

2013-01-23

Performance Evaluation of Localization Techniques for Wireless Sensor Networks

Maddumabandara, Ratagedara Herath M. Achintha

Maddumabandara, R. H. (2013). Performance Evaluation of Localization Techniques for Wireless Sensor Networks (Master's thesis, University of Calgary, Calgary, Canada). Retrieved from <https://prism.ucalgary.ca>. doi:10.11575/PRISM/27295

<http://hdl.handle.net/11023/446>

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

Performance Evaluation of Localization Techniques
for Wireless Sensor Networks

by

Ratagedara H. M. Achintha Maddumabandara

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

CALGARY, ALBERTA

January, 2013

© Ratagedara H. M. Achintha Maddumabandara 2013

Abstract

Localising people can be achieved by using wearable sensor nodes or by listening to speech. Receiver Signal Strength Indicator (RSSI), Time of Arrival (TOA) and Time Difference of Arrival (TDOA) are three main techniques that have been used in sensor node localization. Since lack of restrictions, the TDOA based methods have become more popular. It can be used for both wearable sensor nodes and for localizing speech. Using a wireless sensor network instead a wired microphone array is more convenient and easily deployable. However, performance degrades rapidly with acoustic reverberation and packet loss. Though robust beamforming techniques like SRP-PHAT exist, they are computationally exhaustive.

In this study, performance of both localization using wearable sensor nodes and speech are experimentally evaluated. Effect of packet loss is analysed and ways of minimizing the effects are proposed. Variance of TDOA estimate is derived and a new search algorithm to minimize computational cost of SRP-PHAT is proposed.

Acknowledgements

I would like to express my utmost gratitude to Dr. Henry Leung for his continuous guidance throughout my graduate life. Without his support and invaluable suggestions, this thesis would not have been possible.

I would also like to thank Dr. Chi Tsun (Ben) Cheng, in whom I found both a mentor and a good friend and all my labmates for making the last two years of my life unforgettable.

Many thanks to TinyOS community for their generous technical support and feedbacks. Without their help I would be still struggling to code the nodes. I would like to thank to Donald Knuth and Leslie Lamport . Without \LaTeX , I might be still writing this thesis.

To my sister Ishara Dinali and her husband Sanjeewa Doraliyadda for making this foreign land, home.

And finally, above all, my beloved parents for always been there for me through all the tough times. To their unconditional love and never ending caring.

Table of Contents

Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
List of Abbreviations	x
1 Introduction	1
1.1 Related Work	2
1.2 Aim of this Thesis	5
1.3 Contributions	6
1.4 Structure of this Thesis	6
2 Localizing using Wearable Sensors	7
2.1 Experimental Setup	8
2.2 Using Received Signal Strength Indicator	13
2.2.1 RSSI vs. Distance	15
2.2.2 RSSI vs. Antenna Orientation	16
2.2.3 Localization Performance	19
2.3 Time of Arrival	20
2.3.1 TOA vs. Distance	22
2.3.2 Localization Performance	24
2.4 Time Difference of Arrival	24
2.4.1 Localization Performance	29
2.5 Chapter Summary	30
3 TDOA Based Localization	31
3.1 Acoustic Model	31
3.1.1 Direct Propagation	31
3.1.2 Multipath Propagation	32
3.2 Cross-Correlation based Time Delay Estimation	35
3.2.1 Generalized Cross Correlation	36
3.3 Beamformer Based Localization	41
3.3.1 Steered Response Power with Phase Transformation	42
3.4 Chapter Summary	46
4 Effect of Reverberation and Packet Loss	48
4.1 Reverberation	49
4.1.1 Effect of Reverberation on Time Delay Estimation	50
4.2 Packet Loss	54
4.2.1 Effect of Packet Loss on Time Delay Estimation	56
4.2.2 Minimizing Effects of Packet Loss	66
4.3 Variance of GCC-PHAT TDOA Estimate	69
4.4 Chapter Summary	73
5 Proposed Method and Experimental Results	74

5.1	Proposed Search Algorithm	74
5.2	Modifications to the Experimental Setup	77
5.2.1	Communication	78
5.2.2	Data Compression	79
5.2.3	Time Synchronization	81
5.2.4	Source Detection	82
5.3	Experimental Results	85
5.3.1	Performance of the Proposed Method	85
5.3.2	Effect of Speaker Location	90
5.3.3	Effect of Reverberation	93
5.3.4	Effect of Zero Padding	95
6	Conclusion	100
A	GCC Derivations	103
	Bibliography	105

List of Tables

2.1	Localization performance	30
3.1	Popular weighing functions for GCC: $G_{as}(f)$ is power spectral density of $a(t)$ and $G_{ab}(f)$ represents cross power spectral density of $a(t)$ and $b(t)$. . represents absolute value operator	36
4.1	Probability of getting two streams aligned: In the first column possible number of missing packets is given. In the second column number of combinations where both the streams loose packets simultaneously is given. And in the last column probability of getting simultaneous packet lost in both the streams for given number of lost packets is given.	60
4.2	Maximum number of lost packets allowable to maintain desired N_{A_j} for given lossless observation window size	63
5.1	Statistics of noise energy for 4400 samples	83
5.2	Localization error statistics: For proposed search $\eta = 2$, $\rho = 0.2$ m	88
5.3	Localization error vs. speaker location: Rough estimates	91
5.4	Localization error vs. speaker location: Proposed method	91
5.5	Localization error vs. speaker location: Full search	91
5.6	Localization error vs. speaker location with high reverberation: Rough estimates	94
5.7	Localization error vs. speaker location with high reverberation: Proposed method	94
5.8	Localization error vs. speaker location with high reverberation: Full search	94

List of Figures and Illustrations

2.1	Hardware	10
2.2	Modified sounder	11
2.3	Experimental setup	12
2.4	Propagation of a RF signal: “SN-x” tag represent anchor nodes and “SN-BF” represents the wearable sensor node. Black dotted lines represent distance between antenna of SN-BF and anchor nodes. Blue solid lines represent equidistant points	14
2.5	RSSI vs. distance: “RSSI-i-Ex” refers to experimental RSSI recorded from i^{th} receiving node and “RSSI-i-Th” refers to the theoretical RSSI	16
2.6	RSSI vs. antenna orientation: Lines A1, A2, A3, and A4 represent recorded RSSI values for 0° , 90° , 180° , and 270° antenna orientations respectively	18
2.7	Localization using RSSI: Average localization error = 1.0098 m, variance = 0.3081 m^2 , bias = 0.7922 m	19
2.8	TOA vs. distance: “Node-Ex-i” refers to measured TOA for i^{th} anchor node. “Theoretical” refers to TOA generated using Equation 2.7	23
2.9	Localization using TOA: Average localization error = 0.3280 m, variance = 0.1013 m^2 , bias = 0.0645 m	25
2.10	Comparison between RSS/TOA and TDOA based approach: SN-1, SN-2, SN-3, SN-4 and a blindfolded node (a signal source) are located at (0,0), (2,0), (2,2), (0,2) and (0.5,1.5) respectively. r_i is distance between the blindfolded node and i^{th} anchor node. $TDOA_{ij}$ is TDOA calculated between i^{th} and j^{th} anchor nodes	27
2.11	Propagation of acoustic signals: Blue boxes with “SN-x” tag represent sensor nodes and “Speaker” symbol represents a person speaking. Black dotted lines represent distance between the target and each microphone. Blue solid lines represent equidistant points with given distance away from the target	28
2.12	Localization using TDOA: Average localization error = 0.3320 m, variance = 0.1078 m^2 , bias = 0.0871 m	29
3.1	Image-Source method for room acoustics: The dark rectangle in the middle represents a room with four walls. Other dashed rectangles represent images of the room. The small black square represents a signal source and the red square represents a anchor node. Lines between the two squares represent signal propagation paths.	33
3.2	Room Impulse Response: An empty room of $8.7 \times 7.4 \times 4 \text{ m}$ size is considered. Two sensor nodes and a signal source were placed at (3,1), (1,1) and (5,1) at same height respectively. All the surfaces were assumed to have uniform absorption properties and absorption coefficients were set by fixing reverberation time to 0.1 s.	34

3.3	Comparison between different GCC methods: GCC was performed on two acoustic signals generated from a 8 kHz sampled wave file using the Image-Source model. The wave file contains word “Hello” in male voice. For the model, two sensor nodes were placed 2m apart (47 sample delay) and reverberation was set to 0.1 s.	39
3.4	Percentage of error vs. SNR: Two different acoustic signals were generated from a 8 kHz sampled wave file using the Image-Source model. The two sensor nodes were placed 2 m apart (47 sample delay) and reverberation was fixed at 0.1 s. Noise was added to achieve average SNR between the two signals from -10 to 10 dB. For each SNR value, time delay was estimated using the methods mentioned in Table 3.1. 1000 independent trials were performed for each method and percentage of error was recorded	41
3.5	Comparison between N-1 and N(N-1)/2 pairs	44
3.6	Spacial likelihood function for SRP-PHAT:	45
4.1	Comparison of room impulse response for different reverberation conditions: Two room transfer functions were generated using Image-Source method for reverberation time 0.1 s and 0.3 s. In both the cases distance difference from sound source to two sensor nodes was fixed to 2 m (47 sample delay)	52
4.2	Percentage error vs. reverberation vs. SNR: Two signals were generated using Image-Source method by fixing time delay at 47 samples and varying T_{60} . 500 trials were performed using each GCC method with varying SNR and percentage errors were recorded.	53
4.3	Markov Model for Packet Loss. The blue circles represent two states and arrows represent possible state transitions. “G” is the state of receiving a packet and “B” is losing a packet. Probability of receiving the next packet given that the current one is missing, $P(G B) = g$. Probability of loosing the next packet given that the current one is successfully received, $P(B G) = b$	55
4.4	Received packets at the base station: Stream-1 is samples taken from the first anchor node as a stream of packets and stream-2 is from the other anchor node. 1 : 2 represents 2 nd packet from 1 st anchor node. Each observation window is N_p packets long	57
4.5	Received packets at the base station: First two rows represent transmitted packets and last two rows represent received packets. In the second row, red block represents a missing packet	57
4.6	Received packets at the base station: First two rows represent transmitted packets and last two rows represent received packets. In the first two rows, red blocks represent a missing packets	58
4.7	Packet loss in received signals	59
4.8	Overlap between two signals. The two boxes represent two aligned data blocks and black vertical line in each box represents samples.	61
4.9	For $N_p = 5$, $N_{loss} = 1$, $N_{Aj} = 3$	62

4.10	Percentage of errors vs. percentage of missing packets	65
4.11	Received data from two sensor nodes after inserting empty packets: Here, blue box represents received packets and red boxes on the first two rows represent missing packets. Red boxes on the last two rows represent empty packets	67
4.12	Percentage of errors vs. percentage of packet loss before and after inserting empty packets.	68
4.13	Number of samples after with packet loss	72
5.1	Proposed search area: $\theta = (0.5,1.5)$, $\hat{\theta}_e = (0.9,1.7)$, $l(\theta) = 0.4575$ m, $\eta = 1$, $\rho = \pm 0.1$ m, $\lambda = 0.02$ m	76
5.2	Number of search points vs. η vs. ρ	77
5.3	Data compression: Top four rows represents four samples occupying 8 bytes. Boxes with dark colours represent data bits and light coloured boxes represent “null” bits. The bottom row represents compressed samples.	80
5.4	Effect of filtering	84
5.5	Histogram of filtered noise	85
5.6	2×2 m anchor node configuration	86
5.7	Localization performance	87
5.8	Localization results	89
5.9	Speaker locations	91
5.10	Localization error vs. speaker location	92
5.11	Experimental setup	97
5.12	Localization error in a high reverberant environment	98
5.13	Effect of zero padding	99

List of Abbreviations

Abbreviation	Definition
ADC	Analogue to Digital Converter
BLUE	Best Linear Unbiased Estimator
CCA	Clear Channel Assessment
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
DFT	Discrete Fourier Transformation
DOP	Dilution of Precision
FDMA	Frequency Division Multiple Access
FIFO	First In First Out
GCC	Generalized Cross Correlation
GPS	Global Positioning System
IEEE	Institute of Electrical and Electronic Engineers
ISM	Industrial Scientific and Medical
LLS	Linear Least Squares
MLE	Maximum Likelihood Estimator
MSE	Mean Squared Error
PHAT	Phase Transformation
RADAR	Radio Detection and Ranging
RF	Radio Frequency
RSSI	Received Signal Strength Indicator
SNR	Signal to Noise Ratio
SONAR	Sound Navigation and Ranging
SRP	Steered Response Power
SRP-PHAT	Steered Response Power with Phase Transformation

TDMA	Time Division Multiple Access
TDOA	Time Difference of Arrival
TOA	Time of Arrival
U of C	University of Calgary
WSN	Wireless Sensor Network

Chapter 1

Introduction

Person localization is the process of identifying location of a person. Ability to locate and track a person has many advantages. In a sport event, this can be used to steer video cameras so that it follows the athlete, automatically. In a meeting room, this can be used to turn the microphone towards the speaker to get a better signal reception. And more importantly, in surveillance, it can be used to monitor and track intruders.

Based on the signals being used, localization methods can be categorized primarily into two groups. They are termed “active” methods and “passive” methods. In active methods, a deterministic signal is transmitted towards the target and reflected signal from the target is observed. Location of the target is then estimated comparing the two signals. Radio Detection and Ranging (RADAR) and Sound Navigation and Ranging (SONAR) are two well known localization methods that use this principle [1],[2].

This approach is not transparent to the target, and hence highly undesirable in military and surveillance applications. In contrast, passive methods use signals generated by the target itself. In Passive SONAR, an array of spatially separated hydrophones are used to secretly listen to vibrations generated by operating engines [3]. It has been used widely to locate enemy submarines. However, spending funds on a state-of-the-art passive SONAR system to localize people is neither practical nor realistic. Thus, alternative localization methods have been proposed by various researchers for these kind of situations.

Due to the rapid developments in recent decades in fields like, circuit design and fabrication, sensors have now become comparably cheaper and smaller. There are “cheap” wireless nodes with various sensors, different communication interfaces and compact

microprocessors. Networked wireless sensor nodes or Wireless Sensor Networks (WSN) have been widely used in various fields. As examples, it has been used in military to identify friendly military forces and ammunition [4], in health care, to monitor blood pressure level of patients [5], in environmental studies, to monitor volcanic eruptions [6], in habitat monitoring, to identify sea birds nesting [7], in manufacturing sector, to monitor pressure level of pipelines [8] and in surveillance, to locate snipers [9]. Using wireless sensor nodes for localization makes it easily deployable, and opting for low-cost nodes makes it more economical. Thus, localization using wireless sensor networks has more desirable advantages over dedicated wired microphone arrays.

1.1 Related Work

Person localization using wireless sensor networks can be implemented in two different ways. Option one is using wearable sensor nodes. When it comes to wearable sensors, using Global Positioning System (GPS) is the popular solution that comes to everyone's mind. Wearable GPS sensors have been used to monitor and track moose in natural habitat [10]. But our consideration is localizing people indoors. Since line-of-sight cannot be maintained with the orbiting satellites, GPS cannot be used inside buildings. Thus, other available resources on the sensor nodes have to be utilized.

Decay of signal strength and time of flight of propagating signals can be considered as the two most widely used observations in localization. Received Signal Strength Indicator (RSSI) of a received packet has been used to locate the transmitting sensor node. The difference between those applications and ours is, here the blindfolded sensor node is wearable. In [11] it is used to localize sensor nodes in outdoors. But average achievable accuracy is limited to 7 m. Approach given in [12] has achieved 1.5-2 m accuracy but it also needs tight calibration. In [13] and [14], a Maximum Likelihood Estimator (MLE) for RSSI considering noise is derived. They report 1.8 m median error

with maximum error around 4.2 m. For method proposed in [15], accuracy is around 3 m. The same method is revised for outdoor situation in [16] and average localization error of 1.07 m is reported. To achieve better performance, more anchor nodes are needed for [17] and more data points are needed for [18].

Simplicity and low cost are the main driving forces behind choosing RSSI. However, as long as localization accuracy and resolution are the concerns, RSSI based methods perform poorly [19]. All the published results tell us it is not suitable when fine-grained sub-meter accuracy is needed. Fading and shadowing effects increase RSSI measurement variance [12],[20],[21],[22]. Correlation between the recorded RSSI value and distance become ambiguous when multipath effects exist [15]. Accuracy of measurements deteriorates indoors and extensive calibration is needed [11],[23],[24],[25],[26].

Time of flight of an acoustic beacon has been used to localize a sensor node using Time of Arrival (TOA) and Time Difference of Arrival (TDOA) techniques [27],[28],[29]. To localize people, we can use the same methods with wearable sensors. “Cricket” is a popular TOA based localization method that use ultrasound signals [30]. Accuracy of Cricket is around 0.02 m, which is considerably better than any RSSI based methods [31]. Same technique has been used with an acoustic source in [32]. Though it is not as accurate as Cricket, it has a relatively large operation range [32]. [33] suggests placing a receiver node right next to the transmitter all the time to get time of transmission. Then TOA is taken as difference between signal reception at a node and the node closer to the transmitter. All the methods given have demonstrated sub-meter accuracy.

The TDOA approach is similar to passive SONAR. In SONAR, an array of hydrophones is used to detect underwater marine life, or in military, to detect hostile submarines [3]. In TDOA, an array of sensor nodes with microphones is used to localize acoustic signal sources [34]. An array of microphone is used to position a sound source in [35] and an array mounted on a robot to detect direction of sound arrival is used

in [36], [37] and [38]. Though these methods were implemented on wired microphone arrays they can be adapted to WSN.

Accuracy of TDOA based localization is constrained by accuracy of each TDOA measurements. When clocks in anchor nodes are synchronized, TDOA has fine-grained localization performance similar to TOA [39]. The two sensor nodes involved in calculating the TDOA measurement should accurately detect arrival of the signal. For impulsive signals, an energy detector can be used [40], [41]. For deterministic narrow-band signals, a hardware tone detector can be used. However existing hardware tone detectors are not always reliable [42], [33], [43].

The other option is by listening to sound generated by the person. A talking person can be considered as a wideband signal source. Then localization is performed as localizing a sound source. Since time of transmission is not needed, TDOA technique has gained more attention than TOA. Even commercial products like PictureTel and Polycom have been developed to steer cameras towards the speaker [44],[45]. Based on nature of the algorithm, TDOA based methods can be categorized into two groups as two-step and single-step. In two-step method, pairwise TDOA (or time delay between signals recorded at two sensor nodes) is estimated first. Then using the estimated TDOAs and location information of the sensor nodes, location of the sound source is estimated. For location estimation step a Maximum Likelihood Estimator (MLE) is derived in [46] and a spherical interpolation method is proposed in [47]. A method called “linear intersection” is presented in [48] and a constrained least squares method is given in [49]. However, these methods are vulnerable to reverberation effects [50].

Instead of estimating TDOAs explicitly, methods categorized as single-step employ beamforming techniques to locate the target. Beamforming is the process of performing spatial filtering on a microphone array to enhance the desired signal while suppressing the unwanted [51]. Instead giving a close-form location estimate, it delays and

combines signals from all the sensors to form a “spatial likelihood function”. This is called “delay-sum beamforming” [52]. Location of the target is then given by point in space that maximizes the likelihood function. When there are multiple speakers, the likelihood function produces local maxima at each speaker location. Using cross-correlation between filtered signals instead raw signals is called Steered Response Power (SRP) [53]. Due to promising superior results, it has been widely used in speaker localization [54],[55],[56],[57],[58],[59].

However, beamforming is computationally exhaustive. The beamformer output for a given point only returns likelihood of having the target at that particular point. So, it should be steered at multiple points until the desired maxima is found. Hence, the performance is constrained by number of search points.

1.2 Aim of this Thesis

Though using a microphone array to localize acoustic sources is not a completely new idea, little has been done to implement it in a low cost WSN. Due to communication interference and packet loss, WSN are not as robust as wired microphone arrays. Economical sensor nodes are not computationally powerful. Though fast two-step close-form methods exist, their performance degrade rapidly with acoustic reverberation. More accurate beamforming techniques, which have been proposed for microphone arrays, are computationally exhaustive. Thus, this thesis investigates ways of implementing a fast responding, yet accurate person localization method using desirable properties of both two-step and beamforming methods.

1.3 Contributions

In this thesis, performance of different localization techniques for WSN are evaluated. Each method is implemented in an indoor environment and their performance is compared. Suitability of using acoustic signals for localization over RSSI is experimentally proved. Two main issues when using acoustics in WSN are identified as reverberation and packet loss. Their effects are analysed and possible solutions to minimize the effects are proposed. A faster hybrid speaker localization method is proposed and its performance is compared with existing methods.

1.4 Structure of this Thesis

The rest of the thesis is as follows. In chapter 2, person localization using wearable sensor nodes is summarized. RSSI, TOA and TDOA methods are investigated and their experimental performances are given. Advantages and disadvantages of using each method and reasons behind selecting TDOA approach are discussed. In chapter 3, propagation of an acoustic signal in an indoor environment is modelled. Various methods available to estimate time delay in speech are discussed. Then, theory behind beamforming and its advantages are summarized. In chapter 4, various effects that degrade the time delay estimation are short listed. Reverberation and packet loss are selected as the major issues involved with localising speech using WSN, and their effects are analysed. A method of reducing effect of packet loss is proposed. In chapter 5, modifications are proposed to make the beamformer faster and experimental results with the proposed modifications are given.

Chapter 2

Localizing using Wearable Sensors

There are two ways of looking at the localization problem. It can be treated as a “lost person” (or an object) looking for where he/she is or it can be treated as a someone trying to identify where the “lost person” is. In standard terminology, the former is called a navigation problem and the later is called a tracking problem. The ultimate goal of our research is to send a robot to a person in need of help. Since location of the person is needed to the system, localization is treated as a tracking problem in this study. The person or the object being localized is called “the target”.

With range measurements from at least three spatially separated sensors, position coordinates of the target in 2D can be estimated using trilateration [60]. Ranging is the process of estimating distance between a sensor and a target. In the context of localizing people, heights can be assumed nearly uniform when everyone is sitting or standing. Nevertheless height variation is negligible compared to room dimensions. Since it simplifies the problem formulation, we opted looking only at 2D perspective.

Due to advancements in fields like miniature circuit design and fabrication, sensors have now become comparably smaller. There are off-the-shelf sensor nodes with various sensors, dedicated power sources and wireless connectivity [61]. GPS chips have been embedded in everyday use personal items like mobile phones, shoes and watches [62],[63],[64]. It is certain that a person or an object can be localized using wearable or embedded sensors. Thus, it is reasonable to narrow down the problem as localizing a wearable or embedded wireless sensor node.

A set of wireless sensor nodes with suitable sensors placed at predetermined locations are called “Anchor Nodes”. And the sensor node which is being localized is called

“Blindfolded Node”. Localization is performed by ranging from each anchor node to the blindfolded. Signal strength decay and time of flight of a propagating signal are the two most widely used observations for ranging. Our initial study is identifying the most suitable candidate for our application.

2.1 Experimental Setup

Our experimental setup consists of 4 anchor nodes and a single blindfolded node. All the nodes are IRIS[®] nodes with MTS310CA sensor boards from Crossbow[®]. IRIS[®] has a 16 MHz Atmel[®] ATmega1281 based microprocessor, 2.4 GHz ATmel[®] AT86RF230 radio transceiver and a 51-pin expansion connector for various sensor boards [65]. The microprocessor has a 10 bit 8 channel ADC, a 512 kB flash memory and 8 kB RAM [66]. It can sample at maximum 17 kHz but data logging to the flash memory is limited to 800 Hz [67]. The Zigbee[®] (IEEE 802.15.4) compliant radio transceiver operates at Industrial Scientific and Medical (ISM) frequency band and it is capable of switching between 16 different radio channels and transmitting packets at 250 kbps [68]. The sensor board is equipped with a photoresistor (Clairex[®] CL94L), a thermistor (Panasonic[®] ERT-J1VR103J), a microphone, a piezoelectric buzzer (ARIO[®] S14T40A), a phase-locked loop tone detector (Texas Instrument[®] LMC567), a 2-axis accelerometer (Analog Devices[®] ADXL202JE) and a 2-axis magnetometer (Honeywell[®] HMC1002) [69]. Both the buzzer and the tone detector are configured to 4.3 kHz.

The microprocessor doesn’t have enough resources to process data on-board [66]. Thus all the sensed data was forwarded to a computationally powerful “base station” for processing. Base station was a computer with an IRIS[®] node attached through a MIB520 USB interfacing card. All the nodes were programmed in TinyOS-2.x and MATLAB[®] was used at the base station for processing and analysis.

The on-board buzzer produces only 88 dB sound pressure level at 10 cm [42]. We experimentally found the tone detector can not detect sound generated by the buzzer approximately beyond 1.5 m. Since it is not sufficient for our experiments, a high power sounder has been attached as illustrated in Figure 2.2. An external 4.3 kHz signal generator was implemented using an Arduino[®] board which can be triggered by IRIS[®]. Output of the signal generator was then fed into a speaker with an adjustable volume control. This modification increases buzzer power to 98 dB and extends operating range beyond 5 m.

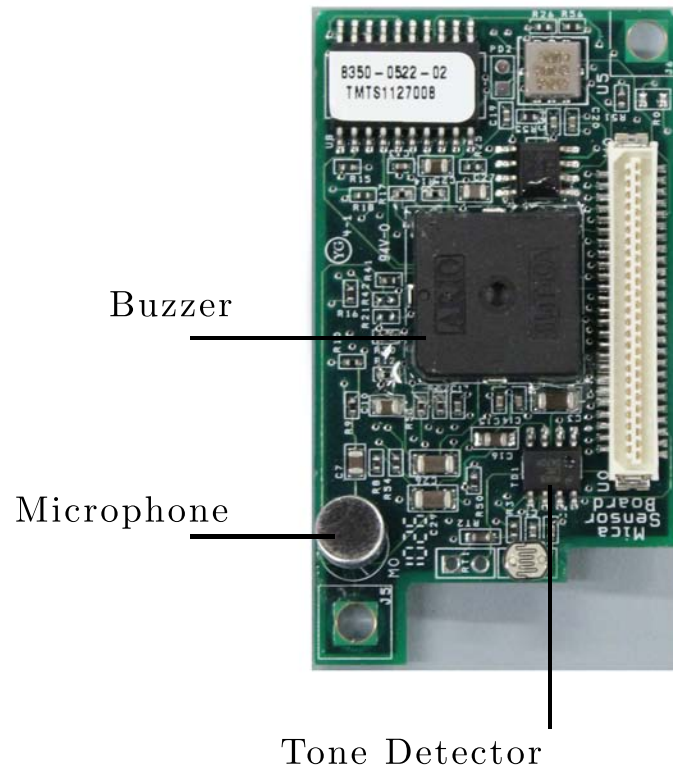
For all the experiments mentioned in this chapter, we created a testbed by placing the four anchor nodes at each corner of a 2×2 m square. The blindfolded node was then placed inside so that it maintains line-of-sight to the anchors all the time. For ease of analysis, all the nodes were placed on the same horizontal plan as illustrated in figure 2.3. Then coordinates of each anchor node and the blindfolded node were taken as (0.0,0.0), (2.0,0.0), (2.0,2.0), (0.0,2.0), and (0.5,1.5) respectively. Considering Dilution of Precision (DOP), this configuration is considered as the optimal [70]. This was considered to relax the effect of node placement in our analysis. We conducted experiments to analyse performance for both signal strength decay and time of flight based methods.



(a) IRIS



(b) MIB520/IRIS



(c) MTS310CA

Figure 2.1: Hardware

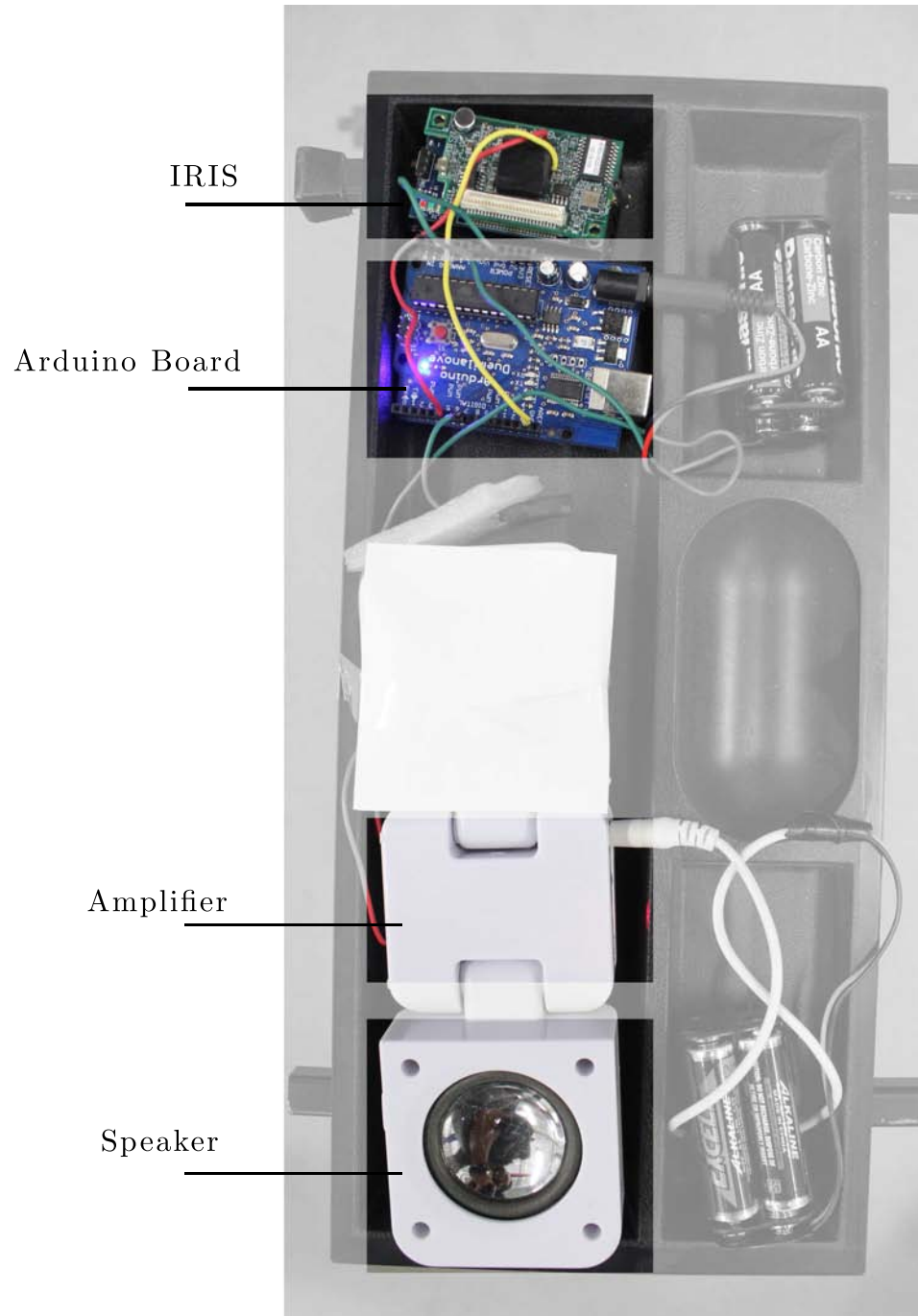


Figure 2.2: Modified sounder



Figure 2.3: Experimental setup

2.2 Using Received Signal Strength Indicator

In communication systems, Clear Channel Assessment (CCA) is a widely used method to minimize packet collisions [71]. Received signal strength of communication channel is monitored every time before each packet transmission to make sure there are no interfering concurrent transmitters. Thus, measuring received signal strength is a basic functionality in every ubiquitous radio transceiver chip. Whenever a packet is received at the radio chip, recorded received signal strength is embedded into metadata field of the received packet. Receiver Signal Strength Indicator (RSSI) is an accessible field in the packet that stores the value [71]. For convenience the term RSSI is also used to represent received signal strength hereafter.

RSSI is inversely related to distance. As distance between transmitter and receiver increases, RSSI decays [72]. Since it is always measured when two sensor nodes are communicating, RSSI is a desirable measurement to estimate distance between the two nodes. The most desirable property of using RSSI is, no additional hardware is needed to get measurements. A simple node with a wireless transmitter and a power source is sufficient. More importantly, smaller components are always easy to embed or less uncomfortable to wear.

If SN-i and SN-j are two sensor nodes communicating with each other, RSSI at i^{th} node for a packet received from j^{th} node is defined as,

$$RSSI_{i,j} = RSSI_0 - 10.n.\log_{10} \left(\frac{d_{i,j}}{d_0} \right) \quad (2.1)$$

where $d_{i,j}$ is Euclidean distance between the two nodes. $RSSI_0$ is measured signal strength at a known distance (d_0) away from the transmitter and n , the path loss exponent, is to represent signal attenuation [72]. Using model given in Equation 2.1, distance

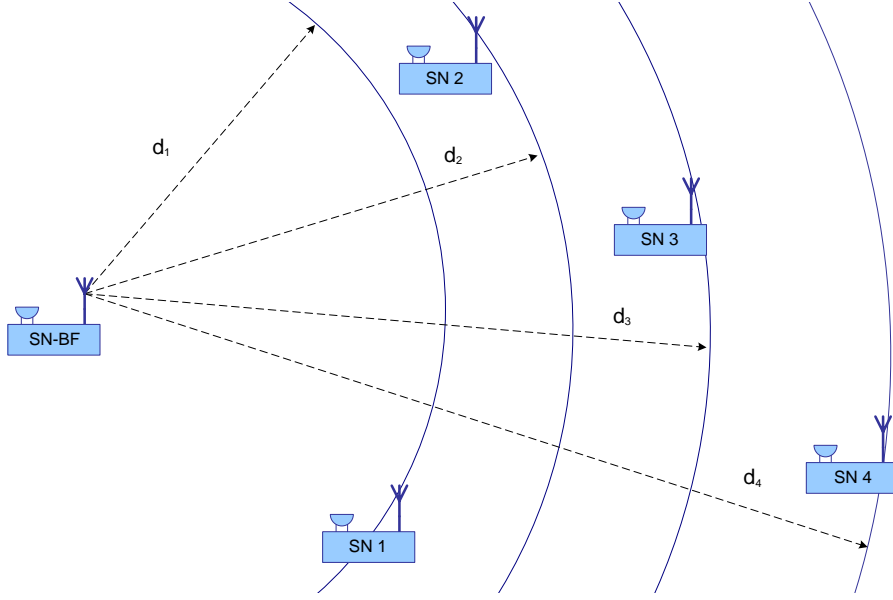


Figure 2.4: Propagation of a RF signal: “SN-x” tag represent anchor nodes and “SN-BF” represents the wearable sensor node. Black dotted lines represent distance between antenna of SN-BF and anchor nodes. Blue solid lines represent equidistant points

between the two nodes ($\hat{d}_{i,j}$) can be estimated as,

$$\begin{aligned} d_{i,j} &= d_0 10^{\frac{RSSI_0 - RSSI_{i,j}}{10n}} \\ \hat{d}_{i,j} &= d_{i,j} + \omega_{ij} \end{aligned} \quad (2.2)$$

ω_{ij} is to compensate ranging errors. Both shadowing and measurement effects cause errors. However, this model only consider measurement noise.

Consider the illustration given in Figure 2.4. SN-BF is a sensor node worn by a person at an unknown location. Suppose location of the person, hence location of SN-BF is (x,y) , and location of i^{th} anchor node is (x_i,y_i) . Then distance between each anchor node and the target can be estimated using model given in Equation 2.2. By replacing d_{BF-j} , Euclidean distance between the j^{th} sensor node and the target, with $\sqrt{(x - x_j)^2 + (y - y_j)^2}$ and neglecting noise we can write.

$$d_{BF-j} = \sqrt{(x - x_j)^2 + (y - y_j)^2} = d_0 10^{\frac{RSSI_0 - RSSI_{B,j}}{10n}} \quad (2.3)$$

squaring both side of the Equation, this can be rearranged as,

$$-2x_jx - 2y_jy + R = d_0^2 10^{\frac{RSSI_0 - RSSI_{B,j}}{5n}} - x_j^2 - y_j^2 \quad (2.4)$$

where $R = x^2 + y^2$. For N number of anchor nodes, considering $\theta = [x \ y \ R]^T$, above Equation can be written in the form of $H\theta = X_{RSSI}$ as,

$$\begin{bmatrix} -2x_1 & -2y_1 & 1 \\ -2x_2 & -2y_2 & 1 \\ \vdots & \vdots & \vdots \\ -2x_N & -2y_N & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ R \end{bmatrix} = \begin{bmatrix} d_0^2 10^{\frac{RSSI_0 - RSSI_{B,1}}{5n}} - x_1^2 - y_1^2 \\ d_0^2 10^{\frac{RSSI_0 - RSSI_{B,2}}{5n}} - x_2^2 - y_2^2 \\ \vdots \\ d_0^2 10^{\frac{RSSI_0 - RSSI_{B,N}}{5n}} - x_N^2 - y_N^2 \end{bmatrix} \quad (2.5)$$

Then using a Linear Least Squares (LLS) estimator, θ can be estimated as,

$$\hat{\theta} = (H^T H)^{-1} H^T X_{RSSI} \quad (2.6)$$

We conducted several experiments to study reliability of RSSI. RSSI was measured for four different sensor nodes inside our lab. The shortest possible packet size is a single byte. Thus, the blindfolded node was configured to broadcast a 8-bit packet in every 500 ms. Then RSSI was recorded at receiving anchor nodes for every packet reception. A total of hundred consecutive readings were taken and averaged to get a single measurement.

2.2.1 RSSI vs. Distance

All the anchor nodes and the blindfolded node were placed on a straight line. Then measurements were taken by moving the transmitting blindfolded node away from receiving anchor nodes. A single measurement was taken by averaging 100 consecutive RSSI readings. Goal of this experiment is to identify a reliable relationship between distance and RSSI. Thus, path loss exponent for the model given in Equation 2.1 was estimated by minimizing mean squared errors for all the anchor nodes. For reference

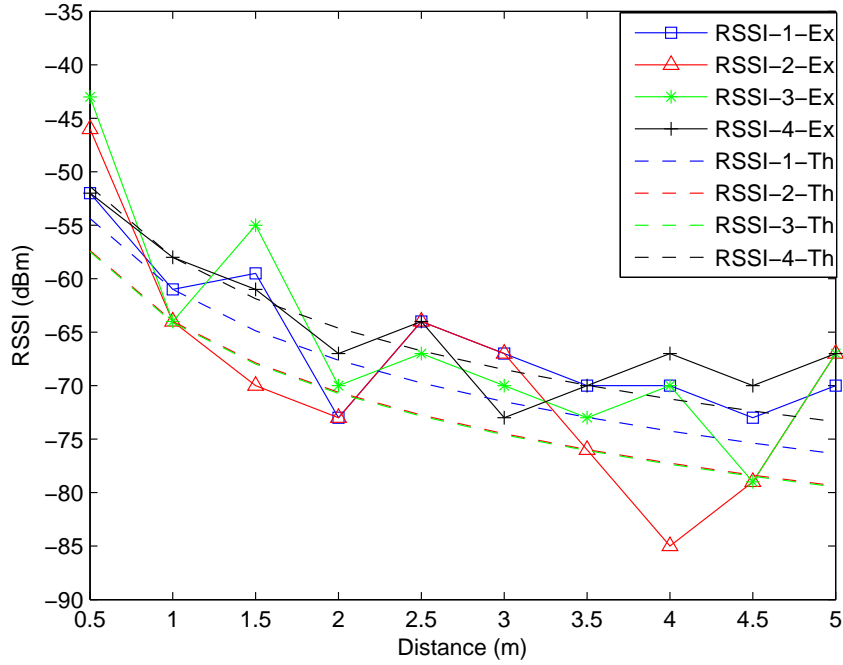


Figure 2.5: RSSI vs. distance: “RSSI- i -Ex” refers to experimental RSSI recorded from i^{th} receiving node and “RSSI- i -Th” refers to the theoretical RSSI

distance (d_0) 1 m, path loss exponent for the model was found as 2.2. Measured RSSI and RSSI generated from Equation 2.1 for varying distance are illustrated in Figure 2.5.

Though all the sensor nodes are identical, each demonstrates different variation. Though an monotonically decreasing variation is expected, it does not behave as it is expected to be. Though RSSI generally decreases with distance, there are sudden increases at random locations. Based on our experimental results we concluded RSSI is not only a function of distance between the two nodes. Since we didn’t consider antenna orientation when conducting the experiment, another experiment was performed to identify the relationship.

2.2.2 RSSI vs. Antenna Orientation

All the anchor nodes were placed 1 m away from the blindfolded node. The nodes were aligned horizontally and their antennas were aligned vertically. Then RSSI measure-

ments were taken by rotating the blindfolded node horizontally on its antenna axis. Results obtained for all the four anchor nodes are shown in Figure 2.6. It is clearly evident RSSI is also dependent on antenna orientations. Recorded RSSI is not uniformly distributed around antenna axis. Each has different variation for each angle. We concluded antennas attached to IRIS[®] nodes are not omnidirectional. Thus, RSSI model given in Equation 2.1 should be calibrated for each antenna orientation.

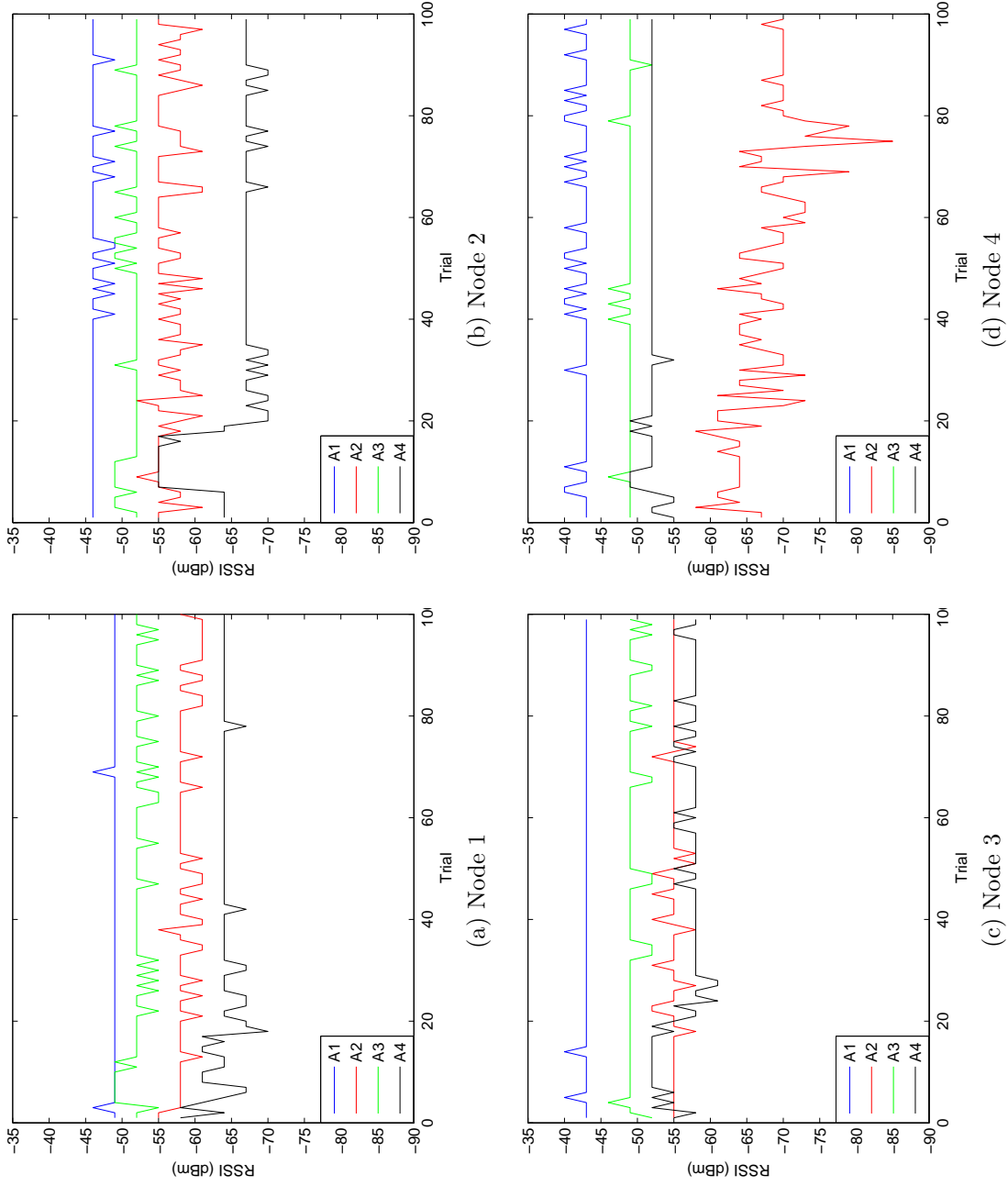


Figure 2.6: RSSI vs. antenna orientation: Lines A1, A2, A3, and A4 represent recorded RSSI values for 0° , 90° , 180° , and 270° antenna orientations respectively

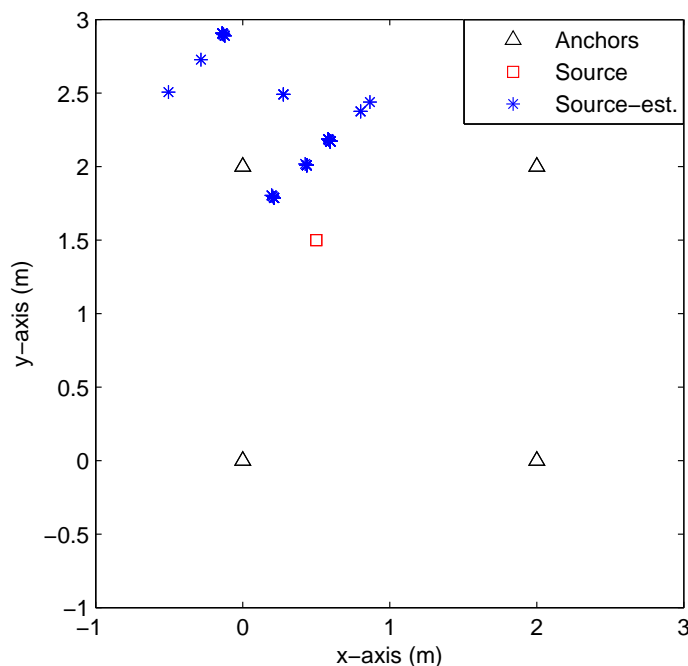


Figure 2.7: Localization using RSSI: Average localization error = 1.0098 m, variance = 0.3081 m^2 , bias = 0.7922 m

2.2.3 Localization Performance

Then we placed anchor nodes on corners of a 2×2 m square as described earlier. The blindfolded node placed at (0.5,1.5) m was localized using RSSI measurements and LLS estimator derived in Equation 2.6. RSSI measurements were taken by averaging 10 consecutive readings. Localization results for 100 estimates are shown in Figure 2.7.

After an intensive calibration for both distance and antenna orientation, average error of localization has been reduced to 1.098 m with 0.3081 m^2 variance. However, there is a 0.7922 m average bias. All the location estimates are biased toward the closest anchor node. Using this results we can only identify location of the blindfolded node to the level of neighbourhood anchor nodes. Since it couldn't demonstrate fine grained results even when line-of-sight was enforced, we concluded RSSI is not reliable for person localization.

2.3 Time of Arrival

A propagating signal takes time to reach its destination. Receivers that are far from the transmitter take more time to receive the signal than receivers that are closer. Propagation time is a function of distance between the transmitter and the receiver. This property is used in Time of Arrival (TOA) based approach. Since measuring time is a common functionality in most of the microprocessors, each sensor node can measure reception time when it receives a signal. If exact time of the signal transmission is known, distance between the transmitter and the receiving sensor node can be estimated.

When sensor node SN-j receives a signal transmitted by node SN-i, distance between the two node can be expressed as,

$$\begin{aligned}
 d_{i,j} &= C.(t_{rx-j} - t_{tx-i}) \\
 &= C.TOA_{i,j} \\
 \hat{d}_{i,j} &= d_{i,j} + w_{i,j} \\
 &= C.TOA_{i,j} + w_{i,j}
 \end{aligned} \tag{2.7}$$

where t_{tx-i} is transmission time of the signal from i^{th} node and t_{rx-j} is reception time at j^{th} node. C is propagation speed of the signal in a given medium. For the same configuration given in Figure 2.4, for a person wearing a sensor node at (x,y), TOA measurements for N number of anchor nodes can be written in the form of $H\theta = X_{TOA}$ as,

$$\begin{bmatrix} -2x_1 & -2y_1 & 1 \\ -2x_2 & -2y_2 & 1 \\ \vdots & \vdots & \vdots \\ -2x_N & -2y_N & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ R \end{bmatrix} = \begin{bmatrix} C^2TOA_{1,B}^2 - x_1^2 - y_1^2 \\ C^2TOA_{2,B}^2 - x_2^2 - y_2^2 \\ \vdots \\ C^2TOA_{N,B}^2 - x_N^2 - y_N^2 \end{bmatrix} \tag{2.8}$$

Then a LLS estimator can be derived to find unknown parameter θ as,

$$\hat{\theta} = (H^T H)^{-1} H^T X_{TOA} \tag{2.9}$$

Since two time measurement are compared from two different nodes, their clocks should be synchronized. To minimize errors, synchronization error should be smaller than clock resolution. Propagation speed of Radio Frequency (RF) signals in free air is approximately 3×10^8 m/s. If RF signals (beacon packets) are used, to get a distance measurement with 1 m resolution a time measurement with 10 ns resolution is needed. Expensive high resolution clocks are needed to achieve time measurements with nanosecond resolution. Though they are available in GPS receivers, we cannot expect that from an ordinary sensor node [73]. Economical nodes can only give stable time measurements with sub-millisecond accuracy [61]. Thus we concluded RF signals are not feasible to use for TOA based localization in WSN.

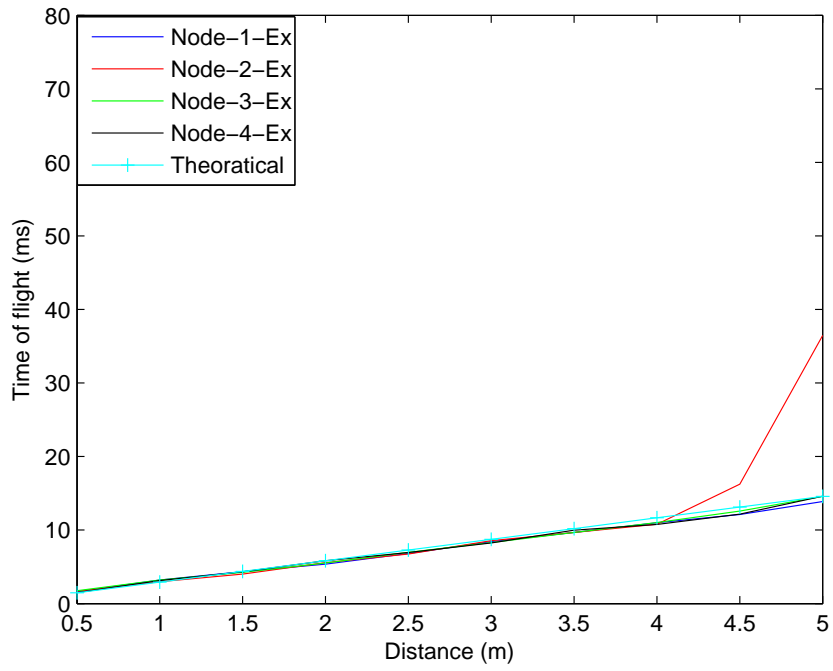
However, acoustic and ultrasound signals propagate at 343 m/s in dry air (at 20 °C) [74]. If acoustic or ultrasound signals were used, to get the same distance resolution, only a time measurement with 3 ms resolution is needed. This is feasible with the existing hardware on IRIS[®] nodes. Since each of our sensor board (MTS310CA) has a 4.3 kHz buzzer and a tone detector configured to the same frequency, we selected acoustics over ultrasound signals.

We conducted similar experiments to study reliability of TOA. Since the buzzer has a limited coverage, we used the modified sounder instead. The blindfolded node was configured to broadcast a 8-bit packet, and after 10 ms, to switch-on the sounder for 100 ms long period of time. The anchor nodes were configured to record reception time and switch-on the tone detector when it receives the packet. Time is again recorded when the tone detector detects the signal. Since our setup is only 2×2 m wide, propagation time of RF signals are negligible. Thus, TOA for each anchor node was estimated taking time difference of RF and acoustic arrivals. A 10 ms delay was maintained to accommodate all the transmission latencies and delays at the anchor nodes when switching-on the tone detector. We experimentally found the delay should be at least 8ms. A total of hundred

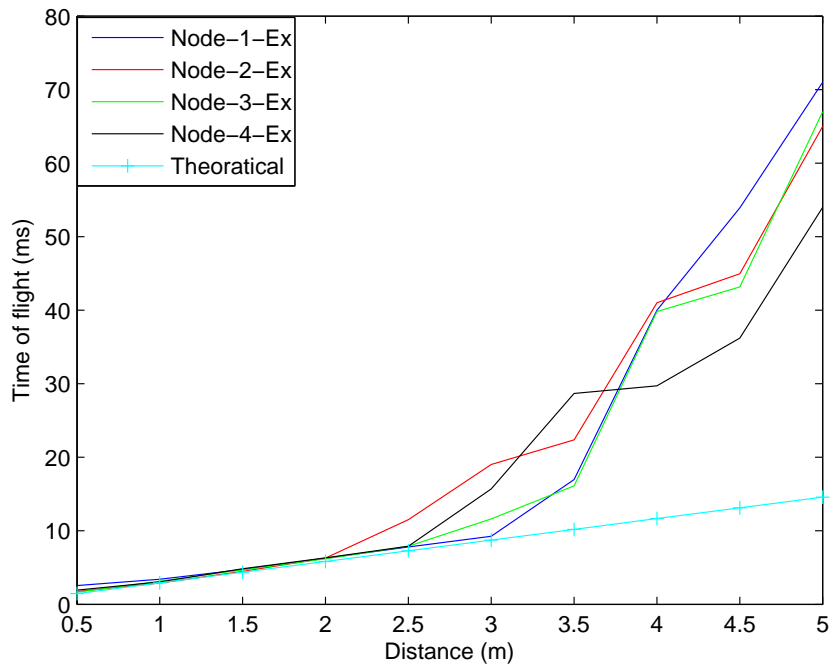
consecutive readings were taken and averaged to get a single TOA measurement.

2.3.1 TOA vs. Distance

All the anchor nodes and the blindfolded node were placed on a straight line. Then TOA measurements were taken by moving the blindfolded node away from the anchor nodes. Since microphones and speakers are directional, two different experiments were performed with/without maintaining line-of-sight between the sounder and the microphone. Speed of sound inside the lab was estimated by minimizing mean squared error for all the sensor nodes. For reference distance (d_0) 1 m, speed of sound for the model given in Equation 2.7 was found as 340.91 m/s. Measured TOA and TOA generated from the model for variable distance are illustrated in Figure 2.8.



(a) Line-of-sight case



(b) Ordinary case

Figure 2.8: TOA vs. distance: “Node-Ex- i ” refers to measured TOA for i^{th} anchor node.

“Theoretical” refers to TOA generated using Equation 2.7

Our experimental results for TOA demonstrates stronger relationship with distance. TOA monotonically increases as it is expected to be. All the anchor nodes demonstrate nearly identical results and TOA generated using theoretical model closely resembles them. However as distance increases, signal strength of the acoustic signal decreases. Thus, signal detection at the tone detector become unreliable. Thus, TOA accuracy decreases. For line-of-sight enforced case, it deviates from the model after 4 m and for ordinary case after 2 m. That is due to directional properties of microphones and speakers. Sound strength is not uniformly distributed in all the directions.

2.3.2 Localization Performance

Then we performed an experiment to study performance of TOA measurements in localization. Four anchor nodes were placed on corners of a square as did before. The blindfolded node with the sounder was placed at (0.5,1.5) m. TOA measurements was taken by averaging 10 consecutive readings. Then the blindfolded node was localized using the LLS estimator derived in Equation 2.9. Localization results for 100 estimates are shown in Figure 2.9.

TOA based approach demonstrate superior results. After calibrating for speed of sound, average localization error of 0.3280 m with 0.1013 m² variance has been recorded. Estimate precision increased considerably. Bias has reduced to 0.0645 m. Since it demonstrated fine grained results, we concluded time of TOA (or time of flight) is a reliable measurement for person localization.

2.4 Time Difference of Arrival

Compared to RSSI approach, TOA gives reliable fine-grained location estimates. However, the time of transmission must be available to accurately estimate time of flight. In previous experiment it was obtained by simultaneously transmitting a RF signal. This

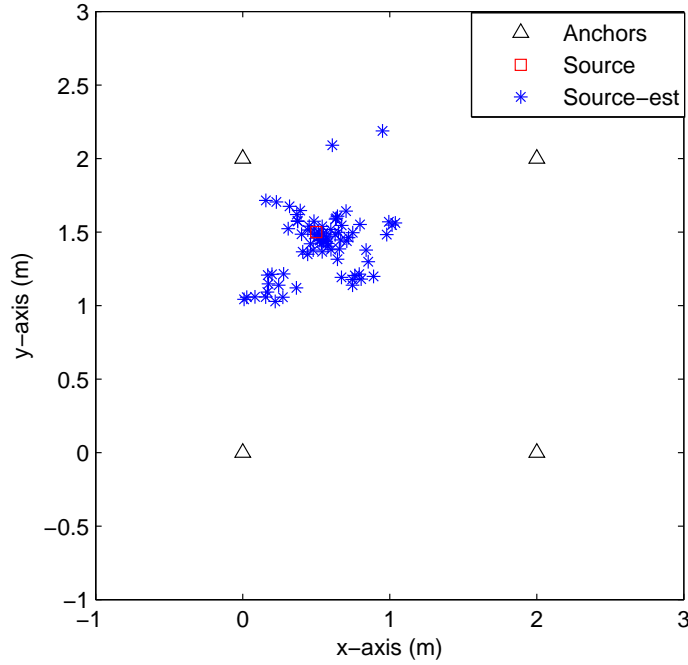


Figure 2.9: Localization using TOA: Average localization error = 0.3280 m, variance = 0.1013 m², bias = 0.0645 m

is only possible when localizing sensor nodes. If time of transmission can be removed from the formulation, any type of acoustic signal source can be localized. [29] proposes a method of taking Time Difference of Arrivals (TDOA) between two anchor nodes. Since time of transmission is common, it cancels out when taking the difference.

A single RSSI or TOA measurement restricts target location to a circle. Exact location is then given by intersecting points of at least three circles drawn for measurements from three different anchor nodes. But as illustrated on Figure 2.10, each TDOA measurement restricts target location into a half hyperbola whose focal points are located at the two anchor nodes involved in calculating the TDOA. Exact location is given by intersection point of at least two hyperbolas. Since measurements are affected by noise, additional sensors are needed to get an accurate location estimate.

Consider the sensor configuration given in Figure 2.11. There are four anchor nodes SN-1, SN-2, SN-3 and SN-4 located at (x_1, y_1) , (x_2, y_2) , (x_3, y_3) and (x_4, y_4) respectively.

A person with a beaconing sensor node (SN-BF) is standing at (x,y). When all the sensors are on the same horizontal plane, $d_{B,1}$, $d_{B,2}$, $d_{B,3}$ and $d_{B,4}$ are the respective distances from each anchor node to the target and TOA_1 , TOA_2 , TOA_3 and TOA_4 are time of flight measurements between the target and each node. TOA is calculated using,

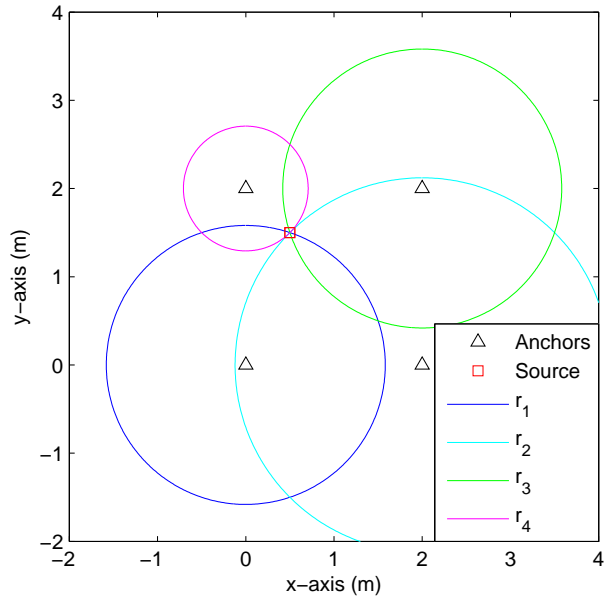
$$TOA_i = \frac{d_{B,i}}{C} = t_{rx-i} - t_{tx-B} \quad (2.10)$$

Each symbol has its usual representation and t_{tx-B} is transmission time of signal from the signal source. TDOA is defined as difference of TOA between two nodes. Thus, it can be calculated w.r.t. SN-1 as,

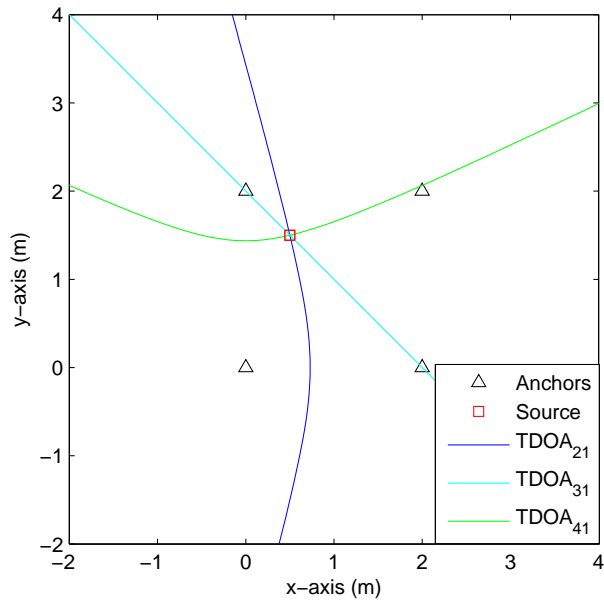
$$\begin{aligned} TDOA_{i,1} &= TOA_i - TOA_1 \\ &= \frac{d_{B,i} - d_{B,1}}{C} \\ TDOA_{i,1} &= t_{rx-i} - t_{rx-1} \end{aligned} \quad (2.11)$$

Time of transmission is then cancelled out since it is the same for all the sensor nodes. Model given in Equation 2.11 can be rewritten as,

$$\begin{aligned} d_{B,i} &= d_{B,1} + C.TDOA_{i,1} \\ \hat{d}_{B,i} &= d_{B,i} + \omega_i \end{aligned} \quad (2.12)$$



(a) Spherical intersection



(b) Hyperbolic intersection

Figure 2.10: Comparison between RSS/TOA and TDOA based approach: SN-1, SN-2, SN-3, SN-4 and a blindfolded node (a signal source) are located at (0,0), (2,0), (2,2), (0,2) and (0.5,1.5) respectively. r_i is distance between the blindfolded node and i^{th} anchor node. $TDOA_{ij}$ is TDOA calculated between i^{th} and j^{th} anchor nodes

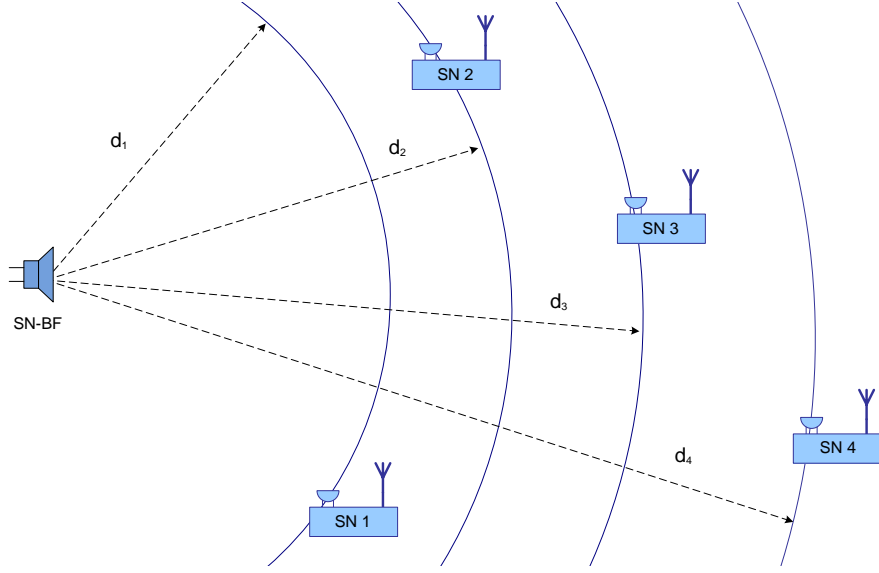


Figure 2.11: Propagation of acoustic signals: Blue boxes with “SN-x” tag represent sensor nodes and “Speaker” symbol represents a person speaking. Black dotted lines represent distance between the target and each microphone. Blue solid lines represent equidistant points with given distance away from the target

Then for N number of anchor nodes, neglecting effects of noise, it can be written in the form of $H\theta = X_{TDOA}$ as,

$$\begin{bmatrix} 2(x_1 - x_2) & 2(y_1 - y_2) & -2TDOA_{2,1}C \\ 2(x_1 - x_3) & 2(y_1 - y_3) & -2TDOA_{3,1}C \\ \vdots & \vdots & \vdots \\ 2(x_1 - x_N) & 2(y_1 - y_N) & -2TDOA_{N,1}C \end{bmatrix} \begin{bmatrix} x \\ y \\ d_{B,1} \end{bmatrix} = \begin{bmatrix} x_1^2 + y_1^2 - x_2^2 - y_2^2 + C^2TDOA_{2,1}^2 \\ x_1^2 + y_1^2 - x_3^2 - y_3^2 + C^2TDOA_{3,1}^2 \\ \vdots \\ x_1^2 + y_1^2 - x_N^2 - y_N^2 + C^2TDOA_{N,1}^2 \end{bmatrix} \quad (2.13)$$

where $\theta = [x \ y \ d_{B,1}]^T$. Then a LLS estimator to find unknown parameter θ from TDOA model can be derived as,

$$\hat{\theta} = (H^T H)^{-1} H^T X_{TDOA} \quad (2.14)$$

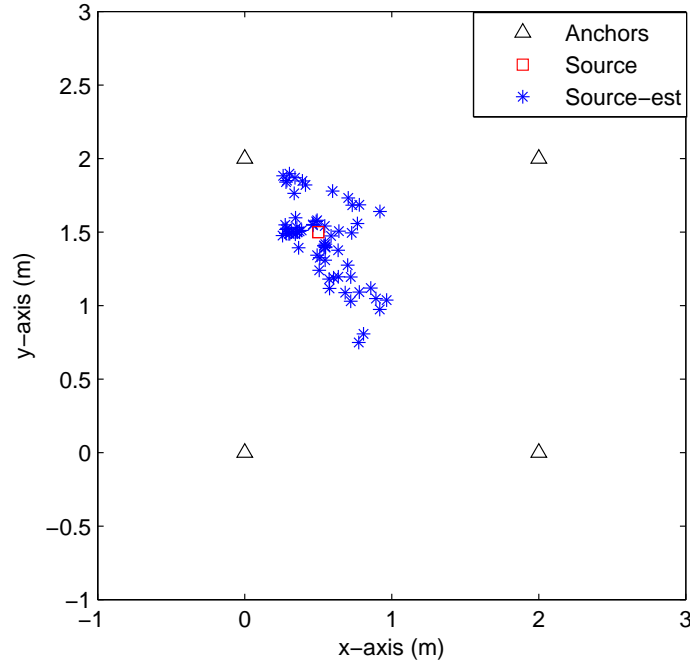


Figure 2.12: Localization using TDOA: Average localization error = 0.3320 m, variance = 0.1078 m², bias = 0.0871 m

2.4.1 Localization Performance

To study performance of TDOA measurements in localization, we performed a similar experiment. Four anchor nodes were placed on corners of a square and the blindfolded node with the sounder at (0.5,1.5) m. TDOA measurements was taken by averaging 10 consecutive readings. Then the blindfolded node was localized using the LLS estimator derived in Equation 2.14. Localization results for 100 estimates are shown in Figure 2.12.

Localization performance is similar to TOA approach. Average localization error of 0.3320 m with 0.1078 m² variance has been recorded. Bias has slightly increased to 0.0871 m. Compared to RSSI based approach TDOA also demonstrates fine-grained results. TDOA is completely independent of transmission time of the signal. No additional attention needs to be given to identify transmission time of the signal. There are no

Method	Average Error (m)	Error Variance (m ²)	Bias (m)
RSSI	1.0098	0.3081	0.7922
TOA	0.3280	0.1013	0.0645
TDOA	0.3320	0.1078	0.0871

Table 2.1: Localization performance

restrictions for the signal source. It could be a wearable sensor node transmitting an acoustic/ultrasound beacon or it could be any other sound source. Thus we concluded TDOA based approach is both reliable and practical.

2.5 Chapter Summary

Person localization can be performed by using wearable sensors. Based on measurement used, it can be categorized into two groups as signal strength decay and time of flight. Decaying signal strength is more vulnerable to environmental changes than time of flight. Fading, multipath and obstacles make the measurements unreliable. Since antennas on sensor nodes are not always omnidirectional, calibration is needed for all the antenna orientations. Thus signal strength decay is not a reliable measurement.

Time of flight measurements are more reliable compared to signal strength. Relationship between distance and time of flight is stronger than that with signal strength. Thus as given in Table 2.1, time of flight based approach results in fine-grained estimates. It can be categorized into two groups as TOA and TDOA. Transmission time of the signal is needed for TOA but it is not needed for TDOA. Since identifying time of transmission cannot be done without prior knowledge or using additional hardware, TDOA method has an advantage over the TOA.

Chapter 3

TDOA Based Localization

The approaches mentioned in the previous chapter need the person being localized to wear a sensor node. Though components need to be worn for RSSI approach are lesser, its accuracy is poorer. Though TOA approach give fine-grained location estimates, knowing time of transmission is crucial. Nevertheless listening to a stream of continuous beaconing can be frustrating.

However, a talking person can be treated as a wideband acoustic signal source [75]. If time of flight for speech can be estimated, localization can be performed without using inconvenient beacons. If TOA is used, the time when the person started speaking must be known. Since it is not deterministic, only way of identifying it is by placing a sensor node right next to speakers mouth, which is intrusive and impractical. However if TDOA approach is used, the speaker can be localized conveniently without expecting him/her to wear any electronic device.

3.1 Acoustic Model

For ease of analysis, we need to have a model for acoustic wave propagation. Here, a person speaking is treated as an acoustic signal source.

3.1.1 Direct Propagation

Received signals at two spatially separated anchor nodes are generally modelled as,

$$\begin{aligned}x_1(t) &= s(t) + \omega_1(t) \\x_2(t) &= s(t + \tau_0) + \omega_2(t)\end{aligned}\tag{3.1}$$

where $s(t)$ is the signal, $\omega_i(t)$ is zero mean uncorrelated additive Gaussian noise at i^{th} anchor node and τ_0 is TDOA or delay between the two signals. Assuming distance between the nodes is negligible compared to distance from the signal source, this model assumes zero relative signal attenuation. It is only possible for far-field signals [76]. Due to hardware limitations, our setup is 2×2 m wide, which is a near-field signal environment. Thus, for near-field signals, this model should be modified compensating attenuation effects.

$$\begin{aligned} x_1(t) &= \alpha_1 s(t) + \omega_1(t) \\ x_2(t) &= \alpha_2 s(t + \tau_0) + \omega_2(t) \end{aligned} \tag{3.2}$$

where α_i is signal attenuation at i^{th} node relative to transmitted signal. When there are multiple concurrent sources, resulted received signal can be treated as summation of signals. Thus, the model can be further extended as,

$$\begin{aligned} x_1(t) &= \sum_{j=1}^M \alpha_{1j} s_j(t) + \omega_1(t) \\ x_2(t) &= \sum_{j=1}^M \alpha_{2j} s_j(t + \tau_j) + \omega_2(t) \end{aligned} \tag{3.3}$$

Here α_{ij} is signal attenuation at i^{th} node for a received signal from j^{th} signal source. τ_j is TDOA for a signal transmitted from j^{th} source and M is number of active concurrent signal sources. Though this model has been widely used, direct path propagation is idealistic [77],[78],[79]. Direct path propagation is only possible in large open areas where there are no reflective surfaces. Our motivation is localizing people indoors using speech. Since multipath effects exist, such as indoor reverberation, this model is not suitable for our application.

3.1.2 Multipath Propagation

By considering multipath effects, propagation delays and signal attenuation, received signal at an anchor node can be modelled as a convolution between the transmitted

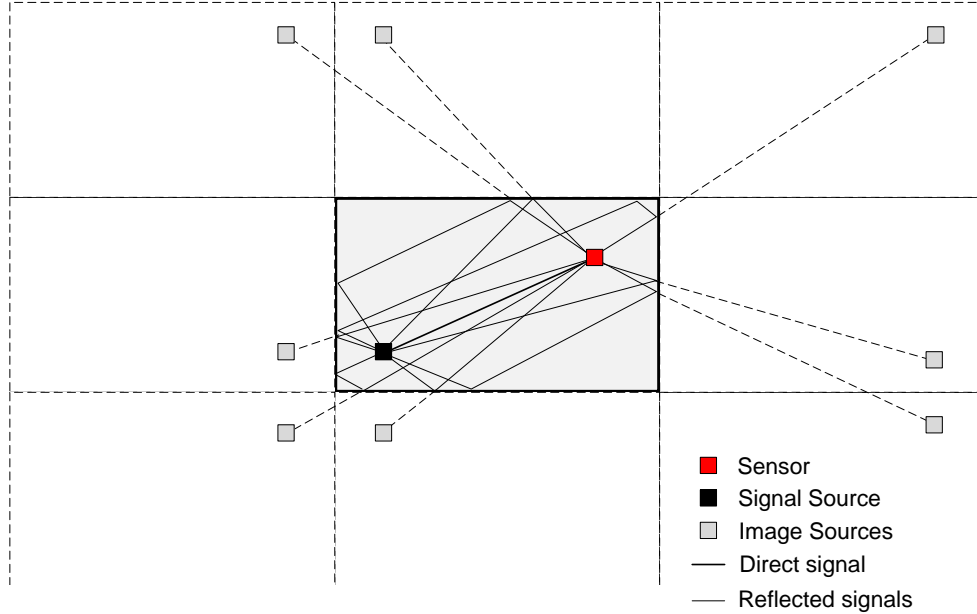


Figure 3.1: Image-Source method for room acoustics: The dark rectangle in the middle represents a room with four walls. Other dashed rectangles represent images of the room. The small black square represents a signal source and the red square represents an anchor node. Lines between the two squares represent signal propagation paths.

signal and a transfer function [80], [81].

$$\begin{aligned} x_1(t) &= h_1(t) * s(t) + \omega_1(t) \\ x_2(t) &= h_2(t) * s(t) + \omega_2(t) \end{aligned} \quad (3.4)$$

where ‘*’ is convolution operator. $h_i(t)$ is impulse response for propagation channel between the signal source and i^{th} anchor node which is also called “Room Transfer Function” [81]. As long as the environment, location of the sound source and anchor nodes are fixed, room transfer function is time invariant [81]. It can be decomposed into two terms [81].

$$h_i(t) = h_i^d(t) + h_i^r(t) \quad (3.5)$$

That is, as direct path propagation component ($h_i^d(t)$) and as diffusive multipath component ($h_i^r(t)$). In [82], a method called “Image-Source method” has been proposed to estimate the room transfer function.

Consider the illustration given in Figure 3.1. An empty rectangular room with a

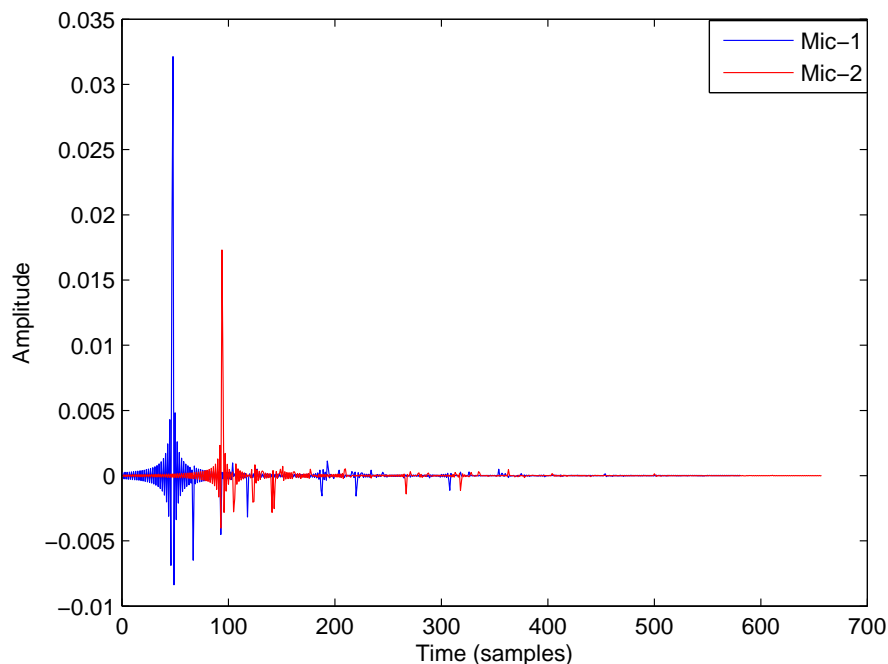


Figure 3.2: Room Impulse Response: An empty room of $8.7 \times 7.4 \times 4$ m size is considered. Two sensor nodes and a signal source were placed at (3,1), (1,1) and (5,1) at same height respectively. All the surfaces were assumed to have uniform absorption properties and absorption coefficients were set by fixing reverberation time to 0.1 s.

signal source and a receiving anchor node is given. Multipath exists because of reflecting surfaces. Hence, all the walls are assumed as mirrors. All the reflecting waves are treated, as if they are coming from images of the signal source on the “mirrors”. Then each wave is attenuated to represent signal absorption at each reflective surface and added together to get the multipath effect [82].

Room transfer functions between a sound source and two sensor nodes generated using Image-Source method are illustrated in Figure 3.2. Two dominant impulses at around 50th and 100th samples represent directly propagated signals. Their amplitude represent signal attenuation. Other minor impulses represent signals receiving as multipath.

For multiple concurrent sound sources, the model given in Equation 3.4 can also be

extended as,

$$\begin{aligned} x_1(t) &= \sum_{j=1}^M h_{1j}(t) * s_j(t) + \omega_1(t) \\ x_2(t) &= \sum_{j=1}^M h_{2j}(t) * s_j(t) + \omega_2(t) \end{aligned} \quad (3.6)$$

where $h_{ij}(t)$ is room transfer function for signal transmission between j^{th} source and i^{th} sensor node.

3.2 Cross-Correlation based Time Delay Estimation

Since hardware tone detectors are tuned for a specific narrowband signal, they cannot be used to detect voice. For wideband signals, cross-correlation has been widely used to estimate time delay between signals [83]. Since speech is also wideband, it is suitable to estimate TDOA between two spatially separated anchor nodes. If τ_0 is time delay between the two signals, the estimator can be formulated as,

$$\hat{\tau}_0 = \arg \max_{\tau} R\{x_1(n), x_2(n - \tau)\} \quad (3.7)$$

where $R\{.\}$ is cross-correlator, $x_i(n)$ is recorded signal at i^{th} anchor node, τ is shift in samples and $\hat{\tau}_0$ is estimated TDOA.

As an example assume there are two recorded signals. One signal is delayed version of the other. Cross-correlation is performed between the two signals iteratively shifting a selected signal at each iteration. Output of the cross-correlator maximizes only when both the signals are aligned. Thus, the shift that maximizes the output is a reliable estimate for time delay between the two original signals [84].

Cross-correlation between a signal and a shifted signal is defined as,

$$R_{x_1x_2}(\tau) = E\{x_1(t).x_2(t - \tau)\} \quad (3.8)$$

where $E\{.\}$ is the expectation operator. For faster implementation, this can be written in frequency domain as,

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} X_1(f).X_2(f)^* e^{j2\pi f\tau} df \quad (3.9)$$

Filter Name	Weighing function, $\psi(f)$	Reference
Cross Correlator	1	[83]
Roth Processor	$1/G_{x_1x_1}(f)$	[87]
Phase Transformation	$1/ G_{x_1x_2}(f) $	[88]
MLE	$ G_{x_1x_2}(f) /[G_{w_1w_1}(f)G_{x_2x_2}(f) + G_{w_2w_2}(f)G_{x_1x_1}(f)]$	[89]

Table 3.1: Popular weighing functions for GCC: $G_{as}(f)$ is power spectral density of $a(t)$ and $G_{ab}(f)$ represents cross power spectral density of $a(t)$ and $b(t)$. $|\cdot|$ represents absolute value operator

where $X_i(f)$ is Fourier transformed $x_i(t)$ and $\{\cdot\}^*$ is complex conjugate. This is Fourier inverse of cross power spectral density [85]. Cross power spectral density of $x_1(t)$ and $x_2(t)$ is defined as,

$$G_{x_1x_2}(f) = X_1(f).X_2(f)^* \quad (3.10)$$

Thus, cross-correlation in frequency domain can be re-written as,

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} G_{x_1x_2}(f)e^{j2\pi f\tau} df \quad (3.11)$$

or

$$R_{x_1x_2}(\tau) = \mathcal{F}^{-1}\{G_{x_1x_2}(f)\} \quad (3.12)$$

3.2.1 Generalized Cross Correlation

Taking cross-correlation between filtered signal is more reliable than taking cross-correlation between raw samples [86]. If $\psi_1(t)$ and $\psi_2(t)$ are impulse responses for respective filter functions, filtered cross-correlation can be written as,

$$R_{x_1x_2}(\tau) = E\{[\psi_1(t) * x_1(t)].[\psi_2(t) * x_2(t - \tau)]\} \quad (3.13)$$

Considering $\Psi(f) = \mathcal{F}\{\psi_1(t)\}\mathcal{F}\{\psi_2(t)\}$, the above Equation can be expressed as,

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} \Psi(f).G_{x_1x_2}e^{j2\pi f\tau} df \quad (3.14)$$

This is called Generalized Cross-Correlation (GCC) or Weighted Cross-Correlation [86]. Popular filter functions of GCC are summarized in Table 3.1. Each filter function has different design considerations. Thus, each performs differently. For ease of analysis consider direct path propagation model given in Equation 3.2. Assuming relative attenuation between the two sensor nodes as α , it can be re-written as,

$$\begin{aligned}x_1(t) &= s(t) + \omega_1(t) \\x_2(t) &= \alpha s(t + \tau_0) + \omega_2(t)\end{aligned}\tag{3.15}$$

and the Fourier transformed version as,

$$\begin{aligned}X_1(f) &= S(f) + W_1(f) \\X_2(f) &= \alpha S(f) \cdot e^{j2\pi f \tau_0} + W_2(f)\end{aligned}\tag{3.16}$$

Then we can write,

$$\begin{aligned}G_{x_1x_1}(f) &= X_1(f) \cdot X_1(f)^* \\&= [S(f) + W_1(f)] \cdot [S(f) + W_1(f)]^* \\&= S(f) \cdot S(f)^* + S(f) \cdot W_1(f)^* + W_1(f) \cdot S(f)^* + W_1(f) \cdot W_1(f)^* \\&= G_{ss}(f) + G_{sw_1}(f) + G_{sw_1}(f) + G_{w_1w_1}(f) \\G_{x_1x_2}(f) &= X_1(f) \cdot X_2(f)^* \\&= [S(f) + W_1(f)] \cdot [\alpha S(f) e^{j2\pi f \tau_0} + W_2(f)]^* \\&= S(f) \cdot \alpha S(f)^* e^{-j2\pi f \tau_0} + S(f) \cdot W_2(f)^* + W_1(f) \cdot \alpha S(f)^* e^{-j2\pi f \tau_0} + W_2(f) \cdot W_1(f)^* \\&= \alpha G_{ss}(f) e^{-j2\pi f \tau_0} + G_{sw_2}(f) + \alpha G_{sw_1}(f) e^{-j2\pi f \tau_0} + G_{w_1w_2}(f)\end{aligned}\tag{3.17}$$

For uncorrelated signals, cross power spectral density is assumed negligible [86]. Thus,

$$\begin{aligned}G_{x_1x_1}(f) &= G_{ss}(f) + G_{w_1w_1}(f) \\G_{x_1x_2}(f) &= \alpha G_{ss}(f) e^{-j2\pi f \tau_0}\end{aligned}\tag{3.18}$$

For example consider GCC with Roth filter (GCC-Roth).

$$R_{x_1x_2}(\tau) = \frac{\int_{-\infty}^{\infty} G_{x_1x_2}(f) e^{j2\pi f \tau} df}{\int_{-\infty}^{\infty} G_{x_1x_1}(f) df}\tag{3.19}$$

Using Equations in 3.18 this can be rewritten as,

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} \frac{\alpha G_{ss}(f)}{G_{ss}(f) + G_{w_1w_1}(f)} e^{j2\pi f(\tau-\tau_0)} df \quad (3.20)$$

Then using properties of convolution with a delta function, this can be written as,

$$R_{x_1x_2}(\tau) = \delta(\tau - \tau_0) * \int_{-\infty}^{\infty} \frac{\alpha G_{ss}(f)}{G_{ss}(f) + G_{w_1w_1}(f)} e^{j2\pi f(\tau)} df \quad (3.21)$$

where $\delta(\cdot)$ is Dirac delta function.

Roth filter gives lower weights for frequency bins with higher noise spectral power density ($G_{w_1w_1}(f)$) and it gives higher weights for frequency bins with lower noise spectral power density. Convolution with a delta function results a delta function only when the function being convoluted is also a delta function. To have an impulse only at τ_0 , the second term of the convolution given in Equation 3.21 has to be an impulse. This happens only when $G_{w_1w_1}(f)$ is an integer multiple of $G_{ss}(f)$. Since it is not common, GCC-Roth output spreads out.

GCC with Phase Transformation (GCC-PHAT) can also be analysed the same way. For the same signal model used above, GCC-PHAT is defined as,

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} \frac{G_{x_1x_2}(f)}{|G_{x_1x_2}(f)|} e^{j2\pi f\tau} df \quad (3.22)$$

As did with GCC-Roth, this can be re-written as,

$$R_{x_1x_2}(\tau) = \delta(\tau - \tau_0) * \int_{-\infty}^{\infty} \frac{G_{ss}(f)}{|G_{ss}(f)|} e^{j2\pi f\tau} df \quad (3.23)$$

PHAT filter “normalizes” all the frequency bins. Every bin is weighted based on magnitude of cross power spectra ($G_{x_1x_2}(f)$). The term $\frac{G_{ss}(f)}{|G_{ss}(f)|}$ returns phase of $G_{ss}(f)$. This makes the second term of the convolution also an impulse. Thus, spreading effects are minimum in GCC-PHAT compared to GCC-Roth processor [86]. Thus, the method results a more distinguishable impulse at τ_0 compared to the former method.

Maximum Likelihood Estimator (MLE) was derived considering probabilistic distribution of noise in the signal as Gaussian [89]. As long as Signal to Noise Ratio (SNR)

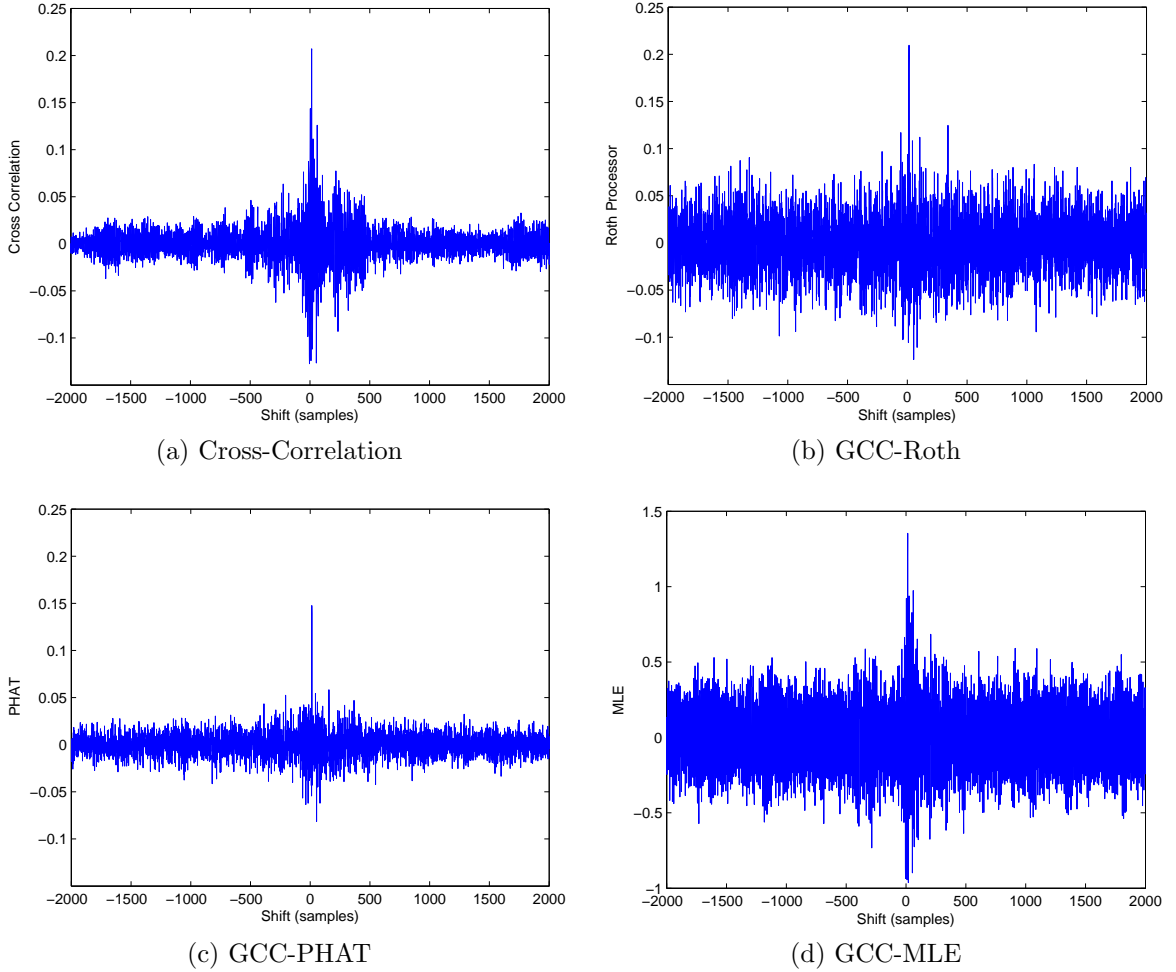


Figure 3.3: Comparison between different GCC methods: GCC was performed on two acoustic signals generated from a 8 kHz sampled wave file using the Image-Source model. The wave file contains word “Hello” in male voice. For the model, two sensor nodes were placed 2m apart (47 sample delay) and reverberation was set to 0.1 s.

is the concern, it outperforms all the other filter functions used here [90]. For MLE to work robustly, assumed statistical properties of noise should be stationary and accurate. In case they are not known or deterministic, calibration is needed.

With suitable filter weights, delay between two signals can be found by finding parameter τ that maximize the GCC function. i.e.,

$$\hat{\tau}_0 = \arg \max_{\tau} \{R_{x_1 x_2}(\tau)\} \quad (3.24)$$

For visual comparison, GCC generated with different filter weights are illustrated in

Figure 3.3. For each GCC, $R_{x_1x_2}(\tau)$ was obtained by varying shift, τ , from -2000 to 2000 samples. i.e. by shifting a selected signal 2000 samples in both the directions relative to the other. For all the cases, SNR was set to 30 dB to neglect effects from noise and reverberation was set at 0.1 s to minimize multipath effects. The time delay estimate, $\hat{\tau}_0$, is the abscissa that results the highest output.

Since each filter weights GCC differently, each produces different output. When $\Psi(f) = 1$, the resulted GCC function is simply the cross-correlation. As illustrated in part (a), it does not produce a “narrow” impulse-like peak at the true signal delay (i.e. at τ_0). Instead, there is a peak with multiple “spikes” around the true delay. When reverberation (multipath effects) is dominant, the spikes are even more severe. Thus, correct peak is less distinguishable and identifying it is even more difficult. However, both Roth and PHAT filters produce more distinguishable impulse-like peaks at τ_0 . As shown in part (b) and (c) respectively, they are more distinguishable. GCC-Roth has highly varying output compared to GCC-PHAT. That is because, Roth puts higher weights for frequencies with lower noise spectra and we set a higher SNR (30 dB). Performance of GCC-MLE is not clearly visible in part (d). Thus, another case was simulated with varying SNR.

Figure 3.4 compares percentage of anomalies with SNR. An anomaly is any delay estimate other than 47 samples. For the two signals generated using Image-Source model, uncorrelated zero mean Gaussian noise were added to get desired SNR. SNR was varied from -10 to 10 dB and at each SNR, time delay between the two signals was estimated using GCC methods discussed. Since true delay between the two signals is 47 samples, an anomaly was defined as any delay estimated other than 47 samples. As shown in the Figure, GCC-MLE outperforms all the methods at lower SNR, given that exact statistical properties of the transmitted signal and noise are known. However, they are not known nor stationary for a person speaking.

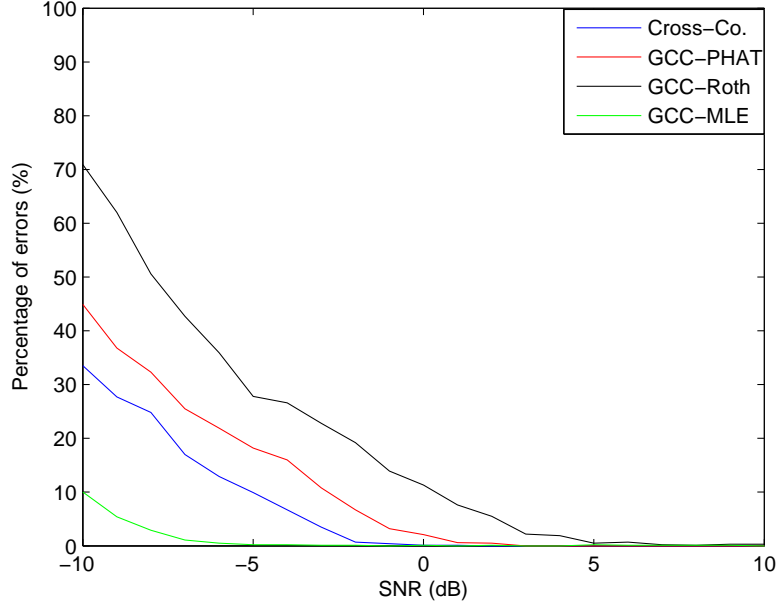


Figure 3.4: Percentage of error vs. SNR: Two different acoustic signals were generated from a 8 kHz sampled wave file using the Image-Source model. The two sensor nodes were placed 2 m apart (47 sample delay) and reverberation was fixed at 0.1 s. Noise was added to achieve average SNR between the two signals from -10 to 10 dB. For each SNR value, time delay was estimated using the methods mentioned in Table 3.1. 1000 independent trials were performed for each method and percentage of error was recorded

3.3 Beamformer Based Localization

The LLS estimator derived for TDOA localization in section 2.4 produces closed-form location estimates. Based on formulation, it can be clearly categorized as a “two-step” localization method. In two-step methods, a set of parameters between the target and the anchor nodes (TDOAs between anchor node pairs in this case) are estimated first and then location of the target is estimated using the estimated parameters. Multi-lateration, LLS, constrained LS, MLE and Best Linear Unbiased Estimator (BLUE) estimators can be considered as well known examples [28],[29],[46],[49]. Since computational cost is constant for a given number of anchor nodes, these methods perform relatively faster. However, for TDOA case, these methods does not effectively use all the available measurements.

Since TDOA is measured between two anchor nodes, a different node pair is needed

for each independent measurement. For an array of N number of anchor nodes, there are ${}^N C_2$ (i.e. $N(N - 1)/2$) number of different sensor pairs. Compare the illustrations given in Figure 3.5. When linearised to get a close-form solution, only TDOAs from $N - 1$ pairs can be used [29]. Information available in $(N - 1)(N - 2)/2$ number of node pairs has to be neglected. When number of anchor nodes are greater than 5, number of discarded TDOAs (6 TDOA measurements) is always higher than the number of TDOAs actually used (4 measurements). Undesirable effects like reverberation make the TDOA measurements unreliable. Thus, more and more measurements are needed. Since these methods do not use all the available measurements, they can be considered inefficient.

In contrast, beamformer based localization methods do not estimate any parameter explicitly. Instead it combines all the measurements to form a “spacial likelihood function” [91]. Points that maximize the spatial likelihood function represent probable candidates for the target location. This approach does not give a closed-form solution nor does it try to linearise any non-linear Equations. Thus, all available measurements from all the anchor nodes can be used in the formulation. Since higher number of measurement increases estimation accuracy, beamformer based methods demonstrate better performance compared to two-step methods.

3.3.1 Steered Response Power with Phase Transformation

Delay-and-sum beamformer is a well-known beamformer used to change directivity of a microphone array [52]. For an array of N number of elements, it is defined as,

$$y(t|\theta) = \frac{1}{N} \sum_{n=1}^N x_n(t - \tau_n(\theta)) \quad (3.25)$$

Beamformer output, $y(t|\theta)$, is summation of individually delayed input signals, $x_n(t)$. $\tau_n(\theta)$, the delay parameter is selected by steering focal point of the beamformer on a point θ , which is in a predefined area. This approach is used to combine measurements in Steered Response Power with Phase Transformation (SRP-PHAT) method [54]. As

the inputs, PHAT weighted cross-correlator output of recorded signals at a node pair is used. For N_p number of node pairs, SRP-PHAT is defined as,

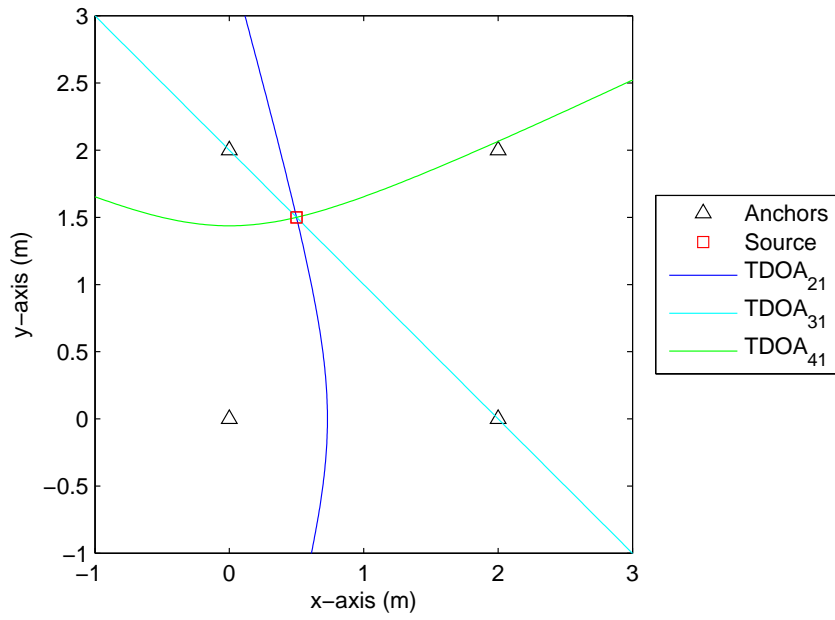
$$SP(\theta) = \frac{1}{N_p} \sum_{n=1}^{N_p} GP(\tau_n(\theta)) \quad (3.26)$$

where symbols SP and GP represent SRP-PHAT and GCC-PHAT respectively. Then location of the target is estimated by searching for point θ that maximizes the SRP-PHAT output.

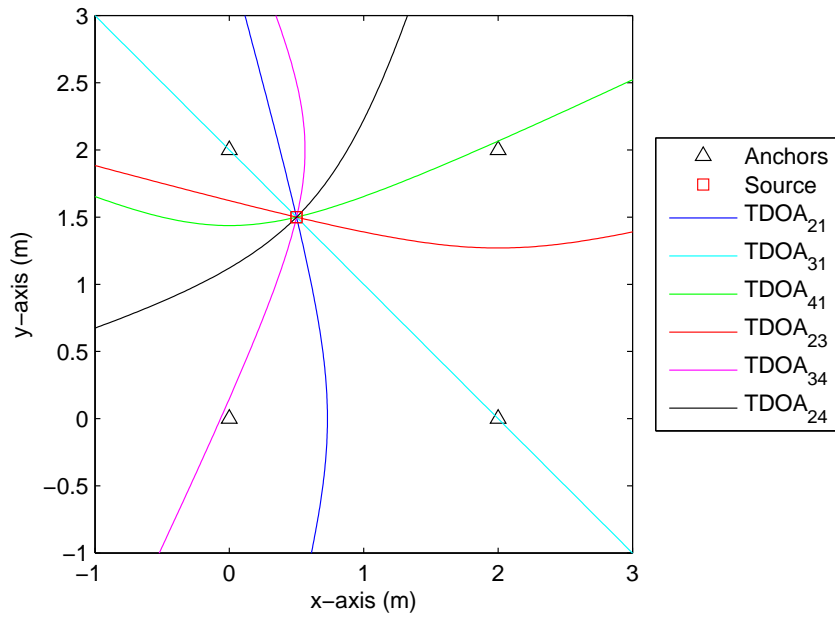
$$\hat{\theta} = \arg \max_{\theta} \{SP(\theta)\} \quad (3.27)$$

A spatial likelihood function generated using SRP-PHAT is illustrated in Figure 3.6. Four anchor nodes and a speaker were placed on (0,0), (2,0), (2,2), (0,2) and (0.5,1.5) points respectively. Acoustic signals received at each anchor node were generated from an 8 kHz sampled wave file (stores voice of a man saying “Hello”) using Image-Source method (reverberation time was fixed to 0.1 s). Uncorrelated additive Gaussian noise were then added to achieve 3 dB average SNR. All the possible node pairs (6 pairs) were selected and GCC-PHAT was performed between each pair. Then GCC-PHAT values were combined using SRP-PHAT beamformer. The spatial likelihood function was obtained steering the beamformer on a 4×4 m point grid (for θ) around the anchor nodes with 1 cm grid resolution.

The spatial likelihood function “visualizes” likelihood of having the sound source at a particular point. The higher the beamformer output, the more likely the sound source at that particular point. As illustrated in Figure 3.6, as the illustration colour gets closer to red, likelihood of having the sound source increases. As it gets closer to blue, likelihood reduces. Global maximum of the likelihood function is clearly visible in part (a). Estimated location of the target is respective x and y coordinates of the maxima. In part (b), traces of six hyperbolas are visible. They are respective hyperbolas of TDOAs for each anchor pair. Stronger the trace, more reliable the TDOA is.

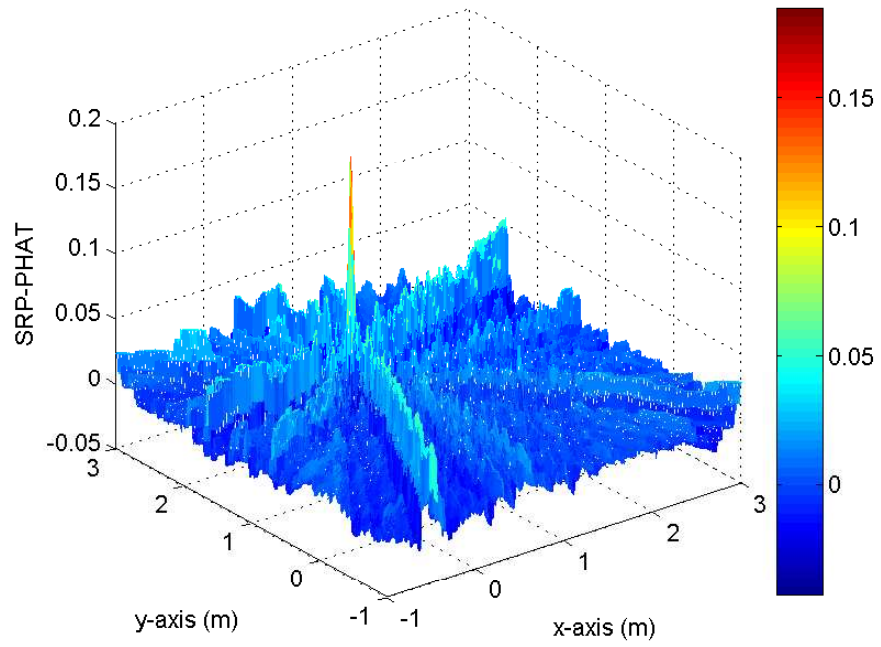


(a) With $N-1$ pairs

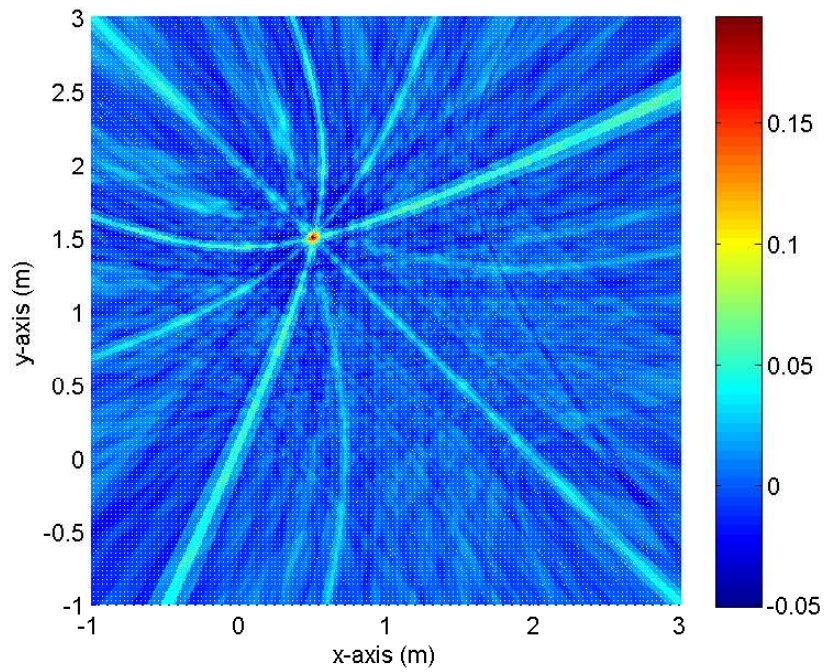


(b) With $N(N-1)/2$ pairs

Figure 3.5: Comparison between $N-1$ and $N(N-1)/2$ pairs



(a) 3D view



(b) 2D view

Figure 3.6: Spatial likelihood function for SRP-PHAT:

This method has been widely used and become the standard beamformer in sound source localization using microphone arrays [55],[56],[92],[93],[59],[94],[95]. Since the beamformer does not provide a close-form solution for the target location, efficiency of the localization is restricted by the search algorithm. The likelihood function is not “smooth”. Thus, traditional optimization methods like Newton’s method cannot be used [96]. In both [92] and [93], a brute force approach has been used. SRP-PHAT is steered on all the possible points in the search space and complete spatial likelihood function is obtained. Then each point is compared to find the maxima. Though this approach provide accurate results, it is computationally exhaustive. Computational cost increases exponentially with size of the area being monitored. In [56] a particle filter based search optimization is proposed. In [59] and [94] a multi-resolution search algorithm is proposed. In both cases, the search space is truncated iteratively until it converges on the maxima. Accuracy of the convergence is restricted by initial search point distribution. In [55], a sector based approach is proposed with a circular microphone array. First, the area being monitored is partitioned into different sectors and “activeness” in each sector is compared. Then the beamformer is steered only on the most “active” sectors. In [95], TDOAs for each node pair is estimated and then inversely mapped to a set of spatial coordinates. Beamformer is then steered only on neighbourhood of those points.

3.4 Chapter Summary

A person speaking can be considered as a wideband signal source. Thus, he/she can be localized using a TDOA based method. This relaxes restrictions on the target to wear sensor nodes. Considering effect of reflection, acoustic wave propagation can be modelled as direct path propagation or multipath propagation. Since reflective surfaces are common, the later model is more suitable for indoors. Received signal at a microphone

can be modelled as a convolution between the transmitted signal and transfer function of the propagation channel. Image-Source method is a widely used model to get the transfer function.

For wideband signals, cross-correlation based methods are used to estimate time delay. Cross-correlation between filtered or pre-whitened signals is more effective than it is between raw signals. PHAT and Roth filters result more distinguishable impulse-like outputs. MLE based filters perform better when effect of lower SNR is dominant. However, statistical properties of transmitted signal (speech) are needed for the MLE formulation. Both PHAT and Roth filters need only power spectra of received signals.

Close-form methods like LLS and BLUE operate in two steps. First TDOAs are estimated from the measurements and then target is localized using the estimated TDOAs. A linearisation process is needed to get the close-form estimate and TDOAs from all the pairs cannot be used in the formulation. Thus, they are more vulnerable to undesirable effects like low SNR and reverberation. In contrast beamformer based single-step method combines measurements from anchor nodes instead explicitly finding TDOAs. Since measurements from all the pairs can be used, they are less vulnerable to undesirable effects. SRP-PHAT is a widely used beamforming technique in microphone arrays to localize sound sources. It performs more accurately than close-form methods. A likelihood distribution is created for each possible point in search space and point with the highest likelihood is taken as the target location. Since computational cost and efficiency is restricted by number of search points, an optimization method need to be utilized.

Chapter 4

Effect of Reverberation and Packet Loss

Performance of time delay estimation of speech using wireless sensor nodes is restricted by various reasons. Some of them are issues involved with propagating acoustic signals and others are issues involved with wireless sensor nodes. In this chapter, two major issues are identified and discussed.

Decay of signal strength and multipath are two major issues involved with propagating signals. When signal strength is smaller, voice signature in the recorded signal is undetectable. When multipath effects are dominant, voice signature in the recorded signal is unrecognisable. Our consideration is localizing speakers in indoors. Since we can place anchor nodes at relatively short distances, assuming all the nodes receive the signal with sufficient strength, effect of signal strength decay can be relaxed. However, as size of the room gets smaller, a signal can be reflected multiple times by each surface. Thus, tendency to have multipath becomes higher. At the anchor node, they arrive as echoes and identifying the original is harder. Thus, we identified multipath as the major issue involved with propagating signals.

In wireless sensor nodes, there are effects of measurement noise, number of anchor nodes, sampling rate, communication interference, congestion, packet collision, etc.. However, effects of measurement noise and number of sensor nodes can be considered trivial. Lower the measurement noise or higher the number of sensor nodes, higher the SNR is. Higher the SNR, better the performance is. Effect of sampling rate is also well studied in literature and we make sure sampling rate in our implementation also satisfies Nyquist-Shannon criterion [97]. However, effect of communication interference, congestion and packet collision are inherent in wireless networks. The ultimate result of

all of them is losing packets. Thus, we selected packet loss as the major issue involved with wireless sensor nodes.

4.1 Reverberation

When an acoustic signal is reflected by a smooth surface at a considerable distance away from the source, the original and the reflected signals arrive at the listener separately. Since human ear can distinguish the two signals, it is called an echo. When there are multiple surfaces at close distances, like inside a room, each wall reflects the original signal as well as the reflected signals from other walls. Since all the multipath signals are closely aligned, the human ear cannot distinguish each signal separately. This effect is called reverberation. Reverberation time, T_{60} , is defined as time it takes to decay strength of reflected signals by 60 dB compared to the directly arrived signal. For a given environment, say inside a room, T_{60} can be estimated using Sabine's Equation [98]. It is defined as,

$$T_{60} = \frac{4 \ln 10^6}{C} \times \frac{V}{\Sigma A_i \nu_i} \quad (4.1)$$

where V is volume of the room (in m^3), C is speed of sound (in m/s) and Σ is summation operator. A_i and ν_i are surface area and sound absorption coefficient of i^{th} surface respectively. Speed of sound in dry air at 20°C is 343 m/s [74]. Thus, above Equation can be approximated as,

$$T_{60} \approx 0.1611 \frac{V}{\Sigma_i A_i \nu_i} \quad (4.2)$$

Absorption and reflection of acoustic wave by a reflective surface are interdependent [98]. They are related as,

$$\nu = 1 - \beta^2 \quad (4.3)$$

β is reflection coefficient of the reflective surface. Thus, one can select appropriate surface material with desired reflection coefficients to have desired reverberation. To calculate

suitable reflection coefficient to get desired T_{60} , Eyring’s Equation can be used [98]. Eyring’s Equation is defined as,

$$\beta = e^{\frac{-13.82}{cT_{60}} \frac{1}{(\frac{1}{L} + \frac{1}{W} + \frac{1}{H})}} \quad (4.4)$$

where L , W and H are length, width and height of the room respectively.

To visualize effect of reverberation, we are investigating transfer function of the propagating channel. A transfer function obtained using the Image-Source method for two different reverberation conditions are illustrated in Figure 4.1. A low reverberation ($T_{60} = 0.1$ s) case and a high reverberation ($T_{60} = 0.3$ s) case are illustrated in part (a) and (b) respectively. In part (a), there are two easily distinguishable impulses with 47 samples separation. Since most of the acoustic energy is absorbed by the reflecting surfaces, reflective energy decays faster. Thus, the transfer function lasts approximately for a period of time less than 650 samples. However in part (b), for $T_{60} = 0.3$ s, it takes three times longer period of time to achieve the same signal decay. Thus, the transfer function lasts for around 2000 sample period of time. Apart from the two main impulses with 47 sample delay, there are other distinguishable impulses representing reflections. Since the GCC algorithm “confuses” with them as different speakers, performance of the time delay estimation is affected.

4.1.1 Effect of Reverberation on Time Delay Estimation

The transfer function for propagation channel can be decomposed into two parts. i.e. direct propagation and reverberation [81]. Thus we can write,

$$\begin{aligned} h_i(t) &= h_i^d(t) + h_i^r(t) \\ h_1^d(t) &= \delta(t) \\ h_2^d(t) &= \alpha \cdot \delta(t + \tau_0) \end{aligned} \quad (4.5)$$

$h_i^d(t)$ and $h_i^r(t)$ are direct and reverberant (or multipath) propagation for i^{th} anchor node respectively. When microphones of sensor nodes are placed at least half wavelengths

away from reflecting walls (i.e. first reflections are not strong) reverberant component can be considered diffusive [98]. For a noise-free channel, when anchor nodes are placed away from walls (reverberation is diffusive), variance of estimated delay around the true delay can be estimated by extending work in [81] as,

$$var(\hat{\tau}_0) = \frac{\beta^2}{A(1 - \beta^2)\pi^2\alpha^2} \frac{\int_{-\infty}^{\infty} \Psi(f)^2 G_{ss}(f)^2 f^2 df}{[\int_{-\infty}^{\infty} \Psi(f) G_{ss}(f) f^2 df]^2} \quad (4.6)$$

where A is total surface area of the room. Here measurement noise is removed to isolate effects from reverberation. Since we are working with samples in discrete time domain, this can be written in discrete frequency domain with N frequency bins as,

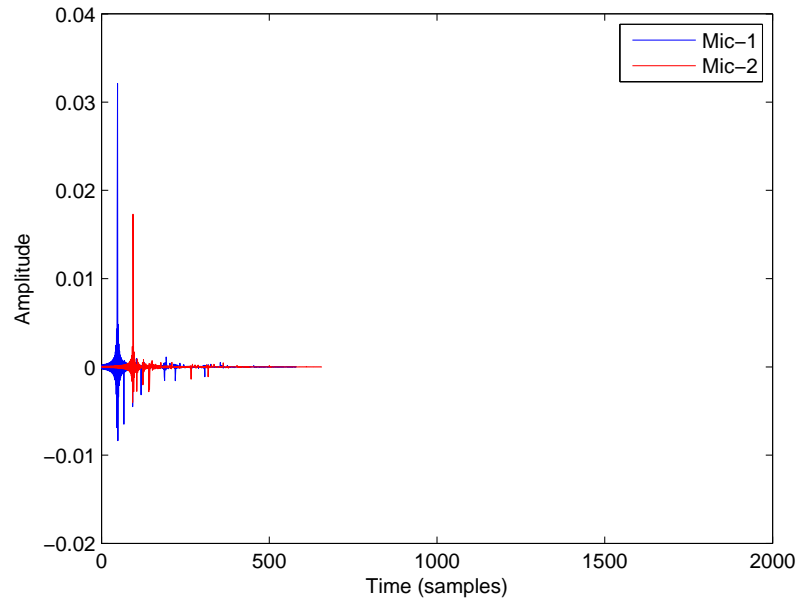
$$var(\hat{\tau}_0) = \frac{\beta^2}{A(1 - \beta^2)\pi^2\alpha^2} \frac{\sum_{n=0}^{N-1} \Psi(f_n)^2 G_{ss}(f_n)^2 f_n^2}{[\sum_{n=0}^{N-1} \Psi(f_n) G_{ss}(f_n) f_n^2]^2} \quad (4.7)$$

f_n is n^{th} frequency bin for Discrete Fourier Transformation (DFT). There is no randomness in the given expression. As long as positions of the speaker, placement of anchor nodes and reflective surfaces are fixed, variance due to reverberation is time invariant. For variance to be minimum, $\Psi(f_n)$ in the previous Equation should satisfy following condition [81].

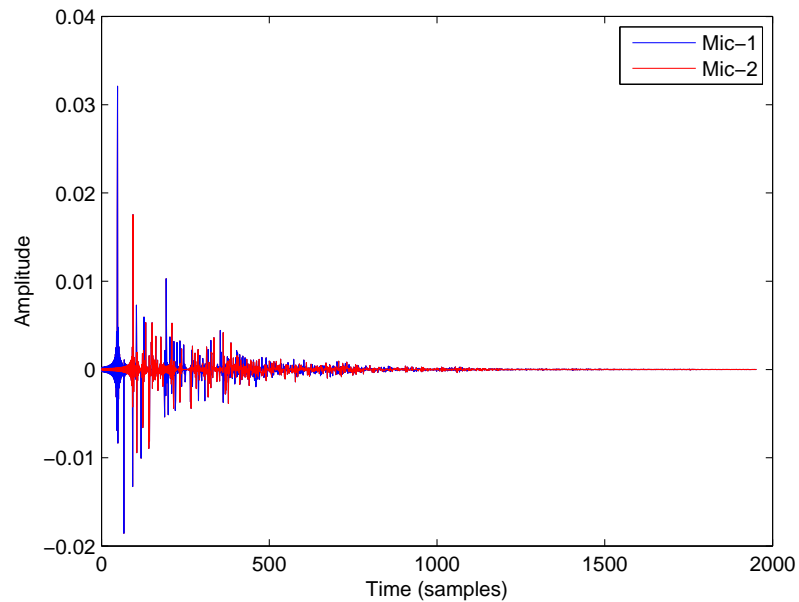
$$\Psi(f_n) = \frac{1}{|G_{ss}(f_n)|} \quad (4.8)$$

When noise power spectra is negligible, $\hat{G}_{ss}(f_n)$ and $G_{x_1x_2}(f_n)$ are the same. That is PHAT weighting function for GCC. Since it minimizes variance, as long as reverberation is the concern, PHAT is the optimum weighting function for GCC formulation.

Percentage error of time delay estimation w.r.t. reverberation is illustrated in Figure 4.2. Four cases of reverberation are simulated and for each reverberation time, percentage error is obtained by changing average SNR (between the node pair). A case of zero reverberation (i.e. direct propagation only) is presented in part (a) for comparison. In part (b), relatively low reverberant case with $T_{60} = 0.1$ s is given. In part (c), T_{60} is increased to 0.3 s to simulate a high reverberant case. Part (d) with $T_{60} = 1.0$ s represents a severe reverberant situation.

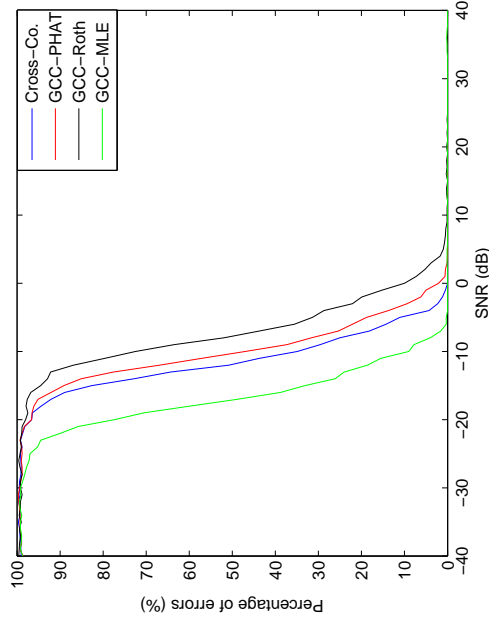


(a) For $T_{60} = 0.1$ s

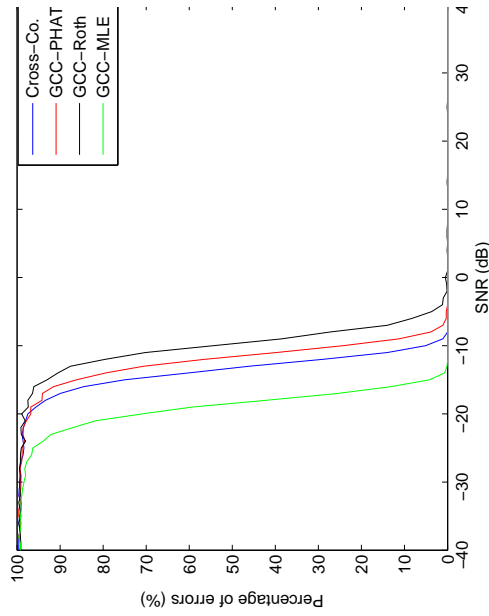


(b) For $T_{60} = 0.3$ s

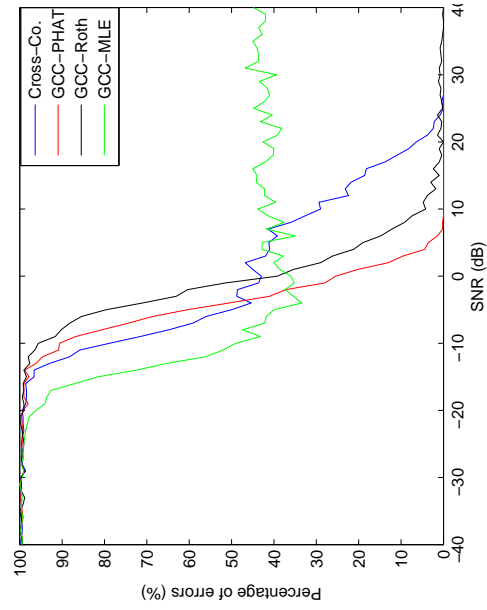
Figure 4.1: Comparison of room impulse response for different reverberation conditions: Two room transfer functions were generated using Image-Source method for reverberation time 0.1 s and 0.3 s. In both the cases distance difference from sound source to two sensor nodes was fixed to 2 m (47 sample delay)



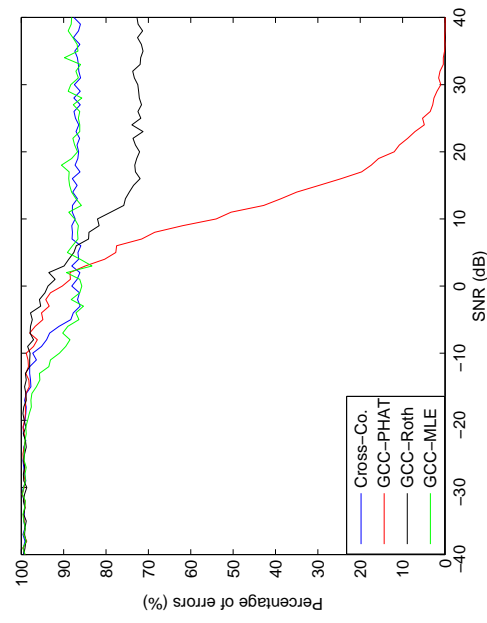
(a) No Reverberation



(b) For $T_{60} = 0.1$ s



(c) For $T_{60} = 0.3$ s



(d) For $T_{60} = 1.0$ s

Figure 4.2: Percentage error vs. reverberation vs. SNR: Two signals were generated using Image-Source method by fixing time delay at 47 samples and varying T_{60} . 500 trials were performed using each GCC method with varying SNR and percentage errors were recorded.

All GCC methods result lower percentage error as SNR increases. However, for a given SNR, percentage error increases when reverberation time increases. For lower SNR values, GCC-MLE always demonstrates lower errors. In other words, MLE is the optimum filter function for GCC when effect of noise is dominant. In part (c), percentage errors for a higher reverberant case is given. When SNR improves, effect of reverberation becomes dominant. As mentioned earlier, variance caused by reverberation is constant for stationary reverberation condition (constant T_{60}). Thus, GCC-MLE demonstrates nearly uniform performance after -5 dB SNR. For SNR values higher than -2 dB, GCC-PHAT performs better than all the other estimation methods. This observation confirms suitability of PHAT weighting for GCC when reverberation is the main concern. In part (d) an environment completely dominated by reverberation ($T_{60} = 1.0$ s) is given. The MLE filter is derived assuming noise is zero mean Gaussian. Since high reverberation violates that assumption, reliability of GCC-MLE estimator completely disappears after -5 dB SNR. Both GCC-MLE and cross-correlator result around 90% errors regardless high SNR. GCC-Roth outperforms GCC-MLE at around 7 dB SNR and after 15 dB, percentage error converges at 70%. It is clearly evident that, GCC-PHAT is the best option to estimate time delay in higher reverberant environments.

4.2 Packet Loss

To perform time delay estimation using GCC, two streams of recorded data is needed. Samples should be taken at least twice the maximum frequency of the signal to accurately reconstruct at the destination [97]. Our interest is speech, which has most of the recognizable frequencies between 340-3400Hz [75]. Since computational power of sensor nodes is not sufficient to perform all the computations, samples are streamed to a computationally powerful base station, which is a computer. Most of the wireless nodes are Zigbee compliant and both Zigbee and WiFi shares the same radio spectrum [71]. When

Packet loss in a communication channel has been modelled as a two state Markov chain in [1],[2]. We can also model packet loss in incoming acoustic data using a Markov Chain.

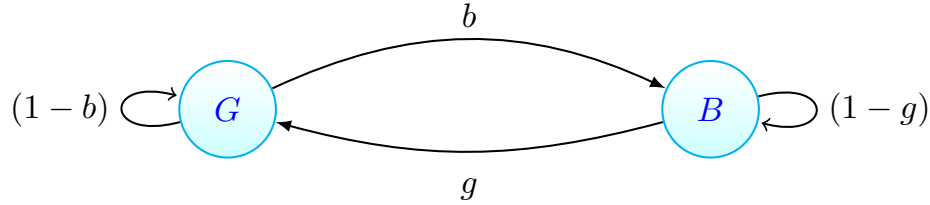


Figure 4.3: Markov Model for Packet Loss. The blue circles represent the two states and arrows represent the possible transition between the states. “G” is the state of receiving a packet and “B” is losing a packet. Probability of receiving the next packet given that the current one is missing, $P(G|B) = g$ while probability of losing the next packet given that the current one is successfully received, $P(B|G) = b$.

there are other coexisting high power transmissions in the same radio spectrum, interference introduced by the high power transmitter impedes the low power transmission. This makes the low power transmission to have packet loss. On the other hand, when the receiver moves towards the outer edge of the coverage area, the low RF signal reception makes the communication unreliable. Thus, losing packets is unavoidable. When a packet loss is detected by comparing the packet sequence number, the standard approach is to request a packet retransmission from the sender. Retransmission adds more communication overheads, delay and decreases throughput. Since low power wireless nodes have relatively slow data rate, lesser overhead is always encouraged. Sacrificing throughput might not be an option for a real-time, fast-response system.

Packet loss in a communication channel has been modelled as a two state Markov chain [99],[100]. As given in Figure 4.3, we can also model packet loss in incoming acoustic data using a Markov Chain. Transition probability matrix of the Markov chain is given by,

$$T = \begin{bmatrix} (1-b) & b \\ g & (1-g) \end{bmatrix} \tag{4.9}$$

Probability of losing the next packet ($P(B)$) and successfully receiving it ($P(G)$)

converge to constants as the number of received packet increases. This is called as stationary distribution. State probabilities at the stationary distribution has two properties [101]. They are,

$$\begin{bmatrix} P(G) \\ P(B) \end{bmatrix} = \begin{bmatrix} (1-b) & b \\ g & (1-g) \end{bmatrix} \begin{bmatrix} P(G) \\ P(B) \end{bmatrix} \quad (4.10)$$

and

$$P(G) + P(B) = 1 \quad (4.11)$$

Using the two properties given above, state probabilities at the stationary distribution can be calculated as,

$$\begin{aligned} P(G) &= \frac{g}{g+b} \\ P(B) &= \frac{b}{g+b} \end{aligned} \quad (4.12)$$

The two state Markov chain does not necessarily follow Bernoulli distribution unless each row of the transition probability matrix has identical rows [101]. In other words, unless $g = 1 - b$.

4.2.1 Effect of Packet Loss on Time Delay Estimation

To make an accurate time delay estimation, samples in the two signals should be accurately aligned in time domain. In other words, i^{th} sample of the recorded signal at the first anchor node and i^{th} sample of the recorded signal at the second anchor node should be two samples taken at the same time at respective nodes. When both the nodes have the same sampling rate and if all the buffers involved in communication are First In First Out (FIFO), sampled signals at the anchor nodes can be reconstructed at the base station as illustrated in Figure 4.4. When there are no packet loss, there are no discontinuities in received signals. If both the nodes start and stop sampling at the

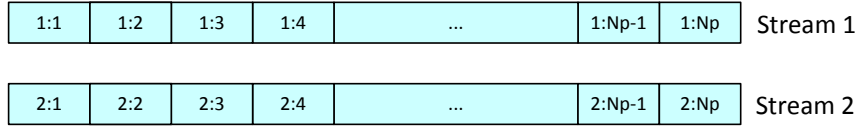


Figure 4.4: Received packets at the base station: Stream-1 is samples taken from the first anchor node as a stream of packets and stream-2 is from the other anchor node. 1 : 2 represents 2nd packet from 1st anchor node. Each observation window is N_p packets long

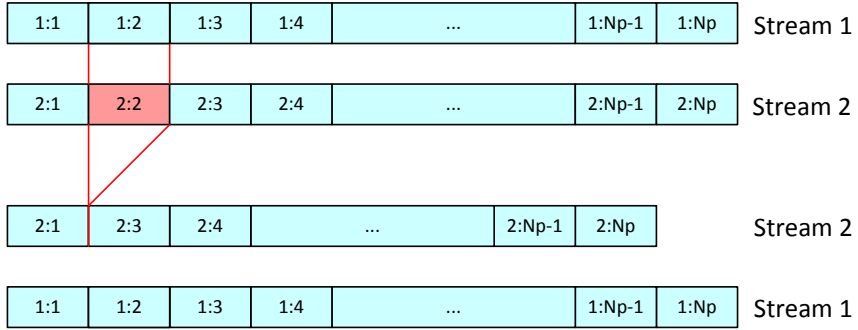


Figure 4.5: Received packets at the base station: First two rows represent transmitted packets and last two rows represent received packets. In the second row, red block represents a missing packet

same time, both the reconstructed signals are time domain aligned. Hence, performance of time delay estimation is restricted mainly by issues involved with acoustic signals.

However, it is not the case when there is packet loss. An example of two received signal with a missing packet is illustrated in Figure 4.5. 2nd packet transmitted from the 2nd anchor node is missing. Thus, rest of the packets in the reconstructed signal are shifted toward left side of the observation window. Samples in the two reconstructed signals are not time domain aligned. When time delay is estimated between the two reconstructed signals, there is always an error. This is independent of content of packets, thus independent of acoustic signals. A packet loss can happen in anywhere in the incoming stream. Thus, each packet loss make the time delay estimation unreliable.

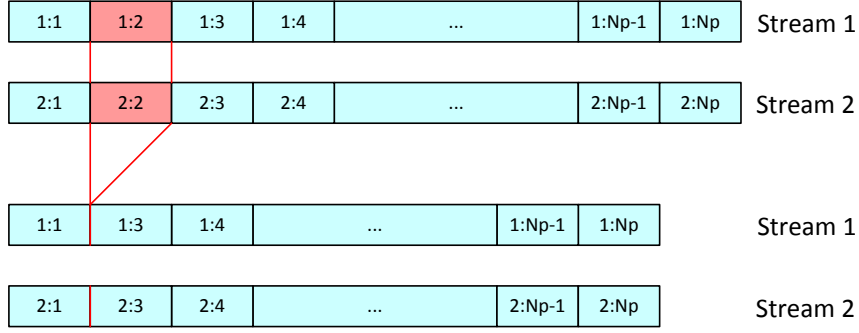


Figure 4.6: Received packets at the base station: First two rows represent transmitted packets and last two rows represent received packets. In the first two rows, red blocks represent a missing packets

In the situation given in Figure 4.6, both the received signals have missing packets. 2nd packet of both the streams are missing. Number of packets in the observation window has reduced by one. However, interestingly, all the samples are still time domain aligned and there is no artificial delay introduced between the two reconstructed signals. Performance of time delay estimation is not affected by packet loss and it is again restricted only by content of the packets. All the possible missing packet combinations for a stream of 5 packets are given in Figure 4.7. With usual symbols, probability of losing the first packet in a stream is $P(G)^{N_p-1}P(B)^1$. Since both the anchor nodes are independent, probability of getting both the streams with the same packet missing is $[P(G)^{N_p-1}P(B)^1]^2$. There are ${}^{N_p}C_1$ number of ways both the streams miss the same packet. Thus, probability of getting two reconstructed signals time domain aligned even with two missing packets is ${}^{N_p}C_1[P(G)^{N_p-1}P(B)^1]^2$. Table 4.1 tabulates probabilities of getting two reconstructed signals aligned with multiple missing packets. Higher the number of missing packets, lesser the number of possible combinations. Total probability of getting all the samples aligned is summation of the third column.

$$P_t = \sum_{k=1}^i {}^{N_p}C_k [P(G)^{N_p-k}P(B)^k]^2 \quad (4.13)$$

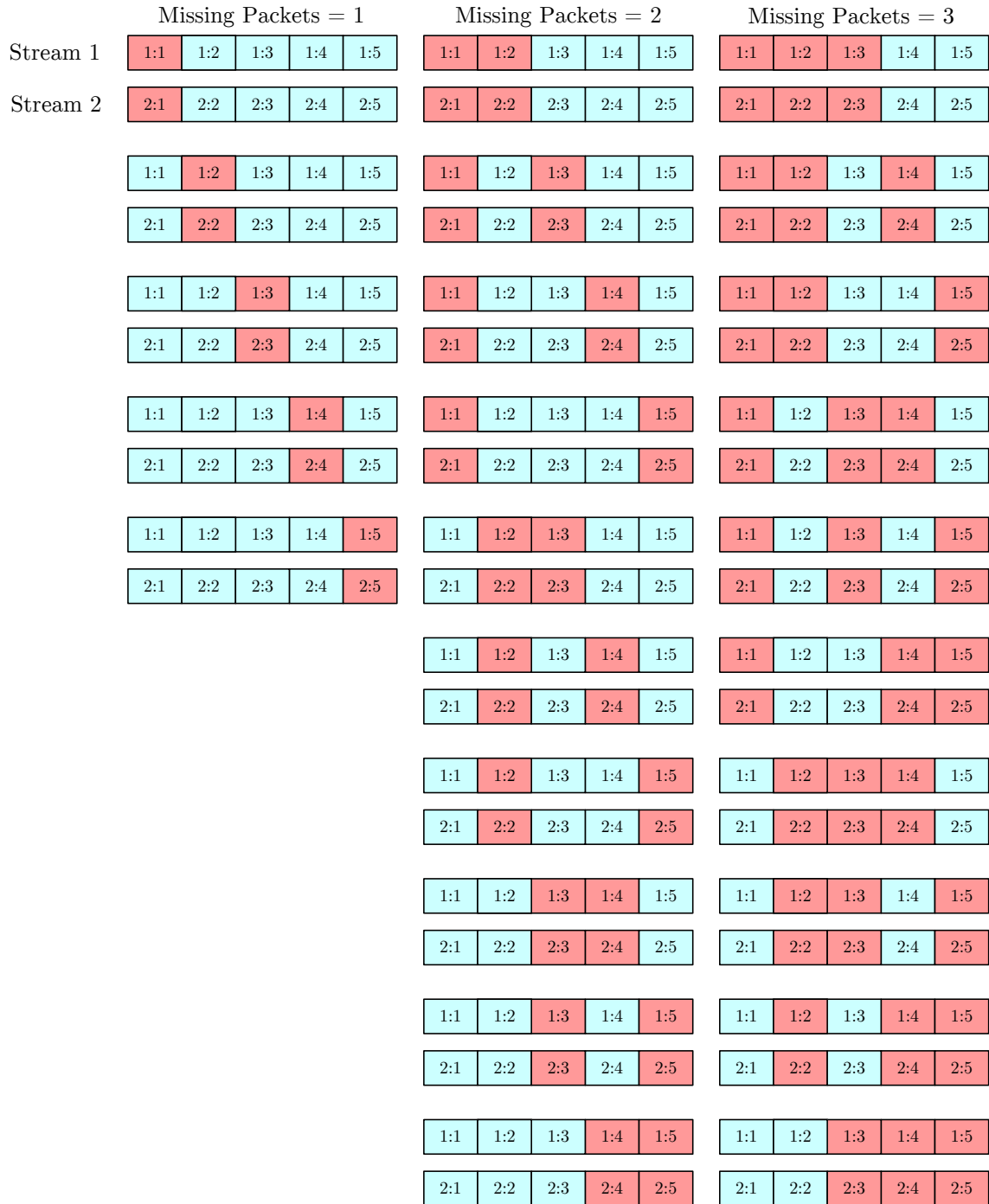


Figure 4.7: Packet loss in received signals

No. of Lost Packets	No. of Combinations	Probability of Getting that
1	${}^{N_p}C_1$	${}^{N_p}C_1 \times [P(G)^{N_p-1} \times P(B)^1]^2$
2	${}^{N_p}C_2$	${}^{N_p}C_2 \times [P(G)^{N_p-2} \times P(B)^2]^2$
3	${}^{N_p}C_3$	${}^{N_p}C_3 \times [P(G)^{N_p-3} \times P(B)^3]^2$
\vdots	\vdots	\vdots
i	${}^{N_p}C_i$	${}^{N_p}C_i \times [P(G)^{N_p-i} \times P(B)^i]^2$

Table 4.1: Probability of getting two streams aligned: In the first column possible number of missing packets is given. In the second column number of combinations where both the streams loose packets simultaneously is given. And in the last column probability of getting simultaneous packet lost in both the streams for given number of lost packets is given.

However, as packet loss increases, amount of data available to perform the cross correlation reduces. Due to lack of information after certain number of missing packets, performance of time delay estimation deteriorates. The threshold number of packet loss is given as i in Table 4.1. From our observations, we found this depends on sampling rate, number of samples in each packet, minimum overlap between two signals and maximum delay possible between the two sensor nodes. Maximum possible delay is direct time of flight between the farthest pair of anchor nodes. If d_{max} is the maximum distance between nodes in the pair, maximum delay possible is,

$$\tau_{max} = \frac{d_{max}}{C} \quad (4.14)$$

If f_s is sampling rate, number of samples needed to represent τ_{max} is,

$$N_\tau = \frac{\tau_{max}}{f_s} = \frac{d_{max}}{C f_s} \quad (4.15)$$

In most of time delay estimation algorithms, delay is assumed to be small compared to observation window [86]. To estimate time delay reliably, both the signals should have some overlap. Consider illustration given in Figure 4.8. ξ is the percentage of overlap between the two reconstructed signals. When $\xi = 0$, there is no overlap between two signals, thus it is impossible to find delay. When $\xi = 1$, both the streams overlap

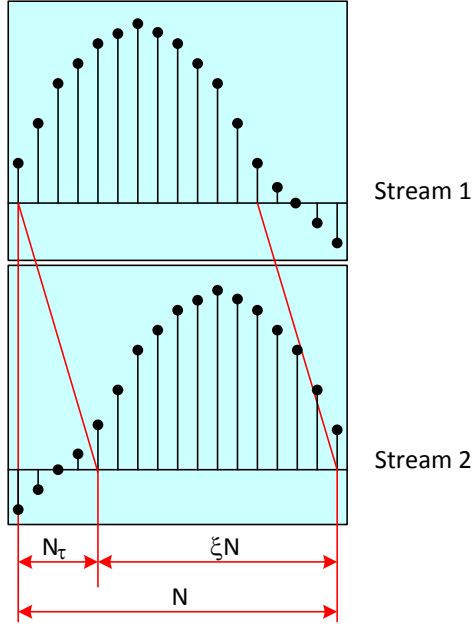


Figure 4.8: Overlap between two signals. The two boxes represent two aligned data blocks and black vertical line in each box represents samples.

completely, but there is no delay. Thus, number of samples needed to represent τ_{max} with a ξ percentage overlap is,

$$N_o = \frac{N_\tau}{1 - \xi} = \frac{d_{max}}{C f_s (1 - \xi)} \quad (4.16)$$

If each packet contains N_s number of samples, number of adjacent packet needed to represent τ_{max} is,

$$N_{Aj} = \frac{N_o}{N_s} = \frac{d_{max}}{C f_s (1 - \xi) N_s} \quad (4.17)$$

Equation 4.17 gives minimum number of adjacent packets needed to make a reliable time delay estimation. Due to lack of information, it is impossible to make a time delay estimation without having N_{Aj} number of packets for each received signals. Existing number of packets should always be more than N_{Aj} .

$$N_p - N_{loss} > N_{Aj} \quad (4.18)$$

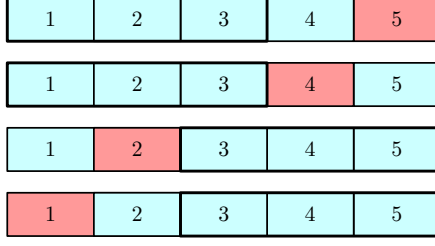


Figure 4.9: For $N_p = 5$, $N_{loss} = 1$, $N_{Aj} = 3$

As the number of missing packets increases, the probability of having the desired number of adjacent packets decreases. Consider the following example which has 5 packets. Suppose there is a packet missing from both the sensor nodes. As explained earlier, there are ${}^5C_1 = 5$ ways for both the reconstructed signals to remain aligned. All the five ways satisfy the minimum number of adjacent packets criterion if it is one. It is always guaranteed to have two adjacent packets. However, when minimum number of adjacent packets needed is three, there are only four situations that it satisfy. This is illustrated in Figure 4.9.

When the desired number of adjacent packets is unavailable, time delay estimation is unreliable. Maximum number of allowable missing packets without disturbing the desired number of adjacent packets is given in Table 4.2. The same Table can be obtained using the following relationship.

$$N_{loss} = RoundDown \left\{ \frac{N_p - N_{Aj}}{N_{Aj}} \right\} \quad (4.19)$$

Finally, probability of getting all the data points aligned without any effect from packet loss can be found by combining Equations 4.13, 4.17 and 4.19. The notation i in the Equation 4.13 is given as N_{loss} in Equation 4.19 and N_{Aj} of 4.19 is given in 4.17.

As discussed earlier, it is obvious that packet loss deteriorates time delay estimation. Missing packets introduce artificial delays between the two signals and as the number of lost packets increases correlation between the two signals become more and more ambiguous. The percentage error increases rapidly as probability of packet loss increases.

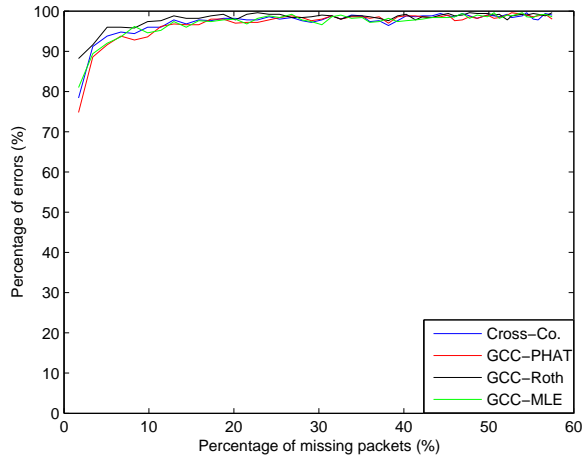
	$N_p = 4$	$N_p = 5$	$N_p = 6$	$N_p = 7$	$N_p = 8$...
$N_{Aj} = 1$	3	4	5	6	7	...
$N_{Aj} = 2$	1	1	2	2	3	...
$N_{Aj} = 3$	0	0	1	1	1	...
$N_{Aj} = 4$	0	0	0	0	1	...
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\ddots

Table 4.2: Maximum number of lost packets allowable to maintain desired N_{Aj} for given lossless observation window size

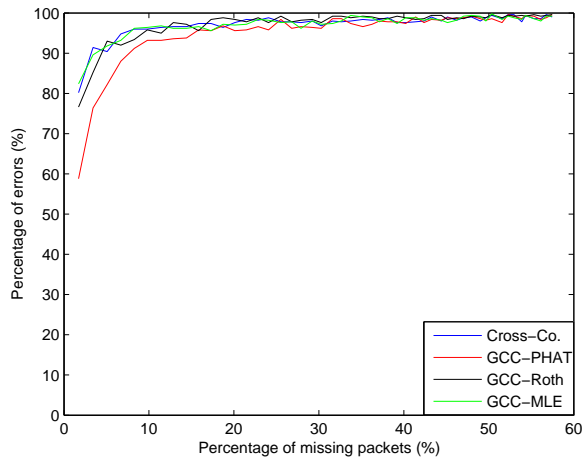
After a certain point, the estimated time delay is completely inaccurate. To analyse the performance of the time delay estimation with packet loss, the following experiment was performed.

Two acoustic signals were generated from a 8 kHz sampled wave file using Image-Source method. The wave file contains a male saying ‘‘Hello’’. In this model, reverberation time and distance between the two nodes were selected as 0.1 s and 2 m respectively. Zero mean uncorrelated white Gaussian noise was added to the signals to achieve desired average SNR. 4000 consecutive samples from both signals were selected and they were decomposed into 50 packets with 80 samples in each packet. Since maximum delay possible between the two sensor nodes is 47 samples, there is 42% sample overlap between any two packets taken simultaneously from the two signals. Then packets were removed from identified locations to form two signals with the same loss rate. When packets were removed, rest of the samples were shifted left to fill the void. Number of missing packets was increased step by step starting from 1 packet. Indexes of to-be-removed packets were selected from two uniform distributions with bounds 1 and 50. Time delay between the two lossy streams were estimated using different GCC methods. For given number of missing packets, 1000 different trials were performed and the average percentage error was recorded. An error is defined when the estimated time delay is not equal to 47 samples.

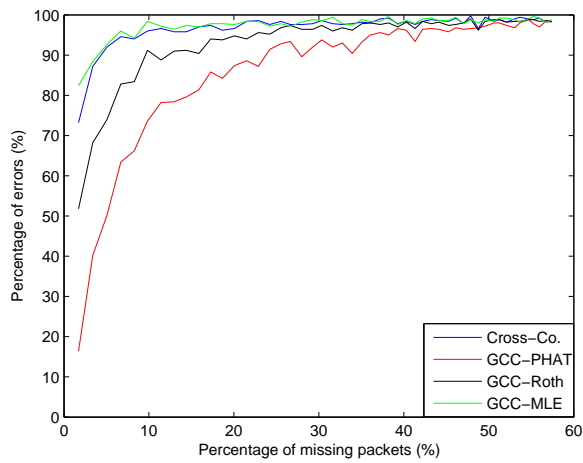
Referring to the results given in part (b) of Figure 4.2, we can say all the GCC methods give less than 10% percentage of errors for zero packet loss for 0 dB SNR and $T60 = 0.1$ s. As given in Figure 4.10, all the time delay estimation methods are severely affected by packet loss. For a single missing packet (2% missing packets), percentage errors for all the GCC methods are above 75%. The percentage of anomalies increases rapidly as the number of missing packets increases. Even for higher SNR like 30 dB, all the methods give more than 50% of errors when packet loss is 5%. It passes 90% errors when packet loss is just 35%. However, when compared to other GCC methods, GCC-PHAT estimator performs robustly. As given in part (c), for 30 dB, GCC-PHAT results only 15% errors for 2% missing packets. For the same percentage of packet loss, GCC-Roth, cross-correlation and GCC-MLE result 50%, 72% and 82% error rate respectively.



(a) For SNR = 0 dB



(b) For SNR = 3 dB



(c) For SNR = 30 dB

Figure 4.10: Percentage of errors vs. percentage of missing packets

4.2.2 Minimizing Effects of Packet Loss

Cross-correlation is used in all the mentioned time delay estimation methods. Losing packets distorts the reconstructed signal. Thus, transmitted signal and received signal become less correlated. When time delay estimation is performed between two distorted signals, it gives anomalous results. Anomalies are there, because samples in both the signals are not time domain aligned. To minimize anomalous results, the sample alignment should be maintained even there is packet loss.

In most of the transmission standards, packet sequence number is inserted into header every time a packet leaves the transmitter. Zigbee standard, which is common in WSN, maintains sequence number in both MAC and network layers [71]. It is usually a 8 bit integer assuming there never be 256 packets from the same node in flight at the same time [71]. This is helpful for the receiver to reconstruct the transmitted data using received packets. Additionally, packet sequence number can be used to detect missing packets. For ordinary lossless data reception, packet sequence number of oncoming packets should always increase by one and only discontinuity occurs when the 8 bit register overflow at 256. Thus, whenever there is a discontinuity in packet sequence number, that must be due to packet loss.

When packet retransmission is disabled, the receiver ignores discontinuities and tries to reconstruct the transmitted data by ordering received packets based on sequence number. This is how the artificial delays are introduced to the signals. Instead, an empty packet with the same length can be inserted when a discontinuity in packet sequence number is detected. When the discontinuity is larger than one, appropriate number of packets (similar to size of the discontinuity) should be inserted. Though some of the data is missing, received samples are always aligned in time domain. By doing this, we can maintain the rest of the oncoming packets less affected by packet loss.

Consider the illustration give in Figure 4.11. All the missing packets are replaced

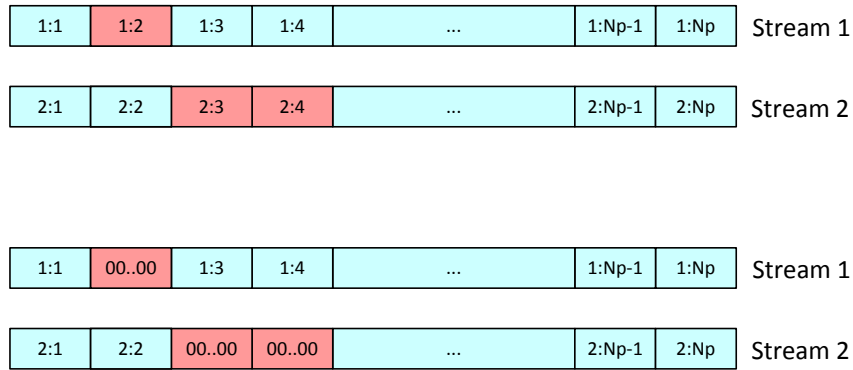
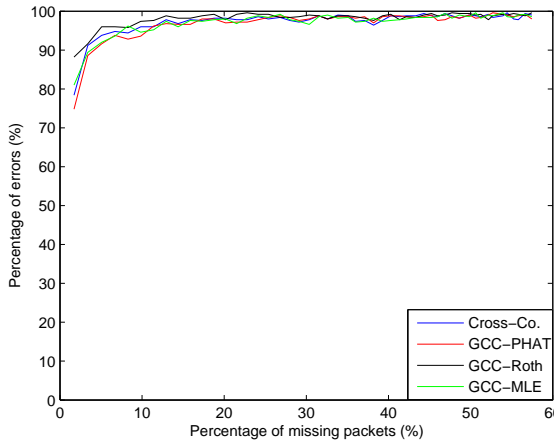


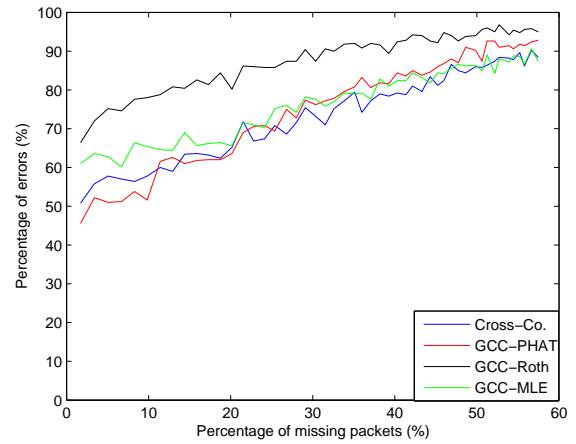
Figure 4.11: Received data from two sensor nodes after inserting empty packets: Here, blue box represents received packets and red boxes on the first two rows represent missing packets. Red boxes on the last two rows represent empty packets

with empty packets. Rest of the samples are completely aligned regardless of missing data. In this case performance of time delay estimation should be better than it is without inserting empty packets. To analyse the effect of this correction, we performed the same experiment. But, all the missing packets were replaced before performing time delay estimation.

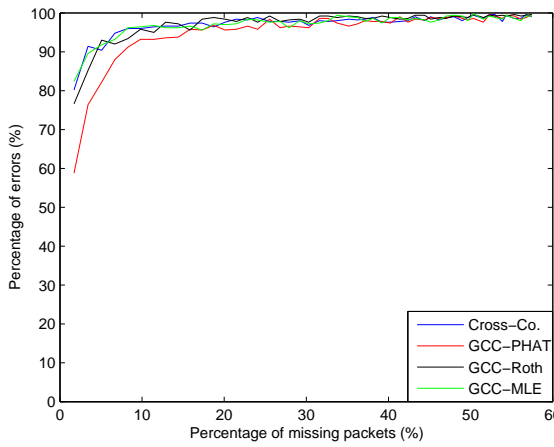
The performance of inserting empty packets in time delay estimation is presented in Figure 4.12. For ease of comparison, result obtained without is also presented. All the GCC methods for all the SNR conditions show performance improvements. The percentage of errors has reduced significantly in all the cases. As given in part (b), for 0 dB SNR, errors for 2% packet loss had reduced by at least 30%. Part (d) demonstrates same improvements for 3 dB. For higher SNR, GCC-PHAT outperforms all the other GCC estimation methods. As shown in part (f), it performs robustly even when percentage of packet loss is 10%.



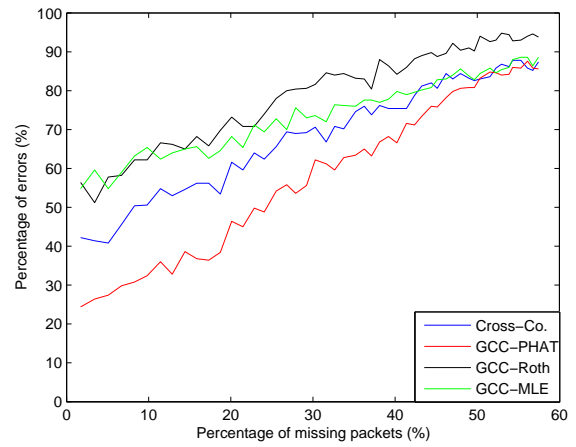
(a) For SNR = 0 dB, before



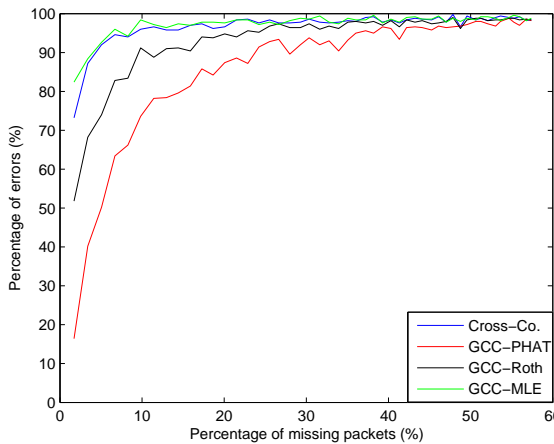
(b) For SNR = 0 dB, after



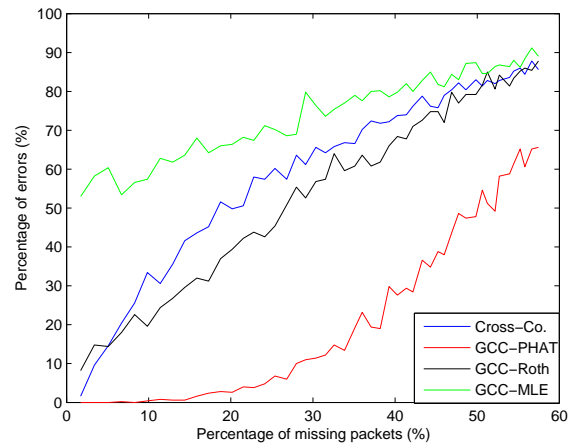
(c) For SNR = 3 dB, before



(d) For SNR = 3 dB, after



(e) For SNR = 30 dB, before



(f) For SNR = 30 dB, after

Figure 4.12: Percentage of errors vs. percentage of packet loss before and after inserting empty packets.

4.3 Variance of GCC-PHAT TDOA Estimate

Various filter functions to improve GCC based time delay estimation are summarized in [86]. In the same study, a general expression for estimated delay variance with general filter function $\Psi(f)$ is given as,

$$var(\hat{\tau}_0) = \frac{\int_{-\infty}^{\infty} |\Psi(f)|^2 (2\pi f)^2 G_{x_1 x_1}(f) G_{x_2 x_2}(f) [1 - |\gamma(f)|^2] df}{T [\int_{-\infty}^{\infty} (2\pi f)^2 |G_{x_1 x_2}(f)| \Psi(f) df]^2} \quad (4.20)$$

where T is observation time and $|\cdot|$ is absolute value operator. $\gamma(f)$ is coherence and it is defined as,

$$\gamma(f) = \frac{G_{x_1 x_2}(f)}{G_{x_1 x_1}(f) G_{x_2 x_2}(f)} \quad (4.21)$$

Coherence is a real valued function with bounds $0 < |\gamma(f)|^2 < 1$ [102]. Since GCC-PHAT demonstrates desirable results for both reverberant and “lossy” environments, we opted to use PHAT weights in our implementation. PHAT filter function is defined as,

$$\Psi(f) = \frac{1}{|G_{x_1 x_2}(f)|} \quad (4.22)$$

Now we can simplify Equation 4.20 by substituting Equations 4.21 and 4.22.

$$var(\hat{\tau}_0) = \frac{1}{4\pi^2 T} \frac{1}{[\int_{-\infty}^{\infty} f^2 df]^2} \int_{-\infty}^{\infty} \frac{1 - |\gamma(f)|^2}{|\gamma(f)|^2} f^2 df \quad (4.23)$$

The Equation is in continuous frequency domain. However, data in discrete-time domain cannot be analysed in continuous-frequency domain. Since recorded acoustic signals are arrays of samples in discrete time domain, the resulted cross power spectra are DFTs. For a N -point DFT, n^{th} frequency bin can be expressed as,

$$f_n = n\Delta \quad (4.24)$$

where $n = 0, 1, \dots, N - 1$ and Δ is frequency resolution. Assuming $\Delta \ll 1$, the Equation 4.23 can be transformed into DFT domain as,

$$var(\hat{\tau}_0) \approx \frac{1}{4\pi^2 T} \frac{1}{[\sum_{n=0}^{N-1} (n\Delta)^2 \Delta]^2} \sum_{n=0}^{N-1} \frac{1 - |\gamma(f_n)|^2}{|\gamma(f_n)|^2} (n\Delta)^2 \Delta \quad (4.25)$$

Since $\sum_{n=0}^{N-1} n^2 = \frac{N(N+1)(2N+1)}{6}$, it can be further simplified to,

$$\text{var}(\hat{\tau}_0) \approx \frac{9}{\pi^2 T \Delta^3 N^2 (N+1)^2 (2N+1)^2} \sum_{n=0}^{N-1} \frac{1 - |\gamma(f_n)|^2}{|\gamma(f_n)|^2} n^2 \quad (4.26)$$

When all the signals are sampled at f_s Hz and when the observation window has M samples, we can write,

$$\begin{aligned} T &= \frac{M}{f_s} \\ \Delta &= \frac{f_s}{N} \end{aligned} \quad (4.27)$$

By using this, we can further simplify the Equation 4.26 as,

$$\text{var}(\hat{\tau}_0) \approx \frac{9N}{\pi^2 M f_s^2 (N+1)^2 (2N+1)^2} \sum_{n=0}^{N-1} \frac{1 - |\gamma(f_n)|^2}{|\gamma(f_n)|^2} n^2 \quad (4.28)$$

Signal model given in 3.4 for acoustics can be transformed into discrete Fourier domain as,

$$\begin{aligned} x_1(m) &= h_1(m) * s(m) + \omega_1(m) \leftrightarrow X_1(f_n) = H_1(f_n)S(f_n) + W_1(f_n) \\ x_2(m) &= h_2(m) * s(m) + \omega_2(m) \leftrightarrow X_2(f_n) = H_2(f_n)S(f_n) + W_2(f_n) \end{aligned} \quad (4.29)$$

Assuming cross power spectral density of two uncorrelated signal is negligible [81][86], we can write

$$\begin{aligned} G_{x_1 x_2}(f_n) &= H_1(f_n)H_2(f_n)^*G_{ss}(f_n) \\ G_{x_1 x_1}(f_n) &= H_1(f_n)H_1(f_n)^*G_{ss}(f_n) + G_{w_1 w_1}(f_n) \\ G_{x_2 x_2}(f_n) &= H_2(f_n)H_2(f_n)^*G_{ss}(f_n) + G_{w_2 w_2}(f_n) \end{aligned} \quad (4.30)$$

Assuming both the signals have same noise power spectral density (i.e. $G_{w_1 w_1}(f_n) = G_{w_2 w_2}(f_n) = G_w(f_n)$), we can rewrite coherence as,

$$|\gamma(f_n)|^2 = \frac{|H_1(f_n)H_2(f_n)^*G_{ss}(f_n)|^2}{|[H_1(f_n)H_1(f_n)^*G_{ss}(f_n) + G_w(f_n)][H_2(f_n)H_2(f_n)^*G_{ss}(f_n) + G_w(f_n)]|} \quad (4.31)$$

Now Signal to Noise Ratio (SNR) for each frequency bin is defined as,

$$\text{SNR}(f_n) = \frac{G_{ss}(f_n)}{G_w(f_n)} \quad (4.32)$$

Again assuming cross power spectral density of two uncorrelated signal is negligible, Equation 4.31 can be rewritten as,

$$|\gamma(f_n)|^2 = \frac{|H_1(f_n)H_2(f_n)^*|^2 SNR(f_n)^2}{|H_1(f_n)H_2(f_n)^*|^2 SNR(f_n)^2 + [H_1(f_n)H_1(f_n)^* + H_2(f_n)H_2(f_n)^*] SNR(f_n) + 1} \quad (4.33)$$

Using this, Equation 4.28 can be expressed as,

$$var(\hat{\tau}_0) \approx \frac{9N}{\pi^2 M f_s^2 (N+1)^2 (2N+1)^2} \sum_{n=0}^{N-1} \frac{[H_1(f_n)H_1(f_n)^* + H_2(f_n)H_2(f_n)^*] SNR(f_n) + 1}{|H_1(f_n)H_2(f_n)^*|^2 SNR(f_n)^2} n^2 \quad (4.34)$$

Variance of TDOA estimated using GCC-PHAT is a function of reverberation ($H_i(f_n)$), SNR, sampling rate (f_s), no of samples collected (M) and number of frequency bins in DFT (N bins). Since the sampling rate is inversely proportional to variance, higher the sampling rate, smaller the variance. Also wide observation window results lower variance.

GCC can be written in time domain as cross correlation between two filtered signals [86]. When $\tilde{x}_1(t)$ and $\tilde{x}_2(t)$ are the two time domain filtered signals of $x_1(t)$ and $x_2(t)$, cross-correlation taken around true delay is given by,

$$R_{x_1 x_2}(\hat{\tau}_0) = E\{\tilde{x}_1(t) \cdot \tilde{x}_2(t - \hat{\tau}_0)\} \quad (4.35)$$

Each sample in $\tilde{x}_1(t)$ is multiplied with that of $\tilde{x}_2(t - \hat{\tau}_0)$ and average is taken. Here, signals were inserted with empty packets to represent packet loss. Samples multiplied by zeros does not contribute to the cross-correlator. Figure 4.13 illustrates “effective number of samples”. Effective number of samples are the samples that are not multiplied by empty packets. Since variance of GCC-PHAT estimated time delay derived in Equation 4.34 is valid only around neighbourhood of true delay, M , samples in observation window can be replaced with effective number of samples (\tilde{M}). When the effective number of samples is larger compared to time delay, this can represent the effect of packet loss in the derived variance expression. Thus, modified expression for variance of GCC-PHAT

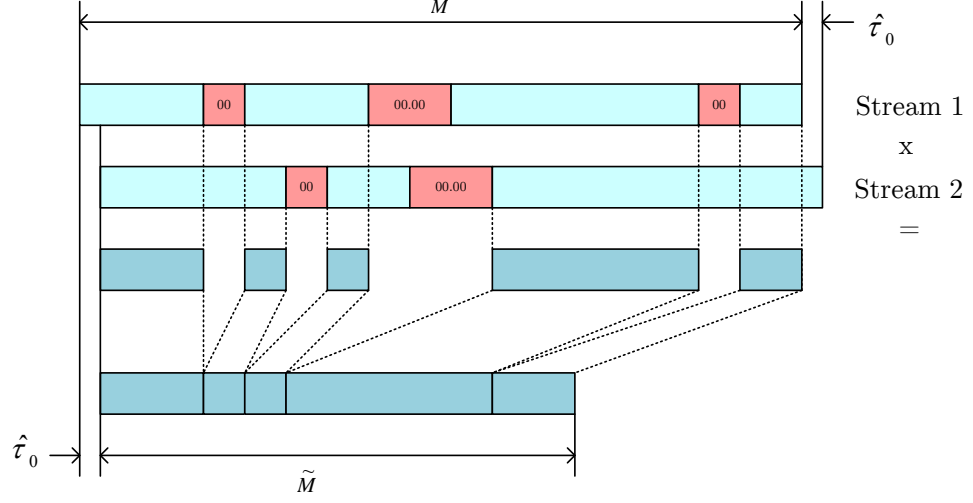


Figure 4.13: Number of samples after with packet loss

estimated time delay involving effects of packet loss can finally be written as,

$$var(\hat{\tau}_0) \approx \frac{9N}{\pi^2 \tilde{M} f_s^2 (N+1)^2 (2N+1)^2} \sum_{n=0}^{N-1} \frac{[H_1(f_n)H_1(f_n)^* + H_2(f_n)H_2(f_n)^*]SNR(f_n) + 1}{|H_1(f_n)H_2(f_n)^*|^2 SNR(f_n)^2} n^2 \quad (4.36)$$

4.4 Chapter Summary

Sound waves reflected from smooth surfaces overlap each other and create reverberant environments. When source and sensor locations are fixed and properties of the room are stationary, reverberation is time invariant. For time delay estimation, reverberation is undesirable. When it's severe, it violates Gaussian noise assumption. Thus, MLE based estimations methods doesn't perform well with reverberation. Among all the GCC filtering methods, PHAT weights perform robustly when reverberation is the main concern.

Packet loss dramatically reduces performance of time delay estimation regardless the GCC method used. The maximum packet loss GCC can accommodate without affecting the performance is restricted by maximum possible delay between two sensors, sampling rate, number of samples per packet and overlap between the two signals. Usually it is less than 2% and the percentage of anomalous estimates increases rapidly as packet loss rate increases beyond 2%. Thus, time delay cannot be estimated accurately in a lossy environment.

In WSN, Zigbee standard is commonly used for communication. It offers sequence number for transmitting frames in MAC layer and packets in network layer. Using sequence number packet loss can easily be detected. In a situation where sequence number is not available, it can be easily embedded to the payload.

Inserting empty packets in places where packet loss is detected helps to maintain samples aligned in time domain. The approach is simple, but the performance is significantly improved. For PHAT, the desired GCC method for reverberant environment, the correction works with negligible errors upto 15% packet loss rate when SNR is high.

Chapter 5

Proposed Method and Experimental Results

Utilizing desirable properties of both approaches, the proposed method is a hybrid version of traditional close-form localization and beamforming. Here, SRP-PHAT beamformer is used to produce fine-grained location estimates, and a close-form LLS estimator is used to minimize search space of the beamformer.

5.1 Proposed Search Algorithm

As mentioned in previous chapters, a TDOA measurement restricts location of the target into a hyperbola. In ideal case, the exact location is given by the intersection point of multiple hyperbolas for different pairwise TDOA measurements. Due to anomalies in collected data, TDOA measurements are prone to errors. Thus, the hyperbolas rarely intersect on a single point. However, it is safe to assume the hyperbola respective to the most reliable TDOA measurement lies closer to the actual speaker location than the hyperbolas respective to other TDOAs. Thus, the source location should be somewhere in the neighbourhood of the most reliable hyperbola. The most reliable hyperbola is represented by the notation $\tilde{B}(\theta)$. Instead steering the beamformer on all the possible points in the space, it can be steered only along $\tilde{B}(\theta)$. This reduces the number of points searched in the proposed search algorithm.

First, location of the target is estimated approximately using a fast, close-form method. It is represented by the notation $\hat{\theta}_e$. Due to simplicity and lack of demand for additional information, we opted using the LLS estimator derived for the TDOA model in section 2.4 as our close-form estimator. Estimation error reduces as accuracy of the TDOA measurement increases. Since accurate estimates lie closer to the actual

location than the less accurate estimates, it is assumed the accurate estimates lie closer to the most reliable hyperbola. Suppose θ_c is the closed point to $\hat{\theta}_e$ on $\tilde{B}(\theta)$. Euclidean distance between $\hat{\theta}_e$ and θ_c can be considered as a reliable estimate for accuracy of $\hat{\theta}_e$. When the distance is smaller, it is assumed $\hat{\theta}_e$ is more accurate. Thus, a shorter portion of the hyperbola is selected. When it is larger, that is when $\hat{\theta}_e$ is less accurate, a longer portion is selected. The selected portion of $\tilde{B}(\theta)$ is centred around θ_c . Exact length of the selected portion can be expressed as,

$$l(\theta) = \eta D\{\hat{\theta}_e, \tilde{B}(\theta)\} \quad (5.1)$$

where η is a proportional constant and $D\{.\}$ is the Euclidean distance operator.

Even the most reliable TDOA measurements are with noise. Thus, the respective hyperbola and the hyperbola respective to the theoretical TDOA does not overlap each other all the time. Searching only along a line can miss the spatial likelihood maximum. Thus, an area covering both side of the hyperbola should be selected allowing tolerance for TDOA variations. If allowable tolerance is $\pm\rho$, the selected search area can be expressed as,

$$A(\theta) = 2l(\theta)\rho \quad (5.2)$$

The search space is now limited to a specific portion of the total area. However, searching all the possible points inside the region is computationally exhaustive and pointless. Thus, a point grid with a specific resolution is considered. If $J(\lambda)$ is a point grid with resolution λ , proposed search points can be expressed as,

$$A(\theta) \cap J(\lambda) \quad (5.3)$$

where \cap is the intersection operator.

An example of search area selected using the proposed algorithm is illustrated in Figure 5.1. Parameters η , ρ and λ were chosen as 1, 0.1 m and 0.02 m respectively.

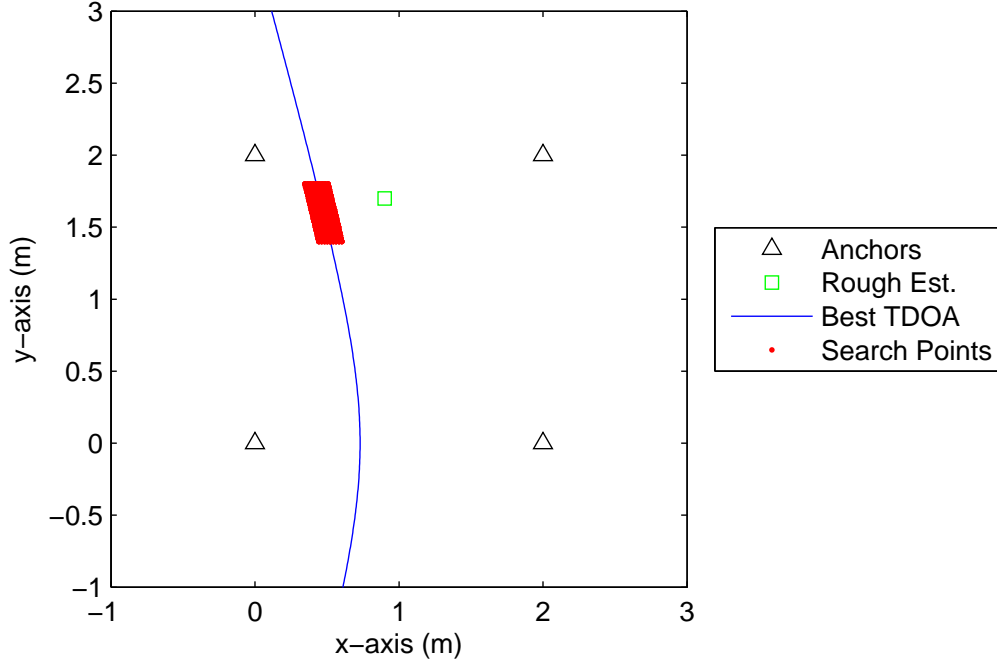


Figure 5.1: Proposed search area: $\theta = (0.5, 1.5)$, $\hat{\theta}_e = (0.9, 1.7)$, $l(\theta) = 0.4575$ m, $\eta = 1$, $\rho = \pm 0.1$ m, $\lambda = 0.02$ m

There are 228 total number of search points. If traditional search method with the same resolution is used, there are 10000 possible search points inside the 2×2 m² area. So in this case, only 2.3% of search points is used.

The number of search points resulted by the proposed method is controlled by accuracy of the rough estimate, grid resolution (λ), tolerance for the hyperbola (ρ) and the proportional constant (η). Average LLS localization error for the configuration given was experimentally found as 0.4575 m. Average cross-section area at the base of the spatial likelihood maximum was experimentally found as 0.01 m². Thus, λ was selected as 0.02 m so that there will be at least 25 grid points inside the peak. For a given average localization error and grid resolution, the average selected search points is restricted

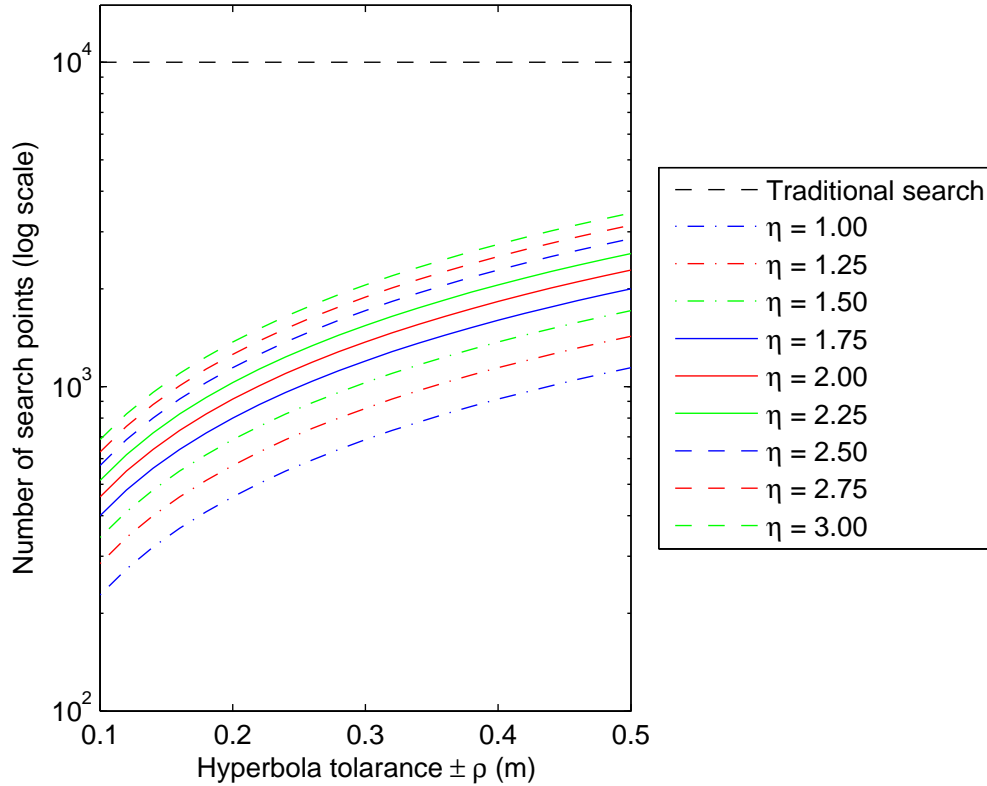


Figure 5.2: Number of search points vs. η vs. ρ

only by parameters η and ρ . Figure 5.2 illustrates variation of search points for different η and ρ values.

Parameter η is varied from 1 to 3 in 0.25 steps. For each η value, ρ is varied from 0.1 m to 0.5 m in 0.02 m steps. For each values, number of selected search points as a percentage of total possible points varies between 2.3% and 34%. Still only two third of search points in the traditional full search is used.

5.2 Modifications to the Experimental Setup

The proposed method was implemented on a network of four IRIS[®] nodes. For ease of comparison, nodes were placed on the same 2×2 m configuration as did with the previous experiments. However, since speech is wideband, hardware tone detectors were not

used. Samples were forwarded to a computer for processing and analysis. Localization is performed at the computer using recorded signals. To satisfy those requirements, we made several modifications to the system architecture.

5.2.1 Communication

Since most of the high power speech content is limited between 300-3400 Hz, microphones have to be sampled at least 7 kHz to avoid aliasing effects. Allowing guard bands as in telecommunication standards, we selected 8 kHz sampling [75]. The 8 kB internal RAM in IRIS[®] is not sufficient to store both the samples and all the other runtime variables [66]. Overall throughput of writing samples into the flash memory is 1.6 kBps (800 16-bit samples per second) which is not fast enough for 8 kHz sampling [67]. However, nodes can communicate with each other at 250 kbps [68]. Thus, we concluded streaming samples real-time to the base station is the only option available in our setup.

However, when all the four nodes stream samples at the same time, congestion occurs. Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) is used by default in TinyOS Medium Access Control (MAC) protocol. However, with CSMA/CA we observed significant amount of delays in the packet transmission. After few seconds, nodes stopped communicating. Since the ADC samples the microphone continuously but the transceiver cannot transmit them at the same speed when there is congestion, we concluded it is due to buffer overflow. When Time Division Multiple Access (TDMA) was implemented, we observed packet loss after few seconds. For TDMA, a tight clock synchronization among all the participating nodes is needed. Since the low cost nodes are prone to clock drifts, we concluded it as the reason [39]. However, when there was only one streaming node, all the packets were successfully delivered. Thus, we selected a Frequency Division Multiple Access (FDMA) approach assigning dedicated channel for each transmitting node.

AT86RF230, the transceiver chip in IRIS[®], operates in 2.4 GHz frequency band. In Zigbee[®] standard, there are 16 different radio channels with 5 MHz spacing defined for 2.4 GHz band [71]. Thus, the four nodes were programmed to transmit packets in four different radio channels. To minimize interference from other coexisting high power WiFi signals, CH-11 (2405 MHz), Ch-15 (2425 MHz), CH-25 (2475 MHz) and CH-26 (2480 MHz) were selected. Since IRIS[®] nodes cannot communicate in different radio channels at the same time, four dedicated interfacing boards for each channel were connected to the base station. Using this technique, we were able to stream samples from nodes to the base stations reliably with minimum losses.

5.2.2 Data Compression

The maximum allowable packet length in IEEE802.15.4/Zigbee[®] standard is 127 bytes [71]. After reserving space for headers, only 100 bytes is available for data samples. Data conversion from analogue to digital by IRIS[®] results 10 bit values [65]. Since byte is the smallest unit of space a microprocessor can access, each sample occupies 2 bytes. Thus, space is available only for 50 samples. But, 6 bits in each sample are “null”. Thus, 2 byte representation of a sample results 37.5% space wastage in each packet.

However, as illustrated in Figure 5.3, samples can be compressed by removing “null” bits. A 10 bit long sample is decomposed into two parts as the least 8 bits and the 2 highest bits. Then, another byte is created using the highest 2 bits from four consecutive samples. The least byte of four samples is inserted into the packet first. Then the newly created byte with 2 bits from each sample is inserted as the fifth byte. By doing this four consecutive samples occupying 8 bytes can be compressed into 5 bytes without losing any data. At the base station, each five consecutive bytes in the incoming packet are dispatched into four 16 bit samples using the opposite of the same mechanism. Hence, 88 samples can be inserted into a single packet.

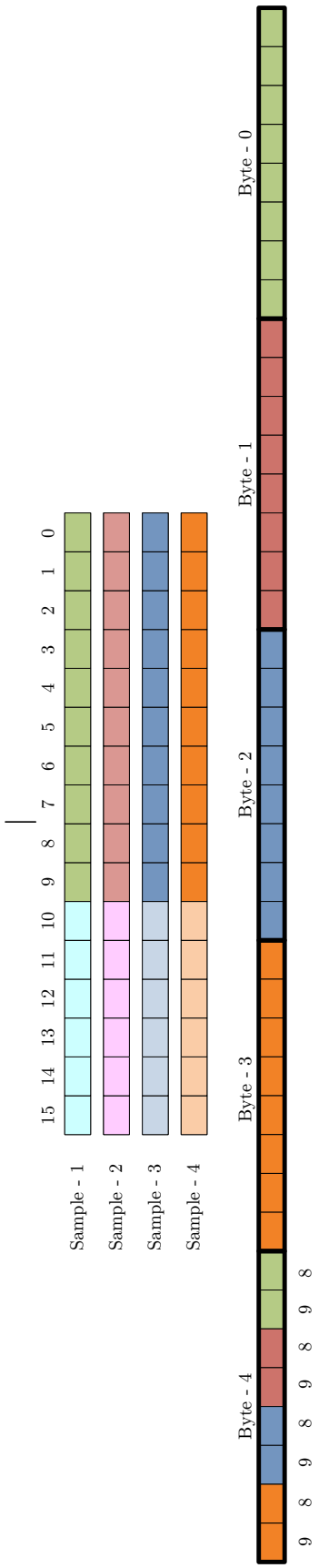


Figure 5.3: Data compression: Top four rows represents four samples occupying 8 bytes. Boxes with dark colours represent data bits and light coloured boxes represent “null” bits. The bottom row represents compressed samples.

5.2.3 Time Synchronization

The time delay estimated between unsynchronized nodes, does not have a strong relationship with the speaker location. Since time delay is estimated between samples recorded at two different anchor nodes, clocks of each node have to be synchronized. “Flooding” is a widely used clock synchronizing technique for wireless sensor nodes [39]. With frequent message passing among the participating nodes, it has achieved sub-microsecond synchronization accuracy. However, in our implementation each microphone is sampled at 8 kHz with 125 μ s sampling interval. Thus, sub-millimetre synchronization accuracy is sufficient. Nevertheless, flooding takes considerable amount of time for convergence. Thus, we opted not to use it.

We synchronized our anchor nodes using a simple technique. First, all the anchor nodes were configured to a listening state. Since we use FDMA, each node uses different radio channels. However, IRIS[®] nodes can change transmission channel both at the programming-time and at the run-time. Thus, all of them were configured to listen to a common channel in the listening state. In our implementation CH-11 (2405 MHz) was selected as the common channel. Then, the base station which operates in CH-11, broadcasts a beacon message. A beacon is a byte long packet with sampling duration as the payload. Whenever each node receives the beacon, it records the reception time and then reads the packet content. TinyOS-2.x provides packet level time stamping [103]. Thus, exact reception time of the packet can be obtained with a minimum error. After 25 ms receiving the beacon, all the nodes start sampling their microphones for the requested sampling duration. Within the 25 ms delay, nodes switch-on their microphones and tune their radios back to the predetermined transmission channels. When the first sample is taken, each node again records the time. Whenever a node have 88 samples, it transmits them to the base station in a single packet. Reception time of the beacon and time when the first sample was taken are piggybacked in the first packet. Once the sampling

duration is elapsed, each node stops sampling and turns back into the listening state.

At the base station, the two time measurements for each node are recorded. Since radio signals travels nearly at the speed of lights, it is assumed all the nodes received the beacon at the same time. Thus, all the beacon reception time measurements reported by individual nodes represent the same event. Then, delay between the beacon reception and the sampling for each node is calculated. All the received samples are then shifted accordingly, so that all of them are aligned in time domain.

5.2.4 Source Detection

There is no point of performing localization if there is no voice signature during the sampling window. Thus, a simple detector was implemented to check whether there is speech content in the received samples. If speech is detected, only then the samples are forwarded to the localizer. We implemented an energy detector considering the noise statistics. Decision threshold was obtained by fixing the false alarm rate to 0.01%. For a given false alarm rate, P_{fa} , decision threshold can be obtained by,

$$Threshold = \sigma_e \cdot Q^{-1}\{P_{fa}\} + \mu_e \quad (5.4)$$

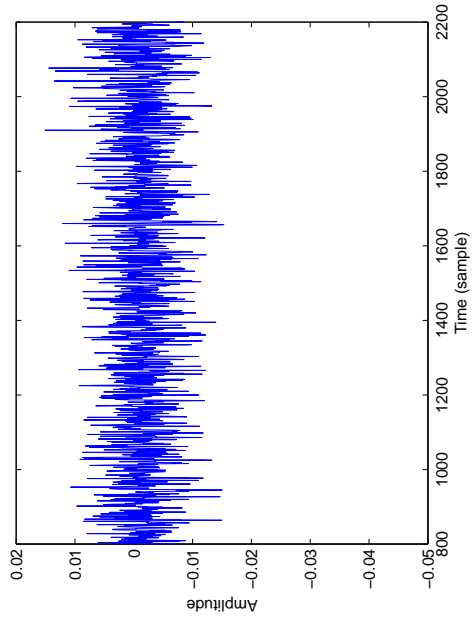
where μ_e and σ_e^2 are mean and variance of energy during the sampling window. Notation $Q\{.\}$ represents the standard Q-function.

To get the statistics, 4400 samples (approximately 500 ms long) were recorded for all the four nodes in a quite environment. We made sure there is no one speaking throughout the experiment. However, as shown in part (a) of Figure 5.4, noise is affected by a periodic signal. Frequency domain analysis given in part (b), confirms it is not white. There are high power frequency components below 300 Hz. From our observations, we concluded it is due to a hum generated by the air conditioning system. Since it is out of the vocal range, it was filtered out using a high pass filter. In part (c) and (d), filtered signal and its power spectrum are given respectively. It is clear the filtered signal is more

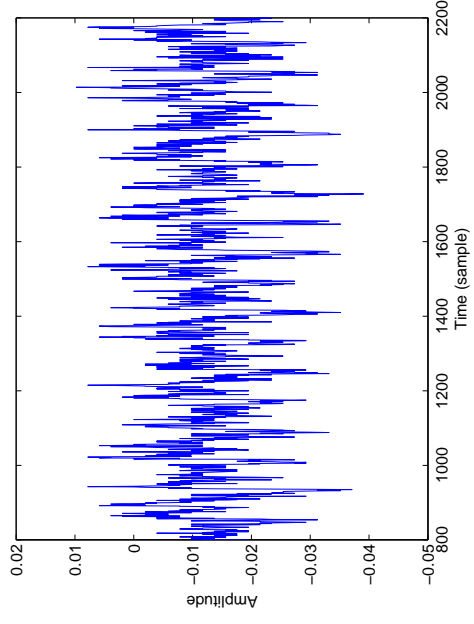
	Node-1	Node-2	Node-3	Node-4
Mean	0.4840	0.3343	0.3570	0.4469
Variance	0.4741×10^{-3}	0.2302×10^{-3}	0.5335×10^{-3}	0.5795×10^{-3}

Table 5.1: Statistics of noise energy for 4400 samples

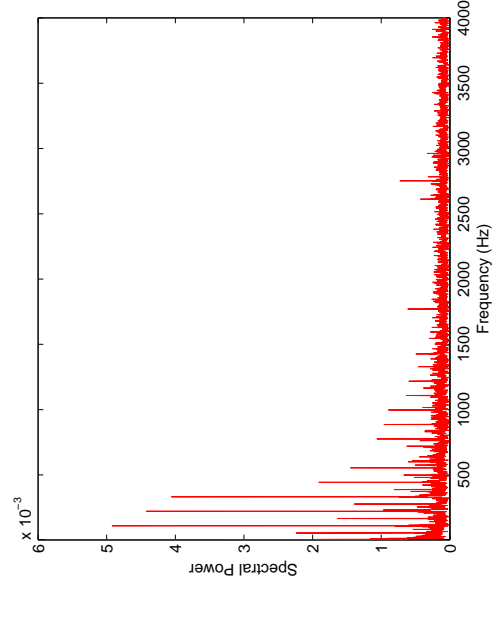
random than the original. Figure 5.5 is a histogram generated for the filtered noise, which is approximately a Gaussian. Thus, it is assumed noise in the filtered samples are additive Gaussian. Then noise energy was measured for 500 trials. Since measurement noise is a function of hardware (IRIS[®] node, sensor board, battery voltages, etc.), mean and variance of the noise energy of each node were recorded separately. They are tabulated in Table 5.1. Using the statistics found, an energy detector was implemented for each anchor node. Localization is performed only when all the four anchor nodes positively detect a signal.



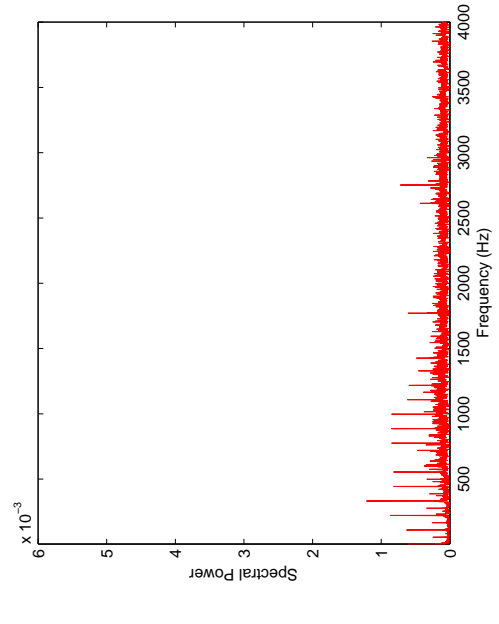
(a) Raw signal



(b) Filtered signal



(c) Power spectrum of raw signal



(d) Power spectrum of filtered signal

Figure 5.4: Effect of filtering

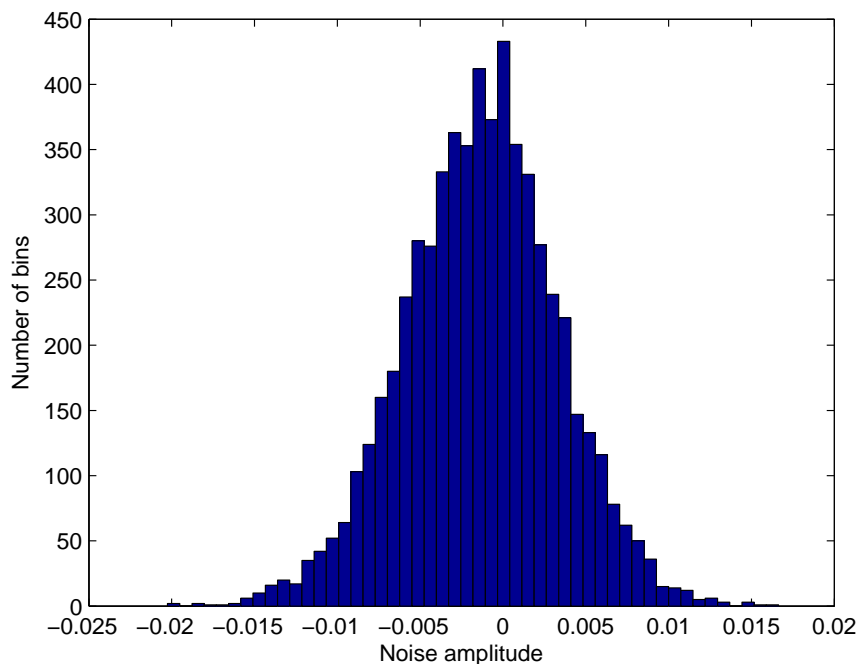


Figure 5.5: Histogram of filtered noise

5.3 Experimental Results

As illustrated in Figure 5.6, the speaker was oriented so that its sound direct inside the node array. Then, a TED[®] talk was played from YouTube[®] to emulate speech. “The paradox of choice” from presenter Barry Schwartz was selected since it has both speech segments as well as considerably amount of silent segments. Each experiment was performed on 4400 samples long (approximately 500 ms) recorded signal segments. For comparison, rough location estimates and estimates by SRP-PHAT with the full search were also recorded.

5.3.1 Performance of the Proposed Method

The speaker was placed at (0.5,1.5) position and localized with the proposed search algorithm for different η (hyperbola tolerance) and ρ (proportional constant) values. For each η and ρ setting, 500 different trials were performed. The localization error for each

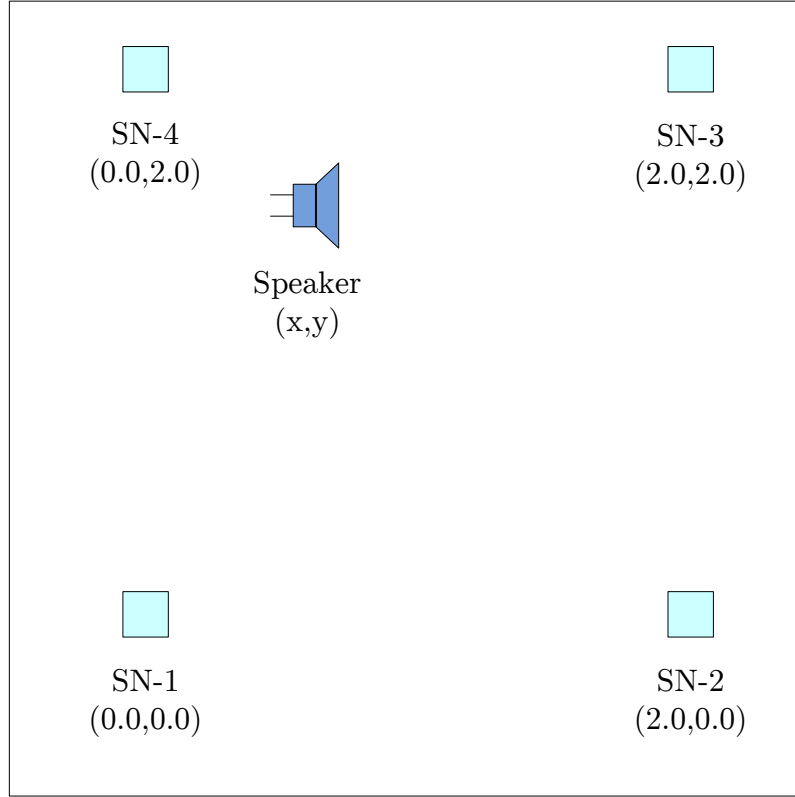


Figure 5.6: 2×2 m anchor node configuration

trial was calculated and their average was taken. Figure 5.7 illustrates recorded average localization error for each setting. Localization performance is always better than the rough estimates. Since all the points are searched, accuracy of SRP-PHAT with the full search is always superior. Thus, performance of SRP-PHAT with the proposed search algorithm is within the full search and the rough estimator. In a scale of 0 to 100 between accuracies of the two methods, accuracy of proposed method is always within 28% from the full search. In other words, the proposed search method is also as robust as the full search. As the hyperbola (or TDOA) tolerance ρ increases, width of the search area increases. When search space increases, probability of covering the maximum of the spatial likelihood also increases. Thus, localization performance increases with ρ . As proportional constant η increases, length of the search area increases. Since this also increases the search space, localization accuracy increases with η .

With $\rho = 0.25$ m and $\eta = 2$, the proposed search algorithm achieves the same accuracy

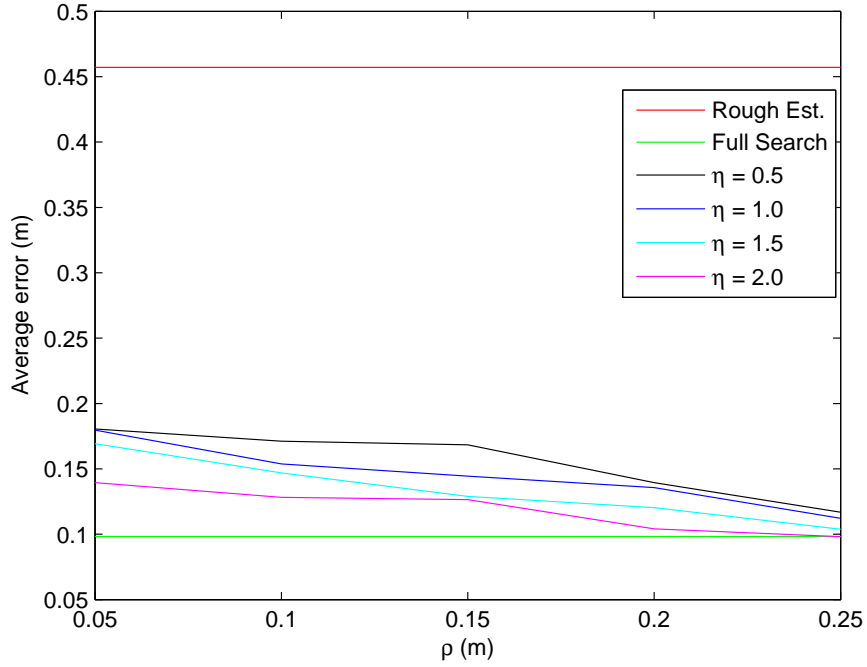


Figure 5.7: Localization performance

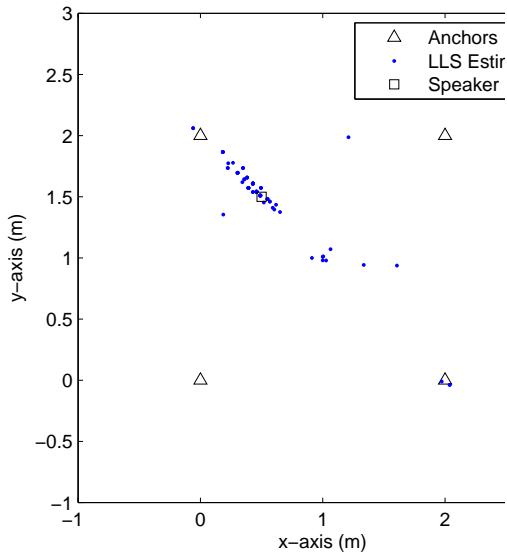
of the full search. Instead searching at all possible points of the $2 \times 2 \text{ m}^2$ area, only $0.4575 \times 2 \times 0.25 \text{ m}^2$ area (i.e. $D\{\hat{\theta}_e, \tilde{B}(\theta)\} \times \eta \times \rho$ considering average LLS localization error 0.4575 m) is searched. Only 11.4% of the total search area is used. In other words, it is a 88.6% improvement. Collecting samples from anchor nodes and performing GCC-PHAT to estimate the pairwise TDOAs also exist in the close-form localization methods. Thus, higher computational cost of the SRP-PHAT localization method is mostly because of the search process. This algorithm uses 88.6% less search points, thus approximately 88.6% less computations. Hence, this method is approximately 88.6% faster than the traditional search methods in the given test configuration. Thus, it is a desirable candidate for a fast responding, yet accurate, SRP-PHAT localization implementation.

Location estimates obtained for $\eta = 2$ and $\rho = 0.2 \text{ m}$ are illustrated in Figure 5.8. Their error statistics are given in Table 5.2. For ease of comparison, only 100 samples

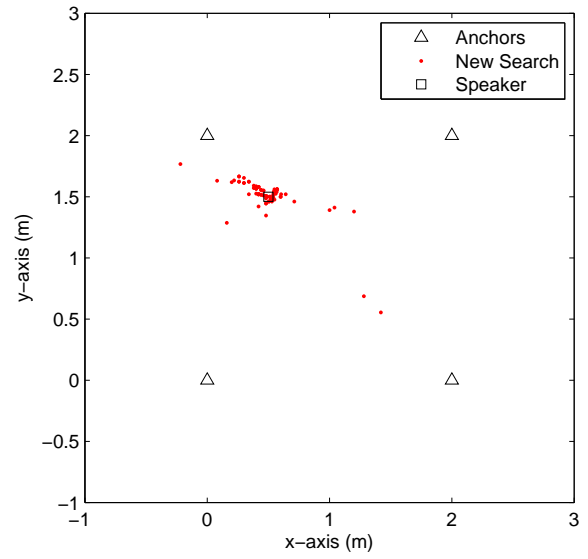
	Mean error (m)	Error variance (m ²)	Bias (m)
Rough estimate	0.4572	0.1756	0.0673
Full search	0.0982	0.0314	0.0104
Proposed search	0.1072	0.0460	0.0227

Table 5.2: Localization error statistics: For proposed search $\eta = 2$, $\rho = 0.2$ m

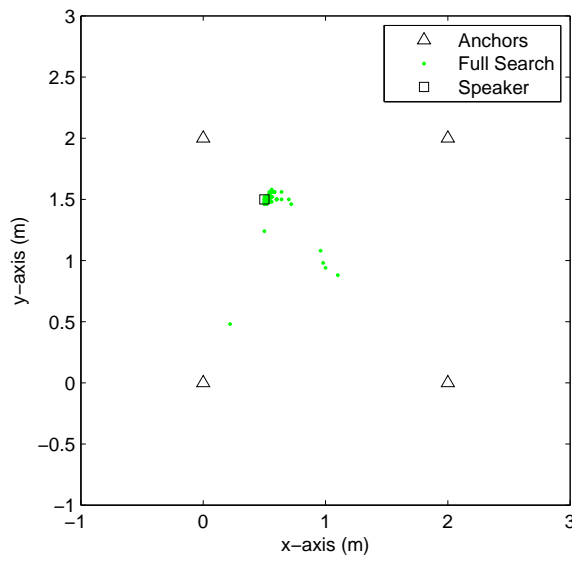
were plotted. LLS estimated rough location estimates are shown in part (a). Since the speaker is closer to the fourth anchor node, it was chosen as the reference node when calculating the TDOAs. Thus, LLS location estimates are biased towards the fourth anchor node. Compared to the results given in part (b) and (c), precision of the LLS estimator is poorer. Part (b) of Figure 5.8 shows location estimated using SRP-PHAT with the proposed search algorithm. Most of the estimates are closer to the actual location. Compared to the LLS method, localization error has reduced and precision has improved. It can clearly be seen that the location estimates from the SRP-PHAT methods are more accurate than the close-form LLS estimates. Precision of the location estimates are considerably higher. The slight bias is mostly because of perturbation when placing the nodes and wide opening of the speaker. The random outliers result when signal content within the sampling window is shorter.



(a) For LLS estimator



(b) For proposed search ($\eta = 2, \rho = 0.2$)



(c) For full search

Figure 5.8: Localization results

5.3.2 Effect of Speaker Location

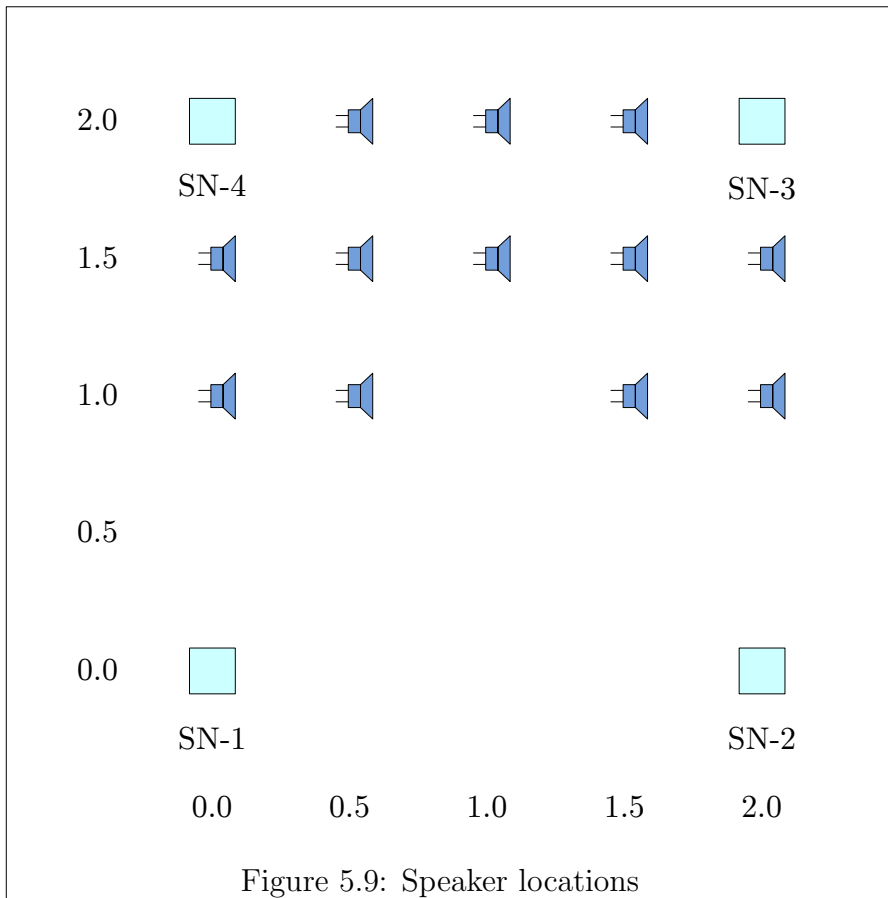
Speech is directional. Thus, when a person speaks, acoustic energy does not propagate omnidirectionally. The space in front of the person receives more acoustic energy than the space behind him. The same theory applies to the hardware speakers as well. Anchor nodes in front of the speaker receives stronger signals than nodes behind the speaker. This affects SNR in recorded signals at each node. Higher the SNR, better the localization performance. Thus, speaker position affects the localization accuracy. To identify performance variation, the following experiment was performed.

The speaker was placed on different locations as illustrated in Figure 5.9. In all the cases, speaker was oriented towards SN-2 and SN-3. Considering symmetrical property, one half of the area was discarded. When the speaker is placed at the center, all the nodes receives the acoustic signal at the same time. Thus, all the TDOA values become zeros. Hence, matrix H , in the LLS estimator defined in Equation 2.14 becomes

$$H = \begin{bmatrix} 4 & 4 & 0 \\ 0 & 4 & 0 \\ 4 & 0 & 0 \end{bmatrix}$$

which is singular. Since H doesn't have an inverse, the LLS estimator doesn't exist. Since it is used as the rough location estimator, the speaker was not placed at the centre of the array.

At each position, the speaker was localized using SRP-PHAT with the proposed search algorithm. To make errors more distinguishable, $\eta = 2$ and $\rho = 0.2$ m were used. For each position, 100 trials were performed and the average localization error was recorded. For ease of comparison, rough location estimates and estimates produced by SRP-PHAT with the full search were also recorded. Localization error w.r.t. speaker location is illustrated in Figure 5.10. Individual values for the rough estimator, the proposed method and the full search are tabulated in Tables 5.3, 5.4 and 5.5 respectively.



	x = 0.0	x = 0.5	x = 1.0	x = 1.5	x = 2.0
y = 1.0	0.9066	0.5947	–	1.7347	2.3931
y = 1.5	0.2325	0.4601	0.6456	1.2293	1.8229
y = 2.0	–	0.7821	1.1711	0.8366	–

Table 5.3: Localization error vs. speaker location: Rough estimates

	x = 0.0	x = 0.5	x = 1.0	x = 1.5	x = 2.0
y = 1.0	0.8237	0.5920	–	0.6219	1.1214
y = 1.5	0.1193	0.1962	0.2529	0.3768	0.7508
y = 2.0	–	0.2462	0.9632	0.3328	–

Table 5.4: Localization error vs. speaker location: Proposed method

	x = 0.0	x = 0.5	x = 1.0	x = 1.5	x = 2.0
y = 1.0	0.6830	0.5017	–	0.4015	1.0833
y = 1.5	0.1105	0.1233	0.1578	0.2178	0.5228
y = 2.0	–	0.1474	0.2775	0.1740	–

Table 5.5: Localization error vs. speaker location: Full search

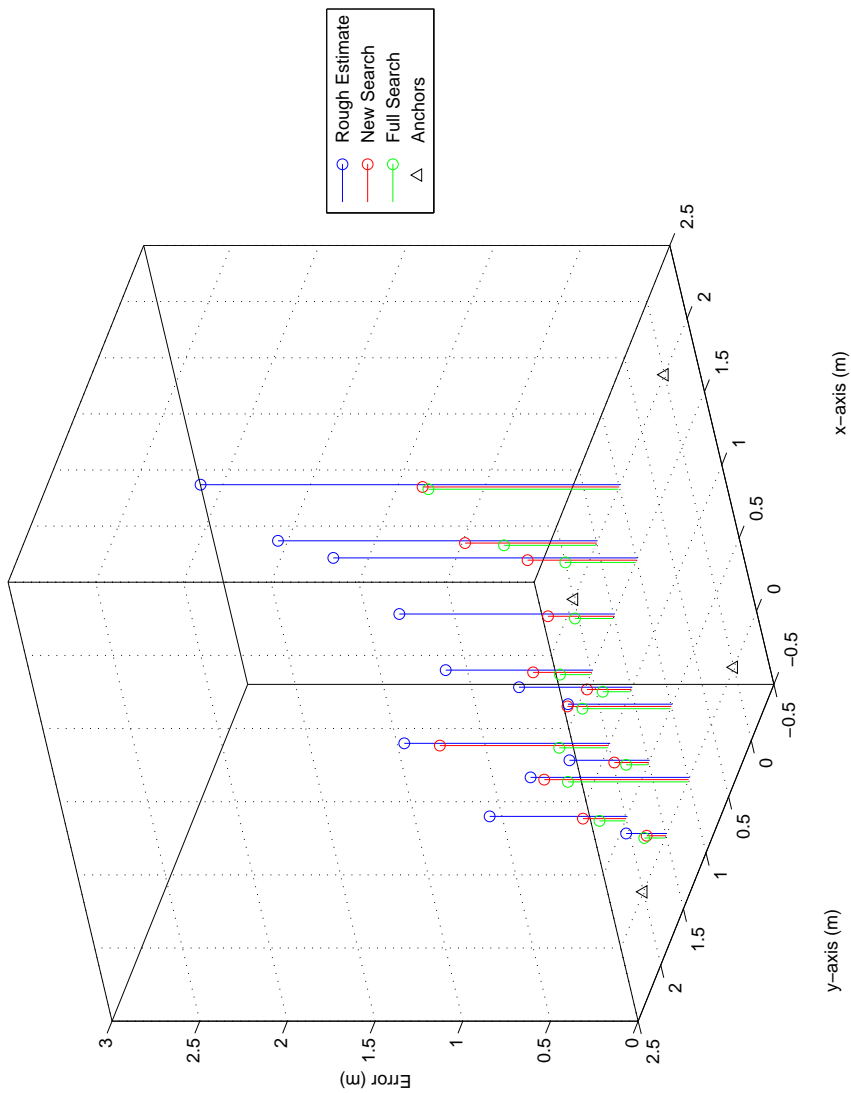


Figure 5.10: Localization error vs. speaker location

The speaker was oriented towards the direction where anchor nodes SN-2 and SN-3 are. As the speaker moves along x-axis, signal reception at SN-2 and SN-3 improves. However, since the speaker is directional, signal reception at SN-1 and SN-4 deteriorates dramatically. This affects the SNR of recorded signals. Since lower SNR deteriorates the time delay estimation performance, it affects the localization accuracy. Thus, localization error increases as the speaker moves along the x-axis. When the speaker is on the edge (i.e. $y = 2$ line), it is relatively away from SN-1 and SN-3 anchor nodes. Due to low SNR, the estimated time delays involving SN-1 and SN-3 are less reliable. Thus, localization performance again deteriorates. When the speaker is located on the symmetrical axis (i.e. $x = 1.0$ or $y = 1.0$ line), node pairs which are at the sides of the axis (i.e. pairs SN-1/SN-2 and SN-3/SN-4 or pairs SN-1/SN-4 and SN-2/SN-3) receive the signal at the same time. Instead crossing each other at location of the speaker, the hyperbolas generated by both the pairwise TDOA measurements completely overlap each other along the symmetric line. This makes the spatial likelihood maximum more like a “hill”. Thus, SRP-PHAT method fails to identify the true peak. This produces considerable amount of errors when the speaker is on a symmetrical axis. Thus, for the given configuration and the speaker orientation, considering all the effects, the best localization estimates result when the speaker is located on $(0.0,1.5)$ or $(0.5,1.5)$ (or considering symmetrical property $(0.0,0.5)$ and $(0.5,0.5)$).

5.3.3 Effect of Reverberation

To illustrate the effect of reverberation on localization performance, the same experiment was performed in a smaller room (width, length and height are $3.5 \times 5.0 \times 4$ m respectively). The actual node configuration in the room is illustrated in Figure 5.11. As size of the room gets smaller, reverberation becomes stronger. Thus, the results obtained in the smaller room reflect the effect of reverberation on acoustic localization.

	x = 0.0	x = 0.5	x = 1.0	x = 1.5	x = 2.0
y = 1.0	2.0720	1.0477	–	1.5953	3.4291
y = 1.5	1.5270	1.2415	1.3425	1.5677	1.0690
y = 2.0	–	3.7371	2.4596	4.4762	–

Table 5.6: Localization error vs. speaker location with high reverberation: Rough estimates

	x = 0.0	x = 0.5	x = 1.0	x = 1.5	x = 2.0
y = 1.0	1.1043	0.7525	–	1.0037	1.3900
y = 1.5	0.4275	0.7921	0.7542	0.7037	0.9015
y = 2.0	–	1.1695	0.7993	0.9950	–

Table 5.7: Localization error vs. speaker location with high reverberation: Proposed method

	x = 0.0	x = 0.5	x = 1.0	x = 1.5	x = 2.0
y = 1.0	0.7950	0.5638	–	0.6509	1.1375
y = 1.5	0.3725	0.3587	0.3124	0.2260	0.8940
y = 2.0	–	0.5339	0.4947	0.5327	–

Table 5.8: Localization error vs. speaker location with high reverberation: Full search

For each speaker position, 100 trials were performed and averaged. Average localization errors of the rough estimator, SRP-PHAT with the proposed method and SRP-PHAT with the full search are tabulated in Tables 5.6, 5.7 and 5.8 respectively. For ease of comparison, they are plotted in Figure 5.12.

It can be clearly seen that all the three methods are affected by the reverberation. Compared to the previous experiment, each method reports higher localization errors at all the tested speaker locations. In the previous experiment, the minimum reported error for the rough estimator was 0.2325 m when the speaker was placed at (0.0,1.5). In the high reverberant environment, it is 1.0477 m at (0.5,1.0), which is more than four times higher. For SRP-PHAT with the proposed search algorithm, the minimum error is 0.4275 m at (0.0,1.5). This is also approximately four times higher than the previous experiment (0.1193 m at (0.0,1.5)). SRP-PHAT with the full search also demonstrates an

increase in the minimum error from 0.1105 m (at (0.0,1.5)) to 0.3124 m ((1.0,1.5)). It is distinguishable that the rough estimator reports higher errors when the speaker is located on the outer edge (i.e. $y = 2$ line). As illustrated in Figure 5.11, SN-3 and SN-4 are closer to a wall. Hence, early acoustic reflections are considerably stronger at those two nodes. Since reverberation become less diffusive with high power reflections, the time delay estimation become less reliable. Thus, the LLS estimator reports higher localization errors when the nodes are closer to reflective surfaces. In conclusion, reverberation deteriorates localization performance.

5.3.4 Effect of Zero Padding

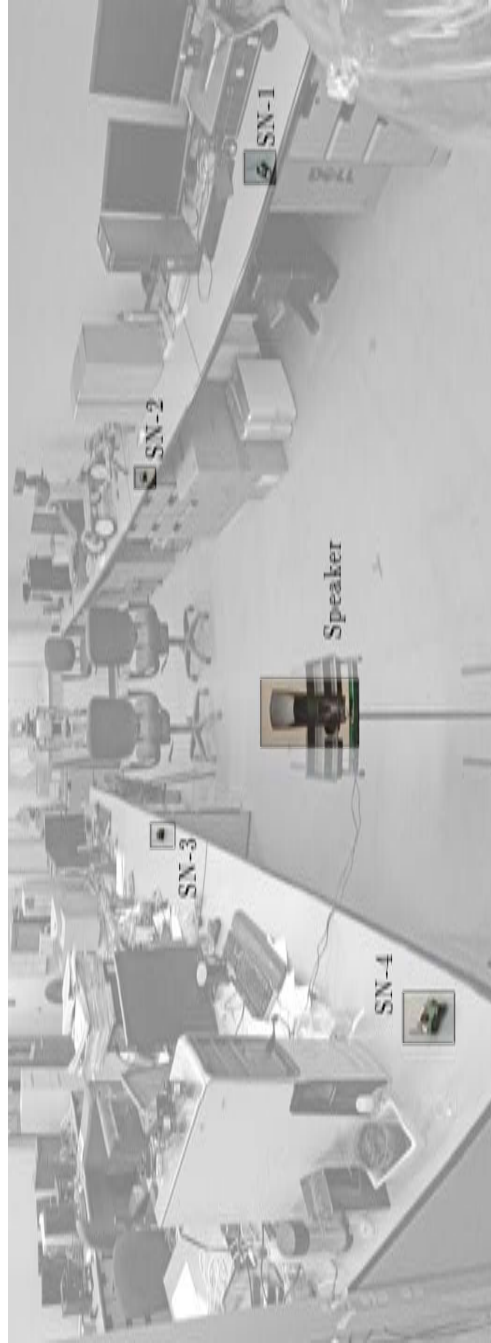
To minimize effect of packet loss, zero padding was introduced in section 4.2.2. Its performance was given in the same section using simulation results. To reconfirm the improvement experimentally, a similar experiment was performed. Four anchor nodes were placed in the low reverberant room as did with the previous experiment. However, to introduce packet loss, batteries of each node were replaced with partially discharged ones. Though we couldn't fix packet loss to a specific value, we observed 1 and 12 missing packets in each trial. Since it needs 50 packets (with 88 samples in each packet) to transmit 4400 samples, there was 2% to 24% packet loss. Then the speaker was placed at (0.5,1.5) position. When dispatching incoming packets into samples at the base station, two sets of samples were maintained, with and without zero padding. Then, 100 localization trials were performed. At each trial, the speaker was localized using both the sample sets. The obtained location estimates are plotted in Figure 5.13.

It can be clearly seen that the estimation precision is higher in the zero padded estimates. For the 100 trials conducted, the LLS estimator completely fails to make accurate estimates without zero padding. It reports 0.8951 m average error without zero padding and 0.5112 m error with zero padding. SRP-PHAT with the proposed

search algorithm reports 0.6943 m error without zero padding. However, with the zero padded sample set, the error has reduced to 0.1402 m. It is evident that the effect of packet loss can be minimized using zero padding.



(a) In the small room



(b) Ordinary setup

Figure 5.11: Experimental setup

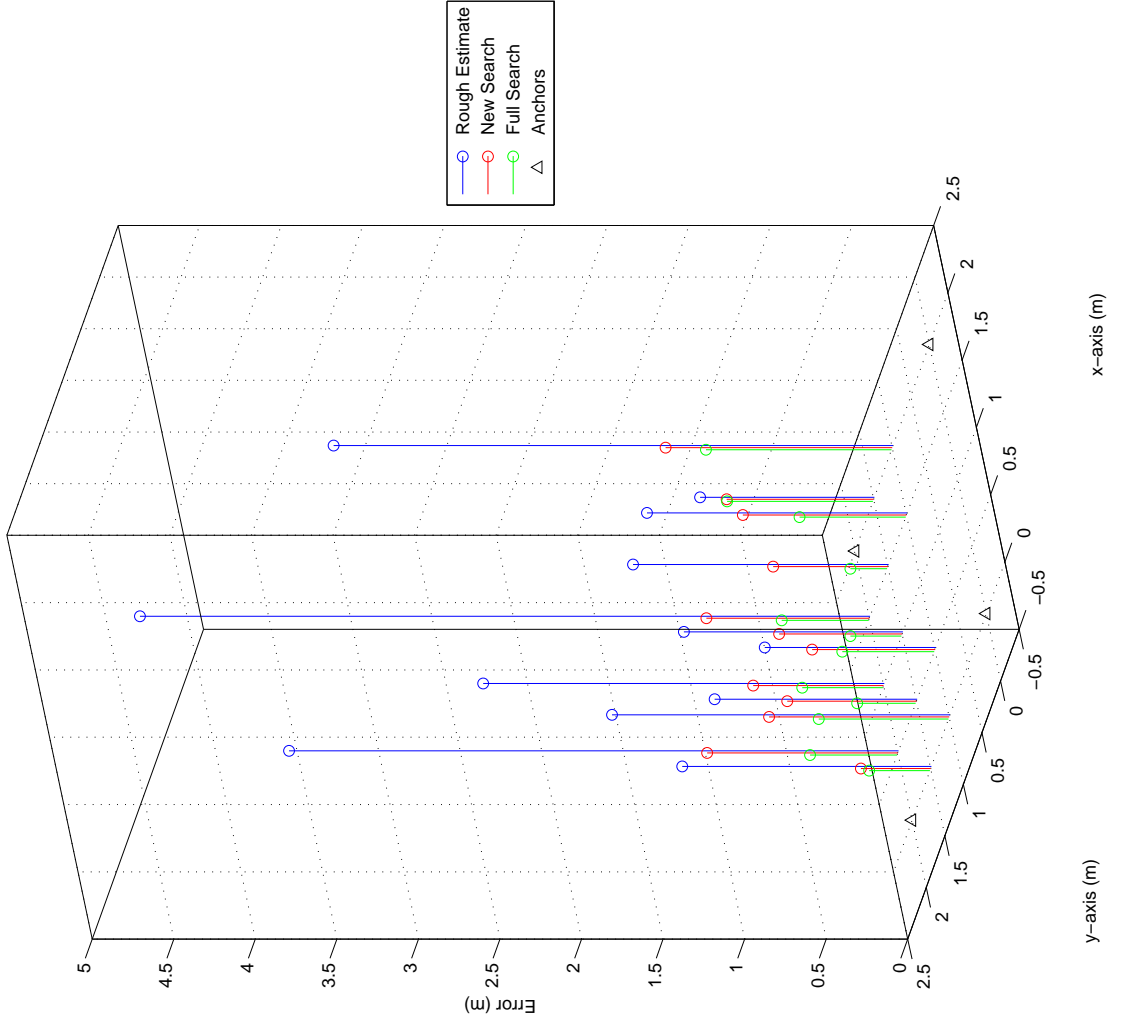
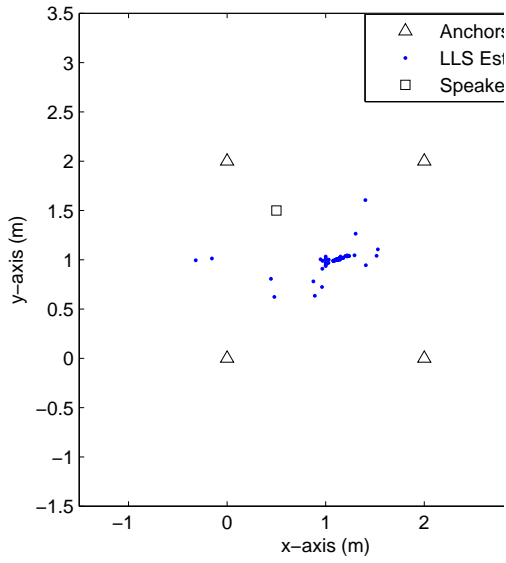
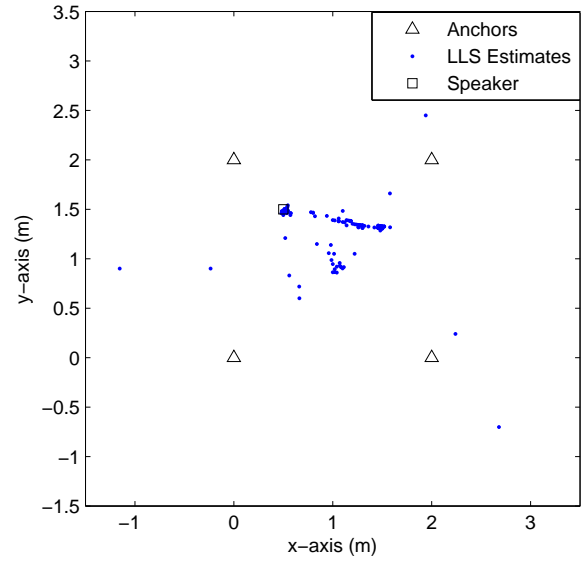


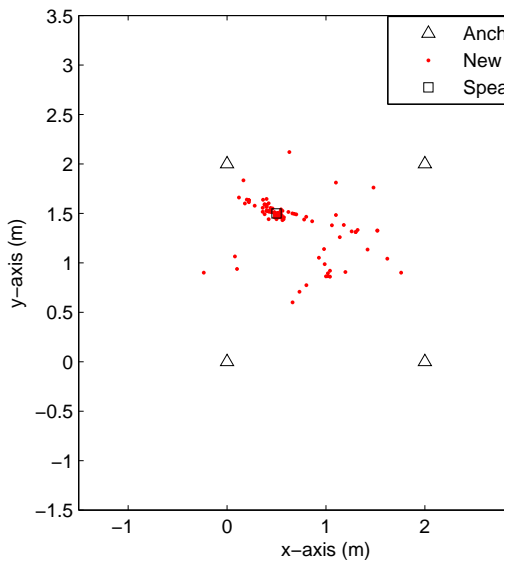
Figure 5.12: Localization error in a high reverberant environment



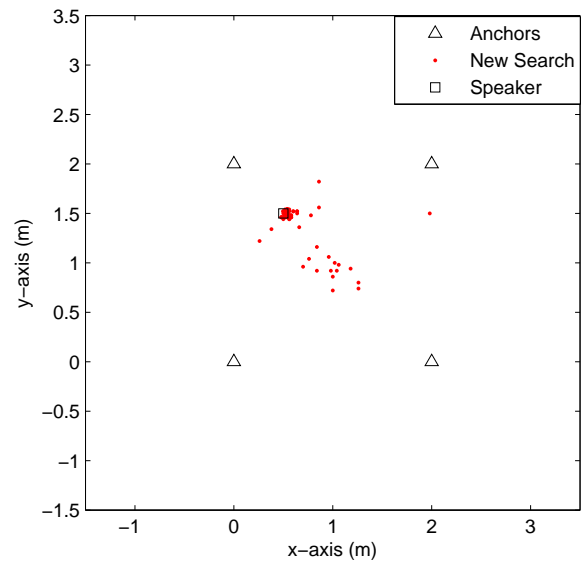
(a) LLS without ZP



(b) LLS with ZP



(c) New search without ZP



(d) New search with ZP

Figure 5.13: Effect of zero padding

Chapter 6

Conclusion

This study is focused on person localization using wireless sensor networks. It can be performed either by using a wearable sensor node or by treating a speaking person as a wideband signal source. To localize wearable sensor nodes, signal strength decay and time of flight are two major observations. However, decaying signal strength is more vulnerable to environmental changes than time of flight. Due to fading and multipath effects, the observations become unreliable. Since antennas in the sensor nodes are not omnidirectional, the signal strength decay also depend on the antenna orientation. Thus, we concluded signal strength decay or using RSSI for person localization is not reliable. Time of flight based approach demonstrates promising sub-metre localization accuracy. Since TDOA approach doesn't need time of transmission, it has more practical usage than the TOA approach.

Since a talking person can be considered a wideband signal source, he/she can be localized without using any wearable sensor node. However, acoustic reverberation degrades localization performance. For estimating time delay between speech signals, GCC methods have demonstrated better results than simple cross-correlation. When noise in acoustic data is additive Gaussian, the MLE filter function is optimal for GCC. However, reverberation effects violate Gaussian noise assumption. Thus, for higher reverberation MLE filter become unreliable. Due to reverberation suppression properties, PHAT filter function perform optimal for high reverberation situations. Statistical properties of both noise and transmitted signal are needed for the MLE. PHAT filter needs only power spectral density of received signal. Thus, using PHAT filter is more practical and convenient.

Packet loss dramatically reduces performance of the time delay estimation process regardless the GCC method used. The maximum allowable packet loss without affecting the time delay estimation is a function of sampling rate, number of samples per packet, the maximum possible delay and overlap between the two signals. Percentage of anomalies increases dramatically as packet loss rate exceed 2%. Time delay estimation is inaccurate in a lossy environment. Replacing missing packets with empty packets can minimize effect of packet loss. Using packet sequence number, missing packets can be identified. An empty packet can be inserted where a loss is detected so that all the samples can be time domain aligned. For higher SNR, when using GCC-PHAT, this correction can make the time delay estimation with negligible errors upto 15% loss.

The close-form TDOA localization methods like LLS and BLUE, operate in two steps. First, TDOAs are estimated from the measurements and then target is localized using the estimated TDOAs. All the information available in the signal is concluded into a single TDOA parameter. Linearisation process involved in getting the close-form solution cannot use TDOAs from all the sensor pairs. Thus, the two-step methods can not use the data efficiently. In contrast, beamformer based single-step methods combine the measurements from anchor nodes instead explicitly finding the TDOAs. With the help of reverberant resistant PHAT filter, SRP-PHAT beamformer uses all the data efficiently. Instead giving a close-form location estimate, it generates a spatial likelihood function combing all the GCC-PHAT functions. Location of the target is given by point that maximize the likelihood function. Since likelihood function is not smooth, a traditional optimization technique cannot be used. Thus, searching for the maximum is computationally exhaustive.

The proposed method uses desirable properties of both the close-form localization and beamforming. A close-form LLS method is used to make a rough location estimate. Then, SRP-PHAT beamformer is used in the neighbourhood to get a fine-grained lo-

cation estimate. Traditional searching approach is searching at all the possible points. Thus, computational cost depends on dimensions of the search space. The proposed method search along the most reliable TDOA measurement. Its computational cost depends on accuracy of the rough estimate, a proportional constant ρ and TDOA error tolerance (η). Our method has achieved the same localization accuracy compared to the full search with $\rho = 2$ and $\eta = 0.25$ m. However, only 11.4% of space is searched. In other words approximately 88.6% less computations, thus 88.6% faster than traditional method. Thus, it is a desirable candidate when a fast responding yet accurate localization method is needed.

Appendix A

GCC Derivations

Consider model of two signals given.

$$\begin{aligned}x_1(t) &= s(t) + \omega_1(t) \\x_2(t) &= \alpha s(t + \tau_0) + \omega_2(t)\end{aligned}\tag{A.1}$$

$\omega_i(t)$ is uncorrelated Gaussian noise. This can be Fourier transformed as,

$$\begin{aligned}X_1(f) &= S(f) + W_1(f) \\X_2(f) &= \alpha S(f).e^{j2\pi f\tau_0} + W_2(f)\end{aligned}\tag{A.2}$$

Delayed cross correlation for two signals $a(t)$ and $b(t)$ is defined as,

$$\begin{aligned}R_{ab}(\tau) &= E\{a(t).b(t - \tau)\} \\&= \int_{-\infty}^{\infty} G_{ab}(f)e^{-j2\pi f\tau} \\G_{ab}(f) &= A(f).B(f)^*\end{aligned}\tag{A.3}$$

where $E\{.\}$ is expectation operation and $G_{ab}(f)$ is Fourier transformed $R_{ab}(\tau)$ [86].

Using model given in A.2 and equation A.3, we can write,

$$\begin{aligned}G_{x_1x_1} &= X_1(f).X_1(f)^* \\&= [S(f) + W_1(f)].[S(f) + W_1(f)]^* \\&= S(f).S(f)^* + S(f).W_1(f)^* + W_1(f).S(f)^* + W_1(f).W_1(f)^* \\&= G_{ss}(f) + G_{sw_1}(f) + G_{sw_1}(f) + G_{w_1w_1}(f) \\&= G_{ss}(f) + G_{w_1w_1}(f)\end{aligned}\tag{A.4}$$

$$\begin{aligned}
G_{x_1x_2} &= X_1(f).X_2(f)^* \\
&= [S(f) + W_1(f)].[\alpha S(f)e^{j2\pi f\tau_0} + W_2(f)]^* \\
&= S(f).\alpha S(f)^*e^{-j2\pi f\tau_0} + S(f).W_2(f)^* + W_1(f).\alpha S(f)^*e^{-j2\pi f\tau_0} + W_2(f).W_1(f)^* \\
&= \alpha G_{ss}(f)e^{-j2\pi f\tau_0} + G_{sw_2}(f) + \alpha G_{sw_1}(f)e^{-j2\pi f\tau_0} + G_{w_1w_2}(f) \\
&= \alpha G_{ss}(f)e^{-j2\pi f\tau_0}
\end{aligned} \tag{A.5}$$

since uncorrelated noise $G_{sw_1}(f), G_{sw_2}(f), G_{w_1w_2}(f) = 0$.

PHAT is defined as [88],

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} \frac{G_{x_1x_2}}{|G_{x_1x_2}|} e^{j2\pi f\tau} df \tag{A.6}$$

Using equations in A.4 and A.5, this can be written as,

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} \frac{\alpha G_{ss}(f)e^{-j2\pi f\tau_0}}{|\alpha G_{ss}(f)e^{-j2\pi f\tau_0}|} e^{j2\pi f\tau} df \tag{A.7}$$

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} \frac{G_{ss}(f)e^{-j2\pi f\tau_0}}{|G_{ss}(f)|} e^{j2\pi f\tau} df \tag{A.8}$$

Using properties of convolution with a delta function, this can be written as,

$$R_{x_1x_2}(\tau) = \delta(\tau - \tau_0) * \int_{-\infty}^{\infty} \frac{G_{ss}(f)}{|G_{ss}(f)|} e^{j2\pi f(\tau)} df \tag{A.9}$$

Bibliography

- [1] M. Skolnik, *Radar handbook*. New York, NY: McGraw-Hill, 1970.
- [2] A. Waite, *Sonar for practising engineers*. John Wiley & Sons, 2002.
- [3] G. Carter, “Time delay estimation for passive sonar signal processing,” *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 29, no. 3, pp. 463 – 470, June 1981.
- [4] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: a survey,” *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.
- [5] B. Lo, S. Thiemjarus, R. King, and G. Yang, “Body sensor network-a wireless sensor platform for pervasive healthcare monitoring,” in *Proc. The 3rd International Conference on Pervasive Computing*, vol. 13, 2005.
- [6] G. Werner-Allen, J. Johnson, M. Ruiz, J. Lees, and M. Welsh, “Monitoring volcanic eruptions with a wireless sensor network,” in *Proc. The 2nd European Workshop on Wireless Sensor Networks*. IEEE, 2005, pp. 108–120.
- [7] J. Polastre, R. Szewczyk, A. Mainwaring, D. Culler, and J. Anderson, “Analysis of wireless sensor networks for habitat monitoring,” *Wireless sensor networks*, pp. 399–423, 2004.
- [8] I. Stoianov, L. Nachman, S. Madden, T. Tokmouline, and M. Csail, “Pipenet: A wireless sensor network for pipeline monitoring,” in *Proc. The 6th International Symposium on Information Processing in Sensor Networks*. IEEE, 2007, pp. 264–273.

- [9] G. Simon, M. Maróti, A. Lédeczi, G. Balogh, B. Kusy, A. Nádas, G. Pap, J. Sallai, and K. Frampton, “Sensor network-based countersniper system,” in *Proc. International Conference on Embedded networked sensor systems*, ser. SenSys ’04. New York, NY, USA: ACM, 2004, pp. 1–12.
- [10] R. Moen, J. Pastor, Y. Cohen, and C. Schwartz, “Effects of moose movement and habitat use on gps collar performance,” *The Journal of wildlife management*, pp. 659–668, 1996.
- [11] M. Sichitiu and V. Ramadurai, “Localization of wireless sensor networks with a mobile beacon,” in *Proc. IEEE International Conference on Mobile Ad-hoc and Sensor Systems*, Oct. 2004, pp. 174–183.
- [12] M. Sugano, “Indoor localization system using RSSI measurement of wireless sensor network based on Zigbee standard,” in *Wireless and Optical Communications*. IASTED/ACTA Press, 2006, pp. 1–6.
- [13] N. Patwari, R. O’Dea, and Y. Wang, “Relative location in wireless networks,” in *Proc. IEEE VTC*, vol. 2, May 2001, pp. 1149–1153.
- [14] N. Patwari, I. Hero, A.O., M. Perkins, N. Correal, and R. O’Dea, “Relative location estimation in wireless sensor networks,” *IEEE Trans. on Signal Processing*, vol. 51, no. 8, pp. 2137–2148, Aug. 2003.
- [15] N. Bulusu, V. Bychkovskiy, D. Estrin, and J. Heidemann, “Scalable, ad hoc deployable RF-based localization,” in *Proc. The Grace Hopper Conference on Celebration of Women in Computing*, Oct. 2002.
- [16] N. Bulusu, J. Heidemann, and D. Estrin, “GPS-less low-cost outdoor localization for very small devices,” *IEEE Personal Communications*, vol. 7, no. 5, pp. 28–34, Oct. 2000.

- [17] C. Savarese, J. Rabaey, and J. Beutel, "Location in distributed ad-hoc wireless sensor networks," in *Proc. ICASSP*, vol. 4, Salt Lake City, UT, 2001, pp. 2037–2040.
- [18] P. Bergamo and G. Mazzini, "Localization in sensor networks with fading and mobility," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 2, Sept. 2002, pp. 750–754.
- [19] A. T. Parameswaran, M. I. Husain, and S. Upadhyaya, "Is RSSI a reliable parameter in sensor localization algorithms an experimental study," in *Proc. Field Failure Data Analysis Workshop*, Sept. 2009, pp. 471–478.
- [20] P. Krishnan, A. Krishnakumar, W.-H. Ju, C. Mallows, and S. Gamt, "A system for LEASE: location estimation assisted by stationary emitters for indoor RF wireless networks," in *Proc. INFOCOM*, vol. 2, Mar. 2004, pp. 1001–1011.
- [21] T. Roos, P. Myllymaki, and H. Tirri, "A statistical modeling approach to location estimation," *IEEE Trans. on Mobile Computing*, vol. 1, no. 1, pp. 59–69, Jan.-Mar. 2002.
- [22] E. Elnahrawy, X. Li, and R. Martin, "The limits of localization using signal strength: a comparative study," in *Proc. IEEE SECON*, Oct. 2004, pp. 406–414.
- [23] P. Bahl, V. Padmanabhan, and A. Balachandran, "Enhancements to the RADAR user location and tracking system," technical report, Microsoft Research, Tech. Rep., 2000.
- [24] A. M. Ladd, K. E. Bekris, A. Rudys, G. Marceau, L. E. Kavasaki, and D. S. Wallach, "Robotics-based location sensing using wireless Ethernet," in *Proc. International Conference on Mobile computing and networking*, ser. MobiCom '02. New York, NY, USA: ACM, 2002, pp. 227–238.

- [25] P. Bahl and V. Padmanabhan, “RADAR: an in-building rf-based user location and tracking system,” in *Proc. INFOCOM*, vol. 2, 2000, pp. 775–784.
- [26] M. Helén, J. Latvala, H. Ikonen, and J. Niittylahti, “Using calibration in RSSI-based location tracking system,” in *Proc. World Multiconference on Circuits, Systems, Communications and Computers*, 2001.
- [27] K. Cheung, H. So, W.-K. Ma, and Y. Chan, “Least squares algorithms for time-of-arrival-based mobile location,” *IEEE Trans. on Signal Processing*, vol. 52, no. 4, pp. 1121 – 1130, Apr. 2004.
- [28] F. Chan, H. So, J. Zheng, and K. Lui, “Best linear unbiased estimator approach for time-of-arrival based localisation,” *IET Signal Processing*, vol. 2, no. 2, pp. 156–162, June 2008.
- [29] M. Walworth and A. Mahajan, “3D position sensing using the difference in the time-of-flights from a wave source to various receivers,” in *Proc. International Conference on Advanced Robotics*, July 1997, pp. 611–616.
- [30] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, “The cricket location-support system,” in *Proc. International Conference on Mobile computing and networking*, ser. MobiCom ’00. New York, NY, USA: ACM, 2000, pp. 32–43.
- [31] A. Smith, H. Balakrishnan, M. Goraczko, and N. B. Priyantha, “Tracking moving devices with the Cricket Location System,” in *Proc. International Conference on Mobile Systems, Applications and Services*, Boston, MA, June 2004.
- [32] J. Sallai, G. Balogh, M. Maróti, Ákos Lédeczi, and B. Kusy, “Acoustic ranging in resource-constrained sensor networks,” in *Proc. ICWN 04, Las Vegas, NV*. CSREA Press, 2004, p. 04.

- [33] J. Zhang, T. Yan, J. A. Stankovi, and S. H. Son, “Thunder: towards practical, zero cost acoustic localization for outdoor wireless sensor networks,” *SIGMOBILE: Mobile Computing Communications Review*, vol. 11, no. 1, pp. 15–28, Jan. 2007.
- [34] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, 1st ed. Berlin, Germany: Springer, 2001.
- [35] F. Gustafsson and F. Gunnarsson, “Positioning using time-difference of arrival measurements,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, Apr. 2003, pp. 553–556.
- [36] J.-M. Valin, F. Michaud, J. Rouat, and D. Letourneau, “Robust sound source localization using a microphone array on a mobile robot,” in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, Oct. 2003, pp. 1228 – 1233.
- [37] S. Lauria, G. Bugmann, T. Kyriacou, and E. Klein, “Mobile robot programming using natural language,” *Robotics and Autonomous Systems*, vol. 38, no. 34, pp. 171 – 181, 2002.
- [38] J.-M. Valin, J. Rouat, and F. Michaud, “Enhanced robot audition based on microphone array source separation with post-filter,” in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, Sept.- Oct. 2004.
- [39] M. Maróti, B. Kusy, G. Simon, and A. Lédeczi, “The flooding time synchronization protocol,” in *Proc. International Conference on Embedded Networked Sensor Systems*, ser. SenSys '04. New York, NY, USA: ACM, 2004, pp. 39–49.
- [40] T. Machmer, J. Moragues, A. Swerdlow, L. Vergara, J. Gosálbez, and K. Kroschel, “Robust impulsive sound source localization by means of an energy detector for temporal alignment and pre-classification,” in *Proc. EURASIP*, 2009.

- [41] R. Morrissey, J. Ward, N. DiMarzio, S. Jarvis, and D. Moretti, “Passive acoustic detection and localization of sperm whales (*Physeter macrocephalus*) in the tongue of the ocean,” *Applied Acoustics*, vol. 67, no. 11, pp. 1091–1105, 2006.
- [42] Y. Kwon, K. Mechitov, S. Sundresh, W. Kim, and G. Agha, “Resilient localization for sensor networks in outdoor environments,” in *Proc. IEEE International Conference on Distributed Computing Systems*, June 2005, pp. 643–652.
- [43] M. van de Goor, “Indoor localization in wireless sensor networks,” Master’s thesis, Radboud Universiteit Nijmegen, the Netherlands, Mar. 2009.
- [44] M. Harnar, S. Brown, and H. Mayall, “Measuring the effect of distance education on the learning experience: Teaching accounting via picturetel.” *International Journal of Instructional Media*, vol. 27, no. 1, pp. 37–49, 2000.
- [45] H. Wang and P. Chu, “Voice source localization for automatic camera pointing system in videoconferencing,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1. IEEE, 1997, pp. 187–190.
- [46] K. Lay and W. Chao, “Mobile positioning based on toa/tsoa/tdoa measurements with nlos error reduction,” in *Proc. International Symposium on Intelligent Signal Processing and Communication Systems*. IEEE, 2005, pp. 545–548.
- [47] J. Abel and J. Smith, “The spherical interpolation method for closed-form passive source localization using range difference measurements,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 12. IEEE, 1987, pp. 471–474.
- [48] R. Schmidt, “A new approach to geometry of range difference location,” *IEEE Trans. on Aerospace and Electronic Systems*, no. 6, pp. 821–835, 1972.

- [49] H. So and S. Hui, “Constrained location algorithm using tdoa measurements,” *IEICE Trans. on fundamentals of electronics, communications and computer sciences*, vol. 86, no. 12, pp. 3291–3293, 2003.
- [50] K. Varma, “Time-delay-estimate based direction-of-arrival estimation for speech in reverberant environments,” Master’s thesis, Virginia Polytechnic Institute and State University, Oct. 2002.
- [51] J. H. DiBiase, “A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays,” Ph.D. dissertation, Brown University, May 2000.
- [52] T. S. N. U. V. Ramesh, *Speech Enhancement in Hearing Aids: with emphasis on Elko’s Beamformer*. LAP LAMBERT Academic Publishing, July 2012.
- [53] J. DiBiase, H. Silverman, and M. Brandstein, “Robust localization in reverberant rooms,” *Microphone Arrays*, pp. 157–180, 2001.
- [54] M. Brandstein, Editor, D. Ward, Editor, M. W. Hoffman, and Reviewer, “Microphone arrays,” *The Journal of the Acoustical Society of America*, vol. 112, no. 3, pp. 793–793, 2002.
- [55] G. Lathoud and M. Magimai-Doss, “A sector-based, frequency-domain approach to detection and localization of multiple speakers,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, Mar. 2005, pp. 265 – 268.
- [56] P. Pertilä, T. Korhonen, and A. Visa, “Measurement combination for acoustic source localization in a room environment,” *EURASIP Journal on Audio Speech Music Processing*, vol. 2008, pp. 3:1–3:14, Jan 2008.

- [57] S. M. Griebel, “A microphone array system for speech source localization denoising and dereverberation,” Ph.D. dissertation, Harvard University, 2002.
- [58] M. O. A. Brutti and P. Svaizer, “Oriented global coherence field for the estimation of the head orientation in smart rooms equipped with distributed microphone arrays,” in *Proc. The 9th European Conference on Speech Communication and Technology*, 2005, pp. 2337–2340.
- [59] H. Do, H. Silverman, and Y. Yu, “A real-time srp-phat source location implementation using stochastic region contraction(src) on a large-aperture microphone array,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, Apr. 2007, pp. 121–124.
- [60] F. Thomas and L. Ros, “Revisiting trilateration for robot localization,” *IEEE Trans. on Robotics*, vol. 21, no. 1, pp. 93–101, 2005.
- [61] “Wireless sensor networks.” Internet: <http://www.memsic.com/products/wireless-sensor-networks.html>,2010 [Sept. 6, 2012].
- [62] M. Jamel, P. Bertagna, and R. Davis Jr, “Footwear with GPS,” U.S. Patent 6788200, Sep. 7, 2004.
- [63] E. Foxlin, “Pedestrian tracking with shoe-mounted inertial sensors,” *Computer Graphics and Applications, IEEE*, vol. 25, no. 6, pp. 38–46, 2005.
- [64] K. Chadha, “The global positioning system: Challenges in bringing GPS to mainstream consumers,” in *Solid-State Circuits Conference, 1998. Digest of Technical Papers. 1998 IEEE International*. IEEE, 1998, pp. 26–28.
- [65] “IRIS,” datasheet, http://www.dinesgroup.org/projects/images/pdf_files/iris_datasheet.pdf,2011, Crossbow Technology, Inc.

- [66] “ATmega640/1280/1281/2560/2561,” datasheet, <http://www.atmel.com/Images/doc2549.pdf>,2012, Atmel Corporation.
- [67] “Lesson 8: Data logging application,” Online, <http://webs.cs.berkeley.edu/tos/api/tinyos-1.x/doc/tutorial/lesson8.html>,2002, Berkeley Wireless Embedded Systems.
- [68] “AT86RF230,” datasheet, <http://www.atmel.com/Images/doc5131.pdf>,2009, Atmel Corporation.
- [69] “MTS/MDA sensor board users manual,” datasheet, http://www.investigacion.frc.utn.edu.ar/sensores/Equipamiento/Wireless/MTS-MDA_Series_Users_Manual.pdf,2007, Crossbow Technology, Inc.
- [70] B. Yang and J. Scheuing, “Cramer-rao bound and optimum sensor array for source localization from time differences of arrival,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, march 2005, pp. 961–964.
- [71] D. Gislason, *Zigbee Wireless Networking*, 1st ed. Burlington, MA: Newnes, Aug. 2008.
- [72] H. Hashemi, “The indoor radio propagation channel,” *Proceedings of the IEEE*, vol. 81, no. 7, pp. 943–968, July 1993.
- [73] K. Brown Jr, “The theory of the gps composite clock,” in *Proc. The 4th International Technical Meeting of the Satellite Division of the Institute of Navigation*, vol. 1, 1991, pp. 223–241.
- [74] E. A. Dean, “Atmospheric effects on the speed of sound,” Atmospheric Science Laboratory, Tech. Rep. ASL-CR-79-0100-4, Aug. 1979.

- [75] I. Titze and D. Martin, “Principles of voice production,” *The Journal of the Acoustical Society of America*, vol. 104, p. 1148, 1998.
- [76] G. Lerosey, J. de Rosny, A. Tourin, and M. Fink, “Focusing beyond the diffraction limit with far-field time reversal,” *Science*, vol. 315, no. 5815, pp. 1120–1122, 2007.
- [77] M. Brandstein, J. Adcock, and H. Silverman, “A closed-form location estimator for use with room environment microphone arrays,” *IEEE Trans. on Speech and Audio Processing*, vol. 5, no. 1, pp. 45–50, 1997.
- [78] H. Schau and A. Robinson, “Passive source localization employing intersecting spherical surfaces from time-of-arrival differences,” *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 35, no. 8, pp. 1223–1225, 1987.
- [79] W. Hahn and S. Tretter, “Optimum processing for delay-vector estimation in passive signal arrays,” *IEEE Trans. on Information Theory*, vol. 19, no. 5, pp. 608–614, 1973.
- [80] B. Champagne, S. Bédard, and A. Stéphenne, “Performance of time-delay estimation in the presence of room reverberation,” *IEEE Trans. on Speech and Audio Processing*, vol. 4, no. 2, pp. 148–152, 1996.
- [81] T. Gustafsson, B. Rao, and M. Trivedi, “Source localization in reverberant environment: Modeling and statistical analysis,” *IEEE Trans. on Speech and Audio Processing*, vol. 11, no. 6, pp. 791–803, Nov. 2003.
- [82] J. Allen and D. Berkley, “Image method for efficiently simulating small-room acoustics,” *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

- [83] A. Nuttall, G. Carter, and E. Montavon, "Estimation of the two-dimensional spectrum of the space-time noise field for a sparse line array," *The Journal of the Acoustical Society of America*, vol. 55, no. 5, pp. 1034–1041, 1974.
- [84] L. Welch, "Lower bounds on the maximum cross correlation of signals (corresp.)," *IEEE Trans. on Information Theory*, vol. 20, no. 3, pp. 397–399, 1974.
- [85] K. Harker and D. Ilic, "Measurement of plasma wave spectral density from the cross-power density spectrum," *Review of Scientific Instruments*, vol. 45, no. 11, pp. 1315–1324, 1974.
- [86] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 24, no. 4, pp. 320–327, Aug. 1976.
- [87] P. R. Roth, "Effective measurements using digital signal analysis," *IEEE Spectrum*, vol. 8, no. 4, pp. 62–70, Apr. 1971.
- [88] G. Carter, A. Nuttall, and P. Cable, "The smoothed coherence transform," *Proceedings of the IEEE*, vol. 61, no. 10, pp. 1497–1498, Oct. 1973.
- [89] M. Brandstein and H. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, Apr. 1997, pp. 375–378.
- [90] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, 1st ed. Prentice Hall, Mar. 1993, vol. 1.
- [91] P. Aarabi, "The fusion of distributed microphone arrays for sound localization," *EURASIP Journal on Applied Signal Processing*, pp. 338–347, 2003.

- [92] M. O. A. Brutti and P. Svaizer, “Multiple source localization based on acoustic map de-emphasis,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2010, 2010.
- [93] A. Brutti, M. Omologo, and P. Svaizer, “Localization of multiple speakers based on a two step acoustic map analysis,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Apr. 2008, pp. 4349–4352.
- [94] H. Do and H. Silverman, “A fast microphone array srp-phat source location implementation using coarse-to-fine region contraction (CFRC),” in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE, 2007, pp. 295–298.
- [95] J. Dmochowski, J. Benesty, and S. Affes, “A generalized steered response power method for computationally viable source localization,” *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 15, no. 8, pp. 2510–2526, 2007.
- [96] R. Ricardo, U. Helmke, and S. Yoshizawa, “Newton’s method for optimization,” internet, <http://europepmc.org/abstract/CIT/573074>.
- [97] C. Shannon, “Communication in the presence of noise,” *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.
- [98] H. Kuttruff, *Room Acoustics*. New York, NY: Wiley, 1973.
- [99] E. Gilbert, “Capacity of a burst-noise channel,” *Bell System Technical Journal*, vol. 39, pp. 1253–1265, 1960.
- [100] E. Elliott, “Estimates of error rates for codes on burst-noise channels,” *Bell System Technical Journal*, vol. 42, pp. 1977–1997, 1963.
- [101] J. Norris, *Markov chains*. Cambridge University Press, 1998.

- [102] J. Bendat and A. Pierso, *Random Data: Analysis and measurement procedures*. New York, N.Y.: Wiley, 1971.
- [103] M. Maróti and J. Sallai, “Packet-level time synchronization,” internet, <http://www.tinyos.net/tinyos-2.1.0/doc/html/tep133.html>, TinyOS Core Working Group, 2008.