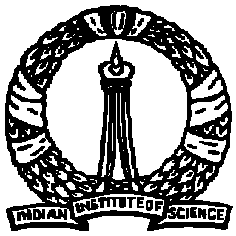
Acoustic Source Localization Using Time Delay Estimation

A Thesis
Submitted for the Degree of
Master of Science (Engineering)
in Faculty of Engineering

by

Ashok Kumar Tellakula



Supercomputer Education and Research Centre
Indian Institute of Science
Bangalore – 560 012 (INDIA)

AUGUST 2007

Dedicated to ...

My Mother Sai.

Acknowledgments

I would like to express my immense gratitude to my advisor Dr. Atanu Mohanty for his constant guidance and support. He has been inspiring and motivating me all through my research work. It has been an enriching experience working with such a knowledgeable supervisor.

I enjoy learning most through discussion and dialog, and I am grateful to those people who have been happy to answer my questions and explore ideas. Particularly to my friends Subbu, Rahul, Rupesh and Chandrika for their cooperation and wonderful company.

I wish to express my profound gratitude to my teacher, Dr. K. J. Vinoy of E.C.E department, and to the other professors who have taught me while I have been at S.E.R.C. Thanks to the chairman of the department Prof. Govindarajan. My special thanks to IISc authorities and, S.E.R.C in particular for providing an un-interrupted computing facilities. I thank Mallika, Shekhar and other staff members of S.E.R.C for their help and support.

Thanks to my friends and family members, for inspiring and motivating me, for making me who I am. I am indebted to my parents: Gurunadha Rao, Ramadevi, sister: Tulasi, brother: Madhu and grandparents for their love and care.

I offer my loving pranams and dedicate this work at the Lotus Feet of Bhagawan Sri Sathya Sai Baba.

Abstract

The angular location of an acoustic source can be estimated by measuring an acoustic direction of incidence based solely on the noise produced by the source. Methods for determining the direction of incidence based on sound intensity, the phase of cross-spectral functions, and cross-correlation functions are available. In this current work, we implement Dominant Frequency SElection (DFSE) algorithm. Direction of arrival (DOA) estimation using microphone arrays is to use the phase information present in signals from microphones that are spatially separated. DFSE uses the phase difference between the Fourier transformed signals to estimate the direction of arrival (DOA) and is implemented using a three-element 'L' shaped microphone array, linear microphone array, and planar 16-microphone array. This method is based on simply locating the maximum amplitude from each of the Fourier transformed signals and thereby deriving the source location by solving the set of non-linear least squares equations. For any pair of microphones, the surface on which the time difference of arrival (TDOA) is constant is a hyperboloid of two sheets. Acoustic source localization algorithms typically exploit this fact by grouping all microphones into pairs, estimating the TDOA of each pair, then finding the point where all associated hyperboloids most nearly intersect. We make use of both closed-form solutions and iterative techniques to solve for the source location. Acoustic source positioned in 2-dimensional plane and 3-dimensional space have been successfully located.

Contents

List of Tables

List of Figures

List of Abbreviations

DOA	Directionof Arrival
GCC	Generalised Cross Correlation
GCCPHAT	GCC Phase Transform
TDE	Time Delay Estimation
TDOA	Time Difference Of Arrival
PL	Position Location
2-D	Two dimensional
3-D	Three dimensional
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
LSE	Least Squares Estimator
MLE	Maximum Likelihood Estimator
PHAT	Phase Transform
SNR	Signal to Noise Ratio
DFSE	Dominant Frequency SElection Algorithm
AED	Adaptive Eigen-value Decomposition algorithm
PC	Personal Computer
A/D	Analog to Digital Converter
QCLS	Quadratic-Correction Least-Squares
LCLS	Linear-Correction Least-Squares
XPSD	Cross-Power Spectral Density

Chapter 1

Introduction

In this thesis we study the localization of acoustic sound source in a two-dimensional plane and three-dimensional space using a microphone array. We study microphone arrays of different configurations.

1.1 Motivation for Research

Microphone arrays have been implemented in many applications, including teleconferencing, speech recognition, position location of dominant speaker in an auditorium. Direction of arrival (DOA) estimation of acoustic signals using a set of spatially separated microphones has many practical applications in everyday life. DOA estimates from the set of microphones can be used to automatically steer cameras to the speaker in a conference room.

Direction of arrival (DOA) estimation using microphone arrays is to use the phase information present in signals picked up by sensors (microphones) that are spatially separated. When the microphones are spatially separated, the acoustic signals arrive at them with differences in time of arrival. From the known array geometry, the Direction of arrival (DOA) of the signal can be obtained from the measured time-delays. The time-delays are estimated for each pair of microphones in the array. Then the best estimate of the DOA is obtained from time-delays and geometry. Applications like Vehicle Location [1], vehicle

monitoring systems [2] use DOA techniques. Other applications that need a mention are distributed robotics [3], sensor networks [4].

Techniques such as the generalized cross correlation (GCC) method, phase transform (GCC-PHAT) are widely used for DOA estimation. The estimated time-delay for a pair of microphones is assumed to be the delay that maximizes the GCC-PHAT function for that pair. Fusing of the pair-wise time delay estimated (TDE's) is usually done in the least squares sense by solving the set of linear equations to minimize the least squared error. The simplicity of the algorithm and the fact that a closed form solution can be obtained has made TDE based methods as a choice for DOA estimation and position location of sound source using microphone arrays.

Accuracy of the DOA estimates obtained using the TDE based algorithms depends on various factors. The hardware used for data acquisition, sampling frequency, number of microphones used for data acquisition, and noise present in the signals captured, determine the accuracy of the estimates. Increase in the number of microphones increases the performance of source location estimation. Many of the conventional microphone array techniques for DOA estimates use large number of microphones typically 10 - 40 microphones. This puts a requirement for large number of data acquisition channels, which implies involvement of huge cost. Especially, applications like automatic camera steering uses bulky and huge microphone arrays. This concludes that there is a need for reducing the size and cost involved in time-delay estimation of acoustic source. It also requires multiple data acquisition channels which have to be synchronized in a centralized manner.

1.2 Acoustic Source Localization

The process of determining the location of an acoustic source relative to some reference frame is known as acoustic source localization. Acoustic source present in the near-field can be localized with knowledge of the time difference of arrivals (TDOAs) measured with pairs of microphones. The speed of sound in the medium in which the acoustic source is

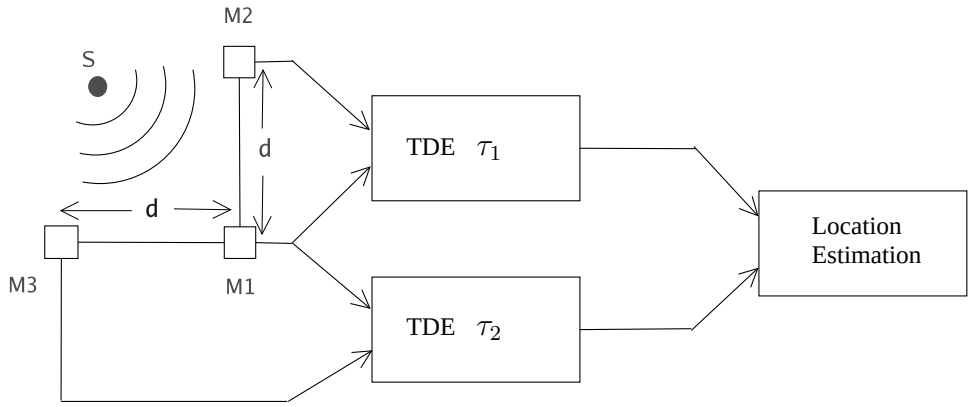


Figure 1.1: A two stage algorithm for sound source localization

present is known.

Passive acoustic source localization techniques can be used to locate and track an active talker automatically and hence are able to eliminate the human camera operators in current video-conferencing systems. The problem of passively localizing acoustic sources accurately is difficult. Microphone array processing is a rapidly emerging technique and, will play an important role in a practical solution. With continued investigation over the last two decades, the time delay estimation (TDE)-based localization has become the technique of choice [5].

Most practical acoustic source localization schemes are based on time delay of arrival estimation (TDOA) for the following reasons: such systems are conceptually simple. They are reasonably effective in moderately reverberant environments. Moreover, their low computational complexity makes them well-suited to real-time implementation with several sensors.

Time delay of arrival-based source localization is based on a two-step procedure as shown in Fig. 1.1.

1. The first stage involves estimation of the time difference of arrival (TDOA) between receivers through the use of time delay estimation techniques. The estimated

TDOA's are then transformed into range difference measurements between sensors, resulting in a set of nonlinear hyperbolic range difference equations.

2. The second stage utilizes efficient algorithms to produce an unambiguous solution to these nonlinear hyperbolic equations. The solution produced by these algorithms result in the estimated position location of the source.

The TDOA between all pairs of microphones is estimated, typically by finding the peak in a cross-correlation or generalized cross-correlation function [6], [7], [8]. For a given source location, the squared error is calculated between the estimated TDOA's and those determined from the source location [9], [10], [11]. The estimated source location then corresponds to that position which minimizes this squared error. [12] provides the summary of the state of the art on localization of an acoustic source. [13] discusses the Kalman filter based TDOA measurements.

In this thesis, the location of the target of interest is found using a two-dimensional-three-element microphone array. A pair of microphones gives the DOA w.r.t. the axis of the microphones. Since the target has two degrees of freedom, the DOA estimated would give only the direction of the source. On coupling a third microphone with previously installed microphones in an 'L' fashion, the direction of arrival (DOA) w.r.t. the axis containing the third and the center microphone can be obtained. In the second stage, the measured DOA's are used to obtain the source position by solving a system of equations.

1.3 Previous work

TDE algorithms suitable for real-time operation include the conventional generalized cross-correlation (GCC) [14], [15] and adaptive eigen-value decomposition algorithm (AED) recently proposed by Huang and Benesty [16], [17]. Compared to GCC, AED applies a more realistic signal propagation channel model and achieves better performance especially in reverberant environments [18].

If the TDOA estimates are assumed to have a Gaussian-distributed error term, it can be shown that the least-squares metric provides the maximum likelihood (ML) estimate of the speaker location. This least-squares criterion results in a nonlinear optimization problem. Several authors have proposed solving this optimization problem with standard gradient-based iterative techniques. Though these techniques typically yield accurate location estimates, they are typically computationally intensive and are ill-suited for real-time implementation [5], [19].

For any pair of microphones, the surface on which the TDOA is constant is a hyperboloid of two sheets. Acoustic source localization algorithms typically exploit this fact by grouping all microphones into pairs, estimating the TDOA of each pair, and then finding the point where all associated hyperboloids most nearly intersect. Several closed-form position estimates based on this approach have appeared in the literature. Chan and Ho [20], [21] discusses the hyperbolic position location systems in depth. They have proposed effective techniques in locating a source based on intersections of hyperbolic curves defined by the time differences of arrival (TDOA) of a signal received at a number of sensors. They gave solutions for hyperbolic position fix in closed-form, valid for both distant and close sources.

[20] illustrated a 2-D localization problem with an arbitrary array manifold. With three sensors to provide two TDOA estimates, an exact solution was obtained. With four or more sensors, the original set of TDOA equations are transformed into another set of equations which are linear in source position coordinates. In chapter 3, position location of sound source using three sensors arranged in 'L' shape and linear array have been discussed. We make use of the solutions proposed by Chan and Ho.

The closed-form position estimates are well-suited for real-time implementations. The point of intersection of two hyperboloids can change significantly based on a slight change in the eccentricity of one of the hyperboloids. Hence, a class of algorithms was developed wherein the position estimate is obtained from the intersection of several spheres. The

first algorithm in this class was proposed by Schau and Robinson [22], and later came to be known as spherical intersection [23], and spherical interpolation [24], [25], [26]. These methods provide closed-form estimates suitable for real-time implementation.

Brandstein et al. [5] proposed yet another closed-form approximation known as linear intersection. Their algorithm proceeds by first calculating a bearing line to the source for each pair of sensors. Thereafter, the point of nearest approach is calculated for each pair of bearing lines, yielding a potential source location. The final position estimate is obtained from a weighted average of these potential source locations.

In this thesis we study a three-element 'L' shaped microphone array, three-element linear microphone array, and a planar 16-microphone array. Acoustic source positioned in a 2-dimensional plane is located using three-element microphone arrays. The 16-microphone array is used to locate an acoustic source positioned in the 3-dimensional space. Areas of application include microphone arrays mounted on handheld instruments or on the body of robots, etc. Additionally applications that often require inexpensive hardware and minimal power consumption while maintaining acceptable DOA estimation performance. Nelder-Mead optimization method has been used to search for the source position in 3-dimensional space.

1.4 Assumptions and Limitations

We assume the following conditions under which location of sound source is estimated:

1. Single sound source, infinitesimally small, omni directional source.
2. Reflections from the bottom of the plane and from the surrounding objects are negligible.
3. No disturbing noise sources contributing to the sound field.
4. The noise source to be located, is assumed to be stationary during the data acquisition period.

5. Microphones are assumed to be both phase and amplitude matched and without self-noise.

6. The change in sound velocity due to change in pressure and temperature are neglected. The velocity of sound in air is taken as 330 m/sec.

7. Knowledge of positions of acoustic receivers and perfect alignment of the receivers as prescribed by processing techniques.

8. The task of locating the sound source is more difficult due to number of complications, some of which are:

    - Finite source dimensions;
    - Reflections from the bottom of the surface and nearby objects.

Perfect solutions are not possible, since the accuracy depends on the following factors:

- Geometry of microphone and source.
- Accuracy of the microphone setup.
- Uncertainties in the location of the microphones.
- Lack of synchronization of the microphones.
- Inexact propagation delays.
- Bandwidth of the emitted pulses.
- Presence of noise sources.
- Numerical round off errors.

1.5 Thesis Organization

The thesis is organized as follows. In chapter 2, we formulate the problem and it initially discusses the array structure and its restrictions. The distance between any pair of microphones in the array should not exceed half the minimum wavelength present in the signal. It then describes the proposed Dominant Frequency SElection (DFSE) algorithm which is used for DOA estimation of the sound source. It then compares with the conventional algorithms like GCC and GCC-PHAT for DOA estimation In chapter 3, the conventional closed form techniques that are used to find the approximate solution, have been discussed. We discuss the linear and 'L'-shaped microphone array structures. It introduces the models for the position location problem and the techniques involved in the hyperbolic position location. In chapter 4, we discuss a method of estimating the source position, using measurements across a number of sensors. By measuring the delay between arrival of sounds for each microphone pair, we can solve for the source location. A 16-microphone array that has been implemented, is described in this chapter. The chapter 5 gives the results from the linear, 'L'-shaped and 16-microphone array. The chapter 6 concludes with the future work and the issues involved in designing a practical system.

Chapter 2

Direction of Arrival Estimation

In this chapter we discuss the estimation of Time Direction Of Arrival (TDOA) of the sound source. Firstly we present the problem statement and then discuss the structure of the microphone array and the conventions used in this thesis. The angular location of a sound source can be estimated by measuring an acoustic direction of incidence based solely on the noise produced by the sound source. Methods for determining the direction of incidence based on sound intensity, the phase of cross-spectral functions, and cross-correlation functions have been studied in literature [14], [6], [7], [27].

2.1 Introduction

The existing strategies of source localization are of two types. One is Time Delay Estimation (TDE) based (also called as indirect methods) and the other is direct methods where the source location is computed in one stage. We have implemented the TDE based method, since they are easily realizable in real-time, unlike the direct methods which are computationally intensive. The initial experiments for DOA estimation were performed using a measurement system that is constructed by the three microphones arranged in 'L' shape. We measure the sound source location using the phase difference of a dominant frequency through three microphones. Normally the frequency components have some fluctuations. Therefore, in DFSE, we reduce the influence of the fluctuations by selecting

the maximum components of the Fast Fourier Transforms (FFT) results of the bandpass filter outputs.

2.2 Problem Statement

Given a set of M acoustic sensors (microphones) in known locations, our goal is to estimate two or three dimensional coordinates of the acoustic sound source. We assume that source is present in a defined coordinate system. We know the number of sensors present and that single sound source is present in the system. The sound source is excited using a broad band signal with defined bandwidth and the signal is captured by each of the acoustic sensors. The Time Difference Of Arrival (TDOA) is estimated from the captured audio signals.

The TDOA for a given pair of microphones and speaker is defined as the difference in the time taken by the acoustic signal to travel from the speaker to the microphones. We assume that the signal emitted from the speaker does not interfere with the noise sources. The measurements from Dominant Frequency SElection (DFSE) algorithm constitute our observations. Based on the measurements we estimate the source position knowing the geometry of the microphone array.

2.3 Time Difference Of Arrival (TDOA)

Let m_i for $i \in [1, M]$ be the three dimensional vectors representing the spatial coordinates of the i^{th} microphone and 's' as the spatial coordinates of sound source. We excite the source 's' and measure the time difference of arrivals. Letting 'c' as the speed of sound in the acoustical medium (air) and $\|\|$ is the Euclidean norm. The TDOA for a given pair of microphones and the source is defined as the time difference between the signals received by the two microphones. Let $TDOA_{ij}$ be the TDOA between the i^{th} and

j^{th} microphone when the source 's' is excited. It is given by

$$TDOA_{ij} = \frac{(\|m_i - s\| - \|m_j - s\|)}{c} \tag{2.1}$$

TDOA's are then converted to time delay estimations (TDE's) and path differences. This is depicted in Fig. 2.1

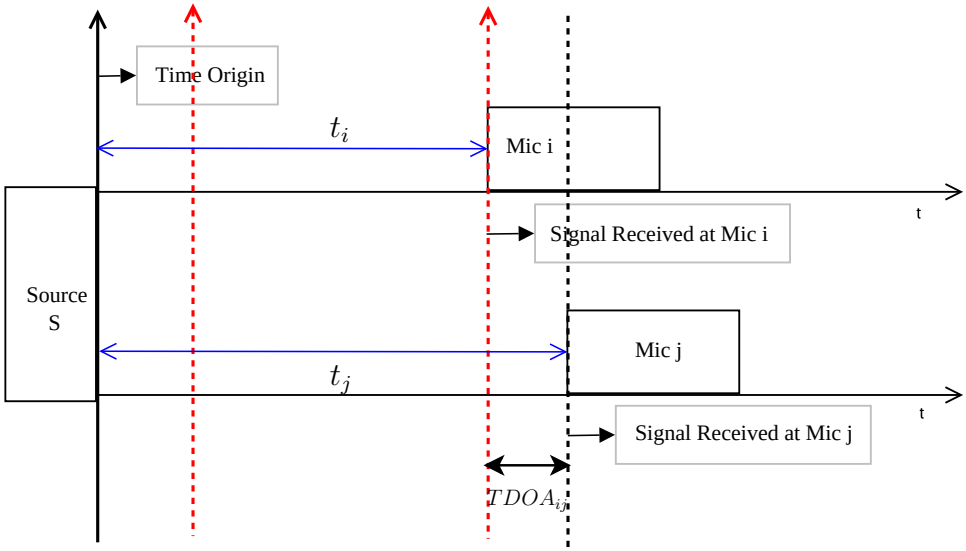


Figure 2.1: Schematic depicting the signal receiving time and TDOA

2.4 Microphone array structure

Preliminary experiments were done using a three-element-two-dimensional microphone array for Direction of Arrival (DOA) estimation. The array consists of three microphones arranged in an 'L' fashion in a 2-dimensional plane. As shown in the Fig. 2.2 the microphones M_3-M_1-M_2 form the array with M_1 being the center microphone. M_1 is at the origin of the coordinate axis, M_1-M_2 form the x-axis, M_1-M_3 the y-axis. The angle of arrival θ_1 is measured in clockwise direction w.r.t the line perpendicular to M_1-M_2 axis and passing through the mid point of the axis. θ_2 is measured in counter clockwise direction w.r.t the line perpendicular to M_1-M_3 axis and passing through the mid point of the axis. This convention is chosen for experimental convenience.

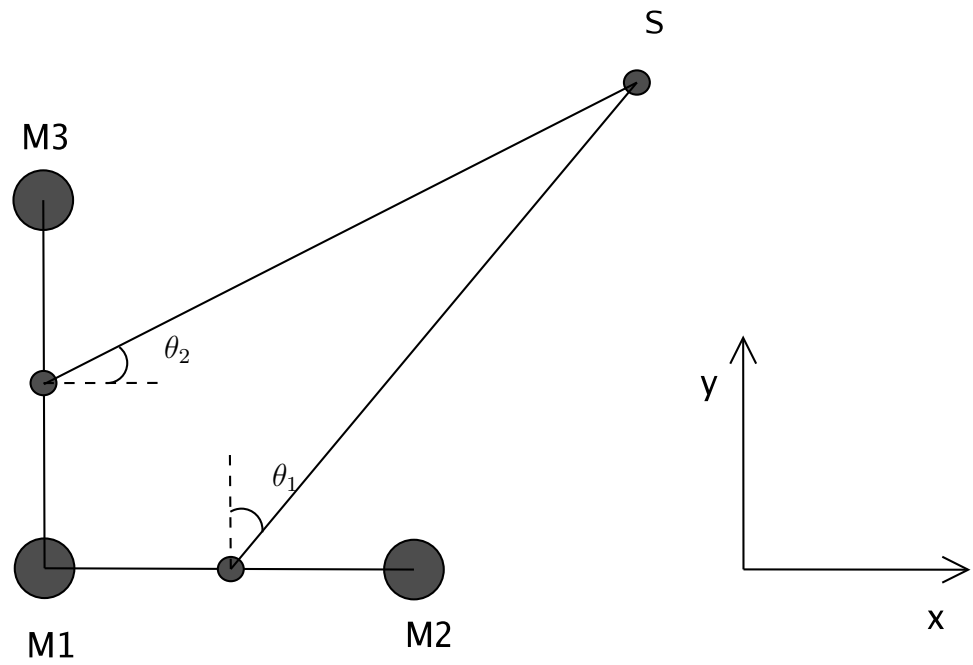


Figure 2.2: Three-element-two-dimensional microphone array

The signal from the source reaches the microphones at different times. This is because the sound wave has to travel different paths to reach the different microphones. Consider the microphone pair M_1-M_2. These microphones form a part of 'L' shaped microphone array with a distance d between adjacent microphones. Also shown in Fig. 2.2 is a sound source S at an angle θ_1 with respect to line perpendicular to x-axis. The extra distance traveled by source signal to reach M_1 as compared to M_2 is $d \sin\theta$. We assume that the microphones are omni directional, which means that the gain of the microphones doesn't change the direction of the acoustic wavefront. We don't have a front-back ambiguity as we are aware that the source is present only in the first quadrant.

2.4.1 Restrictions on the array geometry

In this section we derive a relationship between the frequency content of the incident signal and the maximum allowed separation between each pair of microphones in the array. The maximum phase difference is restricted to $|\pi|$. Any phase difference out of the range of $-\pi$ and π is wrapped around to within this range. This places an important

restriction on the array geometry when performing DOA estimation.

Consider a signal incident on the pair of microphones as shown in Fig. 2.2 at an angle θ^o. Let the broad band signal have a maximum frequency of f_{max}. At f_{max}, if we restrict the phase difference between signals of pair of microphones to be less than or equal to π, then we require

$$2\pi f_{max}\tau \leq \pi \tag{2.2}$$

and

$$\tau = \frac{d\sin\theta}{\nu} \tag{2.3}$$

where

τ = signal time delay between the two microphones,
d = distance between the pair of microphones,
θ = incident angle,
ν = velocity of sound.

Rearranging these terms, we have

$$d \leq \frac{1}{2}\left(\frac{\nu}{f_{max}}\right)\frac{1}{\sin\theta} \tag{2.4}$$

For $\sin\theta|_{max} = 1$ and since $\frac{\nu}{f_{max}}$ is same as λ_{min}, the minimum wave length present in the signal.

$$\Rightarrow \qquad d \leq \frac{\lambda_{min}}{2} = \frac{\nu}{2f_{max}} \tag{2.5}$$

which means that the distance between any pair of microphones in the array should not exceed half the minimum wavelength present in the signal. This condition becomes very important when we perform TDE from phase difference estimates of the signals.

2.5 Conventional DOA methods

There are various techniques that can be used to compute pair-wise time delays, such as the generalized cross correlation (GCC) method. The phase transform (PHAT) is the most commonly used pre-filter for the GCC. Computation of the time delay between signals from any pair of microphones can be performed by first computing the cross-correlation function of the two signals. The lag at which the cross-correlation function has its maximum is taken as the time delay between the two signals.

Consider any two microphones, i and j. Let the signals on these two microphones be $x_i(n)$ and $x_j(n)$ where n is a time-sample index. Let DFT samples of these signals be represented by $X_i(k)$ and $X_j(k)$ where k is a frequency-sample index. The cross-power spectral density (XPSD) between these signals is given by

$$\Phi_{x_i x_j}(k) = X_i(k) X_j^*(k) \tag{2.6}$$

The cross-correlation between these signals is given by the inverse DFT of the XPSD.

$$R_{x_i x_j}(l) = \frac{1}{M} \sum_{k=0}^{M-1} \Phi_{x_i x_j}(k) e^{j\frac{2\pi k l}{M}} \tag{2.7}$$

Here M is the length of the XPSD and l is a lag. $R_{x_i x_j}(l)$ can be computed for the range of possible negative and positive lags. The lag at which $R_{x_i x_j}(l)$ maximizes is the number of samples of delay between the two signals. Hence the time delay is shown by

$$\tau_{ij} = \frac{1}{F_s} \arg\max\left(R_{x_i x_j}(l)\right) \tag{2.8}$$

The DOA of the sound signal can be computed from the measured time delay.

2.6 Description of Dominant Frequency Selection Algorithm (DFSE)

The dominant frequencies present in the broad band signals can be picked and separate DOA estimates can be made on each frequency to arrive at separate estimates. The mean of these independent estimates can be used as a good estimate of the DOA. Normally the frequency components have some fluctuations. Therefore, in DFSE, we reduce the influence of the fluctuations by selecting the maximum components of the Fast Fourier Transforms (FFT) results of the bandpass filter outputs. Experiments were done using a three-element-two-dimensional microphone array with inter-microphone spacing of 10 cm. The broad band signal was incident to a pair of microphones at θ^o. The threshold used to pick dominant frequencies was 25 dB, which picked varied number of frequencies depending on duration of acquisition.

To implement the DOA estimation algorithm, band pass filters were used to extract signals. Here a new approach was used in picking the frequencies from the signals. The signal from the center microphone (M_1) was pre-amplified. Then a band pass filter was introduced that extracts signals in the frequency range of $400 - 1700\,Hz$. The frequencies below $400\,Hz$ were filtered out to avoid the low frequency noise. The frequencies above $1700\,Hz$ were cutoff as they would not meet the required condition.

$$d \leq \frac{1}{2}\left(\frac{\nu}{f_{max}}\right) \tag{2.9}$$

We recall that the spacing between pair of microphones was chosen as 10 cm. With $\nu = 330\,m/sec$, we get $f_{max} = 1715\,Hz$. So, the maximum allowed frequency f_{max} was $1715\,Hz$. The above mentioned reasons justify the choice of the frequency range.

In the experimental setup we meet a shortcoming on the number of data acquisition channels. In order to meet the demand for more number of channels, initially, signals were added in time domain and were demultiplexed in frequency domain. Later on, we made

use of analog multiplexers to accommodate multiple signals. PC sound card was used to acquire the signals from the microphones. The sound card contains two input channels: left and right that can sample the signals independently. But the array consists of three microphone elements.

The algorithm implemented meets this shortcoming. It assumes that the sound source emits a broad band signal containing formant frequencies in the frequency range $400 - 1700\,Hz$ and call this range as Δf. The signal from the center microphone was band pass filtered in the above mentioned frequency range and fed to one channel. M_2 was band pass filtered in the frequency range $400 - 1000\,Hz$. Lets call this frequency range as Δf_1. M_3 was band pass filtered in the frequency range $1100 - 1700\,Hz$ and call this range as Δf_2. That is, Δf was split into Δf_1 and Δf_2. The band pass filtered signals from M_2 and M_3 were summed and fed to the second channel. The overlap between the two frequency bands was kept minimal.

2.7 DOA using Dominant Frequency Selection (DFSE) algorithm

In this section we propose a simple sound source direction estimation method that uses the dominant frequency components of the received signal. The measurement system is constructed by three microphones arranged in 'L' fashion. We use three band pass filters to extract the signals in the desired frequency range. A personal computer is used to capture the filtered signals, then for algorithm implementation for source location estimation.

Using a pair of microphones, we detect the phase difference of the dominant frequency signal component between the microphones. We use this phase difference to calculate the direction of arrival (DOA) with respect to the microphone pair. Then we couple another microphone with the existing pair of microphones to calculate the direction of arrival with respect to the new pair. We can estimate the sound source location using these angles assuming source is in two-dimensional plane.

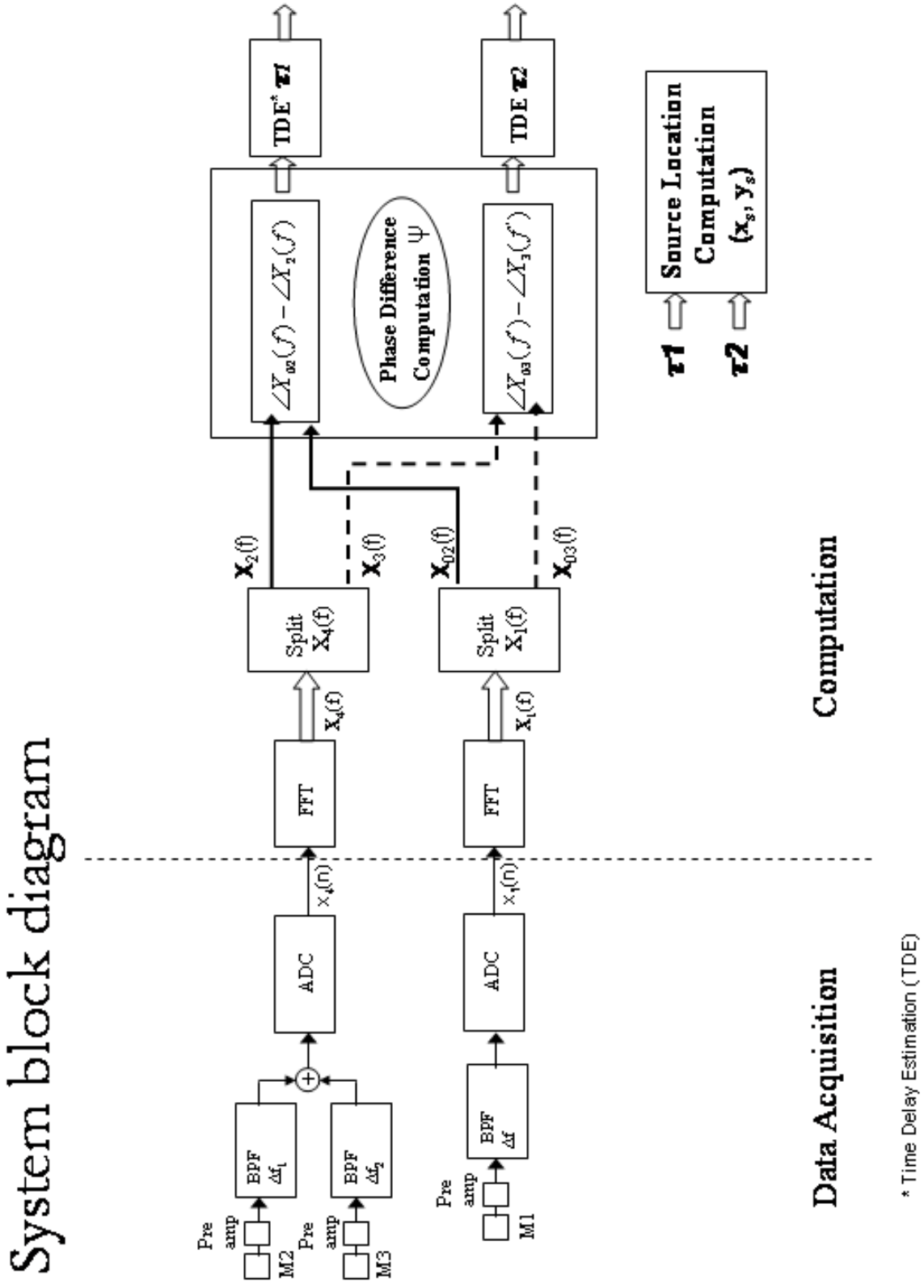


Figure 2.3: System block diagram of DFSE algorithm implemented for three element 'L' shaped microphone array

A system block diagram of the proposed method is shown in the Fig. 2.3. The output signals $x_1(n)$, $x_2(n)$ and $x_3(n)$ of the microphones M_1, M_2 and M_3 are band pass filtered. Signal x_2 from M_2 is band pass filtered in frequency range Δf_1 and the signal x_3 from M_3 is band pass filtered in the frequency range Δf_2. Denote these band pass filtered signals as bx_2 and bx_3. These two signals were summed and fed to one channel of the PC sound card. That is $x_4(n) = bx_2(n) + bx_3(n)$. Signal $x_1(n)$ from M_1 is band pass filtered in the range Δf. We keep Δf as sum of Δf_1 and Δf_2. This signal is captured from the second channel of the PC sound card. The captured signals were post-processed inside the PC.

Fourier transforms of the signals can be represented as

$$x_1(n) \longleftrightarrow X_1(f) = |X_1(f)|e^{jarg[X_1(f)]} \tag{2.10}$$

and

$$x_4(n) \longleftrightarrow X_4(f) = |X_4(f)|e^{jarg[X_4(f)]} \tag{2.11}$$

In the frequency domain $X_4(f)$ is demultiplexed to obtain $X_2(f)$ and $X_3(f)$.

Also, $X_1(f)$ is split into two regions $X_{02}(f)$ and $X_{03}(f)$ in the frequency ranges Δf_1 and Δf_2. Time delay estimation is computed from pairs $X_2(f)$, $X_{02}(f)$ and $X_3(f)$, $X_{03}(f)$. Signals $x_1(n)$, $x_2(n)$ and $x_3(n)$ are all real in time domain. In the frequency domain $X_2(f)$, $X_{02}(f)$ and $X_3(f)$, $X_{03}(f)$ are all complex.

From the amplitude spectrum $|X_2(f)|$ and the phase spectrum $\angle X_2(f)$, the maximum amplitude frequency component $|X_2(f_m)|$ is determined.

$$|X_2(f_m)| = max\{|X_2(f_1)|, ..., |X_2(f_{N_1})|\} \tag{2.12}$$

where f_1, $..., f_{N_1}$ belong to frequency range Δf_1. Similarly, the maximum amplitude frequency component $|X_3(f_m)|$ is determined as

$$|X_3\left(f'_m\right)| = max\{|X_3\left(f'_1\right)|, ..., |X_3\left(f'_{N_1}\right)|\} \tag{2.13}$$

where $f_1^{'}$, ..., $f_{N_1}^{'}$ belong to frequency range Δf_2.

These signals have significant power over a wide range of frequencies. So, it makes sense to use these dominant frequencies to perform DOA estimation. We set a threshold power and pick up the frequency components that are above the threshold. For the experiment we set the threshold at 25 dB so that all the frequency components that are more than 25 dB below the power of the maximum frequency component are rejected. DOA estimations vary significantly with the threshold used to pick dominant frequencies, as the threshold determines the number of frequencies that can be selected. A peak-picking algorithm was run on these frequencies to pick up the dominant frequencies. Narrow band DOA estimation is performed at each of these frequencies. The mean of these estimates form a good approximation to the true DOA.

We pick up the corresponding frequency components from the spectra's of $X_{02}(f)$ and $X_{03}(f)$. Then we compute the phase difference. From the phase spectra's $\angle X_2(f)$, $\angle X_{02}(f)$ and $\angle X_3\left(f^{'}\right)$, $\angle X_{03}\left(f^{'}\right)$ the phase differences between the microphone pairs are calculated by

$$\Psi_1 = \angle X_2(f) - \angle X_{02}(f) \tag{2.14a}$$

$$\Psi_2 = \angle X_3\left(f^{'}\right) - \angle X_{03}\left(f^{'}\right) \tag{2.14b}$$

The calculated phase differences between the signals can be used to obtain the time difference of arrival (TDOA). TDOA is nothing but the time delay involved for the signal to reach the pair of microphones. Time difference of arrival can be calculated by

$$\tau = \frac{\Psi}{2\pi f} \tag{2.15}$$

where Ψ is phase difference between the microphones and f is the corresponding frequency of the narrow band sinusoid.

2.8 Experimental details

In this section we measure the direction of the sound source using a pair of microphones. We use the entire frequency range Δf to identify the dominant frequencies. The signals at the microphones are sampled at 44.1 kHz. The signal frame size is varied from 0.1 *sec* to 2 *sec*. The error in DOA estimation reduces with increase in frame size of signal acquired (duration of acquisition). The FFT length of the signals is kept equal to length of the signal in order to avoid the zero-padding of the signal or truncation of the signal. As the FFT length of the signal increases we get more number of dominant frequencies. Increase in number of frequency components improves the DOA estimation. The Fig. 2.4 shows the performance of the dominant frequency selection algorithm for various signal frame sizes.

Dominant Frequency Selection (DFSE) algorithm is proved by comparing the DOA estimates with that of Generalized Cross Correlation (GCC) and Phase Transform (GCC-PHAT) methods. GCC and GCC-PHAT are well known techniques used to find the direction of a source. GCC works well for a broad band signals whereas GCC-PHAT gives good estimates even in noisy environments. The Figs. 5.1, 5.2 compares the DOA estimations of GCC, GCC-PHAT and DFSE. DFSE showed discrepancies at larger angles.

From the literature we find that GCC-PHAT performs better in environments with reverberation. GCC-PHAT is generally performed using longer signal frames since for short-time frames performance degrades as the signal statistics fluctuate. A more recent approach is the so called Adaptive Eigenvalue Decomposition (AED) algorithm. The AED algorithm performs better than the GCC-PHAT algorithm in highly reverberant environments. The results show that the proposed method DFSE shows better results when the sound source is relatively close to both the microphones those are under consideration. That is, when the sound source is near to the perpendicular line to the axis and passing through the mid point of the axis. When it is away from the perpendicular line there seem to be discrepancies in the estimated time-delays. The estimated directions of

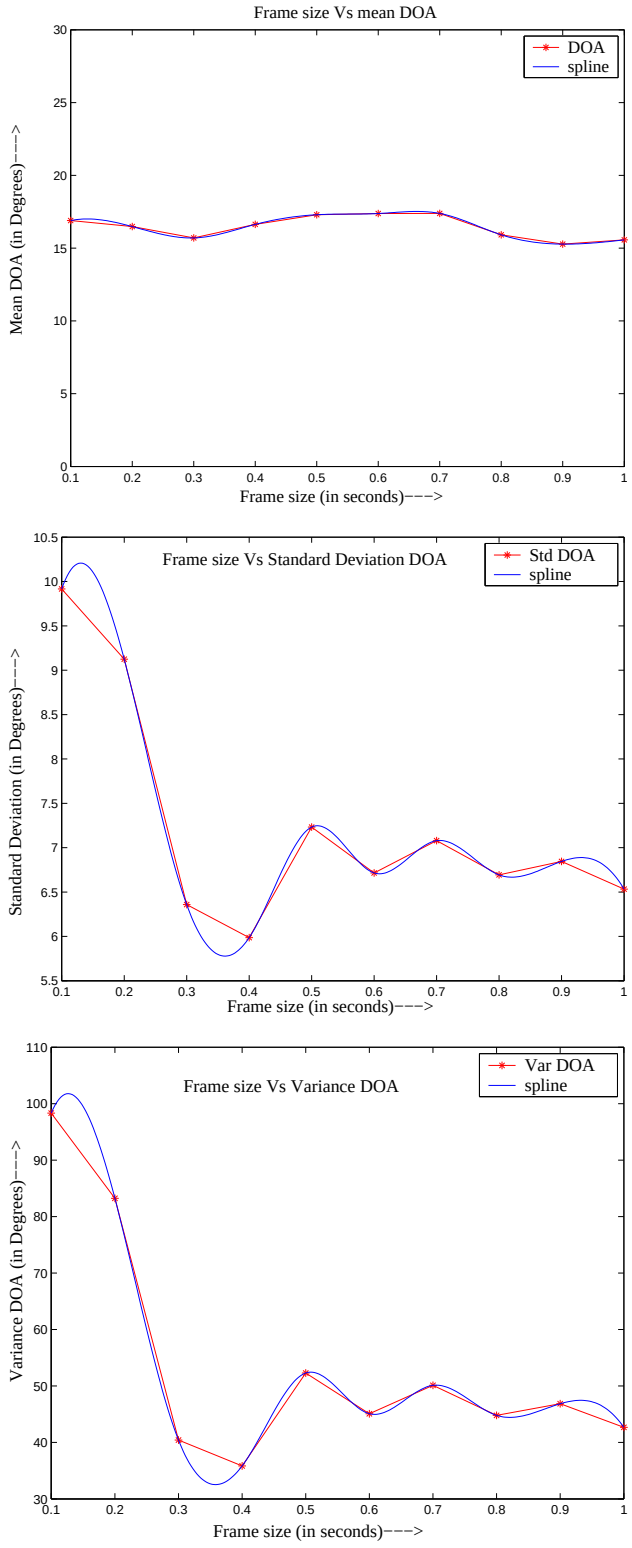


Figure 2.4: Plot showing the decreasing error in the DOA estimation with increase in signal frame size. (a) Mean DOA, (b) Standard Deviation (error) and (c) Variance in DOA.

the moving source by the proposed DFSE method are closer to the actual values that are obtained by the GCC method. This can be easily seen in Figs. 5.1, 5.2.

As described in the previous section, we set a threshold power and pick up the frequency components that are above the threshold. We run a peak picking algorithm to identify the frequencies that are above the cutoff threshold. A lower cutoff would lead to selection of more number of frequency components. This would increase the number of DOA estimates leading to improvement in the estimate. The mean estimate depends on the signal frame size and the threshold cutoff for frequency selection. The Fig. 2.5 shows the DFSE performance.

Discrepancies in DOA estimations for higher threshold cutoffs are observed. A higher cutoff would lead to selection of less number of frequency components. This would decrease the number of DOA estimates leading to decrease in the measurement accuracy. So, at higher cutoffs like -12 dB very less number of frequency components are selected. This would increase the error in the mean measured estimate. This would explain the discrepancy in DFSE at threshold cutoff of -12 dB in Fig. 2.5.

Fig. 2.4 shows that the error in DOA estimation decreases with increase in signal frame size. Fig. 2.5 shows plot with threshold cutoff Vs Mean DOA angle (in Degrees). Each of the plots compares DOA estimation when the source is kept constant. Results in the plots are when the source is kept at 10 degrees, 30 degrees and 40 degrees. Estimates from DFSE with frame sizes of 0.1 sec, 0.5 sec, 1 sec are compared with estimates from GCC and GCC-PHAT keeping the source constant. The discrepancies at -12 dB are explained in experimental details section. Figs. 2.6 - 2.7 shows DOA estimation of fixed source with varying threshold cutoff for frame size of 0.1 sec and 0.5 sec separately. Figs. 2.8 - 2.9 shows DOA estimation using selected frequencies for threshold cutoff of 25 db and 15 db. This plot emphasizes that when the cutoff threshold is less, more number of estimates would be obtained and better DOA estimate could be obtained. For higher threshold cutoffs less number of estimates would be obtained. These estimates are combined in a

least square sense to obtain the DOA.

Least Square Fitting

Time delays can be computed in a manner for all the possible microphone pairs. These time delays can be combined in a least squares sense to obtain the DOA. Let $\boldsymbol{t}$ be a $N \times 1$ vector that contains the time delays for all the microphone pairs. The distance-time relationship is given by

$$\boldsymbol{d} \sin\theta = \nu \boldsymbol{t} \tag{2.16}$$

where $\boldsymbol{d}$ is a $N \times 1$ vector that contains the distances between each pair of microphones. This equation (2.16) represents N different equations that can be solved individually to obtain DOA estimates. Experiments reveal that in an over-determined system of equations produce accurate DOA estimates. This can be solved to obtain a least squares solution.

$$\theta = \sin^{-1}\left[\left(\boldsymbol{d}^T\boldsymbol{d}\right)^{-1}\boldsymbol{d}^T\left(\nu\boldsymbol{t}\right)\right] \tag{2.17}$$

We make use of θ, as obtained above, to compute the source location.

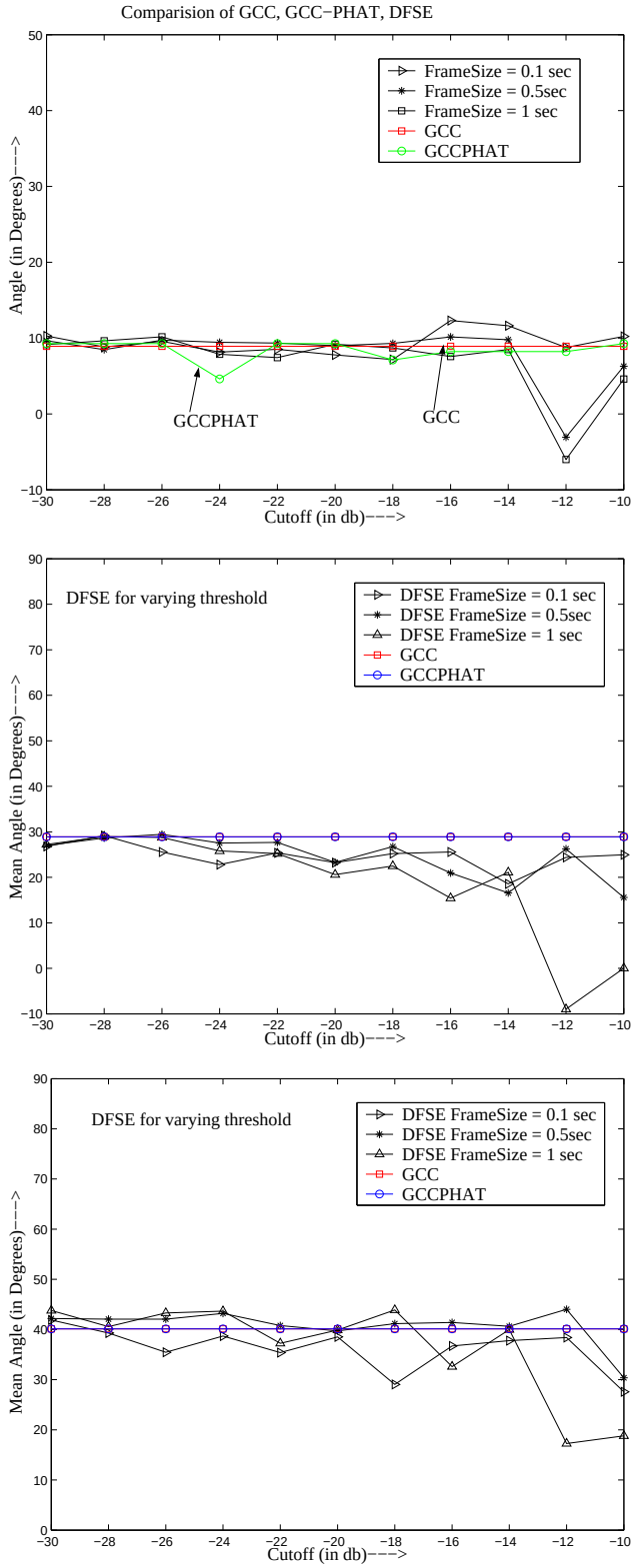


Figure 2.5: Plot showing the improvement in DOA estimate with decreasing threshold cutoff for various signal frame sizes when source is at 10 degrees, 30 degrees and 40 degrees.

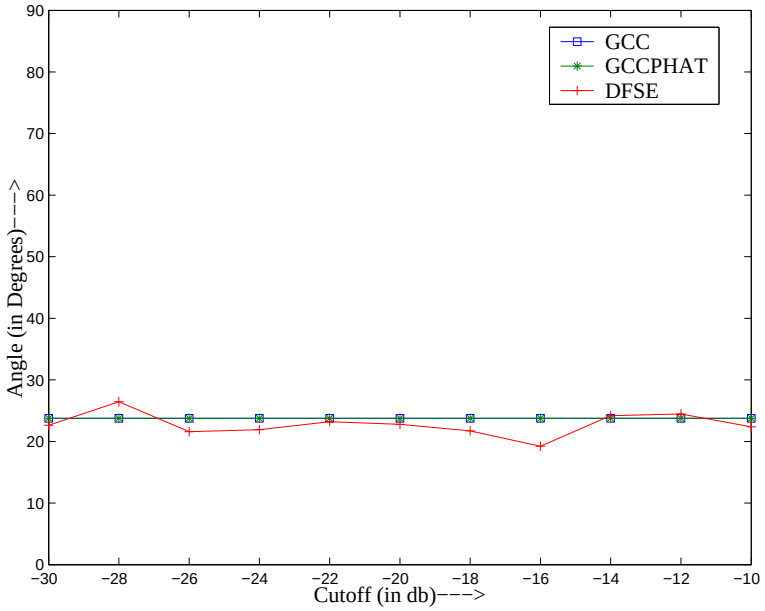


Figure 2.6: DOA estimation of fixed sound source with varying threshold cutoff for 0.1 *sec* frame size

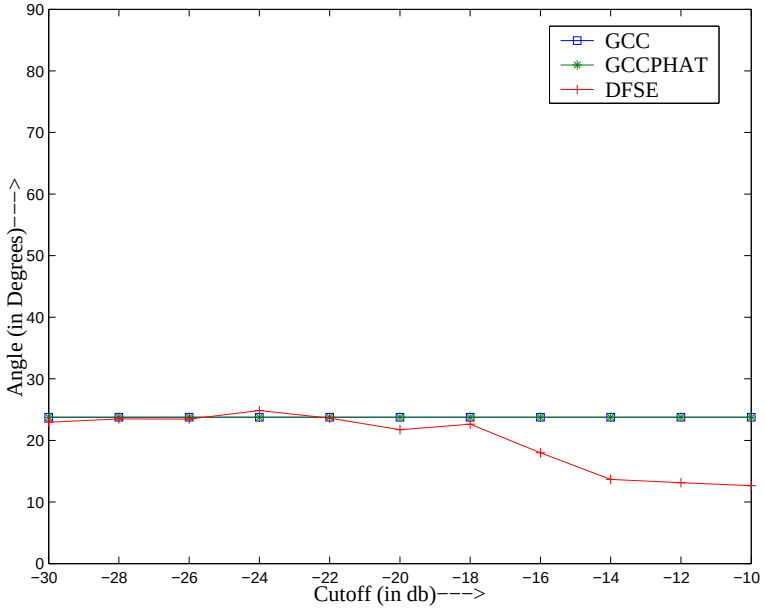


Figure 2.7: DOA estimation of fixed sound source with varying threshold cutoff for 0.5 *sec* frame size

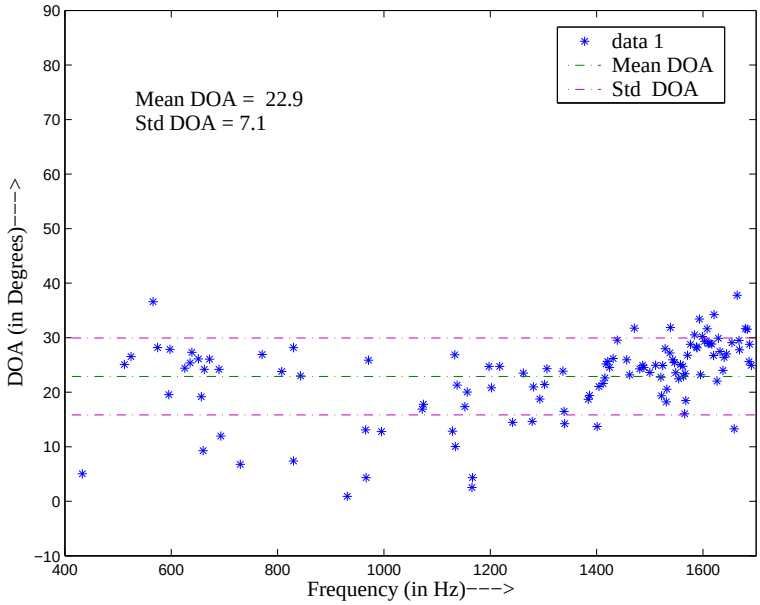


Figure 2.8: DOA estimate using selected frequencies for threshold cutoff of 25 *db*

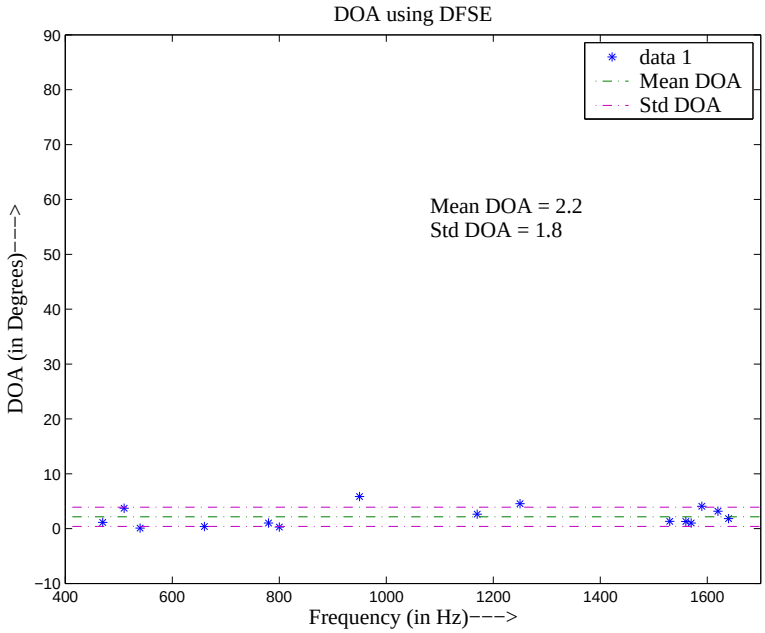


Figure 2.9: DOA estimate using selected frequencies for threshold cutoff of 15 *db*

Chapter 3

Source Localization In 2-Dimensional Space

3.1 Introduction

In this chapter we discuss the acoustic sound source localization in a plane. The TDOA method can be summarized as follows. We assume that the sound source is a point source and the microphones we use have omni-directional receiving pattern (although realistic modeling can be done). So there is a delay in the sound wave received by a pair of microphones. Sound source localization is a two step problem.

- First the signal received by several microphones is processed to obtain information about the time-delay between pairs of microphones. We use the DFSE method for estimating the time-delay which is based on dominant frequencies of received signals.
- The estimated time-delays for pairs of microphones can be used for getting the location of the sound source.

The object localization requires three microphones in a 2-D plane since two coupled microphones can give only one information *i.e.,* direction of source. The acoustic waves

generated by the source reaches the first microphone earlier than the second. The difference in the propagation delay and that the acoustic velocity in air is known, we calculate the path difference of the acoustic waves.

By definition, a hyperbola is the set of all points in the plane whose location is characterized by the fact that the difference of their distance to two fixed points is a constant. The two fixed points are called the foci [28], [29]. In our case the foci are the microphones. Each hyperbola consists of two branches. The emitter is located on one of the branches. The line segment which connects the two foci intersects the hyperbola in two points, called the vertices. The line segment which ends at these vertices is called the transverse axis and the midpoint of this line is called the center of the hyperbola.

We say that the standard equation of a hyperbola centered at the origin is given by:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \tag{3.1}$$

if the transverse axis is horizontal. The vertices are always a unit from the center of the hyperbola, and the distance c of the foci from the center of the hyperbola can be determined by using a, b, and the following equality: $b^2 = c^2 - a^2$

The time-delay of the sound arrival gives us the path difference that defines a hyperbola on one branch of which the emitter must be located. At this point, we have infinity of solutions since we have single information for a problem that has two degrees of freedom. We need to have a third microphone, when coupled with one of the previously installed microphones, it gives a second hyperbola. The intersection of one branch of each hyperbola gives one or two solutions with at most of four solutions being possible. Since we know the sign of the angle of arrivals, we can remove the ambiguity.

3.2 Hyperbolic position location

This chapter discusses the model of the hyperbolic position location system. Hyperbolic position location (PL) estimation is accomplished in two stages. The first stage involves estimation of the time difference of arrival (TDOA) between the sensors (microphones) through the use of time-delay estimation techniques. The estimated TDOAs are then utilized to make range difference measurements. This would result in a set of nonlinear hyperbolic range difference equations. The second stage is to implement an efficient algorithm to produce an unambiguous solution to these nonlinear hyperbolic equations. The solution produced by these algorithms result in the estimated position location of the source. The following sections discuss the techniques used to perform hyperbolic position location of the sound source.

Accurate position location (PL) estimation of a source requires an efficient hyperbolic position location estimation algorithm. Once the TDOA information has been measured, the hyperbolic position location algorithm will be responsible for producing an accurate and unambiguous solution to the position location problem. Algorithms with different complexity and restrictions have been proposed for position location estimation based on TDOA estimates.

When the microphones are arranged in non-collinear fashion, the position location of a sound source is determined from the intersection of hyperbolic curves produced from the TDOA estimates. The set of equations that describe these hyperbolic curves are non-linear and are not easily solvable. If the number of nonlinear hyperbolic equations equals the number of unknown coordinates of the source, then the system is consistent and a unique solution can be determined from iterative techniques. For an inconsistent system, the problem of solving for the position location of the sound source becomes more difficult due to non-existence of a unique solution. Refer appendix A.1 for the general model for two-dimensional (2-D) position location estimation of a source using three microphones. The results from this derivation are used in the further sections.

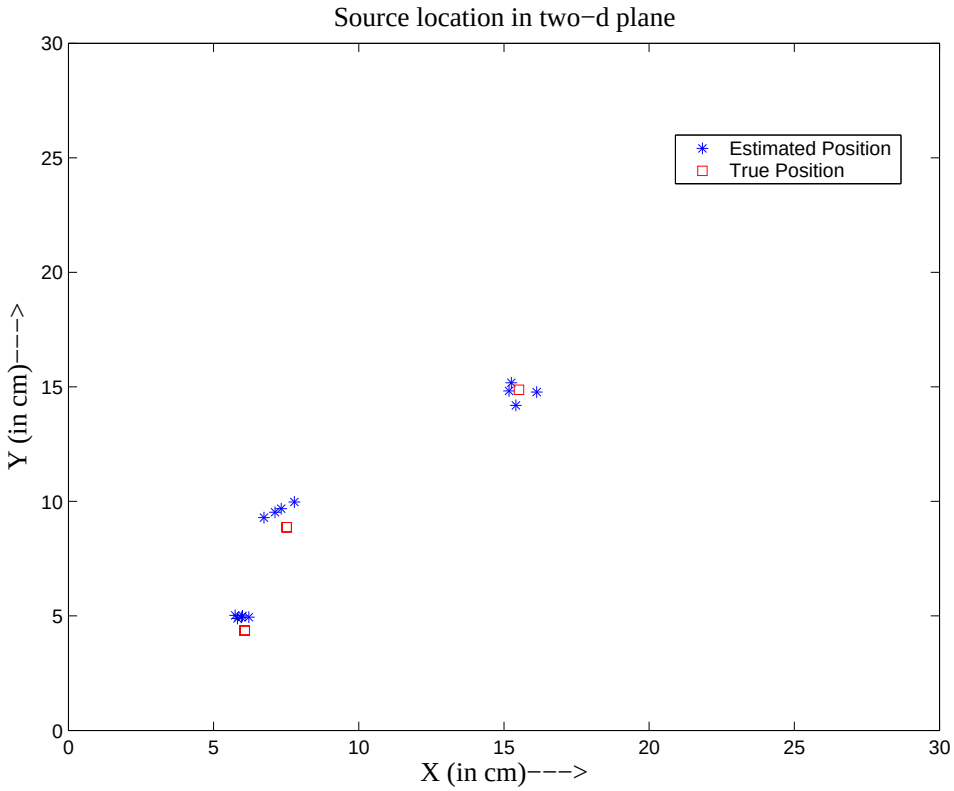


Figure 3.1: True and estimated acoustic source positions using three element 'L'-shaped microphone array

The solutions for hyperbolic position location fixes proposed by Chan and Ho have been utilized for a linear array of three microphones. For a linear microphone array the solution matrices containing source location will be singular because the sensor positions satisfy a linear set of equations. Fang provides an exact solution to the nonlinear equations that forms a consistent system. For a 2-dimensional hyperbolic position location system using three microphones to estimate the source location (x, y), Fang establishes a coordinate system so that the first microphone is located at the origin and the other two microphones located on x and y axes each. However, Fang's method does not make use of redundant measurements made at additional microphones to improve position location accuracy. Thereby, this method cannot be extended to arrays with more than three microphones.

3.2.1 Array of Three Sensors. $(M = 3)$

Chan's method [20] for a three microphone system $(M = 3)$, producing two TDOA's, x and y can be solved in terms of R_1 from the equation (A.1) for $i = 1$. The solution is in the form of

$$\begin{bmatrix} x \\ y \end{bmatrix} = - \begin{bmatrix} x_{2,1} & y_{2,1} \\ x_{3,1} & y_{3,1} \end{bmatrix}^{-1} \mathsf{X} \left\{ \begin{bmatrix} r_{2,1} \\ r_{3,1} \end{bmatrix} r_1 + \frac{1}{2} \begin{bmatrix} r_{2,1}^2 - K_2 + K_1 \\ r_{3,1}^2 - K_3 + K_1 \end{bmatrix} \right\} \tag{3.2}$$

where

$$\begin{aligned} K_1 &= X_1^2 + Y_1^2 \\ K_2 &= X_2^2 + Y_2^2 \\ K_3 &= X_3^2 + Y_3^2 \end{aligned}$$

When the equation (3.2) is inserted into equation above, with $i = 1$, a quadratic equation in terms of r_1 is produced. Substituting the positive root back into the equation (3.2) results in the final solution. Two positive roots may exist from the quadratic equation that can produce two different solutions, resulting in an ambiguity. This problem has to be resolved by using a priori information.

3.2.2 Linear Array

When the sensors are arranged linearly, the matrices containing x_i and y_i will be singular because the sensor positions satisfy $y_i = ax_i + b$, $i = 1, 2, ..., M$ where a and b are some constants. Rewrite (A.9) as

$$-2x_{i,1}\left(x + ay\right) - 2r_{i,1} = r_{i,1}^2 - K_i + K_1 \tag{3.3}$$

Equation (3.3) is linear in $w = (x + ay)$ and r_1.

Here we consider a special case of a linear array of three sensors. Carter [30] has derived an exact formula for source range and bearing. The sensor positions are numbered as $(x_1 = 0, y_1 = 0)$, $(x_2 = -L_1, y_2 = 0)$ and $(x_3 = L_2, y_3 = 0)$ so that r_1 corresponds to the range r of the object. Substituting $a = 0$, $x_{2,1} = -L_1$, $x_{3,1} = L_2$, $K_1 = 0$, $K_2 = L_1^2$ and $K_3 = L_2^2$ into (3.3) yields

$$\begin{bmatrix} -L_1 & r_{2,1} \\ L_2 & r_{3,1} \end{bmatrix} \begin{bmatrix} x \\ r \end{bmatrix} = -\frac{1}{2} \begin{bmatrix} r_{2,1}^2 - L_1^2 \\ r_{3,1}^2 - L_2^2 \end{bmatrix} \tag{3.4}$$

Solving (3.4) for x and r gives

$$r = \frac{L_1 \left[1 - \left(\frac{r_{2,1}}{L_1}\right)^2\right] + L_2 \left[1 - \left(\frac{r_{3,1}}{L_2}\right)^2\right]}{2\left(\frac{r_{3,1}}{L_2} + \frac{r_{2,1}}{L_1}\right)} \tag{3.5}$$

and the bearing angle

$$\theta = \cos^{-1}\left(\frac{x}{r}\right) = \cos^{-1}\left(\frac{L_2^2 - 2rr_{3,1} - r_{3,1}^2}{2rL_2}\right) \tag{3.6}$$

and

$$x = \frac{r_{2,1}L_2^2 - r_{3,1}L_1^2 - r_{2,1}r_{3,1}r_{3,2}}{2\left\{r_{2,1}L_2 + r_{3,1}L_1\right\}} \tag{3.7}$$

and y is obtained from $\sqrt{r^2 - x^2}$ or $y = r\sin(\theta)$

3.2.3 Fang's Method

For microphones placed in an orderly fashion and with a consistent system of equations in which the number of equations equals the number of unknown source coordinates to be solved, Fang [31] provides an exact solution to the nonlinear equations. For a 2-D hyperbolic position location system using three microphones to estimate the source location (x, y), Fang establishes a coordinate system so that the first microphone is located at $(0, 0)$, the second microphone at $(x_2, 0)$ and the third microphone at $(0, y_3)$. Realizing that for the first microphone, where $i = 1$, $X_1 = Y_1 = 0$, and for the second microphone,

where $i = 2$, $Y_2 = 0$ and for the third microphone, where $i = 3$, $X_3 = 0$, the following relationships are simplified

$$\begin{aligned} R_i &= \sqrt{(X_1 - x)^2 + (Y_1 - y)^2} \\ &= \sqrt{x^2 + y^2} \\ X_{i,1} &= X_i - X_1 = X_i \\ Y_{i,1} &= Y_i - Y_1 = Y_i \end{aligned} \tag{3.8}$$

Using these relationships, we can write the following equations:

$$2R_{2,1}R_1 = R_{2,1}^2 - X_i^2 + 2X_{i,1}x \tag{3.9}$$

$$2R_{3,2}R_1 = R_{3,1}^2 - \left(X_3^2 + Y_3^2\right) + 2X_{3,1}x + 2Y_{3,1}y \tag{3.10}$$

Equating the two equations and simplifying results in

$$y = g * x + h \tag{3.11}$$

where

$$\begin{aligned} g &= R_{3,1} - (X_2 R_{2,1}) - X_3 Y_3 \\ h &= X_3^2 + Y_3^2 - R_{3,1}^2 + R_{3,1} * R_{2,1} \left(1 - (X_2 R_{2,1})^2\right) / (2Y_3) \end{aligned}$$

Substituting the above equations in to y results in

$$d * x^2 + e * x + f = 0 \tag{3.12}$$

where

$$d = -\left(1 - (X_2 R_{2,1})^2\right) + g$$
$$e = X_2 * \left(1 - (X_2 R_{2,1})^2\right) - 2g * h$$
$$f = -\left(R_{2,1}^2/4\right) * \left(1 - (X_2 R_{2,1})^2\right)^2 - h^2$$

Fang's method provides an exact solution; however, his solution does not make use of redundant measurements made at additional receivers to improve position location accuracy. Furthermore, his method experiences an ambiguity problem due to the inherent squaring operation. These ambiguities can be resolved using a priori information or through use of symmetry properties.

As was seen for Fang's algorithm, one of the roots of the quadratic equation in R_1 almost always gives negative values for R_1, which is not possible. In some cases when that root gives positive numbers, the numbers are too large and are well above the range values, which is again not possible. Hence, when the quadratic equation in R_1 is obtained in the form

$$aR_1^2 + bR_1 + c = 0, \tag{3.13}$$

only the following root should be considered for source position location.

$$R_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \tag{3.14}$$

Interestingly ambiguities in Fang's and Chan's algorithms are essentially the same. Observations tell that if wrong choices are made for a given case, wrong results yielded by both the algorithms are identical.

When Chan's method is compared to Fang's method, it is seen that Chan's method is the best choice for solving the hyperbolic equations. It is nearly an exact method and is better than iterative methods like Taylor-series method [32] which have the risk of

converging to local minima. Hence we conclude this section making a note that we use Chan's method to generate a solution. This solution will be used as an initial guess for solving the non-linear hyperbolic equations.

Chapter 4

Position Estimation Using A 16-Microphone Array

In this chapter we discuss a method of estimating the positions from which sounds have originated, using measurements across a number of asynchronous sensors (microphones). By measuring the delay between arrivals of sounds at each microphone and solving for the source location. Given here is an overview of the applied method of solving for source position for systems in which time synchronization between pair of microphones has been established.

4.1 Estimating Position Using Multiple Microphones

Phased array source localization has been investigated considerably in the fields of beamforming and radar [33], although the far-field assumptions, based on considerable distance between points, do not hold for room acoustics. Some sound based application examples include passive underwater acoustics [34] and ballistics positioning [35].

Generally the microphones configuration is known and a solution must be found for the range differences [36] calculated from the arrival times of a sound wave at each sensor. Arrival time differences measured at a number of spatially distributed points are related to the array geometry and the originating sound source. For each microphone pair, the

difference in path lengths from a single start point to each of the sensors is fixed. This is proportional to the difference in radial distance from the source to two nodes, termed Time Difference of Arrival (TDOA) and results in a hyperbolic constraint for possible position from each difference calculated.

A hyperbolic curve is defined in a two dimensional case and a hyperboloid surface in three dimensions. A function can be similarly defined for each microphone pair. For noiseless measurements, the solution then lies at the intersection of these hyperboloids. For a system constrained to a surface, three microphones are required, whereas for a system in free space four are needed. The challenge however, is that each of the arrival time measurements is corrupted by noise. The hyperbolic function can vary significantly with small changes, and so a guess must be obtained for the closest point of intersection of each of the curves. This can be solved, and becomes more certain as the number of nodes increases. Measurement uncertainty in practical applications have been addressed in a number of off-line and real-time techniques [37].

Finding a solution for the intersection of hyperboloids corrupted by noise is a nonlinear problem. An additive measurement error model can be used to describe how the data relates to each of the time delays and the uncertainty components. A number of techniques have been proposed for solving this system using maximum likelihood or least-squares approaches. Divide and conquer approach to least-squares estimation has been proposed in [38]. The method employed depends on the constraints implied by the intended application. In particular, how the noise is modeled and whether the system is to run in real-time must be considered. [39], [16], [17], [37] reviews the recent literature on use of the additive measurement error model, assessing criteria including linear approximations as opposed to direct numerical optimization, and iterative or closed-form techniques.

Maximum Likelihood

Use of a maximum likelihood estimator (MLE) has been historically popular, performing well in systems with large numbers of microphones, and reaching a stable solution (asymptotically consistent). It is constrained, however, by requiring assumptions of the distribution of the errors to be made. Often these uncertainties are assumed to be Gaussian, which has been justified by [40] for continuous-time signals, but does not take into account the non-Gaussian sampling error associated with digitizing the signal. Coupled with the difficulty of convergence to the correct minima, without a good initial guess, using typical iterative root finding algorithms such as Newton-Raphson, Gauss-Newton or Least-Mean-Squares linear approximation approaches, has made MLE problematic, especially for real-time use.

4.2 Least Squares Estimator

To support real-time or close to real-time processing, closed form algorithms are desired. Methods that include, and make use of, redundancy in sensor information perform better in cases with measurement error. Time difference of arrival measurements can be modeled and expressed in a least squares form. The Least Squares Estimator (LSE) can be used to minimize the squared error between the model and the measured data. With an accurate model, in the absence of noise, the square error function should be zero. The form of the error criteria derived from the source localization problem can effect the computational complexity of finding a solution.

The hyperbolic error function is formulated from the difference between the observed range differences and those calculated for the estimated source location. As described above, the range difference between a pair of sensors defines a hyperboloid, with a solution for each point on the surface [41]. Minimizing the hyperbolic least squares criterion places a sound source estimate as close to all of the hyperboloids as possible but these functions are unstable in the presence of noise, becoming mathematically intractable as the number

of sensors grow large [42].

In order to overcome these problems a second error function, based on the intersection of spheres centered at each of the microphone locations, can be derived. By substituting in a further unknown variable, the range from each of the microphones to the unknown source position R_1, a linear set of equations is obtained with a quadratic constraint linking the source coordinates and range estimation. Several techniques for solving for the source co-ordinates using this set of equations have been proposed with progressively improved performance and simplicity. Optimum performance is defined by calculating the Cramer-Rao lower bound on the variance of each estimated coordinate. Substituting in for the intermediate variable, R_1 to give $N-2$ linear equations in x, y and z, or solving for the coordinates first, to give equations in terms of R_1 only first, result in solutions which can be shown to be mathematically equivalent. This is termed the Spherical Interpolation (SI) method. Redundant information in the source range is not used with this method, and there is a large variance relative to the Cramer-Rao lower bound.

Chan and Ho [20] suggested using a second least squares estimator to make use of the redundancy in the measurements. Their approach takes advantage of the relation of the source coordinates to the range to improve the estimation efficiency using quadratic correction. This is termed the Quadratic-Correction Least-Squares (QCLS) method. Huang [37] propose further changes to this approach that no longer require an assumption on the measurement error covariance matrix to be made. A perturbation approach in the quadratic correction of QCLS, which results in a limit on the error magnitude, is also overcome. This method is termed Linear-Correction Least Squares (LCLS).

Derivation of Closed Form Estimator

Following the derivations of Chan and Ho [20], Hui and So [43], and Huang and Benesty [37] the problem can be stated as follows. Let the positions of the $N+1$ microphones, in Cartesian coordinates be denoted by

$$\boldsymbol{r_i} \triangleq \left[x_i, y_i, z_i\right]^T, \qquad i = 0, 1, ..., N,$$

where $(.)^T$ denotes the transpose of a vector. The first microphone is set to be the origin, $\boldsymbol{r_0} = [0, 0, 0]^T$ and the source coordinates sought are $\boldsymbol{r_s} \triangleq \left[x_s, y_s, z_s\right]^T$. The range between the i-th microphone and the source is given by the Euclidean norm

$$D_i \triangleq \|\boldsymbol{r_s} - \boldsymbol{r_i}\| = \sqrt{(x_s - x_i)^2 + (y_s - y_i)^2 + (z_s - z_i)^2}. \tag{4.1}$$

The corresponding range differences, for microphones i and j in a pair, to the source are given by

$$d_{ij} \triangleq D_i - D_j, \qquad i, j = 0, ..., N. \tag{4.2}$$

This is related, by the speed of sound ν, to the time difference of arrival, τ_{ij}, measured in synchronized microphone arrays by

$$d_{ij} = \nu.\tau_{ij}.$$

It is noted [37], that there are $(N + 1)\, N/2$ distinct delay measurements, τ_{ij}, excluding the $i = j$ case and repetition arising from $\tau_{ij} = -\tau_{ji}$, but any N linearly independent ν_{ij} values determine all the others in the absence of noise. For simplicity, at the cost of improved accuracy in noisy systems, the N time differences of arrival with respect to the first sensor, τ_{i0}, $i = 1, ..., N$ are selected. The range differences measured between connected microphones, d_{i0} are modeled as the actual value and an additive noise term ϵ_i, assumed to be zero-mean:

$$\hat{d}_{i0} = d_{i0} + \epsilon_i, \qquad i = 1, ..., N \tag{4.3}$$

where,

$$d_{i0} = \|\boldsymbol{r_s} - \boldsymbol{r_i}\| - \|\boldsymbol{r_s}\|.$$

From the range difference (4.2) and distance definitions (4.1) we have

$$\hat{d}_{i0} + \sqrt{(x_s)^2 + (y_s)^2 + (z_s)^2} = \sqrt{(x_s - x_i)^2 + (y_s - y_i)^2 + (z_s - z_i)^2}, \qquad i = 1, ..., N. \tag{4.4}$$

The solution to this set of hyperbolic functions is nonlinear and sensitive to measurement noise. Instead, a spherical error function can be formulated.

The distances from r_0 at the origin to the remaining microphones and to the source are denoted by R_i and R_s respectively, where

$$R_i \triangleq \|\boldsymbol{r_i}\| = \sqrt{x_i^2 + y_i^2 + z_i^2}, \qquad i = 1, ..., N \tag{4.5}$$

$$R_s \triangleq \|\boldsymbol{r_s}\| = \sqrt{x_s^2 + y_s^2 + z_s^2}. \tag{4.6}$$

Squaring both sides of (4.4) and substituting in R_s as an intermediate variable yields a set of linear, spherical signal model, equations:

$$\boldsymbol{r_i}^T \boldsymbol{r_s} + d_{i0} R_s = \frac{1}{2}\left(R_i^2 - d_{i0}^2\right), \qquad i = 1, ..., N. \tag{4.7}$$

Writing this spherical least squares error function in vector form,

$$\boldsymbol{G\theta} = \boldsymbol{h}, \tag{4.8}$$

where,

$$\boldsymbol{G} \triangleq \left[\boldsymbol{S}|\hat{\boldsymbol{d}}\right], \qquad \boldsymbol{S} \triangleq \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_N & y_N & z_N \end{bmatrix}, \tag{4.9}$$

$$\boldsymbol{\theta} \triangleq \begin{bmatrix} x_s \\ y_s \\ z_s \\ R_s \end{bmatrix}, \qquad \boldsymbol{h} \triangleq \frac{1}{2} \begin{bmatrix} R_1^2 - \hat{d}_{10}^2 \\ R_2^2 - \hat{d}_{20}^2 \\ \vdots \\ R_N^2 - \hat{d}_{N0}^2 \end{bmatrix}, \tag{4.10}$$

$$\hat{\boldsymbol{d}} \triangleq \begin{bmatrix} \hat{d}_{10} & \hat{d}_{20} & \dots & \hat{d}_{N0} \end{bmatrix}^T, \tag{4.11}$$

and [.|.] implies adjoining of two matrices. Solving the corresponding least squares criterion is a linear minimization problem

$$\min_{\theta} \quad (\boldsymbol{G\theta} - \boldsymbol{h})^T (\boldsymbol{G\theta} - \boldsymbol{h}) \tag{4.12}$$

subject to the quadratic constraint

$$\boldsymbol{\theta}^T \boldsymbol{\Sigma} \boldsymbol{\theta} = 0, \tag{4.13}$$

where $\Sigma \triangleq diag\,(1, 1, 1, -1)$ is a diagonal and orthonormal matrix. The technique of Lagrange multipliers can be used yielding the constrained least squares estimate:

$$\hat{\boldsymbol{\theta}} = \left(\boldsymbol{G}^T \boldsymbol{G} + \lambda \boldsymbol{\Sigma}\right)^{-1} \boldsymbol{G}^T \boldsymbol{h}, \tag{4.14}$$

where λ is still to be found.

An unconstrained spherical least squares estimator can be derived by not enforcing the quadratic constraint (4.13) which is equivalent to assuming that the source coordinates and distance to the origin x_s, y_s, z_s, and R_s are mutually independent. An estimate of θ is then given by

$$\hat{\boldsymbol{\theta}}_1 = \boldsymbol{G}^{\dagger} \boldsymbol{h}, \tag{4.15}$$

where

$$\boldsymbol{G}^{\dagger} = \left(\boldsymbol{G}^T \boldsymbol{G}\right)^{-1} \boldsymbol{G}^T$$

is the pseudo-inverse of the matrix $\boldsymbol{G}$. This estimate has been shown [39] to be accurate in some cases, but the source range information redundancy (4.13) can be exploited to improve this estimate.

4.3 Implementation of 16-Microphone Array

The microphone array consists of 16 microphones. The reason for building such a large array is the performance of a microphone array improves linearly as the size of the array grows. This is well established in the literature on microphone arrays [44], [45]. To date, the largest microphone array known to us [46] has been a 1020-element array and work in microphone arrays of this size has been extremely limited.

In this section we describe the implementation of the 16-element microphone array and the algorithm utilized for computation of source position. We first outline the hardware design of the array components and the connections to the Personal Computer (PC). We then present the array geometry that we have used and propose an algorithm to improve the performance of the microphone array.

Hardware

Our microphone array feeds signals into the PC sound card. Each microphone in the 16-element array is biased via an amplifier. The setup is designed for applications requiring real-time source location estimation. We use PC sound card for acquisition. We have the limitation on the number of acquisition channels. The PC sound card consists of two i/p channels. The two i/p channels explores the 16-element microphone array through two 16-channel analog multiplexers [47].

The 16 microphones are arranged on a square planar cardboard as shown in the Fig. 4.3 (The microphones used are electret condenser microphones). We have opted for small microphone modules to ensure low cost, not compromising on the precision of estimation. The sound card samples the signals at 44.1 kHz, with four triggers per second.

Geometry

Many array geometries have been suggested in past work, from linear to rectangular to circular, and similarly, many microphone spacing schemes have been suggested, from uniform to logarithmic. While many geometrical configurations of the array are possible and potentially desirable, our final 16-element microphone geometry is a square array with 4 microphones width by 4 microphones breadth. Arrays spatially sample at the intra-microphone spacing wavelength, and any source signal component with a wavelength shorter than twice the spacing will be aliased (as per the Nyquist criterion).

For this work, we have chosen to use uniform spacing at 10 cm (meaning spatially sampling the waveform at $\frac{330m/s}{0.1m} = 3300\ Hz$, meaning frequencies above $1,650\ Hz$ will be aliased). This decision was due both to practicality reasons, as well as preliminary experiments with various spacing. The 10 cm spacing is maintained in both length and breadth directions. The sound source (speaker) is positioned on a movable stand (movable in x, y and z directions). The surface of the sound source is kept parallel to the planar microphone array. i.e., the speaker faces directly to the planar array. This is to avoid the angular errors in source orientation.

Electronics

The electronics section of the experimental setup mainly consists of 16 signal amplifiers and two 16-channel analog multipliers (Fig. 4.1. The circuit diagram of the amplifier is shown in the Fig. 4.4. Uniformity is maintained in signal amplification as this could impact the intensity of the acquired signal. The output of the amplifier is fed to the i/p channels of analog multiplexers. The interface circuit diagram for a 16-microphone array is shown in the Fig. 4.5. We select the channels using a parallel port which in turn is controlled by personal computer (PC).

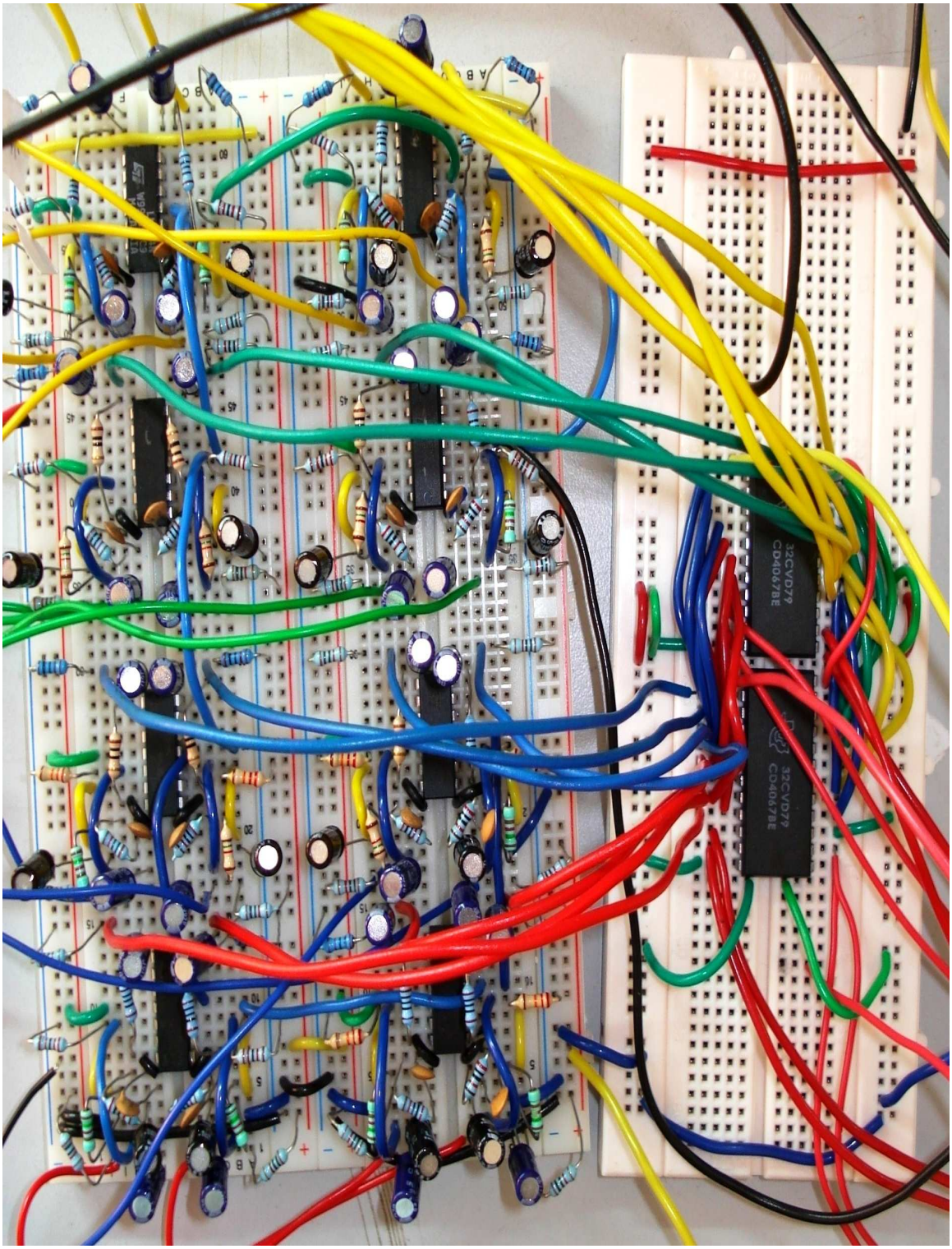


Figure 4.1: Interfacing circuit for 16-microphone array: amplifiers (left) and analog multiplexers (right)

Software

In order to utilize the microphone array to sample the signals synchronously, we selectively pick a pair of microphones at ones using analog multiplexer, which is controlled by parallel port. Initially, an application was developed in Win32 Programming to acquire signals and for further processing. Later, we utilized the Data Acquisition Toolbox available in MATLAB software to control the parallel port, sound card and for further signal processing and computations.

The problem of accurately localizing a source is crucial, but rather separate from the problem of amplifying signal from a particular pair of microphones. For the work presented in this thesis, we assume that the positions of the microphones are known. Using the MATLAB software, we set the sampling rate of A/D of PC sound card to 44.1 KHz. Since we have two channels (left and right channels of sound card), signals from a pair of microphones are acquired synchronously.

Signal acquisition is done for a duration of 0.25 seconds (meaning 4 triggers per second). We repeat the above procedure of acquiring signals for microphones M_i, $i = 2 \ldots 15$. The first microphone ($i = 0$) is regarded as the reference and placed at origin of the coordinate system $(0, 0, 0)$. So, per each trigger (acquisition), we acquire signal from 1^{st} and i^{th} microphone, where $i = 2 \ldots 15$. (Actually, we require just one 16-channel analog multiplexer and other signal from 1^{st} microphone could be directly fed to one of the channels of sound card). We fix the position of the source until we acquire the signals from all the pairs of microphones.

Evaluation

We have conducted preliminary experiments with the 16 element microphone array. The experiments involved acquiring signals from a speaker which is triggered using signal generator. The source (speaker) is located in space and its location is estimated using a planar 16-microphone array which is arrested on a table top. Several noise sources are

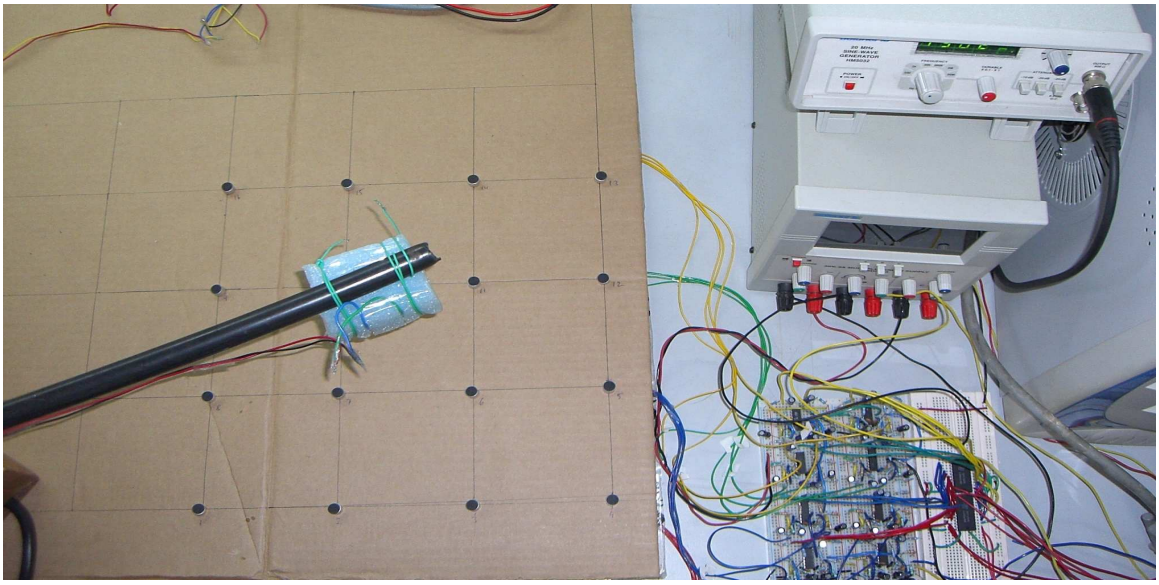


Figure 4.2: Experimental setup

present in the room where the experiments are conducted. The main noise sources are the cooling fans of computers and the loud air conditioner.

As a preliminary step to source localization, we have used a narrow band source (speaker fed by a sinusoid signal). We kept the signal frequency at 1500 Hz to meet the condition $\frac{330m/s}{2*0.1m} = 1650\ Hz$. In this manner we have avoided the use of analog filters (which are cumbersome in their design and implementation).

As mentioned earlier, we have assumed a fixed position for our subject. The difference in the amount of time required for the sound to travel from the position to each microphone in the array is determined with the DFSE algorithm.

4.4 Non-Linear Optimization

By using a large number of microphones we can formulate the problem as one which minimizes an error criterion. Specifically with the fact that for a given microphone pair, the estimated path difference d_{i0} and the actual path difference b_{i0} will never be equal and the true location would be one which minimizes this difference among all microphone

pairs.

Section 2.3 discusses about TDOA between a pair of microphones. Equation (2.1) gives the TDOA which can be converted to path differences. Assuming the velocity being constant, the estimated path differences d_{i0} and the actual path differences b_{i0} will never be equal because the estimated delays are corrupted by noise. Given M pairs of sensors, their spatial coordinates and the estimated delays, the source location X is one which minimizes the error between the actual and the estimated path differences over all microphone pairs.

$$\hat{X} = \arg_X \left(\min \left(E \left(X \right) \right) \right) \tag{4.16}$$

where $E\left(X\right)$ could be absolute difference of estimated and actual path differences or square of the difference.

$$E\left(X\right) = \sum_{i=1}^{M} \| d_{i0} - b_{i0}\left(X\right) \| \tag{4.17}$$

or

$$E\left(X\right) = \sum_{i=1}^{M} \left[d_{i0} - b_{i0}\left(X\right) \right]^2 \tag{4.18}$$

These do not have a closed-form solution since they are non-linear functions of X. We will have to use optimization methods here. We have compared the results of estimated source location using both the above functions (4.17), (4.18).

The geometrical interpretation is as follows. Intersection of all the hyperboloids corresponding to each microphone pair would be a unique point. Due to the errors in the estimation of time delays we do not have a unique intersection point. So, by using the redundant microphones we try to find the source location which best belongs to all the hyperboloids in the least square sense.

The functions to be minimized are non-linear. We evaluate Nelder Mead optimization technique to minimize the functions. The simulations were done using MATLAB. Much of the details given here are taken from *Optimization Toolbox User's Guide* and [48].

Nelder Mead Simplex Method

This is a direct search method that does not use numerical or analytic gradients. A simplex in n-dimensional space is characterized by the $n + 1$ distinct vectors that are its vertices. In two-space, a simplex is a triangle. In three-space, it is a pyramid. At each step of the search, a new point in or near the current simplexes generated. The function value at the new point is compared with the function's values at the vertices of the simplex and, usually, one of the vertices is replaced by the new point, giving a new simplex. This step is repeated until the diameter of the simplex is less than the specified tolerance. It can handle discontinuities. Suitable for nonlinear problems with a large number of discontinuities. Does not require the evaluation of gradient or Hessian. But this method has a slow convergence due to large number of iterations. We use the *MATLAB function* ***fminsearch***.

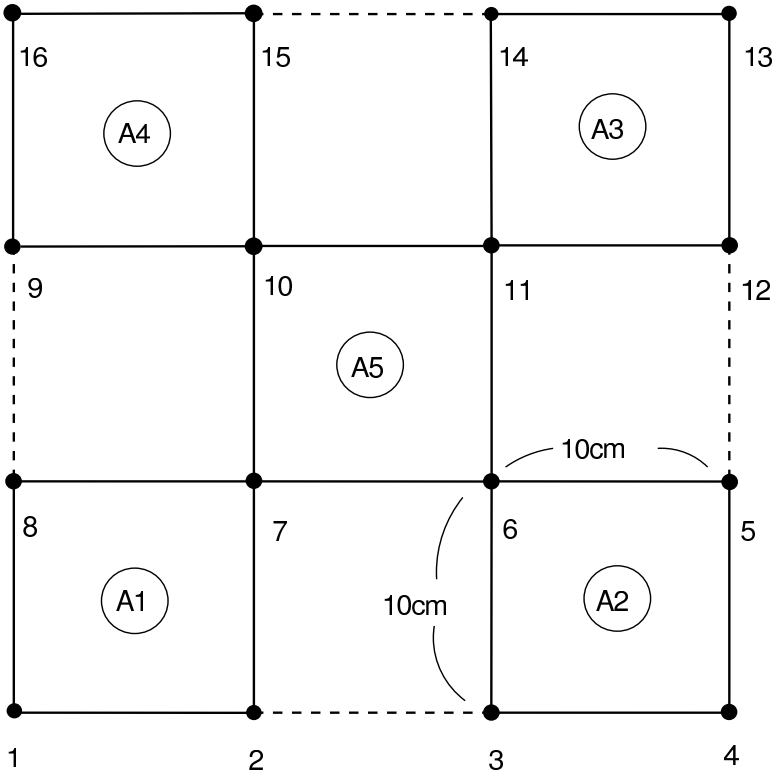


Figure 4.3: 16-Microphone array

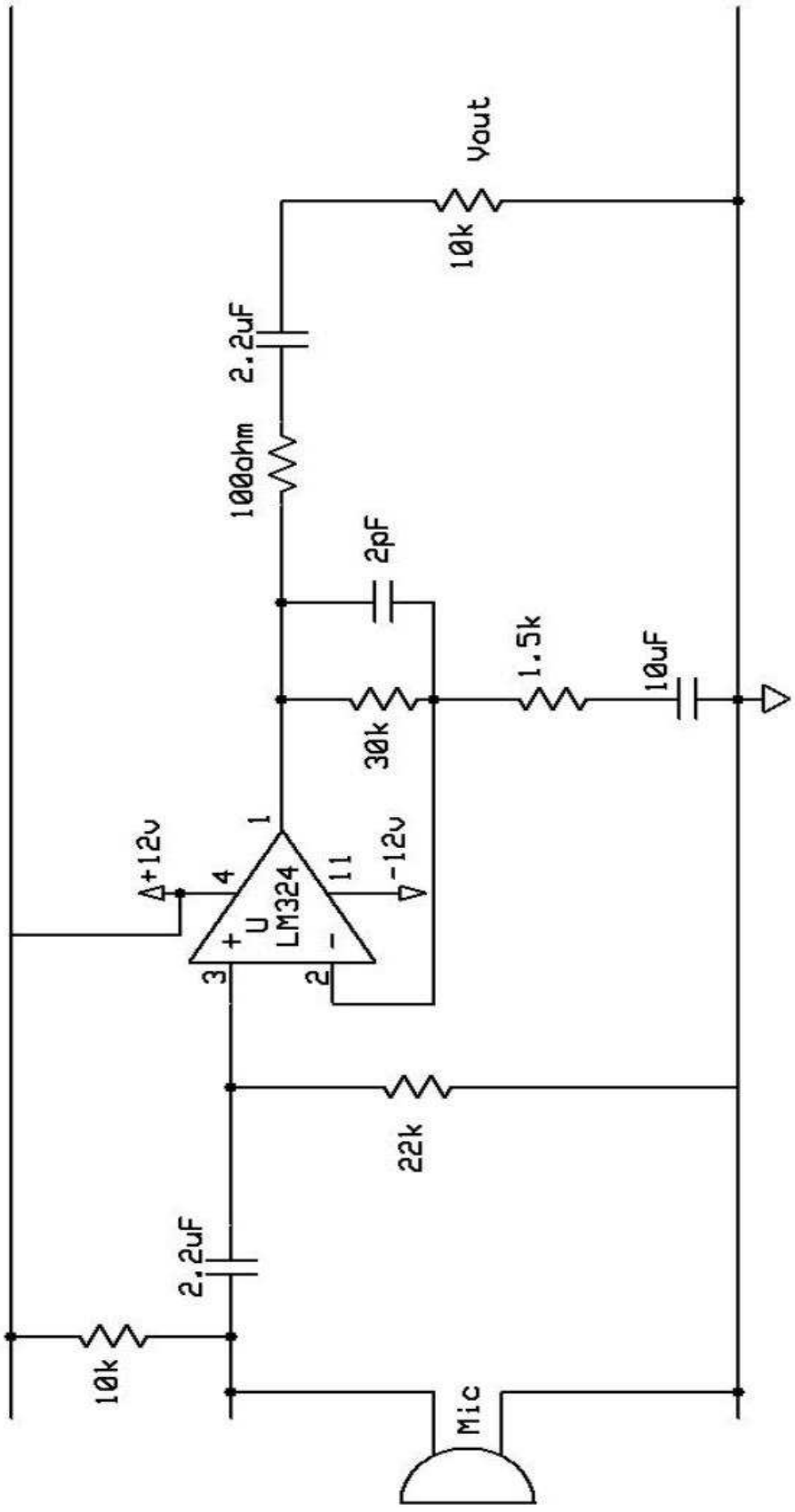


Figure 4.4: Microphone amplifier circuit diagram

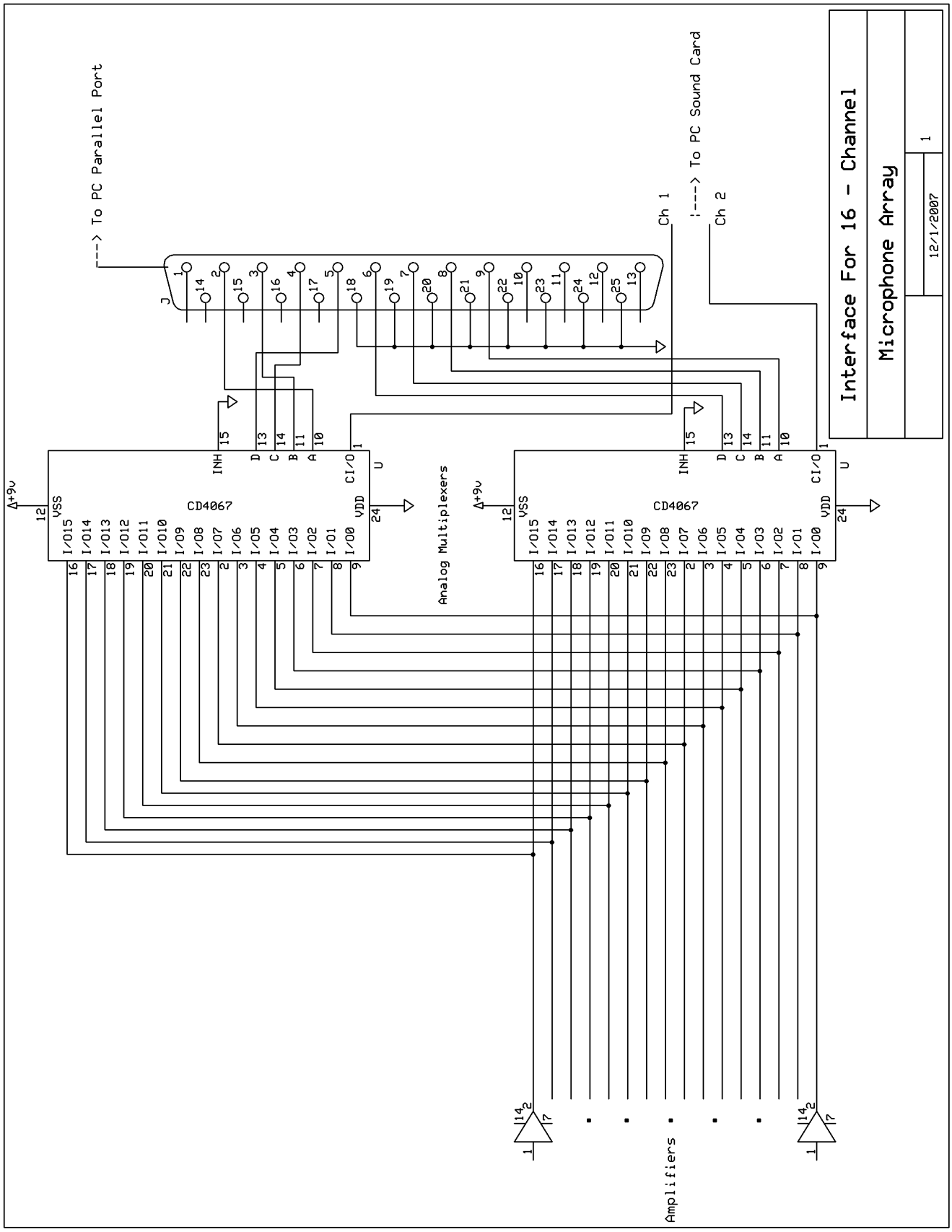


Figure 4.5: Interface for a 16-microphone array

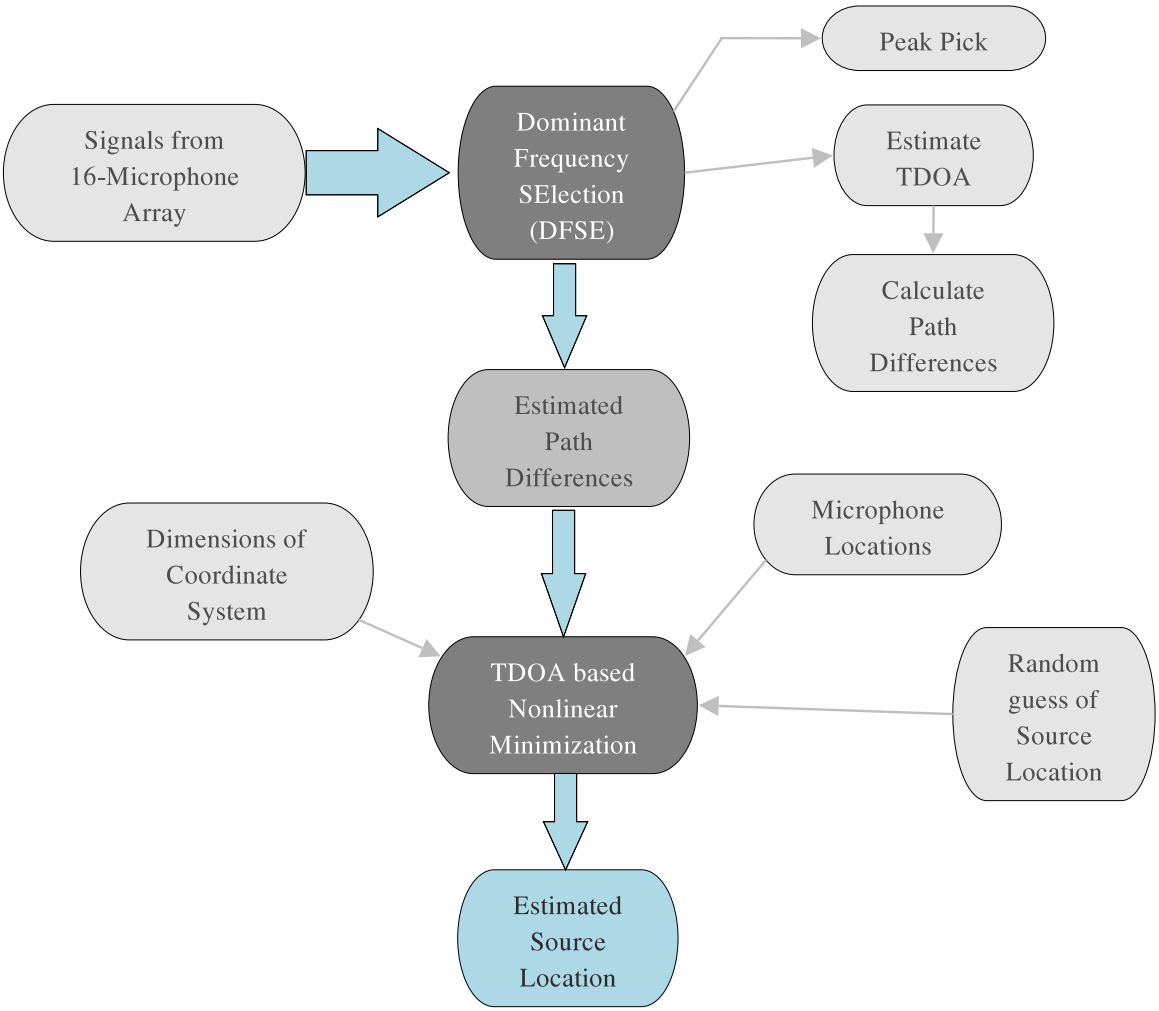


Figure 4.6: Flow chart of the complete algorithm

Chapter 5

Results and Discussion

5.1 Results From DFSE Algorithm

Experiments were done, using the algorithms described in the previous chapter, in order to be able to gain insight into the operation of the system. A localization error for each scenario was measured as the difference between the true angle, calculated from the center of the array to the primary source, and the estimated angle as predicted by the time delays. For this, it was assumed that the source was far away, compared to the size of the array, and that the source could therefore fall on a straight line from the array. This assumption was made and the errors calculated for both the azimuthal and altitudinal angles of incidence and for each time-delay estimation routine implemented.

By its definition, the altitudinal angle may vary from +90 to -90. The maximum angular error is therefore 180. Similarly, the azimuthal angle may vary from 0 to 360. Because the measurement system is circular, a maximum error of 180 also exists for the azimuth prediction. Comparisons could then be made of the performance of the various time-delay estimation techniques:

- Dominant Frequency Selection algorithm (DFSE),
- Maximizing the cross-correlation coefficient function (GCC),

- Maximizing the cross-correlation coefficient with phase transform as the pre-filter (GCC-PHAT).

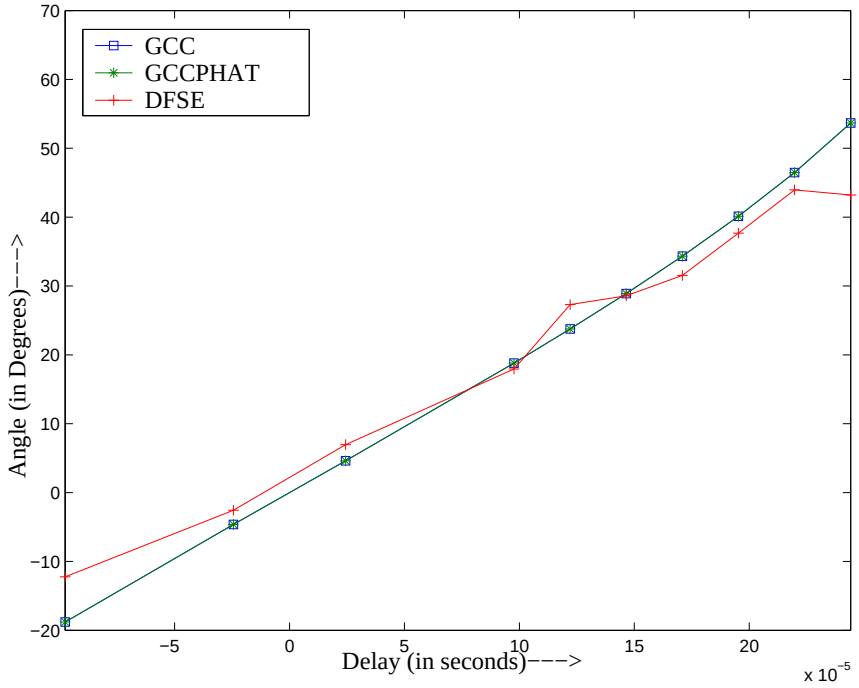


Figure 5.1: DOA estimation for frame size = 1 sec.

Each can produce different results under various circumstances, and some are more sensitive to error than others. This will be discussed in this section. A number of complications limit the potential accuracy of the system. Some of these are due to physical phenomena that can never be corrected, and others are due to inherent errors built into the processing, due to the design of the system. As mentioned in the introduction, complications in locating the sound source that exist outside of perfect conditions.

A microphone separation distance of 10 *cm* was used for all the experiments. This value was chosen for practical purposes related to the required physical limitations on the device's size. An additional reason for this choice was that, also for practical reasons, this distance was chosen for use in laboratory. Reasons for this selection were discussed in chapter 2.

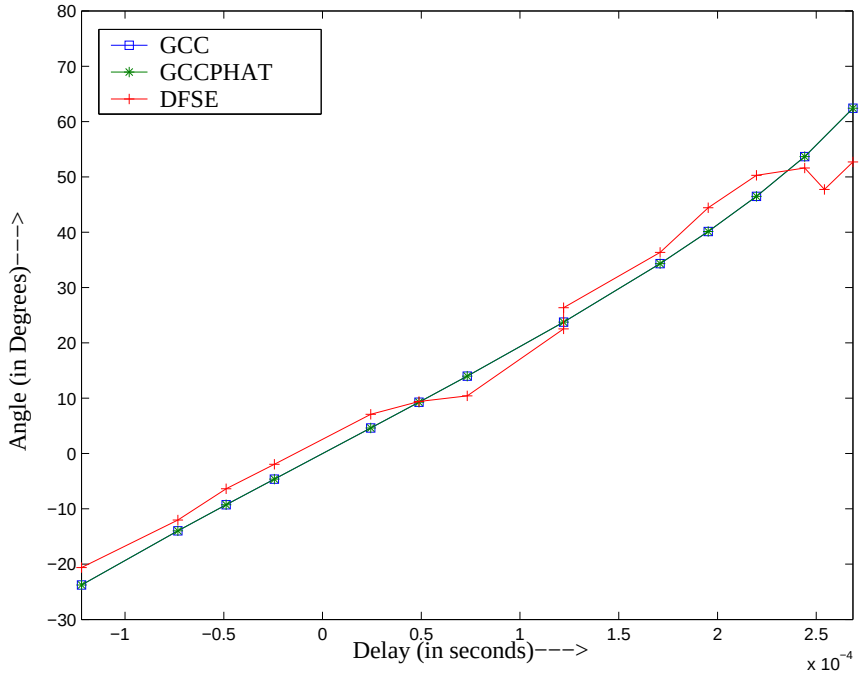


Figure 5.2: DOA estimation for frame size = 2 sec.

Each of the techniques has predicted the angle of incidence with an error of less that 1 deg. The DOA estimation of the sound source for various angles is shown in the Fig. 5.1 and Fig. 5.2. In the case of three-microphone array, it is confirmed that the phase is a good indicator of time delay in the absence of other noise sources. Results from the cross-correlation (GCC), GCC-PHAT, and DFSE techniques have produced identical results for various angles. With this simple, well-behaved function, the TDE from the dominant frequency is at a similar time delay as that of the cross-correlation techniques.

5.2 Source Localization Using Linear and 'L' Shaped Microphone Array

A very simple experiment was first done with a set of three microphones arranged in a linear fashion. We examine the performance and compare the time-delay estimation with GCC, GCC-PHAT. A single, broadband source was positioned at an angle of $\phi_r = 30$ and approximately 20 cm from the array. Here, ϕ_r represents the true angles at which

the source was positioned.

Direction of the source was estimated from the measured sharp peak created by cross-correlation of microphone pairs. For this experiment, there was no bandpass filtering of the signal. The measured phase differences and path differences for a source at various positions using a three element linear microphone array are produced in the table 5.1. The table also shows the estimated source location and the direction of the source. As mentioned earlier, the spacing between the microphones was 10 cm. Source localization using 'L' shaped microphone array is shown in the Fig. 3.1.

Table 5.1: Measured phase differences and path differences at various positions.

Actual (x, y)(in cm)	Measured (x, y)(in cm)	Measured (r, θ)	Phdiff21, d21	Phdiff32, d32	Phdiff31, d31
(0,10)	(-0.5, 13.2)	(13.2, 92.1)	0.88, 3.1	0.15, 0.53	1.05, 3.66
(-10, 10)	(-8.9, 11.1)	(14.2, 128.7)	-0.88, -3.1	3.01, 10.5	2.2, 7.7
(10, 20)	(8.7, 19.6)	(21.45, 66.17)	1.62, 5.63	-2.08, -7.25	-0.51, -1.79
(0, 20)	(-0.9, 18.4)	(18.37, 92.92)	0.6, 2.09	0.23, 0.81	0.86, 2.99
(20, 20)	(20.8, 22.2)	(30.44, 46.85)	2.16, 7.55	2.56, 8.9	-1.65, -5.74
(-20, 10)	(-17.5, 10.7)	(20.46, 148.6)	-2.13, -7.4	-1.5, -5.23	2.58, 9
(-10, 30)	(-7.2, 33.8)	(34.56, 102.1)	-0.19, -0.65	1.29, 4.5	0.97, 3.37
(-10, 30)	(-6.8, 28.8)	(29.61, 103.3)	-0.18, -0.61	1.37, 4.78	1.08, 3.74
(0, 25)	(-1.6, 26.5)	(26.55, 93.4)	0.36, 1.26	0.3, 1.05	0.68, 2.38
(10, 15)	(9.5, 13.2)	(16.25, 54.1)	2.09, 7.3	-2.99, -10.42	-0.88, -3.07
(-10, 15)	(-8.3, 12.8)	(15.24, 123.1)	-0.68, -2.36	2.73, 9.53	2.03, 7.09
(-10, 40)	(-5.9, 33.88)	(34.3, 99.93)	-0.08, -0.27	0.95, 3.31	0.87, 3.05
(-10, 40)	(-6.6, 40.9)	(41.56, 99.1)	-0.11, -0.38	0.93, 3.25	0.78, 2.7
(0, 40)	(-0.4, 41.9)	(41.93, 90.53)	0.31, 1.09	0.07, 0.25	0.37, 1.27
(10, 40)	(8.4, 38.2)	(39.1, 77.63)	0.94, 3.29	-1.28, -4.47	-0.25, -0.87
(20, 40)	(15.9, 41.9)	(44.83, 69.24)	1.27, 4.44	-1.92, -6.68	-0.72, -2.5
(-20, 20)	(-17.2, 19.9)	(26.27, 130.8)	-1.47, -5.14	2.96, 10.32	2.12, 7.4
(-20, 30)	(-17.4, 25.7)	(30.96, 124.3)	-1.24, -4.32	3.05, 10.62	1.88, 6.56

5.3 Results From 16-Microphone Array

The source is positioned at various places in 3-D space. The table 5.2 produces the true, estimated source locations of the sound source. As mentioned in the chapter 4 Nelder Mead nonlinear optimization method is utilized for solving the nonlinear equations. The source position is estimated with both the expressions given in (4.17) and (4.18). The frequency had to be reduced as the sound source was moved away from the origin microphone. This is to increase the resolution of the source position estimation. Fig. 5.3 shows clearly the true and estimated source positions.

Table 5.2: True and estimated source position

No.	True Position	Estimated Position Absolute	Estimated Position Squared	Freq (Hz)
1	(10,10,12)	(10.35,9.96,11.18)	(11.47,9.59,11.9)	1500
		(10.5,8.1,10.99)	(12,6.3,15.5)	1500
		(10.45,8.35,10.5)	(12.1,7.24,16.24)	1500
2	(20,10,12)	(18.44,8.36,10)	(18.42,8.19,9.6)	1300
		(18.43,8.08,10.46)	(18.39,7.96,10.06)	1300
		(18.57,7.77,10.78)	(18.52,7.89,10.11)	1300
		(18.6,7.96,10.7)	(18.4,8.14,9.58)	1300
3	(10,20,13)	(9.5,18.9,11.4)	(10.3,18.6,12.62)	1100
		(9.39,19,11.2)	(9.05,19.15,12.91)	1100
		(10.18,18.5,14.73)	(10.47,17.74,13.3)	1300
		(10.1,17.8,13.32)	(10.31,17.47,12.62)	1300
4	(20,20,14)	(19.62,20.58,13.4)	(doesn't converge)	1500
		(19.74,20.35,11.84)	(doesn't converge)	1200
		(20.2,17.9,20)	(19.41,18.2,19.6)	800
		(19.16,19.44,18.57)	(22.03,21.4,19.2)	1050
		(21.65,17.1,21.85)	(21.6,17.3,21.59)	800
		(21.48,17.15,19.65)	(19.67,18.5,18.47)	850

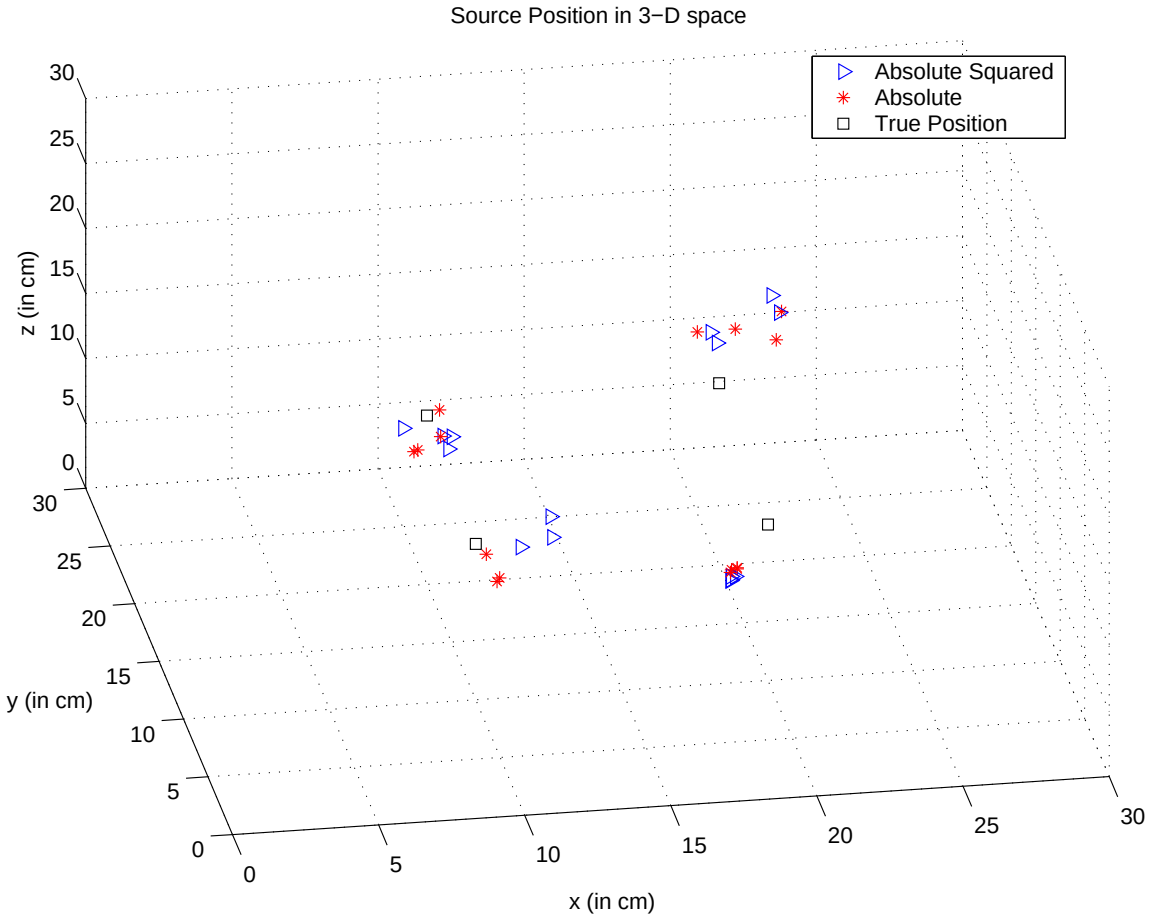


Figure 5.3: Distribution of source location estimates around the true position

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this chapter we described the problem of locating the acoustic source in 2-D plane and 3-D space using microphone arrays. Our approach of determining the time-delays using the proposed Dominant Frequency SElection (DFSE) algorithm produces results that are comparable to the conventional time-delay methods such as GCC, GCC-PHAT. The acoustic positioned in a plane and space have been successfully located. Together with the DFSE for TDOA estimation, the source could be localized. Experiments on DFSE were done using both narrow band and broad band signals.

6.2 Future Work

The hyperbolic position location techniques presented in this thesis provides a general overview of the capabilities of the system. Further research is needed to evaluate the Dominant Frequency SElection (DFSE) algorithm for hyperbolic position location system. If improved TDOA's could be measured, the source position can be estimated very accurately. Improving the performance of the DFSE algorithm for TDOA measurements reduces the TDOA errors. The DFSE discussed for TDOA measurements is in its simplest form. DFSE for targeting a broad band source has a poor performance. This could

further be improvised for accurate TDOA measurements. Experiments were performed assuming that the source is stationary until all the microphones have finished sampling the signals. Sophisticated multi-channel sampling devices could be used to get rid of this stationary condition.

While the accuracy of the TDOA estimate appears to be a major limiting factor in the performance of the hyperbolic position location system, the performance of the hyperbolic position location algorithms is equally important. Position location algorithms which are robust against TDOA noise and are able to provide unambiguous solution to the set of nonlinear range difference equations are desirable. For real-time implementations of source localization, closed-form solutions or iterative techniques with fast convergence to the solution could be used. The trade-off between computational complexity and accuracy exist for all position location algorithms. The trade-off analysis through performance comparison of the closed-form and iterative algorithms can be performed.

Appendix

Appendix A

Hyperbolic Positions Location Estimation

Chan and Ho [20], [21] have proposed the solutions for hyperbolic position fixes. This appendix discusses the derivation.

A.1 General Model

A general model for the two-dimensional (2-D) position location estimation of a source using three microphones is developed. All TDOAs are measured with respect to the center microphone M_1 and let index $i = 2, 3$. $i = 1$ represents the microphone M_1. Let (x, y) be the source location and (X_i, Y_i) be the known location of the $i\ th$ microphone. The squared range difference between the source 'S' and the $i\ th$ microphone is given as

$$\begin{aligned} R_i &= \sqrt{(X_i - x)^2 + (Y_i - y)^2} \\ &= \sqrt{X_i^2 + Y_i^2 - 2X_i x - 2Y_i y + x^2 + y^2} \end{aligned} \tag{A.1}$$

The range difference between center microphone M_1 and $i\,th$ microphone is

$$\begin{aligned} R_{i,1} &= c\,\tau_{i,1} = R_i - R_1 \\ &= \sqrt{(X_i - x)^2 + (Y_i - y)^2} - \sqrt{(X_1 - x)^2 + (Y_1 - y)^2} \end{aligned} \tag{A.2}$$

where c is velocity of sound, $R_{i,1}$ is the range difference distance between M_1 and $i\ th$ microphone, R_1 is the distance between M_1 and sound source and $\tau_{i,1}$ is the estimated TDOA between M_1 and $i\ th$ microphone. This defines the set of nonlinear hyperbolic equations whose solution gives the 2-D coordinates of the source.

A.2 Position Estimation Techniques

An effective technique in locating a source based on intersections of hyperbolic curves defined by the time differences of arrival of a signal received at a number of sensors. The non-iterative approaches give an explicit solution.

We discuss a closed-form solution, valid for both distant and close sources when TDOA estimation errors are small. We consider a special case of a linear array, with three sensors to provide two TDOA estimates.

The development of solution is in a $2-D$ plane with three microphones. Extension of this would be to introduce more number of microphones. This would give more than two TDOA measurements whose number is greater than the number of unknowns (coordinates of sound source). We make use of these extra measurements with extra sensors to improve position accuracy. Assume that there are M sensors placed arbitrarily in a $2-D$ plane. Let the sampled observations at sensor i be $u_i(k) = s(k - d_i) + n_i(k)$, $i = 1, 2, ..., M$. Where $s(k)$ is the signal radiating from the source, d_i the time delay associated with receiver i and $n_i(k)$ be the additive noise. The signal and noises are assumed to be mutually independent, zero mean stationary Gaussian random processes.

To localize the source, we first estimate TDOA of the signal received by sensors i and j using DFSE technique proposed in chapter 2. The technique measures TDOA's w.r.t. the first receiver, $d_{i,1} = d_i - d_1$ for $i = 2, 3, ..., M$. TDOA between receivers i and j are computed from

$$d_{i,j} = d_{i,1} - d_{j,1} \qquad i, j = 2, 3, ..., M \tag{A.3}$$

Let $\mathbf{d} = [d_{2,1}, d_{3,1}, ..., d_{M,1}]^T$ be the estimated TDOA vector. Denote the noise free value of $\{*\}$ as $\{*\}^o$. TDOA $d_{i,j}$ will then be

$$d_{i,j} = d^o_{i,j} + n_{i,j} \qquad i, j = 1, 2, ..., M \tag{A.4}$$

with $n_{i,j}$ representing the noise component.
Define the noise vector as $\mathbf{n} = [n_{2,1}, n_{3,1}, ..., n_{M,1}]^T$. Since TDOA is unbiased, the mean of $\mathbf{n}$ is zero.

Let $i = 2, ..., M$ and source be at unknown position (x, y) and sensors at known locations (x_i, y_i). The squared distance between the source and sensor i is

$$\begin{aligned} r_i^2 &= (x_i - x)^2 + (y_i - y)^2 \\ &= K_i - 2x_i x - 2y_i y + x^2 + y^2, \qquad i = 1, 2, ..., M. \end{aligned}$$

$$where \quad K_i = x_i^2 + y_i^2 \tag{A.5}$$

$$\tag{A.6}$$

If c is the speed of sound propagation, then

$$r_{i,1} = cd_{i,1} = r_i - r_1 \tag{A.7}$$

define a set of nonlinear equations whose solution gives (x, y).

Solving these nonlinear equations is difficult. Linearizing (A.7) by Taylor-series expansion [32] and then solving iteratively is one possible way. An alternative is to first

transform (A.7) into another set of equations. From (A.7) $r_i^2 = (r_{i,1} + r_1)$ so that (A.5) can be rewritten as

$$r_{i,1}^2 + 2r_{i,1}r_1 + r_1^2 = K_i - 2x_i x - 2y_i y + x^2 + y^2 \tag{A.8}$$

Subtracting (A.5) at $i = 1$ from (A.8), we obtain

$$r_{i,1}^2 + 2r_{i,1}r_1 = -2x_{i,1}x - 2y_{i,1}y + K_i - K_1 \tag{A.9}$$

The symbols $x_{i,1}$ and $y_{i,1}$ stand for $x_i - x_1$ and $y_i - y_1$ respectively. We note that (A.9) is a set of linear equations with unknowns x, y and r_1. To solve for x and y, [49] eliminates r_1 from (A.9) and produces $(M - 2)$ linear equations in x and y. The intermediate result is inserted back into (A.9) to generate equations in the unknown r_1 only.

Bibliography

[1] H. Staras and S. N. Honickman, "The accuracy of vehicle location by trilateration in a dense urban environment," *IEEE Transaction on Vehicular Technology*, vol. VT-21, no. 1, pp. 38–44, February 1972.

[2] J. W. S. Turin. G. L and T. S. Johnston, "Simulation of urban vehicle monitoring systems," *IEEE Transaction on Vehicular Technology*, vol. VT-21, no. 1, pp. 9–16, February 1972.

[3] M. Hebert and S. Drulhe, "Source localization for distributed robotic teams." *Robotics Institute of Carnegie Mellon University*, 2002.

[4] B. C. Dalton, "Audio-based localisation for ubiquitous sensor networks," Master's thesis, Massachusetts Institute of Technology, September 2005.

[5] A. J. E. Brandstein. M. S and H. F. Silverman, "A closed form location estimator for use with room environment microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 1, pp. 45–50, 1997.

[6] M. Omologo and P. Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'94)*, vol. 2, pp. 273–276, April 1994.

[7] M. M. P. Svaizer and M. Omologo, "Acoustic source localization in a three-dimensional space using crosspower spectrum phase." *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'97)*, 1997.

[8] M. Omologo and P. Svaizer, "Acoustic source localization in noisy and reverberant environments using csp analysis," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'96)*, pp. 901–904, 1996.

[9] F. Gustafsson and F. Gunnarsson, "Positioning using time-difference of arrival measurements," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2003)*, vol. VI, pp. 553–556, 2003.

[10] Y. Yu and H. F. Silverman, "An improved tdoa-based location estimation algorithm for large aperture microphone arrays," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2004)*, vol. IV, pp. 77–80, 2004.

[11] G. D. A. Xuehai Bian, James M. Rehg, "Sound source localization in domestic environment," *GVU center, Georgia Inst. of Technology*, 2004.

[12] P. Minero, "State of the art on localization and beamforming of an acoustic source," June 21 2004, summary of Localization Techniques.

[13] T. G. Ulrich Klee and J. McDonough, "Kalman filters for time delay of arrival-based source localization," *EURASIP Journal on Applied Signal Processing*, pp. 1–15, 2006.

[14] G. C. Knapp. C. H., Carter, "The generalized correlation method for estimation of time delay." *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-24, no. 4, pp. 320–327, August 1976.

[15] G. C. Carter, "Coherence and time delay estimation," *Proc. IEEE*, vol. 75, pp. 236–255, February 1991.

[16] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *Bell Labs Tech. Memo.*, 1998.

[17] J. B. Yiteng Huang and G. W. Elko, "Adaptive eigenvalue decomposition algorithm for realtime acoustic source localization system," *IEEE*, pp. 937–940, 1999.

[18] J. H. DiBiase, "A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays," Ph.D. dissertation, Brown University, May 2000.

[19] M. S. Brandstein, "A framework for speech source localization using sensor arrays," Ph.D. dissertation, Brown University, Prividence, RI, USA, 1995, ph.D.thesis.

[20] Y. T. Chan, "A simple and efficient estimator for hyperbolic location," *IEEE Trans. on Signal Processing*, vol. 42, no. 8, pp. 1905–1915, August 1994.

[21] Y. T. Chan and K. C. Ho, "An efficient closed-form localization solution from time difference of arrival measurements," *IEEE*, 1994.

[22] J. C. Schau and A. Z. Robinson, "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," *IEEE Transactions on Acoustic, Speech, and Signal Processing*, vol. 29, no. 4, pp. 984–995, July 1989.

[23] J. O. Smith and J. S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 12, pp. 1661–1669, December 1987.

[24] J. S. Abel and J. O. Smith, "The spherical interpolation method for closed-form passive localization using range difference measurements," *Proc. ICASSP-87 (Dallas, TX)*, pp. 471–474, 1987.

[25] J. O. Smith and J. S. Abel, "The spherical interpolation method for source localization," *IEEE Journal of Oceanic Engineering*, vol. OE-12, no. 1, pp. 246–252, January 1987.

[26] J. S. Abel and J. O. Smith, "Source range and depth estimation from multipath range difference measurements," *IEEE Transaction on Acoustic, Speech, and Signal Processing*, vol. 37, no. 8, pp. 1157–1165, August 1989.

[27] B. M. H. Rabinkin. D. V, "A dsp implementation of source location using microphone arrays." *Journal of Acoustical Society of America*, vol. 99, no. 4, April 1996.

[28] EKSLER, "Time delay estimate method for detection of acoustic source location," in *In Proceedings of 9th Conference and Competition STUDENT EEICT*, vol. 2. FEKT, 2003, pp. 216–220.

[29] P. A. G. Bendat. J. S., *Engineering Applications of Correlation and Spectral Analysis*. John Wiley and Sons, Inc., 1980.

[30] G. C. Carter, "Time delay estimation for passive sonar signal processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 3, pp. 463–470, June 1981.

[31] B. T. Fang, "Simple solutions for hyperbolic and related position fixes," *IEEE Trans. on Aerosp. Electron. Systems*, vol. 26, no. 5, pp. 748–753, September 1990.

[32] W. H. Foy, "Position-location solutions by taylor-series estimation," *IEEE Trans. Aerosp. Electron. Systems*, vol. AES-12, no. 12, pp. 187–194, March 1976.

[33] P. Mailloux, *Phased Array Antenna Handbook*. Artech House, 1994.

[34] D. Carevic, "Tracking target in cluttered environment using multilateral time-delay measurements." *The Journal of the Acoustical Society of America*, 2003.

[35] C. L. G. Ferguson. B. G and L. K. W, "Locating far-field impulsive sound sources in air by triangulation," *Journal of Acoustical Society of America*, 2002.

[36] R. O. Schmidt, "A new approach to geometry of range difference location," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-8, no. 6, pp. 821–835, November 1972.

[37] G. W. E. Yiteng Huang, Jacob Benesty and R. M. Mersereau, "Real-time passive source localization: A practical linear-correction least squares approach." *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, November 2001.

[38] J. S. Abel, "A divide and conquer approach to least-squares estimation," *IEEE Transaction on Aerospace and Electronic Systems*, vol. 26, pp. 423–427, March 1990.

[39] J. B. Yiteng Huang and G. W. Elko, "Passive acoustic source localization for video camera steering," *Acoustics, Speech, and Signal Processing, 2000.ICASSP'00.Proceedings*, vol. 2, 2000.

[40] W. R. Hahn and S. A. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays." *IEEE Transaction in Information Theory*, 1973.

[41] W. K. M. K. W. Cheung, H. C. So and Y. T. Chan, "Least squares algorithms for time-of-arrival-based mobile location," *IEEE Transactions on Signal Processing*, vol. 52, no. 4, pp. 1121–1128, April 2004.

[42] S. M. S. Bechler. D and K. Kroschel, "System for robust 3d speaker tracking using microphone array measurements," in *Proc. 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sep 28 - Oct 2 2004, pp. 2117–2122.

[43] H. C. So and S. P. Hui, "Constrained location algorithm using tdoa measurements," *IEICE Trans. Fundamentals*, vol. E86-A, no. 12, December 2003.

[44] W. R. P. H. F. Silverman and J. L. Flanagan., "The huge microphone array (hma)-part i." *IEEE Transactions on Concurrency*, vol. 6, no. 4, pp. 36–46, October-December 1998.

[45] ——, "The huge microphone array (hma)-part ii." *IEEE Transactions on Concurrency*, vol. 7, no. 1, pp. 32–47, January-March 1999.

[46] A. A. Eugene Weinstein, Kenneth Steele and J. Glass, "Loud: A 1020-node modular microphone array and beamformer for intelligent computing spaces," *MIT/LCS Technical Memo*, 2004.

[47] *CMOS Analog Multiplexers/Demultiplexers*, Texas Instruments, June 2003, data Sheet.

[48] W. T. V. W. H. Press, S. A. Teukolsky and B. P. Flannery, *Numerical Recipes in C*. Cambridge University Press, New York, 1992.

[49] B. Friedlander, "A passive localization algorithm and its accuracy analysis," *IEEE J. Ocean Eng.*, vol. OE-12, pp. 234–245, January 1987.