ACOUSTIC BEAMFORMING: DESIGN AND DEVELOPMENT OF STEERED RESPONSE POWER WITH PHASE TRANSFORMATION (SRP-PHAT)

AJOY KUMAR DEY
AND
SUSMITA SAHA

THIS THESIS IS PRESENTED AS PART OF DEGREE OF MASTER OF SCIENCE IN ELECTRICAL ENGINEERING WITH EMPHASIS ON SIGNAL PROCESSING

BLEKINGE INSTITUTE OF TECHNOLOGY
AUGUST, 2011
ABSTRACT

Acoustic sound source localization using signal processing is required in order to estimate the direction from where a particular acoustic source signal is coming and it is also important in order to find a solution for hands free communication. Video conferencing, hand free communications are different applications requiring acoustic sound source localization. These applications need a robust algorithm which can reliably localize and position the acoustic sound sources. The Steered Response Power Phase Transform (SRP-PHAT) is an important and robust algorithm to localize acoustic sound sources. However, the algorithm has a high computational complexity thus making the algorithm unsuitable for real time applications. This thesis focuses on describe the implementation of the SRP-PHAT algorithm as a function of source type, reverberation levels and ambient noise. The main objective of this thesis is to present different approaches of the SRP-PHAT to verify the algorithm in terms of acoustic environment, microphone array configuration, acoustic source position and levels of reverberation and noise.
ACKNOWLEDGEMENT

First of all, we would like to thank from deep of our heart to Dr. Nedelko Grbic for his remarkable contribution and the help provided throughout the thesis work. We believe that without his help and guidelines, we could not able to complete our thesis work.

We would also like to thank God for to give the ability to do this thesis and also give thanks to our families and friends for all their support throughout our study period.
### CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>I</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>II</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>III</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>VII</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>VIII</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>11</td>
</tr>
<tr>
<td>1.1 Acoustic Source Localization</td>
<td>11</td>
</tr>
<tr>
<td>1.1.1 Acoustic Source Localization and Source Tracking</td>
<td>12</td>
</tr>
<tr>
<td>1.1.2 Acoustic Source Localization Methods</td>
<td>12</td>
</tr>
<tr>
<td>1.2 Time Difference of Arrival (TDOA) Estimation</td>
<td>13</td>
</tr>
<tr>
<td>1.3 Methods of Steered Beam forming</td>
<td>14</td>
</tr>
<tr>
<td>1.4 Hypothesis</td>
<td>15</td>
</tr>
<tr>
<td>1.5 Organization of the Thesis</td>
<td>15</td>
</tr>
<tr>
<td>2. ACOUSTIC MODEL</td>
<td>17</td>
</tr>
<tr>
<td>2.1 Active Sound Sources</td>
<td>17</td>
</tr>
<tr>
<td>2.2 Multipath Propagation Model of Acoustic Sound Waves</td>
<td>17</td>
</tr>
</tbody>
</table>
2.3 Measurement Scenarios................................................................. 18
2.4 Direction of Arrival................................................................. 19
2.5 Summary of the Acoustic Model............................................... 21

3. TIME DIFFERENCE OF ARRIVAL APPROACHES ....................... 22
  3.1 Time Difference of Arrival.................................................... 22
  3.2 Estimation to Time Difference of Arrival (TDOA) with LMA.......... 23

4. GENERALIZED CROSS CORRELATION (GCC) USING THE PHASE
   TRANSFORMATION METHOD (GCC-PHAT)..................................... 25
  4.1 Generalized Cross- Correlation (GCC).................................... 25
  4.2 Derivation of the Generalized Cross Correlation....................... 26
  4.3 The Phase Transform (PHAT)................................................. 28
  4.4 Generalized Cross Correlation Phase Transform (GCC-PHAT)....... 29
  4.5 Generalized Cross Correlation Phase Transform (GCC-PHAT) in the System.. 29

5. STEERED RESPONSE POWER (SRP) USING THE PHASE TRANSFORM
   (SRP-PHAT).............................................................................. 31
  5.1 Overview of SRP-PHAT.......................................................... 31
  5.2 Beam forming for Steered Response of Power......................... 31
  5.3 The Steered Response of Power............................................. 34
5.4 The Phase Transform (PHAT) .......................................................... 36
5.5 SRP-PHAT .................................................................................. 37
5.6 SRP-PHAT as a Source Localization Method ................................. 37

6. EXPERIMENTAL DESIGN AND MODELING ................................. 39
  6.1 Requirements ........................................................................... 39
  6.2 Architecture ........................................................................... 40
  6.3 Test Environment .................................................................... 40
  6.4 Test Signals Used ................................................................... 42

7. IMPLEMENTATION OF SRP-PHAT ALGORITHM ....................... 44
  7.1 Implementation of Steered Response Power Phase Transform Algorithm ................................. 44
  7.2 Acoustic Sound Source Localization by using Linear Microphone Array ........................................... 44
  7.3 Implementation of Two-Dimensional Speaker Position with SMA .............................................. 46
  7.4 Implementation of Three-Dimensional SMA ........................................................................ 47
  7.5 Optimization ........................................................................... 48
    7.5.1 Stochastic Region Contraction in SRP-PHAT ................................................................. 48
    7.5.2 Coarse to Fine Region Contraction (CFRC) in SRP-PHAT ............................................. 50
  7.6 Computational Cost ................................................................... 52
LIST OF TABLES

Table 1: Summery of the Acoustic Experimental Room Setup for Data Acquisition….. 41
Table 2: Summary of the Signals Used in the System to Drive the Acoustic Source…… 43
Table 3: The SRC algorithm for Finding the Global Maximum……………………… 49
Table 4: The CFRC algorithm for Finding the Global Maximum……………………… 51
Table 5: A Simple Energy Discriminator………………………………………………. 56
Table 6: Performance Evaluation of SRP-PHAT using Full Grid Search Over all Frames……………………………………………………………………………….... 58
LIST OF FIGURES

**Figure 1.1:** Sound Source Localization system using Time Difference of Arrival (TDOA)................................................................................................................................. 14

**Figure 2.1:** Four Microphones in an Array for the Near Field Acoustic Condition of Direction of Arrival................................................................................................................................. 19

**Figure 2.2:** Determination of Azimuth Angle $\theta$ and Elevation Angle $\phi$ from the far field condition in the Direction of Arrival................................................................. 20

**Figure 3.1:** Acoustic wave arrival to microphones and relative time delay between the consecutive microphone pairs................................................................. 23

**Figure 4.1:** Time Difference of Arrival (TDOA) estimation between two microphones approaches................................................................................................................................. 25

**Figure 5.1:** The Steered Response Power algorithm using the delay sum beam forming method................................................................................................................................. 32

**Figure 6.1:** Ideal Reverberant Room Conditions.................................................................................................................................................................................................... 40

**Figure 6.2:** Graphical Representation of Microphone Array design and the Sound Source Position.............................................................................................................................................. 42

**Figure 7.1:** An Acoustic Environment with the Linear Microphone Array Geometry.............................................................................................................................................. 45

**Figure 7.2:** In sixteen speaker positions the SMA geometry in two dimensional ways.... 45

**Figure 7.3:** Position of a speaker on different locations in three dimensional SMA systems.............................................................................................................................................. 46

**Figure 7.4:** Two Dimensional Example of SRC.................................................................................................................................................................................................. 47

**Figure 7.5:** Two Dimensional Example of Coarse to Fine Region Contraction (CFRC). 51

**Figure 7.6:** Top View of the Microphone Array Where Indicating the Source Location and Panels.............................................................................................................................................. 54

**Figure 7.7:** SRP-PHAT Surface (a) Without Interpolation, (b) with Filter Interpolation, (c) With Cubic Interpolation.............................................................................................................................................. 55

**Figure 7.8:** The Simple Energy Discriminator for (a) Source 4, (b) Source 3, (c) Source 4, (d) Source 1.................................................................................................................................................................................................. 57
Figure 7.9: Performance of SRP-PHAT using CFRC Relative to Grid Search

Figure 7.10: Performance of SRP-PHAT using (a) SRC-I, (b) SRC-II, (c) SRC-III Relative to Grid Search
CHAPTER 1

INTRODUCTION

1.1 Acoustic Source Localization

Now a day, new technologies make our daily life easy, convenient and comfortable and also give the assurance our quality of living. Modern technologies update very fast to meet the demand of fifth generation applications. In communication research field, Sound Source localization takes a huge place of research and day by day its increase its importance in new generation communication research application like automobile speech enhancement, active noise cancellation for audio and voice communication, teleconferencing, speech recognition and localization, true source detection, talker characterization and voice capture in reverberant and various environment [11,15, 25]. There are lot many higher and specialized applications part directly and partially involving with this modern day’s technology like artificial intelligence robot navigation, speech separation and sound tracking, security surveillance systems [7, 10, 16]. There are lots many approaches already have under research and development and in coming days we can look and use the applications of this technology.

Linear systems or distributed microphone array approaches have many applications; most of them are directly related our real life stuffs. Final goal of these applications is sound source detecting and locating in most cases. For a instance, in a meeting or conference environment, it is very important that to detect and locate all voices and beam forming each voice to create and measure the independent channels for each speaker [34, 19]. If it is not work perfectly sort out the active source in such kind of environment then system will be significantly fail to maintain the performance level of such systems [27]. We can use Sound Source Localization methods in various operations and to measure the performance level of many applications but the basic aim of this method is to confirm the acceptable and desirable performance level in various operational circumstances.

In real world application of active Sound Source Localization needs to meet the different reliability constraints. In practical situation, we are not able to get good performance level of Sound Source Localization all the time, because off the poor environment of room condition, different scale of noise, relative movement of sound sources, different design of microphone array and robust Sound Source Localization Algorithm [26]. But we can get the better performance of Sound Source Localization in terms of some conditions [2], which are mentioned bellow,

i. Have to use the quality full microphones and enough amounts of microphones to create the perfect environment for detect the Sound Source Localization.

ii. Have to use the perfect and useful microphones placement geometry.

iii. Enough number of active sources in the environment.
iv. Ambient noise and reverberation levels.

The detection and detection decision process of Sound Source Localization methods mostly depend on this above factors. If we increase the number of microphones in the microphone array geometry, averagely we can achieve improved performance in adverse environmental conditions. Some important key factors for to get the optimal solution about the geometry of an array are perfect layout of the experimental room, existing acoustic conditions, number and type of sources [8, 12]. For this reasons, now a day’s some important factors like specific application conditions, hardware availability and other cost effective factors are take into consideration for to design many acoustic Sound Source Localization systems.

1.1.1 Acoustic Source Localization and Source Tracking

Sensor Configuration and efficient microphone placement geometry have a strong effect on performance to get the best accuracy of main objective of acoustic localization and sound source tracking systems [20, 29, 46]. Many factors we have to consider, when we design a perfect environment and systems for specific acoustic source localization like perfect layout of the experimental room design, source placement, source movement, speaking scenarios, acoustic conditions and environment, considerable noise level and prevailing environment [31, 36].

Moreover, we have to consider more many things that are related to the specific objectives of the sound source localization. For individual specification (Single Source Specification/ Multi Source Specification), we have to consider different approaches to find out the actual position of acoustic source.

1.1.2 Acoustic Source Localization Methods

In acoustic source localization methods, there are two effective stage of algorithm we have consider in our real systems in terms of their performance in the real life acoustic system, they are two stage and one stage algorithm [27, 29, 45]. Conditions of the environment, the data being processed, the ways of algorithm implementation process and the active acoustic source itself are the main issue that can be effect the performance and the computational costs of these two categories.

There are two effective steps algorithmic processes have to be completed for the two stage algorithm. The system produces the pair wise time difference of arrivals (TDOA’s) of speech sounds between the pairs of acoustic microphones, in the first step of the process which is commonly known as the Time Delay Estimation (TDE) step [9, 21]. The available time delay estimation and the microphone position date are generated the hyperbolic curve, is the second step. Maximum likelihood estimation, the least square error method, a linear intersection method, spherical interpolations have been proposed methods already to solve this task. On the other hand, pair wise manner is not the perfect process for the one stage algorithm but in order to overcome the limitation associated with some of the making early decisions and reverberation, the one stage algorithm exploits the magnitude of the microphones, in contrast. For an instance, the acoustic beam forming is the correct approach for to support this one. To reinforce the signal with respect to noise or waves propagations in different shorts of directions, beam forming is the
process where the systems delaying the outputs of the microphones in the particular array approach and the adding all those together after delaying. In a particular system, we used to use the beam forming methods for to scan or steer over the predefined region for to find out the active acoustic source localization or all possible acoustic source positions [15, 20]. The actual source would be located at this point, where the system can get the maximum power of beam forming, this one the most possible acoustic source location. This method is well known as the Steered Response Power (SRP) [1, 4]. On short data segment the Steered Response Power (SRP) is also allows one to process efficiently by the integrating the data process from many. One stage method is fairly able to recognize and localize the multiple simultaneous acoustic sources, another remarkable advantage [27, 36]. In that kind of case of operation, multiple times corresponding to the locations of multiple talkers will be peaked by the steered beam former.

From the overall discussion, we can categorized the acoustic source localization procedure in three general group in terms of their approaches,

i. Approaches employing the Time Difference of Arrival (TDOA) information
ii. High resolution spectral estimation concepts techniques adopting.
iii. Maximizing the Steered Response Power (SRP) of a beam former.

The method of estimation and application environment is the main considerable thing and the base part of the broad classifications. The source location has been calculated from a set of delay estimation measured across various pattern of acoustic microphone array model procedure is the first category [19, 37]. In the second method, for the source position estimation, any localization methods can be referred associated with the application of the single correlation matrix. The last approach, any situation where the source position is estimated directly is referred from a filtered, weighted and summed version of the signal data received at the sensors.

1.2 Time Difference of Arrival (TDOA) Estimation

The Time Difference of Arrival (TDOA) approaches are the most widely used acoustic source localization approach [21]. Two step procedures can be adapted with this source localization strategy. In the Time Difference of Arrival (TDOA) approach the acoustic microphone pairs are placed in the system in a particular array model and used to determine the Time Delay Estimation (TDE) of the signal from a point of source [9]. This kind of value and data with the perfect knowledge of the position of microphone establishment in a particular array is used to determine estimation for the source localization. In the system, a specific delay can be mapped to a number of different spatial points along a hyperbolic curve and at the same time, to arrive at source location estimation, the curves are then intersected in some optimal sense [47]. Mainly the bulk of passive talker localization methods are used in the Time Difference of Arrival based model in terms of their computational practicality and reasonable performance under amicable conditions are consider.

For to achieve the effective and perfect acoustic source localization, we have to acquiring good Time Division of Estimation (TDE) of the received speech signals is important. Background noise and the channel multipath due to the room reverberation process is the two major considerable problems which are responsible for the signal degradation of the system and
which can be complicated the acoustic source estimation problem. The inability of the accommodate multi source scenarios is the another noticeable common limitation. By operating at short analysis intervals, some Time Difference of Arrival (TDOA) based methods are used to track several individuals like the sense of the presence of multiple simultaneous acoustic sources, extreme presence of ambient noise, moderate to high reverberation process in the particular acoustic field and typically get the poor Time Difference of Arrival (TDOA) results and the unreliable location estimation where the algorithms assume a single source model [22]. The validity and the accuracy of the delay and location estimation is basic requires for the Time Difference of Arrival (TDOA) based locator for to locate any acoustic source.

1.3 Methods of Steered Beam forming

The microphone array has the special capability of the converging or focusing on the specific location and any specific direction in terms of the volume of signal generation. Such
kind of capability is referred to as a beam forming [15]. This kind of beam forming method containing the sound source localization can be used to steer over the region. This kind of output is known as Steer Response. When the focus direction matches perfectly the true and actual source location in the systems, the Steered Response Power (SRP) will be peak in the system where the delay and sum method is the basic part of any simplest beam former [4].

When the Phase Transform (PHAT) filter is used by the Steered Response beam former, defines a one stage method which is commonly known as the Steered Response Power using the Phase Transform, shortly SRP-PHAT [1, 2, 4]. In the high noise environment and the high reverberation process, Steered Response Power using the Phase Transform (SRP-PHAT) has been shown to be more robust than the two stage methods in the system. The focus of this thesis is improving the Steered Response Power using the Phase Transform (SRP-PHAT) method with its different approach and works more efficiently in real time though it has more expensive computational cost problem of one stage algorithm.

1.4 Hypothesis

For enhancing detection and localization of targets, the research work acoustic sound source localization method has focused on algorithm. The main objective of this thesis is to present different approaches of the Steered Response Power Phase Transform (SRP-PHAT), verify the algorithm in terms of the acoustic environment, microphone array algorithm, different acoustic source position and different band of reverberation and noisy environment and after that evaluation and justification the best approach for to get perfect acoustically active source localization [14, 18]. To study the performance level, separate experiments are designed with the respect to active sound source detection in high reverberant condition and noisy rooms and produced an effective methodology for its solution.

For an efficient evaluation of the acoustic sound source localization, this thesis is particularly focus on the performance and implementation of the Steered Response Power using the Phase Transform (SRP-PHAT) algorithm as a effective function of source type, reverberation levels and ambient noise rather focus on the change in specific environmental scenario in the system and the microphone geometry. When the aim of the experiments is to analysis the different source scenario in the system, the above technique are applicable in the system with little bit modification in terms of situation demand.

1.5 Organization of the Thesis

The organization of the thesis is focused on the design and development of Steered Response Power using the Phase Transform (SRP-PHAT). For this reason, we have modeled and emphasizing our paper with different approaches of SRP-PHAT and their design and how we can develop the algorithm.

In the chapter 2, Acoustic Model we have described about the related issue of active sound source, multipath propagation model, measurement scenario and direction of arrival and letter part we get a summary of the acoustic model and acoustic environment. And in the chapter
3. Time Difference of Arrival focused on the basic Time difference of arrival and the estimation between the time differences of arrival with LMA.

In the chapter 4, Generalized Cross Correlation with Phase Transformation (GCC-PHAT) emphasizing the basic model of basic Generalized Cross Correlation (GCC) and the derivation of GCC, the phase transformation and GCC-PHAT and later part of this chapter we describe about the GCC-PHAT in the system. And in the chapter 5, Steered Response Power with Phase Transformation (SRP-PHAT) are really focused on the over view of Steered Response Power and the beam forming of the Steered Response Power, and the basic derivation of SRP and after that we discuss about the Phase Transformation (PHAT) and letter part we discuss about the Steered Response Power with Phase Transformation (SRP-PHAT) and SRP-PHAT as a source localization method.

In the chapter 6, the experimental design and modeling, here we discussed about the requirements, architecture, test environment and test signal are necessary for to create the actual environment of the acoustic sound source localization with using the Steered Response Power with Phase Transformation (SRP-PHAT) algorithm method. And in chapter 7, implementation of the Steered Response Power with Phase Transformation (SRP-PHAT) algorithm, here we discussed about the basic implementation of SRP-PHAT, acoustic source localization by using linear microphone array approaches, implementation of two dimensional speaker position with SMA and the letter part we discuss about the three dimensional SMA algorithm and in chapter 9, we discuss about the conclusion and the future work of Steered Response Power with Phase Transformation (SRP-PHAT).

In the Next Chapter, we will discuss about the different acoustic model of SRP-PHAT, where we will briefly discuss regarding the different propagation model for different acoustic sound wave, their measurement scenario for different field behavior and their direction of arrival.
CHAPTER 2

ACOUSTIC MODEL

2.1 Active Sound Sources

In real world, we can get different types and different levels of acoustic speech from different kinds of acoustic sources. An acoustic source, either a human talker or a mechanical effective sound source, is not an ideal spherical radiator. In real time acoustic environments, it possesses directionality and spatial attenuation [6]. It is very natural scenario that the phase of the microphones are facing towards the active sources or talker will receive stronger signals than the microphones which are not facing towards or off to the side or behind the source. For to get the simplified process of this system, we can easily assume that the acoustic sources can be effectively modeled as point sources [37, 38].

To make the whole model simple enough, there are some efficient assumptions needed to be made in there,

I. Spherical Sound Waves are emitted by the Source- The complex radiation patterns of the human head models, we never incorporate.

II. Homogeneous Medium- The acoustic sound propagation is non- refractive which gives the assurance of speed of sound where \( c \) is constant everywhere.

III. The Lossless Medium- The medium is lossless; this kind of condition ensures that the medium does not absorb energy from the propagation waves.

In different experimental approaches, the speed of sound \( c \) can be change in order to changes of temperature which is varying from one experiment to another experiment. In a single experiment, \( c \) does not change.

2.2 Multipath Propagation Model of Acoustic Sound Waves

First we consider an ideal acoustic environment, where the propagation of sound waves is interfered by objects like room walls, furniture and people. This kind of interference formed a reverberation environment and creates the multi path propagation of the waves [30, 32]. The performance levels of the microphones array severely affected by this reverberation process. So we have to consider this reverberation environment and the multi path propagation of the wave’s model into the acoustic model to best cope with the realistic conditions.
For both direct path and reflected paths from the active sound source at $\vec{d}_s$ to microphone $m$ at location $\vec{d}_m$, we consider the room impulse response which is denoted through the $h'(\vec{d}_m, \vec{d}_s, t)$. For to describe the characteristics of microphone $m$, $v(\vec{d}_s, t)$ is response we use in this systems. The response function are mostly depends on the source location $\vec{d}_s$, where the location and orientation of microphones $m$ are known and fixed as well in the system. Now we can be modeled the microphone’s signal at microphone $m$ are as follows,

$$x_m(t) = s(t) * h'(\vec{d}_m, \vec{d}_s, t) * v(\vec{d}_s, t) + n_m(t)$$  \hspace{1cm} (2.1)$$

Where $s(t)$ denotes the source signal and $n_m(t)$ denotes the noise corresponding to the $m^{th}$ channel as well. And in the same time $*$ symbolize the linear convolution. From the above expression we can easily assume that noise corresponding $n_m(t)$ is uncorrelated to the source signal $s(t)$.

$h'(\vec{d}_m, \vec{d}_s, t)$ and $v(\vec{d}_s, t)$ is the convolution of the impulse response from the source output to the microphone output. As we know that the impulse response function depends on the source location where as the microphone $m$ is located at a fixed point forever. Denote this response by $h(\vec{d}_s, t)$ and the equation becomes,

$$x_m(t) = s(t) * h(\vec{d}_s, t) + n_m(t)$$  \hspace{1cm} (2.2)$$

From this equation, we can easily assume that the signal which are received at microphone $m$, where the multipath propagation channel’s impulse response and uncorrelated noise factor are taken into account.

2.3 Measurement Scenarios

Microphone setting and phase of the microphone facing towards the pressure wave must be taken into the account, when we want to get the characteristics of an ideal transducer [2, 10, 32]. We can get different types of measurement in order the difference between the active source and microphone. We can divided the measurement scenarios in the following way,

i. **Free Field Behavior** - In free field scenario, the microphone or the array of microphones is situated is an ideal anechoic chamber. In ideal anechoic chamber it has a good and convenient conditions of free field since it has no reverberation or multipath propagation conditions. This kind of conditions mean that the fixed point source like microphone, situated in the same place of the anechoic chamber, which is bound to fulfill the condition of decreasing its sound level 6dB each time the measurement distance doubles.

ii. **Near Field Behavior** - In this kind of near filed scenario, the microphone is located much closed to the effective source; and the distance between the source
and microphones is not further than the 50 cm. In this case, the wave length of the
signal has the same order of the magnitude as the size of the source which creates
the almost lossless condition in the system. The spherical character of the acoustic
field is increased by this measurement process, which is one of the primary goals
of this measurement project, since some microphones are present in the different
response when the spherical divergence increases.

iii. **Far Field Behavior** - In this kind of scenario, the microphone and the source are
situated far away from each other. And this difference is more than 50 cm where
the distance size does not affect the results of the measurement.

### 2.4 Direction of Arrival

We know that the near field situation, the microphones are situated very close to the
acoustic effective source and in the same time for the $M$ element of microphone array, there are
$M$ directions of arrival which is commonly used in DOA (*Direction of Arrival*) [10, 21]. Each
and every DOA is the direct path from the microphones point to the acoustic source.
Mathematically, we can express the full process by a point on the unit vector expression.

$$\overrightarrow{D}_m = \frac{\overrightarrow{d}_m - \overrightarrow{d}_s}{|\overrightarrow{d}_m - \overrightarrow{d}_s|} \quad (2.3)$$

Where, $m = 1, 2, ..., M$. Now we consider an ideal situation for 4 (*Four*) microphones
array design for the near field acoustic condition.

![Figure 2.1: Four Microphones in an Array for the Near Field Acoustic Condition of Direction of Arrival.](image)

In far field condition, we know that the microphones are located far away from the
acoustic sources and all microphones in the array design maintained the same Direction of
Arrival (DOA), which is commonly chosen as the path of system from the origin of the array design to the active acoustic source [21, 22]. We can express the full process mathematically, where the origin of the array is expressed as O in the coordinate system.

\[
\vec{D}_o = \frac{\vec{d}_o - \vec{d}_s}{|\vec{d}_o - \vec{d}_s|}
\]  

(2.4)

In this process, we can express the Direction of Arrival (DOA) through the standard Azimuth Angle \(\theta\) and the Elevation Angle \(\phi\).

In this angle measurement, we can express it educationally in the following way,

\[
\vec{D}_o = \begin{pmatrix} 
\cos \phi & \sin \theta \\
\cos \phi \cos \theta \\
\sin \theta 
\end{pmatrix}
\]  

(2.5)

In the far field condition, the distance or the range between the microphone array and the effective acoustic source cannot be determined in the acoustic source localization problems. The Direction of Arrival is the only spatial information about the Source.
2.5 Summary of the Acoustic Model

For an ideal case condition, we would like to localize a single acoustic source in a perfect room condition for an instance. For this particular work, our experimental acoustic model should be the following way,

i. An acoustic sound source is effectively modeled as a point source.

ii. For single acoustic source localization, the near field condition is applied where we can easily estimate the source position in three dimensional (3D) ways.

iii. For a single course of experiment the speed of sound $c$ is constant.

iv. For to create the best realistic acoustic conditions, reverberation effects, multipath effects and various real life noise should be taken into account.

v. Source positions are tested with the real data collected from the human sources.

In the next chapter, we will discuss about the Time Difference of Arrival Approaches, where we will briefly discuss about the TDOA estimation, estimation with LMA, where we can get the idea of different arrival approaches of SRP.
CHAPTER 3

TIME DIFFERENCE OF ARRIVAL APPROACH

3.1 Time Difference of Arrival

For to find out the position of the effective acoustic source, we have to use the perfect acoustic source localization algorithm. Time Delay Estimation (TDE) or Time Difference of Arrival (TDOA) Technique is the key part of the acoustic source localization algorithm [21]. From good knowledge about the perfect positioning geometry of microphone and also from the source signal we get the time difference of arrival at different microphones pairs used in the geometry, we get the estimation of the acoustic source localization technique. The spatial coherence of the acoustic signal reaching the sensors is one of the key parts of the reliability of a time delay estimation which is sometimes influenced by the distance between the two acoustic microphones, the background noise level, and the reverberation volume of the room.

From the delayed microphone pair signals we can entirely estimate the maximum Generalized Cross-Correlation (GCC), is one of the base parts of maximum Time Difference of Arrival (TDOA) scheme [22]. We can perfectly estimate the time delays by very popular Generalized Cross Correlation (GCC) method [33]. Low computational complexity which is achieved by the Fast Fourier Transform (FFT) implementations is one of causes of the popularity of Generalized Cross Correlation. The mathematical expression of the cross correlation between 2 microphone channel, at $t^{th}$ microphone the signal is denoted by the $x_i(t)$, $X_1(\omega)$ be its Fourier transform over a finite interval $0 \leq t \leq T$.

$$\hat{R}_{GCC}(\tau) = \int_{-\infty}^{\infty} |U(\omega)|^2 \hat{P}_{12}(\omega) e^{j\omega \tau} \tag{3.1}$$

From the above expression, $\hat{P}_{12}(\omega)$ denote the cross power spectrum and $|U(\omega)|$ denote the weighting function.

$$\hat{P}_{12}(\omega) \triangleq X_2(\omega)X_1^*(\omega) \tag{3.2}$$

Here the (*) denotes the complex conjugate. If the weighting function $|U(\omega)|$ is set to “1” in the first equation of the cross correlation, the estimated time delay $\hat{\tau}$ of the Generalized Cross Correlation method will be,

$$\hat{\tau} = \arg \max_{\tau} (\hat{R}_{GCC}(\tau)) \tag{3.3}$$
For the multi path propagation, multi source in the system, high level of background noise in the system and reverberation condition can be destroyed the performance level of the Generalized Cross Correlation (GCC) method [3]. In this kind of condition, the Generalized Cross Correlation with Phase Transform (GCC-PHAT) methods gives the considerably better performance compare with the conventional acoustic sound source localization approaches for Time Difference of Arrival (TDOA) based Sound Source Localization systems. For this condition, the weighting function is designed below for GCC-PHAT as below,

\[ |U(\omega)|^2 = \frac{1}{|P_{12}(\omega)|} \]  \hspace{1cm} (3.4)

3.2 Estimation to Time Difference of Arrival (TDOA) with LMA

The steered response of power is maximized by the linear microphone array approach which is performs for to localized the acoustic source based on the search for a time [39]. Different time will be needed to arrive to individual microphone depends upon the acoustic source location. For computation of GCC the relative delay between pairs of microphones is useful.

Figure 3.1: Acoustic wave arrival to microphones and relative time delay between the consecutive microphone pairs.

From the above figure we can observe that, from the two different acoustic sources position in the particular scenario generates the relative time delays of incoming acoustic signal.
Above figure also depicts that, the sound wave produced directly from the acoustic source position 1 arrives to the microphones array and also generated the relative time delay equal to zero. On the other hand, the sound wave generated from the source position 2 arrives to the microphone array with the different time delays [5]. So from the above condition, if we want to get the maximum relative time delay between the microphones pair, we have set the acoustic sound source at the right most position in the same line of the linear array of microphones [39].

It is very important to take perfect acoustic source localization algorithm and also important to judge the choice of distance between the microphones in the array [42]. Because those important things make an imperative role of the source localization process and increase the computational load as well. From the time delay estimation we can assume the searching interval which one is depends upon the distance between the distances between the consecutive microphones. The maximum time delay estimation which is characterize the search interval is,

$$\tau_{\max} = d \frac{F_s}{c}$$  \hspace{1cm} (3.5)

From the above mathematical expression, sampling frequency is denoted by the $F_s$ and $c$ is the speed of sound, and $d$ denotes the distance between the microphones of the particular system. And from the expression, relative time delay is denoted by the $\tau_{\max}$, which one is maximizes the output power of the steered beam former. To find out the maximum Generalized Cross Correlation (GCC), we have to use a set of time delays which is forced by the distance of the two consecutive microphones in the pair [41].

$$\tau \in (-\tau_{\max}, \tau_{\max}) = D$$ \hspace{1cm} (3.6)

The Time Difference of Arrival (TDOA) is the process of a time delay which one is maximizes the Generalized Cross Correlation (GCC). Having the Time Difference of Arrival (TDOA) estimate, Direction of Arrival (DOA) of coming from the acoustic sound source is computed by the following way,

$$\theta = a \cos \left( \tau \frac{c}{F_s} \right)$$ \hspace{1cm} (3.7)

The direction of arrival to the microphone array of the incoming acoustic sound signal is denoted by the parameter $\theta$.

The next chapter, we will discuss about the Generalized Cross Correlation (GCC), derivation of GCC-PHAT, GCC-PHAT in the system, where we can get the idea of derivation of SRP-PHAT.
CHAPTER 4

GENERALIZED CROSS- CORRELATION (GCC)
USING THE PHASE TRANSFORMATION METHOD
(GCC-PHAT)

4.1 Generalized Cross- Correlation (GCC)

Averaging the multi microphones speech signals, we can assume the delay and sum approaches. Coherence is achieved by focused on an acoustic source, which is specified by the estimated Time Difference of Arrival (TDOA). When we have to find out the TDOA approaches between two microphones in a pair environment, Generalized Cross Correlation (GCC) has been a power full method to determine this approach [3, 33]. For to create a general environment for time delay of arrival approaches, we consider here 4 microphones linear array and a source.

\[
\tau_{mn} = \frac{r_m - r_n}{c}
\]

Figure 4.1: Time Difference of Arrival (TDOA) estimation between two microphones approaches.

We denotes all microphones with \( M_1, M_2, M_3 \) and \( M_4 \) and all are placed in a linear microphone array where the distance between the two microphones consider as \( d_1, d_2 \) and \( d_3 \).
Now if we consider the time delay which one is traveling time from the acoustic sound source to microphone array is

$$\tau_m = \frac{r_m}{c} \quad (4.1)$$

Let two microphones $m$ and $n$ are in the system. The Time Difference of Arrival (TDOA) between these two microphones can be defined as the following way,

$$\tau_{mn} = \tau_m - \tau_n = \frac{r_m - r_n}{c} \quad (4.2)$$

The above equation is conveying the relationship between the Time Difference of Arrival (TDOA) and the effective distances from the acoustic source to the microphones, $r_m$. Several techniques like linear intersection, spherical interpolation etc is used for to estimate the active acoustic sound source localization for Multiple Time Difference of Arrivals (TDOA) [22].

Now we discuss the derivation of Generalized Cross Correlation (GCC) and how do we define the Time Difference of Arrival (TDOA) from the Generalized Cross Correlation (GCC) [1].

### 4.2 Derivation of the Generalized Cross Correlation

For both direct path and reflected paths from the active sound source at $d_s$ to microphone $k$ at location $d_k$, we consider the room impulse response which is denoted through the $h'(d_k, d_s, t)$. For to describe the characteristics of microphone $k$, $v(d_s, t)$ is response we use in this systems. The response function are mostly depends on the source location $d_s$, where the location and orientation of microphones $k$ are known and fixed as well in the system. Now we can be modeled the micro phone’s signal at microphone $k$ are as follows,

$$x_k(t) = s(t) * h'(d_k, d_s, t) * v(d_s, t) + n_k(t) \quad (4.3)$$

Where $s(t)$ denotes the source signal and $n_k(t)$ denotes the noise corresponding to the $k^{th}$ channel as well. And in the same time $*$ symbolize the linear convolution. From the above expression we can easily assume that noise corresponding $n_k(t)$ is uncorrelated to the source signal $s(t)$.

$h'(d_k, d_s, t)$ and $v(d_s, t)$ is the convolution of the impulse response from the source output to the microphone output. As we know that the impulse response function depends on the source location where as the microphone $k$ is located at a fixed point forever. Denote this response by $h(d_s, t)$ and the equation becomes,

$$x_k(t) = s(t) * h(d_s, t) + n_k(t) \quad (4.4)$$
Now consider the another signal at another microphones \( l \), we get

\[
x_k(t) = s(t - \tau_{kl}) * h(d, t) + n_k(t)
\] (4.5)

For to get accurate calculation, we have to include the time delay \( \tau_{kl} \) factor into the source signal \( s(t) \). The delayed version of the source signal at microphone \( k \) is denoted by the \( s(t - \tau_{kl}) \). The time delay estimation from the acoustic sound source to the microphone \( k \), \( \tau_k \) is 0 because we normalized the condition. In this situation, the relative time difference of arrival \( \tau_{kl} \) between the two acoustic microphones \( k \) and \( l \) is the main concern of us [3].

We will get a peak at the time lag when these two microphone signals are cross correlated with each other where these two shifted signals are aligned and corresponds to the Time Difference of Arrival (TDOA), \( \tau_{kl} \). The mathematical expression of the cross correlation of two signals \( x_k(t) \) and \( x_l(t) \) is derived as follows,

\[
c_{kl}(\tau) = \int_{-\infty}^{\infty} x_k(t)x_l(t + \tau)dt
\] (4.6)

Now, we have to take the Fourier Transform of the cross correlation results in a cross power spectrum,

\[
C_{kl}(\omega) = \int_{-\infty}^{\infty} c_{kl}(t) e^{j\omega\tau} dt
\] (4.7)

Now again, we have to apply the convolution properties of the Fourier Transform and we get the following expression,

\[
C_{kl}(\omega) = X_k(\omega)X_l^*(\omega)
\] (4.8)

From the above expression, we get the \( X_l^*(\omega) \), which one is the Fourier Transform of \( x_l(t) \) and the (*) express the complex conjugate form.

Now, if we do the inverse Fourier Transform of the above equation (4.8), we get the cross correlation function in terms of the Fourier Transform of the microphone signals,

\[
c_{kl}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X_k(\omega)X_l^*(\omega)e^{j\omega\tau} d\omega
\] (4.9)

The Generalized Cross Correlation (GCC) is the filtered versions of the cross correlation of \( x_k(t) \) and \( x_l(t) \). Let \( W_k(\omega) \) and \( W_l(\omega) \) is the Fourier Transform version of these two filter
$x_k(t)$ and $x_l(t)$. So, we can express the Generalized Cross Correlation (GCC) in the following way,

$$R_{kl}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (W_k(\omega)X_k(\omega))(W_l(\omega)X_l(\omega))^* e^{j\omega \tau} d\omega$$ \hspace{1cm}(4.10)

Here, $(W_l(\omega)X_l(\omega))^*$ is the complex conjugate part of the $(W_k(\omega)X_k(\omega))$. Now we define the combined weighting function $\Psi_{kl}(\omega)$ are given as follows,

$$\Psi_{kl}(\omega) = W_k(\omega)W_l^*(\omega)$$ \hspace{1cm}(4.11)

We will get Generalized Cross Correlation (GCC) after substituting the equation 4.10 and the equation 4.11,

$$R_{kl}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Psi_{kl}(\omega)X_k(\omega)X_l^*(\omega)e^{j\omega \tau} d\omega$$ \hspace{1cm}(4.12)

The time leg $\tau$ is the Time Difference of Arrival (TDOA) between the two microphones $k$ and $l$ which one maximizes the Generalized Cross Correlation (GCC) $R_{kl}(\tau)$ in the real range.

$$\hat{\tau}_{kl} = \arg\max_{\tau} R_{kl}(\tau)$$ \hspace{1cm}(4.13)

In the real life application, $R_{kl}(\tau)$ has many local maxima which one sometimes make it harder to find out the global maximum. The performance level of Generalized Cross Correlation (GCC) can be affected by the choice of the Weighting functions, $\Psi_{kl}(\omega)$ [33].

### 4.3 The Phase Transform (PHAT)

It has been proven that in real life acoustic environment, the Phase Transform (PHAT) weighting function is robust. In reverberation free conditions, the Phase Transform (PHAT) is the sub-optimal to the maximum likelihood weighting function [2]. So the mathematically we can express the Phase Transform (PHAT) in the following way,

$$\Psi_{kl}(\omega) = \frac{1}{|X_k(\omega)X_l^*(\omega)|}$$ \hspace{1cm}(4.14)
4.4 Generalized Cross Correlation Phase Transform (GCC-PHAT)

Now for to create mathematical expression of the Generalized Cross Correlation Phase Transform (GCC-PHAT), we have to apply the weighting function Phase Transform (PHAT) from the equation 4.14 into the another expression of Generalized cross correlation in the equation no 4.12 and we get the GCC-PHAT expression between the two microphones \( k \) and \( l \) [2, 3, 10] The Expression is defined as below,

\[
R_{kl}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{|X_k(\omega)X^*_l(\omega)|} X_k(\omega)X^*_l(\omega)e^{j\omega\tau} \, d\omega \quad (4.15)
\]

Here, from the above mathematical expression we know that \( \frac{1}{|X_k(\omega)X^*_l(\omega)|} \) is the Phase Transform part of the expression and apart of this one, rest of the expression belong the Generalized Cross Correlation expression.

In a linear \( M \) microphone array system, mathematically we can assume that there are \( \frac{M \times (M-1)}{2} \) pairs of microphone in the system. Assume that the \( D \), is the subset of these pair of microphone. Now for to estimates the Time Difference of Arrival (TDOA) of any subset \( D \) pairs of microphones, we have to use the Generalized Cross Correlation Phase Transform (GCC-PHAT). In the experimental room condition surrounding with the active acoustic sound source, \( \hat{x} \) is the hypothesized point in 3D space in the room for to calculate the true Time Diffe\( \bar{r} \)rence of Arrival (TDOA) for \( D \) pairs of microphones. We can mathematically express the root mean square (RMS) error estimation from the estimated Time Difference of Arrival (TDOA), \( \hat{\tau}_D(\hat{x}) \) and the true Time Difference of Arrival (TDOA), \( \tau_D(\hat{x}) \) as the following way,

\[
E_{RMS}(\hat{x}) = \sqrt{\left\{ (\hat{\tau}_D(\hat{x}) - \tau_D(\hat{x}))^2 \right\}} \quad (4.16)
\]

And \( \hat{x} \) is the estimation of the source location. We can express it mathematically in the following way,

\[
\hat{x} = \arg \min_{\hat{x}} E_{RMS}(\hat{x}) \quad (4.17)
\]

4.5 Generalized Cross Correlation Phase Transform (GCC-PHAT) in the System

LEMS algorithm which one is the Generalized Cross Correlation Phase Transform (GCC-PHAT) based acoustic source localization algorithm, can be used in any real time acoustic sound system. In huge microphone array systems like which has 512 (Five Hundred Twelve)
microphones, which are able to implement 8 (eight) simultaneous LEMS algorithm locators in real time acoustic environment. 16 (sixteen) pairs of acoustic microphones are selected per locator manually in LEMS algorithm [6, 11, 18]. We select three groups of microphones from the 24 (twenty four) microphones of each locator, taking two pairs from each of the group. From the orthogonal sections of the array, maximum acoustic microphones are selected from this zone, that from panels near a corner of the linear array. Microphone pairs are different by its characteristics and complementary sensitivity of their Time Difference of Arrival (TDOA) to the acoustic source direction which are selected on the orthogonal planes and the directional discrimination are improved by the exploiting this effect [44, 48].

The reverberation process and noise are the key factors which can make a good role in the performance level of LEMS Algorithm where we know that we can get good performance level on LEMS algorithm when the reverberation and the noise are relatively low. In real time scenario we can get the long latency from the LEMS algorithm, when its implemented uses over 200 ms of data. It will degrade its performance level in the comparatively high noise and high reverberation conditions.

Steered Response of Power Phase Transform (SRP-PHAT) is the one-stage sound source localization process which one is fulfill our need for perfect acoustic sound source location estimation in the presence of high noise and high reverberation process [1]. In next few chapters, we will discuss about the SRP-PHAT and its implementation in different acoustic environment.
CHAPTER 5

STEERED RESPONSE POWER (SRP) USING THE PHASE TRANSFORM (SRP-PHAT)

5.1 Overview of SRP-PHAT

We naturally use the speech array applications like linear array microphone set up in the system for the voice capture when we applied the sound source localization based beam forming. Again when we applied this one for to get acoustic source localization, the output of the beam former is maximized when the acoustic array is focused on the target location. In order to conquer the limitation in estimation accuracy of Time Difference of Arrival (TDOA) based approaches, the Steered Response Power (SRP) algorithm uses the multitude of microphones, in the presence of noise and reverberation [23]. The spatial filtering capability of a microphone array is used by the Steered Response Power which one is used for the further increment of its applicability for the sound source localization problem. The selective enhancement of the signal from the source of interest is enabled by the Steered Response Power (SRP) [4]. This property of Steered Response Power (SRP) algorithm makes its more robust application for sound source localization approaches.

Compare to the Time Difference of Arrival (TDOA) with the Steered Response Power, the improved feature of SRP make its better in terms of high robustness to reverberation condition for the acoustic sound source localization problem [2]. In this chapter, we will discuss the Steered Response Power (SRP) with the Phase Transform (PHAT), which applies a magnitude normalizing weighting function to the cross spectrum of two microphone signals.

5.2 Beam forming for Steered Response of Power

For both direct path and reflected paths from the active sound source at \( \hat{d}_s \) to microphone \( m \) at location \( \hat{d}_m \), we consider the room impulse response which is denoted through the \( h'(\hat{d}_m, \hat{d}_s, t) \). For to describe the characteristics of microphone \( m \), \( v(\hat{d}_s, t) \) is response we use in this systems. The response function are mostly depends on the source location \( \hat{d}_s \), where the location and orientation of microphones \( m \) are known and fixed as well in the system [17, 49]. Now we can be modeled the micro phone’s signal at microphone \( m \) as are follows,

\[
x_m(t) = s(t) \ast h'(\hat{d}_m, \hat{d}_s, t) \ast v(\hat{d}_s, t) + n_m(t)
\]  

(5.1)

Where \( s(t) \) denotes the source signal and \( n_m(t) \) denotes the noise corresponding to the \( m^{th} \) channel as well. And in the same time \( \ast \) symbolize the linear convolution. From the above
expression we can easily assume that noise corresponding \( n_m(t) \) is uncorrelated to the source signal \( s(t) \).

\[
h'(\vec{d}_m, \vec{d}_z, t) \text{ and } v(\vec{d}_z, t) \text{ is the convolution of the impulse response from the source output to the microphone output. As we know that the impulse response function depends on the source location where as the microphone } m \text{ is located at a fixed point forever. Denote this response by } h(\vec{d}_z, t) \text{ and the equation becomes,}
\]

\[
x_m(t) = s(t) * h(\vec{d}_z, t) + n_m(t) \tag{5.2}
\]

From this equation, we can easily assume that the signal which is received at microphone \( m \), where the multipath propagation channels’s impulse response and uncorrelated noise factor are taken into account.

Delaying the microphone signals \( x_m(t) \) with the appropriate steering delays \( \delta_m \) where, \( m = 1, 2, \ldots, M \) can be creating the unitarily weighted delay and sum beam former in a \( M \) microphone acoustic array system to make them aligned in time, and then summing all these time aligned signals together.

![Figure 5.1: The Steered Response Power algorithm using the delay sum beam forming method](image)

In the above figure, we can see that an array of \( M \) microphones, \( x_i(t) \) defines that a delayed and filtered version of the source signal exist in the microphone channel and we get this delayed versions of \( x_i(t) \) by the time aligning. While the uncorrelated signals present in the \( n_m(t) \), the resulting signals can be summed together so that all copies add constructively.

By the setting of some steering delays equal to the negative values of the propagation delays plus some constant delays, \( \tau_0 \), the copies of the \( s_i(t) \) at each of the individual microphones can be time aligned,
\[ \delta_m = \tau_m - \tau_0 \quad (5.3) \]

Where, we can get the value of \( m \) is \( 1, 2, \ldots, M \), the phase center of the acoustic array design is denoted by the \( \tau_0 \), and among all the microphone array design it is set a largest propagation delay, making all the steering delays greater than or equal to zero \([43]\). All shifting operations are causal, this idea implies by the above feature and the requirement for practical implementation in a system satisfied. This idea also makes the steering delay values relative to one acoustic microphone. \( \tau_m \) is the time delay factor from the source to microphone \( m \). So mathematically we can express the output equation for delay and sum beam former in the following way,

\[ y(t, \delta_1, \delta_2, \ldots, \delta_M) = \sum_{m=1}^{m=M} x_m(t - \delta_m) \quad (5.4) \]

Where, \( \delta_1, \ldots, \delta_m \) are the defined as the steering delay \( M \), the source’s spatial position or the direction are focused or steer by this steering delay. The signal received at the \( m^{th} \) microphone is denoted by the \( x_m \).

Now, in terms of the source signal, the channel’s impulse response and the noise, we can mathematically express the output of a delay and sum beam former \([49]\). So now from the equation no 5.4 the delay and sum beam former output \( y(t, \delta_1, \delta_2, \ldots, \delta_M) \) can be expressed in terms of the microphone signal model \( x_m(t) \) and the steering delays \( \delta_m \). The expression are given below,

\[ y(t, \delta_1, \delta_2, \ldots, \delta_M) = s(t) \ast \sum_{m=1}^{m=M} h(\vec{d}_s, t - \tau_m + \tau_0) + \sum_{m=1}^{m=M} n_m(t - \tau_m + \tau_0) \quad (5.5) \]

A filter and sum beam former is achieved when an adaptive filter is applied to the delay and sum beam former. Considering \( h \) is the impulse response of the individual microphone channels to approximate a band pass filter. With the amplitude of \( M \), the output of the beam former will be a band limited version of \( s(t) \), which one is the larger signal that the signal from any single acoustic microphone. Separating the noise issue from the equation 5.5, we will get the following expression,

\[ y(t, \delta_1, \delta_2, \ldots, \delta_M) = s(t) \ast \sum_{m=1}^{m=M} h(\vec{d}_s, t - \tau_m + \tau_0) \quad (5.6) \]

From the above equation (5.6), we get the output of an \( M \) element which one is the delay and sum beam former in time domain. In the frequency domain, the filter and sum beam former output is,
Where, the $X_m(\omega)$ is the Fourier Transform of the microphone signal $x_m(t)$, and the Fourier Transform of the filter response is the $G_m(\omega)$.

5.3 The Steered Response of Power

A function of $M$ steering delays $\delta_1, ..., \delta_M$ is the general form of steered response. To find out the beam former at a particular position or direction in the space, the steering delays are used. By sweeping the focus of the beam former, the steered response is obtained. The time aligned signals in the acoustic microphone channels add up and the power of the steered response reaches the maximum due to the constructive interference when the main focus of the beam forming corresponds to the source location [4, 6, 15].

In general, the output power of the filter and sum beam former, we get the Steered Response Power (SRP) when steering the beam former over all points $\hat{x}$ in a predefined region. In the frequency domain, we can express its mathematically in the following way,

$$ P(\delta_1, ..., \delta_M) = \int_{-\infty}^{\infty} Y(\omega, \delta_1, ..., \delta_M) Y^*(\omega, \delta_1, ..., \delta_M) \, d\omega $$  \hspace{1cm} (5.8)

From the above equation, $Y^*(\omega, \delta_1, ..., \delta_M)$ is the complex conjugate part of the $Y(\omega, \delta_1, ..., \delta_M)$, which one is the output of the filter and sum beam former. Now substituting the equation 5.8 into the equation 5.7 and we get the following expression,

$$ P(\delta_1, ..., \delta_M) = \int_{-\infty}^{\infty} \left( \sum_{k=1}^{M} G_k(\omega) X_k(\omega) e^{-j\omega\delta_k} \right) \left( \sum_{l=1}^{M} G_l^*(\omega) X_l^*(\omega) e^{j\omega\delta_l} \right) \, d\omega $$  \hspace{1cm} (5.9)

Now we rearranging the whole expression and we get the following one,

$$ P(\delta_1, ..., \delta_M) = \int_{-\infty}^{\infty} \sum_{k=1}^{M} \sum_{l=1}^{M} (G_k(\omega)G_l^*(\omega)) \, (X_k(\omega)X_l^*(\omega)) e^{j\omega(\delta_l-\delta_k)} \, d\omega $$  \hspace{1cm} (5.10)

From the equation 5.3, we can express it like as follows,
\[ \delta_i - \delta_k = \tau_i - \tau_k \quad (5.11) \]

Now inserting the equation no 5.11 into the equation no 5.10 and we will get the following expression from this two equations,

\[ P(\delta_1, ..., \delta_M) = \int_{-\infty}^{\infty} \sum_{k=1}^{k=M} \sum_{l=1}^{l=M} (G_k(\omega)G_l^*(\omega)) \left( X_k(\omega)X_l^*(\omega) \right) e^{j\omega(\tau_i - \tau_k)} d\omega \quad (5.12) \]

Here the information should be notified that the microphone signals and the filter have finite energy because of its integral convergence. Here we can interchanged the summation with the integral and the form of expression are given as follows,

\[ P(\delta_1, ..., \delta_M) = \sum_{k=1}^{k=M} \sum_{l=1}^{l=M} \int_{-\infty}^{\infty} (G_k(\omega)G_l^*(\omega)) \left( X_k(\omega)X_l^*(\omega) \right) e^{j\omega(\tau_i - \tau_k)} d\omega \quad (5.13) \]

The combined weighting function are defines as below,

\[ \Psi_{kl}(\omega) = (G_k(\omega)G_l^*(\omega)) \quad (5.14) \]

And again from the equation, we get

\[ \tau_i - \tau_k = \tau_{lk} \quad (5.15) \]

Now final expression of the equation no 5.14 and 5.15, we will put in the equation no 5.13 and we will get the following expression,

\[ P(\delta_1, ..., \delta_M) = \sum_{k=1}^{k=M} \sum_{l=1}^{l=M} \int_{-\infty}^{\infty} \Psi_{kl}(\omega) \left( X_k(\omega)X_l^*(\omega) \right) e^{j\omega(\tau_{lk})} d\omega \quad (5.16) \]

Now we can recall the Generalized Cross Correlation (GCC),

\[ R_{kl}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Psi_{kl}(\omega) X_k(\omega)X_l^*(\omega) e^{j\omega\tau} d\omega \quad (5.17) \]
The Steered Response Power (SRP) and the Generalized Cross Correlation (GCC) have the almost identical expressions, except that the Steered Response Power (SRP) is summed over all pairs of microphones and also have a constant offset of $2\pi$. For this reason, by summing the Generalized Cross Correlation (GCC) of all pairs of microphones in the array, we can calculate the Steered Response Power (SRP) of a microphone array [2].

5.4 The Phase Transform (PHAT)

The filter and sum beam forming operation is one of the key factors of Steered Response power (SRP) which are directly involved with the noise power reduction proportional to the number of uncorrelated acoustic microphone channels [1]. For beam forming, the correlated noise present great challenge than the uncorrelated noise. The approaches taken by the correlated noise always belongs the independent sources and various types of spectral weighting are included by the reverberation process and directly involving with the Generalized Cross Correlation (GCC) [33].

We can get deemphasize low Signal to Noise Ratio (SNR) spectral regions by the developing the maximum likelihood weights, if the noise spectrum is known. A Phase Transform (PHAT) can be introduced which one effectively whitens the signal spectrum when the noise spectrum is not known [35]. This approach is very much accepted and as well these correlations are done for the Steered Response Power (SRP) likelihood functions and estimate the time delay factor in the system. Different reverberant environments are the key factors which one can play a role in the performance level of Phase Transform (PHAT) estimation [13]. The optimal weighting strategy for minimizing the variance of the time delay estimate is the factor which one is performed by the Phase Transform (PHAT).

In generally, we can express the Phase Transform (PHAT) function mathematically in the following way,

$$\Psi_{kl}(\omega) = \frac{1}{|X_k(\omega)X_l^*(\omega)|}$$  \hspace{1cm} (5.18)

From the above mathematical expression, we get that $\Psi_{kl}(\omega)$ is the weighting function aimed at emphasizing the actual acoustic sound source over the undesired signal and $X_k(\omega)$ is the signal spectrum. This filter whitens the acoustic microphone signal spectrum in the Phase Transform (PHAT) factor [4,15]. The acoustic microphone signal spectrum is effectively flattens by this whitening technique. The Steered Response Power (SRP) can be perfectly introduced in the effective way into the microphone array applications in the acoustic system by whitening the acoustic microphone signals.

The effective comparison between the Phase Transform (PHAT) and other similar weighting function in terms of the effect on Steered Response Power (SRP), we get that the PHAT effect on SRP perform better to compare with the other under the realistic reverberant
operating conditions [26]. Within the high noise and high reverberation condition, Steered Response Power Phase Transform (SRP-PHAT) will peak at the actual acoustic sound source location, which one is the hypothesis of the SRP-PHAT.

5.5 SRP-PHAT

From the above discussion, we know that we can get the Steered Response Power with using the Phase Transform (SRP-PHAT) when the weighting function Phase Transform, PHAT is applied in the Steered Response of Power. By the equation, we can express it in the same way if we take the Phase Transform (PHAT) from the equation 5.18 and applied it in the equation no 5.16 and we get the mathematical expression of SRP-PHAT which one is express in below,

\[ P(\delta_1, ..., \delta_M) = \sum_{k=1}^{k=M} \sum_{l=1}^{l=M} \int_{-\infty}^{\infty} \frac{1}{|X_k(\omega)X_l^*(\omega)|} (X_k(\omega)X_l^*(\omega))e^{j\omega(\tau_{lk})}d\omega \]  \hspace{1cm} (5.19)

It is the perfect expression for the Steered Response Power Phase Transform (SRP-PHAT).

5.6 SRP-PHAT as a Source Localization Method

The above SRP-PHAT function formed by the summing of the elements and this SRP-PHAT form a symmetric matrix with fixed energy terms on the diagonal since the Generalized Cross Correlation (GCC) between the microphone k and the microphone l is the same factor as the Generalized Cross Correlation (GCC) between the microphone l and k [3, 4]. Therefore, the upper part of the matrix or the lower part of the matrix that can be changed the part of the Steered Response Power Phase Transform (SRP-PHAT) with the \( \hat{x} \).

On the other hand, the part of the Steered Response Power Phase Transform (SRP-PHAT) that can be change with the \( \hat{x} \) can be computed by the Generalized Cross Correlation (GCC) with the elements summing factor of not all pairs of the M acoustic microphone array design, but only a subset \( D \) of the pairs, where, \( D = [k, l], \forall k \in [1, ..., M - 1], M \geq l > k \),

\[ P'(\delta_1, ..., \delta_M) = \sum_{k=1}^{k=M} \sum_{l=k+1}^{l=M} \int_{-\infty}^{\infty} \frac{1}{|X_k(\omega)X_l^*(\omega)|} (X_k(\omega)X_l^*(\omega))e^{j\omega(\tau_{lk})}d\omega \]  \hspace{1cm} (5.20)

So, that the acoustic sound source location is in that place, where the points give the maximum weighted output power the Steered Response Power Phase Transform (SRP-PHAT) of the beam former. The location estimates \( \hat{x}_s \) for a single sound source measurement is,

\[ \hat{x}_s = \arg\max_{\hat{x}} P'(\hat{x}) \]  \hspace{1cm} (5.21)
Where Steered Response Power Phase Transform (SRP-PHAT) is denoted by the $P'(\tilde{x})$ at point $\tilde{x}$ which one is the described in the equation no 5.20. We know that functional evaluation is the calculation of any particular point of $P'(\tilde{x})$. 
CHAPTER 6

EXPERIMENTAL DESIGN AND MODELING FOR SRP-PHAT

6.1 Requirements

Need hardware and software both support for capturing acoustic sound data and computerized simulation result for implementation of acoustic sound source localization. Different types of microphone array geometry like linear array and square array geometries are commonly used in beam forming in terms of hardware part. Each array contains four, eight or sixteen microphones according to the array size or acoustic system model [18, 40].

In the real acoustic system, for amplification the sound energies for to reach up to the microphone layer, we use a pre-amplifier. But in real time data processing, the sound signal naturally in weak when it is captured by the microphone and requires amplification factor before processing them. For to get good recoding signals from the acoustic microphone array design, we need a good sound recording system which one can be connected with both computer and a speaker as well. In the laboratory, we need an acoustic speaker which sometimes acts as a source in a acoustic environment [47].

For to implement the proposed algorithm, we can use our personal laptop or desktop computer both in laboratory and the simulation based environment. We can use the MATLAB latest version in terms of software and organizing the simulation. For interfacing the device with a computer and sound recording we need a device driver and recoding software respectively.

6.2 Architecture

The Steered Response Power Phase Transform (SRP-PHAT) simulation are occupies to generate the white noise and the captured signals are modeled to come from the different angles and from the different sources as well [50]. The acoustic microphones are also simulated to capture an effective acoustic sound signal by using different algorithm and estimation process like Steered Response Power Phase Transform algorithm and Difference of Arrival estimation.

Data acquisition is another important part of this kind of experiment where we can use a number of devices for data acquisition in a real time reverberant biometric acoustic environment. Acoustic Speaker, Effective Microphone array model, Pre-amplifier, Sound recorder and a computer laptop or desktop is the main required device for the efficient data acquisition.

In the real acoustic experimental environment, the speaker placed in different position in the acoustic system and act as an active sound sources and this generated sound is captured by
the array of microphones. The sound recorder is recording the sound in the acoustic environment. For this reason we use the sound recording software in our personal computer for recording the sound and the speaker is directly connected with this computer and after recording the sound we can play through the computer and we get the sound through the speaker and that time speaker act as active sound source in this kind of acoustic experimental environment [44, 45]. An array comprising of four, eight or sixteen microphones according to the array size or acoustic system model are connected with the pre-amplifier, which is involves the amplifying the weak signals in the system which are ready for recording. After that this amplified good quality recorded sound is ready for the computer based simulation system for acoustic sound source localization which is done by some high level software like MATLAB.

6.3 Test Environment

Let we think an acoustic experimental room set up for data collection with the audio lab facilities where acoustic treatments can be mounted on the wall of the particular acoustic system to comprehend various kind of noise and reflective properties [24].

The dimension of the acoustic experimental room is 3.66 m for both its length and width side and the height of this room is 2.22m. Sound from the predetermined effective acoustic source location, using the measured delay of arrival between the 2 (two) microphones the average speed of sound can be estimated. In our particular acoustic sound environment let we think that the speed of sound is around 346.2 meter/second.

Figure 6.1: Ideal Reverberant Room Conditions.
For the data collection, we have to consider a acoustic microphone array design where the acoustic microphone are placed in various array model like linear array design or the square array design, consisting of four, eight or sixteen Omni-directional microphones according to the array size or acoustic system model [35, 38]. Let, in this microphone array design, the spacing between the 2 (two) microphones are 1.284m. Let each microphones are placed in the 1 meter 57 centimeter above from the ground level, so the effective height of the microphones in this system is 1.57 meter and microphones are placed in the 28 centimeter perpendicular from the cage surface [27]. We can also use the new technique like laser measuring device for to verify the actual microphone position in the system. The summarization regarding the room condition are given below,

Table 1: Summery of the Acoustic Experimental Room Setup for Data Acquisition.

<table>
<thead>
<tr>
<th>Acoustic Experimental Environment</th>
<th>Effective Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acoustic Room Condition</td>
<td></td>
</tr>
<tr>
<td>Length and Width</td>
<td>3.66 meter</td>
</tr>
<tr>
<td>Height</td>
<td>2.22 meter</td>
</tr>
<tr>
<td>Sound</td>
<td></td>
</tr>
<tr>
<td>Velocity of Sound</td>
<td>346.2 meter/second</td>
</tr>
<tr>
<td>Microphone Array Model</td>
<td></td>
</tr>
<tr>
<td>Microphone Spacing</td>
<td>1.284 meter</td>
</tr>
<tr>
<td>Source Height</td>
<td>1.57 meter</td>
</tr>
</tbody>
</table>

The acoustic sound sources which are the placed predetermined location are moved inside the fixed part of the region during the data capture time. For to get the various level of data from the experiment, we can change the acoustic environment with vary of the room reverberation levels [13, 23]. We can change the environment through the use of different kinds of material in the room’s wall to control the reverberation level. For to create the low reverberation level we can use the acoustic soundproof foam and Plexi glass is use for to create the high reverberation acoustic environment [24].

When we use the acoustic soundproof glass in the acoustic experimental room’s wall, the soundproof foam increased its absorption power that it can absorb the multipath signals in the system. This kind of soundproof foam decreased the reverberation level which one is depending on the thickness of the foam. Low frequencies signals are pass through the soundproof foam while the others signal are attenuated [26].

Again when we use the Plexi glass in the acoustic experimental room’s wall, the room walls act as very good reflector in the environment and create some worse case of multipath scenario. For to increase the reverberation condition effectively, we have to use Plexi glass in the defined acoustic system.

So from the above discussion, we know that we can switched our acoustic system from low reverberation environment to high reverberation environment with the effective change of
the room construction materials and we can get different types of data acquisition for the
different types of acoustic environment. Efficiently measured and control the acoustic
environment is one of important parts of sound source localization method [35].

Figure 6.2: Graphical Representation of Microphone Array design and the Sound Source
Position.

6.4 Test Signals Used

In our expected acoustic test environment, 2 (two) input signals used in the system to
derive the sound source speaker. Impulse response to a Butterworth filter of order 4 with 3dB
lower cutoff frequency at 400 Hz and upper cutoff frequency at 600Hz for the narrow band
signal is one of the input signals in the system and another input signal is 5600Hz which one is
for broadband signal [38]. For its maximally flat spectrum in the pass, we choose the
Butterworth impulse response and for the uniform distribution of the spectral power while the
impulse response is a causal signal with perfect phase transform, we choose the stop band signal.

From a white nose source using 3dB lower cutoff frequency at 400 Hz and upper cutoff
frequency at 600Hz for the narrow band signal and 5600Hz for broadband signal, a colored noise
signal is generated. As a test signal, colored noise is selected because off its power spectrum
covered all frequencies in the range interest [37].

In terms of the signal that is totally spread out in time and that which exists only for a
small time interval in the system, the choice of impulse and colored noise signal sources helps in
analyzing the performance level. Again, in terms of the signal that have the different spectral
characteristics, the broadband and narrowband variations help analyze the performance. All signal used in the system are generated at the sampling rate of 32 (thirty two) KHz. For to reduce the actual size of the audio data file saved in the computer hard drive, the signals are down sampled at 16 KHz in the system. The down sampling rate 16 KHz does not do any affect of the performance level because offs the bandwidth range, which one is the 300 Hz to 6 KHz.

Table 2: Summary of the Signals Used in the System to Drive the Acoustic Source

<table>
<thead>
<tr>
<th>Types of the Signal</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Narrowband Signal</td>
</tr>
<tr>
<td>Impulse Signal</td>
<td>400Hz- 600Hz</td>
</tr>
<tr>
<td>Colored Noise</td>
<td>400Hz- 600Hz</td>
</tr>
</tbody>
</table>
CHAPTER 7
IMPLEMENTATION OF SRP-PHAT ALGORITHM

7.1 Implementation of Steered Response Power Phase Transform Algorithm

In this particular section, we will discuss and demonstrate about the acoustic sound source localization system and implementation the other supporting algorithms. We also discuss the implementation of Steered Response Power Phase Transform (SRP-PHAT) through the different types of microphone geometrical array model [1, 4, 17].

Now we will discuss the Steered Response Power Phase Transform (SRP-PHAT) implementation for to localize the active acoustic sound source position by using different types of microphone array model like Linear Microphone Array model, implementation in two-dimensional speaker position with SMA, implementation in three dimensional SMA [7, 8].

7.2 Acoustic Sound Source Localization by using Linear Microphone Array

An acoustic environment speaker generated sound that travel omni directional. These sound signal propagating in space region and travel the distance and reached in the linear microphone array approaches [18]. The speaker volume is adjusted accordingly to the signal to noise ratio (SNR).

If any weak microphone is captured by any microphone among the acoustic microphone linear array model, the microphone transmitted the signal to a pre-amplifier state for to amplify the signal again and later it transmits to the acoustic sound recorder. The acoustic sound recorder is connected to the personal computer like laptop or desktop which has an Steered Response Power (SRP) algorithm [41]. In an ideal acoustic experimental environment, the Standard distance from the acoustic source to the linear microphone array model should be in 160 cm whereas the distance between the two consecutive microphones in the array should be in 4 cm.

We know that, the sound which one we use for the speaker in the system in here is called white noise. For the Linear microphone array approach, white noise should be created with a length of 2, 40,000 samples and the sampling rate should be in 8 (Eight) KHz. The White noise is used in the system as a input portion of the main function where it trigged the speaker and commanding the recording system for to start the system [44]. Both the speaker and the recording materials used to start the system at the same time.

The Steered Response Power Phase Transform (SRP-PHAT) basically involves the active acoustic sound source localization in a particular acoustic environment where mainly SRP-PHAT is related with some important major functions like Welch function, relative delay function, SRP function and they are mostly simulated by the MATLAB scripts. The active
relative delay function is involves to produce the relative time delay estimation between the different pairs of microphones in linear microphone array approach. The SRP function is used to compute the steered power for each and every microphone with their respective relative time delays. In this system, we use in the system another important function named Welch function where we use a block size of 1024 samples and an overlapping block of 512 samples.

Figure 7.1: An Acoustic Environment with the Linear Microphone Array Geometry.

The Hanning window is used by the Welch algorithm where this Hanning window is used the FFT length of 1024 (One Thousand twenty four) samples. For to maximize the steered response power (SRP), the time delay is divided into 1024 elements for linear microphone array approach [47, 48].

In this acoustic experimental environment, there are 6 (six) pairs of microphone are usually involved in acoustic system with generalized linear microphone array approach where the microphones are also used for computation of Generalized Cross Correlation (GCC) as well [33]. From the above figure 7.1, we can easily assume that the microphone pairs which are directly involve in the system are M1 and M2, M1 and M3, M1 and M4, M2 and M3, M2 and M4, M3 and M4.

Zero degree angles are considered in the system in the direction of positive x-axis and angles are increasing in the counter clock wise direction. When the acoustic source signals are arrived in the microphones which are placed in the linear microphone array approaches in the system, there are positive angles can be generated which we can defined with $\alpha$. 

45
7.3 Implementation of Two-Dimensional Speaker Position with SMA

There are two and three dimensional ways that we can implement the SMA geometry in the acoustic system. The white noise that we used in the linear microphone array approach, the same function we have to use in this particular system [45]. Again the Welch function that we used in the linear microphone array approach, the same Welch function we also used in there. There are two and three dimensional ways, Steered Response Power (SRP) can be implemented in the system as well.

In the above figure, 7.2 we can observe that the microphone array approach and the acoustic sound sources are placed in the same locations on the conference table and we observed that the both are placed in the XY plane.

From 0 to 359 degree angles are uses for the steering angle, used by the two dimensional SMA geometry. We can easily estimated the azimuth angle, because off when we implement the SMA in two dimensional acoustic speaker position environments, it is always involves for scanning the azimuth angle. From the above illustration, we observed that there are sixteen
different positions where we can get the acoustic speaker and from the each location there are ten measurements can be recorded [38].

### 7.4 Implementation of Three-Dimensional SMA

After successfully established the two dimensional acoustic speaker position, the three dimensional SMA is nothing but an extension of the two dimensional SMA which has used for active acoustic source localization [28]. After perfectly scanning both the azimuth angle and the elevation angle, we can determine the acoustic source localization for the three dimensional SMA. The degrees of angles are pre set for the azimuth and elevation angle, 0 to 359 degree and 0 to 90 degree accordingly.

Optimal non parametric Welch’s algorithm is using in there for to estimate the power spectrum. In this research, the cross correlation between the acoustic microphone pairs are estimated by the spectrum estimation, which is very important, turn are used to compute the Steered Response Power (SRP) [4, 8]. It is very natural scenario that in real life experiment that always we have to majorly consider the noise problem which is associated with the signal source. To bind the noise problem in a tolerance range, we have to use the Welch’s algorithm in the system, where the noise and acoustic signal spectra both estimated [12]. For the suppression of the noise power perturbing the signal uses the Wiener filter which can increase the performance level of acoustic source localization.

![Diagram of speaker positions in three-dimensional SMA](image)

**Figure 7.3:** Position of a speaker on different locations in three dimensional SMA systems.
The above figure, we can observed that the various acoustic source establishment in three dimensional way always involve to search the azimuth angle and the elevation angle with the computation complexity of algorithm [21]. In many different cases, we use the statistical standard deviation for to average the measurement of acoustic source localization.

7.5 Optimization

7.5.1 Stochastic Region Contraction in SRP- PHAT

Stochastic Region Contraction (SRC) is commonly known as the technique to optimize the acoustic microphone array’s pattern. By minimizing the Power Spectral Density Function (PSD) we can optimize the problem for placements and gains of microphones in an array. The nonlinear optimization technique, SRC has been shown it’s robustness in finding the global optimum [2]. Given an initial rectangular search volume where contains the desired global optimum and so many local maxima or minima in an iterative process, contract the original volume until a sufficiently small sub volume is reached in which the global optimum is trapped, is the basic idea of SRC [2, 3, 5]. A basic stochastic exploration of the SRP-PHAT usually controls the contraction operation on iteration $i$ where the surface, $P'(x)$ is functional in current sub volume. As the first step, we have to determine the number of random points, $j_0$, where we have to evaluated this point to make sure that the one or more is likely to be in the volume, $V_{peak}$, of higher values surrounding the global maximum of $P'(x)$.

Figure 7.4: Two Dimensional example of Stochastic Region Contraction (SRC). $j$ is the iteration index. The rectangular regions show the contracting search regions.
When the initial search volume is \( V_{room} \), than the probability of a random point hitting \( V_{peak} \) is,

\[
P(\text{hit } V_{peak}) = \frac{V_{peak}}{V_{room}} \tag{7.1}
\]

On the other side, we can also calculate the probability of a random point missing \( V_{peak} \) is,

\[
P(\text{miss } V_{peak}) = 1 - \frac{V_{peak}}{V_{room}} \tag{7.2}
\]

Throwing a random point is independent from one to another point. So the \( V_{peak} \), for the probability of throwing \( n \) random points missing will be,

\[
P(n - \text{misses}) = (1 - \frac{V_{peak}}{V_{room}})^n \tag{7.3}
\]

After taking the logarithm parts in both sides, we get,

\[
n = \frac{\log P(n - \text{misses})}{\log(1 - \frac{V_{peak}}{V_{room}})} \tag{7.4}
\]

How many random points \( n \) needed to throw to ensure that the missing probability, \( P(n - \text{misses}) \), we can determine from the relationship between the probability of throwing \( n \) random points missing \( V_{peak} \) and the basic ratio between the \( V_{peak} \) and \( V_{room} \). In our experiment, \( V_{room} = 400cm \times 100cm \times 600cm \ (24 \times 10^6 cm^3) \) and \( V_{peak} = 12 \times 10^4 cm^3 \) which we get from the preliminary experimental results for a low signal to noise ration situation and this one makes \( \frac{V_{peak}}{V_{room}} \approx 0.005 \).

Now we have to define the \( J_i \), as the number of random points evaluated for iterations \( N_i \), \( V_{i+1} \) is the number of points used to define the new source volume, having a rectangular boundary vector, \( \vec{B}_{i+1} = [x_{max}(i+1)x_{min}(i+1)y_{max}(i+1)y_{min}(i+1)z_{max}(i+1)z_{min}(i+1)] \), and the number of iteration \( I \), \( FE_i \) the total number of evaluated iteration \( i \), where \( \Phi \) is the maximum number of allowed to be computed [2, 5]. Now as the above description about the \( J_0 \) and \( N_i \), we can fixed the SRC algorithm like as below for to find the global maximum,

**Table 3: The SRC algorithm for finding the global maximum**

<table>
<thead>
<tr>
<th>Category</th>
<th>( i = 0, J_0 = 3000, N_i = 100, V_0 = V_{room} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluate</td>
<td>( P'(\vec{x}) ) for ( J_i ) points</td>
</tr>
<tr>
<td>Sort</td>
<td>The best $N_i \ll J_i$ points</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Contract</td>
<td>The search region to the smaller region $V_{i+1, B_{i+1}}$ that contains these $N_i$ points.</td>
</tr>
<tr>
<td>Test</td>
<td>If $V_{1+i} &lt; V_u$, or $FE_i &gt; \Phi$ and $V_{i+1} &lt; T_1 V_u$, where $T_1$ is a parameter (about 10); determine $x_{z}(i^*)$, $l = i$, Stop, Keep Result. Else if $FE_i &gt; \Phi$, Stop, Discard result. Else: Among the $N_i$ points, keep a subset $G_i$ points that have values greater that the mean, $\mu_i$ of the $N_i$ points.</td>
</tr>
<tr>
<td>Evaluate</td>
<td>$J_{i+1}$ new random points in $V_{i+1}$.</td>
</tr>
<tr>
<td>Form</td>
<td>The set of the $N_{i+1}$ as the union of $G_i$ and the best $N_{i+1} - G_i$ points from the $J_{i+1}$ just evaluated. This gives $N_{i+1}$ high points for iteration $i + 1$.</td>
</tr>
<tr>
<td>Iterate</td>
<td>$i = i + 1$. Go to Step 4.</td>
</tr>
</tbody>
</table>

Now depending on $J_i$, we can find out three variants of SRC algorithm,

1. SRC-I: Let $J_i$ is that the number of random points needed to find out $N_i - G_i$ points greater than $\mu_i$ where use the finite value of $\Phi$. Guarantees increasing of $\mu_i$.
2. SRC-II: Let $J_i$ is that the number of random points needed to find out $N_i - G_i$ points higher than $N_i$, which one is the minimum of the full set where almost all iterations $\mu_i$ increases. Here also uses the finite value of $\Phi$.
3. SRC-III: Now we have to fix the $J_i$ is equal to the $J$. For each and every iteration, keep the highest points of $N_i - G_i$. There are no guarantees of increasing $\mu_i$. And set the parameter $\Phi \rightarrow \infty$.

### 7.5.2 Coarse to Fine Region Contraction (CFRC) in SRP- PHAT

Another idea of region of contraction is Coarse to fine region contraction (CFRC), which is very much similar to the stochastic region contraction (SRC). In the current sub volume, the iteration $i$ is the based on the sub grid search of the $P'(\hat{x})$ is functional is CFRC. For a low Signal to Noise ratio case, $V_{peak}$ is determined to be $28 \text{ cm} \times 10 \text{ cm} \times 30 \text{ cm}$, from our preliminary experimental data and our $V_{room}$ is $400 \text{ cm} \times 100 \text{ cm} \times 600 \text{ cm}$. Now in $x$ axis we have to evaluate $\frac{400}{28} \approx 15$ grid points and in the same way, in $y$ axis we have to evaluate $\frac{100}{10} = 10$ grid points and in $z$ axis we have to evaluate $\frac{600}{30} = 20$, implying $J_0 = 15 \times 10 \times 20 = 3000$ equally spaced grid points in 3D to have at least a grid point in $V_{peak}$.
Figure 7.5: Two Dimensional example of Coarse to Fine Region Contraction (CFRC). $j$ is the iteration index. The rectangular regions show the contracting search regions.

We can defined $N_i, V_{i+1}, \overline{B}_{i+1}, FE_i$ and $\Phi$ for CFRC, the same way as in SRC. Naturally, by the different methods used to update $f_i$ and $V_i$ for each iteration, CFRC can be implemented in many different ways. The general algorithm is,

Table 4: The CFRC algorithm for finding the global maximum

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initialize</td>
<td>$i = 0, J_0, N_0, V_0 = V_{\text{room}}$</td>
</tr>
<tr>
<td>Evaluate</td>
<td>$P'(\hat{x})$ for $J_0$ points.</td>
</tr>
<tr>
<td>Sort</td>
<td>The best $N_1 \ll J_0$ points.</td>
</tr>
<tr>
<td>Contract</td>
<td>The search region to the smaller region $V_{i+1}, \overline{B}_{i+1}$ that contains these $N_i$ points.</td>
</tr>
</tbody>
</table>
| Test       | If $V_{i+1} < V_u$, or $FE_i < \Phi$; determine $\bar{x}_g$, **Stop, Keep Result.**  
Else if $FE_i > \Phi$, **Stop, Discard Result.**  
Else: Among the $N_i$ points, keep a subset $G_i$ points that have values greater than the mean, $\mu_i$ of the $N_i$ points. |
| Evaluate   | $J_{i+1}$ new grid points in $V_{i+1}$.                                   |
| Form       | To obtain the set of the $N_{i+1}$ as the union of $G_i$, the $J_{i+1}$ just evaluated. |
| Iterate    | $i = i + 1$. **Go to Step 4.**                                            |
For to select, $f_i$ and $N_i$ we have to chose the simplest algorithm from our data set. For to give the perfect performance level with the lowest cost, $j$ was selected, $j = 10 \times 5 \times 15 = 750$ points and again $N = 100$ turned out to preserve perfect performance at a low cost [3, 5, 10]. Now, SRC and CFRC, two efficient algorithms, to see how they improve the SRP-PHAT, we would like to see their significantly reduced computational cost as well as their correct performance relativity to a full grid search. Now, using full grid search as well as the SRC and CFRC, we have to calculate the computational cost of the SRP-PHAT.

7.6 Computational Cost

7.6.1 Signal Processing Cost

To do the Phase Transformation, Steered Response Power Phase Transformation (SRP-PHAT) requires the frequency domain processing [3]. If the number of microphones $M$ used in particular location, Phase Transformation $Q = \frac{M(M-1)}{2}$ is required by the computation of the $P'(\hat{x})$. Counting additions and multiplications as separate arithmetic operations, for a DFT size of $L$, an FFT takes $\frac{L}{2} \log_2 L$ complex multiplications and $L \log_2 L$ complex addition.

1. **DFT**- When we do an Fast Fourier Transform (FFT) for $M$ number of microphones, the total DFT cost will be $M \times 5L \log_2 L$.

2. **Spectral Processing**- We can calculate the cost for spectral processing in the following way, for each pair of microphones,
   i. Cross Power Spectrum: A Complex Multiplication for $L$ point cross power spectrum, or $(4+2) L = 6L$ operations.
   ii. Phase Transform: $L$ point over the magnitude of the cross power spectrum, which costs $L$ operations.

   According to this process, $7QL$ is the total spectral processing cost for $Q$ pairs of microphones.

3. **IDFT**- $5QL \log_2 L$ is required by the $Q$ pairs of microphones.

7.6.2 Cost per Functional Evaluation, $fe$

The following steps are required for each point of $\hat{x}(fe)$.

1. Obtain the $M$ (where $M = 24$) Euclidean distances, $d^n_j(\hat{x})$, from $\hat{x}$ to each microphone. Cost: 3 multiplications, 5 additions and 1 square root ($\approx 12 \text{ ops}$). Cost: 20ops/mic or 480ops/fe.
2. Determine $Q$ (where $Q = 276$) TDOA’s. $T_{ij}^n(\hat{x}) = (d_{ij}^n(\hat{x}) - d_{ij}^m(\hat{x}))/C$ where the inverse off the speed of sound is denoted by the $C$. Cost: 3ops/pair or 728ops/fe.

3. Now sum up the PHAT values and requires a multiplication, addition and truncation to determine TDOA value. Cost: 5ops/ pair or 1380ops/fe.

In this way, we need $(480+ 728+ 1380)$ ops/fe = 2588 ops/fe.

7.6.3 Cost of Full Grid Search

In this case, our search volume is $400cm \times 100cm \times 600cm$. For $V_u = 1cm$, we required $2.4 \times 10^7$fe’s. Therefore, the cost of full grid search is, Grid Search Cost= $24 \times 10^6$ (2588) = 62,112 mo/frame. The computational cost of full grid search is greater compare with the computational costs of signal processing and interpolation [1, 3].

7.6.4 Cost of SRC and CFRC

The same signal processing and interpolation is required for SRP-PHAT when it is using SRC or CFRC and on the other side, to find the global maximum, we need the significantly reduces the number of fe’s [2, 3].

A few small additional computations are required for both SRC and CFRC. These are determine each random point (in the particular case of SRC) which is cost about 21 ops/fe and grid point (CFRC) which is cost about 12 ops/fe [2, 3].

7.7 Experiments and Result Analysis

For our experiment, we consider the Huge Microphone Array approach which can support up to 512 microphones in real time [3]. Only the Huge Microphone Array system and experimental room with a $T_{60} = 0.45s$ that we used in our experiments have been described below

7.8 Experimental System

Approximately facing the locator microphones, A human talker repeated the first four seconds from four locations with the distances and SNR as shown in the following figure,
7.9 Preliminary Processing for SRP-PHAT Data

7.9.1 Interpolation

The problem of SRP-PHAT is it is very discontinuous near the peak point, so some interpolation is really needed in order to make the surface smoother in this particular region [3, 5, 12]. Two interpolation techniques, low pass FIR filter interpolation and cubic interpolation on the spectral samples are implemented on the SRP-PHAT surface. For to get over than 8001 interpolated points, 131 long low pass FIR filter interpolates almost 824 original samples [3]. After interpolation with the low pass FIR filter interpolation technique, the surface becomes more smoother (Figure 7.7(b)), compare to the original (Figure 7.7(a)). The cubic interpolation also gives almost the same smoothing effects (Figure 7.7(c)) as the low pass filter interpolation (Figure 7.7(b)).
7.9.2 Energy Discriminator

When we get the reliable estimation from the SRP-PHAT, using the full grid search we would like to compare the performance level of SRP-PHAT with using Stochastic Region Contraction (SRC) and Coarse to Fine Region Contraction (CFRC) [2, 3]. This kind of comparison would be proved the effectiveness of the SRC and CFRC in SRP-PHAT. Now the question is how we would say one estimation of SRP-PHAT as a reliable estimation [3].

From our observation regarding the SRP-PHAT methods, whenever the speech information are present sufficiently in the fame, Steered Response of Power Phase Transformation (SRP-PHAT) seems to make good estimates all the time [1, 3, 8]. SRP-PHAT only fails or cannot show its performance of correct estimation on frames containing a large
portion of silence data. That’s why, if we get the good frames, where the speech information is present properly, using the full grid search we can compare the performance level of SRP-PHAT using the SRC and CFRC [2, 3, 8, 10].

It is very normal scenario that the speech frames normally consumed the higher energy compare to the non speech frames and at the same time, it is very difficult to detect the low energy gathered events in highly noisy conditions [3].

Now we use a discriminator based on energy power can be used to detect the frames containing the speech or non speech information. In this particular research work, we have created the energy discriminator as in the following way,

Table 5: A Simple Energy Discriminator

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initialize</td>
<td>Energy threshold, ( \theta ) to background energy (in our case we use ( \approx -30dB ))</td>
</tr>
<tr>
<td>Determine</td>
<td>Energy of all frames in the processed data, normalizing highest frame to 0dB.</td>
</tr>
<tr>
<td>Select</td>
<td>A set of frames, ( G ), that have energy greater than ( \theta ).</td>
</tr>
<tr>
<td>Calculate</td>
<td>The ( x, y, z ) standard deviations, ( \sigma_x, \sigma_y, \sigma_z ) of the source locations estimates given by SRP-PHAT using a full grid search on ( G ) frames.</td>
</tr>
<tr>
<td>Test</td>
<td>IF ( \sigma_x, \sigma_z \leq 0.05m ) and ( \sigma_y \leq 0.1m ), STOP. Determine ( \hat{G} = G ), and ( \hat{\theta} = \theta ). KEEP ( \hat{G}, \hat{\theta} ).</td>
</tr>
<tr>
<td>Increase</td>
<td>Threshold ( \theta ) to a higher value.</td>
</tr>
<tr>
<td>Iterate</td>
<td>Go to step 2.</td>
</tr>
</tbody>
</table>

We nearly obtained \( \hat{\theta} = -7dB \), after implementing this discriminator over 4 positions with different SNR and we also use \( \theta = -7dB \) as the threshold in our energy discriminator.

For the SRP-PHAT, all frames \( G \) that have energy \( E_G \geq -7dB \) consider as a good frames and on those good frames, we only can see the comparison the performance of SRC and CFRC with the full grid search. If the implementation of SRC and CFRC achieve the global maxima everywhere, the performance level will be listed as 100% [3, 7, 11].

56
Figure 7.8: The Simple Energy Discriminator for (a) Source 4, (b) Source 3, (c) Source 2, (d) Source 1.

7.10 Result Analysis

Results are shown in the following table for the percentage of correct approaches or accuracy and the average number of functional evaluation used for grid search on all frames, CFRC and SRC on the good frames. For our evaluation, frames of 102.4 ms, advancing 25.6 ms within the speech were used for experiment and if it were either off by more than 5 cm in $x$ or $z$ or 10 cm in $y$, than the estimation was consider an error estimation [3, 7].
Table 6: Performance evaluation of SRP-PHAT using full grid search over all frames; CFRC and three SRC (SRC-I, SRC-II, SRC-III) over good frames for 4 different locations.

<table>
<thead>
<tr>
<th>Source No</th>
<th>SNR</th>
<th>ALGORITHM</th>
<th>Grid Search</th>
<th>CFRC</th>
<th>SRC-I</th>
<th>SRC-II</th>
<th>SRC-III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>% Corr 100</td>
<td>%Corr 100</td>
<td>%Corr 100</td>
<td>%Corr 100</td>
<td>%Corr 100</td>
</tr>
<tr>
<td>Source 1</td>
<td>7.9dB</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Fe’s</td>
<td></td>
<td></td>
<td>2.4×10^7</td>
<td>Fe’s</td>
<td>9123</td>
<td>Fe’s</td>
<td>7486</td>
</tr>
<tr>
<td>Source 2</td>
<td>5.7dB</td>
<td></td>
<td>96.6</td>
<td>%Corr 100</td>
<td>100</td>
<td>%Corr 100</td>
<td>99.1</td>
</tr>
<tr>
<td>Fe’s</td>
<td></td>
<td></td>
<td>2.4×10^7</td>
<td>Fe’s</td>
<td>9960</td>
<td>Fe’s</td>
<td>6531</td>
</tr>
<tr>
<td>Source 3</td>
<td>3.47dB</td>
<td></td>
<td>87.8</td>
<td>%Corr 100</td>
<td>100</td>
<td>%Corr 100</td>
<td>97.5</td>
</tr>
<tr>
<td>Fe’s</td>
<td></td>
<td></td>
<td>2.4×10^7</td>
<td>Fe’s</td>
<td>15889</td>
<td>Fe’s</td>
<td>22568</td>
</tr>
<tr>
<td>Source 4</td>
<td>1.9dB</td>
<td></td>
<td>67.3</td>
<td>%Corr 100</td>
<td>100</td>
<td>%Corr 100</td>
<td>98.4</td>
</tr>
<tr>
<td>Fe’s</td>
<td></td>
<td></td>
<td>2.4×10^7</td>
<td>Fe’s</td>
<td>19273</td>
<td>Fe’s</td>
<td>33601</td>
</tr>
</tbody>
</table>

The Visualization of the performance level using CFRC and three variants of SRC (SRC-I, SRC-II, SRC-III) are given below.

![Performance of CFRC Relative to Grid Search](image)

Figure 7.9: Performance of SRP-PHAT using CFRC relative to grid Search
Figure 7.10: Performance of SRP-PHAT using (a) SRC-I, (b) SRC-II, (c) SRC-III relative to grid Search.
CHAPTER 8

CONCLUSION AND THE FUTURE WORK

8.1 Conclusion

Especially under the higher noise conditions, real time, two stage location estimation algorithm, we have evaluated here that Steered Response Power Phase Transform (SRP-PHAT) is superior. In relative to a full grid search (From the Table 6), SRC and CFRC both shown the perfect performance level and at the same time, we can minimize the SRP-PHAT’s large computational cost by using the SRC and CFRC.

In our research, at worst case conditions ($SNR \approx 1.9dB$), using CFRC, we can able to get the perfect accuracy of SRP-PHAT with a computational advantage of 1217:1. And on the other side, at our best case ($SNR \approx 7.9dB$), CFRC gives full accuracy with a computational advantage of 2631:1, if the conditions are less noisy. Now, using SRC-I ($SNR \approx 1.9dB$) gives the computational advantage of 714:1 and SRC-I ($SNR \approx 7.9dB$) gives the computational advantage of 3206:1 to the full grid search. It is very important to see that the CFRC is less costly under the noisy conditions ($SNR \approx 1.9dB, 3.4dB$) and at the same time under the less noisy case ($SNR \approx 7.9dB, 5.7dB$) the system are more expensive. So, when the noise is comparatively high, we have to use CFRC rather use to SRC and on the other side, when the noise is low, we have to use the SRC to get better performance and reduce the computational cost.

There is lots of optimizing technique that we can be used in Steered Response Power Phase Transformation (SRP-PHAT) for to reduce the computational cost of SRP-PHAT, among all those two global optimizing technique Stochastic Region Contraction (SRC) and Coarse to Fine Region Contraction (CFRC) are more efficient and robust which can make the SRP-PHAT more practical in real time by the three orders of magnitude relative to a full grid search [2]. Over the fast, two stage method and three dimensional ways, both optimizing techniques have been shown to preserve perfectly over the performance of the Steered Response Power Phase Transformation (SRP-PHAT) for to implement the real time in our system.

In the real life conditions, the algorithmically Steered Response Power Phase Transformation (SRP-PHAT) is important for to localize the active sound source localization where Phase Transformation (PHAT) itself allowed the parametric variation conditions which directly influence to the original spectral amplitude on the final coherent power values [4, 20]. Under the different mode of operating conditions, the choose of perfect parameter is important for to get the justification of the impact of Phase Transformation which is based on the operating conditions and for to make a performance path of the Phase Transformation (PHAT) more robust over a narrower range and the broadband source [2].
8.2 Future work

For to fulfillment the overall goal of this research Acoustic Beam forming, Design and Development of Steered Response Power Phase Transformation (SRP-PHAT), the described simulated environment and the condition are almost remain same in terms of their effectiveness of the partial weighting factor, the experimental setup and the test conditions. For to get and evaluate the more comprehensive performance, different mode of setup can be investigated in terms of the following issues,

a. Number of acoustic microphones can be changed in the array model.
b. Rearrangement of the spacing of the acoustic microphones.
c. Development of the microphone array geometry.

From the different phase of investigation, Generalized Cross Correlation (GCC) with pitch based weighting system is the more robust application than any conventional Phase Transformation (PHAT) under fully noise environment and shows the same efficiency level as the Phase Transformation (PHAT) under reverberation conditions. As the result of it, under reverberation conditions, than the maximum likelihood situations, it is shows its’ robustness and perform better than the Phase Transform (PHAT) under the noise only situation. It would be interesting to investigate to see that if this kind of weighting function is applied to the Steered Response of Power (SRP) will make any kind of improvement or not.
REFERENCES


[23] Ping Zou, Zheng Huang and Jianhua Lu, “Passive Stationary Target Positioning Using Adaptive Particle Filter with TDOA and FDOA Measurements”, Joint Conference of the


