

INDOOR IMPULSIVE ACOUSTIC SOURCE LOCALIZATION METHODS UTILIZING SPARSE SIGNAL RECONSTRUCTION ALGORITHMS

BY

MUHAMMAD OMER

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

ELECTRICAL ENGINEERING

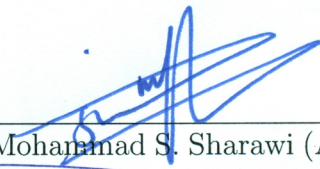
January, 2012

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN 31261, SAUDI ARABIA

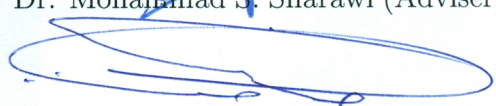
DEANSHIP OF GRADUATE STUDIES

This thesis, written by **MUHAMMAD OMER** under the direction of his thesis adviser and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN ELECTRICAL ENGINEERING**.

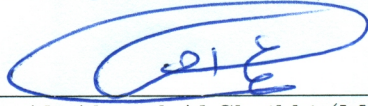
Thesis Committee



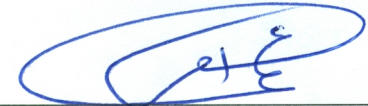
Dr. Mohammad S. Sharawi (Adviser)



Dr. Tareq Y. Al-Naffouri (Member)



Dr. Ali Ahmad Al-Shaikhi (Member)



Dr. Ali Ahmad Al-Shaikhi
Department Chairman



Dr. Salam A. Zummo
Dean of Graduate Studies



17/1/12
Date

INDOOR IMPULSIVE ACOUSTIC SOURCE
LOCALIZATION METHODS UTILIZING
SPARSE SIGNAL RECONSTRUCTION
ALGORITHMS

by

MUHAMMAD OMER

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

In Partial Fulfillment of the Requirements
for the degree

MASTER OF SCIENCE

IN

ELECTRICAL ENGINEERING

KING FAHD UNIVERSITY
OF PETROLEUM & MINERALS

Dhahran, Saudi Arabia

JAN 2012

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **MUHAMMAD OMER** under the direction of his thesis adviser and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN ELECTRICAL ENGINEERING**.

Thesis Committee

Dr. Mohammad S. Sharawi (Adviser)

Dr. Tareq Y. Al-Naffouri (Member)

Dr. Ali Ahmad Al-Shaikhi (Member)

Dr. Ali Ahmad Al-Shaikhi
Department Chairman

Dr. Salam A. Zummo
Dean of Graduate Studies

Date

Dedicated to Mama & Papa

ACKNOWLEDGMENTS

In the name of Allah, the Most Beneficent, the Most Merciful

All praise be to Allah (The One and The Only Creator of everything) for His limitless blessings. May Allah bestow peace and His choicest blessings on His last prophet, Hazrat Muhammad (Peace Be Upon Him), his family (May Allah be pleased with them), his companions (May Allah be pleased with them) and his followers.

First and foremost gratitude is due to my esteemed university, King Fahd University of Petroleum and Minerals (KFUPM) for my admittance, and to its exceptional faculty members for imparting quality learning and knowledge with their valuable support and able guidance that led my way through this point of undertaking my research work. Also, I want to thank King Abdullah City of Science and Technology (KACST) for supporting the research work through generous funding.

I would like to express my profound gratitude and appreciation to my thesis committee chairman and adviser, Dr. Mohammad S. Sharawi, whose expertise, understanding, and patience, added considerably to my graduate experience. I appreciate his vast knowledge and skills in many areas. Dr. Sharawi is the one

professor who truly made a difference in my life. It was under his tutelage that I developed a focus and became interested in research. He provided me with direction and intellectual support, first during course work and then during the research phase, and became more of a mentor than a professor. It was through his persistence, understanding and kindness that I completed my degree. I owe him my eternal gratitude.

Special thanks to my committee member Dr. Tareq Al-Naffouri for his assistance and guidance in the mathematical part of the thesis. The continuous conversations and discussions with him has provided me with better understanding of the underlying theory of compressive sensing. I would also like to thank Dr. Ali Shaikhi for being part of my thesis committee. I highly acknowledge his supportive comments regarding my efforts.

I would like to say special thanks Mr. Ahmad Abdul Quadeer for his valuable support in my thesis. Without his support, I might not have done such a magnitude of work.

Finally, I would like to thank my parents for their prayers and support from the core of my heart. I would also like to thank my sisters for the moral support they provided and my brother Mr. Muhammad Ali Khan for his support during my MS, and his useful advices for my future academic career. Last but not the least, I would like to thank my fiancé for all the encouragement she provided through such crucial phase of my life.

May Allah help us in following Islam according to Quran and Sunna!(Ameen)

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
ABSTRACT (ENGLISH)	x
ABSTRACT (ARABIC)	xii
NOMENCLATURE	xiii
CHAPTER 1 INTRODUCTION	1
1.1 Problem Description	3
1.2 Previous Work	8
1.3 Motivation for Research	9
1.4 Thesis Organization	11
CHAPTER 2 ACOUSTIC FUNDAMENTALS	12
2.1 Acoustic Wave Equation	12
2.2 Types of Acoustic waves	15
2.2.1 Longitudinal Waves	15
2.2.2 Transverse Waves	16
2.3 Basic Acoustic Parameters	16
2.4 Sound Propagation in the Atmosphere	18
2.4.1 Spreading Losses	18
2.4.2 Atmospheric Absorption	19

2.4.3	Ground Effects	20
2.4.4	Diffraction and Barriers	20
2.4.5	Wind and Temperature Gradient Effects	21
2.5	Room Acoustics	21
2.5.1	Reverberation Time	22
2.5.2	Early Decay Time	23
2.5.3	Room Impulse Response	23
2.6	Summary and Conclusions	24

CHAPTER 3 TIME DELAY ESTIMATION BASED LOCALIZATION TECHNIQUES

26

3.1	Signal Models for TDE	27
3.1.1	Ideal Propagation Model	27
3.1.2	Multipath Model	28
3.1.3	Reverberant Model	28
3.2	Time Delay Estimation Techniques	29
3.2.1	Cross Correlation based TDE	29
3.2.2	Generalized Cross Correlation based TDE	30
3.2.3	Adaptive Eigenvalue Decomposition based TDE	31
3.3	Localization Techniques	33
3.3.1	Direction of Arrival (DOA)	33
3.3.2	Time Difference of Arrival (TDOA)	37
3.4	Summary and Conclusions	41

CHAPTER 4 TIME DELAY ESTIMATION BASED ON ORTHOGONAL CLUSTERING

42

4.1	Signal Model for RIR Estimation	43
4.2	Sparse Signal Estimation Techniques	45
4.3	Proposed Method of RIR Estimation	47
4.3.1	MMSE Estimation of a Sparse Signal	50
4.3.2	Structure based Bayesian Recovery Approach	53

4.3.3	The orthogonal clustering (OC) Algorithm for RIR Estimation	55
4.4	Orthogonal Clustering based TDE	56
4.5	Summary and Conclusions	58
CHAPTER 5	EXPERIMENTAL SETUP AND RESULTS	59
5.1	Hardware Components	60
5.1.1	Acoustic Sensors and Pre-Amplifiers	60
5.1.2	Data Acquisition Device	61
5.1.3	Processing Platform	63
5.2	Time Delay Estimation Results	64
5.2.1	TDE Simulation Results	64
5.2.2	TDE Experimental Results	69
5.3	Source Localization Results	73
5.3.1	DOA based Source Localization	73
5.3.2	TDOA based Source Localization	74
5.3.3	Discussion on the Results	103
5.4	Summary and Conclusions	105
CHAPTER 6	CONCLUSIONS AND FUTURE WORK	108
VITAE		119

LIST OF TABLES

5.1	Comparison of time-delay estimates obtained using CC and the proposed technique based on OC algorithm.	72
5.2	Mic. configuration I results, for the source location of (100,100). .	83
5.3	Mic. configuration I results, for the source location of (120,60). . .	84
5.4	Mic. configuration I results, for the source location of (60,140). . .	85
5.5	Mic. configuration I results, for the source location of (-100,100). .	86
5.6	Mic. configuration I results, for the source location of (-140,60). .	87
5.7	Mic. configuration II results, for the source location of (100,100). .	95
5.8	Mic. configuration II results, for the source location of (100,200). .	96
5.9	Mic. configuration II results, for the source location of (200,200). .	97
5.10	Mic. configuration III results, for the source location of (100,200). .	106
5.11	Mic. configuration III results, for the source location of (200,100). .	107

LIST OF FIGURES

1.1	The process of sound localization	3
1.2	Impulsive acoustic signal from a toy gun	4
1.3	Impulsive acoustic signal from a hand clap	5
2.1	Direct path signal and reflections in a reverberant environment . .	22
2.2	Concept of room acoustics using image source model	24
3.1	Microphone array for DOA estimation	34
3.2	DOA based localization, intersection of lines of bearing	37
3.3	TDOA based localization, intersecting hyperbolic curves	39
3.4	Fixed range difference hyperbolic curve	40
4.1	Flowchart of the OC algorithm	57
5.1	Electret microphone mounted on the back side of MAX9814 board.	60
5.2	Top side of MAX9814 LNA evaluation kit.	61
5.3	PCI Data Acquisition Device from MCC.	62
5.4	Interface board, DBK202 from MCC.	62
5.5	USB Data Acquisition Device from MCC.	63
5.6	Shoe box room model of dimensions $8 \times 6 \times 3$ meters.	64
5.7	(a) Typical impulsive acoustic source signal, (b) Room impulse re- sponse, (c) Received signal at the microphone.	65
5.8	Example of simulated RIR estimation (low reverberations).	66
5.9	Example of simulated RIR estimation (dense reverberation).	67

5.10	MSE in time-delay estimation using the OC algorithm for different sub-sampling rates at SNR = 30 dB and 40 dB.	67
5.11	Experimental setup for indoor TDE.	70
5.12	Experimental RIR estimate with microphone at the room center. .	71
5.13	Experimental RIR estimate with microphone close to the wall. . .	72
5.14	Microphone configuration I for source localization.	76
5.15	Experimental setup of microphone configuration I.	77
5.16	Mic. configuration I, localization results for source at(100,100). . .	78
5.17	Mic. configuration I, localization results for source at (120,60). . .	79
5.18	Mic. configuration I, localization results for source at (60,140). . .	80
5.19	Mic. configuration I, localization results for source at (-100,100). .	81
5.20	Mic. configuration I, localization results for source at (-140,60). .	81
5.21	Localization failures in CC and OC for microphone configuration I.	82
5.22	Microphone configuration II for source localization.	89
5.23	Experimental setup of Mic. configuration II.	90
5.24	Mic. configuration II, localization results for source at (100,100). .	91
5.25	Mic. configuration II, localization results for source at (100,200). .	92
5.26	Mic. configuration II, localization results for source at (200,200). .	93
5.27	Localization failures in CC and OC for microphone configuration II.	94
5.28	Microphone configuration III for source localization.	99
5.29	Experimental setup of Mic. configuration III.	100
5.30	Mic. configuration III, localization results for source at (100,200).	101
5.31	Mic. configuration III, localization results for source at (200,100).	102
5.32	Localization failures in CC and OC for microphone configuration III.	103

THESIS ABSTRACT

NAME: Muhammad Omer

TITLE OF STUDY: Indoor Impulsive Acoustic Source Localization Methods
utilizing Sparse Signal Reconstruction Algorithms

MAJOR FIELD: Electrical Engineering

DATE OF DEGREE: Jan 2012

Acoustic source localization is a fairly developed field and current research is focused on building robust systems based on more realistic sound propagation models and low complexity hardware. The performance of these systems rely on the time delay estimation (TDE) techniques used.

This work presents an efficient technique for TDE using low rate samples in a room environment excited by an impulsive acoustic source. Sampling at low rate decreases the communication between the microphones and the centralized location which results in increased battery life of the microphones. The time delays are found from a pair of microphones. The estimates are obtained using a recently proposed sparse reconstruction technique known as Orthogonal Clustering (OC). The time delay estimates are then used in localization methods to locate the

impulsive acoustic source in 2D.

In addition, an impulsive source signal localization prototype is implemented. The performance of the experimental results of this new technique is compared against the methods that appeared in literature in this context. The effect of sampling rates, number of microphones and their configuration on the accuracy of the impulsive source localization system is investigated.

الخلاصة

الاسم : محمد عمر

عنوان : تحديد مكان مصدر صوت انفجاري داخل غرفة باستخدام متناثر

الخوارزميات إعادة الإعمار إشارة.

التخصص الرئيسي : الهندسة الكهربائية

التاريخ : يناير 2012

إن مجال تحديد مكان مصادر الأصوات يعتبر معروف لدى الباحثين و مجالات البحث الجديدة في هذا النطاق تكاد تنحصر في بناء أنظمة فائقة الدقة على أساس اعتماد نماذج دقيقة لانتشار الصوت أو تصميم أنظمة ذات تعقيد قليل. وفعالية هذه الأنظمة تعتمد على طريقة تحديد الأزمان المتعاقبة.

هذا المشروع (العمل) يقدم طريقة حديثة و فعالة لتحديد الأزمنة المتعاقبة باستخدام معدل عينات منخفض داخل غرفة مغلقة لتحديد مكان مصدر صوت انفجاري. إن استخدام معدل عينات منخفض يقلل من كمية الاتصال ما بين المضخات الصوتية (الميكروفونات) ومعالج البيانات داخل جهاز الحاسوب , مما يؤدي إلى زيادة عمر البطارية المستخدمة. إن وقت التأخير تم رصده من ميكروفونات حين تمت معالجة الاشارات باستخدام خوارزمية جديدة تدعى المجموعة المتعامدة. هذه الاوقات تم استخدامها لتحديد مكان الصوت الانفجاري في بعدين في الفضاء.

اضافة لذلك , تم بناء نظام كامل لتحديد مكان مصادر الصوت الانفجارية في بعدين باستخدام الكترونيايات و خوارزميات متكاملة في نظام واحد. النتائج التي تم الحصول عليها من هذا النظام و الخوارزمية الجديدة تمت مقارنتها مع طرق اخرى معروفة في هذا المجال. تم ايضاً فحص النتائج من هذا النظام بعد تغيير عدد المايكروفونات المستخدمة. عدد العينات المنتقاء , وكيفية توزيع المايكروفونات على صحة النتائج المأخوذة من هذا النظام الجديد.

Nomenclature

Abbreviations

AED	: Adaptive Eigenvalue Decomposition
ADC	: Analog to Digital Converter
AGC	: Automatic Gain Control
ASL	: Acoustic Source Localization
AWGN	: Additive White Gaussian Noise
BLUE	: Best Linear Unbiased Estimate
CC	: Cross Correlation
CS	: Compressive Sensing
DAQ	: Data Acquisition
DI	: Directivity Index
DLOS	: Direct Line of Sight
DOA	: Direction of Arrival
DSL	: Digital Subscriber Lines
EDT	: Early Decay Time
FBMO	: Fast Bayesian Matching Pursuit
FIR	: Finite Impulse Response
GCC	: Generalized Cross Correlation
GPS	: Global Positioning System
IASL	: Impulsive Acoustic Source Localization
IL	: Insertion Loss

LNA	:	Low Noise Amplifier
MAP	:	Maximum a Posteriori
ML	:	Maximum Likelihood
MMSE	:	Minimum Mean Square Error
MRI	:	Magnetic Resonance Imaging
MSE	:	Mean Square Error
OC	:	Orthogonal Clustering
OFDM	:	Orthogonal Frequency Domain Multiplexing
PDF	:	Probability Density Function
PHAT	:	Phase Transform
RADAR	:	Radio Detecting and Ranging
RIP	:	Restricted Isometry Property
RIR	:	Room Impulse Response
RSS	:	Received Signal Strength
SCOT	:	Smoothed Coherence Transform
SIL	:	Sound Intensity Level
SNR	:	Signal to Noise Ratio
SPL	:	Sound Pressure Level
SONAR	:	Sound Navigation and Ranging
SIMO	:	Single Input Multiple Output
TDE	:	Time Delay Estimation
TDOA	:	Time Difference of Arrival
UWB	:	Ultra Wideband

CHAPTER 1

INTRODUCTION

Localization has a significant historical legacy. When people were traveling by sea, they used to determine their locations by measuring the angles of the celestial objects from the horizon at a known time. At first, man made landmarks like light houses were built high enough to be spotted from far. These landmarks served as references for proximity estimation of distance. Modern localization systems still rely on the idea of extracting location information from the observable angles or distances measured from certain references.

Localization systems are divided into active and passive systems. The active systems both transmit and receive signals. Examples of active system are the radio detecting and ranging (RADAR) and active sound navigation and ranging (SONAR) for underwater applications. The passive systems only receive signals. The global positioning system (GPS) and microphone arrays are the examples passive systems. Active systems suffer from some drawbacks. First, the emitter location is revealed to the receiver and second, higher energy is consumed in signal

transmission and reception. The passive systems only receive energy and hence they consume less energy and thus are more suitable for surveillance purposes. The active systems are suitable for tracking moving objects as they can transmit the signal at any instant and measure the reflections from the object. On the other hand, the passive systems rely on continuous signal reception from the source and thus they are unable to track the objects if the source signal terminates. In this thesis, we deal with passive systems.

Electrical localization techniques such as GPS and RADAR utilizes electromagnetic signals. However, in many natural occurring events like tracking an active speaker for automatic camera control in a video conferencing system, an explosion, a gun shot or an electric spark, the use of sound waves for localization becomes a necessity.

Acoustic source localization (ASL) uses sound waves to estimate the source location in a two-dimensional ($2D$) or three-dimensional ($3D$) space with respect to some known references in a coordinate system. ASL is a well developed feature of the human auditory system. Two ears separated by some distance resembles an array of microphones which can resolve the direction of an acoustic event with remarkable precision. The human ears in conjunction with the brain can accurately localize and track sources in a sound field around the head.

This thesis is concerned with the localization of impulsive acoustic sources in the environment. Examples of such impulsive acoustic sources are explosions, arms fire, collisions, accidents etc. Consider a scenario of a gun shot by a thief or

a burglar in a shopping mall. Considering large size of the malls, it becomes very critical to locate the position of the gun shot immediately. Another example can be of an oil and gas refinery where the exact location of explosion in pipe is quite significant to carry out necessary measurements immediately.

1.1 Problem Description

The process of impulsive acoustic source localization (IASL) in a room environment is shown in Figure 1.1. The four different stages of the problem can be identified as: sound emission from the impulsive source, sound propagation with multiple reflections, sound measurement from multiple microphones, followed by the localization process. The first three phases of this process are physical phenomena while the fourth stage employs various signal processing algorithms to determine the source location. The focus of this thesis will be on the last stage, that is developing an efficient signal processing algorithm and prototype for indoor source localization. We start by briefly discussing each of these stages involved in the localization process.

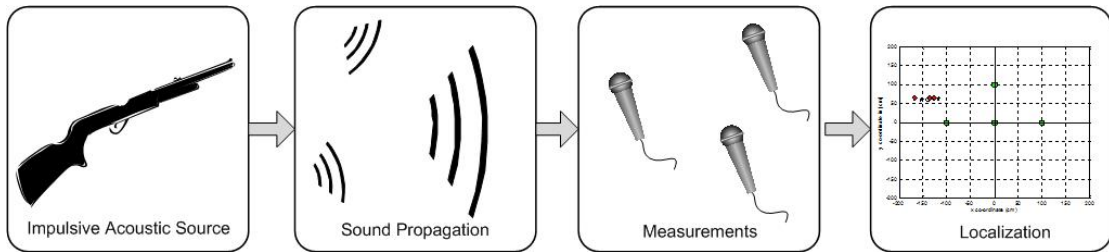


Figure 1.1: The process of sound localization

I. Impulsive Acoustic Source

Localization systems usually make assumptions about the source type based on the application under consideration. An impulsive source signal is a wide-band signal of a very short duration. Examples of impulsive sources include gun shots, explosions, electric sparks, slamming doors, or breaking glass. Figures 1.2 and 1.3 show examples of impulsive acoustic signals from a toy gun (triggered without cracker) and a human clap, respectively. The sound source may possess directional characteristic that may vary with frequency. In this thesis we do not consider the directional nature of the sources as the methods used do not exploit the source directivity.

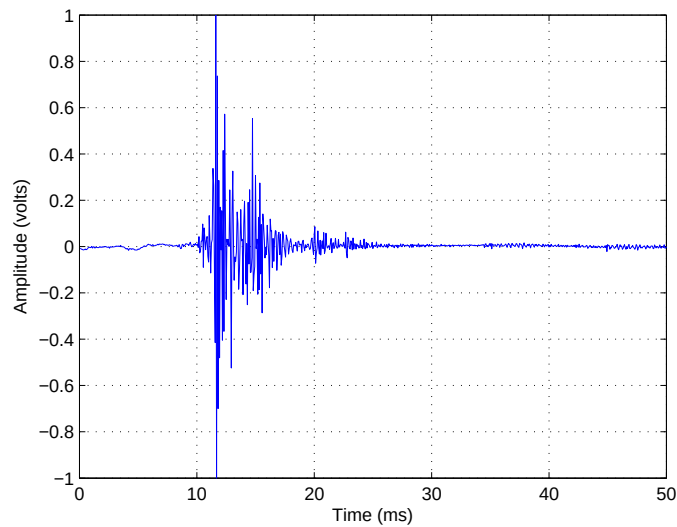


Figure 1.2: Impulsive acoustic signal from a toy gun

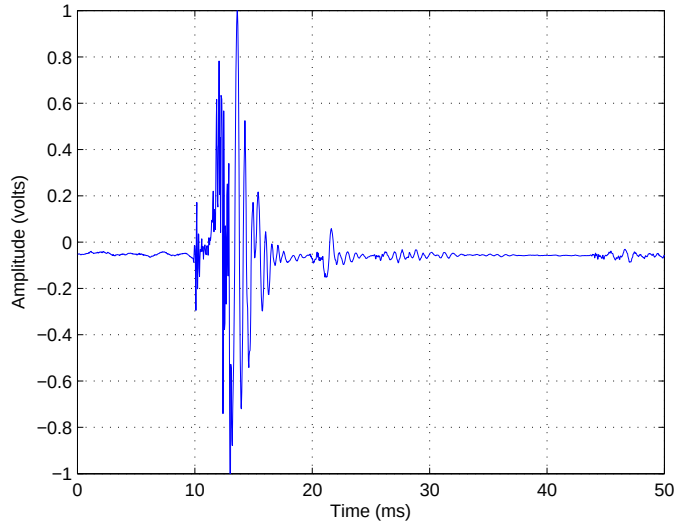


Figure 1.3: Impulsive acoustic signal from a hand clap

II. Sound Propagation

Sound waves propagate in air by compression and rarefaction, a characteristic of longitudinal waves. In an indoor environment the sound wave is disturbed by the background noise and suffers from multipath propagation. Multipath propagation is caused by sound wave reflections from objects larger than its wavelength. It also diffracts or bends around the objects that are smaller than its wavelength. Indoor environments, like halls or rooms, are characterized by the reverberation time which is the time it takes for the sound pressure to fall by 60 dB from its original level at the time of source excitation [1]. The extent of reverberation depends upon the absorption coefficients of the room surfaces which are in turn a function of the incident angles, frequency and the material properties. Thus, realistic modeling of sound propagation from the source to the acoustic receiver is not a trivial task and requires knowledge of physical properties of sound propagation.

III. Sound Measurements

Microphones measure the sound waves. They use transducers to convert the pressure change caused by the incident sound waves into corresponding voltage changes. An analog to digital converter (ADC) converts the voltage change which is an analog signal into a digital signal to be processed by digital computers. For localization purposes, multiple spatially located microphones are used to capture the sound signal. When arranged in a certain geometry, multiple microphones make microphone array. The array dimension is defined as the placement of microphones along single or multiple axis. To locate sound source in $2D$ or $3D$ space, multiple microphone arrays are used. In this thesis, we assume that microphone radiation patterns are omni-directional and their locations are known.

IV. Source Localization

The location information of the acoustic source can be extracted from various parameters of the signal measured by microphone arrays. The design of the localization technique depends upon the signal propagation conditions, background noise, source signal type and source directivity. On this basis, source localization can be performed by any of the following methods.

a) Received Signal Strength Based Method

Sound pressure is inversely proportional to the square of the distance from the source. Assuming an isotropic source and noiseless background, the ratio of the

received signal strength (RSS) at a pair of microphones gives the ratio of two squared distances. This ratio determines the possible source location on a circle. Adding one more microphone gives two such distance ratios and hence two circles which intersect at two points. The location ambiguity can be resolved by adding a fourth microphone. RSS systems perform well in short range scenarios while they are very sensitive to low signal to noise ratio (SNR) that causes large estimation errors. One such signal energy based method is described in [2].

b) Beamforming Based Method

The direction of the incoming sound can be tracked by microphone arrays. The delay sum is the simplest beamforming technique in which the response of the microphone array is steered in a desired direction by applying phase weights to the input channels. The direction that maximizes the directive response of the array corresponds to the source direction. The estimate of the source direction from spatially separated microphone arrays are used to find a least-square estimate for the unknown sensor location [3]. A review of various beamforming methods, their advantages and drawbacks are discussed in [4].

c) Time Delay Estimation Based Method

The time delay estimation (TDE) based localization methods are suitable for wideband source signals and many algorithms [5–14] have been proposed for TDE based localization. The coherent sound wavefronts arrive at the microphones at time instants depending on the microphone location, shape and direction of the

arriving wavefront. The temporal similarity between the microphone signals is used for time delay estimation.

TDE based localization systems can be divided into two categories. The first one assumes the presence of the sound source in the far-field of the microphone array. The sound wavefronts arriving at the microphones are planar and a TDE from a pair of microphones can be used to find a direction of arrival (DOA) of the signal [15]. Several DOA estimates from spatially separated microphone arrays are used to find the source location.

Another TDE based method is called the time difference of arrival (TDOA) and it assumes the presence of the sound source in the near field of the microphones [16–19]. The TDE from a pair of microphones is converted into range difference from which range difference equation is formed. Multiple range difference measurements lead to multiple equations which can be solved for the unknown source position [20,21]. With an increased number of microphones, the ambiguity of the source location can be resolved by the least squares solution for the multiple range difference equations.

1.2 Previous Work

The focus of this thesis will be on TDE based localization methods. An accurate TDE is vital for both relative angles (DOA) and range based (TDOA) localization schemes. Among the most basic TDE method are the cross-correlation (CC) based ones [7].

The CC method assumes an ideal sound propagation model which makes it suitable for low noise and reverberant free outdoor environments. The improved version of the CC method is the generalized cross-correlation (GCC) method [6]. The GCC is a family of algorithms that incorporates different knowledge to improve the TDE performance [5, 6]. Despite much improvement in GCC over the CC method, the GCC methods suffers performance degradation in dense reverberant environments [22]. The performance of GCC method under real reverberant conditions has been analyzed in [23] and several improvements have been proposed. The fundamental problem with CC and GCC methods is their assumption that the microphones receive only the direct path signal. Recently, an adaptive eigenvalue decomposition (AED) [11] algorithm has been proposed that assumes more realistic sound propagation models for TDE in a room reverberant environment. It first estimates the room impulse response (RIR) between the source and the pair of microphones. The TDE estimate is determined by identifying the direct paths from the two RIR. The adaptive RIR estimation is a computational intensive task which makes AED unsuitable for tracking acoustic sources.

1.3 Motivation for Research

The performance of the cross correlation based algorithms and their variants deteriorate in reverberant indoor environments, thus more robust algorithms are required to alleviate these effects. In addition, these TDE techniques requires high sampling rates and their performance further worsens at low sampling frequencies

caused by the missing time information [24]. The typical sampling frequency of 44.1 kHz makes the localization process not only computationally intensive but also puts a strain on the hardware requirements. An overview of problems and associated with cross correlation based TDE and possible solutions have been discussed in [14].

The drawbacks of the present TDE techniques motivate us to find a better solution for TDE. In this work, we apply the recently developed sparse signal reconstruction algorithm in [25] for the TDE problem. This algorithm has already been applied successfully to impulsive noise estimation and cancellation in DSL lines [25]. The objectives of this work are:

1. Tune the algorithm in [25] and apply it to enhance the robustness of indoor acoustic source localization systems against dense room reverberation.
2. Work at extremely low sampling rates to relax the computational and hardware complexity.
3. Build a hardware setup and investigate the algorithm performance in a real reverberant environment via actual measurements.
4. Benchmark the algorithm performance in terms of localization accuracy and computation time against the CC based method.

The method we propose finds the time delays from the estimated RIR. This makes it robust against the room reverberations. We consider the RIR as a sparse phenomenon and employ the sparse signal reconstruction technique called Orthog-

onal clustering (OC) to estimate it. The method reconstructs the sparse signals from the under sampled data which reduces the hardware and computational complexity. The arrival time of the direct path signal at a pair of microphones is identified from the estimated RIR and their difference yields the desired TDE. These estimates are then used for localization purpose. Simulation results are shown and measurements from a hardware setup are presented.

1.4 Thesis Organization

The thesis is organized as follows. In Chapter 2, we discuss some acoustic signal fundamentals including the acoustic wave equation, basic acoustic parameters, wave propagation and room acoustics. In Chapter 3 we review various TDE techniques that exist in literature and the methods of estimating the source position, using measurements from a number of microphones. Closed form solutions are presented to find the approximate source location. The details of the proposed TDE method is discussed in detail in Chapter 4. In Chapter 5 we show the simulation results of the proposed algorithm and the experimental results of three of microphone geometries, followed by results analysis. Concluding remarks, future work and some issues involved in designing a practical system are discussed in the Chapter 6.

CHAPTER 2

ACOUSTIC FUNDAMENTALS

Sound is produced as a result of the mechanical vibration of particles and propagates as longitudinal wave. It requires a medium to propagate which in our case will be air. Accurate sound propagation models require the knowledge of acoustic fundamentals.

This chapter briefly reviews fundamental acoustic concepts which are essential for acoustic source localization. We derive the sound wave equation in Section 2.1. The two basic types of waves are discussed in Section 2.2 followed by the definitions of basic acoustic parameters in Section 2.3. The effects of sound propagation through atmosphere are studied in Section 2.4. The acoustical properties of a room which characterize the room environment are discussed in Section 2.5.

2.1 Acoustic Wave Equation

A sound wave is a disturbance that propagates through a medium with some velocity. Acoustic waves passing through a medium causes pressure and density

variations. A crest of an acoustic wave passing through a small area ΔA of a medium causes an increase in the density in that region (compression) whereas a trough causes a decrease in the density (rarefaction) in that region. Thus in a small volume of a gas, a change in the pressure $\Delta P = P - P_0$ causes a change in the density from an equilibrium value ρ_0 to ρ .

Acoustic processes occur without heat exchange between adjacent volumes of gas. Such processes are called adiabatic or isentropic processes. The relation between the pressure variation and the corresponding density variation for such processes is given as

$$\frac{P}{P_0} = \left(\frac{\rho}{\rho_0} \right)^\gamma \quad (2.1)$$

where γ is the ratio of specific heat of gas at constant pressure to specific heat at constant volume. Using the Taylor series expansion, the pressure variation caused by fluctuations in density is given as

$$P = P_0 + \left[\frac{\partial P}{\partial \rho} \right]_{\rho_0} (\Delta \rho) + \frac{1}{2} \left[\frac{\partial^2 P}{\partial \rho^2} \right]_{\rho_0} (\Delta \rho)^2 + \dots \quad (2.2)$$

where $\Delta \rho = \rho - \rho_0$. Neglecting higher order terms and rearranging equation (2.2)

$$\Delta P = P - P_0 = \left[\frac{\partial P}{\partial \rho} \right]_{\rho_0} (\Delta \rho) = B \frac{\Delta \rho}{\rho_0} \quad (2.3)$$

where $B = \rho_0 \left[\frac{\partial P}{\partial \rho} \right]_{\rho_0}$ is the adiabatic bulk modulus of the gas.

We have expressed the variation in pressure in terms of density variations of the medium. Consider a physical motion of a fluid as an acoustic wave passes

through an infinitesimal volume element dV that is fixed in space. The effect of the flux of mass entering and leaving the volume is to change the density of the fluid contained within that volume as the total mass is conserved. Thus the rate of change of density of the volume can be expressed as

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho u) \quad (2.4)$$

where u is the fluid velocity. The rate of mass change in that volume is

$$\rho_o dV = -\nabla \cdot (\rho u) dV \quad (2.5)$$

Using Newton's second law, we can find the force acting on the small amount of mass dm contained within that infinitesimal region dV given by $dF = dm * a$ where a is the particle acceleration. Following from Newton's second law, one can easily derive the relationship between particle velocity and acoustic pressure. This relation is called the Euler equation.

$$\rho_o \frac{\partial u}{\partial t} = -\Delta P \quad (2.6)$$

Combining adiabatic pressure condition with the continuity equation and Euler equation, the acoustic wave equation finally takes the following form [1]:

$$\nabla^2 P = \frac{1}{c^2} \frac{\partial^2 P}{\partial t^2} \quad (2.7)$$

where $c = \sqrt{\frac{B}{\rho_o}}$ is the sound velocity. The solution of the differential equation (2.7) gives us the wave equation given by [1]

$$A = A_o e^{i(kx - \omega t)} \quad (2.8)$$

where A_o is the amplitude, k is wave number, ω is the angular frequency and t is the propagation time of the signal. As sound travels over a long distances energy losses are expected to take place causing significant decrease in the amplitude. This attenuation effect can also be accounted for in the solution of the wave equation.

$$A = A_o e^{-\alpha x} e^{i(kx - \omega t)} \quad (2.9)$$

where α is known as the absorption coefficient. The resulting equation (2.8) is modulated by a decreasing exponential function.

2.2 Types of Acoustic waves

Sounds wave are divided into two major types based on the direction of particle motion in the medium relative to the direction of the wave propagation [1].

2.2.1 Longitudinal Waves

Longitudinal waves are produced when wave particles move in the same direction as that of the wave propagation. A compressed spring generates longitudinal waves when released to move back and forth (compression and condensation)

causing compressional wave to travel down the length of the spring. Sound waves are also longitudinal waves.

2.2.2 Transverse Waves

Transverse waves are produced when the motion of wave particles is at right angle to the direction of wave propagation. A string fastened to a wall at one end and driven from the other end at right angle to the plane of the string, generates transverse waves. The disturbance travels down the length of the string whereas wave particles of the string move at right angle to the direction of propagation of wave.

2.3 Basic Acoustic Parameters

The power of an acoustic wave is defined as the energy carried by the acoustic wave per unit time. Another experimentally convenient measure is defined, known as the power density or the sound intensity (I). The power density (sound intensity) is given by:

$$I = \frac{dP}{dA} \quad (2.10)$$

where dP is the small fraction of power that interacts with the dA area of a sound detector placed perpendicular to the direction of the onset acoustic wave.

To measure the strength of a sound wave, two relative measures are defined sound intensity level (SIL) and sound pressure level (SPL). As the range of sound

intensity and pressure is quite high, it is easier to work with these relative measures in decibel scale.

$$SIL = 10 \log \left(\frac{I}{I_{ref}} \right) \quad (2.11)$$

where I is the intensity of the sound. For air the reference intensity is $I_{ref} = 10^{-12} \text{ W/m}^2$.

$$SPL = 10 \log \left(\frac{P}{P_{ref}} \right) \quad (2.12)$$

where P is the acoustic pressure. For air the reference pressure is $P_{ref} = 20^{-6} \mu Pa$.

Periodic acoustic waves can be described by combinations of sinusoidal or cosinusoidal waves. The process of expressing waves in terms of sine and cosine waves is called Fourier decomposition. Mathematically the wave shape or its profile can be written as

$$y(x) = A \sin(kx + \phi) \quad (2.13)$$

where A is the amplitude, k is the wave number, x is the distance wave travels and ϕ denotes the shift in the phase of the wave. For traveling waves, the wave profile is also a function of time t . The wave profile then takes the following form

$$y(x, t) = A \sin(k(x - ct) + \phi) \quad (2.14)$$

where c is velocity of the sound propagation in air.

Two other velocities are also defined, they are the phase velocity and group velocity. Infinitely continuous waves given in equation (2.14) propagate through

a medium with a phase velocity, whereas in wave packets – waves that consist of finite number of cycles – the center of the packet travels with a phase velocity and the ends travel with a group velocity.

2.4 Sound Propagation in the Atmosphere

Sound waves propagating through the atmosphere suffer from spreading losses, air absorption, ground effects, diffraction across barriers, and refraction with wind and temperature gradients.

2.4.1 Spreading Losses

The sound wavefront spreads with the increasing distance from the sound source. The intensity of a sound waves radiated by an isotropic source varies inversely with the square of the distance from the source.

$$I = \frac{P}{4\pi R^2} \quad (2.15)$$

where I is the sound intensity, P is the sound pressure and R is the distance from the source. In logarithmic form, the sound pressure level (SPL), L_P is related to the sound power through [1]

$$L_p = L_w - 20 \log(R) - 11dB \quad (2.16)$$

where L_w is the sound power in decibels. This implies that the sound losses 6dB of power as distance is doubled from the sound source. If the source is directional, then SPL also includes the directivity factor of the source [1]

$$L_p = L_w + DI - 20 \log(R) - 11dB \quad (2.17)$$

where DI is the directivity of the sound source defined as the ratio of sound intensity in a given direction to the intensity of an isotropic source. It can be found from:

$$DI = 10 \log \left(\frac{P/\pi r^2}{P/4\pi r^2} \right) = 10 \log \left(\frac{4}{r^2} \right) \quad (2.18)$$

where r is the radius of the circular beam of a directive source.

2.4.2 Atmospheric Absorption

Atmospheric absorption caused by heat conduction, shear viscosity and molecular relaxation causes the sound wave to attenuate. This attenuation depends upon frequency, humidity and temperature. Air absorption increases at high frequencies and thus acts as a low pass filter. A signal of frequency 0.25 KHz at 20° C temperature and 70% humidity experience an attenuation of 1.1dB/Km whereas an 8 KHz signal suffers from 76.6dB/Km attenuation [1].

2.4.3 Ground Effects

Acoustic waves radiated from sources lying close to the ground does not reach the receiver alone, in fact the received wave is the resultant of the original wave and its reflected version from the ground. The resultant wave can either have increased or diminished amplitude depending upon the constructive or destructive interference between the two waves. Non porous surfaces are acoustically hard surfaces. Reflections from these surfaces have an effect of doubling the sound pressure over a wide range of frequencies. Porous surfaces are termed as acoustically soft surfaces. Such surfaces attenuates high frequency signals as they tend to penetrate through the pores. On the other hand low frequency signals are reflected without much change in the wave amplitude or phase.

2.4.4 Diffraction and Barriers

Sound waves bend around the objects that are smaller than the sound wavelength. Barriers like large objects blocks the direct path from the sound source to the receiver. It gets diffracted around the barrier edges to reach the receiver. The sound field recorded by the receiver depends upon the relative position of the source, receiver and the barrier. The attenuation caused by the barrier is known as the insertion loss IL and is given as [1]

$$IL = 20 \log \left(\frac{p_w}{p_{w0}} \right) \quad (2.19)$$

where p_w and p_{wo} are the total sound field with and without the presence of the barrier respectively [1].

2.4.5 Wind and Temperature Gradient Effects

There is a constant motion in the atmosphere caused by wind shear and uneven heating of the earth surface. Sound fields present within the surface layer experience wind and temperature gradient effects. As height increases, temperature usually decreases and the velocity of sound propagation increases. When sound waves flow against the wind they bend upwards and the wind speed is subtracted from the speed of the sound resulting in a negative gradient in the speed of sound. Usually the wind effect dominates over the temperature effect.

2.5 Room Acoustics

In a closed environment, the sound wave front at the receiver end consists of the direct path component which is the uninterrupted sound wave as well as the reflected waves of the room boundaries or the objects within it as shown in Figure 2.1. The reflected components are weak in intensity because of the energy absorption from the obstacles and attenuation caused by traveling longer distances as compared to the direct path. The behavior of an acoustic source in a room environment can be totally different than what one expects it to be. Therefore certain set of parameters have been defined to quantify acoustic wave behavior in reverberant environments [1].

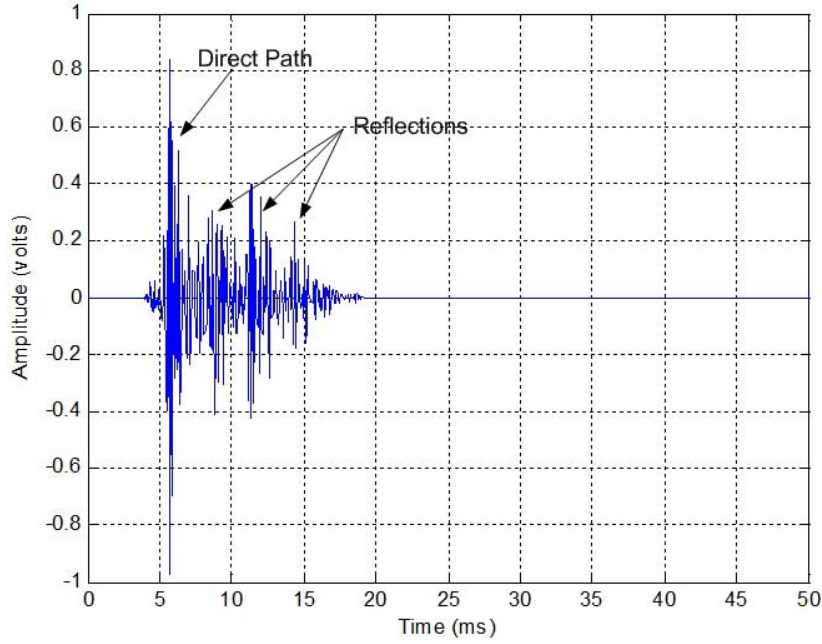


Figure 2.1: Direct path signal and reflections in a reverberant environment

2.5.1 Reverberation Time

With time, the reflection density in the room builds up while diminished in magnitude. This whole process is marked as room reverberations. Room reverberation can bring several noticeable effects. It may cause the speech to lose its intelligibility (bad effect) and it can add pleasant effects by bonding adjacent notes (good effect). The reverberation time determines the duration the reverberation lasts and is defined as the 'Time it takes for the sound level in a room to decrease by 60dB after continuous sound source has been shut off' [1]. It is found using:

$$T_{60} = 60dB \frac{t_{-X} - t_{-Y}}{(-X) - (-Y)} \quad (2.20)$$

where t_{-X} is the time when the decay has decreased X dB below its initial value and t_{-Y} is the time when decay has decreased Y dB below its initial value.

$$T_{60} = \frac{60dB}{A dB/sec} \quad (2.21)$$

where A signifies the attenuation rate and given as

$$A = \frac{(-X) - (-Y)}{t_{-X} - t_{-Y}} \quad (2.22)$$

2.5.2 Early Decay Time

Another measure of the room reverberation is the Early Decay Time (EDT). Like room reverberation time, EDT is also the decay rate evaluated from initial levels i.e. time it takes for the sound level to drop from 0 dB by -10 dB [1].

$$EDT = \frac{60dB}{A(0 \rightarrow -10)dB/sec} \quad (2.23)$$

EDT varies throughout the room unlike reverberation time T . Moreover, several room acoustic parameters like clarity, center time and others can be expressed in terms of the room reverberation time which makes it more important than EDT.

2.5.3 Room Impulse Response

The channel (acoustic space) between the microphone and the sound source can be characterized by an impulse response. The impulse response of the room can

be modeled from geometrical perspective with the concept of image source [26]. The reflected sound originates from the mirror image of the source as shown in Figure 2.2. The magnitude of the reflections depend on the surface absorption coefficients of the barrier while the distance between the mirror source and the microphone determines the propagation delay of the reflections. From the impulse response both the direct path signal and reflections can be identified.

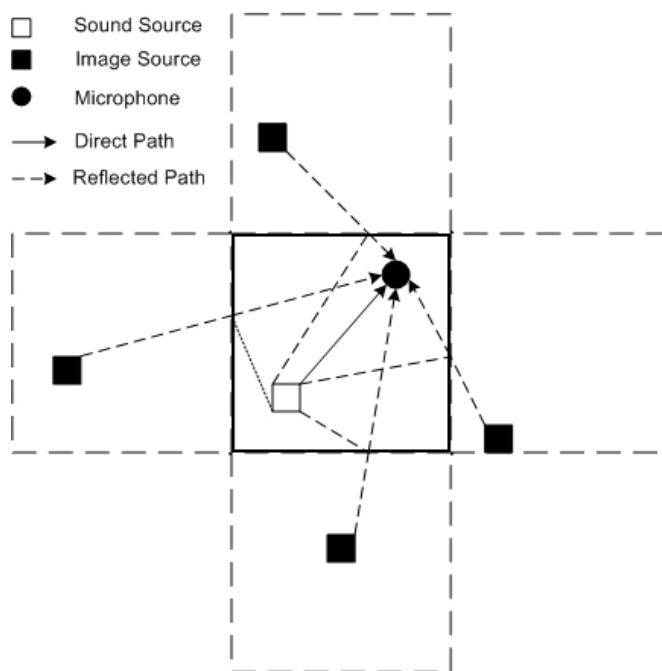


Figure 2.2: Concept of room acoustics using image source model

2.6 Summary and Conclusions

Fundamental concepts related to acoustic waves were covered in this chapter. The solution of the wave equation was derived which gives insight into the wave

propagation phenomena. We discussed basic wave types and important wave terminologies. Wave propagation through the atmosphere undergoes various distortions caused by spreading losses, atmospheric absorptions, ground effects, diffraction from obstacles and wind and temperature effects. Room acoustic parameters are discussed which includes room reverberation time and early decay time and impulse response. The room impulse response identifies the direct path signal and reflections which are very important parameters for accurate indoor source localization. This will be further discussed in the following Chapters.

CHAPTER 3

TIME DELAY ESTIMATION BASED LOCALIZATION TECHNIQUES

Time delay estimation (TDE) serves as the front end for detection, identification and localization of sound sources. Acoustic source localization techniques based on TDE are favorable because for their conceptual simplicity, low computational and low hardware complexity as compared to the signal energy or beamforming based localization methods. Majority of the practical acoustic source localization systems are based on TDE and some of these are presented in [5–14]. Such techniques perform reasonably well in moderately reverberant environments. However, in the noisy environment or when the reverberation is considerable, the source localization problem becomes challenging and many existing algorithms deteriorate and hence require more robust algorithms.

The outline of this chapter is as follows. We cover the acoustic signal propagation models in Section 3.1. In Section 3.2 we discuss the widely used TDE techniques that appeared in the literature. The TDE based localization methods are discussed in Section 3.3.

3.1 Signal Models for TDE

Modeling the signal propagation environment between source and the sensor is essential for developing and analyzing TDE algorithms. Three signal models are found in the literature which include the ideal single path model, multiple-path model and the reverberation model. We consider the case of a sound field from a single source.

3.1.1 Ideal Propagation Model

Ideal signal models assume that the signal received at a microphone is a delayed and scaled version of the original source signal plus some additive noise [14], [27].

Mathematically the signal received $r(t)$ can be expressed as:

$$r(t) = \alpha s(t - \tau) + n(t) \quad (3.1)$$

where $s(t)$ is the source signal, α is the attenuation factor, τ is the time delay of the direct path and $n(t)$ is assumed to be a zero mean the Additive White Gaussian Noise (AWGN).

3.1.2 Multipath Model

The multipath signal model takes into account multiple delayed and attenuated replicas of the original signal caused by reflections of wavefronts from boundaries and objects [14], [27]. Mathematically the received signal is described as

$$r(t) = \sum_{l=0}^{L-1} \alpha_l s(t - \tau_l) + n(t) \quad (3.2)$$

where L is the number of paths capturing significant multipath energy, α_l is the attenuation factor, τ_l is the delay of the l^{th} path and $n(t)$ is assumed to be the zero mean AWGN.

3.1.3 Reverberant Model

The multiple reflections of a sound in an indoor environment can also be described by a reverberant model. In this, the reflections of the sound wavefront from objects are accounted by considering the acoustic space between the source and the microphone as a finite impulse response (FIR) filter [14], [27]. For a single source in the sound field, the system basically acts like single input multiple output (SIMO) system. The received signals are expressed as

$$r(t) = h(t) * s(t) + n(t) \quad (3.3)$$

where $*$ denotes the convolution, $h(t)$ is the channel impulse response between the source and the sensor and $n(t)$ is assumed to be the zero mean AWGN.

3.2 Time Delay Estimation Techniques

In broad sense, time delay estimation is divided into two categories namely Time of Arrival estimation (TOA) and Time Difference of Arrival estimation (TDOA). The former technique is used by active systems like radars and active SONAR to determine the time delay between the transmitted pulse and the reception of its echo. Time synchronization is required in TOA, in case a transmitter and a receiver are used instead of a single transceiver. Whereas, the later is used in passive systems such as passive sonar or microphone array to determine the time delay in the received signal at two spatially separated receivers. The time synchronization with the transmitter is not required, however the receivers must be perfectly synchronized.

We have seen three channel models that are considered in literature. Now we focus on the methods for TDE between pairs of microphones. These TDE techniques are based on simple cross correlation, generalized cross correlation, and a more sophisticated techniques based on channel estimation.

3.2.1 Cross Correlation based TDE

The cross correlation (CC) technique is the most straight forward method of finding the time delay [7]. It assumes the ideal signal propagation model and finds the time index that maximizes the cross correlation of received signals r_1

and r_2 at a pair of microphones respectively. The time delay $\hat{\tau}^{CC}$ is found using:

$$\hat{\tau}^{CC} = \arg \max_p R_{r_1 r_2}^{CC}(p) \quad (3.4)$$

where $R_{r_1 r_2}^{CC}$ is the cross correlation function between the two received signals r_1 and r_2 , and is defined as:

$$R_{r_1 r_2}^{CC}(p) = \mathbb{E}[r_1(k)r_2(k+p)] \quad (3.5)$$

where $\mathbb{E}[\cdot]$ is the mathematical expectation operator.

Cross correlation based TDE is greatly influenced by a number of factors which may cause multiple peaks to appear resulting in ambiguity in finding the dominant peak corresponding to the true time delay. CC is usually used in open environments.

3.2.2 Generalized Cross Correlation based TDE

A class of algorithms called generalized cross correlation (GCC) employ different weighting factors to maximize the sharpness of the cross correlation peak [6]. The time delay $\hat{\tau}^{GCC}$ is found using:

$$\hat{\tau}^{GCC} = \arg \max_p R_{r_1 r_2}^{GCC}(p) \quad (3.6)$$

where $R_{r_1 r_2}^{GCC}$ is the generalized cross correlation function between the two received signals and is defined as:

$$R_{x_1 x_2}^{GCC}(p) = \int_{-\infty}^{\infty} \nu(f) \phi_{r_1 r_2}(f) e^{i2\pi f p} df \quad (3.7)$$

where $\psi(f) = \nu(f) \phi_{r_1 r_2}$ is the generalized cross spectrum, $\phi_{r_1 r_2}$ is the cross spectrum and ν is the weighting function. The GCC becomes the frequency domain implementation of the simple cross correlation if the weighting function ν is set a constant with different basis of the weighting function ν , the GCC algorithms takes various other forms e.g. the smoothed coherence transform (SCOT) [5], the phase transform (PHAT) [6], the maximum-likelihood (ML) processor [6] to name a few.

GCC methods are well suited for moderately noisy and non-reverberant environments [7], [22]. They are computationally efficient and provide instantaneous TDE and further modifications have been proposed in [28, 29] to increase the robustness of these methods to noise.

3.2.3 Adaptive Eigenvalue Decomposition based TDE

An estimate of the RIR is important to determine the extent of room reverberation [26, 30]. The adaptive eigen value decomposition (AED) algorithm estimates the time delay in a room reverberant environment [11, 14]. It estimates the channel impulse responses from the source to the two sensors in an iterative manner and finds the time delays by identifying the two direct paths from the estimated

channel impulse responses. The time delay estimate $\hat{\tau}^{AED}$ is given by:

$$\hat{\tau}^{AED} = \arg \max_l |\hat{h}_{l,1}| - \arg \max_l |\hat{h}_{l,2}| \quad (3.8)$$

where $h_{l,1}$ and $h_{l,2}$ are the estimated channel impulse response of the channel from source to the first and the second microphones respectively.

The performance of the most widely used TDE methods that appeared in the literature were compared in [14]. It was shown that due to the fundamental deficiency in the signal model which GCC methods assume, their performance deteriorates in the case of high room reverberations. These methods are based on the ideal signal propagation model and thus are unable to combat the reverberant room environments. In contrast, the adaptive filtering approach of blind channel identification for TDE, cope well with room reverberations. The adaptive estimation of the RIR is computationally intensive and thus suffers from poor tracking capabilities. As a result, the performance degrades in a rapidly varying acoustic environment.

In addition, these TDE techniques suffers from a time resolution problem at low sampling frequencies. The effects of under-sampling on the performance of these algorithms are discussed in [24]. The high sampling frequency requirements makes the localization process not only computationally intensive but it also requires sophisticated hardware. In [14] the author provides a brief overview, performance comparison and problems associated with widely used TDE methods.

The CC requires the acquired data to be sent to a central processing unit for

TDE. This put stress on the communication link as at high sampling rates, more samples have to be communicated from the sensor to the central processing unit.

3.3 Localization Techniques

An acoustic source present in the sound field of microphones can be localized with the knowledge of the time delays measured with pairs of microphones. It is assumed that the speed of sound propagation in the medium in which the acoustic source is present is known. In this chapter we discuss the two broad categories of acoustics source localization methods.

3.3.1 Direction of Arrival (DOA)

Microphone arrays can be used to enhance signals coming from a desired direction and suppress noise or unwanted signals. Another useful functionality of microphone arrays is the direction of arrival (DOA) estimation and the ability to locate the sound source using that information [15]. Consider N identical omnidirectional microphones with known arbitrary geometry as show in Figure 3.1 where d is the inter microphone spacing, ϕ and θ are the azimuth and the elevation angles of the direct line connecting the sound source location and the center of the array (reference microphone). The sound wave field consists of a single far-field wide band source in the presence of additive white Gaussian noise. The far-field assumption is necessary to make sure that the sound wave front arriving at the sensors is planar. Location of the i^{th} sensor in the Cartesian coordinate

system is denoted by r_i . Taking the first microphone as a reference, the signals arriving at the other microphones are the time delayed versions of the signal received at the reference microphone as the sound has to cover some extra distance to reach them. With these time delays we can form a time delay vector given as:

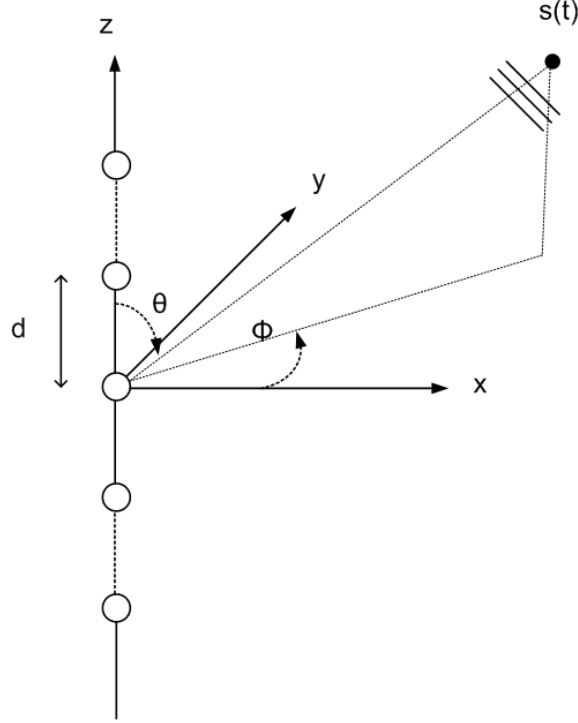


Figure 3.1: Microphone array for DOA estimation

$$\boldsymbol{\tau} = \begin{bmatrix} \tau_{12} & \tau_{13} & \cdots & \tau_{1N} \end{bmatrix} \quad (3.9)$$

where $\tau_{ij} = \tau_j - \tau_i$ is the time delay between microphones i and j . We are interested in finding the direction of arrival vector from which we can easily extract the azimuth and elevation angles. In a 3D scenario, the DOA vector can be written

as:

$$\mathbf{K} = \begin{bmatrix} K_x \\ K_y \\ K_z \end{bmatrix} = \begin{bmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{bmatrix} \quad (3.10)$$

The time delay between a pair of sensors is related to the direction vector by the projection of a distance vector along $\mathbf{K}$ divided by the sound velocity

$$\tau = -\frac{\mathbf{R}\mathbf{K}}{c} \quad (3.11)$$

where $\mathbf{R}$ is the direction vector and c is the sound propagation velocity. $\mathbf{R}$ is given by:

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}_2 - \mathbf{r}_1 \\ \mathbf{r}_3 - \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_N - \mathbf{r}_1 \end{bmatrix} \quad (3.12)$$

where r_i and r_j are locations (in $3D$ plane) of the microphones i and j respectively.

With a microphone array of 3 or more elements, we can find the least square solution of equation (3.11) to estimate the direction vector $\mathbf{K}$ as [15]

$$\begin{aligned} \hat{\mathbf{K}} &= -c (\mathbf{R}^T \mathbf{\Lambda}_\tau^{-1} \mathbf{R})^{-1} \mathbf{R}^T \mathbf{\Lambda}_\tau^{-1} \hat{\boldsymbol{\tau}} \\ &= -c \mathbf{B} \hat{\boldsymbol{\tau}} \end{aligned} \quad (3.13)$$

where $\mathbf{\Lambda}_\tau$ is the covariance matrix of the time delay measurements and the matrix $\mathbf{B}$ is given by

$$B = (\mathbf{R}^T \mathbf{\Lambda}_\tau^{-1} \mathbf{R})^{-1} \mathbf{R}^T \mathbf{\Lambda}_\tau^{-1}$$

$\mathbf{B}$ depends on the array geometry and the covariance matrix of the time delay which under spatially uncorrelated noise assumption is known in advance and can be evaluated offline.

With the measured time delays $\hat{\boldsymbol{\tau}}$, the direction vector $\hat{\mathbf{K}}$ can be estimated. The angle estimates $\hat{\phi}$ and $\hat{\theta}$ can thus be evaluated from the following expressions [15]:

$$\hat{\phi} = \tan^{-1} \left(\frac{\hat{K}_y}{\hat{K}_x} \right) \quad (3.14)$$

$$\hat{\theta} = \cos^{-1} \left(\hat{K}_z \right) \quad (3.15)$$

Accurate time delay estimation between a pair of microphones is important in determining the DOA estimate. A single DOA estimate resolves the source position to a line which makes the estimated angle with a reference axis called bearing line. A minimum of two DOA estimates from spatially separated microphone arrays are required to find the source location in a two dimensional plane. The location of the source is the point of intersection of the two bearing lines as shown in Figure 3.2.

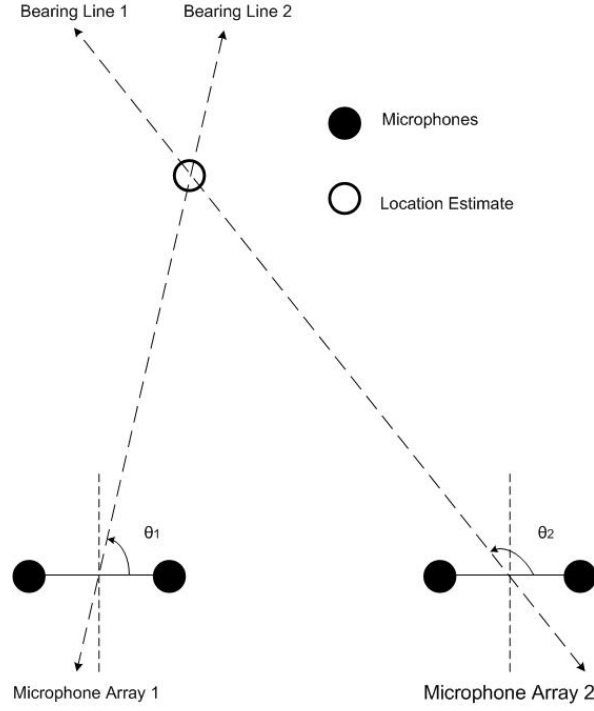


Figure 3.2: DOA based localization, intersection of lines of bearing

3.3.2 Time Difference of Arrival (TDOA)

In this section we consider another type of localization system that also uses time delay estimates to locate the source which is TDOA [16, 17, 19]. Time difference of arrival based source localization is a two-step procedure. In the first stage, the time difference of signal arrival between a pair of microphones is estimated. With the knowledge of the propagation velocity of sound, the estimated TDOA measurement is transformed into range difference measurement from which hyperbolic range difference equation is formed [20]. The second stage utilizes efficient algorithms to produce an unambiguous solution to the hyperbolic equations obtained from multiple pair of microphones [21]. The solution produced by these

algorithms result in the estimated source position. The position information can be used to track an active talker and hence steer the cameras automatically in a video conferencing systems. Further, it can be employed to control the office equipments, such as air conditioner for effective operation. One such method has been presented in [18] which give the location of a person based on TDOA of non speech signals.

We start with an assumption of a single sound source in the field of $N + 1$ spatially separated acoustic sensors at known locations as shown in Figure 3.3. We take the first microphone as a reference located at the origin $\mathbf{m}_0 = [0, 0, 0]^T$ and other microphones at $\mathbf{m}_k$, where $k = 1, 2, \dots, N$. The signal arriving at N non referenced microphones are the time delayed versions of the signal received at the reference. With a known velocity of sound, the range difference d_k between the reference microphone and the k^{th} microphone from the unknown source position $\mathbf{r}$ can be written as

$$\Delta\tau_{k.c} = d_k = \|\mathbf{r} - \mathbf{m}_k\| - \|\mathbf{r}\| \quad (3.16)$$

where $\|\cdot\|$ denotes the Euclidean vector norm. Equation (3.16) is referred to as a hyperbolic equation since for a single range difference measurement, the possible source location forms a hyperbolic curve as shown in Figure 3.4. From equation (3.16) it follows

$$\|\mathbf{r} - \mathbf{m}_k\|^2 = (d_k + \|\mathbf{r}\|)^2 \quad (3.17)$$

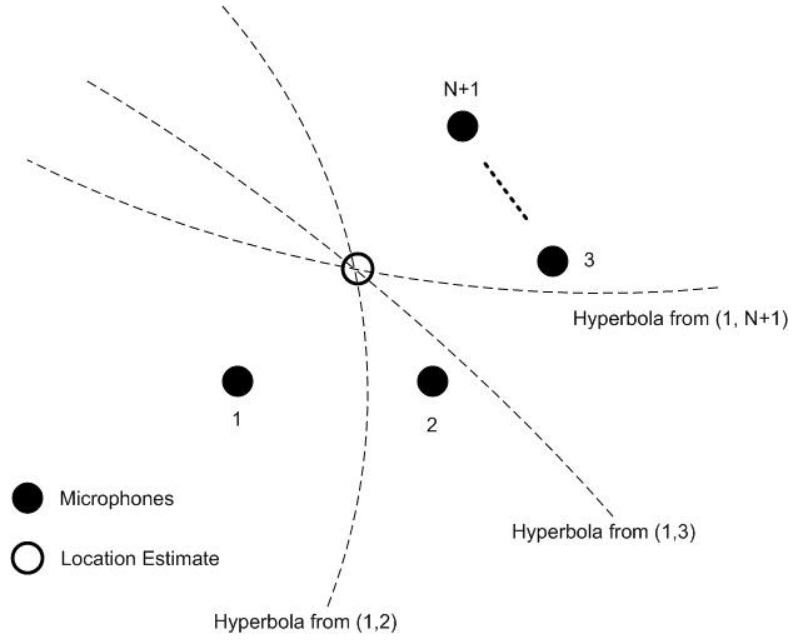


Figure 3.3: TDOA based localization, intersecting hyperbolic curves

from which we get

$$d_k ||\mathbf{r}|| + \mathbf{m}_k^T \mathbf{r} = b_k \quad (3.18)$$

where

$$b_k = \frac{||\mathbf{m}_k||^2 - d_k^2}{2} \quad (3.19)$$

With $N+1$ spatially located microphones we get N range difference measurements, each corresponding to estimated time delays between the reference microphone and N other microphones respectively. These N range difference equations as in

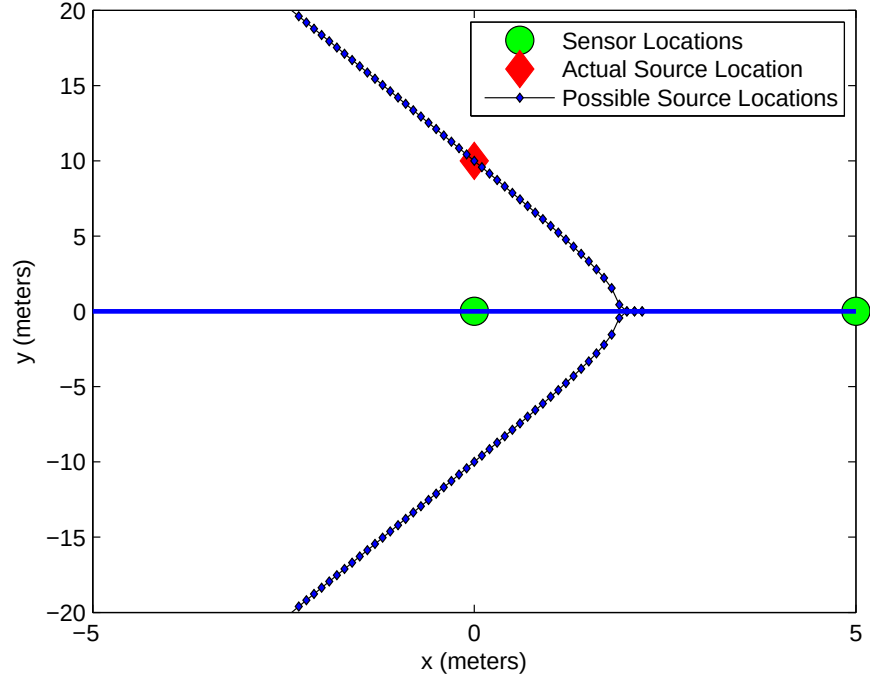


Figure 3.4: Fixed range difference hyperbolic curve

equation (3.18) can be written in a vector form as

$$\mathbf{y} = \begin{bmatrix} \mathbf{R}_0 \\ \mathbf{r} \end{bmatrix}, \quad \Phi = \begin{bmatrix} d_1 & \mathbf{m}_1^T \\ d_2 & \mathbf{m}_2^T \\ \vdots & \vdots \\ d_N & \mathbf{m}_N^T \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix} \quad (3.20)$$

Thus, we may write this system of equations in the following form

$$\mathbf{y}\Phi = \mathbf{b} \quad (3.21)$$

$\mathbf{R}_0$ is the distance of the source from the reference. The method of least squares is applied to find a solution to the system of equations (3.21) for the unknown

source location $\mathbf{r}$. This is obtained from [21]:

$$\hat{\mathbf{y}} = (\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{\Phi}^T \mathbf{b} \quad (3.22)$$

3.4 Summary and Conclusions

This chapter covered various TDE techniques which form the basis of TDE based localization methods. Different signal propagation models starting with the ideal model to more realistic multipath and reverberant models were presented. Simple and advanced TDE methods were explained. TDE based localization methods and their closed form solutions were presented. It was shown that the TDE from a pair of microphones can locate the source by either finding the direction of the incoming source wavefronts (DOA) or from the range difference measurements (TDOA).

CHAPTER 4

TIME DELAY ESTIMATION BASED ON ORTHOGONAL CLUSTERING

Time delay estimation (TDE) is a quite challenging task in reverberant environments. The space between the source-microphone pair separated by some distance can be modeled as a finite impulse response (FIR) filter with a finite response termed as the room impulse response (RIR). The channel taps of this filter represent the multi-path components in the received signal. From the RIR, the direct line of sight signal (DLOS) signal can be identified from the reflected ones. Thus the problem of TDE is equivalent to an accurate RIR estimation at a pair of microphones and by identifying the DLOS components [14, 30, 31].

In this Chapter, we present a new method of TDE from the RIR estimates obtained at a pair of microphones. This method utilizes a recently proposed

compressed sensing (CS) method called orthogonal clustering (OC) [25]. Section 4.1 describes the signal model that we use for RIR estimation. We cover several sparse signal reconstruction methods in Section 4.2. The details of the proposed method of RIR estimation are given in Section 4.3. The procedure of finding the TDE from the estimated RIR is discussed in Section 4.4.

4.1 Signal Model for RIR Estimation

The signal received at the microphone due to a known excitation signal $s(t)$ can be described by the multi-path signal propagation model as discussed in Section 3.1.2. The received signal $r(t)$ is given by:

$$r(t) = \sum_{l=0}^{L-1} \alpha_l s(t - \tau_l) + n(t) \quad (4.1)$$

where L is the number of paths capturing most of the multi-path energy, α_l and τ_l are the scaling magnitude factor and the time shift respectively. $n(t)$ is the additive white gaussian noise (AWGN).

The discrete time representation of the model given in equation (4.1) can be written in a matrix-vector form as

$$\mathbf{r} = \Phi \boldsymbol{\alpha} + \mathbf{n} \quad (4.2)$$

where $\mathbf{r}$ and $\boldsymbol{\alpha}$ are the N length discrete time received and RIR vectors respectively, $\mathbf{n}$ is the AWGN of length N with zero-mean and covariance matrix

$\mathbf{C}_n = \sigma_{\mathbf{n}}^2 \mathbf{I}$. The $N \times N$ matrix Φ is the sensing matrix whose columns consists of N discretized and delayed versions of the source signal $s(t)$, it is given by:

$$\Phi = \begin{bmatrix} s(0 - \Delta) & s(0 - 2\Delta) & \dots & s(0 - N\Delta) \\ s(1 - \Delta) & s(1 - 2\Delta) & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ s((N - 1) - \Delta) & s((N - 1) - 2\Delta) & \dots & s((N - 1) - N\Delta) \end{bmatrix} \quad (4.3)$$

where $i\Delta = i \times \Delta$ ($i = 1, \dots, N$) is the amount of time-shift incurred upon the source signal. The time shift ($\Delta = \frac{1}{F_s}$) $\ll T$, where T is the duration of the source signal $s(t)$ and F_s is the sampling frequency. Let us denote the sampling frequency of the microphones as F_m where $F_m < F_s$ (as we are interested in low sampling rates). The received signal at the microphone is then given by

$$\mathbf{r}_m = \Psi \boldsymbol{\alpha} + \mathbf{n}_m \quad (4.4)$$

where $\mathbf{r}_m$ is the sub-sampled received signal of length M , $\boldsymbol{\alpha}$ is the RIR of length N , and $\mathbf{n}_m$ is the AWGN of the same mean and covariance matrix as of $\mathbf{n}$. The matrix Ψ (of size $M \times N$) is a uniformly sub-sampled version of the sensing matrix Φ where $M \ll N$ and the sub-sampling ratio¹ is $\frac{N}{M} = \frac{F_s}{F_m}$. As $M \ll N$, (4.4) is an under-determined system of equations and thus is ill-posed.

¹Note that it is necessary for the sub-sampling ratio to be less than T to avoid missing the source signal completely.

The impulse response of a room α can be considered as a sparse signal. The sparsity information of the RIR helps in signal reconstruction from less measurements obtained from sampling the received signal at lower rates. In the following section we provide a brief overview of sparse signal reconstruction techniques. After discussing several drawbacks of the existing techniques, we present the details of our proposed method.

4.2 Sparse Signal Estimation Techniques

Sparse signal reconstruction has largely been facilitated since the advent of CS. Being fairly a new area of research, CS has already found a number of applications in signal processing, communication, geophysics, seismology, medical imaging and radar design. As the name suggests, the scheme acquires a signal at compressed sampling rates by randomly projecting it onto a subspace much smaller than the signal dimension.

Most of the naturally occurring signals are sparse in some domain and thus CS techniques are able to reconstruct a signal sampled at Sub-Nyquist rates. This has successfully been achieved in peak to average power reduction in orthogonal frequency domain multiplexing (OFDM), image processing [32], impulse noise estimation and cancellation in power-line communication and Digital Subscribers Lines (DSL) [25], Magnetic Resonance Imaging (MRI) [33], channel estimation in communication systems [34], Ultra-Wideband (UWB) channel estimation [35], Direction of Arrival (DOA) estimation [36] and radar design [37].

The number of measurements M are less than the number of the unknowns N in equation (4.4) which renders the system as under-determined. This implies that an infinite number of $\boldsymbol{\alpha}$'s satisfy equation (4.4). This makes the problem ill-posed and the CS approach of reconstructing a sparse signal is applied. CS solves an ℓ_0 optimization problem using only $M \geq 2K$ measurements. The ℓ_0 minimization problem is NP-hard and its not very practical [38].

An alternate CS approach called convex relaxation, solves a relaxed ℓ_1 optimization problem by linear programming instead of ℓ_0 minimization with a penalty on the number of observations. Given that the sensing matrix $\boldsymbol{\Psi}$ obeys the Restricted Isometry property (RIP), ℓ_1 optimization gives accurate estimates of the sparse signal. Estimators of this type are discussed in [39, 40]. These methods do well in recovering the sparse signals but at the same time they suffer from a number of draw backs. These drawbacks are:

1. Solving an ℓ_1 minimization problem with linear programming is computationally complex. Many greedy algorithms have been proposed for fast recovery [41–43].
2. The structure of the sensing matrix is harmful to these methods as the best estimation results are obtained when the sampling is close to random.
3. They can not make use of a priori statistical information about the signal supports and additive noise.
4. Their performance can not be easily quantified in terms of a mean square

error (MSE). Thus, it can not be compared with the conventional estimation techniques e.g. maximum a posteriori (MAP), minimum mean square error (MMSE), or maximum likelihood (ML).

5. The convex relaxation methods do not exhibit a trade-off between increased computational complexity and improved recovery thus making it impossible to reduce the estimation complexity at the expense of high rate sampling.

These drawbacks motivates us to make use of 1) a priori statistical properties, 2) sparsity information, and 3) the structure of the sensing matrix Ψ , to develop a low complexity sparse signal reconstruction method.

4.3 Proposed Method of RIR Estimation

We begin the discussion of the sparse signal reconstruction algorithm from the system model as given in equation (4.2) where the sparse signal α is modeled as

$$\alpha = \alpha_B \odot \alpha_G \quad (4.5)$$

where α_B entries are i.i.d Bernoulli with success probability p and α_G entries are also i.i.d from some distribution with a marginal pdf $f(\alpha)$. When the supports (indexes with non-zero amplitude) of α are known, we may write equation (4.2) as

$$\mathbf{r}|\mathbf{S} = \Psi_S \alpha_S + \mathbf{n}_S \quad (4.6)$$

where Ψ_S is the sub-matrix formed by the columns $\psi_s : s \in \mathbf{S}$ indexed by the support $\mathbf{S}$.

The estimation problem is of fundamental importance in signal processing and specially in channel estimation [44–46]. An estimate of a random variable can be obtained by using the mean square error (MSE) criteria

$$\underset{c}{\text{minimize}} \quad \mathbf{e} = \mathbb{E} [(\mathbf{r} - c)^2] \quad (4.7)$$

where c is a constant. The MMSE criterion leads to tractable results to estimate $\mathbf{r}$. We consider two cases of estimating a random variable $\mathbf{r}$.

Case I: Estimate α by a constant

Consider a simplest case of estimating the random variable α with a constant c .

$$\begin{aligned} \mathbf{e} &= \mathbb{E} [|\alpha - c|^2] \\ &= \int_{-\infty}^{\infty} (\alpha - c)^2 f(\alpha) d\alpha \end{aligned} \quad (4.8)$$

The best value of c is the value that minimizes the mean square error.

$$\begin{aligned} \underset{c}{\text{minimize}} \quad \mathbf{e} &= \mathbb{E} [|\alpha - c|^2] \\ \frac{d\mathbf{e}}{dc} &= \frac{d}{dc} \int_{-\infty}^{\infty} (\alpha - c)^2 f(\alpha) d\alpha \\ 0 &= -2 \int_{-\infty}^{\infty} (\alpha - c) f(\alpha) d\alpha \end{aligned} \quad (4.9)$$

Solving equation (4.9) for c yields

$$c = \mathbb{E}[\boldsymbol{\alpha}] \quad (4.10)$$

which is the mean of the random variable $\boldsymbol{\alpha}$. The error in estimating a random variable $\boldsymbol{\alpha}$ by a constant equal to the mean $\mathbb{E}[\boldsymbol{\alpha}]$ is the variance of the random variable $\boldsymbol{\alpha}$. This error is given by:

$$\mathbb{E}[(\boldsymbol{\alpha} - \mathbb{E}[\boldsymbol{\alpha}])^2] = \sigma_{\alpha}^2 \quad (4.11)$$

Case II: Estimate α by a function $\mathbf{g}(\mathbf{r})$

Estimating α with a constant c requires α to be directly measurable. However, we observe only a received signal $\mathbf{r}$. Using mean square criterion we find the function $\mathbf{g}(\mathbf{r})$ that minimizes the MSE.

$$\begin{aligned} \mathbf{e} &= \mathbb{E}[|\boldsymbol{\alpha} - \mathbf{g}(\mathbf{r})|^2] \\ &= \int_{-\infty}^{-\infty} \int_{-\infty}^{-\infty} (\alpha - g(r))^2 f(\alpha, r) d\alpha dr \\ &= \int_{-\infty}^{-\infty} \int_{-\infty}^{-\infty} (\alpha - g(r))^2 f(\alpha|r) f(r) d\alpha dr \\ &= \int_{-\infty}^{-\infty} f(r) \left(\int_{-\infty}^{-\infty} (\alpha - g(r))^2 f(\alpha|r) d\alpha \right) dr \end{aligned} \quad (4.12)$$

The optimum function $\mathbf{g}(\mathbf{r})$ is the one that minimizes the mean square error.

$$\underset{\mathbf{e}}{\text{minimize}} \quad \mathbf{e} = \mathbb{E}[|\boldsymbol{\alpha} - \mathbf{g}(\mathbf{r})|^2] \quad (4.13)$$

For each $\mathbf{r}$, we want to minimize the inner integral of equation (4.12) by choosing $\mathbf{g}(\mathbf{r})$ properly. Following from equation (4.9), we conclude that the best estimator of $\boldsymbol{\alpha}$ given the received signal $\mathbf{r}$, in the mean square sense is given by

$$\mathbf{g}(\mathbf{r}) = \mathbb{E}[\boldsymbol{\alpha}|\mathbf{r}] \quad (4.14)$$

4.3.1 MMSE Estimation of a Sparse Signal

In the last section we found that the optimum estimate of $\boldsymbol{\alpha}$ given the observation $\mathbf{r}$ is the MMSE estimate. Knowing the signal supports, the MMSE estimate can be expressed as

$$\begin{aligned} \hat{\boldsymbol{\alpha}}^{MMSE} = \mathbb{E}[\boldsymbol{\alpha}|\mathbf{r}] &= \mathbb{E}[\mathbb{E}[\boldsymbol{\alpha}|\mathbf{r}, \mathbf{S}]] \\ &= \sum_{\mathbf{S}} \mathbb{E}[\boldsymbol{\alpha}|\mathbf{r}, \mathbf{S}] p(\mathbf{S}|\mathbf{r}) \end{aligned} \quad (4.15)$$

where the sum is evaluated over all possible supports set $\mathbf{S}$ of $\boldsymbol{\alpha}$. However, for large N there will be 2^N such sets and it would become hard to evaluate this sum. To reduce in computational complexity we look for ways of approximating equation (4.15). In the following we discuss the way of calculating various terms in equation (4.15).

A) Calculating $\mathbb{E}[\boldsymbol{\alpha}|\mathbf{r}, \mathbf{S}]$

When it is known a priori that $\boldsymbol{\alpha}$ conditioned on its support $\mathbf{S}$ is Gaussian, then from the linearity of the model as in equation (4.6) we know that $\mathbf{r}$ and $\boldsymbol{\alpha}$ are

jointly Gaussian. In this case the $\mathbb{E}[\boldsymbol{\alpha}|\mathbf{r}, \mathbf{S}]$ is equivalent to the linear MMSE estimate of $\boldsymbol{\alpha}$ given $\mathbf{r}$ and $\mathbf{S}$. The posterior pdf $f_{\boldsymbol{\alpha}|\mathbf{r}}(\boldsymbol{\alpha}|\mathbf{r})$ is also jointly Gaussian with a mean $\boldsymbol{\eta}_{\boldsymbol{\alpha}|\mathbf{r}}$

$$\boldsymbol{\eta}_{\boldsymbol{\alpha}|\mathbf{r}} = \mathbb{E}[\boldsymbol{\alpha}] + \mathbf{R}_{\alpha r} \boldsymbol{\Lambda}_r^{-1} (\mathbf{r} - \mathbb{E}[\mathbf{r}]) \quad (4.16)$$

Assuming $\boldsymbol{\alpha}$ and $\mathbf{r}$ to be zero mean, the MMSE estimate of $\boldsymbol{\alpha}$ given $\mathbf{r}$ becomes

$$\begin{aligned} \mathbb{E}[\boldsymbol{\alpha}|\mathbf{r}] &= \boldsymbol{\eta}_{\boldsymbol{\alpha}|\mathbf{r}} \\ \hat{\boldsymbol{\alpha}}^{MMSE} &= \mathbf{R}_{\alpha r} \boldsymbol{\Lambda}_r^{-1} \mathbf{r} \end{aligned} \quad (4.17)$$

where $\boldsymbol{\Lambda}_r$ is the covariance matrix of $\mathbf{r}$

$$\begin{aligned} \boldsymbol{\Lambda}_r &= \frac{1}{\sigma_n^2} \mathbb{E}[\mathbf{r} \mathbf{r}^H | \mathbf{S}] \\ &= \mathbf{I} + \frac{\sigma_\alpha^2}{\sigma_n^2} \boldsymbol{\Psi}_S \boldsymbol{\Psi}_S^H \end{aligned} \quad (4.18)$$

and $\mathbf{R}_{\alpha r}$ is the cross correlation matrix of $\boldsymbol{\alpha}$ and $\mathbf{r}$

$$\mathbf{R}_{\alpha r} = \sigma_\alpha^2 \boldsymbol{\Psi}_S^H \quad (4.19)$$

When $\boldsymbol{\alpha}_S$ is non-Gaussian or its statistics are unknown then $\mathbb{E}[\boldsymbol{\alpha}|\mathbf{r}, \mathbf{S}]$ is hard to find and thus we replace it by the best linear unbiased estimate (BLUE)

$$\boldsymbol{\alpha}^{MMSE} = (\boldsymbol{\Psi}_S \boldsymbol{\Psi}_S^H)^{-1} \boldsymbol{\Psi}_S^H \mathbf{r} \quad (4.20)$$

B) Calculating $p(\mathbf{S}|\mathbf{r})$

The posterior probability $p(\mathbf{S}|\mathbf{r})$ can be obtained from the prior $p(\mathbf{S})$ using Bayes rule

$$p(\mathbf{S}|\mathbf{r}) = \frac{p(\mathbf{r}|\mathbf{S})p(\mathbf{S})}{\sum_{\mathbf{S}} p(\mathbf{r}|\mathbf{S})p(\mathbf{S})} \quad (4.21)$$

According to our model, the elements of $\mathbf{x}$ are obtained from the Bernoulli process with success probability p , so $p(\mathbf{S})$ is given by

$$p(\mathbf{S}) = p^{|\mathbf{S}|}(1-p)^{n-|\mathbf{S}|} \quad (4.22)$$

The only probability that remains to be found is $p(\mathbf{r}|\mathbf{S})$. Here once again we will deal with two cases based on whether $\boldsymbol{\alpha}|\mathbf{S}$ is Gaussian or not.

1. **$\boldsymbol{\alpha}|\mathbf{S}$ is Gaussian:** As mentioned earlier, when $\boldsymbol{\alpha}|\mathbf{S}$ is Gaussian, $\mathbf{r}|\mathbf{S}$ is also Gaussian with zero mean and covariance matrix as given in equation (4.18).

The MAP metric $p(\mathbf{r}|\mathbf{S})$ can be written as

$$p(\mathbf{r}|\mathbf{S}) = \frac{\exp(-\frac{1}{\sigma_n^2}\mathbf{r}^H\boldsymbol{\Lambda}_r^{-1}\mathbf{r})}{\det(\boldsymbol{\Lambda}_r)} \quad (4.23)$$

2. **$\boldsymbol{\alpha}|\mathbf{S}$ is Unknown::** Alternatively if we consider the distribution of $\boldsymbol{\alpha}|\mathbf{S}$ to be as unknown arbitrary, then all we can say about $\mathbf{r}$ is that it is the vector in the subspace spanned by the columns of $\boldsymbol{\Psi}_S$

$$p(\mathbf{r}|\mathbf{S}) \sim \exp(-\frac{1}{\sigma_n^2}\|\mathbf{P}_S^\perp\|^2\mathbf{r}) \quad (4.24)$$

where

$$\mathbf{P}_S^\perp = \mathbf{I} - \Psi_S(\Psi_S^H \Psi_S)^{-1} \Psi_S^H \quad (4.25)$$

Evaluating equations (4.15) and (4.21) over all possible supports (2^N such sets) requires a lot of computations. Instead, we can reduce the search space 2^{S_r} points by finding the most probable supports $\mathbf{S}_r$ of $\boldsymbol{\alpha}$. The convex relaxation tools [38–40] can be used to find the most probable support of the sparse vector $\boldsymbol{\alpha}$. Another method, fast bayesian matching pursuit (FBMP) algorithm is presented in [47] that finds the dominant support and the MMSE estimate of the sparse vector jointly.

The two of the methods of finding the dominant supports of a sparse vector $\boldsymbol{\alpha}$ do not make use of the structural information of the sensing matrix. We will investigate the use of this structure in reducing the computational complexity of the MMSE estimation.

4.3.2 Structure based Bayesian Recovery Approach

The sensing matrix Ψ_S in equation (4.6) has a Toeplitz structure which is encountered in many signal processing applications of channel estimation [34], UWB channel estimation [35], and DOA estimation [36]. The columns of a sensing matrix Ψ_S are not orthogonal since it is a fat matrix ($M \ll N$). However, in the aforementioned applications, we can find the subset of columns which are orthogonal and spans the column space of Ψ_S . These columns can be collected to make a square matrix Ψ_M and group the remaining $N - M$ columns around these

orthogonal columns to form semi-orthogonal clusters.

Let $\mathbf{S}$ be a possible support of $\boldsymbol{\alpha}$. The columns of $\boldsymbol{\Psi}_S$ can be grouped into a maximum of P semi-orthogonal clusters i.e. $\boldsymbol{\Psi}_S = [\boldsymbol{\Psi}_1 \boldsymbol{\Psi}_2 \dots \boldsymbol{\Psi}_P]$. We can then rewrite equation (4.6) as

$$\mathbf{r}|\mathbf{S} = \begin{bmatrix} \boldsymbol{\Psi}_1 \boldsymbol{\Psi}_2 \dots \boldsymbol{\Psi}_P \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_P \end{bmatrix} + \mathbf{n} \quad (4.26)$$

Thus the MMSE of $\boldsymbol{\alpha}$ in equation (4.15) can be written as

$$\begin{aligned} \hat{\boldsymbol{\alpha}}^{\text{MMSE}} &= \begin{bmatrix} \hat{\boldsymbol{\alpha}}_1 \\ \hat{\boldsymbol{\alpha}}_2 \\ \vdots \\ \hat{\boldsymbol{\alpha}}_P \end{bmatrix} = \begin{bmatrix} \mathbb{E}[\boldsymbol{\alpha}_1|\mathbf{r}] \\ \mathbb{E}[\boldsymbol{\alpha}_2|\mathbf{r}] \\ \vdots \\ \mathbb{E}[\boldsymbol{\alpha}_P|\mathbf{r}] \end{bmatrix} \\ &= \begin{bmatrix} \sum_{\mathbf{S}_1} p(\mathbf{S}_1|\mathbf{r}) \mathbb{E}[\boldsymbol{\alpha}_1|\mathbf{r}, \mathbf{S}_1] \\ \sum_{\mathbf{S}_2} p(\mathbf{S}_2|\mathbf{r}) \mathbb{E}[\boldsymbol{\alpha}_2|\mathbf{r}, \mathbf{S}_2] \\ \vdots \\ \sum_{\mathbf{S}_P} p(\mathbf{S}_P|\mathbf{r}) \mathbb{E}[\boldsymbol{\alpha}_P|\mathbf{r}, \mathbf{S}_P] \end{bmatrix} \end{aligned} \quad (4.27)$$

where $\mathbf{S}_i$ is the support set corresponding to the i^{th} cluster. This implies that the MMSE estimate, $\hat{\boldsymbol{\alpha}}^{\text{MMSE}}$ can be obtained in a divide and conquer manner by separately evaluating the MMSE estimates $\hat{\boldsymbol{\alpha}}_i$ corresponding to each cluster.

4.3.3 The orthogonal clustering (OC) Algorithm for RIR Estimation

In the previous section we presented the details of the OC method that combines a priori statistical information, sparsity and structure of the sensing matrix to develop a fast and low complexity sparse RIR reconstruction algorithm. In this section, we describe the steps involved in RIR estimation using OC algorithm. The steps involved in RIR estimation using OC algorithm are shown in Figure 4.1 and explained in the following:

I. Determine Dominant Supports

The correlation of the received signal with the columns of the sensing matrix Ψ gives us the regions where the dominant supports of the sparse RIR vector, α are located.

II. Form Semi-Orthogonal Clusters

The index i with the largest correlation is selected and a cluster of size L is formed with the i^{th} index at the center. The length of the cluster is selected on the basis of the correlation between the columns of the sensing matrix Ψ . Following in this manner, we form P such clusters. The number of clusters can be increased for RIR estimation in an environment with large reverberation time.

III. Determine Dominant Supports within Clusters

Within each cluster, the most probable supports of size $\ell = 1, 2, \dots P_c$ are found. This is done by evaluating the likelihood of all supports of the size $\ell = 1, 2, \dots P_c$ using equation (4.23) or (4.24). The expected value of α given $\mathbf{r}$ is evaluated using equation (4.17) or (4.20). Each cluster is processed independently due to the semi-orthogonality between the clusters. To estimate closely spaced reflections within a cluster, a high value of P_c is selected.

IV. Evaluate an estimate of α

The MMSE estimate of α can easily be evaluated once we have found the dominant supports for each cluster, their likelihoods, and the expected value of α given $\mathbf{r}$.

4.4 Orthogonal Clustering based TDE

The proposed technique for time-delay estimation that utilizes the OC algorithm consists of the following steps:

1. Each microphone acquires a signal at low data rates and transmits it to the centralized workstation. The received signal at the i^{th} microphone is given by

$$\mathbf{r}_i = \Psi_i \alpha_i + \mathbf{n}_i \quad (4.28)$$

where $\mathbf{r}_i$ is the sub-sampled received signal at the i^{th} microphone (M length received signal vector), α_i is the impulse response of the channel between

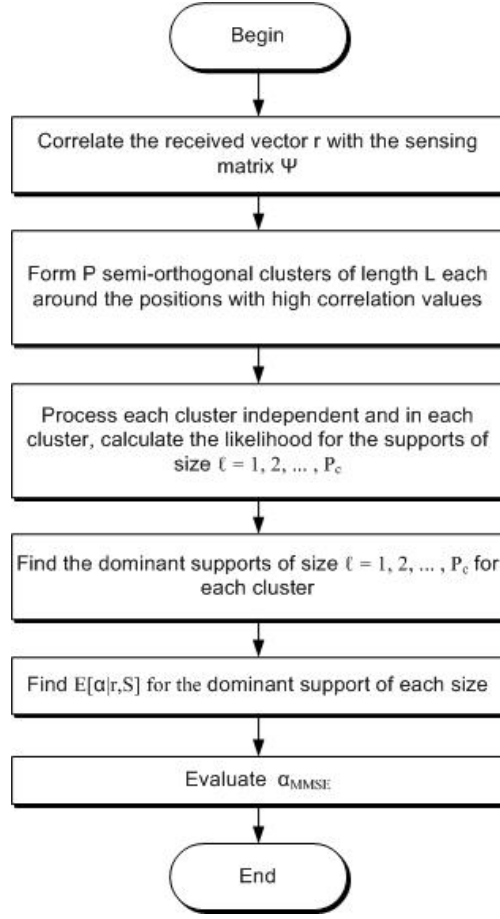


Figure 4.1: Flowchart of the OC algorithm

source and i^{th} microphone (N length RIR vector), and $\mathbf{n}_i$ is the AWGN of the same mean and covariance matrix as of $\mathbf{n}$ in equation (4.2). The matrix Ψ_i (of size $M \times N$) is a uniformly sub-sampled version of the dictionary matrix Ψ where $M \ll N$.

2. The RIR of each microphone is estimated using the OC algorithm [25, 48].

The OC algorithm makes a collective use of the structure present in the sub-sampled Ψ_i matrix and the sparsity information of the RIR signal to

obtain its minimum mean square error (MMSE) estimate given by

$$\hat{\boldsymbol{\alpha}}_i^{MMSE} = \mathbb{E}[\boldsymbol{\alpha}_i | \mathbf{r}_i] = \sum_S \mathbb{E}[\boldsymbol{\alpha}_i | \mathbf{r}_i, \mathbf{S}] p(\mathbf{S} | \mathbf{r}_i) \quad (4.29)$$

3. The DLOS component from each RIR is identified and the time delay is calculated by taking the difference of the time of arrival of the DLOS component of each microphone. With multiple TDEs between a number of microphones and a reference microphone, we may apply localization methods (DOA or TDOA) to locate the acoustic source.

4.5 Summary and Conclusions

In this Chapter, we presented the details of the proposed TDE method based on the OC algorithm. The proposed technique finds the time delay from the RIR estimates obtained at a pair of microphones. The RIR is considered as a sparse signal and several sparse signal reconstruction methods along with their drawbacks were discussed. The new method of RIR estimation based on OC was presented which combines a priori statistical information, sparsity and structure of the sensing matrix for fast and low complex RIR estimation. The procedure of the OC based TDE was summarized in the end.

CHAPTER 5

EXPERIMENTAL SETUP AND RESULTS

This chapter presents an extensive performance analysis of the proposed method through simulations and real hardware experiments. The OC based TDE method is based on realistic sound propagation models. The results obtained favor the applicability of the proposed method in building a robust and practical solution for real time indoor localization system.

In Section 5.1, the hardware components used for the prototype implementation and their features are listed and discussed. The performance of the OC based algorithm for TDE is analyzed in Section 5.2 through simulations conducted for the reverberant environments and under various sub-sampling rates followed by the experimental TDE results. The time delay experiments are extended further for acoustic source localization using multiple microphones. Localization results based on a conventional TDE method and the proposed method are analyzed and

compared for three different microphone configurations in Section 5.3.

5.1 Hardware Components

The prototype implementation of the localization system required setting up an acoustic experiment with various electronic equipment. Among many devices that were used, the most important components were the acoustic sensors, pre-amplifiers, data acquisition modules and the processing unit.



Figure 5.1: Electret microphone mounted on the back side of MAX9814 board.

5.1.1 Acoustic Sensors and Pre-Amplifiers

To acquire an acoustic signal we used electret microphones. The directivity pattern of the microphones is omni-directional. These electret microphones are mounted on the back of the preamplifier board. The MAX9814 Evaluation board was used as shown in Figure 5.1. The evaluation kit uses the MAX9814 low-noise

amplifier IC, an output amplifier, a microphone bias-voltage generator, and automatic gain-control (AGC) internal circuitry. The overall gain of the microphone amplifier is selectable and can be set to 40 dB, 50 dB, or 60 dB. A set of externally connected batteries provide voltage to the amplifier. Figures 5.2 and 5.1 show the top and back side of the preamplifier board, respectively. The output of the preamplifier connects with the data acquisition device via audio jacks and cables.



Figure 5.2: Top side of MAX9814 LNA evaluation kit.

5.1.2 Data Acquisition Device

Two different data acquisition (DAQ) devices are used to facilitate the data acquisition process in various scenarios. The DaqBoard 2000 from Measurement and Computing shown in Figure 5.3 provides a sampling resolution of 16-bit, maximum sampling frequency of 200 kHz, and 16 single-ended analog inputs. The DAQ board connects to the PCI slot of the computer. An external connector

board DBK202 (Figure 5.4) provides an interface between the analog input and the DAQ Board.



Figure 5.3: PCI Data Acquisition Device from MCC.

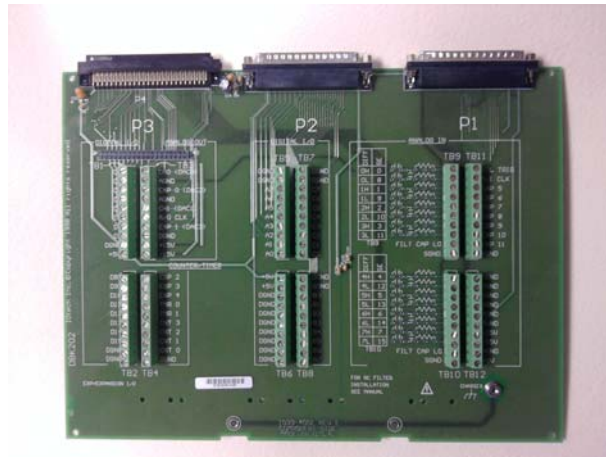


Figure 5.4: Interface board, DBK202 from MCC.

Another device used for data acquisition is a USB module (USB-1608FS) from Measurement and Computing shown in Figure 5.5. It provides 16-bit resolution, 8 single-ended analog input channels and up to 200 kHz overall throughput with a maximum of 50 kHz sampling rate for any channel. The main purpose of using



Figure 5.5: USB Data Acquisition Device from MCC.

the USB based DAQ device was the mobility it provided for the data acquisition process in both indoor and outdoor environments.

5.1.3 Processing Platform

The DAQ device is interfaced with a computer with high end processing capabilities. It uses an INTEL Quad-Core, 2.4 GHz processor and is complemented with a high physical memory of 8 Gbytes. The data acquisition tasks and signal processing for source localization is done using MATLAB. The data acquisition toolbox and a device adaptor (provided by the device manufacturer) are installed within MATLAB. The adaptor provides an interface between the data acquisition toolbox and the hardware driver. PCI-DAQ device uses 'iotech' adaptor while USB-DAQ device uses 'mcc' adaptor.

5.2 Time Delay Estimation Results

Time delay based source localization methods depend on accurate TDE which in turn rely on accurately identifying direct path components from the RIR. In the following, we analyze the performance of our proposed TDE method by both simulations and experiments.

5.2.1 TDE Simulation Results

The performance of the proposed TDE method will be analyzed in simulations by creating a virtual room environment. The impulse response of the channel between a source-microphone pair placed in a room with some length, width, and height is obtained using an image-source model as presented in [26]. Figure 5.6 shows the shoe box model of a room with dimensions $8 \times 6 \times 3$ meters ($x \times y \times z$).

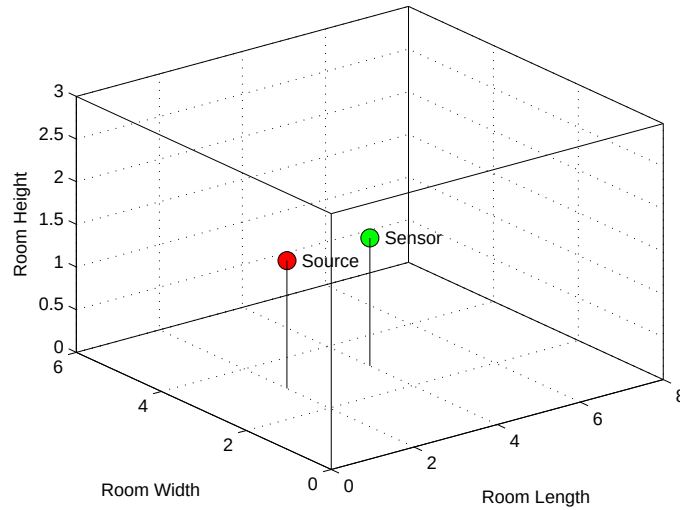


Figure 5.6: Shoe box room model of dimensions $8 \times 6 \times 3$ meters.

Two reverberant environments are considered with different extent of reverber-

ations. A dense reverberant environment is created by setting high wall reflection coefficients: $[x_1 = 0.75, x_2 = 0.75, y_1 = 0.8, y_2 = 0.8, z_1 = 0.85, z_2 = 0.9]$. For a less reverberant environment the reflection coefficients are set low: $[x_1 = 0.2, x_2 = 0.2, y_1 = 0.3, y_2 = 0.25, z_1 = 0.3, z_2 = 0.5]$. The reverberation time T_{60} is set to 0.25 sec.

The sampling frequency is initially set to 10 KHz. Figure 5.7a shows a typical impulsive source signal (a toy gun shot). The room impulse response generated using the model in [26] is shown in Figure 5.7b. The signal received at the microphone is the convolution of source signal with the RIR as shown in Figure 5.7c. The sparsity of the RIR is observable in Figure 5.7b.

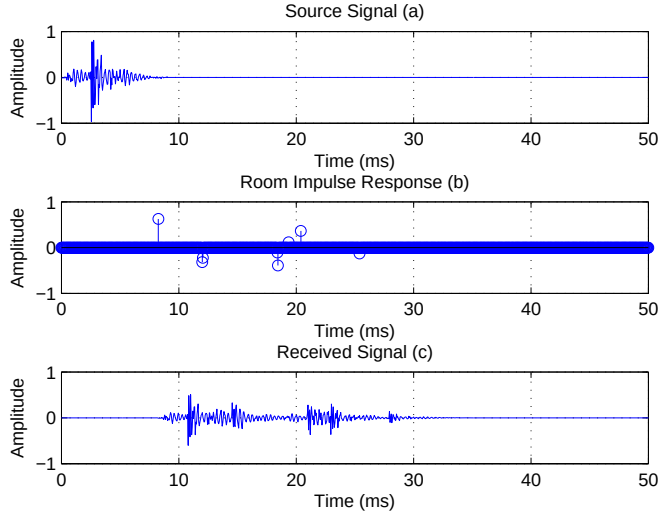


Figure 5.7: (a) Typical impulsive acoustic source signal, (b) Room impulse response, (c) Received signal at the microphone.

An example of the RIR estimation for a low reverberant room environment is shown in Figure 5.8. The OC algorithm is applied for sparse RIR estimation. The direct path signal and most of the reflected signals are correctly estimated

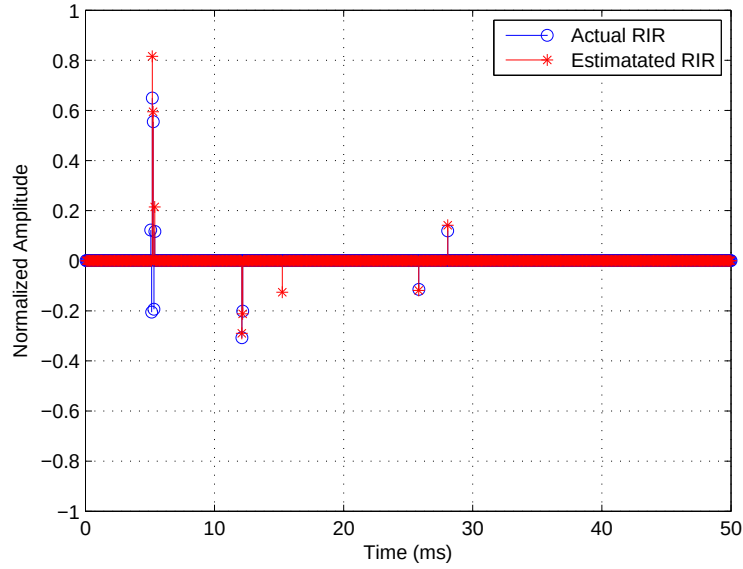


Figure 5.8: Example of simulated RIR estimation (low reverberations).

while few early reflections are missed. Figure 5.9 shows an example of a dense reverberant environment. The OC method of sparse RIR reconstruction is applied. Here, the direct path signal is accurately estimated as well as a number of early and late reflections.

The simulation was run for 500 iterations. For each iteration, the position of the microphones and the acoustic source are randomly varied within the room boundaries and RIR is generated using image source model in [26]. The OC algorithm is applied for TDE and mean square error (MSE) in time delay is evaluated for different sub-sampling rates. The performance is demonstrated in Figure 5.10 for two SNR values of the impulsive source, 30 dB and 40 dB. It can be seen that for the 40 dB case, the MSE at low sub-sampling rates is quite low and the performance degrades gradually for higher sub-sampling rates while there is a significant degradation in performance for sub-sampling rate greater than 2

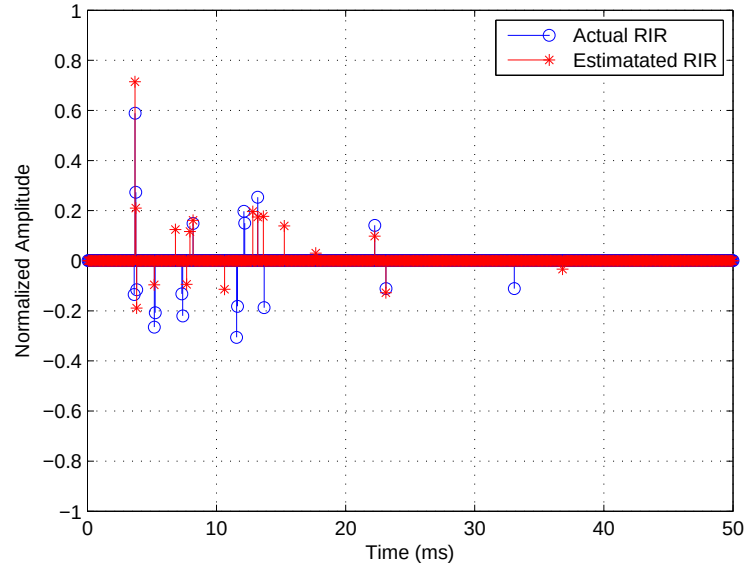


Figure 5.9: Example of simulated RIR estimation (dense reverberation).

in the 30 dB case.

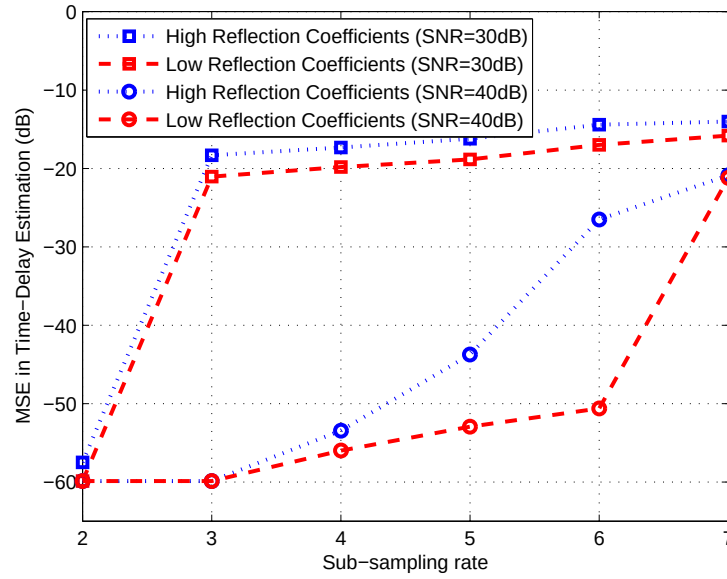


Figure 5.10: MSE in time-delay estimation using the OC algorithm for different sub-sampling rates at SNR = 30 dB and 40 dB.

With less measurements (high sub-sampling factor), the performance of the OC

based RIR reconstruction degrades which eventually resulted in increased MSE in TDE. Similarly, the OC algorithm requires more measurements to estimate the sparse RIR at low values of SNR. Thus, at equal sub-sampling factors the MSE in TDE is less for the SNR value of 40 dB as compared to 30 dB.

5.2.2 TDE Experimental Results

Figure 5.11 shows the actual hardware setup for TDE in a hall room of dimensions $8 \times 6 \times 3$ meters. The microphones are secured with metallic stands placed 100 cm apart. The electret microphones are mounted on a printed circuit board (PCB) with appropriate electronics. Each PCB has a MAX 9814 low-noise amplifier (LNA) IC whose gain is set to 40 dB. With this gain, we get an $SNR \geq 30$ dB at the output of the LNAs which are connected to a 16-bit, 8 channel data acquisition (DAQ) device via audio jacks and cables. The DAQ communicates with a PC through the data acquisition tool box within MATLAB. A toy gun was used as an impulsive source with a duration of approximately 10ms as shown in Figure 5.7. The dictionary matrix Ψ (equation (4.3)) required for the RIR estimation using the OC algorithm is constructed by averaging several instances of the impulsive source signal at $F_s = 16$ KHz.

The first step in finding the time delay based on the proposed OC method is to estimate the sparse RIR. To illustrate this intermediate step of RIR estimation, we discuss two instances of the RIR estimated for different microphone positions inside the room. In the first case, the microphone is placed at the room center and with the source lying close to it. The estimated RIR shows the direct path, early reflections along with few late reflections as shown in Figure 5.12. Thus, we observe that the effect of room reverberations is quite low when microphone is placed away from sound obstacles (room walls). Another example of RIR estimation is conducted with the microphone placed close to the room walls. The

estimated RIR contains a direct path along with number of reflections as shown in Figure 5.13. The presence of the walls close to the microphone builds dense reverberations. This effect is apparent in the high number of room reflections of the estimated RIR.

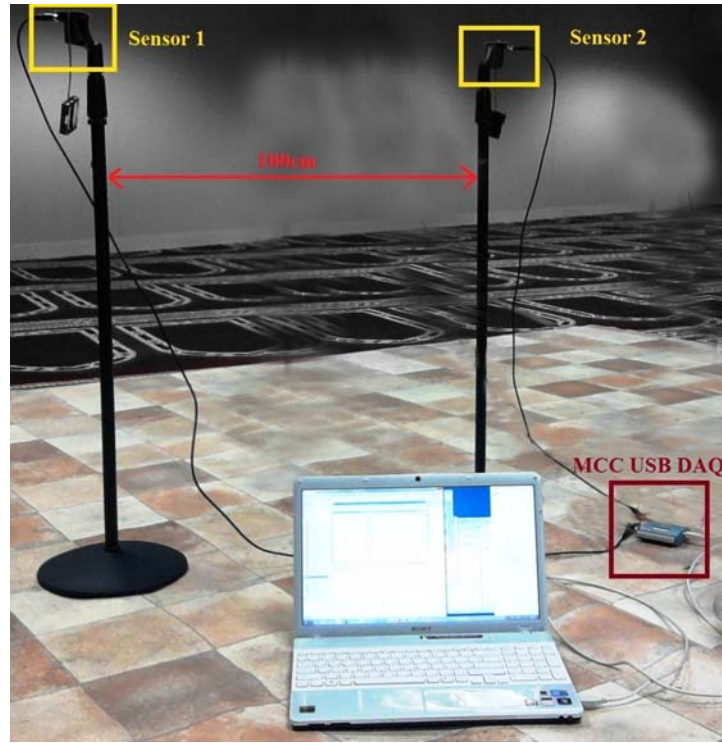


Figure 5.11: Experimental setup for indoor TDE.

The real time functionality of the algorithm for TDE has been verified by placing the source at known locations around the microphones and acquiring the source signal at various sampling rates. Table 5.1 shows the time delays corresponding to three known source locations:

1. Case I: Source positioned at a point on the line that passes through the two microphones.

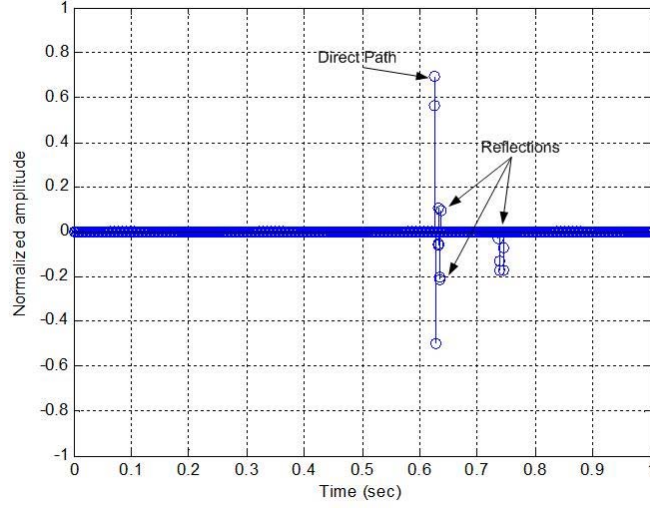


Figure 5.12: Experimental RIR estimate with microphone at the room center.

2. Case II: Source positioned close to microphone 1, on the vertex of an isosceles triangle formed by the microphones and the source.
3. Case III: Source positioned in the middle of the line joining the two microphones.

The corresponding estimated time delays using CC and the proposed algorithm for three sampling rate values of 16KHz, 8KHz and 4KHz are shown in Table 5.1. It can be seen that the proposed algorithm gives closer time delay estimates when compared to CC in the case of low sampling rates. This demonstrates the superior performance of the OC based technique used in [31]. In addition, the run time needed to provide TDE of the proposed algorithm are faster than CC at low sampling rates, which gives it a computational advantage as well.

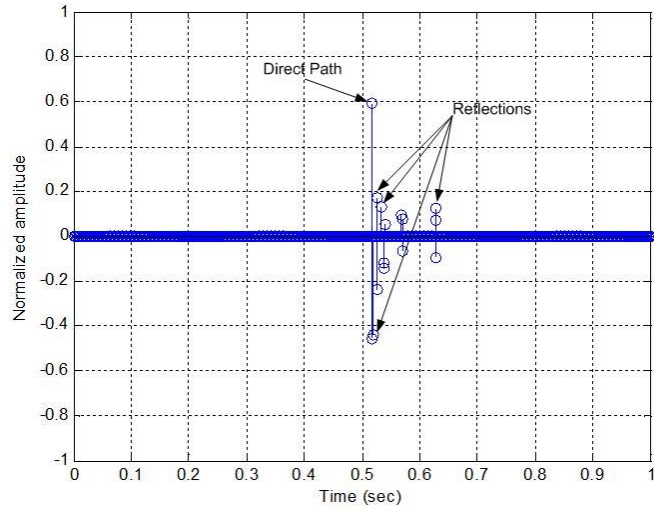


Figure 5.13: Experimental RIR estimate with microphone close to the wall.

Table 5.1: Comparison of time-delay estimates obtained using CC and the proposed technique based on OC algorithm.

	Freq. (KHz)	True TD (ms)	TDE CC (ms)	Run Time CC (sec)	TDE OC (ms)	Run Time OC (sec)
Case I	16	2.941	2.875	66	2.875	70
	8	2.941	3.125	17	2.875	7
	4	2.941	2.250	4.5	3.312	3.5
Case II	16	1.218	1.250	66	1.250	70
	8	1.218	0.875	17	1.312	7
	4	1.218	1.000	4.5	1.250	3.5
Case III	16	0.000	0.000	66	0.000	70
	8	0.000	0.250	17	0.000	7
	4	0.000	0.750	4.5	0.187	3.5

5.3 Source Localization Results

As discussed earlier, the time delay measured with a pair of microphones contains the essential source location information. In this section we present the results of time delay based localization methods. The inherent limitations of DOA based localization method that we face, specifically for indoor environment and under low rate sampling are initially discussed. We apply the TDOA method for localization as it suits our application. Results of several localization experiments for three different microphones geometry using OC based TDE will be shown.

5.3.1 DOA based Source Localization

The TDE is an integer multiple of the sampling period T_s . The estimated delays suffer from poor time resolution at low sampling frequencies (large sampling period) [14]. Such low time resolution is not suitable for DOA based localization scheme where the spatial resolution between the microphones (inter microphone spacing) is also low. Consider a setup for DOA estimation consisting of an array of two microphones separated by a small distance d and a source that lies in the far field of the array. The bearing angle θ is related to the time delay τ by equation (3.11), which for 1D microphone array is expressed as [15]

$$\theta = \cos^{-1} \left(\frac{c\tau}{d} \right) \quad (5.1)$$

where c is the sound propagation velocity.

The angular resolution of an array determines the number of different DOA measurements between 0 and π . It directly corresponds to the sampling frequency F_s (sampling time T_s). For instance, a microphone array with a 30 cm inter microphone spacing, can provide only 14 distinct DOA measurements at the sampling frequency $F_s = 16$ KHz, 7 measurements at $F_s = 8$ KHz and 4 measurements at $F_s = 4$ KHz between 0 and π respectively.

Increasing the sampling rate improves the TDE resolution which in turn lead to a higher DOA resolution [14]. However, this approach will increase the complexity of both the TDE algorithm and hardware. Other ways to improve the DOA resolution is by increasing microphone spacing d which will increase the size of the array. Also the far field requirement for DOA estimation makes it hard to implement in a room environments with a limited space. Moreover, large microphone spacing may cause spatial aliasing. It may also lead to higher computational complexity since more data has to be acquired for extended time durations to account for the larger delay range.

5.3.2 TDOA based Source Localization

The drawbacks of DOA method for indoor localization made us focus on using the TDOA based localization method. The TDOA based method does not make the far field assumption (i.e. plane waves arriving at microphones). Moreover, the microphones can be separated by large distances which provide higher spatial resolution for TDE at low sampling frequencies. In the following we consider three

different microphone geometries. For each geometry, we apply the CC and OC algorithm for TDE and apply the TDOA method for $2D$ source localization.

1. Microphone Configuration I

The microphone configuration shown in Figure 5.14 consists of a $2D$ microphone array placed in the xy plane. Microphone $M1$ placed at the center of the array, is taken as a reference of the array. Other microphones $M2$, $M3$, and $M4$ are placed at an equidistant spacing of $d = 100cm$ from the reference shown in Figure 5.14. The TDOA based localization in $2D$ plane requires a minimum of two time delay estimates obtained from three microphones. However, adding an extra microphone $M4$ gives a redundant time delay measurement which can be used to improve the location estimate as discussed in Section 3.3.2.

The experimental setup for microphone configuration I is shown in Figure 5.15. The setup is made on a 4×4 m floor mat placed in the room center. The floor mat has a printed grid of 20×20 cm which helps in microphones and source placement. The acoustic sensors are secured on top of metallic stands raised to a certain height. The height of the stands form a $2D$ plane in which the acoustic source lies (a person firing a gun). Sensor 1 serves as the reference node and is placed at the origin $(0,0)$. Sensor 2 is placed at $(100,0)$ and Sensor 4 at $(-100,0)$ along the horizontal axis, while Sensor 3 is placed at $(0,100)$ along the vertical axis of the xy plane. Outputs of the pre-amplifiers are connected via audio leads and jacks to four different channels on the USB data acquisition device. MATLAB is

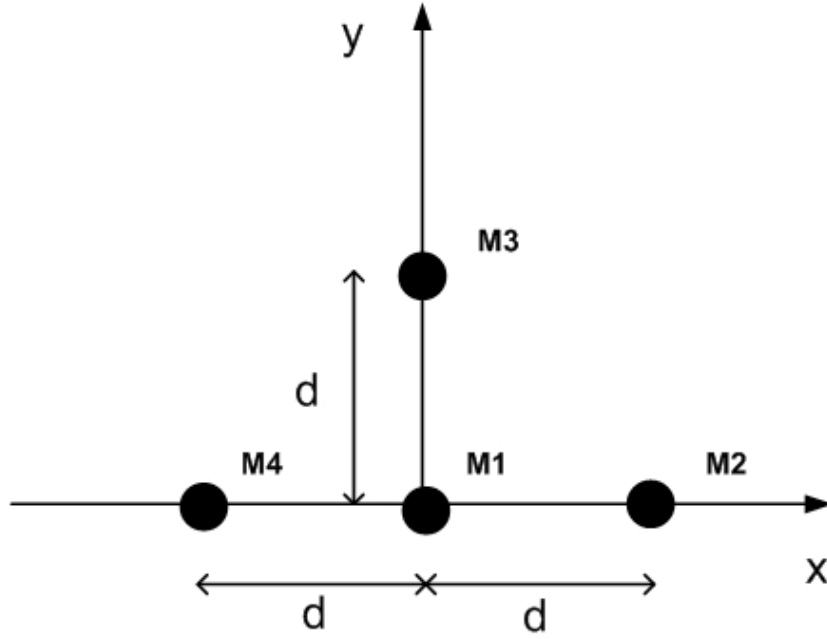


Figure 5.14: Microphone configuration I for source localization.

used for data acquisition and further signal processing for the source localization algorithm.

The CC and OC based methods of TDE have been applied on the data acquired for different source locations. In order to assess the consistency of both algorithms, five instances of the source signal are acquired at each source location at sampling rates of 4, 8, and 16 KHz respectively. TDOA based source localization is used to determine the source location based on the estimated time delays.

The estimated locations using the proposed OC method with sub-sampling, for the source positioned at various locations in the vicinity of the acoustic sensors are shown in Figures 5.16 - 5.20. The figures show the position of the acoustic

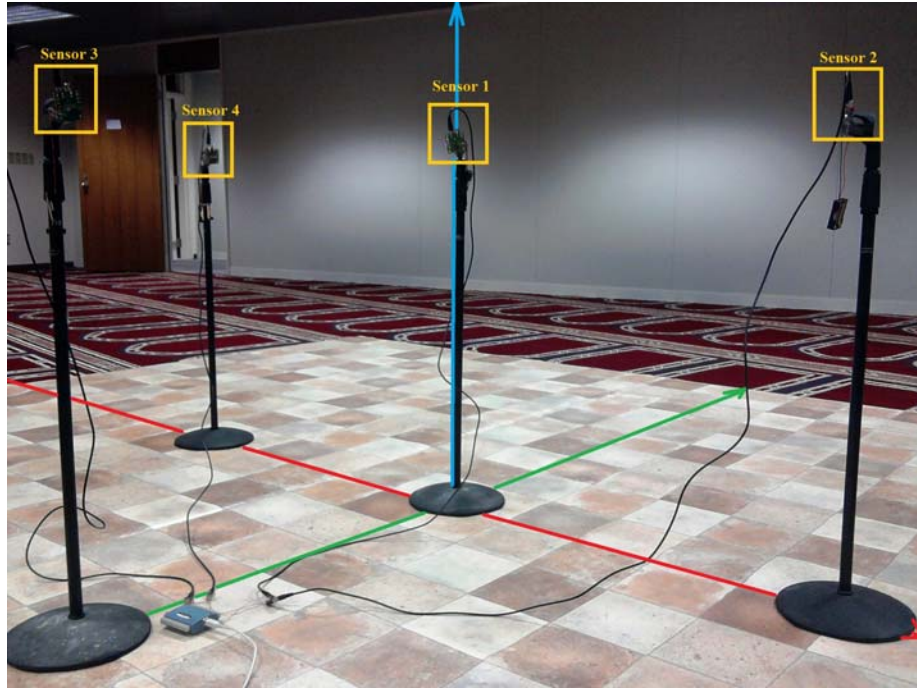


Figure 5.15: Experimental setup of microphone configuration I.

sensors (filled squares), actual (hollow circles) and the estimated source locations (stars for 8 KHz and diamond for 4 KHz). The plot area represents the region of the floor mat. From the figures, it can be seen that the source locations for the 8 KHz measurements lie in close proximity of the true location. The location estimates for the 4 KHz measurements are observed to be lying relatively farther from the true source location than with the 8 KHz measurements. This shows the trade off that exists between the accuracy of location estimates and the sub sampling rates as shown earlier in Figure 5.10.

The experimental results of the CC and OC based time delay and source location estimates for the Microphone configuration I are also tabulated in Ta-

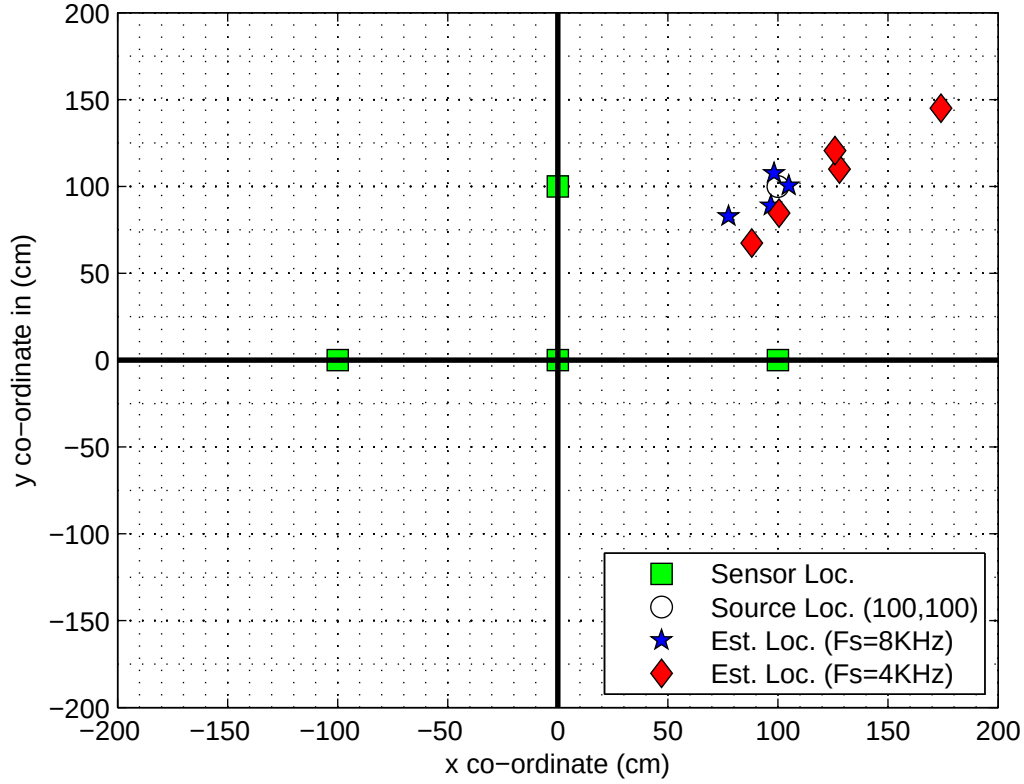


Figure 5.16: Mic. configuration I, localization results for source at(100,100).

bles 5.2–5.6. Each table provides the results of the estimated time delays, source location, and computation time of both TDE techniques for a single source location sampled at three different sampling rates. In the following, we discuss our observations based on the tabulated results:

- Close time delay and location estimates are observed with CC based method for the source signal sampled at 16 KHz. On the other hand, OC could not be applied to 16 KHz measurements as it exceeded the available physical memory (8 Gbytes) of the processing unit to load the dictionary matrix and the data acquired from the acoustic sensors.

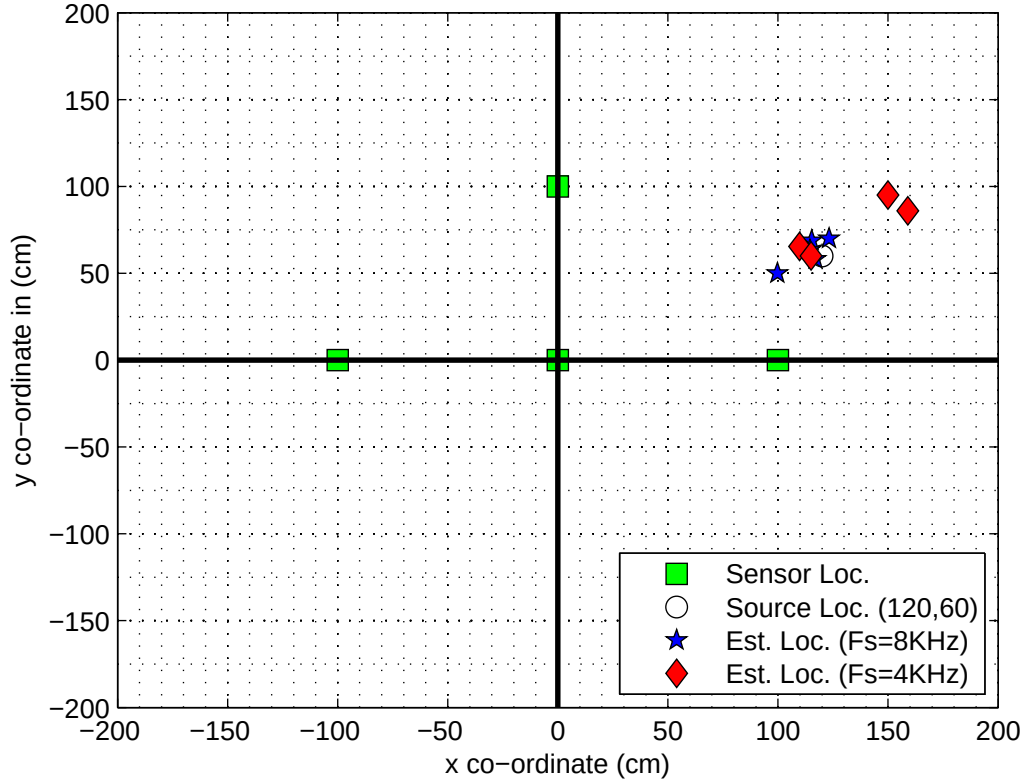


Figure 5.17: Mic. configuration I, localization results for source at (120,60).

- The benefit of the proposed method becomes obvious when source signal is sub-sampled. With the 8 KHz measurements, the proposed OC based TDE method gives better and more consistent location estimates as compared to the CC which fails several number of times to locate the source in an acceptable region (4×4 meters). Figure 5.21 shows that CC fails¹ more than 50% of the times while the failure rate of OC is less than 30%. The OC algorithm computes TDE in 7 sec which is much less than 17 sec taken by CC.
- Sampling the source signal at even lower rate of 4 KHz deteriorates the per-

¹Localization Failure is the ratio of the number of times the algorithm fails to locate the source in an acceptable region to the total number of measurements conducted at a specific F_s .

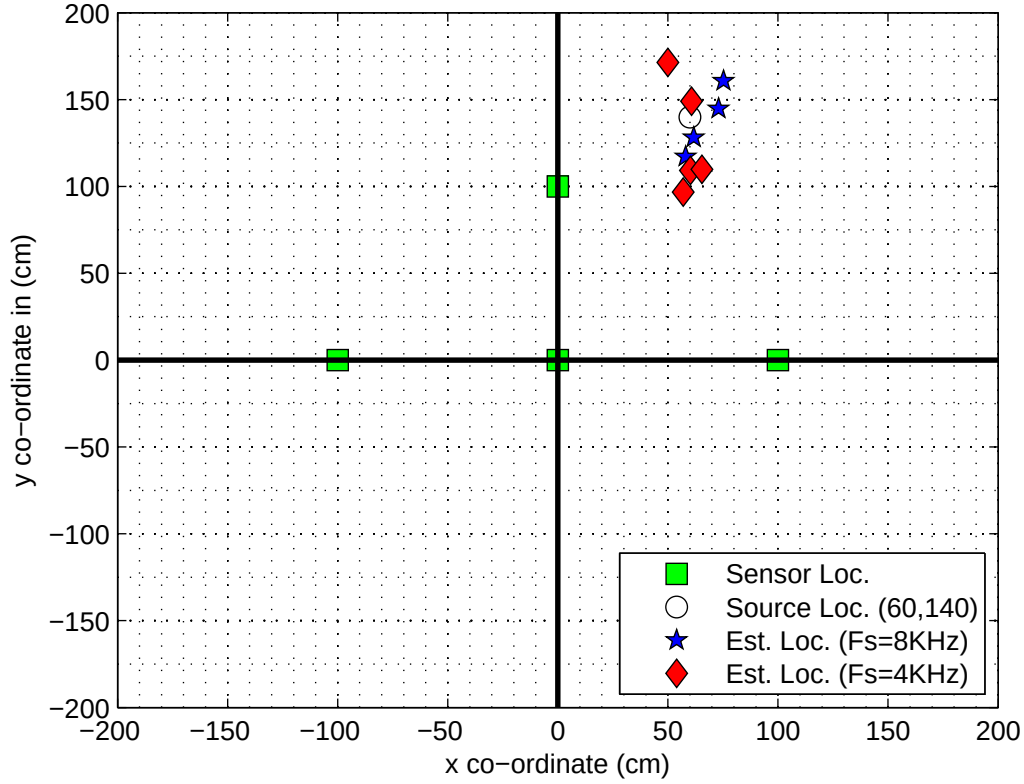


Figure 5.18: Mic. configuration I, localization results for source at (60,140).

formance of the CC method to a much larger extent, resulting in an increased percentage of localization failures up to 85%. In contrast, OC gives acceptable location estimates in many cases with relatively less percentage, 40% of localization failures. The computation time of OC is 3.5 sec which is less than 4.5 sec taken by CC.

It must be noted that the acoustic source used is not an ideal point source, but in fact it is a person exciting an acoustic source while standing at the true location. Hence, we expect the location estimates to lie within an area around the source (± 25 cm).

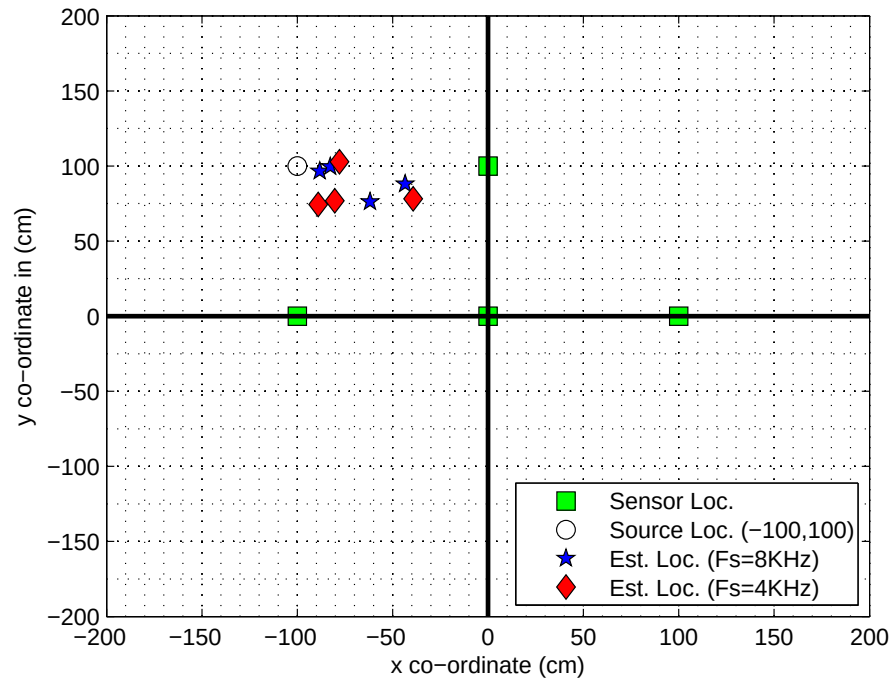


Figure 5.19: Mic. configuration I, localization results for source at (-100,100).

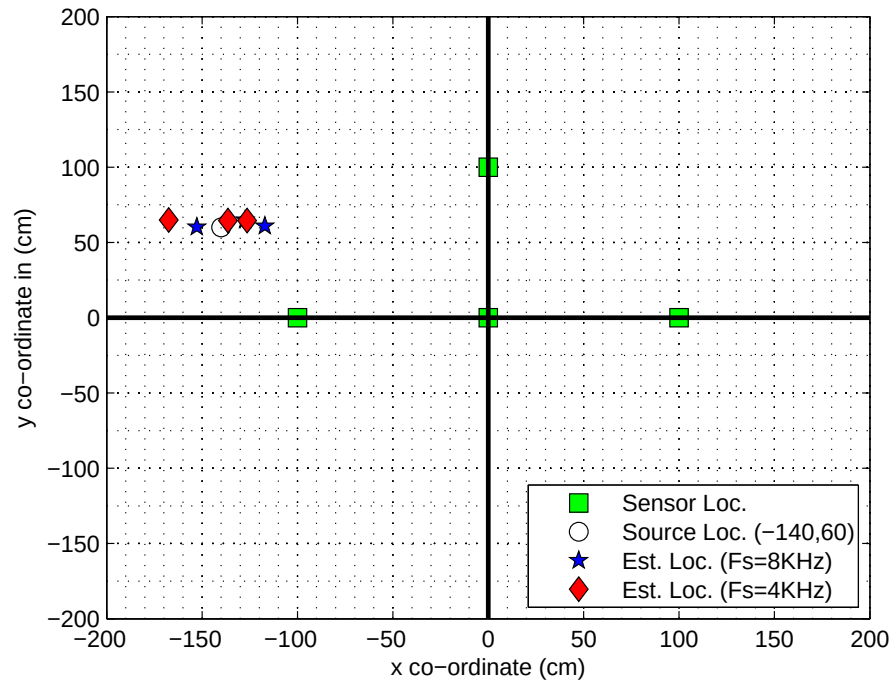


Figure 5.20: Mic. configuration I, localization results for source at (-140,60).

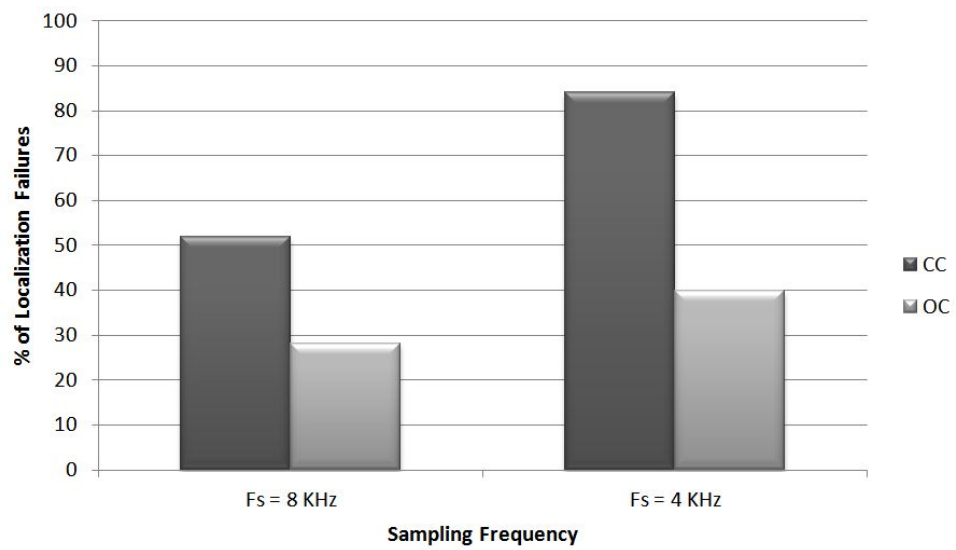


Figure 5.21: Localization failures in CC and OC for microphone configuration I.

Table 5.2: Mic. configuration I results, for the source location of (100,100).

Fs	True Time Delay	Est. Tau CC	Est. Source Location CC	Run Time CC	Est. Tau OC	Est. Source Location OC	Run Time OC
KHz	msec	msec	cm	sec	msec	cm	sec
16	-1.2183 -1.2183 2.4172	-1.2500 -0.9375 2.4375	95.5770 85.8761	~66	N.P.	N.P.	N.P.
		-1.1875 -0.9375 2.5000	92.5913 84.9794				
		-1.1875 -0.9375 2.4375	86.9566 90.7633				
		-1.1875 -1.1875 2.3125	97.5482 97.5482				
		-1.3750 -1.1875 1.5625	108.8394 102.1028				
8	-1.2183 -1.2183 2.4172	-1.8750 -0.8750 7.5000	134.3960 94.4423	~17	-1.1250 -1.3750 2.3125	98.2323 107.6000	~7
		-1.5000 -1.0000 -0.1250	111.0789 93.6093		-0.8125 -1.0000 2.4375	77.5116 82.7767	
		-1.2500 -1.3750 1.7500	106.2121 110.8399		-1.2500 -1.0000 2.5000	96.8472 88.9227	
		-1.3750 -1.8750 2.0000	145.0245 174.1601		-1.8125 -1.8750 1.7500	L.F.	
		-1.2500 -1.3750 2.6250	106.2121 110.8399		-1.3125 -1.1875 2.4375	104.8504 100.4845	
4	-1.2183 -1.2183 2.4172	-1.0000 -1.0000 2.2500	85.1748 85.1748	~4.5	-1.6250 -1.1875 2.5000	127.9521 109.9679	~3.5
		-1.7500 -0.7500 1.7500	118.8219 83.8301		-1.3750 -0.8750 2.5625	100.5135 84.6737	
		-1.5000 0.2500 3.0000	85.4607 41.5611		-1.5000 -1.3750 2.4375	125.8946 120.5635	
		-5.0000 -0.7500 1.7500	L.F.		-1.8125 -1.3750 2.4375	174.6575 145.2548	
		-0.7500 -0.5000 1.5000	71.3549 64.9591		-1.2500 -0.5000 2.6250	88.1248 67.4174	

N.P: Not Possible, L.F: Localization Failure.

Table 5.3: Mic. configuration I results, for the source location of (120,60).

Fs	True Time Delay	Est. Tau CC	Est. Source Location CC	Run Time CC	Est. Tau OC	Est. Source Location OC	Run Time OC
KHz	msec	msec	cm	sec	msec	cm	sec
16	-2.0858 -0.2257 2.7609	-2.0625 -0.1250 2.6875	114.7356 55.3232	~66	N.P.	N.P.	N.P.
		-2.3750 -0.2500 2.6875	145.3617 63.1087				
		-2.6250 -0.2500 2.0625	189.3225 66.7007				
		-1.9375 0.1250 2.6875	101.4815 45.1885				
		-2.0625 0.1250 2.3125	106.4661 44.9973				
8	-2.0858 -0.2257 2.7609	-2.0000 0.3750 5.0000	97.6042 35.9264	~17	-2.0000 -0.4375 2.6875	123.2663 69.9782	~7
		-2.2500 0.3750 3.0000	106.4129 34.9082		-2.0625 -0.1875 2.7650	117.0727 58.1295	
		-1.5000 0.1250 2.6250	87.2184 45.7244		-1.8125 0.0000 2.7500	99.8009 50.0000	
		-2.2500 -0.2500 2.6250	133.1186 62.1254		-1.8750 -0.4375 2.6875	115.2950 68.8706	
		-1.6250 0 2.6250	93.1461 50.0000		-1.9375 -0.3125 2.7500	114.5015 63.3386	
4	-2.0858 -0.2257 2.7609	-2.2500 -1.0000 3.0000	204.1984 125.7576	~4.5	-2.7500 -0.5625 2.9875	L.F.	~3.5
		-2.0000 3.5000 2.2500	L.F.		-2.2500 -0.6250 2.6875	159.2642 86.4700	
		-2.7500 -0.2500 3.0000	243.4423 71.1982		-2.0625 -0.8125 2.6250	150.0715 95.5162	
		-1.2500 3.2500 4.7500	L.F.		-1.8125 -0.3750 2.6875	109.7932 65.4868	
		-0.7500 0.0000 2.7500	68.3420 50.0000		-2.0000 -0.2500 2.7500	115.8076 60.7372	

N.P: Not Possible, L.F: Localization Failure.

Table 5.4: Mic. configuration I results, for the source location of (60,140).

Fs	True Time Delay	Est. Tau CC	Est. Source Location CC	Run Time CC	Est. Tau OC	Est. Source Location OC	Run Time OC
KHz	msec	msec	cm	sec	msec	cm	sec
16	-0.1974 -2.3590 1.7732	-0.2500 -2.3125 1.6250	62.5809 138.8058	~66	N.P.	N.P.	N.P.
		-0.2500 -1.9375 2.0000	60.4858 112.3669				
		-0.3750 -1.9375 1.6250	66.3058 116.7484				
		-0.5625 -2.0000 4.4375	76.8934 129.0034				
		-0.2500 -2.3125 0.6250	62.5809 138.8058				
8	-0.1974 -2.3590 1.7732	0.0000 -2.5000 1.6250	50.0000 141.1880	~17	-0.1875 -2.0625 1.9375	58.1295 117.0727	~7
		-0.2500 -2.1250 1.6250	61.3690 123.6067		-1.4375 -2.3125 1.7500	L.F.	
		0.0000 -2.1250 1.3750	50.0000 113.5233		-0.4375 -2.2500 1.8750	73.0146 144.7892	
		-0.1250 -1.8750 4.2500	54.9670 105.5393		-0.2500 -2.1875 1.8750	61.7254 128.0996	
		-0.3750 -2.2500 3.8750	69.1688 140.6282		-0.4375 -2.3750 1.7500	75.2907 160.6951	
4	-0.1974 -2.3590 1.7732	-0.2500 -1.7500 2.2500	59.7859 103.3290	~4.5	-0.2500 -1.8750 2.0625	60.2365 109.1628	~3.5
		-0.5000 -0.2500 2.2500	62.9782 56.8503		-0.3750 -1.8125 2.0625	65.4868 109.7932	
		-0.5000 -3.0000 4.5000	L.F.		-0.1875 -1.6250 2.1250	56.9557 96.7811	
		2.5000 -2.5000 0	L.F.		-0.2000 -2.4375 1.6875	60.7092 148.9952	
		2.7500 -2.5000 2.2500	L.F.		0 -2.6875 1.5000	50.0000 171.3807	

N.P: Not Possible, L.F: Localization Failure.

Table 5.5: Mic. configuration I results, for the source location of (-100,100).

Fs	True Time Delay	Est. Tau CC	Est. Source Location CC	Run Time CC	Est. Tau OC	Est. Source Location OC	Run Time OC
KHz	msec	msec	cm	sec	msec	cm	sec
16	2.4172 -1.2183 -1.2183	2.3125 -1.3750 -1.3125	-82.6551 99.5696	~66	N.P.	N.P.	N.P.
		2.3125 -1.3125 -1.5625	-75.7456 93.8690				
		2.3125 -1.6250 -1.2500	-144.4132 149.6317				
		2.3125 -1.3125 -1.5625	-75.7456 93.8690				
		2.3125 -1.1250 -1.3125	-61.1917 81.7409				
8	2.4172 -1.2183 -1.2183	4.2500 -1.2500 -1.2500	L.F.	~17	3.1250 -0.6250 -0.6250	L.F.	~7
		2.1250 -1.2500 -2.2500	-39.2931 78.1410		2.3750 -0.9375 -0.3750	-61.8647 76.2075	
		4.5000 -1.1250 -1.0000	L.F.		2.3750 -1.2500 -1.0000	-88.1888 96.5403	
		4.2500 -1.0000 -7.2500	L.F.		2.3125 -1.3750 -0.8125	-82.6551 99.5696	
		2.1250 -1.5000 -0.2500	-49.7821 89.0056		2.0625 -1.5625 0.1875	-43.4303 88.0422	
4	2.4172 -1.2183 -1.2183	1.2500 -2.7500 -2.5000	-0.1508 96.7517	~4.5	2.2500 -1.5000 -0.6875	-77.9321 102.7756	~3.5
		2.5000 -1.5000 -2.0000	L.F.		3.6500 -1.5625 -0.5000	L.F.	
		-0.7500 -1.7500 -1.2500	83.8301 118.8219		2.1250 -1.2500 0.3125	-39.2931 78.1410	
		3.7500 -0.7500 -1.2500	L.F.		2.5625 -0.6875 -1.1875	-89.0434 74.3896	
		2.2500 -0.5000 -0.5000	-30.9224 60.0353		2.5000 -0.8125 -0.9375	-80.3923 76.8808	

N.P: Not Possible, L.F: Localization Failure.

Table 5.6: Mic. configuration I results, for the source location of (-140,60).

Fs	True Time Delay	Est. Tau CC	Est. Source Location CC	Run Time CC	Est. Tau OC	Est. Source Location OC	Run Time OC
KHz	msec	msec	cm	sec	msec	cm	sec
16	2.7962 -0.1974 -2.3590	2.6875 -0.0625 -2.6875	-75.3977 51.9228	~66	N.P.	N.P.	N.P.
		2.5625 -0.5000 -2.4375	-73.8499 65.3152				
		2.6875 -0.5000 -2.4375	-118.5871 72.1532				
		2.6875 -0.2500 -2.4375	-89.8379 58.7635				
		2.9375 0.1250 -3.0625	L.F.				
8	2.7962 -0.1974 -2.3590	3.5000 -0.8750 -3.3750	L.F.	~17	2.2500 -0.6250 -2.3750	L.F.	~7
		5.6250 -0.5000 -5.3750	L.F.		2.8125 -0.1875 -2.4375	-152.6952 60.2618	
		3.8750 -0.2500 -2.1250	L.F.		2.7500 -0.2500 -2.0000	-117.1182 60.8576	
		1.8750 -0.8750 -3.3750	-12.8050 65.4009		2.7500 -0.3185 -2.1250	-127.4766 64.9059	
		4.2500 0.7500 0.2500	L.F.		2.8750 -1.2500 -1.7500	L.F.	
4	2.7962 -0.1974 -2.3590	7.0000 -0.2500 -8.2500	L.F.	~4.5	2.5462 -0.6349 -1.2965	L.F.	~3.5
		6.7500 -0.5000 -2.2500	L.F.		2.8125 -0.2500 -2.5000	-167.4766 64.9059	
		4.5000 2.7500 0	L.F.		1.9212 -0.9474 -1.1715	L.F.	
		4.0000 -0.2500 -3.0000	L.F.		2.7500 -0.3125 -2.1250	-126.3907 64.5128	
		7.2500 0.7500 -5.0000	L.F.		2.7693 -0.2926 -2.2250	-136.3907 64.5128	

N.P: Not Possible, L.F: Localization Failure.

2. Microphone Configuration II

Another microphone configuration adopted for source localization experiments is shown in Figure 5.22. It consists of a 2D microphone array placed in an xy plane. Microphone $M1$ is taken as an array reference and is placed at the origin. Microphones $M2$ and $M3$ are placed at an equidistant spacing of $d = 100$ cm from the reference along the horizontal and vertical axis respectively. Microphone $M4$ is placed 200 cm apart from the reference along y axis. The fourth microphone $M4$ provides an extra time delay measurements which in turn improves the location estimates.

Figure 5.23 shows the experimental setup for microphone configuration II. In order to assess the performance of the proposed algorithm under dense room reverberations, the setup is placed close to the walls. The acoustic sensors are mounted on top of the metallic stands and are raised to the same height. Sensor 1 placed at origin $(0,0)$ serves as the reference node. Sensor 3 is placed at $(0,100)$, and Sensor 4 at $(0,200)$ along the vertical axis, while Sensor 2 is placed at $(100,0)$ along the horizontal axis in the xy plane. The setup is placed on a floor mat of 4×4 m. The mat has a printed grid of 20×20 cm that helps in sensors and source placement for localization purpose. Output of the pre-amplifiers are connected to four different channels of the USB data acquisition device via audio leads and jacks. The data acquisition toolbox of MATLAB is used for the data acquisition process.

The proposed OC based method and existing CC method of TDE have been

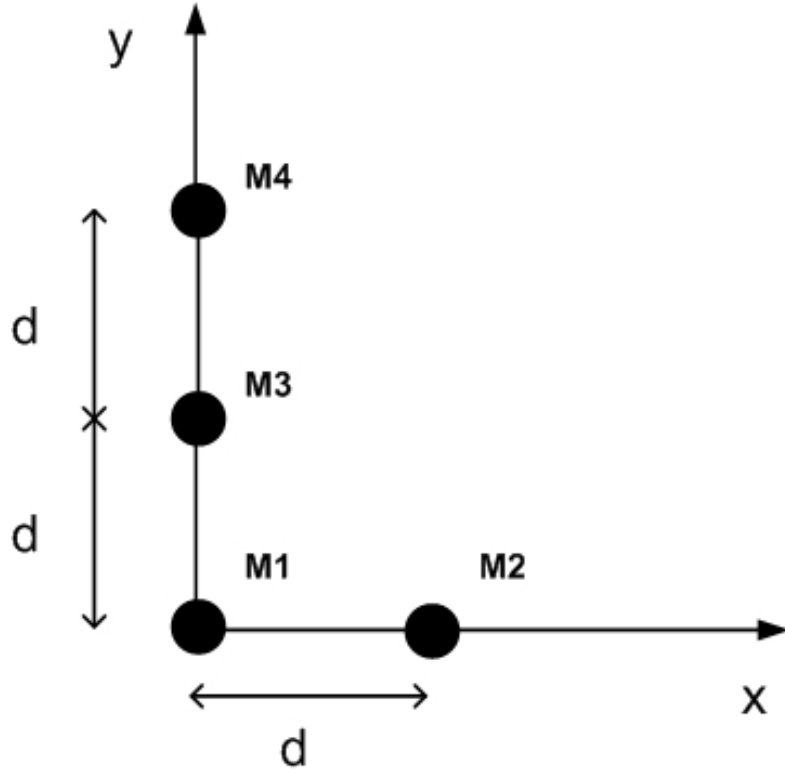


Figure 5.22: Microphone configuration II for source localization.

applied on the acquired data for three different source locations. For each source location, the signal is acquired at sampling rates of 4, 8, and 16 KHz respectively. The source localization method based on TDOA determines the source location from the estimated time delays. The location estimates for the source positioned at various locations around acoustic sensors are shown in Figures 5.24–5.26. Each figure shows the position of the acoustic sensors (filled squares), actual (hollow circles) and the estimated source locations (stars for 8 KHz and diamond for 4 KHz). The plot area represents the region of the floor mat.



Figure 5.23: Experimental setup of Mic. configuration II.

From the figures we observe that with the sampling rate of 8 KHz, a high percentage of estimated source locations are in close proximity with the true location. Decreasing the sampling frequency to 4 KHz, the algorithm finds close location estimates in many cases while in others the estimates lie a bit far from the true location. Thus, a trade off can be achieved between the accuracy of the source location estimates and less computations required with low data rates as shown earlier in Figure 5.10.

The results of the CC and OC based time delay and source location estimates for the Microphone configuration II are tabulated in Tables 5.7–5.9. Each table

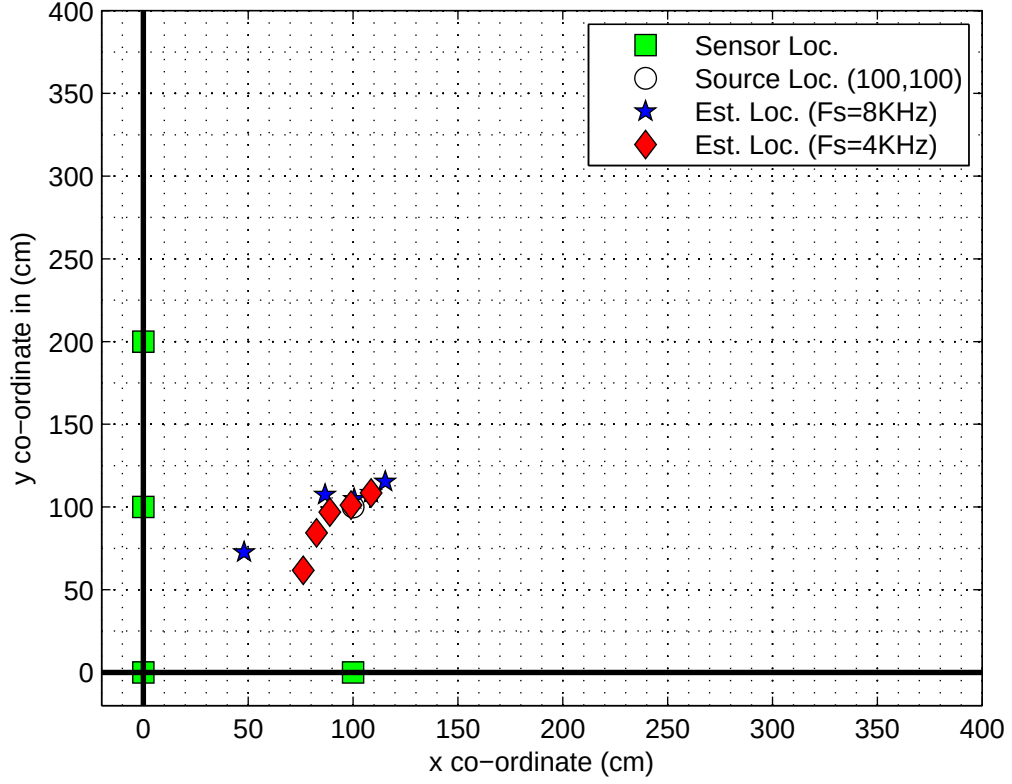


Figure 5.24: Mic. configuration II, localization results for source at (100,100).

provides the results of estimated time delays, source location, and computation time of both TDE techniques for a single source location sampled at three different sampling rates. We make following observations about the results obtained:

- The close to wall microphone configuration results in an increased level of room reverberations. As a result, the CC based location estimates deteriorate even at the high sampling rate of 16 KHz. The proposed OC based method could not be applied to 16 KHz measurements as the available physical memory could not store the dictionary matrix and the data acquired from the acoustic sensors all at once.

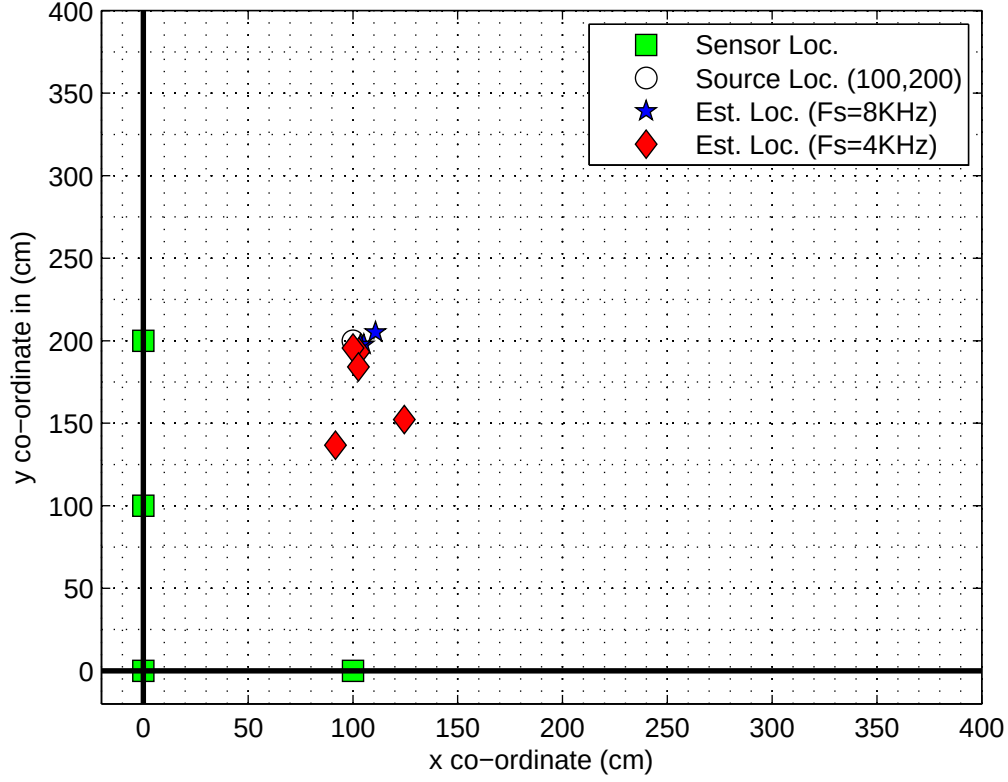


Figure 5.25: Mic. configuration II, localization results for source at (100,200).

- At 8 KHz, the proposed OC based method is observed to be working effectively under dense room reverberation conditions. Comparing this case with the center room configuration, the localization failure² has significantly decreased from 30% to 12%. In contrast, the CC method suffers both from the increased level of reverberations and higher under sampling factors. These reverberations result in an increased rate of localization failures from 50% (center room configuration) to 60%. A comparison of localization failures of both CC and OC with respect to the sampling frequencies is shown in Figure 5.27. Moreover, the execution time of OC method is 7 sec which is much faster than 17 sec consumed by CC.

²Localization Failure is the ratio of the number of times the algorithm fails to locate the source in an acceptable region to the total number of measurements conducted at a specific F_s .

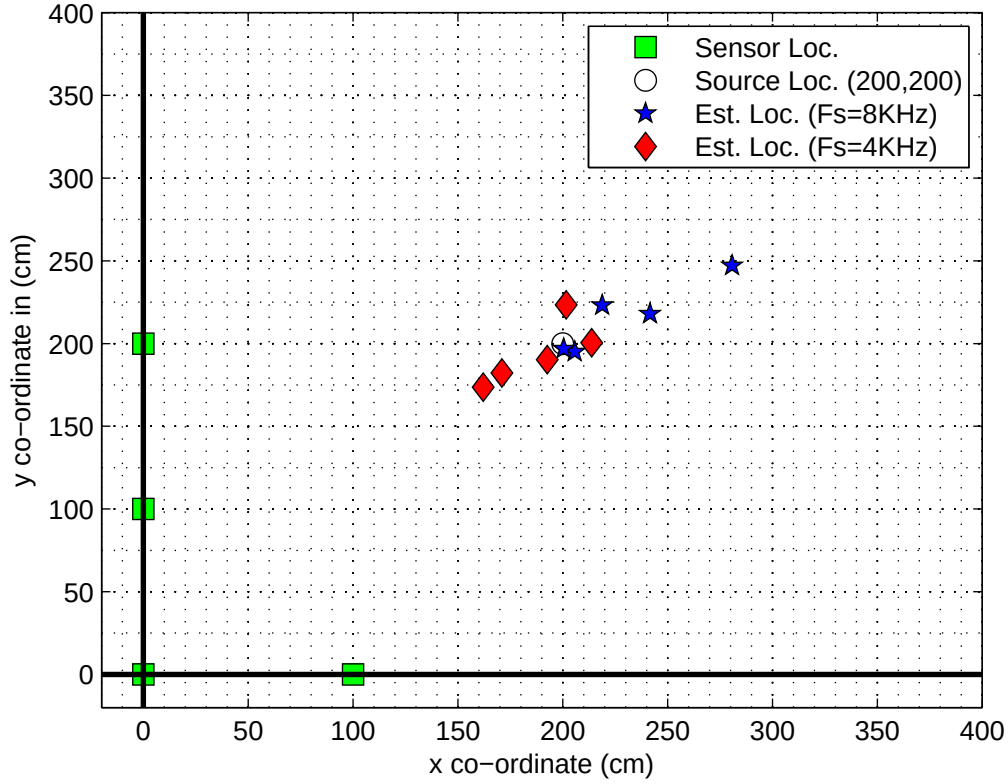


Figure 5.26: Mic. configuration II, localization results for source at (200,200).

- At 4 KHz, the OC method performs consistently well under high sub-sampling rate with a low percentage 20% of localization failures as compared to 40% for center room configuration. On the other hand, the performance of the CC method deteriorates further at such low sampling rate. A high percentage of localization failures of more than 75% are observed with CC. The computation time of OC is 3.5 sec, faster than 4.5 sec of CC.

In the experiments, the acoustic source was not an ideal point source but in fact it was a person standing at the true location and exciting the source in plane of the microphones. The location estimates are expected to lie within a certain region (± 25 cm).

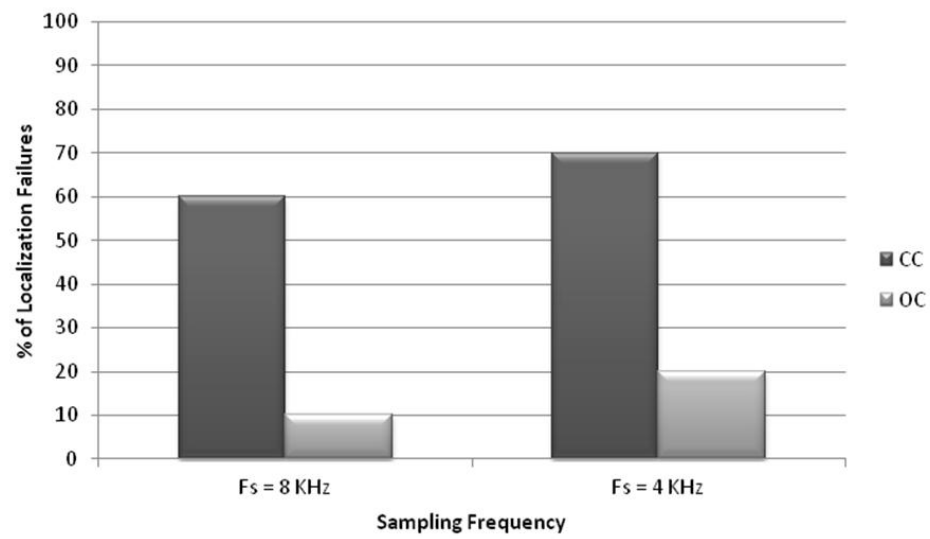


Figure 5.27: Localization failures in CC and OC for microphone configuration II.

Table 5.7: Mic. configuration II results, for the source location of (100,100).

Fs	True Time Delay	Est. Tau CC	Est. Source Location CC	Run Time CC	Est. Tau OC	Est. Source Location OC	Run Time OC
KHz	msec	msec	cm	sec	msec	cm	sec
16	-1.2183 -1.2183 0	-0.9375 -1.5625 -0.500	090.3828 111.6319	~66	N.P.	N.P.	N.P.
		-1.1875 -3.7500 -0.1250	056.0564 114.4621				
		-4.0625 -1.1875 -3.3750	L.F.				
		-1.1875 -1.1875 -0.1250	100.9126 102.3301				
		-1.1875 -2.0625 -0.1250	089.0451 107.9577				
8	-1.2183 -1.2183 0	-1.2500 -0.8750 -0.1250	107.5024 100.2307	~17	-1.3750 -1.3750 -0.5625	115.3065 115.3065	~7
		-1.0000 2.3750 -0.5000	148.7315 078.9918		-0.8750 -1.5000 -0.3125	086.6040 107.3310	
		-1.1250 -1.2500 -0.1250	097.4418 102.5462		-1.3125 -1.3125 -0.3125	108.5450 108.5450	
		-1.2500 -1.2500 -0.3750	107.0015 108.6246		0.0659 -0.9364 1.4375	048.0229 072.6615	
		-1.2500 -1.2500 0.8750	089.8146 083.7002		-1.1875 -1.3125 -0.1875	100.4845 104.8504	
4	-1.2183 -1.2183 0	-1.5000 -0.2500 -1.2500	173.5141 132.6710	~4.5	-1.1875 -1.2500 0.0625	098.9704 101.0963	~3.5
		-1.5000 -2.0000 -0.5000	109.6773 117.3646		-0.9375 -0.3750 1.7500	076.1590 061.6828	
		-4.2500 -2.5000 -2.7500	L.F.		-0.9375 -1.0000 0.6875	082.5345 084.3423	
		-1.7500 -1.0000 -0.5000	143.8900 113.4975		-1.0000 -1.2500 0.1250	088.9227 096.8472	
		-1.5000 -1.2500 3.0000	081.5315 050.9293		-1.3125 -1.3125 -0.3125	108.5450 108.5450	

N.P: Not Possible, L.F: Localization Failure.

Table 5.8: Mic. configuration II results, for the source location of (100,200).

Fs	True Time Delay	Est. Tau CC	Est. Source Location CC	Run Time CC	Est. Tau OC	Est. Source Location OC	Run Time OC
KHz	msec	msec	cm	sec	msec	cm	sec
16	-0.6493 -2.4172 -3.6355	-4.6250 -2.5000 -4.0000	L.F.	~66	N.P.	N.P.	N.P.
		0.5000 -1.1875 -2.2500	-070.8853 147.3618				
		-3.2500 -2.3125 -3.8750	L.F.				
		-0.8125 -2.5000 -5.5000	279.9991 772.3082				
		-0.7500 -4.0625 -3.4375	059.5287 176.5458				
8	-0.6493 -2.4172 -3.6355	1.5000 -5.1250 -3.3750	-10.2078 157.3181	~17	-0.7500 -2.3750 -3.5000	105.1786 197.6586	~7
		-1.6250 -3.3750 -7.0000	L.F.		-0.6875 -2.5625 -3.6250	112.1000 253.6925	
		-3.3750 -4.3750 -3.8750	L.F.		-0.5625 -2.7500 -3.5625	125.4193 383.9465	
		-3.6250 -2.8750 -4.1250	L.F.		-0.8125 -2.3750 -3.5625	110.8345 205.1088	
		-0.2500 -2.0000 -7.7500	L.F.		-0.7500 -2.3750 -3.5000	103.6175 197.4815	
4	-0.6493 -2.4172 -3.6355	-0.7500 2.0000 -8.2500	L.F.	~4.5	-0.7500 -2.3750 -3.3750	102.8201 193.2715	~3.5
		-3.2500 -3.7500 -4.0000	L.F.		-1.2500 -1.8125 -1.8750	124.5632 152.2237	
		2.7500 -2.5000 -2.2500	-118.8193 163.3382		-0.6875 -2.3750 -3.5000	100.0156 195.6426	
		-0.7500 4.0000 -4.5000	L.F.		-0.7500 -2.3125 -3.1250	102.6414 184.1070	
		1.7500 -4.7500 -5.2500	-027.5165 202.0913		-0.8125 -1.9375 -1.5625	091.6643 136.7548	

N.P: Not Possible, L.F: Localization Failure.

Table 5.9: Mic. configuration II results, for the source location of (200,200).

Fs	True Time Delay	Est. Tau CC	Est. Source Location CC	Run Time CC	Est. Tau OC	Est. Source Location OC	Run Time OC
KHz	msec	msec	cm	sec	msec	cm	sec
16	-1.7422 -1.7422 -2.4366	0.3750 -1.6875 -2.8125	053.7274 153.3640	~66	N.P.	N.P.	N.P.
		0.6250 -1.6875 -2.4375	-042.3980 145.7961				
		-1.7500 -4.7500 -3.4375	130.9132 232.9110				
		-1.5000 -1.6875 -2.5000	171.5535 189.6116				
		-1.5000 -1.6875 -4.4375	444.6270 576.5677				
8	-1.7422 -1.7422 -2.4366	-2.1250 -0.6250 -2.5000	L.F.	~17	-1.8125 -1.6875 -2.2500	205.6924 195.0554	~7
		-4.1250 -5.1250 -1.7500	483.4654 358.0039		-1.9375 -1.6875 -2.4375	241.6508 217.9416	
		1.1250 -5.6250 -4.1250	-005.3300 170.0618		-1.7500 -1.8750 -2.6875	218.8773 223.1588	
		-2.7500 -2.7500 -4.3750	L.F.		-2.0000 -1.6785 -2.6250	280.7596 247.1121	
		-0.5000 -0.8750 -1.6250	097.7642 132.6208		-1.7500 -1.6875 -2.3750	200.3877 196.8731	
4	-1.7422 -1.7422 -2.4366	-0.5000 -1.5000 0.2500	067.5485 095.4094	~4.5	-1.5625 -1.7500 -2.1250	162.0508 173.6003	~3.5
		-1.7500 -2.7500 -2.0000	141.0712 168.7447		-1.5625 -1.7500 -2.3125	170.9637 182.1650	
		-1.7500 1.0000 -2.0000	L.F.		-1.6250 -1.8750 -2.6875	201.5788 223.4459	
		-1.7500 -3.5000 7.2500	073.6345 017.5420		-1.7500 -1.6875 -2.3125	192.5786 190.3407	
		-1.7500 -0.2500 -5.0000	L.F.		-1.8125 -1.6875 -2.3125	213.8114 200.5022	

N.P: Not Possible, L.F: Localization Failure.

3. Microphone Configuration III

Figure 5.28 shows the third microphone configuration for source localization experiments. It represents a 2D microphone array placed in an xy plane. Microphone $M1$ is taken as an array reference and is placed at the origin. Microphones $M2$ and $M3$ are placed at an equidistant spacing of $d = 100$ cm from the reference along the horizontal and vertical axis respectively. Microphone $M4$ is placed 100 cm apart from the reference in both x and y axis. The fourth microphone $M4$ provides an extra time delay measurements which in turn improves the location estimates.

The experimental setup for the microphone configuration III is shown in Figure 5.29. Another close to the wall experiment is conducted to gather convincing evidence of better performance of the proposed OC based method as observed in the last section. The acoustic sensors are mounted on top of the metallic stands and are raised to the same height. Sensor 1 placed at origin $(0, 0)$ serves as the reference node. Sensor 2 is placed at $(0, 100)$ along the horizontal axis, and Sensor 3 at $(100, 0)$ along the vertical axis. Sensor 4 is placed at location $(200, 200)$ in the xy plane. The setup is placed on a floor mat of 4×4 m. The mat has a printed grid of 20×20 cm that helps in sensors and source placement for localization purposes. A USB data acquisition device interfaces the output of the pre-amplifiers with the data acquisition toolbox of MATLAB via audio leads and jacks.

The proposed OC based and existing CC method of TDE have been applied

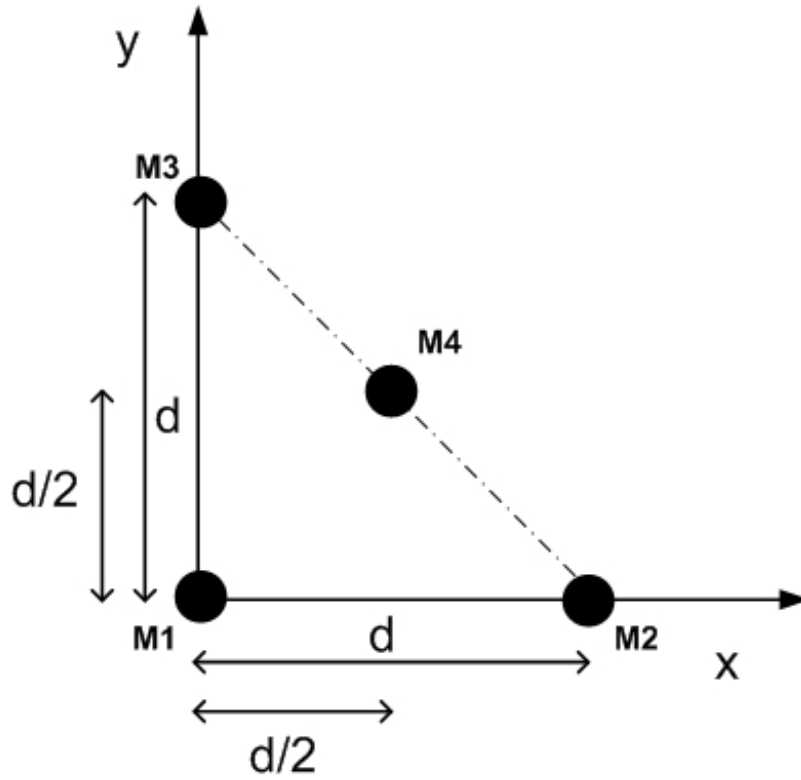


Figure 5.28: Microphone configuration III for source localization.

on the acquired data for two different source locations. For each source location the signal is acquired at sampling rates of 4, 8, and 16 KHz respectively. TDOA based source localization method is used to determine the source location from the estimated time delays. The estimated locations for the source positioned at various locations around acoustic sensors are shown in Figures 5.30–5.31. The figures show the position of the acoustic sensors (filled squares), actual (hollow circles) and the estimated source locations (stars for 8 KHz and diamonds for 4 KHz). The plot area represents the region of the floor mat.



Figure 5.29: Experimental setup of Mic. configuration III.

The location estimates for 8 KHz are observed to be lying close to the true source location. At 4 KHz the location estimates are scattered in a small region around true location. A trade off can be achieved between the accuracy of the source location estimates and less computational time by sampling at low data rates as shown earlier in Figure 5.10.

Tables 5.10–5.11 tabulates all the experimental results conducted for microphone configuration III. It shows the time delay, source location estimates, and computation times for both OC and CC methods. From the results we observe the following

- Similar effect of performance degradation is observed for the CC based method at 16 KHz, as was observed in previous close wall configuration. The OC

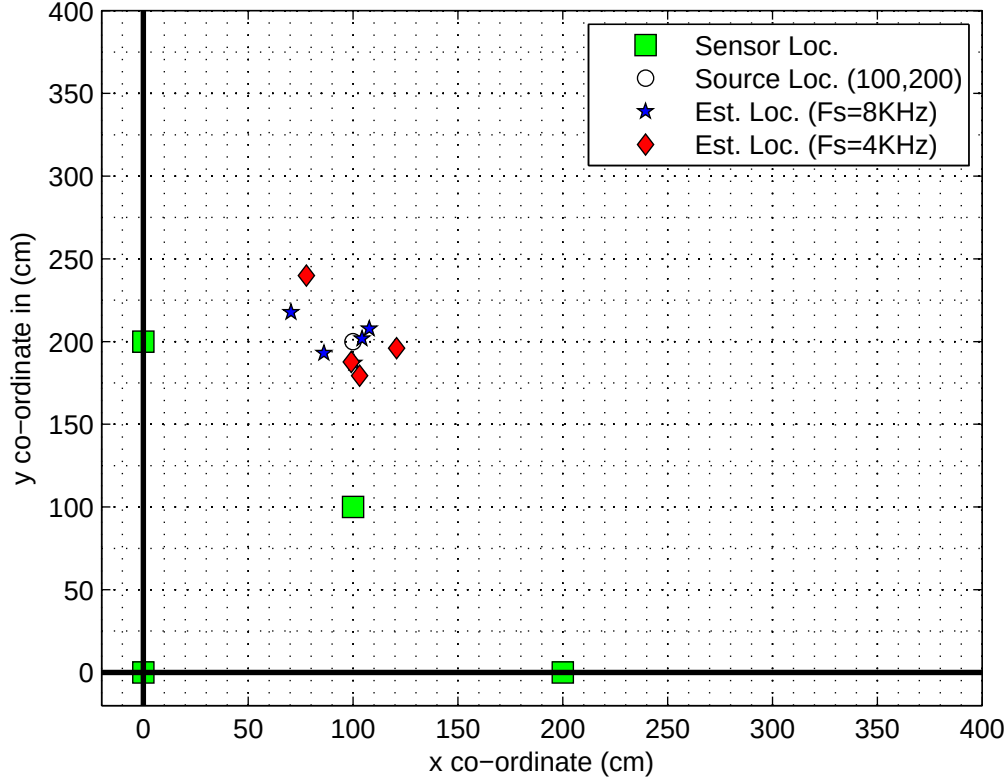


Figure 5.30: Mic. configuration III, localization results for source at (100,200).

method could not be applied to 16 KHz measurements because of physical memory limitations.

- At 8 KHz, the OC method works better under increased level of reverberations as less localization failure rate³ of 10% is observed as compared to 30% for center room configuration. Once again, we observe performance degradation of CC method with high failure rate of CC of 60% caused by increased reverberations and high under sampling factor. Figure 5.27 shows the comparison between the localization failures of OC and CC under different sampling rates. The

³Localization Failure is the ratio of the number of times the algorithm fails to locate the source in an acceptable region to the total number of measurements conducted at a specific F_s .

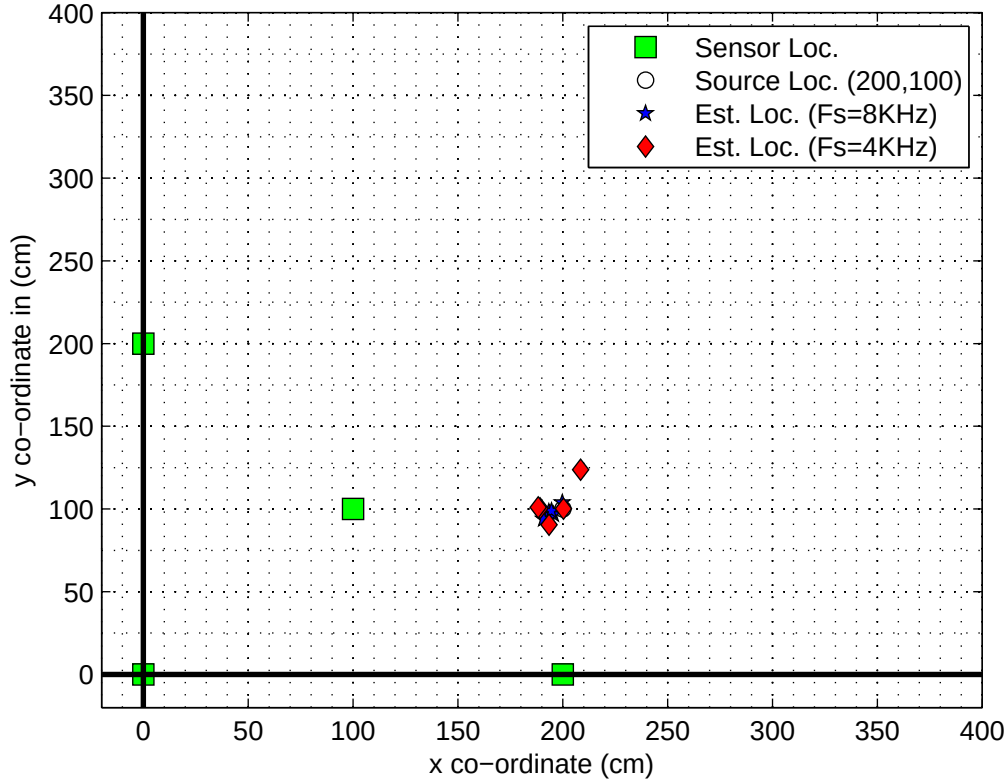


Figure 5.31: Mic. configuration III, localization results for source at (200,100).

execution time of OC method is 7 sec which is much less than 17 sec consumed by CC.

- At 4 KHz, the OC method performs effectively under dense reverberations. Comparing with center room configuration, less localization failure rate of 20% is achieved. In contrast, CC fails to locate the source 70% of the times due to high sub-sampling rate and increased reverberations. Figure 5.27 shows the rate of localization failure for both OC and CC methods under different sub-sampling rates. The computation time of OC is 3.5 sec, faster than 4.5 sec of CC.

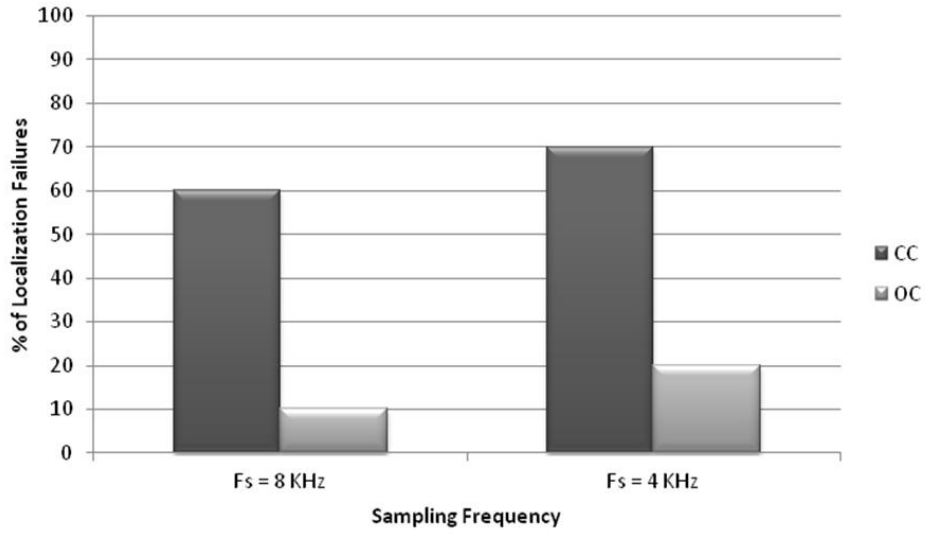


Figure 5.32: Localization failures in CC and OC for microphone configuration III.

5.3.3 Discussion on the Results

In the previous sections, we observed improved performance of the proposed OC based source localization method through simulations and experimental results.

On the basis of these results we make the following conclusions:

1. The time delays for an indoor environment are extremely small (few milliseconds), where the source and the microphones lie close to each other. In this case, accurate TDE requires high time resolution (low sampling time). The method we proposed reconstructs a sparse RIR signal of length N from sub-sampled signal of length M where $M \ll N$ using signal statistics, sparsity information, and problem structure. Although, decreasing the number of measurements (sampling rate) makes the RIR estimation less accurate however, this does not decrease the time resolution. In contrast, CC based

TDE method depends largely on the sampling rate. At low sampling rates, time resolution (large sampling time) gets poor which makes it hard to find the correlation peak close to the true time delay.

2. The proposed method estimates the time delays from RIR. The accuracy of TDE depends largely on estimating the impulse which corresponds to the direct path signal. Since RIR is estimated from under-sampled signal, there is a possibility to miss the impulse corresponding to the true time delay. The TDE in such case will rely on estimating the closest reflection. Thus, increased level of reverberations in the close wall configuration favored the proposed TDE method as it suffered from low localization failures as compared to the center room configuration. On the other hand, reverberations have an adverse effect on CC method. The reflected signals cause multiple peaks to appear in the correlation resulting in performance degradation.
3. Increase in the sampling rate improved the performance of OC based TDE method, as more measurements favored the sparse RIR signal reconstruction. Thus, OC based localization method performed better at the sampling rate of 8 KHz (localization failure $\leq 30\%$) as compared to 4 KHz (localization failure $\leq 40\%$).
4. Placing the microphones close to the wall, improved the performance of the proposed OC based method. Among the three microphone configurations that we experimented, the least localization failures ($\leq 20\%$) were experienced with the Microphone configuration II. Similar failure rate was ob-

served in Microphone configuration III, while the center room configuration resulted in large rate of localization failure ($\leq 40\%$).

5.4 Summary and Conclusions

In this chapter, simulated and experimental results of the proposed OC based method of source localization were presented. The hardware components required for the prototype implementation were first discussed. The effect of the room reverberations and sub-sampling rates on the existing CC and proposed OC method of TDE were analyzed by both simulations and experiments. For the present application, the limitations of DOA based localization method were discussed. With multiple TDEs, TDOA based method was applied for source localization. Three different microphone configurations were considered and the effects of under-sampling rates and reverberation on source localization were analyzed. For this purpose, two microphone configurations were placed close to the wall to increase the level of room reverberation. In the end, conclusions were drawn on the basis of the results thus obtained.

Table 5.10: Mic. configuration III results, for the source location of (100,200).

Fs	True Time Delay	Est. Tau CC	Est. Source Location CC	Run Time CC	Est. Tau OC	Est. Source Location OC	Run Time OC
KHz	msec	msec	cm	sec	msec	cm	sec
16	0 -3.6355 -3.6355	-0.0625 -4.5000 -4.8125	117.0585 272.2130	~66	N.P.	N.P.	N.P.
		1.9375 -3.5625 -4.0000	52.7996 182.8618				
		0.3125 -3.5625 -5.3750	88.5602 306.8881				
		0.3125 -5.8125 -6.6875	L.F.				
		-0.0625 -3.5625 -5.7500	111.4618 477.4808				
8	0 -3.6355 -3.6355	0 -4.6250 -7.6250	L.F.	~17	-0.0625 -3.8125 3.6250	104.3252 201.6832	~7
		0.5000 2.1250 -5.3750	L.F.		0.4375 -3.5625 -3.6875	86.0641 193.0024	
		1.5000 -7.0000 -5.3750	96.5360 259.3756		-0.2500 -3.5625 -3.6875	107.7073 207.6842	
		4.8750 -2.1250 5.3750	L.F.		-0.0625 -3.5625 -3.2500	100.3672 186.9554	
		1.2500 -4.5000 -5.3750	72.5232 258.3508		1.6250 -4.6875 -4.7500	70.3753 217.6093	
4	0 -3.6355 -3.6355	0.5000 -3.2500 -3.7500	79.5541 191.9231	~4.5	2.0000 -0.1875 -3.1875	L.F.	~3.5
		-3.2500 -2.0000 -1.2500	L.F.		-0.6250 -3.6250 -3.1875	120.7007 195.9869	
		0.2500 -4.0000 -3.0000	96.7635 175.6021		-0.0625 -3.8125 -3.0000	103.0997 179.3659	
		1.0000 -2.5000 -3.0000	57.2434 161.5191		0.3750 -3.0000 -4.8125	77.7432 239.8572	
		0.2500 -6.7500 -4.0000	L.F.		-0.1875 -3.1875 -3.2500	99.1637 187.7241	

N.P: Not Possible, L.F: Localization Failure.

Table 5.11: Mic. configuration III results, for the source location of (200,100).

Fs	True Time Delay	Est. Tau CC	Est. Source Location CC	Run Time CC	Est. Tau OC	Est. Source Location OC	Run Time OC
KHz	msec	msec	cm	sec	msec	cm	sec
16	-3.6355 -3.6355 0	-3.6875 -3.8125 0.5000	193.3042 87.9511	~66	N.P.	N.P.	N.P.
		-3.2500 -3.5625 -2.3750	256.8164 218.5378				
		-3.5625 -3.5625 -3.3125	389.1696 370.1791				
		-7.0000 -7.0000 0.3750	L.F.				
		-5.5000 -4.6250 -0.1250	362.7588 128.1782				
8	-3.6355 -3.6355 0	-5.5000 -2.6250 0	L.F.	~17	-3.4375 -3.2500 -0.1250	193.3707 98.3185	~7
		-5.5000 -2.0000 -3.2500	L.F.		-3.5625 -3.6250 -0.1250	199.8387 103.8353	
		-5.5000 -3.3750 7.3750	L.F.		-3.5000 -3.7500 0.2500	190.8271 93.8818	
		-5.5000 -3.1250 0.6250	298.3537 67.4221		-3.5000 -3.0000 0.1875	195.5959 96.7441	
		-6.8750 -3.3750 2.0000	310.3569 15.4702		-3.5000 -3.5625 0	194.5911 98.7931	
4	-3.6355 -3.6355 0	-5.7500 -3.5000 -4.2500	L.F.	~4.5	-3.6250 -3.6250 0.3125	193.4242 90.4284	~3.5
		-5.5000 -3.7500 0.5000	L.F.		-3.6250 -3.5000 -0.0625	200.3173 100.1645	
		-5.5000 -4.0000 -5.7500	L.F.		-3.5000 -4.0000 -0.5625	208.5094 123.7067	
		-3.2500 -7.0000 -2.2500	L.F.		-3.3125 -3.5625 -0.0625	189.1115 100.4742	
		-3.0000 -4.2500 -3.7500	L.F.		-3.2500 -3.3125 -0.1875	188.2947 100.9982	

N.P: Not Possible, L.F: Localization Failure.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

This thesis presents a novel approach of time delay estimation (TDE) for indoor impulsive source localization. The thesis gives a brief overview of acoustic fundamentals relevant to the problem of source localization. Time delay based source localization methods were considered. On the basis of sound propagation models, various existing methods of TDE were discussed including simple CC and GCC which performs well in outdoor environments with low noise levels and moderate reflections but deteriorates largely in dense reverberation of the indoor environments. The AED method of TDE for reverberant environments was discussed. It estimates the impulse response of the acoustic channel for TDE. High computations involved in adaptive estimation of RIR make this method not favorable for real time tracking of the acoustic sources. The high sampling rate requirement of the existing techniques makes the localization process computationally intensive

and requires sophisticated hardware. Moreover, the need of centralized processing in CC based methods puts strain on the communication link between sensors and the processing unit.

This motivated us to develop a more robust TDE method, based on the RIR estimation to enhance the robustness of indoor acoustic source localization. In addition, we proposed a method that worked at extremely low sampling rates and hence decreased the computation time and hardware complexity. Moreover, the distributed nature of the proposed algorithm enables it to perform localization at the sensor level which eliminates the need of centralized processing. The performance of the proposed method for TDE and localization was analyzed by both simulations and experimental setup. The results showed the improved performance of the proposed method over the existing CC method. Several microphone configurations were considered for source localization experiments. Through experimental evidences and theoretical understanding, it was found that the close wall configuration favored the proposed method.

In future, this work can be extended with the aim of improving the localization estimates and reducing the failure rate. In the following we list possible directions:

1. Increasing the number of microphones to achieve more accurate location estimates from redundant TDE measurements.
2. Close reflections favors the OC based localization process. This effect can be simulated by installing multiple closely spaced microphones.
3. Extending the $2D$ source localization method to $3D$ by using three dimensional

microphone arrays.

4. Eliminating the effect of uncertainty of the dictionary matrix in RIR estimation. Total least square (TLS) provides a possible solution in this regard. In our current implementation, we considered the dictionary matrix to be free of uncertainties.

REFERENCES

- [1] T D Rossing, *Springer handbook of acoustics*, Springer, 2007.
- [2] Xiaohong Sheng and Yu-hen Hu, “Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks,” *IEEE Transactions on Signal Processing*, vol. 53, no. 1, pp. 44–53, 2005.
- [3] J.C. Chen, K. Yao, T.L. Tung, C.W. Reed, and D. Chen, “Source Localization and Tracking of a Wideband Source Using a Randomly Distributed Beamforming Sensor Array,” *International Journal of High Performance Computing Applications*, vol. 16, no. 3, pp. 259–272, Aug. 2002.
- [4] Tshilidzi Mukwevho, Jaco Jordaan, and Guillaume Noel, “Advanced beamforming techniques for acoustic source localization,” in *AFRICON 2009*, Sept. 2009, pp. 1–6.
- [5] G.C. Carter, A.H. Nuttall, and P.G. Cable, “The smoothed coherence transform,” *Proceedings of the IEEE*, vol. 61, no. 10, pp. 1497–1498, 1973.

- [6] C. Knapp and G. Carter, “The generalized correlation method for estimation of time delay,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, Aug. 1976.
- [7] J. Ianniello, “Time delay estimation via cross-correlation in the presence of large estimation errors,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 30, no. 6, pp. 998–1003, Dec. 1982.
- [8] G.C. Carter, “Coherence and time delay estimation,” *Proceedings of the IEEE*, vol. 75, no. 2, pp. 236–255, 1987.
- [9] G. Jacovitti and G. Scarano, “Discrete time techniques for time delay estimation,” *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 525–533, 1993.
- [10] Michael S. Brandstein, “Time-delay estimation of reverberated speech exploiting harmonic structure,” *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2914, 1999.
- [11] Jacob Benesty, “Adaptive eigenvalue decomposition algorithm for passive acoustic source localization,” *The Journal of the Acoustical Society of America*, vol. 107, no. 1, pp. 384–392, 2000.
- [12] Simon Doclo and Marc Moonen, “Robust Adaptive Time Delay Estimation for Speaker Localization in Noisy and Reverberant Acoustic Environments,” *EURASIP Journal on Advances in Signal Processing*, vol. 2003, no. 11, pp. 1110–1124, 2003.

- [13] Zhen Zhao and Zi-qiang Hou, “The generalized phase spectrum method for time delay estimation,” in *ICASSP '84. IEEE International Conference on Acoustics, Speech, and Signal Processing*. 2003, vol. 9, pp. 459–462, Institute of Electrical and Electronics Engineers.
- [14] Jingdong Chen, Jacob Benesty, and Yiteng (Arden) Huang, “Time Delay Estimation in Room Acoustic Environments: An Overview,” *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. i, pp. 1–20, 2006.
- [15] Baruch Berdugo, Miriam A. Doron, Judith Rosenhouse, and Haim Azhari, “On direction finding of an emitting source from time delays,” *The Journal of the Acoustical Society of America*, vol. 105, no. 6, pp. 3355, 1999.
- [16] Ulrich Klee, Tobias Gehrig, and John McDonough, “Kalman Filters for Time Delay of Arrival-Based Source Localization,” *EURASIP Journal on Advances in Signal Processing*, vol. 2006, pp. 1–16, 2006.
- [17] Paulin Buaka Muanke and Christopher Niezrecki, “Manatee position estimation by passive acoustic localization.,” *The Journal of the Acoustical Society of America*, vol. 121, no. 4, pp. 2049–59, Apr. 2007.
- [18] Yasuhiro Teshima, Jun-ya Takayama, and Shinji Ohyama, “Person localization using TDOA of non-speech sound signal based on multiplexed CSP analysis,” 2010.

- [19] Tao Han, Xiaochun Lu, and Qi Lan, “Pattern recognition based Kalman filter for indoor localization using TDOA algorithm,” *Applied Mathematical Modelling*, vol. 34, no. 10, pp. 2893–2900, Oct. 2010.
- [20] Y.T. Chan and K.C. Ho, “A simple and efficient estimator for hyperbolic location,” *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 1905–1915, 1994.
- [21] Petre Stoica and Jian Li, “Lecture Notes - Source Localization from Range-Difference Measurements,” *IEEE Signal Processing Magazine*, vol. 23, no. 6, pp. 63–66, Nov. 2006.
- [22] B. Champagne, S. Bedard, and A. Stephenne, “Performance of time-delay estimation in the presence of room reverberation,” *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 2, pp. 148–152, Mar. 1996.
- [23] Jingdong Chen, Jacob Benesty, and Yiteng (Arden) Huang, “Performance of GCC- and AMDF-Based Time-Delay Estimation in Practical Reverberant Environments,” *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 1, pp. 25–36, 2005.
- [24] Omar M. Bouzid, Gui Yun Tian, Jeff Neasham, and Bayan Sharif, “Investigation of sampling frequency requirements for acoustic source localisation using wireless sensor networks,” *Applied Acoustics*, Feb. 2011.
- [25] Tareq Y. Al-Naffouri, Ahmed A. Quadeer, and Giuseppe Caire, “Impulsive noise estimation and cancellation in DSL using orthogonal clustering,”

- in *2011 IEEE International Symposium on Information Theory Proceedings*, July 2011, pp. 2841–2845.
- [26] Eric A Lehmann and Anders M Johansson, “Prediction of energy decay in room impulse responses simulated with an image-source model,” *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 269–77, July 2008.
- [27] Jacob Benesty, M. Mohan Sondhi, and Yiteng (Arden) Huang, Eds., *Springer Handbook of Speech Processing*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
- [28] M. Omologo and P. Svaizer, “Acoustic source location in noisy and reverberant environment using CSP analysis,” in *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, 1996, vol. 2, pp. 921–924.
- [29] M. Omologo and P. Svaizer, “Acoustic event localization using a crosspower-spectrum phase based technique,” in *Proceedings of ICASSP ’94. IEEE International Conference on Acoustics, Speech and Signal Processing*, 1994, vol. ii, pp. 273–276.
- [30] Rama Ratnam, Douglas L Jones, Bruce C Wheeler, William D O’Brien, Charissa R Lansing, and Albert S Feng, “Blind estimation of reverberation time,” *The Journal of the Acoustical Society of America*, vol. 114, no. 5, pp. 2877–92, Nov. 2003.

- [31] M. Omer, A. A. Quadeer, M. S. Sharawi, and T. Y. Al-Naffouri, “Time delay estimation in a reverberant environment by low rate sampling of impulsive acoustic sources,” in *Submitted to IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2012.
- [32] M.F. Duarte, M.A. Davenport, D. Takhar, J.N. Laska, K.F. Kelly, and R.G. Baraniuk, “Single-Pixel Imaging via Compressive Sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, Mar. 2008.
- [33] Michael Lustig, David Donoho, and John M Pauly, “Sparse MRI: The application of compressed sensing for rapid MR imaging.,” *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–95, Dec. 2007.
- [34] Jarvis Haupt, Waheed U. Bajwa, Gil Raz, and Robert Nowak, “Toeplitz Compressed Sensing Matrices With Applications to Sparse Channel Estimation,” *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5862–5875, Nov. 2010.
- [35] Jose L. Paredes, Gonzalo R. Arce, and Zhongmin Wang, “Ultra-Wideband Compressed Sensing: Channel Estimation,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 3, pp. 383–395, Oct. 2007.
- [36] Geert Leus and Ashish Pandharipande, “Direction estimation using compressive sampling array processing,” in *2009 IEEE/SP 15th Workshop on*

Statistical Signal Processing, Aug. 2009, pp. 626–629.

- [37] M.A. Herman and T. Strohmer, “High-Resolution Radar via Compressed Sensing,” *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2275–2284, June 2009.
- [38] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.
- [39] J.A. Tropp, “Just relax: convex programming methods for identifying sparse signals in noise,” *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.
- [40] Emmanuel J. Candes and Paige A. Randall, “Highly Robust Error Correction by Convex Programming,” *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 2829–2840, July 2008.
- [41] Deanna Needell and Roman Vershynin, “Uniform Uncertainty Principle and Signal Recovery via Regularized Orthogonal Matching Pursuit,” *Foundations of Computational Mathematics*, vol. 9, no. 3, pp. 317–334, June 2008.
- [42] D. Needell and J.A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, May 2009.

- [43] Joel A. Tropp and Anna C. Gilbert, “Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit,” *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [44] K.M. Zahidul Islam, Tareq Y. Al-Naffouri, and Naofal Al-Dhahir, “Asymptotically MMSE-optimum pilot design for comb-type OFDM channel estimation in high-mobility scenarios,” in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 3776–3779.
- [45] Manoj Sharma and Garima Saini, “Low complexity MMSE based channel estimation technique in OFDM systems,” in *2010 IEEE International Conference on Computational Intelligence and Computing Research*, Dec. 2010, pp. 1–4.
- [46] Kewen; Liu and Ke Xing, “Research of MMSE and LS channel estimation in OFDM systems,” in *The 2nd International Conference on Information Science and Engineering*, Dec. 2010, pp. 2308–2311.
- [47] Philip Schniter, Lee C. Potter, and Justin Ziniel, “Fast bayesian matching pursuit,” in *2008 Information Theory and Applications Workshop*, Jan. 2008, pp. 326–333.
- [48] A. A. Quadeer, S. F. Ahmed, and T. Y. Al-Naffouri, “Structure Based Bayesian Sparse Reconstruction Using Non-Gaussian Prior,” in *Conference on Communications, Control and Computing*, USA, 2011.

Vitae

- Muhammad Omer
- Nationality: Pakistani
- Current Address: P.O. Box 8651, K.F.U.P.M., Dhahran 31261, Saudi Arabia
- Permanent Address: House 284-A, Rewaz Garden, Lahore, 54000, Pakistan
- Telephone: (+966) 554 938 702
- Email: bromio47@gmail.com
- Born in Rawalpindi, Pakistan on July 1, 1987
- Received Bachelor of Science (B.S.) in Electrical Engineering from University of Engineering and Technology, Lahore, Pakistan in 2009.
- Joined King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia as a Research Assistant in January 2010.
- Completed Masters of Science (M.S.) in Electrical Engineering in January 2012.