorrelation matrix for the Data Covariance matrix x22

### 7.2.3 Outputs for the given data:

1. Receiver's Antenna (Microphone) Spacing - d
2. EigenValues - e11,e12,e21,e22
3. EigenValue Decomposition - [V11,D11], [V12,D12], [V21,D21], [V22,V22]
4. EigenValue and EigenVector - eigenVal, eigenVec

Name	Value	Size	Class
d	170	1×1	double
D11	[0,0;0,0.0015]	2×2	double
D12	[7.5977e-04,0;0,0.0027]	2×2	double
D21	[7.5977e-04,0;0,0.0027]	2×2	double
D22	[0,0;0,0.0053]	2×2	double
e11	[0;0.0015]	2×1	double
e12	[0.0008;0.0027]	2×1	double
e21	[0.0008;0.0027]	2×1	double
e22	[0;0.0053]	2×1	double
eigenVal	[0,0;0,0.0053]	2×2	double
eigenVec	[-0.7071,0.7071;0.7071,...	2×2	double

Fig 7.10: Outputs for the inputs given in the code (Part I)

5. Added Noisy Space - Enoise
6. Length of angle (theta) - i
7. Number of Signal sources - K

8. Wavelength - lamda
9. Peak Locations - locs
10. MUSIC Estimation - MUSIC\_Estim
11. Number of Receiver's Antenna (Microphone) - Nr
12. Local Maxima - pks
13. MUSIC Spectrum - PMusic



Name	Value	Size	Class
eigenVal	[0,0;0,0.0053]	2×2	double
eigenVec	[-0.7071,0.7071;0.7071,...	2×2	double
Enoise	2×500 complex double	2×500	double (complex)
i	3601	1×1	double
K	1	1×1	double
lamda	340	1×1	double
locs	0	1×1	double
MUSIC_Es...	0	1×1	double
Nr	2	1×1	double
pks	Inf	1×1	double
Pmusic	1×3601 double	1×3601	double

Fig 7.11: Outputs for the inputs given in the code (Part II)

14. Value of the formula for MUSIC Spectrum (denominator) - PP
15. Output for the values used for the finding the Steering matrix - s
16. Output for the values used for the finding the Pmusic - SS
17. Steering Matrix - SV
18. Number of Snapshots - T
19. Angle Assigned (-90 to 90) - theta

Workspace			
Name	Value	Size	Class
pks	Inf	1×1	double
Pmusic	1×3601 double	1×3601	double
PP	2.0000	1×1	double
s	1×500 complex double	1×500	double (complex)
SNR	10	1×1	double
SS	[1 + 0i;-1 - 0i]	2×1	double (complex)
SV	2×3601 complex double	2×3601	double (complex)
T	500	1×1	double
theta	1×3601 double	1×3601	double
V11	[-0.7071,0.7071;0.7071,...	2×2	double
V12	[-1,0.0089;0.0089,1]	2×2	double

Fig 7.12: Outputs for the inputs given in the code (Part III)

20. Output for the values used for the finding the Steering matrix -  $V_j$
21. Output for the values used for the finding the Music Spectrum -  $V_n$
22. Input Signal -  $x_1$

Workspace			
Name	Value	Size	Class
T	500	1×1	double
theta	1×3601 double	1×3601	double
V11	[-0.7071,0.7071;0.7071,...	2×2	double
V12	[-1,0.0089;0.0089,1]	2×2	double
V21	[0.0089,-1;-1,-0.0089]	2×2	double
V22	[-0.7071,0.7071;0.7071,...	2×2	double
Vj	2.2361	1×1	double
Vn	[-0.7071;0.7071]	2×1	double
x1	1538090×1 double	1538090×1	double
x11	[7.5992e-04,7.5992e-04...	2×2	double
x12	[7.5992e-04,0;1.6966e-...	2×2	double

Fig 7.13: Outputs for the inputs given in the code (Part IV)



Name	Value	Size	Class
Vn	[-0.7071;0.7071]	2×1	double
x1	1538090×1 double	1538090×1	double
x11	[7.5992e-04,7.5992e-04...	2×2	double
x12	[7.5992e-04,0;1.6966e-...	2×2	double
x2	1538090×1 double	1538090×1	double
x21	[0.0027,1.6966e-05;0,7....	2×2	double
x22	[0.0027,0.0027;0.0027,0...	2×2	double
xa	3×4 double	3×4	double
xb	3×4 double	3×4	double
xc	3×4 double	3×4	double
xd	3×4 double	3×4	double

Fig 7.14: Outputs for the inputs given in the code (Part V)

## 7.2.4 Steering Matrix for the Data Covariance Matrix :

	1	2	3	4	5	6
1	0.9999 - 0.0167i	0.9999 - 0.0167i	0.9999 - 0.0167i	0.9999 - 0.0167i	0.9999 - 0.0167i	0.9999 - 0.0167i
2	0.9999 - 0.0167i	0.9999 - 0.0167i	0.9999 - 0.0167i	0.9999 - 0.0167i	0.9999 - 0.0167i	0.9999 - 0.0167i
3						
4						

Fig 7.15: Steering matrix for the Data Covariance matrix x11

## 7.2.5 MUSIC Spectrum (Pmusic) :

	1795	1796	1797	1798	1799	1800	1801	1802	1803
1	38.6875	40.2711	42.2092	44.7080	48.2298	54.2504	Inf	54.2504	48.2298
2									
3									
4									

Fig 7.16: MUSIC Spectrum (Pmusic) for x11

“Inf” indicates the direction of the source from where the signals are received to the microphones.

### 7.2.6 DOA Estimation based on MUSIC Algorithm Plotting Diagram :

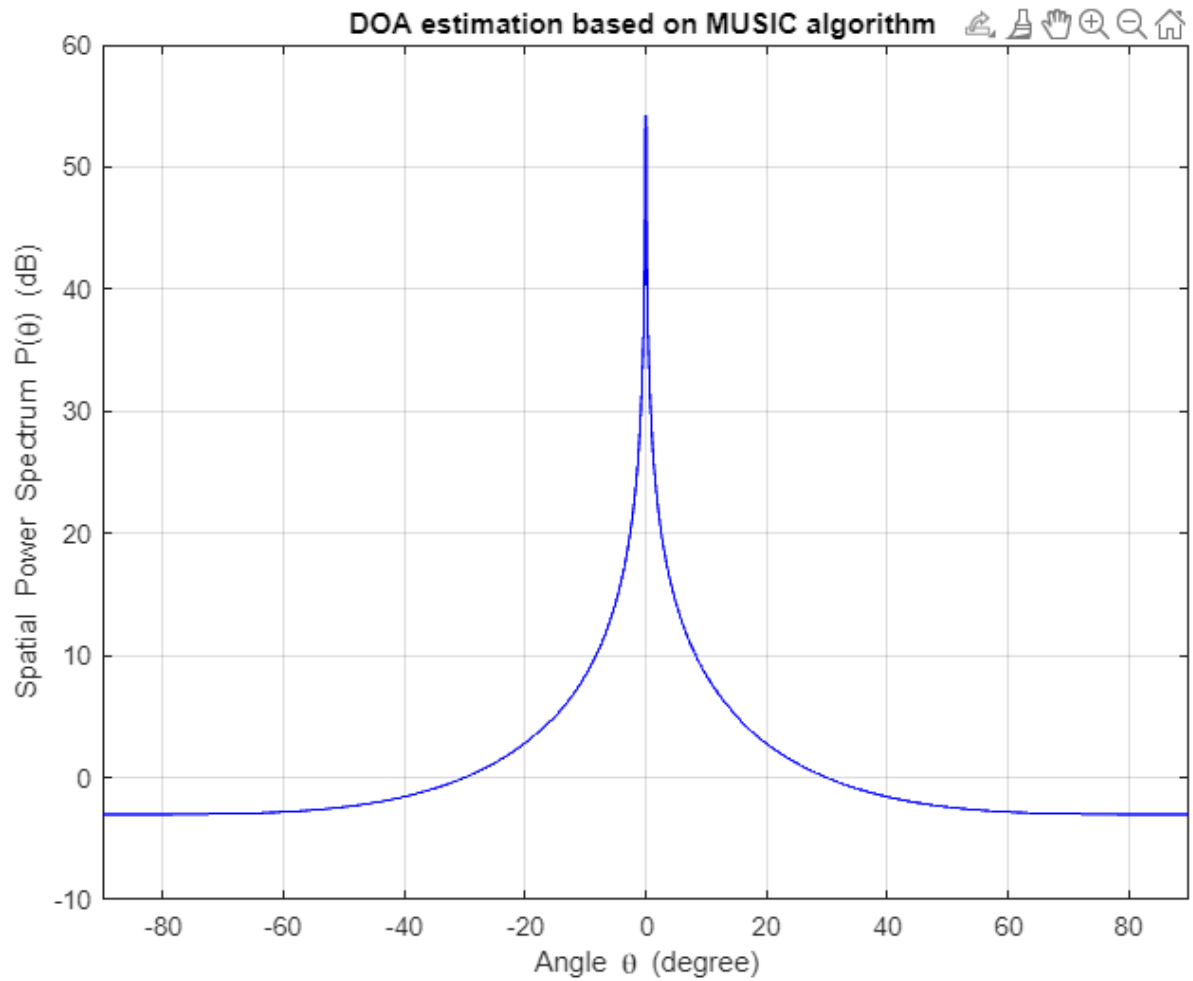


Fig 7.17: DOA Estimation using MUSIC algorithm Plotting Diagram

The Direction of Arrival (DOA) of the two input signals from the Source is at 0 degree angle because the source is perpendicular to the 2 microphones which are aligned in a Unidirectional linear Array fashion.

**Therefore, the Direction of Arrival when tried to estimate using both the GCC-PHAT and MUSIC algorithm for the sound source is proven to be at zero degree angle.**

# Chapter 8

## Summary And Conclusion

The primary objective of the **Sound Source localization module (SSL)** is to automatically estimate the position of sound sources. In the Humanoid, this functionality is useful in several situations, for example: to locate a human speaker or to locate the source of the sound and act accordingly.

The principal reason for studying and using the **ReSpeaker USB mic Array** is that this microphone has different algorithms like Direction of Arrival (DOA), Beamforming, etc., which are used mainly for the Sound Source Localization module (SSL).

Our primary concentration is to Estimate the **Direction of Arrival (DOA)** of the sound source. To implement that we used the **Generalized Cross Correlation with Phase Transform (GCC-PHAT)** and **Multiple Signal Classification (MUSIC) Algorithm** which comes under the Adaptive Beamforming classification of the Beamforming.

### First Method : GCC-PHAT

To Estimate the Direction of Arrival using the **Generalized Cross Correlation with Phase Transform (GCC-PHAT)**, we first find the cross-correlation matrix of the input signal and then estimate the **Time Delay of Arrival (TDOA)** i.e.,  $\tau$  of each input signal coming from the sound source and estimate the time delay between each microphone when receiving the signals from the sound source. We then use the GCC-PHAT inbuilt function to find the Direction of Arrival of the sound source.

## Second Method : MUSIC Algorithm

To Estimate the Direction of Arrival using the **MULTiple Signal Classification (MUSIC) Algorithm**, we first find the Data covariance matrix of each input signal coming from the sound source and we then convert it to Autocorrelation matrix. Using the Data Covariance matrix of each signal, we find the eigenvalues and eigenvectors of each matrix. Then, we input a Noisy Subspace matrix to correlate the eigenvalues and eigenvectors of the covariance matrix. After performing the EigenValue Decomposition method, we create a Steering matrix using the input theta angles. And then we calculate the MUSIC Spectrum (Pmu) using the Steering matrix and plot for the same.

# Chapter 9

## Appendices

### Appendix I : Code for GCC PHAT

```
clc;
clear all;

%Downsampling

fs1 = 48000;
[y,fs1] = audioread('wav_mic1.wav');
[y,fs1] = audioread('wav_mic2.wav');
[y,fs1] = audioread('wav_mic3.wav');

% code to resample audio

%fs1_new = 4000;
%[Numer, Denom] = rat(fs1_new/fs1);
%y_new = resample(y, Numer, Denom);

[x1, fs1] = audioread('wav_mic1.wav');
[x2, fs1] = audioread('wav_mic2.wav');
[x3, fs1] = audioread('wav_mic3.wav');

x12 = xcorr(x1,x2);
x23 = xcorr(x2,x3);
x31 = xcorr(x3,x1);

figure(2)
plot(x12)
subplot(3,1,1)
plot(x23)
subplot(3,1,2)
plot(x31)
subplot(3,1,3)

tau12 = gccphat(x1,x2,fs1);
tau23 = gccphat(x2,x3,fs1);
tau31 = gccphat(x3,x1,fs1);

[tau12, R12, Lag12] = gccphat(x1,x2,fs1);
[tau23, R23, Lag23] = gccphat(x2,x3,fs1);
[tau31, R31, Lag31] = gccphat(x3,x1,fs1);
```

```

tau21 = gccphat(x2,x1,fs1);
tau32 = gccphat(x3,x1,fs1);
tau13 = gccphat(x1,x3,fs1);

```

```

[tau21, R21, Lag21] = gccphat(x2,x1,fs1);
[tau32, R32, lag32] = gccphat(x3,x2,fs1);
[tau13, R13, Lag13] = gccphat(x1,x3,fs1);

```

```

%Direction Of Arrival

```

```

m = 3;
d = 0.0458;
c = 343;

```

```

p1 = (tau12*c)/((m-1)*d);
disp(p1);
q1 = acos(p1);
disp(q1);
N1 = abs(q1);
Ph1 = angle(q1);
disp(Ph1);
disp(rad2deg(Ph1));

```

```

p2 = (tau23*c)/((m-1)*d);
disp(p2);
q2 = acosd(p2);
disp(q2);
N2 = abs(q2);
Ph2 = angle(q2);
disp(Ph2);
disp(rad2deg(Ph2));

```

```

p3 = (tau31*c)/((m-1)*d);
disp(p3);
q3 = acosd(p3);
disp(q3);
N3 = abs(q3);
Ph3 = angle(q3);
disp(Ph3);
disp(rad2deg(Ph3));

```

```

%Reverse Tau's Direction of Arrival

```

```

p4 = (tau21*c)/((m-1)*d);
disp(p4);
q4 = acosd(p4);
disp(q4);
N4 = abs(q4);
Ph4 = angle(q4);
disp(Ph4);
disp(rad2deg(Ph4));

```

```

p5 = (tau32*c)/((m-1)*d);
disp(p5);
q5 = acosd(p5);
disp(q5);
Ph5 = angle(q5);
disp(Ph5);
disp(rad2deg(Ph5));

p6 = (tau13*c)/((m-1)*d);
disp(p6);
q6 = acosd(p6);
disp(q6);
Ph6 = angle(q6);
disp(Ph6);
disp(rad2deg(Ph6));

```

## Appendix II : Code for MUSIC Algorithm

```

clc;
clear all;
close all;

format compact;
T = 500; %No. of snapshots
K = 1; %No. of signal sources
Nr = 2; %No. of receiver's antenna
lamda = 340; %Wavelength
d = lamda/2; %Receiver's antenna spacing
SNR = 10; %Signal to noise ratio

x1 = audioread('wav_mic1.wav');
x2 = audioread('wav_mic2.wav');

x11 = cov(x1,x1); %Data covariance matrix
x12 = cov(x1,x2);
x21 = cov(x2,x1);
x22 = cov(x2,x2);

% To calculate Autocorrelation matrix

xa = xcorr(x11); %autocorrelation matrix
xb = xcorr(x12);
xc = xcorr(x21);
xd = xcorr(x22);

e11 = eig(x11); %Eigenvalues
e12 = eig(x12);
e21 = eig(x21);
e22 = eig(x22);

```

```

[V11,D11] = eig(x11); %Eigenvalue decomposition
[V12,D12] = eig(x12);
[V21,D21] = eig(x21);
[V22,D22] = eig(x22);

% To calculate the Steering Matrix

theta = -90:0.05:90; %Angle assigned between -90 to 90
SV = zeros(Nr,K);
SV = steervec(x11,theta);
Vj = diag(sqrt((10.^(SNR/10))/2));
s = Vj*(randn(K,T) + 1j*randn(K,T));

Enoise = sqrt(1/2)*(randn(Nr,T)+1j*randn(Nr,T)); %Noisy subspace

%MUSIC

[eigenVec,eigenVal] = eig(x11);
Vn = eigenVec(:,1:Nr-K);

%Calculating the MUSIC Spectrum

for i = 1:length(theta)
    SS = zeros(Nr,1);
    SS = exp(-1j*2*pi*d*(0:Nr-1)'*sind(theta(i))/lamda);
    PP = SS'*(Vn*Vn')*SS;
    Pmusic(i) = 1/ PP;
end

Pmusic = real(10*log10(Pmusic)); %Spatial Spectrum function
[pks,locs] =
findpeaks(Pmusic,theta,'SortStr','descend','Annotate','extents');
MUSIC_Estim = sort(locs(1:K));

figure; %Plot MUSIC Spectrum
plot(theta,Pmusic,'-b',locs(1:K),pks(1:K),'r*'); hold on
text(locs(1:K)+2*sign(locs(1:K)),pks(1:K),num2str(locs(1:K)))
xlabel('Angle \theta (degree)'); ylabel('Spatial Power Spectrum P(\theta) (dB)')
title('DOA estimation based on MUSIC algorithm ')
xlim([min(theta) max(theta)])
grid on

```



# Chapter 10

## Future Scope

Some aspects to work upon this project would be :

MUSIC algorithm can get good performance on DOA estimation when the communication environment is normal (SNR isn't very low). However, this algorithm will gradually deteriorate when the SNR continues to decrease.

To solve this problem, we utilize the **wavelet operator** to improve the SNR of received signals during this section. And so it's applied to the MUSIC algorithm named **WMUSIC algorithm**.

We could additionally work on this calculation by adding highlights like **Finding and computing the SNR(Signal-to-Noise ratio)** of the input signal and utilizing fitting calculations when in low or high SNR conditions. Also, moreover we could utilize appropriate **Beamforming techniques** like **MVDR** to further develop and improve the Sound Source Localisation Module.

# Chapter 11

## Workflow

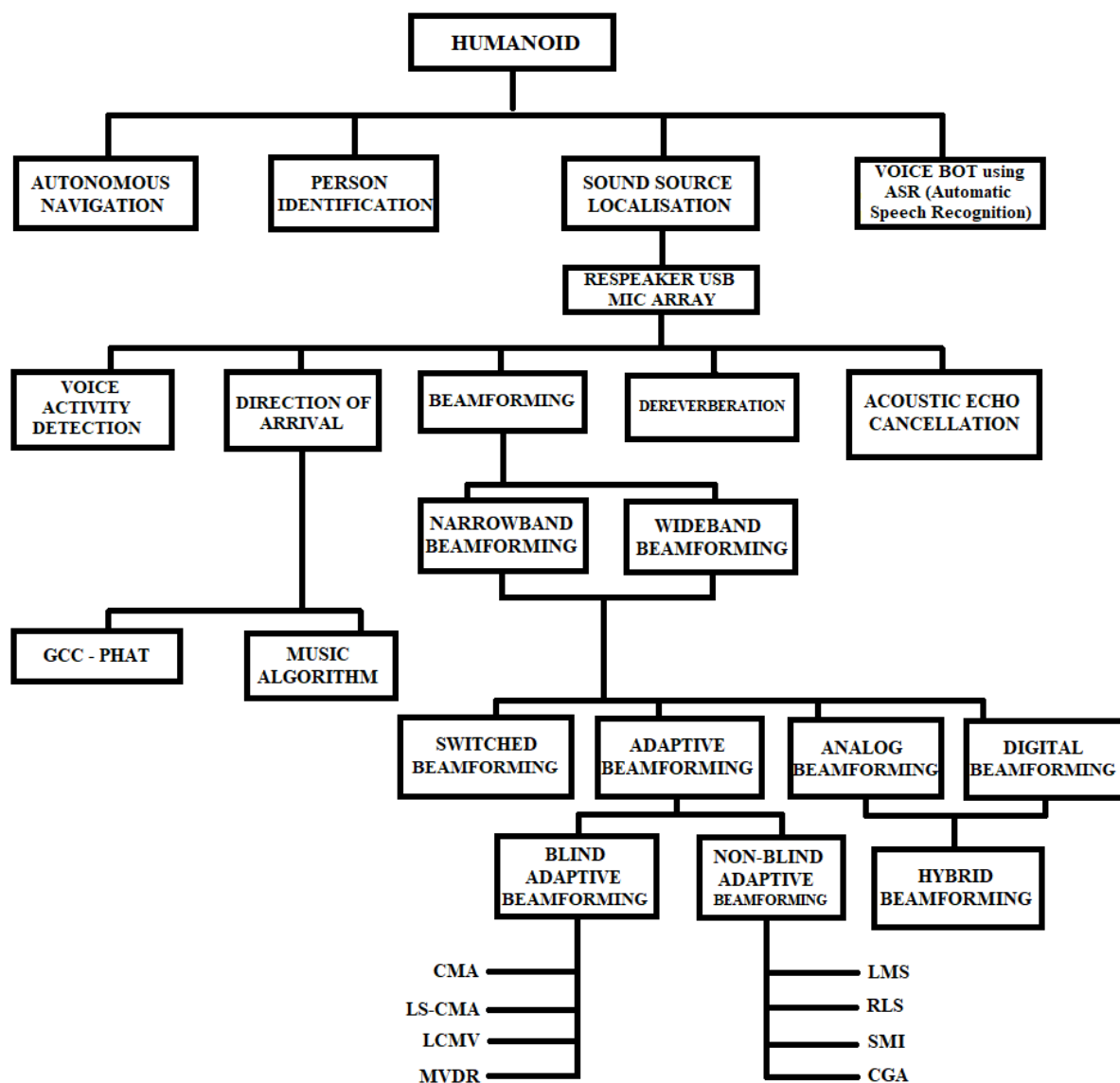


Fig 11.1 Sound Source Localisation Workflow Block Diagram

# Chapter 12

## Reference Papers

- [1] Adaptive Beamforming Using LMS Algorithm, by IJRET: International Journal of Research in Engineering and Technology, May-2014.
- [2] Beamforming Algorithms - Beamformers, by Jørgen Grythe, Norsonic AS, Oslo, Norway. - Published 2015
- [3] Beamforming techniques for massive MIMO systems in 5G: overview, classification, and trends for future research, Ehab ALI†, Mahamod ISMAIL, Rosdiadee NORDIN, Nor Fadzilah ABDULAH, Frontiers of Information Technology & Electronic Engineering.
- [4] Beamforming: A Versatile Approach to Spatial Filtering, Barry D. Van Veen and Kevin M. Buckley, IEEE ASSP Magazine, April 1988
- [5] Calculating Time Delays of Multiple Active Sources in Live Sound, by Alice Clifford and Josh Reiss, Center for Digital Music, Queen Mary, University of London, London, E1 4NS, UK, Presented at the 129th Convention 2010 November 4–7 San Francisco, CA, USA.
- [6] Fixed Beamforming, J. Benesty, I. Cohen, and J. Chen, Fundamentals of Signal Enhancement and Array Signal Processing, Wiley-IEEE Press, 2017.
- [7] Comparative Analysis of Direction of Arrival Estimation Algorithms, by S. S. Jadhav, D. G. Ganage and S. A. Wagh, From International Journal of Advanced Research in Computer and Communication Engineering, July 2016
- [8] Comparison of Direction of Arrival (DOA) Estimation Techniques for Closely Spaced Targets, by Nauman Anwar Baig and Mohammad Bilal Malik, From International Journal of Future Computer and Communication, December 2013
- [9] Detection Sound Source Direction in 3D Space Using Convolutional Neural Networks, 2018 First International Conference on Artificial Intelligence for Industries
- [10] Direction of Arrival Estimation Based on Phase Differences Using Neural Fuzzy Network, IEEE Transactions On Antennas And Propagation, July 2000
- [11] Direction of Arrival Estimation Using MUSIC Algorithm, IJRET: International Journal of Research in Engineering and Technology, March 2014

- [12] On the application of the LCMV Beamformer to Speech Enhancement, 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics.
- [13] Linearly Constrained Minimum Variance Beamforming, LCMV Beamforming.
- [14] Linear Prediction based Dereverberation with advanced speech enhancement and recognition technologies for the reverb challenge, REVERB Workshop 2014.
- [15] MVDR Algorithm Based on Estimated Diagonal Loading for Beamforming, Hindawi Mathematical Problems in Engineering Volume 2017
- [16] ODAS: Open embedded Audition System, EESS.AS, March 2017.
- [17] Robust Adaptive LCMV Beamformer Based on an Iterative Suboptimal Solution, Radio Engineering, June 2015.
- [18] Sound Source Localization Based on GCC-PHAT With Diffuseness Mask in Noisy and Reverberant Environments, IEEE Access, 2020
- [19] Speech Dereverberation, Patrick A. Naylor and Nikolay D. Gaubitch, Imperial College London.
- [20] Speaker localization and tracking with a microphone array on a mobile robot using von Mises distribution and particle filtering - November 2010 - [Robotics and Autonomous Systems](#) 58(11):1185-1196
- [21] XMOS XCore VocalFusion Speaker Datasheet.
- [22] ReSpeaker Mic Array v2.0 Far Field w/4 PDM Microphones Product brief.
- [23] ReSpeaker XVF3000 3100 TQ128 Datasheet
- [24] A Microphone Array and Voice Algorithm based Smart Hearing Aid - August 2019
- [25] Passive Acoustic Array Design for Environmental Monitoring - Fall 2020 - University of Connecticut.
- [26] XMOS Microphone Array Board Support library 2.2.0.
- [27] Acoustic Source Localization and Beamforming: Theory and Practice - EURASIP Journal on Applied Signal Processing 2003:4, 359–370  
c 2003 Hindawi Publishing Corporation.
- [28] Beamforming Techniques for Multichannel audio Signal Separation - [Adel Hidri](#), [S. Meddeb](#), [A. Alaqeeli](#), [H. Amiri](#) - 30 November 2012
- [29] Sound and Ultrasound Source Direction of Arrival Estimation and Localization - Vitaliy Kunin - Chicago, Illinois - December 2010

- [30] GEV Beamforming Supported by DOA Based Masks Generated on Pairs of Microphones - François Grondin, Jean-Samuel Lauzon, Jonathan Vincent, François Michaud - 5th August 2020
- [31] Robotics and Autonomous Systems - Localization of Sound Sources in Robotics: A Review - Caleb Rascon, Ivan Meza - 5th August 2017
- [32] Outdoor Sound Localization using Tetrahedral Array - Ashwin Saraf, Maxime Demurgur - June 7th 2018
- [33] A Study of the LCMV and MVDR Noise Reduction Filters - Mehrez Souden, Jacob Banesty, and Sofiene Affes - 9th September 2010
- [34] Source Localization and Beamforming - Joe C. Chen, Kung Yao, and Ralph E. Hudson - March 2002
- [35] Selecting Sound Source Localization Techniques for Industrial applications - June 2010