Non-Field-of-View Acoustic Target Estimation

"*Hearing is Believing*"

Kuya Takami

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Mechanical Engineering

Tomonari Furukawa (Chair)
Ricardo Burdisso
Alexander Leonessa
Robert Parker
Daniel Stilwell

September 18, 2015
Blacksburg, Virginia

Keywords: non-field-of-view localization, acoustic target estimation, grid-based, recursive Bayesian estimation

Non-Field-of-View Acoustic Target Estimation

Kuya Takami

Abstract

This dissertation proposes a new framework to Non-field-of-view (NFOV) sound source localization and tracking in indoor environments. The approach takes advantage of sound signal information to localize target position through auditory sensors, combination with other sensors within grid-based recursive estimation structure for tracking using nonlinear and non-Gaussian observations.

Three approaches to NFOV target localization are investigated. These techniques estimate target positions within the Recursive Bayesian estimation (RBE) framework. The first proposed technique uses a numerical fingerprinting solution based on acoustic cues of a fixed microphone array in a complex indoor environment. The Interaural level differences (ILDs) of microphone pair from a given environment are constructed as an *a priori* database, and used for calculating the observation likelihood during estimation. The approach was validated in a parametrically controlled testing environment, and followed by real environment validations. The second approach takes advantage of acoustic sensors in combination with an optical sensor to assist target estimation in NFOV conditions. This hybrid of the two sensors constructs observation likelihood through sensor fusion. The third proposed model-based technique localizes the target by taking advantage of wave propagation physics: the properties of sound diffraction and reflection. This approach allows target localization without an *a priori* knowledge database which is required for the first two proposed techniques.

To demonstrate the localization performance of the proposed approach, a series of parameterized numerical and experimental studies were conducted. The validity of the formulation and applicability to the actual environment were confirmed.

Dedication

To my parents,

Nagako and Shuichi Takami *Sheila and John Schulz*

Acknowledgments

To begin, I thank Professor Furukawa Tomonari for his continuous guidance and teaching the way of "research" throughout my Ph.D. study. It was not only academically, but also physically by training and running several half/ full marathons together. I am grateful to his family for their warmth and bringing their energetic kids' power to the lab every weekends.

For guidance, I must thank Dr. Kumon, Dr. Dissanayake and my committee members, Dr. Burdisso, Dr. Leonessa, Dr. Parker, and Dr. Stilwell, for their professional and critical comments, providing additional perspective to my research.

I must thank the CMS Lab members for great friendship and excited discussions on many topics, not just in research. Members include, Kun Jin, Xianqiao, Bonson, Kim, Boren, Varun, Howard, Affan, Luan, Rich, Chris, Tian, Yazhe, Orson, Hangxin, Jason, Murat, and Yoon.

I would like to thank my family for their continuous support and encouragement. My parents, Nagako & Shuichi and Sheila & John, proved me with the opportunities and wise advice which allowed me to this accomplishment. My siblings, Sukoyaka & Diego, Mumu, Kaosu, Raki, Kyle, Kalia, Alaina, Galen and Riley gave me the strength to pursue my goals and filled my life with laughter.

Finally, I would like to thank all the great friends around the world, Berman family, Justin, Juliana, Louie, Moon, Myra & Greg, Yosuke, Stan, UW-Madison/EC-friends, VT slackliners, and special thanks to Cate for correcting my papers.

Contents

List of Figures

List of Tables

Abbreviations

2D	Two-dimensional.
3D	Three-dimensional.
ADC	Analogue-to-digital converter.
AoA	Angle of arrival.
CCD	Charge-coupled device.
CMOS	Complementary metal-oxide-semiconductor.
DAQ	Data acquisition.
DSLR	Digital single-lens reflex.
FFT	Fast Fourier transform.
FOV	Field-of-view.
FPGA	Field programmable gate array.
GCC	Generalized cross-correlation.

GPS Global positioning System.

HRI Human-robot interaction.

I/O Input/output.

ILD Interaural level difference.

IMU Inertial measurement unit.

IPD Interaural phase difference.

ITD Interaural time difference.

KL Kullback-Leibler.

LOS Line-of-sight.

mic Microphone.

NFOV Non-field-of-view.

NLOS Non-line-of-sight.

PC Personal computer.

PDF Probability density function.

RBE Recursive Bayesian estimation.

RF Radio-frequency.

RMS Root mean squared.

RSS Received signal strength.

SAR Search-and-rescue.

SLAM Simultaneous localization and mapping.

SNR Signal-to-noise ratio.

TDOA Time-difference-of-arrival.

TOA Time-of-arrival.

ToF Time of flight.

USB Universal serial bus.

WGN White Gaussian noise.

WSN Wireless sensor network.

Nomenclature

$\tilde{(\cdot)}$	: an instance of $(\cdot)$.
$(\cdot)^\top$	: transpose of $(\cdot)$.
$f((\cdot))$	: function of $(\cdot)$.
$p(\alpha\|\beta)$	: probability of α given condition β.
$l((\cdot))$	: likelihood of $(\cdot)$.
ω	: a frequency of sound.
$\mathbf{\Omega}$	: a frequency bin vector, $[\omega_1, ..., \omega_N]^\top$.
k	: time index.
$\mathbf{x}_k^\alpha$	: state of the α at time step k.

CHAPTER 1

Introduction

Our environment is surrounded by acoustics or sound science. It is a very broad discipline and has been continuously studied throughout human history. The sound is transmitted as vibration through mediums such as air or water. It can be used as a means of communication by imparting or exchanging information. Perception of multiple simultaneous sound sources can be complex. In addition, sound generation and propagation further increase complexity.

This dissertation presents a comprehensive study of indoor localization framework outside of the optical Field-of-view (FOV), or Non-field-of-view (NFOV). NFOV defines the region beyond the extent of the observable world by optical sensors. This NFOV localization takes advantage of sound information to localize target position through auditory sensors which contain a wealth of information

beyond FOV.

1.1 Background

A robot makes an action based on its perception. The capability of mobile robot perception has been increasing with the introduction of new sensing technologies and algorithms [3, 4, 5]. One of the main roles of perception is to estimate the state of a subject of interest, a target, where the target could be human or other mobile robots. Target localization and tracking, or target estimation, within an FOV/ NFOV target is performed by humans and animals that use sound. Similar to vision, audition has multi-functional capabilities. Auditory sensors are used not only as a means of communication, but also as a target estimation. These abilities motivated researchers to mimic similar performance using auditory sensors, primarily Microphones (mics) combined with algorithms.

With the continuous effort by electrical engineering and digital signal processing, robot audition has been successful in the area of speech recognition, classification and source localization. Especially the recent development of humanoid and Human-robot interaction (HRI) has led to extensive development and implementation. However, the auditory target estimation capability of humans has not been sufficiently investigated. This is partly because the vision and other optical sensors have all the data necessary and superbly achieve target estimation when the target is in the sensor's FOV. It is also partly because the target outside the FOV can be captured inside the FOV by robot movement. However, the audition could provide the information outside the vision or optical sensors. Figure 3.3 illustrates this issue. If the robot relies entirely on vision or optical sensing, the robot will never be able to achieve target estimation efficiently for this example. Indoor environments are so

constrained by structures that the FOV in indoor environments is significantly limited, leaving most of the field as NFOV. Audition thus becomes the key technology for target estimation in complex indoor environments.

1.2 Objectives

The primary objectives of this dissertation are as follows:

- To propose a localization technique that can localize NFOV sound targets
- To develop a testing system capable of parametrically and quantitatively studying the performance of proposed techniques

1.3 NFOV Acoustic Estimation

In order to achieve these objectives, novel localization techniques and their systems are developed. The first proposed technique uses a fingerprint technique to construct the location-dependent database from a microphone array. Following the database collection, the NFOV acoustic target location is estimated by calculating the observation likelihood.

The second approach takes a hybrid of visual/auditory information within the Recursive Bayesian estimation (RBE) framework. Auditory sensors are used to assist visual sensors when the event of "no detection" by an optical sensor is converted into an observation likelihood and utilized to positively update probabilistic belief on the target that is dynamically maintained by the RBE. Acoustic observation likelihoods are constructed based on *a priori* acoustic cues in the given environment.

The last proposed technique focuses on wave propagation properties based on

sound physics by extracting the first-arrival diffraction and reflection signals. Unlike radio signals, sound signals reflect significantly without penetrating into different media, while they also diffract at low frequencies. The proposed approach takes signal characteristics of first-arrival diffraction and reflection signals obtaining a time-domain and a frequency domain signal of a relatively impulsive sound at each microphone. The diffraction signal is used to identify the so-called diffraction point by deriving the Time-difference-of-arrival (TDOA) for each pair of microphones and further estimating the direction of target sound beyond the diffraction point from the loss of sound energy through diffraction. An observation likelihood is eventually calculated by additionally estimating the distance from the sound's magnitude and characteristics. The reflection signal estimates the target direction directly from the TDOA by mirroring and creating a virtual target. It also creates an observation likelihood with the distance estimate by considering the sound's magnitude and characteristics and the environmental properties. A joint observation likelihood is finally created by the fusion of the diffraction and reflection observation likelihoods.

1.4 Principal Contributions

The technical contributions of this thesis are summarized as follows:

- A formulation of localization within the RBE framework
- A probabilistic formulation of acoustic observation likelihood
- A new approach and formulation of Non-line-of-sight (NLOS) target localization using optical and acoustic sensor information
- The application of NLOS target localization based on fixed microphones

- The application of NLOS target localization using a camera and a stereo microphone pair on a mobile robot
- A formulation of model-based NFOV acoustic localization
- Developed an NLOS acoustic localization parametric study experimental environment

1.5 Publications

The work presented in this thesis has been described in the following publications:

1.5.1 Journal Articles:

[1] Jean M Whitney, **Takami, Kuya**, Scott T Sanders, and Yasuhiro Okura. Design of system for rugged, low-noise fiber-optic access to high-temperature, high-pressure environments. *Sensors Journal, IEEE*, 11(12):3295–3302, 2011

[2] Xinliang An, Thilo Kraetschmer, **Takami, Kuya**, Scott T Sanders, Lin Ma, Weiwei Cai, Xuesong Li, Sukesh Roy, and James R Gord. Validation of temperature imaging by h 2 o absorption spectroscopy using hyperspectral tomography in controlled experiments. *Applied optics*, 50(4):A29–A37, 2011

[3] **Takami, Kuya**, Tomonari Furukawa, Makoto Kumon, Daisuke Kimoto, and Gamini Dissanayake. Estimation of a nonvisible field-of-view mobile target incorporating optical and acoustic sensors. *Autonomous Robots*, pages 1–17, 2015

[4] **Takami, Kuya**, Tomonari Furukawa, Makoto Kumon, and Gamini Dissanayake. Non-field-of-view acoustic target estimation in complex indoor environment. *Springer Tracts in Advanced Robotics*, 2015

[5] **Takami, Kuya**, Tomonari Furukawa, and Daniel Watman. High-resolution measurement system for rotating deformable body. *Journal of Visualization*, 2015. Submitted

[6] **Takami, Kuya**, Tomonari Furukawa, Makoto Kumon, and Lin Chi Mak. Non-field-of-view acoustic target estimation using reflection and diffraction. In preparation

1.5.2 Conference Articles:

[1] Makoto Kumon, Daisuke Kimoto, **Takami, Kaya**, and Tomonari Furukawa. Acoustic recursive bayesian estimation for non-field-of-view targets. In *Image Analysis for Multimedia Interactive Services (WIAMIS), 2013 14th International Workshop on*, pages 1–4. IEEE, 2013

[2] Makoto Kumon, Daisuke Kimoto, **Takami, Kuya**, and Tomonari Furukawa. Bayesian non-field-of-view target estimation incorporating an acoustic sensor. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 3425–3432. IEEE, 2013

[3] **Takami, Kuya**, Saied Taheri, Mehdi Taheri, and Tomonari Furukawa. Prediction of railroad track foundation defects using wavelets. In *2013 Joint Rail Conference*, pages V001T01A008–V001T01A008. American Society of Mechanical Engineers, 2013

[4] **Takami, Kuya**, Tomonari Furukawa, Makoto Kumon, and Lin Chi Mak. Non-field-of-view indoor sound source localization based on reflection and diffraction. 2015

[5] **Takami, Kuya**, Tomonari Furukawa, Makoto Kumon, and Gamini Dissanayake. Non-field-of-view acoustic target estimation in complex indoor environment. 2015

[6] Tomonari Furukawa, **Takami, Kuya**, Xianqiao Tong, Daniel Watman, Abbi Hamed, Ravindra Ranasinghe, and Gamini Dissanayake. Map-based navigation of a autonomous car using grid-based scan-to-map matching. 2015

[7] **Takami, Kuya** and Tomonari Furukawa. High-resolution deformation measurement system for fast rotating tires towards noise prediction. 2015

[8] **Takami, Kuya** and Tomonari Furukawa. Study of tire noise characteristics with high-resolution synchronous images. 2015

1.6 Organizations

- Chapter 2 surveys past work related to the development of localization techniques. The main technologies and approaches for indoor localizations are classified. Further, the prominent studies that focus on NLOS acoustic target localization are summarized.

- Chapter 3 describes detail formulation of grid-based RBE and derivation of observation likelihood.

- Chapter 4 presents the NLOS localization technique including formulation, system design, implementation, and performance evaluation.

- Chapter 7 summarizes the major conclusion of the dissertation and discusses areas which would benefit from future work.

CHAPTER 2

Literature review

2.1 Introduction

This chapter presents a review of the primary approaches to localization with emphasis on past contributions to indoor environment localization. The investigation leads to theoretical, computational and experimental approaches in conjunction with their theory, methods, and techniques. This includes different technologies, Radio-frequency (RF), optics, and acoustics, with common location dependent distance-based and fingerprinting-based metrics. Additional focus is placed especially on Global positioning System (GPS)-denied complex conditions. This investigation of state-of-the-art techniques unearths an unexplored capability of NFOV acoustic localization.

2.2 Mobile Target Localization

In the field of mobile robot localization, TDOA-measurement-based GPS technology is often used in estimation for most of the outdoor environments. However, it has limited or no localization capabilities indoors. This leads to the development of a large variety of technologies and approaches which extend the localization capabilities for indoor scenarios. One example of taxonomy summarizing available technologies that allow localization of targets in GPS denied environments are shown in Fig. 2.1. Especially with the increasing development and ubiquitousness of mobile devices and location-aware information services, together with easier access to inexpensive sensors, have exponentially increased. The needs and potentials to indoor localization.

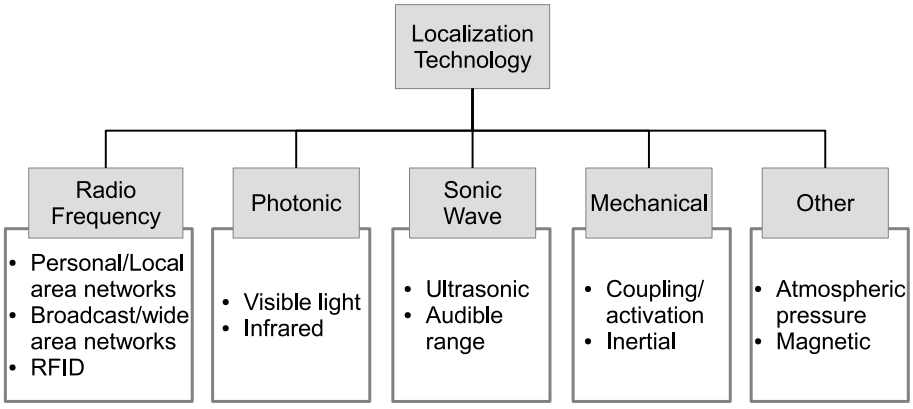


Figure 2.1: Taxonomy of NFOV localization technologies adapted from [1]

While different sensor modalities lead to their own advantages and disadvantages, methods can be classified based on the location-dependent metrics. For the parametric indoor localization, geometry-based, fingerprinting-based and proximity measurements are used. Geometry-based measurement techniques most commonly include triangulation from measurements and trilateration. On the other hand, the fingerprinting-based approach mainly uses Received signal strength (RSS) or

frequency profiling techniques. The geometric properties are used to estimate the target using RSS,Time-of-arrival (TOA), and Angle of arrival (AoA) in most Line-of-sight (LOS) conditions [20, 21]. Contrary to those methods, fingerprinting methods use the complexity of the environment to their advantage, exploiting the signal distinct patterns caused by the effect of the signal property [22, 23, 24, 25]. Additional factors such as computational cost, required infrastructure, accuracy, dynamic environment, and privacy-oriented requirements contribute to the selection of these methods.

2.3 Indoor Localization

2.3.1 Related Past Work and Current Limitations

The two main types of waves used for localization are electromagnetic and mechanical vibration waves, radio wave, and acoustic wave, respectively. While both the signals experience the phenomena of absorption, reflection, diffraction and refraction, these signals are transverse and longitudinal waves, respectively, and each exhibit unique characteristics. The first major difference is the speed of the wave, where electromagnetic waves travel at the speed of light and acoustic waves propagate relatively slower, requiring a transmitting medium. Another different characteristic to be noted is the signal behavior. Electromagnetic signals have a higher penetrating power than non-metallic materials and a lower path loss, so they transmit over a large area by showing all the phenomena of absorption, reflection, diffraction and refraction [26, 27, 28]. Sound signals, on the other hand, have less capability in penetration, and this results in more reflections. With reference to these characteristics, past work conducted for a target estimation is divided into the following three main

techniques.

- **Model-based techniques**

  Wave physical properties such as propagation speed and distance path loss are incorporated into the geometrical models based localization. Main techniques calculate location from power-distance mapping, TOA, TDOA, and AoA. This scheme does not require *a priori* database; however, it is at the cost of decreasing localization accuracy.

- **Fingerprinting-based techniques**

  The principle of a fingerprint matching scheme is commonly used for location determination and thrives in complex environments. Generally, this technique requires a survey of every location of interest to build the location specific surrounding signature database as *a priori*. The location is then estimated by mapping the measured fingerprint to the database.

- **Proximity-based technique**

  The known anchor node devices with limited range are distributed in an environment, and an unknown target is localized by finding the nearest anchor node. This technique requires short range and dense deployment of anchor nodes because the accuracy is limited by the size of the node cells.

The table 2.1 shows a few examples of systems or approaches made by researchers in past couple decades. It is categorized into radio wave and acoustic wave approaches as well as three different technique categories.

Table 2.1: Indoor localization category based on wave type and techniques

	Radio wave		Acoustic wave	
Category	System	Principle	System	Principle
Geometric	Active Badge		Knapp and Carter [29]-1976	GCC
	UWB [30]	TOA	Scott and Dragovic [31]-2005	TOA
	RSP [32]	DOA	Active Bat [33]-1997	Ultrasonic TOA
			Cricket [34]-2000	RF, Ultrasonic, T/TDOA
Fingerprinting	RADAR [35]	WLAN(RSS), kNN	Tarzia et al. [36]	ABS
	LANDMARC [37]	Active RFID, kNN	SoundSense [38]	SL
	GSM fingerprinting [39]	GSM,RSS, Weighted kNN	Aloui et al. [40]	TOA
Proximity	SAL [41]-2010	N/A		
	Bekkelien [42]	Bluetooth		
	Song et al. [43]-2007	RFID		

2.3.2 Geometric: Model-based

The first approach utilizes the principle of the LOS target estimation by multiple sensor anchor nodes. When each pair of sensors derives a difference in some measured quantity, the quantity is used to estimate the distances or angles from the target to the anchor nodes. The accumulation and comparison of these quantities can geometrically estimate the target location. If the quantity is distance-based or angle-based, the technique is classified as trilateration, and triangulation, respectively. The use of a few sensors with the LOS assumption will create a large estimation error known as the NLOS error, but the increased deployment of a number of sensors reduces the error and gives an improved estimate. Both sound and radio signals have been used within the framework.

In the use of sound signals, the TDOA is the most appropriate quantity for dynamic target tracking because of the slow speed of sound [44]. Acoustic TDOA-based target estimation by a set of microphones can be found in [45, 46, 47, 48, 49] where the advantage of the acoustic TDOA-based approach is high accuracy. One of the two key issues of the approach is the accurate determination of TDOA, and the Generalized cross-correlation (GCC) proposed by Knapp and Carter [29] has been predominantly used in the last four decades. Another key issue is the choice of the technique to map the time measurement to estimated sound source positions in LOS conditions. The most common and computationally simple method is the least squares method [50], despite its sensitivity to background noise. A variety of the alternatives have been proposed to improve the reliability [51, 52, 53, 48].

Similarly to sound signals, radio signals can also be used to localize a NFOV target by processing received signals with the LOS assumption [54, 55, 37, 56, 57, 58, 59]. Model-based location techniques have been widely used in wireless local

area networks [27, 60], cellular networks [61, 62, 63] and ultra-wideband radio systems [64, 65, 26]. Out of these, ultra-wideband radio systems can achieve the most accurate and reliable radiolocation due to the extraordinarily wide bandwidth of the ultra-wideband radio signals [64, 65, 66, 26, 67]. A variety of approaches to improve positioning accuracy have been further proposed. Denis et al. [68] compared the first-arrival and strongest signals in LOS conditions and showed the superiority of the first-arrival signals. Study by Lee and Yoo [69] indicated the importance of the threshold value, and various techniques have been proposed for the threshold selection [70, 71, 72, 73], though the NLOS error cannot be essentially eliminated. Identification of NLOS base stations has thus been further explored by several researchers [74, 75, 76, 77, 78, 79], and all the techniques were able to identify the NLOS base stations. The identification of the NLOS base stations has mitigated the negative NLOS effect [78, 76, 77], but the removal of NLOS base stations does not improve the accuracy. Various geometrical models were developed to estimate the NLOS error [80, 81, 62, 82, 83, 84, 85, 61, 86, 63, 87, 88]. A variety of hybrid approaches [89, 90, 91, 92, 93, 94, 95] that use the angle of arrival in conjunction with the TDOA have also been proposed to correct the NLOS error. In addition to these hybrid approaches, there are other hybrid systems, Active Bat [33] and Cricket [34] , that use two wave types, radio and acoustic waves. The use of the difference in RF and ultrasound signal propagation times.

Further, Sasaki et al. [96] and Valin et al. [97, 98, 53, 47] used a beamformer whereas DiBiase [99], DiBiase et al. [100] combined a beamformer and GCC for more accurate target estimation. The experimental results in the work by Silverman et al. [46] prove that the steered response power GCC achieves 16 to 52 mm positioning accuracies in LOS conditions. The computational cost of the steered response power

GCC was significantly reduced by utilization of stochastic region contraction [45].

2.3.3 Feature: Fingerprinting-based

As an alternative to the model-based geometric techniques, fingerprinting-based estimation techniques were effectively utilized for target localization. This technique was widely used with radio signals [37, 101, 102]. Whilst this arrangement could achieve higher accuracy, the critical problem inherent in the RF based approach is its applicability only to near-NFOV target estimation due to the limited capability of radio signals in capturing information on a NFOV target [103, 104, 93, 105]. The fingerprinting, in the case of the radio signal, is the RSS. Similarly to some of the acoustic techniques, the target position can be estimated by matching the measured RSS and the stored RSSs at different positions [106, 35, 107, 108, 109]. Various attempts have been made to further to improve the performance of the approach. Some of them reduced the number of RSSs to be stored [35, 110, 111, 112] since the main drawback of the approach is the time-consuming acquisition of prior measurements [113, 114, 115]. Some others proposed empirical techniques that can more accurately locate the target [24, 116, 108, 117, 118]. Recently, Yang et al. [119] proposed fingerprinting with little human intervention to minimize the effort of database construction. Ferris et al. [107] applied Gaussian processes for RSS-based localization and Gaussian processes latent variable models [113] for WiFi-SLAM. Another challenge in the radio signal approach is its unreliability due to RSS data with disturbances from dynamic temperature, humidity and obstacles [110, 120, 121, 60, 122, 123, 124]. Several noise reduction techniques, including probabilistic filter [125, 113, 126, 116, 108] and singular value decomposition [126] have been used to improve the localization accuracy. Jourdan et al. [127], Jourdan

[128], Evennou and Marx [129], Lee and Mase [130] and You et al. [124] used Inertial measurement units (IMUs) in conjunction with RSS for localization.

With the use of sound waves for fingerprinting-based technique, sound features associated with artificial pinnae were used to determine the angle of arrival of a sound source. The studies by Hofman et al. [23] experimentally proved that pinnae play an important role in the human auditory system for sound localization. Their experimental results show that humans can learn and adapt to modified pinnae easily. Several researchers have achieved sound localization using microphones and pinnae, in which neural networks by Takanishi et al. [131] and spectral cues [132, 133] caused by reflection on the pinnae have both been employed to determine the angle of arrival. Kumon et al. [132] and Shimoda et al. [133] successfully estimated the direction of a white noise sound source in pitch and roll angles. Likewise, Saxena and Ng [134] employed a machine learning method for monaural localization using a pinna and localized various sounds on a horizontal plane with different pinna shapes. The main advantage of these techniques is the reduced number of microphones required in their systems due to the additional directional information extracted from the sound features associated with pinnae.

Acoustic NLOS target estimation can be localized by spectral cues. The approach creates and collects a unique power signature at each location in advance. The target location can then be estimated by matching the measured signature with those in a database. Wu et al. [135] utilized a Gaussian mixture model to model the phase difference and magnitude ratio between the signals received by a pair of microphones for localization in both LOS and NLOS conditions. Hu et al. [136] extended the work by Wu et al. [135] through the exploitation of two speakers on an unmanned ground vehicle. The limitation of these techniques is their prerequisite

of a time-consuming collection of training data to construct the database.

2.3.4 Proximity-based

Both the geometry-based and feature-based techniques can be utilized with sound and radio signals for NLOS target estimation. The bottleneck of the LOS estimation approach by a sensor network is that the NLOS cannot be completely eliminated by the LOS assumption and without considering NLOS signal propagation.

Chawathe [137] used a limited short range of Bluetooth beacons to indoor localization. Their focus was to reduce the time to compute the localization latency with prior knowledge of location graphs and a beacon location map. Although the modeling of this algorithm is sound, a validation in a real complex environment is still required.

Aksu and Krishnamurthy [41] developed a calibration-free system that localizes based on the proximity-based technique. Because it is a discrete binary operation if the target is within the range of given anchor nodes, implementation is simple and computationally less expensive compared to the fine-grained localizations described in above sections. However, the disadvantage of this course-grain calibration-free method results in low accuracy.

2.3.5 Other Techniques

Unlike the aforementioned LOS estimation by a sensor network, the use of TOA signals, which could contain information on the NLOS target, can estimate the target location without NLOS errors [44, 138]. Mak and Furukawa [139] considered the diffraction characteristics of low-frequency sound in addition to reflection characteristics in target estimation. This enables the truly NFOV target estimation owing

to the use of reflected and diffracted signals based on two assumptions. Firstly, the sound is assumed to be controlled synchronously. Secondly, the map of the environment must be known *a priori*. Both the assumptions are impractical since the robot cannot control the sound of humans and may not know the surrounding environment at all.

The majority of sound localization challenges have been focused on the direction of sound rather than its position due to the complexity of sound wave propagation [140, 22, 141]. Recently, acoustic physical models are used to improve the localization capability by investigating the reflection of sound in a known environment. Narang et al. [142] detected reflected sound by a combination of an image model and dynamic environment map. This approach estimates the direction of the sound source, however, it does not estimate the target position in NFOV and does not fuse the vision information for localization. Even et al. [143] localized the sound source by a ray tracing method based on the reflection signal arrival directions. However, their approach resulted in NFOV estimation with meter order accuracy and some inconsistent variation in trials.

The past effort for NFOV target estimation has been reviewed. Although each approach exhibits advantages, none of the approaches, as a sole technique, enables true NFOV target estimation in real-world scenarios. This reconfirms the need for developing a new approach to estimating the NFOV target by using the physics of sound wave propagation associated with the NFOV target.

2.3.6 Applications

Fundamental research into auditory NFOV target estimation proposed in this project could open new possibilities and lead to various applications. The most challenging

but important application would be the Search-and-rescue (SAR) in urban disaster scenarios where victims are trapped in non-visible highly constrained environments. If victims are alive, a mobile robot with the proposed auditory approach will allow their locations to be identified simply by communicating with them and approaching them. Since their locations can be identified before the debris is removed, debris can be removed safely after identifying the victim locations.

Another important application is home healthcare and rehabilitation where an elderly person or a disabled person is to be serviced for daily life and continuously monitored for an emergency at home. With microphones fixed to walls and ceilings as part of the home infrastructure, their movement can be monitored by the sound they create such as footstep or sounds from the speaker they carry. A service robot, if available, could be used further to locate the target person without the microphone infrastructure or to provide services to the target person. Environments with privacy such as personal homes are not supposed to be equipped and monitored with cameras, so that localization with no camera becomes an indispensable solution no matter whether it is with a fixed infrastructure or with a mobile robot.

The acoustic target estimation can be also used for indoor security and maintenance. With microphones fixed to walls and ceilings as part of the building infrastructure, the movement of an intruder can be monitored again by the sound of footstep. The existence of vermin can also be detected. Microphones are much cheaper than cameras and can be sparsely installed due to their NFOV target estimation capability. Since complex indoor structures introduce numerous blind spots while the area to cover could be large, the acoustic NFOV target estimation can be considered as the best solution.

2.4 Chapter Summary

This chapter has presented the extensive efforts of engineers and researchers in their state-of-the-art investigation to localization-particularly those work on GPS denied indoor environments. These efforts were classified based on the technologies used and model-based/fingerprinting-based approaches. Most of the localization approaches focused on and developed for LOS cases where the target needed to be equipped with a transmitter. With the increase of mobile robot applications and the continuous improvement in technology, the true NFOV localization approaches are receiving more interest.

The results of this literature review have identified the need for an efficient approach to NFOV acoustic localization in GPS-denied environments. The objectives of this dissertation aim to address these issues with the current unexplored localization capability in NFOV conditions.

CHAPTER 3

Grid-based Recursice Bayesian Estimation with Acoustic Observation

This chapter describes the detailed formulation of grid-based RBE framework. The motion and sensor models are derived for both optical and acoustic sensors. Definition of the "non-detected" region for the optical sensor and fusion of the sensor information are also explained. The NFOV highly nonlinear and non-Gaussian process shows the importance of this generalized approach.

3.1 Recursive Bayesian Estimation

Recursive Bayesian estimation is a general probabilistic approach, which recursively estimates non-linear non-Gaussian models. For the successful RBE, target and sensor models are required.

3.1.1 Target Motion Models and Sensor Models

Within RBE framework, accurate target and sensor models are fundamental. Consider a target, t, of concern, the motion of which is discretely given by

$$\mathbf{x}_{k+1}^{t} = \mathbf{f}^{t}\left(\mathbf{x}_{k}^{t}, \mathbf{u}_{k}^{t}, \mathbf{w}_{k}^{t}\right) \tag{3.1}$$

where $\mathbf{x}_{k}^{\alpha} \in \mathcal{X}^{t}$ is the state of the target at time step k, $\mathbf{u}_{k}^{t} \in \mathcal{U}^{t}$ is the set of control inputs of the target, and $\mathbf{w}_{k}^{t} \in \mathcal{W}^{t}$ is the "system noise" of the target including environmental influences. For simplicity, the target state describes the two-dimensional position.

Similarly, sensor platform, s, is described, assuming the states are accurately known by use of sensors such as IMU and/or laser range scanners. The sensor platform at time k with state $\mathbf{x}_{k}^{s} \in \mathcal{X}^{s}$ and control inputs $\mathbf{u}_{k}^{s} \in \mathcal{U}^{s}$ is therefore given by

$$\mathbf{x}_{k+1}^{s} = \mathbf{f}^{s}\left(\mathbf{x}_{k}^{s}, \mathbf{u}_{k}^{s}, \mathbf{w}_{k}^{s}\right) \tag{3.2}$$

As a part of the indoor infrastructure, let this moving target t be observed by a fixed optical sensor such as a surveillance camera s_c where the global state of the optical sensor is assumed to be accurately known as $\tilde{\mathbf{x}}^{s} \in \mathcal{X}^{s}$. Note that $\tilde{(\cdot)}$ is an instance of $()$. The FOV, or more precisely the "observable region", of the optical

sensor can be expressed with the probability of detecting the target $P_d\left(\mathbf{x}_k^t|\tilde{\mathbf{x}}^s\right)$ as

$$ {}^{s_c}\mathcal{X}_o^t = \left\{\mathbf{x}_k^t|0 < P_d\left(\mathbf{x}_k^t|\tilde{\mathbf{x}}^s\right) \leq 1\right\}. $$

Accordingly, the target position observed from the optical sensor, ${}^{s_c}\mathbf{z}_k^t \in \mathcal{X}^t$, is given by

$$ {}^{s_c}\mathbf{z}_k^t = \begin{cases} {}^{s_c}\mathbf{h}^t\left(\mathbf{x}_k^t, \tilde{\mathbf{x}}^s, {}^{s_c}\mathbf{v}_k^t\right) & \mathbf{x}_k^t \in {}^{s_c}\mathcal{X}_o^t \\ \varnothing & \mathbf{x}_k^t \notin {}^{s_c}\mathcal{X}_o^t \end{cases} \tag{3.3} $$

where ${}^{s_c}\mathbf{v}_k^t$ represents the observation noise, and $\varnothing$ represents an "empty element", indicating that the optical observation contains no information about the target or that the target is unobservable when it is not within the observable region.

The acoustic sensor, such as a microphone, can, on the other hand, observe a target on the NLOS or even in the NFOV, though accuracy is limited due to the complex behavior of sound signals including reflection, refraction and diffraction. Because of its broad range, the observable region of the acoustic sensor could be considered unlimited when compared to that of the optical sensor. The acoustic sensor model s_a can be therefore constructed without defining an observable region, unlike the optical sensor model:

$$ {}^{s_a}\mathbf{z}_k^t = {}^{s_a}\mathbf{h}^t\left(\mathbf{x}_k^t, \tilde{\mathbf{x}}_k^s, {}^{s_a}\mathbf{v}_k^t\right), \tag{3.4} $$

3.1.2 Recursive Bayesian Estimation

Figure 3.1 shows the probabilistic information flow of a recursive Bayesian process. The RBE updates the belief on a dynamical system, given by a probability density, in both time and observation. It maintains the posterior distribution and prediction

Probability density functions (PDFs). The posterior distribution of the current state is calculated from a prior estimation of the state and a new observation at the present time in the update stage. The prediction stage then calculates the PDF of the next state, using the posterior density of the current state and probabilistic Markov motion model.

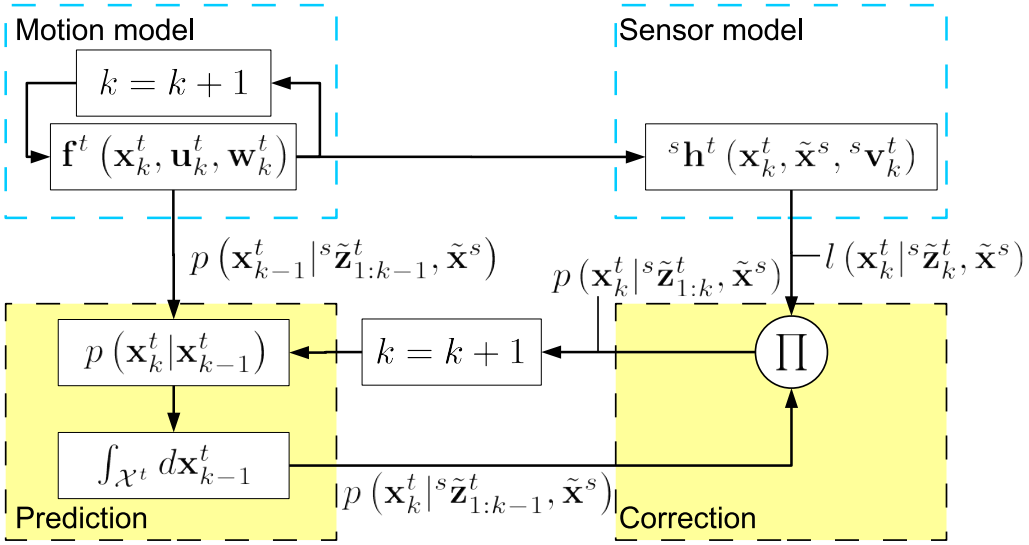


Figure 3.1: Recursive Bayesian estimation framework

Let a sequence of observations of a moving target t, by a stationary sensor system s from time step 1 to time step k, be

$$ {}^s\tilde{\mathbf{z}}_{1:k}^t \equiv \left\{ {}^s\tilde{\mathbf{z}}_\kappa^t | \forall \kappa \in \{1, ..., k\} \right\}. \tag{3.5}$$

Given the initial belief $p\left(\mathbf{x}_0^t\right)$, the sensor platform state $\tilde{\mathbf{x}}^s$ and a sequence of observations ${}^s\tilde{\mathbf{z}}_{1:k}^t$, the belief on the target at any time step k, $p\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k}^t, \tilde{\mathbf{x}}^s\right)$, can be estimated recursively through the two stage equations: update and prediction.

In the prediction process, the target belief is updated in time. The target belief $p\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \tilde{\mathbf{x}}^s\right)$ is updated from that in the previous time step $p\left(\mathbf{x}_{k-1}^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \tilde{\mathbf{x}}^s\right)$

by Chapman-Kolmogorov equation as

$$p\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_{1:k-1}^t,\tilde{\mathbf{x}}^s\right) = \int_{\mathcal{X}^t} p\left(\mathbf{x}_k^t|\mathbf{x}_{k-1}^t\right)\, p\left(\mathbf{x}_{k-1}^t|{}^s\tilde{\mathbf{z}}_{1:k-1}^t,\tilde{\mathbf{x}}^s\right)\, d\mathbf{x}_{k-1}^t, \tag{3.6}$$

where $p\left(\mathbf{x}_k^t|\mathbf{x}_{k-1}^t\right)$ is a probabilistic form of the motion Markov model (3.1). Note that $p\left(\mathbf{x}_{k-1}^t|{}^s\tilde{\mathbf{z}}_{1:k-1}^t,\tilde{\mathbf{x}}^s\right) = p\left(\mathbf{x}_0^t\right)$ when $k = 1$. The correction process, on the other hand, updates the belief in observation. The target belief $p\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_{1:k}^t,\tilde{\mathbf{x}}^s\right)$ is corrected from the corresponding state estimated with the observations up to the previous time step $p\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_{1:k-1}^t,\tilde{\mathbf{x}}^s\right)$ and a new observation ${}^s\tilde{\mathbf{z}}_k^t$ as

$$p\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_{1:k}^t,\tilde{\mathbf{x}}^s\right) = \frac{q\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_{1:k}^t,\tilde{\mathbf{x}}^s\right)}{\int_{\mathcal{X}^t} q\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_{1:k}^t,\tilde{\mathbf{x}}^s\right) d\mathbf{x}_{k-1}^t}, \tag{3.7}$$

where

$$q\left(\cdot\right) = l\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}^s\right) p\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_{1:k-1}^t,\tilde{\mathbf{x}}^s\right)\,, \tag{3.8}$$

and $l\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}^s\right)$ represents the observation likelihood of $\mathbf{x}_k^t$ given ${}^s\tilde{\mathbf{z}}_k^t$ and $\tilde{\mathbf{x}}^s$, which is a probabilistic version of the sensor model. An example would be Equation (3.3), if the sensor is optical. It is to be noted that the likelihood does not need to be a probability density since the normalization in Equation (6.2) makes the output belief a probability density regardless of the formulation of the likelihood.

3.2 Grid-based RBE

3.2.1 Grid-based Representation of RBE

Handling the heavily nonlinear non-Gaussian no-detection likelihood necessitates the grid-based method for RBE. As shown in the Fig. 3.2, the grid-based method represents the belief space in terms of regularly aligned grid cells. Let the cell of concern be positioned at the lth and the mth partitions in x and y directions as $\mathbf{x}^t = [x^t, y^t]^\top \in \mathcal{X}^t$. At the grid cell $[l, m]$, the prediction and the correction are

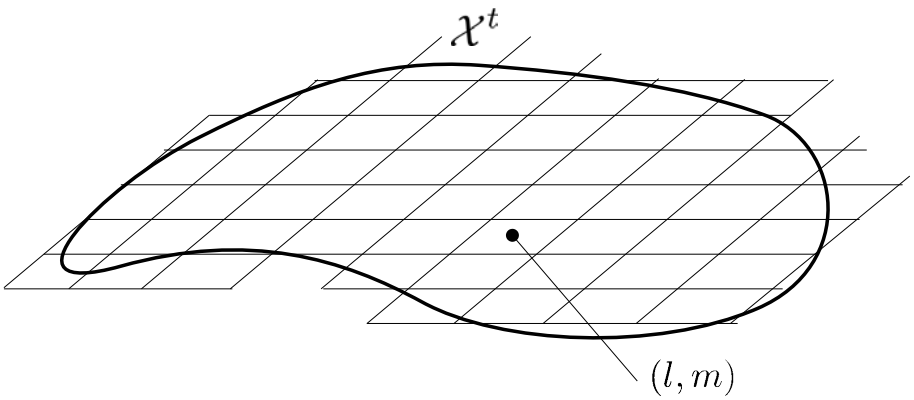


Figure 3.2: Grid-based representation of the target domain

processed independently. The prediction requires the numerical evaluation of the Chapman-Kolmogorov equation in Equation (6.1) at each grid cell. Given the belief $p^{l,m}_{\mathbf{x}^t_{k-1}}\left({}^s\tilde{\mathbf{z}}^t_{1:k-1}\right)$ at time step k as well as the motion model $p^{l,m}_{\mathbf{x}^t_k|\mathbf{x}^t_{k-1}}$ constructed in the $I^t_x \times I^t_y$ matrix form as the convolution kernel, the target belief at the grid cell $[l, m]$ can be predicted as

$$p^{l,m}_{\mathbf{x}^t_k}\left({}^s\tilde{\mathbf{z}}^t_{1:k-1}\right) = p^{l,m}_{\mathbf{x}^t_{k-1}}\left({}^s\tilde{\mathbf{z}}^t_{1:k-1}\right) \otimes p^{l,m}_{\mathbf{x}^t_k|\mathbf{x}^t_{k-1}} \tag{3.9}$$

$$= \sum_{\alpha=0}^{I^t_x} \sum_{\beta=0}^{I^t_y} p^{\alpha,\beta}_{\mathbf{x}^t_k|\mathbf{x}^t_{k-1}} p^{l-\alpha,m-\beta}_{\mathbf{x}^t_{k-1}}\left({}^s\tilde{\mathbf{z}}^t_{1:k-1}\right), \tag{3.10}$$

where $\otimes$ indicates the convolution of the last belief with the motion model.

The correction requires the computation of Equation (6.2) at each grid cell. Given the predicted belief $p^{l,m}_{\mathbf{x}^t_k}\left({}^s\tilde{\mathbf{z}}^t_{1:k-1}\right)$ and the observation likelihood $l^{l,m}_{\mathbf{x}^t_k}\left({}^s\tilde{\mathbf{z}}^t_k\right)$, the target belief at the grid cell $[l, m]$ can be corrected as

$$p^{l,m}_{\mathbf{x}^t_k}\left({}^s\tilde{\mathbf{z}}^t_{1:k}\right) = \frac{q^{l,m}_{\mathbf{x}^t_k}(\cdot)}{\Delta x_r \Delta y_r \sum_{\alpha}\sum_{\beta} q^{\alpha,\beta}_{\mathbf{x}^t_k}(\cdot)}, \tag{3.11}$$

where

$$q^{l,m}_{\mathbf{x}^t_k}\left({}^s\tilde{\mathbf{z}}^t_{1:k}\right) = l^{l,m}_{\mathbf{x}^t_k}\left({}^s\tilde{\mathbf{z}}^t_k\right) p^{l,m}_{\mathbf{x}^t_k}\left({}^s\tilde{\mathbf{z}}^t_{1:k-1}\right). \tag{3.12}$$

and $[\Delta x_r, \Delta y_r]$ is the size of the grid. For the complete formulation refer to Furukawa et al.'s work [144].

3.3 Observation

Whilst the generalized optical observation likelihood allows belief update and maintenance regardless of whether the target has been detected, the RBE with the optical observation likelihood does not update and maintain the belief effectively. Figure 3.3 illustratively depicts this problem where the FOV and the NFOV are given by the light blue and the white colors respectively. When the configuration of the target space is constrained complicatedly, the FOV becomes significantly limited compared to the target space. This makes the belief dominantly updated by the observation likelihood with no detection. If no detection continuously takes place, the belief remains spread out with predictions and becomes highly uncertain and unreliable. The next section will describe the proposed target estimation, incorporating an

acoustic sensor to solve this problem.

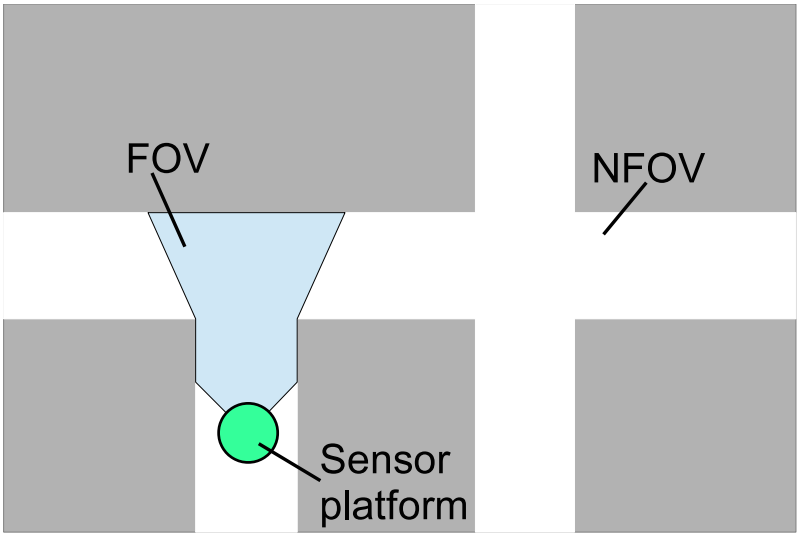


Figure 3.3: Optical observation FOV and NFOV

3.3.1 Modeling of Observation Likelihood

Optical and acoustic sensors

The optical observation likelihood is modeled by first defining the "detectable region". Due to the existence of uncertainty, the observation of a non-empty element does not necessarily indicate that the target has been reliably detected. The detectable region of the optical sensor s_c that describes the region within which the sensor confidently detects the target is thus defined as:

$$^{s_c}\mathcal{X}_d^t = \left\{\mathbf{x}_k^t | \epsilon_c^t < P_d\left(\mathbf{x}_k^t | \tilde{\mathbf{x}}^s\right) \leq 1\right\} \subset {}^{s_c}\mathcal{X}_o^t,$$

where ϵ_c^t is a positive threshold value which judges the detection of the target. Given the observation ${}^s\tilde{\mathbf{z}}_k^t$, the optical observation likelihood is resultantly stated as

$$l_i^c\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}_k^s\right)=\begin{cases} p\left({}^s\tilde{\mathbf{z}}_k^t|\mathbf{x}_k^t,\tilde{\mathbf{x}}^s\right) & \exists {}^s\tilde{\mathbf{z}}_k^t\in {}^{s_c}\mathcal{X}_d^t \\ 1-P_d\left(\mathbf{x}_k^t|\tilde{\mathbf{x}}^s\right) & \nexists {}^s\tilde{\mathbf{z}}_k^t\in {}^{s_c}\mathcal{X}_d^t, \end{cases} \tag{3.13}$$

where, depending on whether there exists a target within the detectable region, the upper and lower formulas return likelihoods with detection and no-detection events, respectively. The detectable region of the acoustic sensor is similarly defined as:

$${}^{s_a}\mathcal{X}_d^t=\left\{\mathbf{x}_k^t|\epsilon_a^t<P_d\left(\mathbf{x}_k^t|\tilde{\mathbf{x}}^s\right)\leq 1\right\}\subset {}^{s_a}\mathcal{X}_o^t.$$

The acoustic observation likelihood, with a nearly unlimited observable region, is represented as a generic equation:

$$l_j^a\left(\mathbf{x}_k^t|{}^{s_a}\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}_k^s\right)=p\left({}^{s_a}\tilde{\mathbf{z}}_k^t|\mathbf{x}_k^t,\tilde{\mathbf{x}}_k^s\right). \tag{3.14}$$

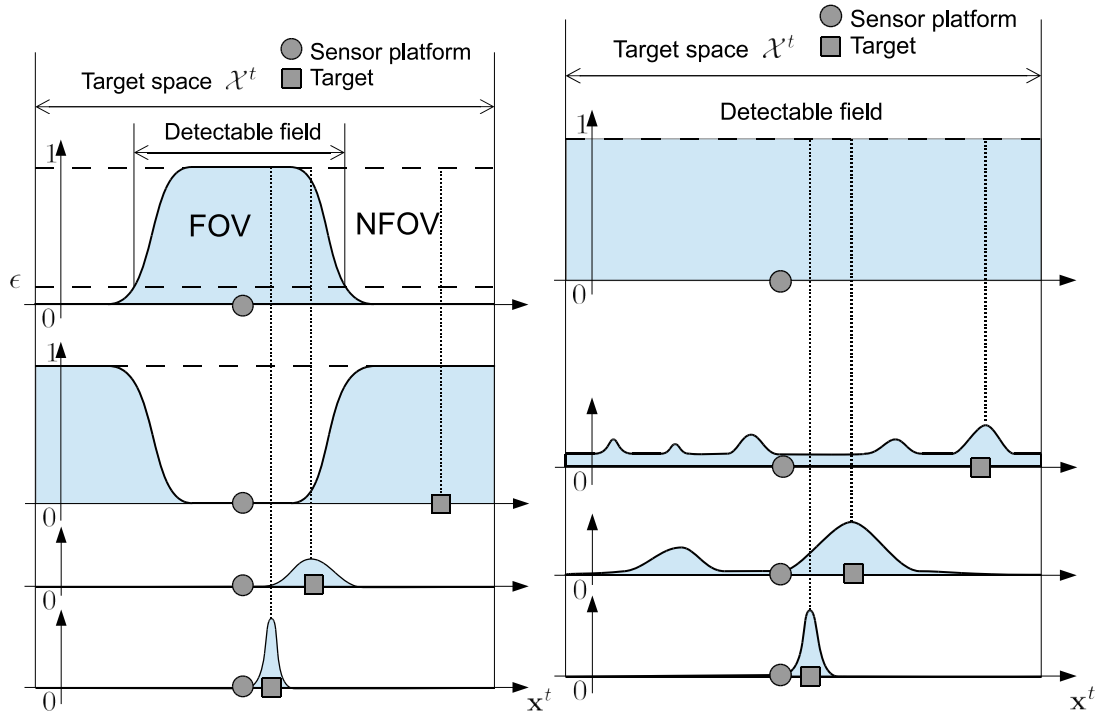
Sensor belief fusion

Given the observation likelihood of the optical sensor $l_i^c\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}_k^s\right)$ and that of the acoustic sensor $l_j^a\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}_k^s\right)$, the joint likelihood is given by multiplying the two likelihoods

$$l\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}_k^s\right)=l_i^c\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}_k^s\right)l_j^a\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}_k^s\right) \tag{3.15}$$

in accordance to the canonical data fusion formula.

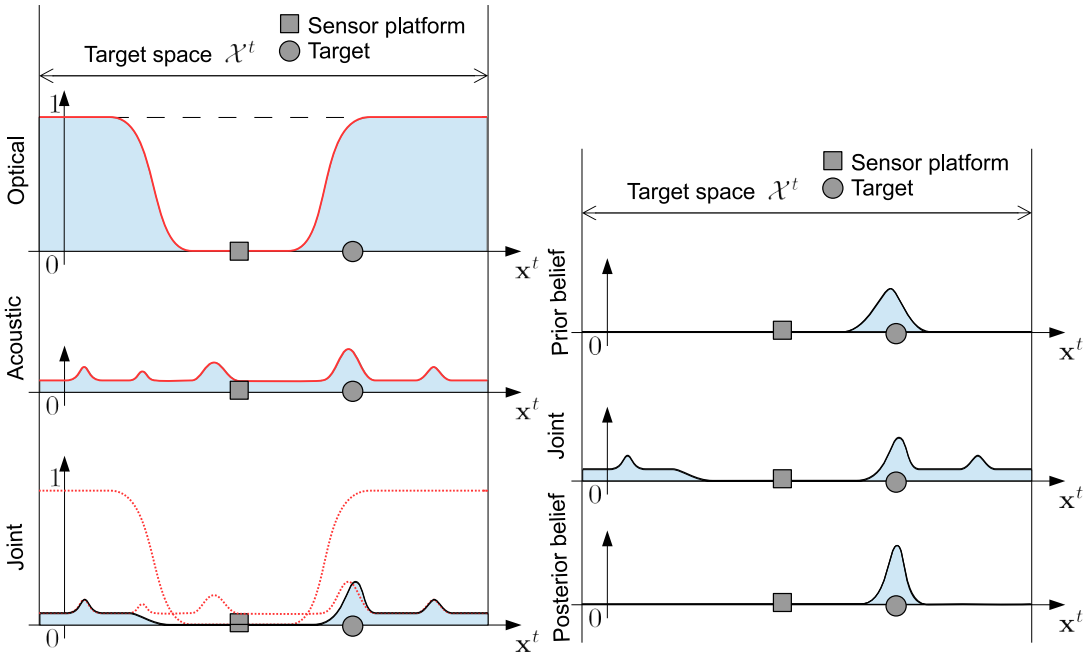
Figure 3.5 illustrates the strength and weakness of the generalized optical/acoustic RBE. Figure 3.5a shows the joint optical/acoustic observation likelihood when the



(a) Joint likelihood (b) RBE with acoustic sensor

Figure 3.4: Optical and acoustic sensor models

target is in the NFOV. The acoustic observation likelihood constructed for a NFOV target becomes heavily non-Gaussian with poor accuracy. However, the possible locations of the target are narrowed down since the optical likelihood with no detection clears out any likelihood in the detectable region and drops some peak(s) as shown in the figure. With the joint likelihood, the posterior belief could be a near-Gaussian probability density if the prior belief is near-Gaussian as depicted in Figure 3.5b. On the contrary, this means that this approach provides a good estimate of the target location only if prior belief has been constructed under the strong influence of optical observations, and is thus near-Gaussian. The approach fails in indoor environments where optical sensors are not allowed or ineffective. Such environments have to depend on acoustic sensing with improved accuracy.



(a) Joint likelihood (b) RBE with acoustic sensor

Figure 3.5: Generalized optical/acoustic RBE

CHAPTER 4

Non-Field-of-View Acoustic Localization

This chapter presents an NFOV target estimation approach that incorporates optical and acoustic sensors. An optical sensor can accurately localize a target in its FOV whereas the acoustic sensor could estimate the target location over a much larger space but only with limited accuracy. An RBE framework where observations of the optical and acoustic sensors are probabilistically treated and fused is proposed in this chapter. A technique to construct the observation likelihood when two microphones are used as the acoustic sensor is also described. The proposed technique derives and stores, the Interaural level difference (ILD) of observations from the two microphones for different target positions, and constructs the likelihood through correlation. A

parametric study of the proposed acoustic sensing technique in a controlled test environment and experiments with an NFOV target in an actual indoor environment, are presented to demonstrate the capability of the proposed technique.

4.1 Indoor NFOV Localization with Fixed Microphone Array

4.1.1 Introduction

This section presents a new acoustic approach to estimate a NFOV mobile target, as well as application to and implementation in complex indoor environments. In the approach, cameras are not applicable because of the privacy of the person who coexist in the environment. Only microphones are sparsely installed in the indoor environment of concern. In this structural implementation of the sensors, as a prior process to the estimation, the ILD of observations acquired by every combination of microphone pairs is derived for different sound target positions and stored as the "fingerprints", or acoustic cues. With the acquisition of a new sound from the target, an acoustic observation likelihood is computed for each pair of microphones by quantifying the correlation of new observation ILD to the stored ILDs. The joint likelihood is then created by fusing the acoustic observation likelihoods, and the NFOV target is estimated by recursively updating the belief within the RBE framework, using the joint likelihood.

4.2 NFOV Acoustic Target Estimation

4.2.1 Indoor Installation

Figure 4.1 shows a schematic hardware installation of the proposed acoustic target estimation approach. As shown in the figure, microphones are placed with some distance in the indoor environment. This is an environment where, whether they are allowed or not, optical sensors would not be used effectively due to complications in the environment due to personal privacy. A number of optical sensors must be placed to cover the entire space. Microphones, on the other hand, can collect information on the NFOV. A much smaller number of inexpensive sensors need to be installed for this reason, making the installation efficient in both time and cost.

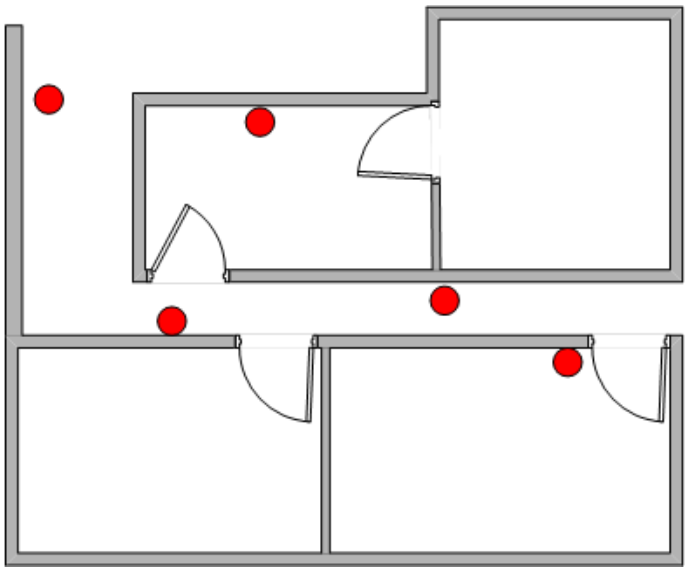


Figure 4.1: Schematic hardware installation of proposed approach

4.3 Modeling of Acoustic Observation Likelihood

In accordance with the preliminary investigations of the authors [145], the theoretical approach proposed in this chapter constructs acoustic cues of the target in the environment of concern *a priori* to create an acoustic observation likelihood. Figure 5.1 shows a schematic diagram of the proposed approach. We assume that the target emits sound with white noise and uses it to create the acoustic observation likelihood. A white noise sound emitted at a specific position by the target for a certain time period is first recorded by the multiple microphones fixed to the environment. After applying Fast Fourier transform (FFT), the difference between the frequency-domain amplitude responses, known as the ILD, is then derived for every combination of two microphones and further sampled to form an ILD vector within the frequency range of interest. The ILD vector is created with various target positions and each is saved as an acoustic cue. The acoustic observation likelihood modeling corresponds to creating the set of ILD vectors. When a target has emitted a white noise sound, the ILD vector of the sound observation is compared to all the acoustic cues between microphone combination pairs. The degree of similarity is then used to develop a correlation map indicating where the target is likely to be. The correlation map is the acoustic likelihood of the particular sound observation. Finally, the correlation maps of the pairs of microphone are collected and fused to yield a joint acoustic observation likelihood. This fusion process only considers a few dominant microphone pairs for scalability of the system.

Mathematically, let the frequency-domain sound level of the target at the ith position $(\tilde{\mathbf{x}}_k^t)_i$, which is observed by $(\cdot)$th microphone at $\tilde{\mathbf{x}}_k^s$, be $s_{(\cdot)}\left(\omega|\left(\tilde{\mathbf{x}}_k^t\right)_i\right)$ where ω is a frequency of sound. The ILD of the jth microphone pair $\{j1, j2\}$ for the ith

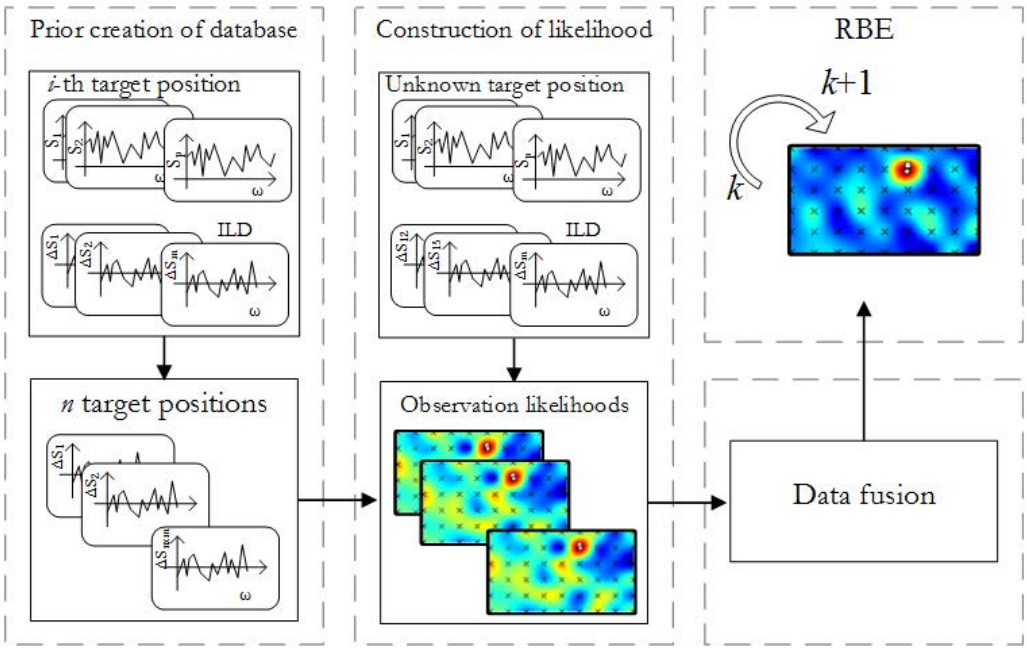


Figure 4.2: Schematic diagram of proposed approach

position $(\tilde{\mathbf{x}}_k^t)_i$, $\Delta S_i^j(\omega)$, is then given by

$$\Delta S_i^j(\omega) = 20\log\left|s_{j1}\left(\omega\middle|\left(\tilde{\mathbf{x}}_k^t\right)_i\right)\right| - 20\log\left|s_{j2}\left(\omega\middle|\left(\tilde{\mathbf{x}}_k^t\right)_i\right)\right|. \tag{4.1}$$

If the ILD is sampled at N frequencies $\mathbf{\Omega} = [\omega_1, ..., \omega_N]^\top$, the ILD vector can be described as

$$\mathbf{S}_i^j(\mathbf{\Omega}) = \left[a_1^j \Delta S_i^j(\omega_1), ..., a_N^j \Delta S_i^j(\omega_N)\right]^\top, \tag{4.2}$$

where

$$a_i^j = \left\langle \min\{\left|s_{j1}\left(\omega_N\middle|\left(\tilde{\mathbf{x}}_k^t\right)_i\right)\right|, \left|s_{j2}\left(\omega_N\middle|\left(\tilde{\mathbf{x}}_k^t\right)_i\right)\right|\} - \epsilon \right\rangle. \tag{4.3}$$

In the equation, $\langle\cdot\rangle$ is Macaulay brackets, and $\min\{\cdot,\cdot\}$ returns the smaller value of the two entities. The acoustic observation likelihood modeling results in the ILD

vectors for n target positions, i.e., $\mathbf{S}_i^{j*}(\mathbf{\Omega}), \forall i \in \{1, ..., n\}$. They are essentially the acoustic cues to be prepared in advance and used to create the acoustic observation likelihood.

Given the ILD vector $\mathbf{S}^j(\mathbf{\Omega}|\mathbf{x}_k^t)$ created from ${}^s\tilde{\mathbf{z}}_k^t$ with the unknown target position $\mathbf{x}_k^t$, the proposed technique quantifies its degree of correlation to the ith ILD vector as

$$X\left(\mathbf{S}^j\left(\mathbf{\Omega}|\mathbf{x}_k^t\right), \mathbf{S}_i^{j*}(\mathbf{\Omega})\right) = \frac{1}{2}\left\{\frac{\mathbf{S}^j(\mathbf{\Omega}|\mathbf{x}_k^t)^\top \mathbf{S}_i^{j*}(\mathbf{\Omega})}{|\mathbf{S}^j(\mathbf{\Omega}|\mathbf{x}_k^t)|\left|\mathbf{S}_i^{j*}(\mathbf{\Omega})\right|} + 1\right\} \tag{4.4}$$

where $0 \leq X(\cdot) \leq 1$. The acoustic observation likelihood with the particular $\mathbf{S}_m(\mathbf{\Omega}|\mathbf{x}_k^t)$ can be finally calculated as

$$l_j^a\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t, \tilde{\mathbf{x}}_k^s\right) = \sum_{i=1}^{n} \mu_i^j\left(\mathbf{x}_k^t\right) X\left(\mathbf{S}^j\left(\mathbf{\Omega}|\mathbf{x}_k^t\right), \mathbf{S}_i^{j*}(\mathbf{\Omega})\right), \tag{4.5}$$

where $\mu_i^j(\mathbf{x}_k^t)$ is a basis function developed by adjacent measurements. One of the suited basis function is a T-spline basis function where $\mu_i^j(\mathbf{x}_k^t)$ in a T-mesh in parameter space (s,t) can be represented as

$$\mu_{im}(s,t) = N(s)N(t) \tag{4.6}$$

where $N(s)$, and $N(t)$ are the cubic B-spline basis functions. Further detailed formulations are found in [146]. It is to be noted that we will not specify the best shape function to use in this chapter since every shape function has advantages and disadvantages. Similarly to $X(\cdot)$, $l_j^a(\cdot)$ is also bounded as $0 \leq l_m^a(\cdot) \leq 1$ due to the use of the shape function.

Finally, the joint likelihood is derived by the canonical data fusion formula:

$$l_j^a\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}_k^s\right)=\prod_j l_j^a\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}_k^s\right). \tag{4.7}$$

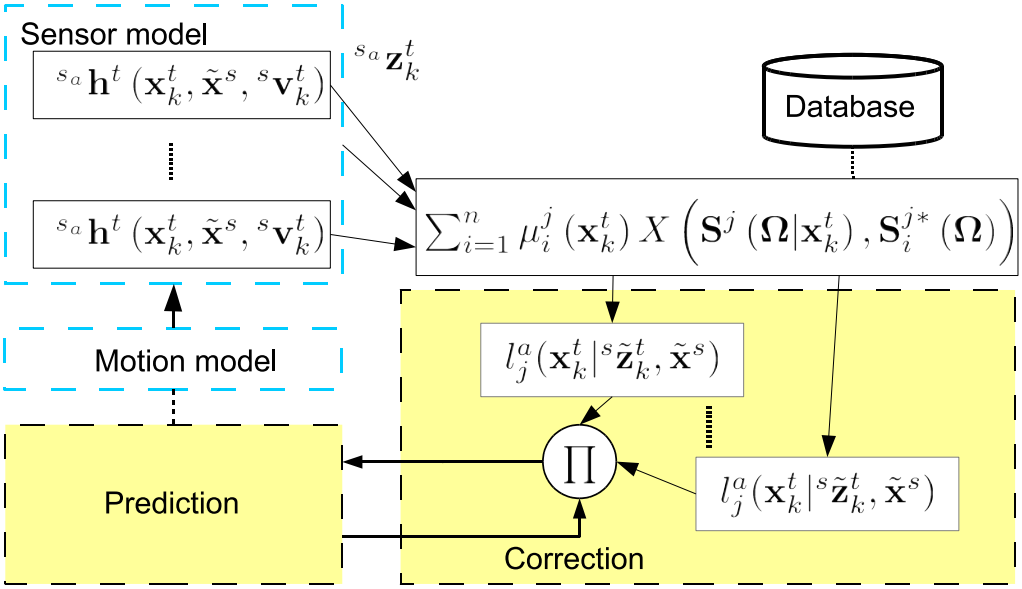


Figure 4.3: RBE framework-given sensor information from acoustic signals

4.3.1 Numerical and Experimental Analysis

The efficacy of the proposed approach was examined experimentally in two steps. The first step was aimed at studying the capability and limitation of the proposed acoustic sensing technique. This was achieved by parametrically changing the complexity of the environment where the experimental system, with a speaker array and a movable/replaceable wall, was developed specifically for this study. After verifying the feasibility of the acoustic sensing for NLOS target localization, the applicability of the proposed approach in a practical indoor scenario was investigated. The investigation looked not only into the performance of the proposed approach but also

into its comparison to a conventional approach.

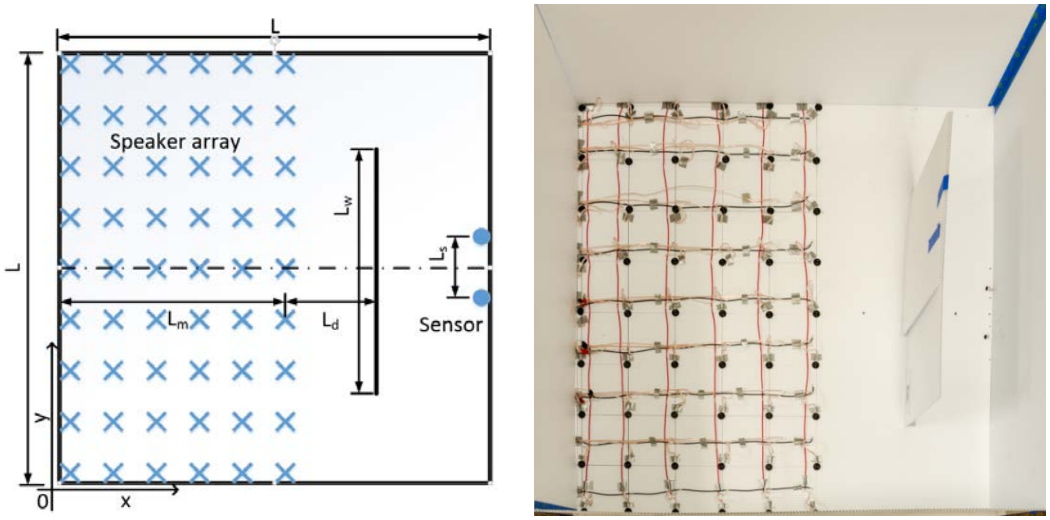
4.4 Acoustic Observation of NLOS Target

Figure 4.4a shows the design of the experimental system that changes the complexity of the environment for the evaluation of the proposed acoustic sensing technique. An acoustic sensor, consisting of two microphones, is fixed next to an outer wall and faces an open space where a speaker array and movable/replaceable wall(s) are placed. The complexity of the environment can be changed by varying the two parameters of the movable/replaceable wall: the distance of the wall to the edge of speaker array L_d and the length of the wall L_w. The longer the distance and/or the larger the length, the more complex the environment is, since sound from the speakers result in more reflections.

Shown in the figure as blue crosses are the speaker locations. A microcontroller controls speakers so that each speaker sequentially emits white noise sound for a programmed period. A set ILDs for a wall setting can be thus collected automatically. Once the ILDs are collected, the ability of the proposed acoustic sensing technique is evaluated by emitting sound from a speaker at some location within the area of the speaker array and identifying the location in the form of observation likelihood. The location is not where one of the speakers of the speaker array is located. The use of a different location demonstrates the ability of the proposed technique in identifying the target at an arbitrary position.

Figure 4.4b shows the developed experimental system whereas the dimensions and other parameters used in the experiments are listed in Table 4.1. Sound was sampled and represented in 8,192 frequency bins within the audible range to capture its behavior accurately. 54 speakers were aligned to cover the open space. The

distance and the length of the wall were varied to introduce both lightly NLOS and heavily NLOS environments. A case of two walls ($n_w = 2$) was tested in addition to the single wall case to make the environment more complex. The distance of only the wall closer to the acoustic sensor was varied.



(a) Schematic design (b) Developed system

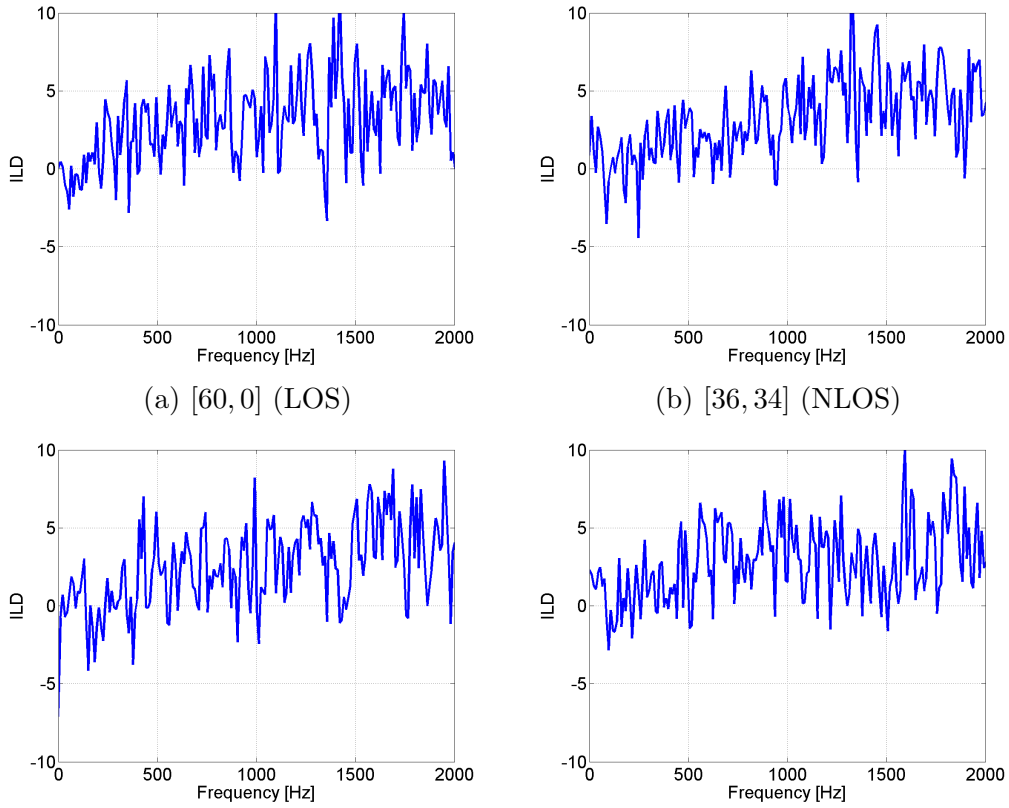
Figure 4.4: Experimental system for investigating the influence of environmental complexity

Table 4.1: Dimensions and other parameters in the experiments

Parameter	Value	Parameter	Value
$\tilde{\mathbf{x}}^t$	$[42, 34]$ cm	L	90 cm
ω_1	0 Hz	L_m	50 cm
ω_N	22 kHz	L_s	10 cm
N	8,192	L_d	$\{0, 10, 20, 30\}$ cm
ϵ	0.01	L_w	$\{50, 60, 70\}$ cm
n	54	n_w	$\{1, 2\}$

Figure 4.5 shows the six ILDs, each observed with a target at one of the 54 positions when $[L_d, L_w, n_w] = [50, 40, 1]$. Two positions are on the LOS, and four are on the NLOS. The configuration of the ILD varies depending on the target

position. The configuration is different even when the target is on the NLOS. This indicates that the ILD contains information on the target position whether or not the target is on the LOS.



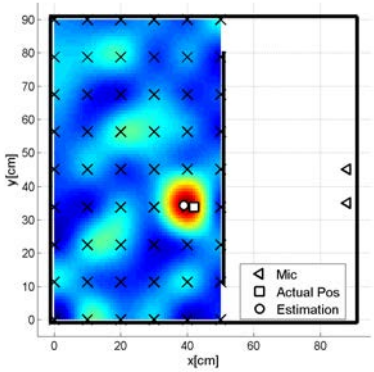
(a) $[60, 0]$ (LOS) (b) $[36, 34]$ (NLOS)

(c) $[24, 56]$ (NLOS) (d) $[12, 90]$ (LOS)

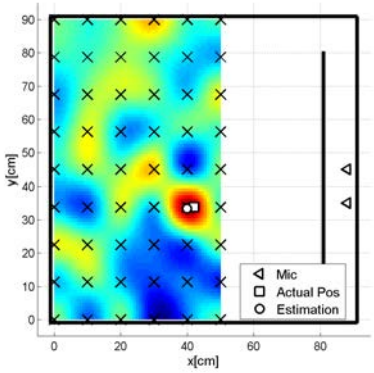
Figure 4.5: ILDs at six of the 54 positions

Figure 4.6 shows four resulting acoustic observation likelihoods, each with a different condition. The former two cases are with a single wall having different distances and the same length whereas the latter two cases are with two walls having different lengths and the same distance. The result indicates that the target location is well estimated when the distance is short (Figure 4.6a) or the length is small (Figure 4.6c). The target is closer to the LOS in these conditions since sound reaches the acoustic sensor with a small number of reflections. The identification of

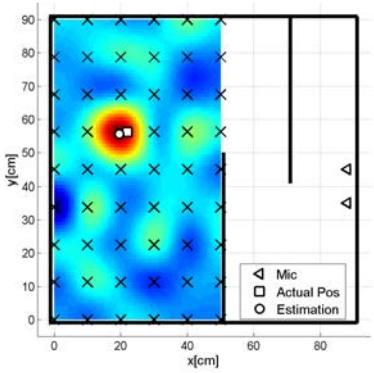
the target location in the remaining two cases (Figures 4.6b and 4.6d) is difficult due to a number of sound reflections. The identification with two walls (Figures 4.6c and 4.6d) is seen to be more laborious than that with single walls (Figures 4.6a and 4.6b) for the same reason. While the acoustic observation likelihood is heavily multi-modal in these cases, the target location is still captured by the highest peak (Figures 4.6a-4.6c), or at least by one of the peaks (Figure 4.6d). This demonstrates the ability of the proposed acoustic sensing technique to identify the location of the NFOV target. However, it is with limited accuracy.



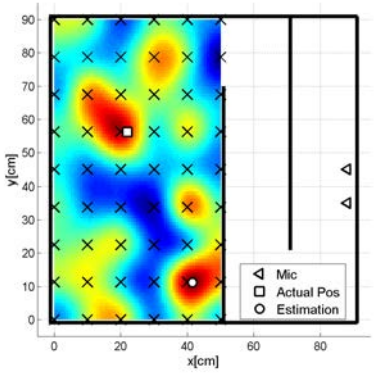
(a) $L_d = 0, L_w = 70, n_w = 1$



(b) $L_d = 30, L_w = 70, n_w = 1$



(c) $L_d = 20, L_w = 50, n_w = 2$



(d) $L_d = 20, L_w = 70, n_w = 2$

Figure 4.6: Acoustic observation likelihoods for different environmental complexities

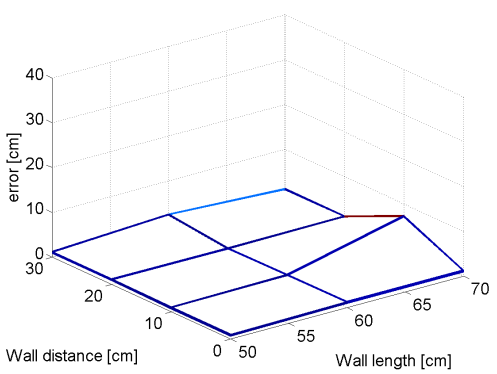
Figure 4.7a and 4.7b show the mean error of the acoustic observation likelihood

when the distance and the length were varied for single and double wall cases. The mean error is the distance of the nearest peak of the acoustic observation likelihood to the true target location. The result of the mean error shows that the proposed technique could locate the target within a 2 cm error in most (11) of the 12 cases for the single wall case. The estimation is particularly good when the wall length was small. The mean error with two walls, meanwhile, is over 20 cm in four of nine cases. This indicates that the proposed acoustic sensing technique is not sufficient enough for NLOS localization. Figure 4.7c shows the variation of the differential entropy with respect to the number of walls and the wall length. It is seen that the addition of a wall dominantly increases uncertainty. While the proposed acoustic sensing technique is effective enough in relatively simple NLOS environments, enhancement is necessary when target estimation in more complex environments is pursued.

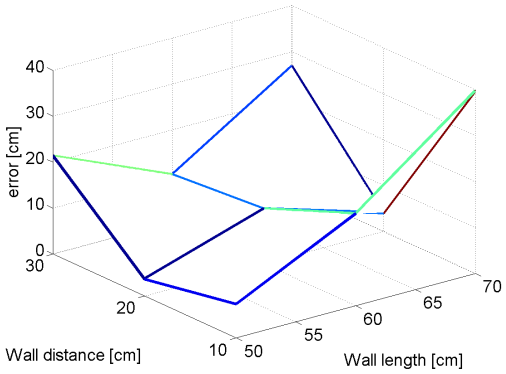
4.5 Applicability to Practical Indoor Scenario

4.5.1 Practical Indoor Scenario

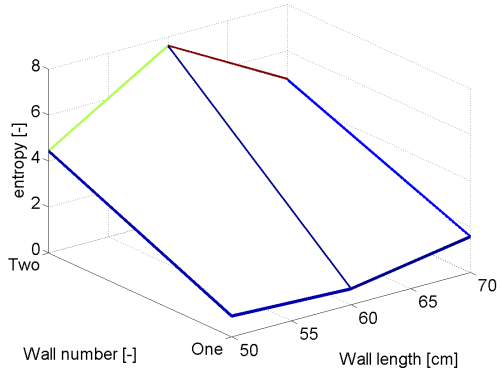
With the acoustic sensing ability validated, the applicability of the proposed approach in NFOV target estimation to a practical indoor scenario was investigated. Figure 4.8 shows the actual indoor environment used for the investigation. This is an apartment of an elderly person who needs a home health care service. As shown in the figure, the environment with five separate rooms is so complicated that it is not possible to cover the entire area by cameras. In addition, this is a personal home, so cameras are not able to be installed. The approximate dimensions of the apartment are 7.1 m in width, 10.4 m in length and 2.5 m in height. Six microphones, shown as red dots, were fixed to cover the entire space. The target person



(a) Single wall mean error



(b) Double wall mean error



(c) Diferential entropy for single and double wall

Figure 4.7: Mean error and differential entropy of the acoustic observation likelihood with a single and double wall

held a speaker which emitted sound with white noise. Parameters used for acoustic target estimation are listed in Table 4.2.

Table 4.2: Parameters for acoustic observation likelihood

Parameter	Value
ω_1	0 [Hz]
ω_N	2.7 [kHz]
N	2,000
ϵ	0.01
n	255

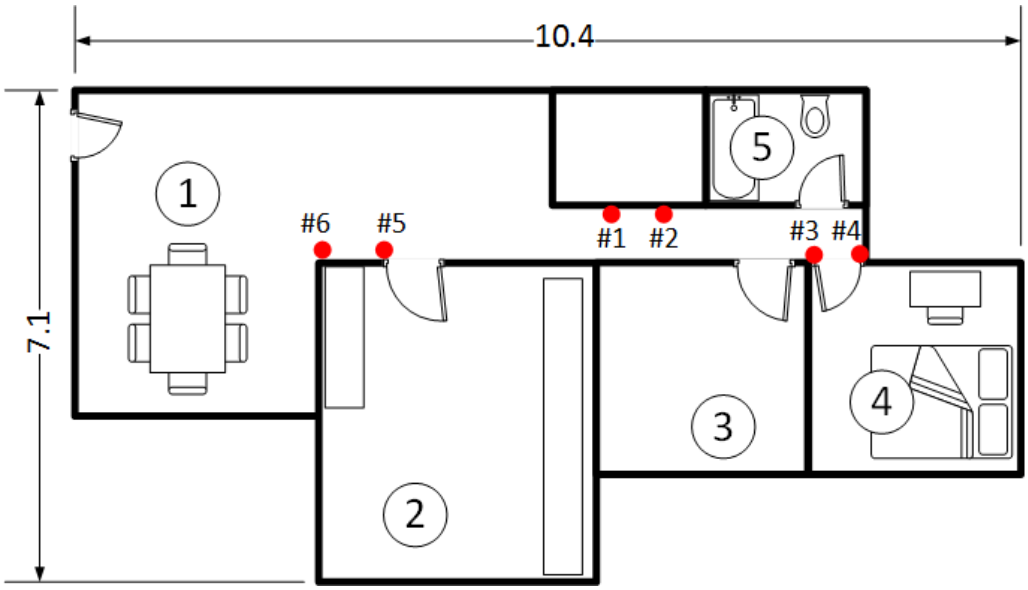
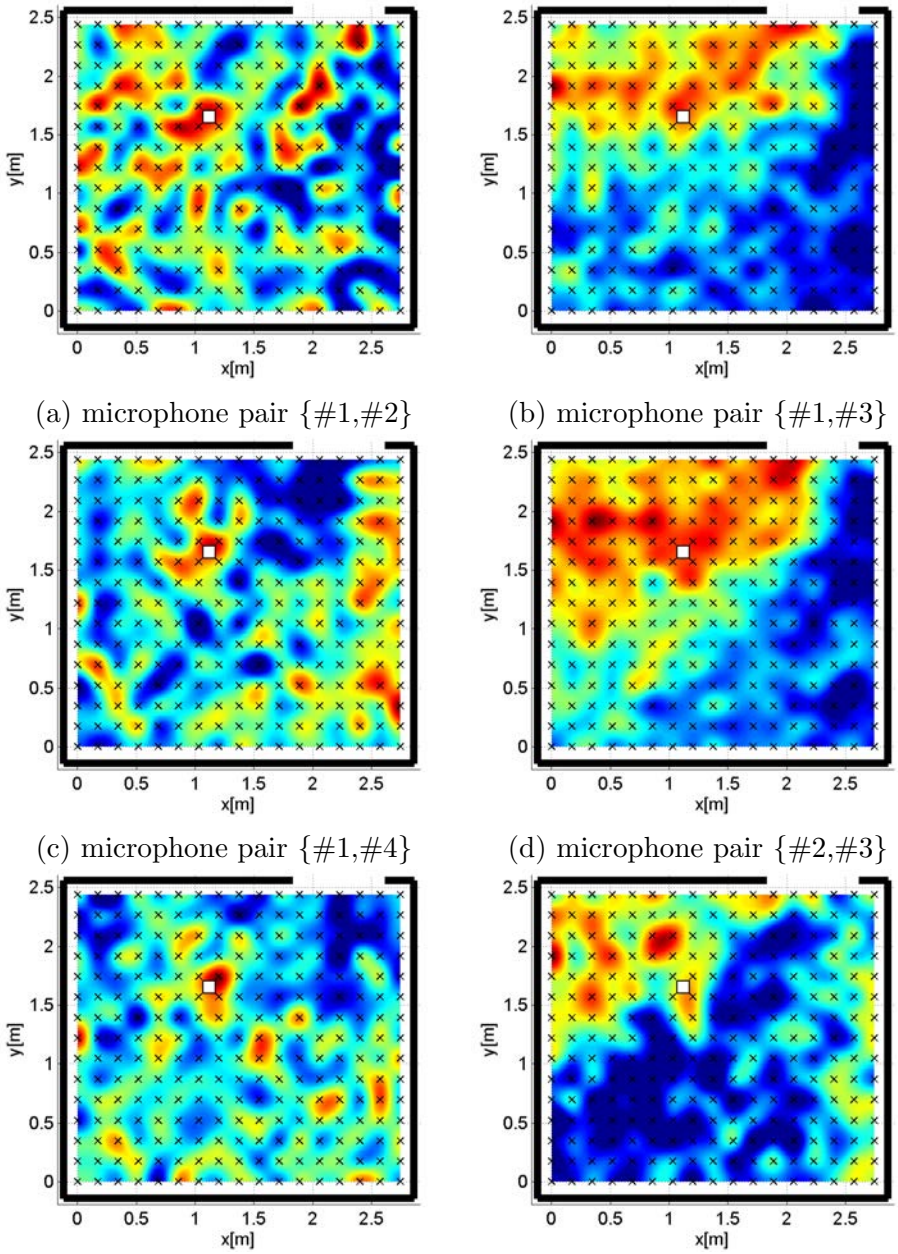


Figure 4.8: Map of the test environment dimensions[m] and other details

4.6 Results

Figure 4.9 shows the acoustic observation likelihoods created by microphone pairs when the target person walked in Room 3. The square dot indicates the true target position. Only the likelihoods with microphones 1-4 are shown since those with microphones 5 and 6 did not contribute well to the target estimation due to the long distance. Identified best of the combinations are pairs 2,3 and 1,3. Microphones 1-3 have the most direct LOS to Room 3, so the result matches well with the logical explanation. Figure 4.10 shows the resulting joint likelihood. The target location is accurately identified by filtering uncertainties. Additional examples of the fused observation likelihoods can be found in the Appendix.

The result of RBE when the target person walked around is shown in Figure 4.12 with the true position again indicated by a square dot. It is seen that the proposed approach accurately tracks the target. The estimated position was less than 15 cm to the true target position 83 % of the time. Cameras and RF re-



(a) microphone pair {#1,#2} (b) microphone pair {#1,#3}

(c) microphone pair {#1,#4} (d) microphone pair {#2,#3}

(e) microphone pair {#2,#4} (f) microphone pair {#3,#4}

Figure 4.9: Acoustic likelihood in room 3 from multiple sensor combinations

ceivers/transmitters cannot be used for such a highly constrained environment, so the conventional acoustic sensing based on two microphones was tested as the only comparable approach. As shown in Figure 4.11, the conventional approach was not

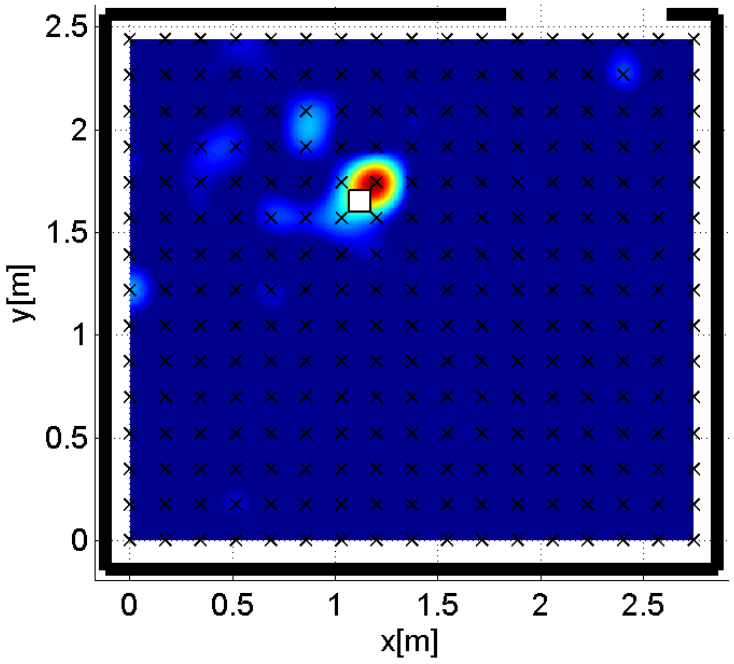


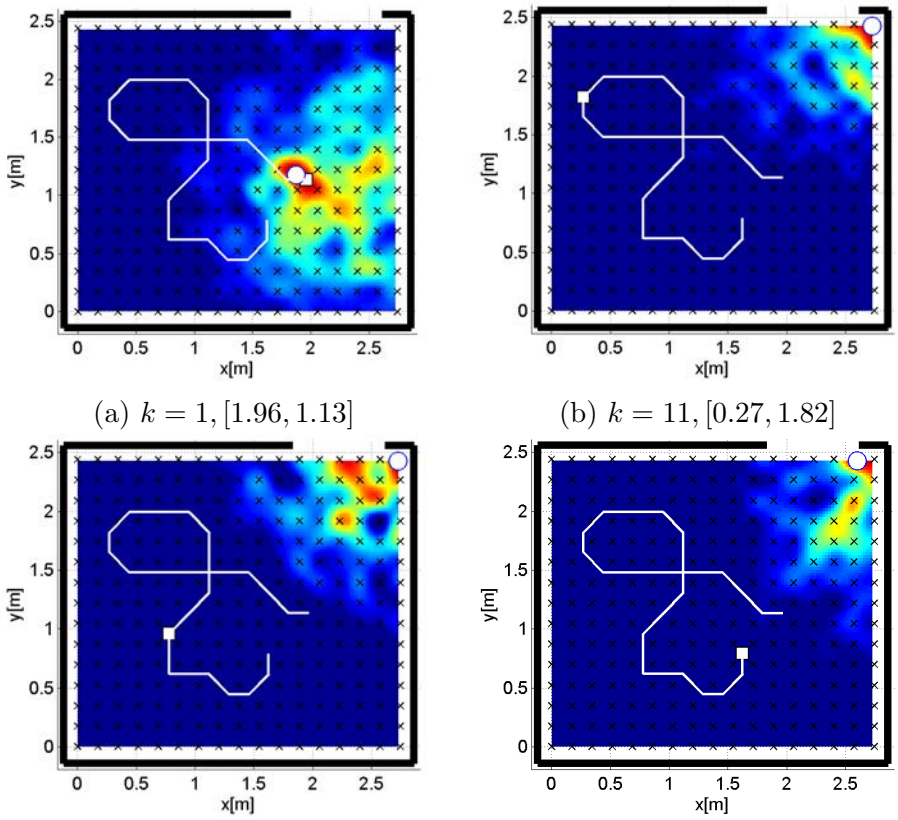
Figure 4.10: Joint acoustic observation likelihood

able to identify the target location once it had failed in the localization.

The above mentioned tracking is further examined in Figure 4.13. The figure shows the error between the estimation with and without a motion model, within the RBE framework demonstrating the effectiveness of the motion model. The worst estimation case at step $k = \{14, 24, 26\}$ is selected and shown in Figure 4.14. Multiple peaks from the fused acoustic observation likelihood are corrected for all the cases; however, step 14 does not provide significant difference in those two cases. This is because the prior position of the target contributes to the prediction

4.7 Chapter Summary

This section has presented a new approach with a set of microphones which localizes and tracks a mobile NFOV target, as well as application to and implementation in

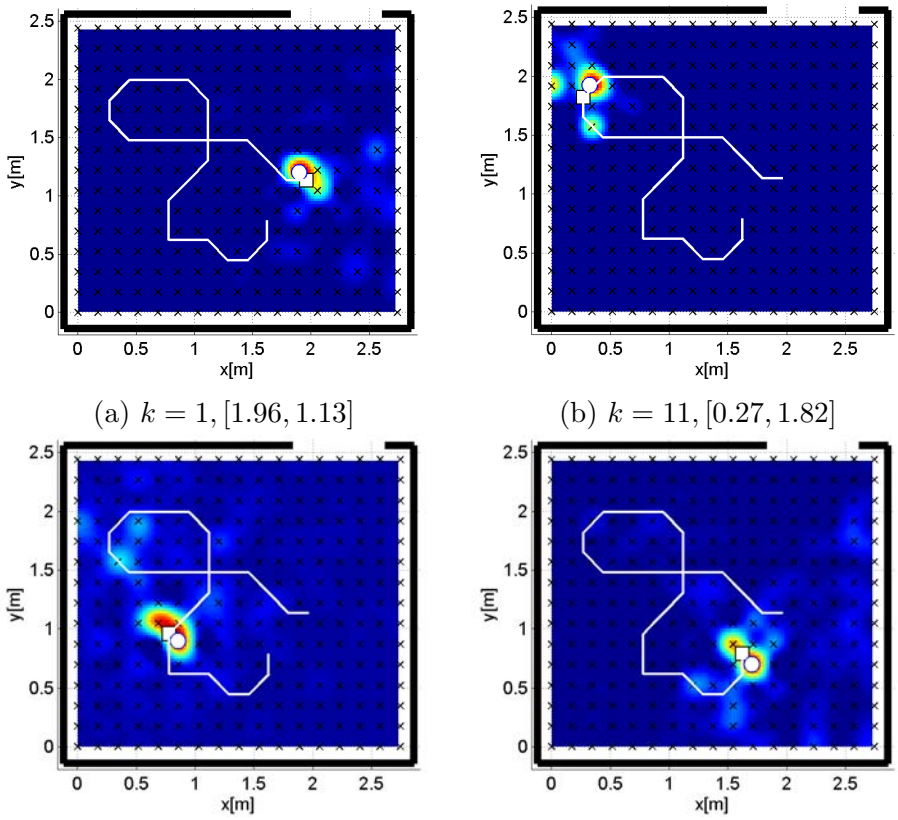


(a) $k = 1, [1.96, 1.13]$ (b) $k = 11, [0.27, 1.82]$

(c) $k = 21, [0.78, 0.96]$ (d) $k = 29, [1.63, 0.79]$

Figure 4.11: Acoustic likelihood in room 3 with one microphone pair

complex indoor environments. The proposed approach derives the ILD of observations from each set of two microphones for different target positions and stores the ILDs as acoustic cues. Given a new sound, an acoustic observation likelihood is computed for each set of two microphones by correlating ILDs. The joint likelihood is then created by fusing the acoustic observation likelihoods, and the NFOV mobile target is estimated by the RBE. Following the experimental parametric studies, the proposed approach was applied to the tracking of an elderly person needing home health care service. The estimation was successful within the 15cm accuracy at 83 % of the estimation. The results have conclusively demonstrated the potential of the proposed approach for practical target localization.



(a) $k = 1, [1.96, 1.13]$ (b) $k = 11, [0.27, 1.82]$

(c) $k = 21, [0.78, 0.96]$ (d) $k = 29, [1.63, 0.79]$

Figure 4.12: Joint acoustic likelihood in room 3 with RBE

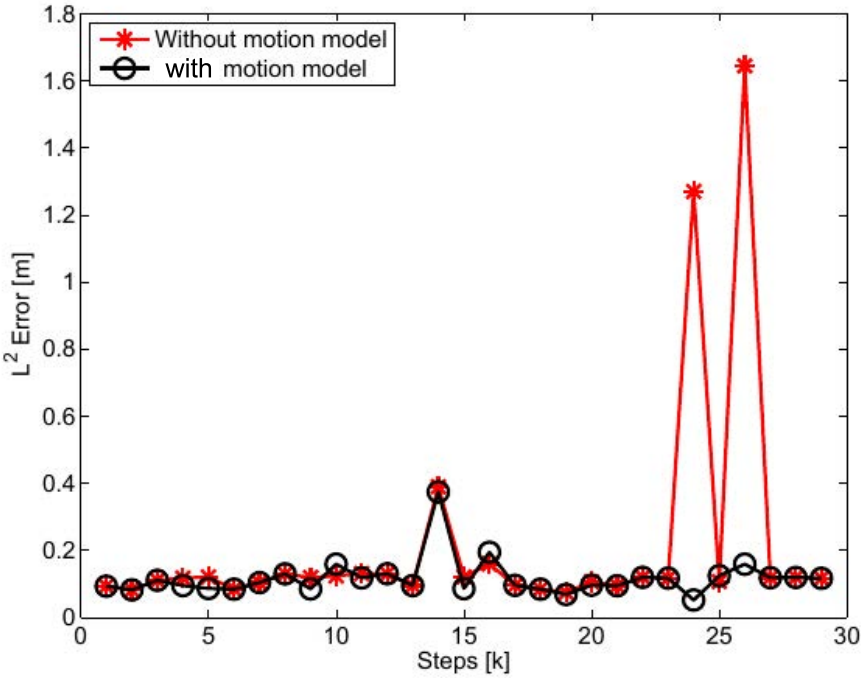
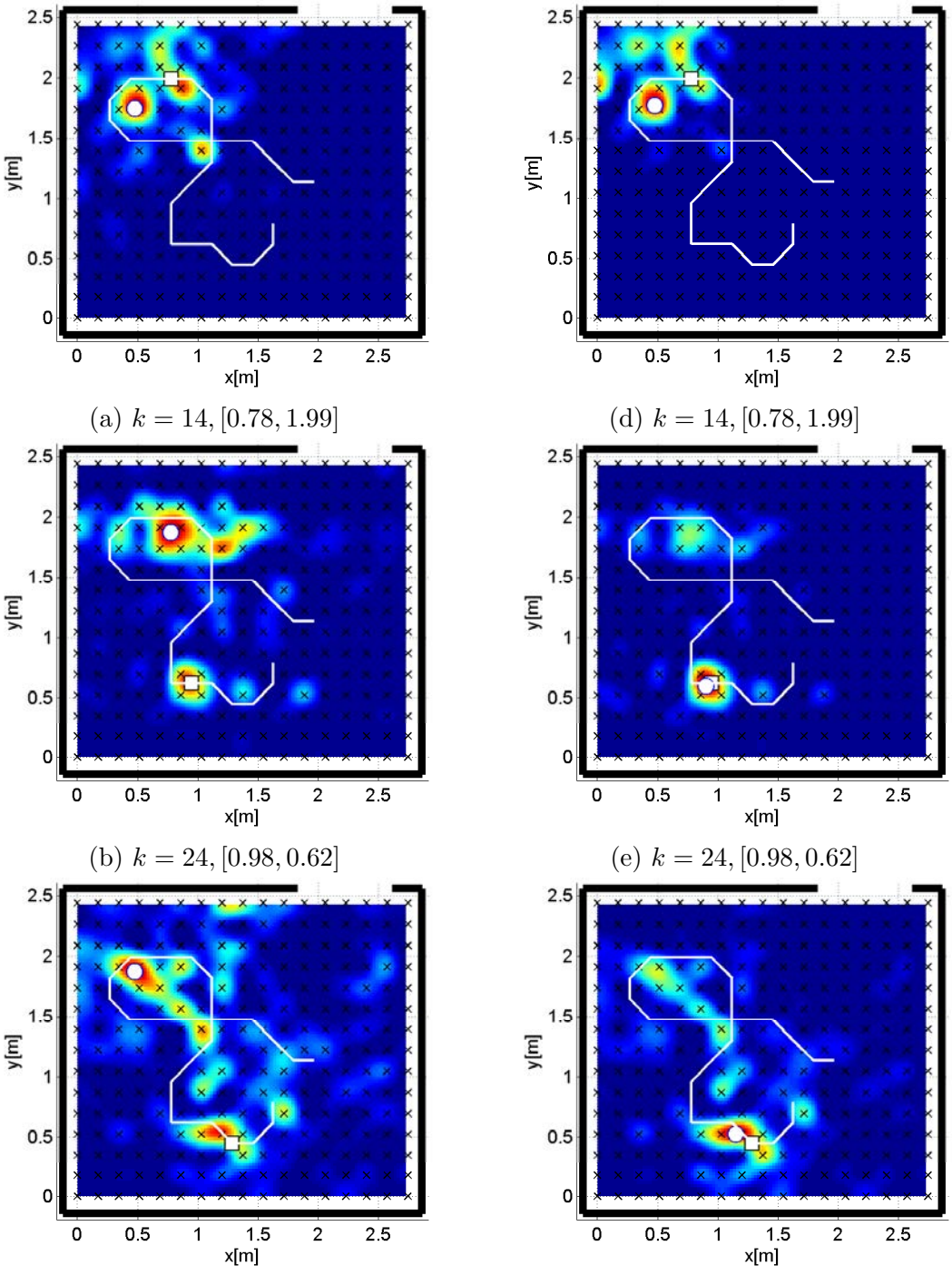


Figure 4.13: Error with and without motion model in RBE



(a) $k = 14, [0.78, 1.99]$

(d) $k = 14, [0.78, 1.99]$

(b) $k = 24, [0.98, 0.62]$

(e) $k = 24, [0.98, 0.62]$

(c) $k = 26, [1.29, 0.45]$

(f) $k = 26, [1.29, 0.45]$

Figure 4.14: Worst tracking case without (a)-(c) and with motion model (d)-(f)

CHAPTER 5

Visual/Auditory Hybrid RBE

5.1 Modeling of Acoustic Observation Likelihood

The technique proposed in this chapter extends the previous approach to more general mobile robot applications, while incorporating both optical and acoustical observation. A model of an acoustic observation likelihood uses two microphones as an acoustic sensor and constructs acoustic cues of the target in the environment of concern *a priori*. This is due to the potential of the proposed technique for NLOS target estimation through the preliminary investigations of the authors [145, 147, 148] and the inability of the aforementioned existing techniques. Figure 5.1 shows a schematic diagram of the proposed technique to model an acoustic observation likelihood. We assume that the target emits sound with white

noise, and we indeed use it to create the acoustic observation likelihood. A white noise sound emitted at a specific position by the target for a certain time period is first recorded by two microphones. After applying FFT, the difference between the frequency-domain amplitude responses, known as the ILD, is then derived and further sampled to form an ILD vector within the frequency range of interest. The ILD vector is created with various target positions and each saved as an acoustic cue. The acoustic observation likelihood modeling essentially corresponds to creating the set of ILD vectors. When a target emits a white noise sound, the ILD vector of the sound observation is compared to all the acoustic cues. The degree of similarity is then used to develop a correlation map indicating where the target is likely to be. The correlation map is the acoustic likelihood of the particular sound observation.

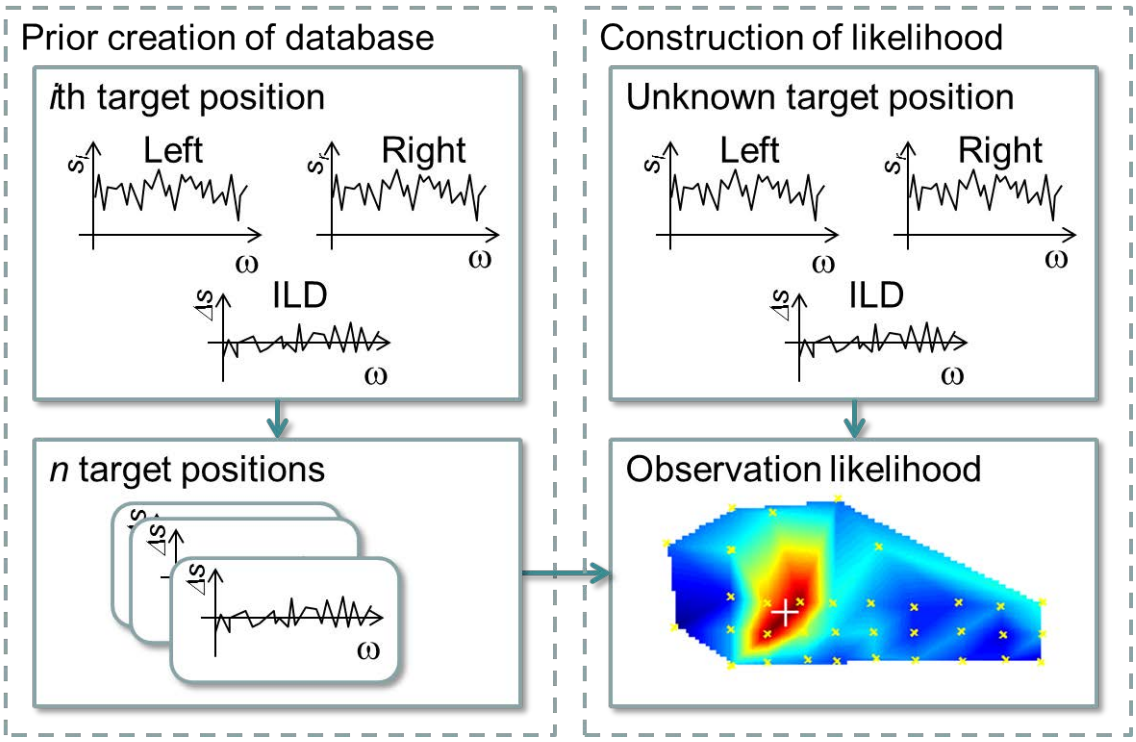


Figure 5.1: Acoustic sensor

Mathematically, let the frequency-domain sound level of the target at the ith

position $(\mathbf{x}_k^t)_i$, which is observed by the left and right microphones, be $s_l(\omega|(\mathbf{x}_k^t)_i)$ and $s_r(\omega|(\mathbf{x}_k^t)_i)$ where ω is a frequency of sound. The ILD for the ith position $(\mathbf{x}_k^t)_i$, $\Delta S(\omega|(\mathbf{x}_k^t)_i)$, is then given by

$$\Delta S\left(\omega|\left(\mathbf{x}_k^t\right)_i\right) = 20\log\left|s_l\left(\omega|\left(\mathbf{x}_k^t\right)_i\right)\right| - 20\log\left|s_r\left(\omega|\left(\mathbf{x}_k^t\right)_i\right)\right|. \tag{5.1}$$

If the ILD is sampled at N frequencies $\mathbf{\Omega} = [\omega_1, \ldots, \omega_N]^\top$, the ILD vector can be described as

$$\begin{aligned}&\mathbf{S}\left(\mathbf{\Omega}|\left(\mathbf{x}_k^t\right)_i\right)\\ &\quad= \left[a_1\Delta S\left(\omega_1|\left(\mathbf{x}_k^t\right)_i\right), \ldots, a_N\Delta S\left(\omega_N|\left(\mathbf{x}_k^t\right)_i\right)\right]^\top,\end{aligned} \tag{5.2}$$

where

$$a_j = \left\langle \min\left\{\left|s_l\left(\omega_j|\left(\mathbf{x}_k^t\right)_i\right)\right|, \left|s_r\left(\omega_j|\left(\mathbf{x}_k^t\right)_i\right)\right|\right\} - \epsilon\right\rangle. \tag{5.3}$$

In the equation, $\langle\cdot\rangle$ is Macaulay brackets, and $\min\{\cdot,\cdot\}$ returns the smaller value of the two entities. When the ILD vector is created with a set of frequencies $\tilde{\mathbf{\Omega}}$ at n known target positions, i.e., $\left(\tilde{\mathbf{x}}_k^t\right)_i, \forall i \in \{1, ..., n\}$, the acoustic cues to be prepared in advance and used to create the acoustic observation likelihood become $\mathbf{S}\left(\tilde{\mathbf{\Omega}}|\left(\tilde{\mathbf{x}}_k^t\right)_i\right), \forall i \in \{1, ..., n\}$.

Given the ILD vector $\mathbf{S}\left(\tilde{\mathbf{\Omega}}|{}^s\tilde{\mathbf{z}}_k^t\right)$ with observation ${}^s\tilde{\mathbf{z}}_k^t$ at unknown target position $\mathbf{x}_k^t$, the proposed technique quantifies the degree of correlation of the ith ILD vector to that of the unknown target position as

$$\begin{aligned}&X\left(\left(\tilde{\mathbf{x}}_k^t\right)_i|{}^s\tilde{\mathbf{z}}_k^t\right)\\ &\quad= \frac{1}{2}\left\{\frac{\mathbf{S}\left(\tilde{\mathbf{\Omega}}|{}^s\tilde{\mathbf{z}}_k^t\right)^\top \mathbf{S}\left(\tilde{\mathbf{\Omega}}|\left(\tilde{\mathbf{x}}_k^t\right)_i\right)}{\left|\mathbf{S}\left(\tilde{\mathbf{\Omega}}|{}^s\tilde{\mathbf{z}}_k^t\right)\right|\left|\mathbf{S}\left(\tilde{\mathbf{\Omega}}|\left(\tilde{\mathbf{x}}_k^t\right)_i\right)\right|} - 1\right\}\end{aligned} \tag{5.4}$$

where $0 \leq X(\cdot) \leq 1$. The acoustic observation likelihood with the unknown target position

$\mathbf{x}_k^t$ can be finally calculated as

$$l_j^a\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}_k^s\right)=\sum_{i=1}^{n}\mu_i\left(\boldsymbol{\xi}_k^t\right)X\left(\left(\tilde{\mathbf{x}}_k^t\right)_i|^s\tilde{\mathbf{z}}_k^t\right), \tag{5.5}$$

where $\mu_i\left(\boldsymbol{\xi}_k^t\right)$ is a basis function of natural coordinates $\boldsymbol{\xi}_k^t$, which are transformed from $\mathbf{x}_k^t$. Based on the characteristics of the likelihood, the proposed technique uses the T-spline basis function such that $\mu_i\left(\boldsymbol{\xi}_k^t\right)$ in a T-mesh, for the two-dimensional parameter space case $\boldsymbol{\xi}_k^t=\left[\xi_k^t,\eta_k^t\right]^\top$, can be represented as

$$\mu_i\left(\boldsymbol{\xi}_k^t\right)=\mu_i^1\left(\xi_k^t\right)\mu_i^2\left(\eta_k^t\right) \tag{5.6}$$

where $\mu_i^{(\cdot)}$ is a cubic B-spline basis function. Further detailed formulations are found in [146]. It is to be noted here that other basis functions are also possible while the proposed technique uses T-spline basis functions.

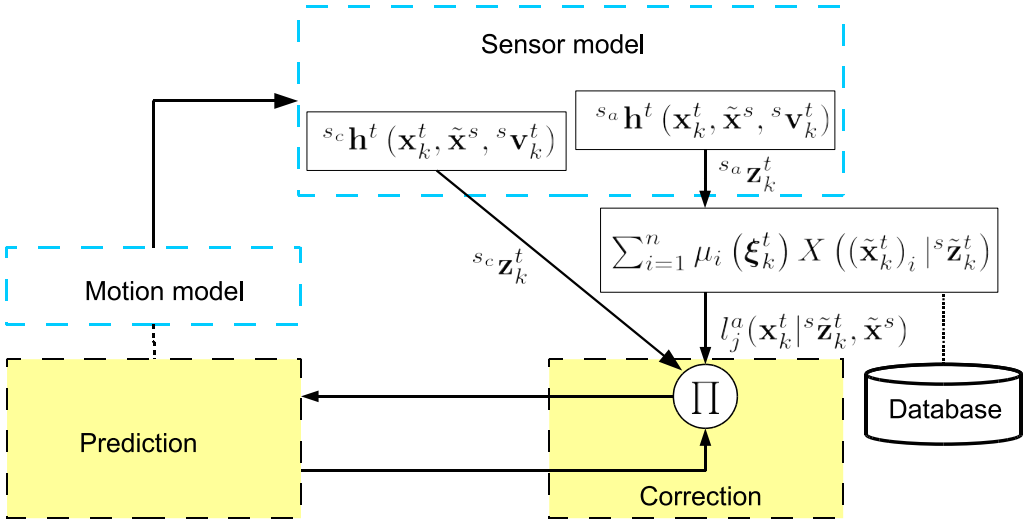


Figure 5.2: RBE framework schematics with optical and acoustical sensors

5.2 Numerical and Experimental Analysis

The efficacy of the proposed approach was examined experimentally in two steps. The first step was aimed at studying the capability and limitation of the proposed acoustic sensing technique. The study was conducted by parametrically changing the complexity of the environment where the experimental system, with a speaker array and a movable/replaceable wall, was developed specifically for this study. After verifying the feasibility of the acoustic sensing for NLOS target localization, the applicability of the proposed approach to the estimation of a NFOV target in a complex practical environment was investigated. The investigation looked into the performance of both the joint optical/acoustic observation likelihood and the RBE with the joint likelihood.

5.3 NFOV Target Estimation in Complex Practical Environments

This subsection investigates the applicability of the proposed approach to NFOV target estimation in complex practical environments by first enhancing the acoustic sensing with the optical sensing and then executing the RBE with the joint optical/acoustic observation likelihood. Figure 5.3 shows the map of the indoor environment used to demonstrate the practical applicability of the proposed approach, as well as the details of the demonstration. A sensor platform, with a camera and two microphones for optical and acoustic observations, was located in a corridor and faced with the open space so the target could move around. The environment is complex, containing walls and structures, so that the FOV of the camera is significantly limited compared to the target space. Shown in the figure as yellow crosses are the target locations at which sound was emitted to collect ILDs. After the collection, the target was then moved along the lines indicated in the figure, and it emitted sound. The observation and estimation was examined at the four

positions marked by red circles.

Figure 5.4 shows the target and the sensor platform. The sensor platform is with a camera and two microphones, as aforementioned, whereas the target is a wireless speaker. The same speaker was used to construct the acoustic observation likelihood. White noise was emitted from the speaker, and the parameters used to construct the acoustic observation likelihood are listed in Table 5.1.

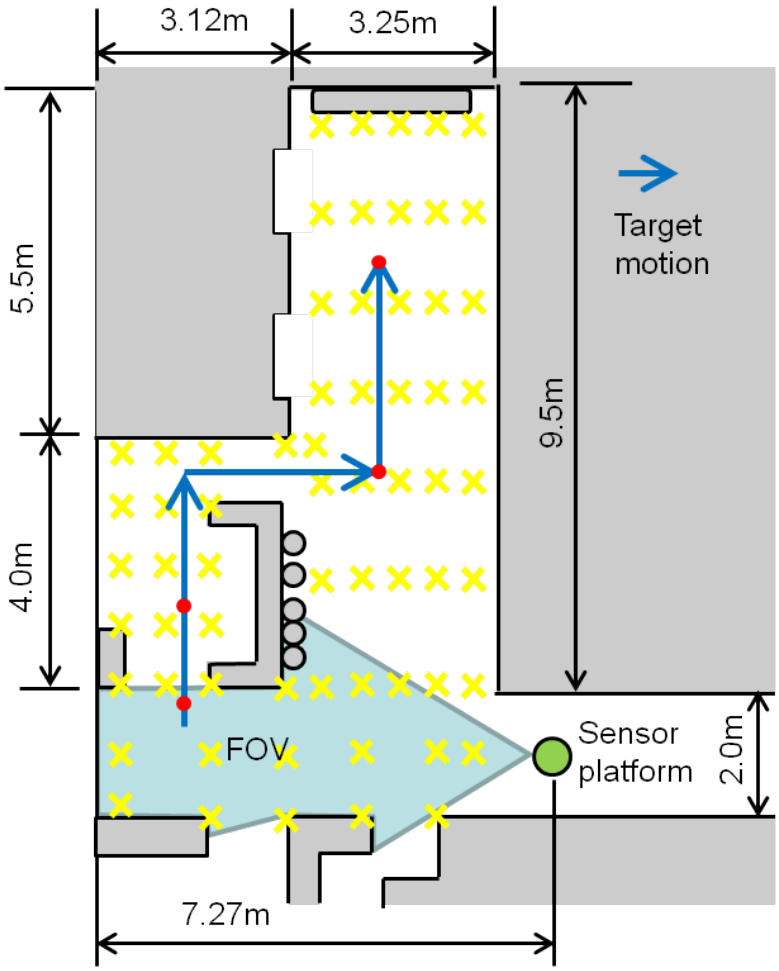


Figure 5.3: Map of the test environment and demonstration

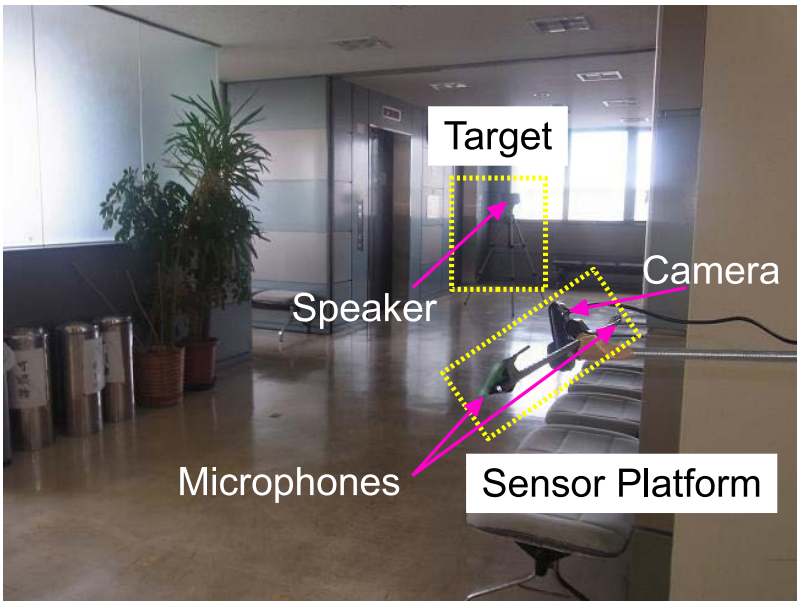


Figure 5.4: Target and sensor platform

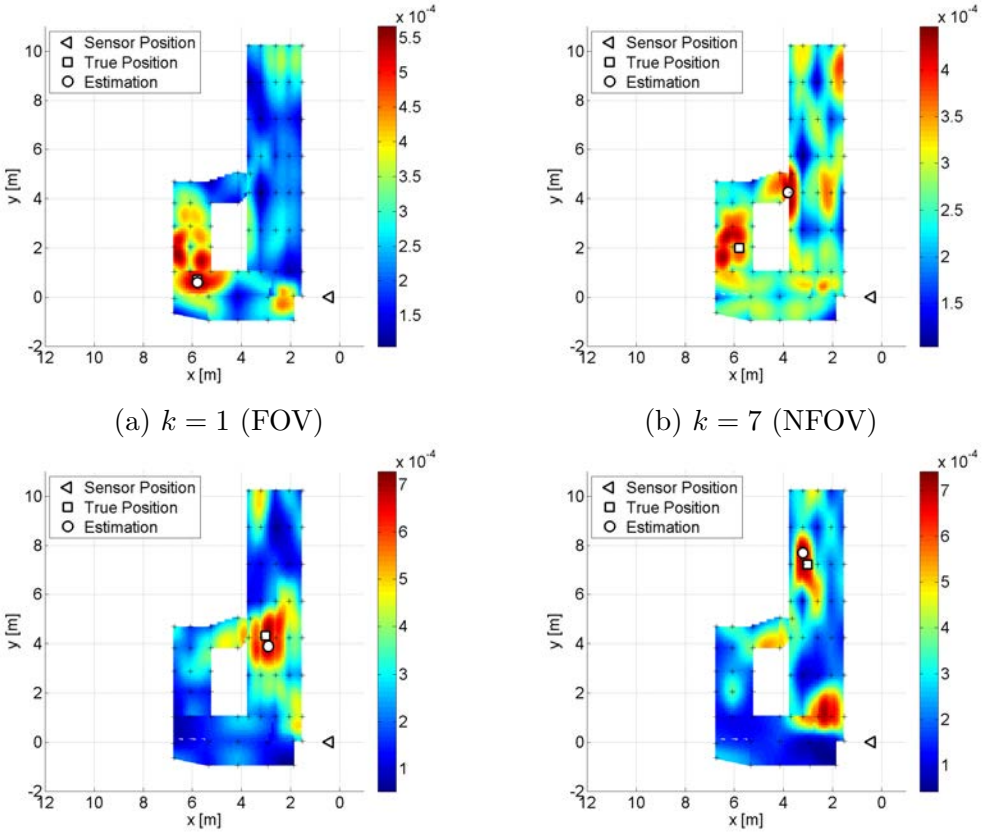
Table 5.1: Parameters for acoustic observation likelihood

Parameter	Value
ω_1	0 [Hz]
ω_N	386 [Hz]
N	100
ϵ	0.01
n	65

5.3.1 Joint Optical/Acoustic Observation of NFOV Target

Figure 5.5 shows the acoustic observation likelihoods when the target moved and emitted sound at the four marked target positions, which are at the 1st, 7th, 35th and 51st steps. The target is in the FOV only at the 1st step. The acoustic observation likelihood is seen to be multi-modal due to sound reflection even when the target is within the FOV and thus in the LOS. The proposed acoustic sensing technique has, however, been able to accurately identify the true target position at one of the peaks and successfully detect it in every step but the 7th step. Failure in the 7th step is a result of the limitation of

acoustic sensing shown and concluded in the last subsection, but it is to be importantly noted that the target position is captured near the second highest peak.



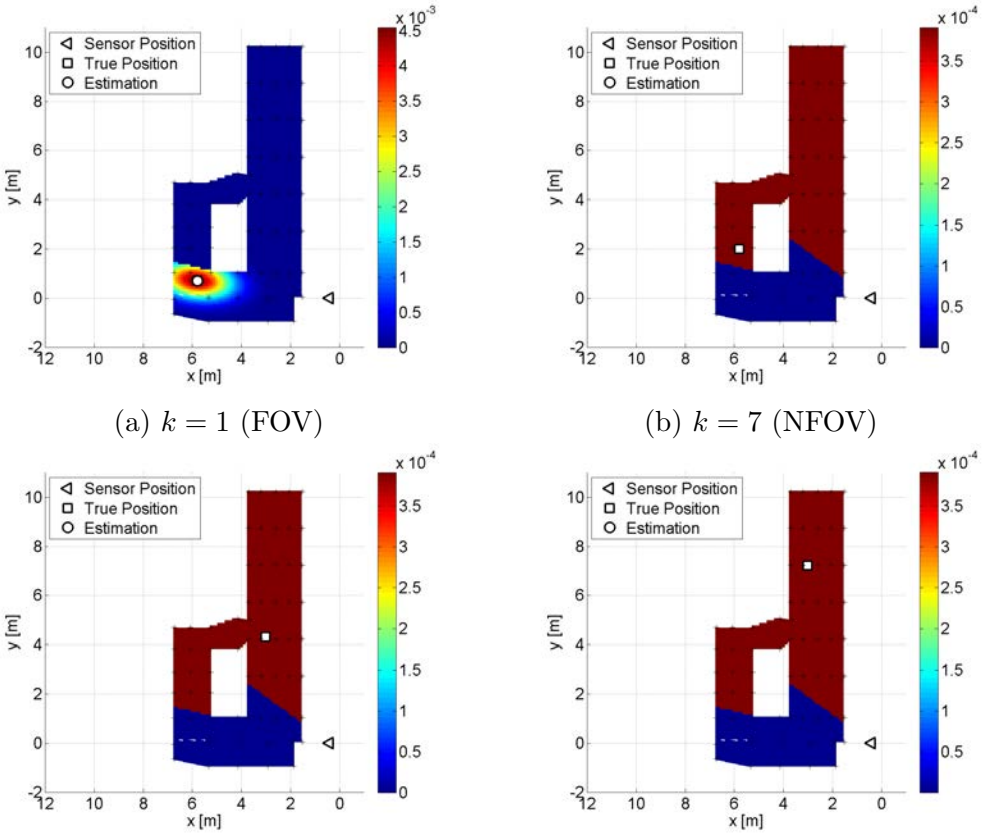
(a) $k = 1$ (FOV) (b) $k = 7$ (NFOV)

(c) $k = 35$ (NFOV) (d) $k = 51$ (NFOV)

Figure 5.5: Acoustic observation likelihoods

The effectiveness of the proposed acoustic sensing technique is further understood comparatively by seeing optical observation likelihoods in Figure 5.6. The optical sensor can identify the target accurately when it is within the FOV. The observation likelihood with the target outside the FOV can however provide no positive information on the target. Finally, Figure 5.7 shows the joint optical/acoustic observation likelihoods. The joint observation likelihoods narrow down the possible target locations the most by detecting the target dominantly with the optical observation likelihood, when the target is within the FOV, and with the acoustic observation likelihood, when the target is outside the

FOV. The wrong computation at the 7th step, however, remains, and necessitates RBE for target estimation.



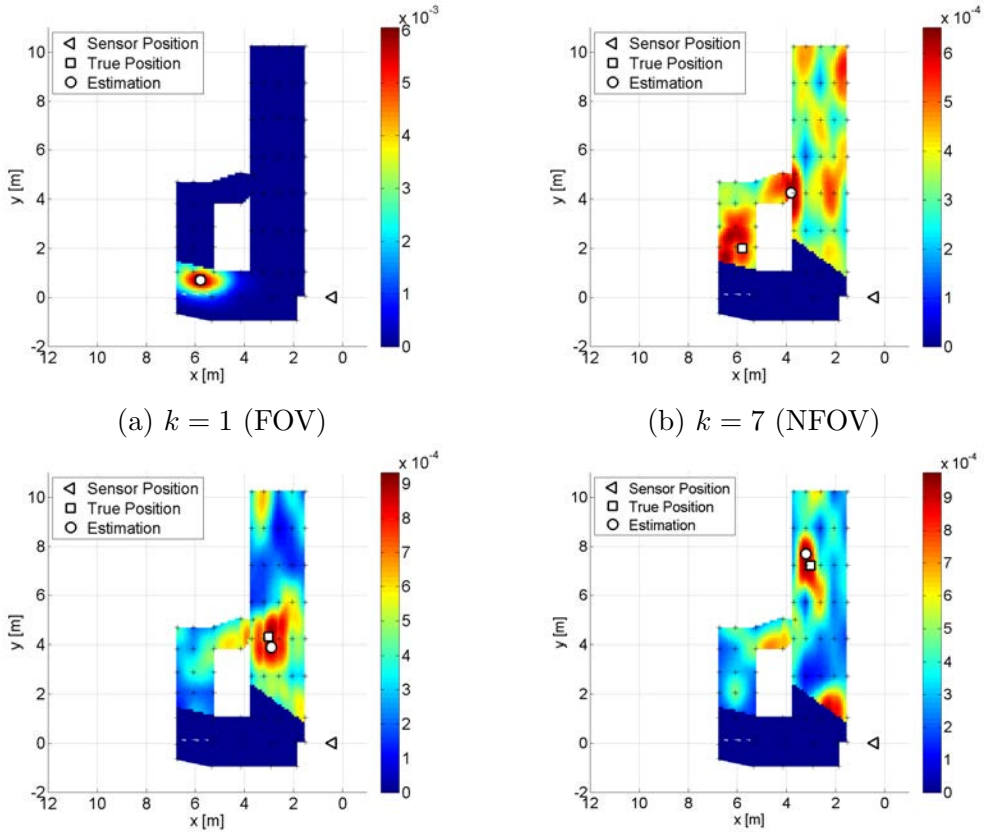
(a) $k = 1$ (FOV) (b) $k = 7$ (NFOV)

(c) $k = 35$ (NFOV) (d) $k = 51$ (NFOV)

Figure 5.6: Optical observation likelihoods

5.3.2 RBE with Joint Optical/Acoustic Observation Likelihoods

Having understood the limitation of the target detection with observations in the last section, the effectiveness of the proposed RBE with the joint optical/acoustic observation likelihoods was investigated using the same test data. Without a good understanding of the target motion, the target motion model was given by a random walk model, assuming that the target is a human who could move to any direction with equal probability.



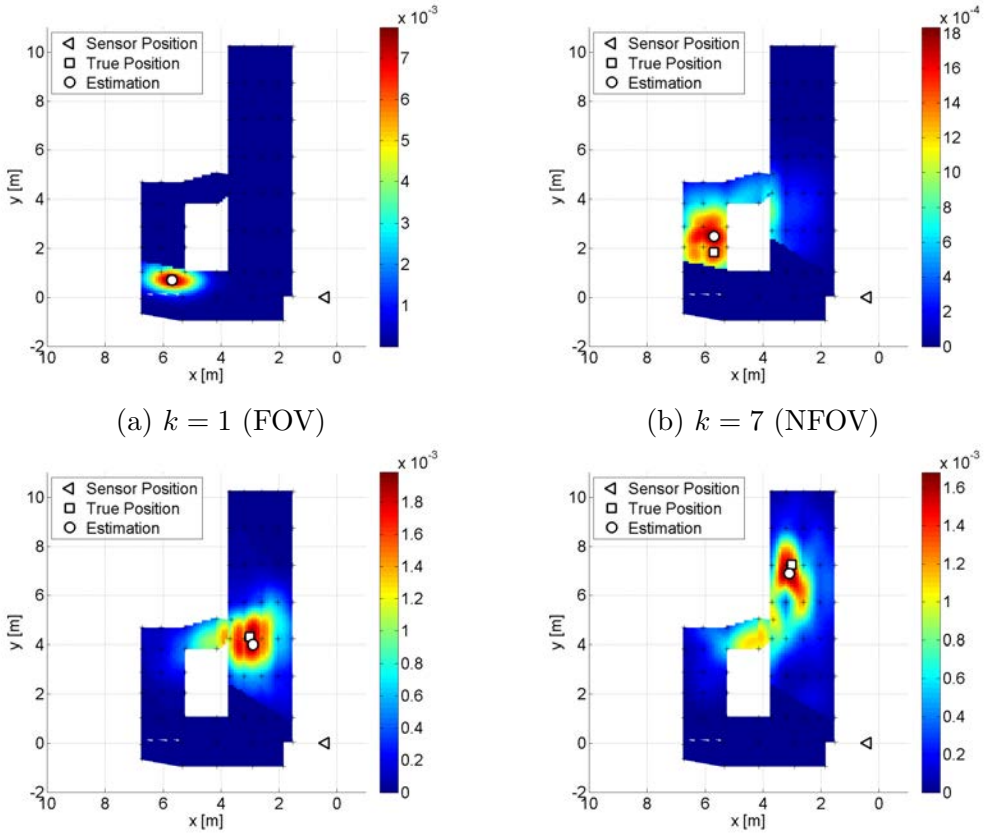
(a) $k = 1$ (FOV) (b) $k = 7$ (NFOV)

(c) $k = 35$ (NFOV) (d) $k = 51$ (NFOV)

Figure 5.7: Joint optical/acoustic observation likelihoods

Figure 5.8 shows the target belief estimated via RBE with the joint optical/acoustic observation likelihoods. The result shows that the target position is well estimated during all the time steps, including the 7th step, where the joint observation likelihood did not detect the target with the highest peak. Because the target was initially observed by the optical sensor, strong and accurate prior belief was constructed. Since the prior knowledge is updated by prediction with the random walk model and correction with the joint observation likelihood, the proposed approach can eliminate wrong detections and estimate the target position near the true position.

Figure 5.9 shows the target belief estimated conventionally with only the optical observation likelihoods to comparatively verify the effectiveness of the proposed approach.



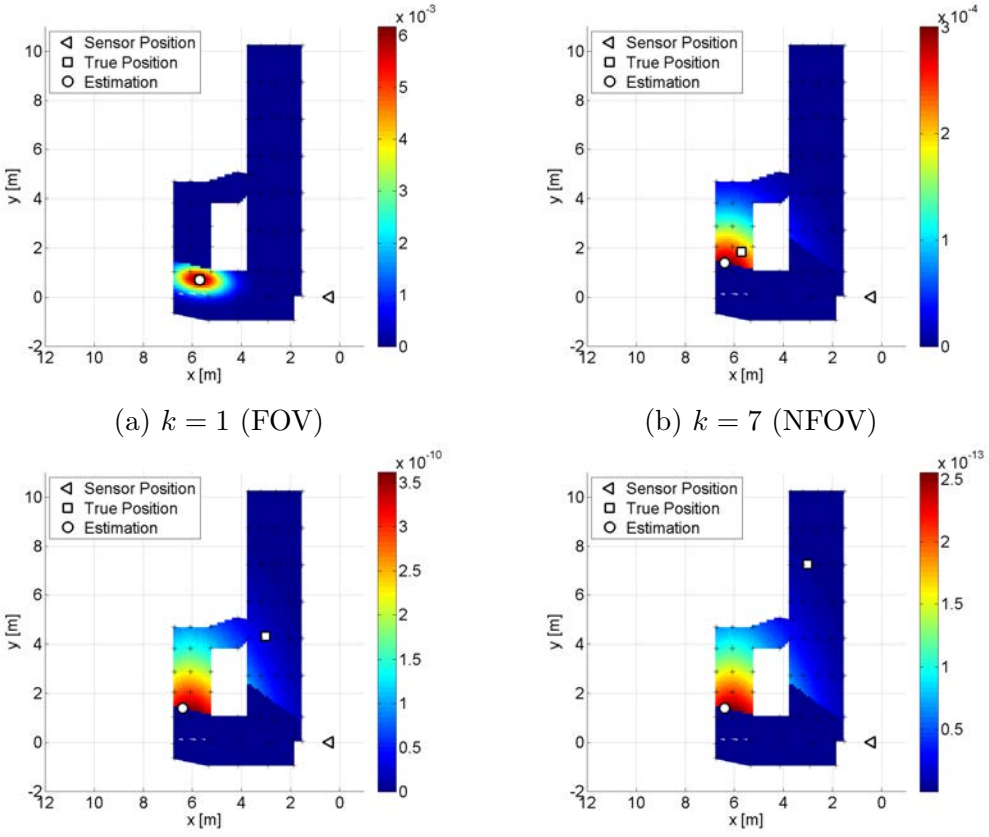
(a) $k = 1$ (FOV) (b) $k = 7$ (NFOV)

(c) $k = 35$ (NFOV) (d) $k = 51$ (NFOV)

Figure 5.8: Proposed optical/acoustic target estimation

The result with the optical observation likelihoods is seen to estimate the target position wrongly when the target is in the NFOV. Because an inaccurate random-walk motion model is used, the target is estimated continuously at the location where it was lost. Finally, Figure 5.10 shows the results quantitatively evaluating the performance of the proposed approach. Figure 5.10a shows the transition of the mean error of the estimated target position from the true position whereas the transition of the Kullback-Leibler (KL) divergence is exhibited in Figure 5.10b. The error transition indicates that the proposed approach maintains error within $1m$. This is the case even when the target has not been lost from the FOV for some time whilst the conventional RBE with optical observation likelihoods increases the error with high gradient. The KL divergence transition also shows

this behavior, indicating that the proposed approach maintains target information with the use of the acoustic sensor.



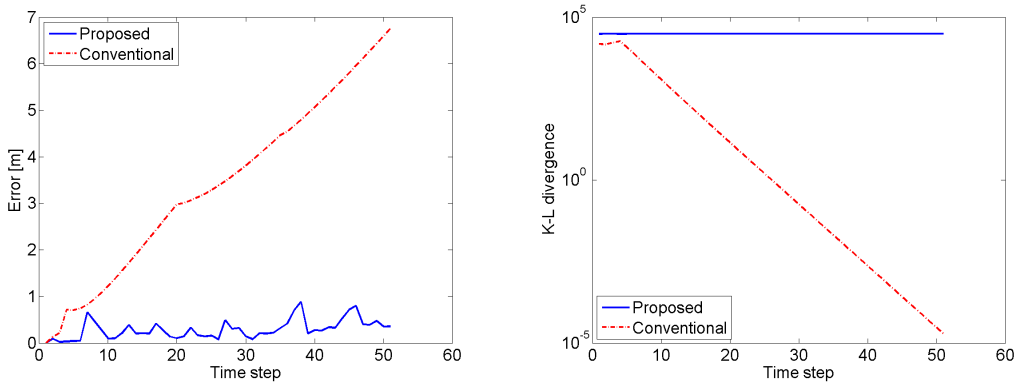
(a) $k = 1$ (FOV) (b) $k = 7$ (NFOV)

(c) $k = 35$ (NFOV) (d) $k = 51$ (NFOV)

Figure 5.9: Conventional optical target estimation

5.4 Chapter Summary

This section has presented a NFOV target estimation approach which incorporates optical and acoustic sensors. The proposed approach performs RBE with joint optical/acoustic observation likelihoods. Although the acoustic observation likelihood could be multi-modal with high uncertainty, the target belief updated in the past with sharply unimodal optical likelihoods effectively acts as strong prior knowledge and enables accurate NFOV target



(a) Error vs time step (b) KL divergence vs time step

Figure 5.10: Quantitative analysis

estimation. A technique to construct an observation likelihood using an acoustic sensor composed of two microphones has also been proposed. The acoustic observation likelihood is created by correlating the ILD of the new acoustic observation with the ILDs collected in advance.

The first experiment studied the capability and limitation of the proposed acoustic sensing technique by parametrically changing the complexity of the environment. It has been found that the proposed technique can identify the location of a NLOS target, but with limited accuracy, particularly when the complexity of the environment is severe. The applicability of the proposed RBE approach with joint optical/acoustic observation likelihoods to the estimation of a NFOV target in a complex practical environment was the investigated as the second experiment. The result shows that the target position is well estimated at all the time steps, even when the joint observation likelihood does not identify the target location well due to the use of prediction with a motion model, and strong prior belief constructed with the past optical observation.

The approach has demonstrated the new concept for NFOV target estimation, and many challenges are still open for future study. One of the improvements to the approach is the enhancement of acoustic sensing by incorporating the Interaural time difference

(ITD) and the Interaural phase difference (IPD), as well as the use of non-white noise sound so that the approach could be used for various applications. Other ongoing work includes the implementations of the proposed approach to the mobile sensor platform and to the infrastructure. The proposed approach can be used for various practical applications including home security, home heath care and urban search-and-rescue by the extension.

CHAPTER 6

Reflection and Diffraction based approach

This section presents a new acoustic approach to locate a mobile target outside the FOV, or the NFOV of an optical sensor, based on the reflection and diffraction signals. With an increase in coexisting environments of robots and humans, it is crucial to extend the approach towards human-robot interaction. In this approach, a sensor platform determines a reflection TDOA and frequency dependent diffraction resulting two distinct observation likelihoods from a single target's sound. The fusion of these likelihoods, a joint acoustic observation likelihood, estimates the NFOV target probabilistically within the RBE framework. The approach was formulated and derived mathematically. Through parametric studies in simulation, the potential of the proposed approach for practical implementation has been demonstrated by the successful

localization of the sound source. Finally, a preliminary validation of sound separation was performed in a controlled experimental environment showing the difference between diffraction alone and combination of diffraction and reflection signals.

6.1 Introduction

For a numerical technique, existing approaches enhance NFOV target estimation by including a sensor with a limited FOV, such as an optical sensor. In the previous section, which discussed a generalized numerical solution technique, the event of "no detection" is converted into an observation likelihood and utilized to positively update probabilistic belief on the target. This belief is dynamically maintained by the RBE. The technique, however, has been found to fail in target estimation unless the target is re-discovered within a short period after being lost. The hybrid method incorporated an acoustic sensor to maintain belief with no optical detection and more reliability. Nevertheless, the technique performed poorly unless the target re-entered the optical FOV since the acoustic sensing is only conducted in an assistive capacity. The fixed microphone array approach which focused more towards a complex indoor environment using ILD *a priori* fixed microphone array knowledge. However, this approach requires prior data collection and can not be applied to the dynamic sensor platforms.

This section presents a new acoustic approach to estimate a NFOV mobile target using a sound wave's physical properties, reflection and diffraction. In the approach, a sound source reflection and diffraction signal TDOA, and a frequency dependent property, construct two distinct observation likelihoods from target sound. The fusion of these likelihoods, a joint acoustic observation likelihood, is derived to perform the target estimation. The location of the NFOV target is finally estimated within the recursive Bayesian estimation framework. This process of target estimation was derived mathematically. Following the formulation, the proposed approach was tested through simulation and parametrically

studied under multiple conditions. Finally, an experimental environment was constructed to identify the distinct differences between a diffraction signal alone, and a combination of diffraction and reflection signals.

6.2 Non-Field-of-View (NFOV)

6.2.1 Auditory Recursive Bayesian Estimation

The proposed approach is mathematically described as follows. Let the state of the robot s at time step k be $\bar{\mathbf{x}}_k^s \in \mathcal{X}^s$. Consider a target t, the state is given by $\mathbf{x}_k^t \in \mathcal{X}^t$, and a sequence of observations of the target t by the robot s from time step 1 to time step k given by ${}^s\tilde{\mathbf{z}}_{1:k}^t \equiv \left\{ {}^s\tilde{\mathbf{z}}_\kappa^t | \forall \kappa \in \{1, ..., k\} \right\}$. The RBE represents belief on the target in the form of a probability density function and iteratively updates the belief in time and observation. Let the belief, given a sequence of observations and the robot state at time step $k-1$, be $p\left(\mathbf{x}_{k-1}^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \bar{\mathbf{x}}_{k-1}^s\right)$. The Chapman-Kolmogrov equation updates the prior belief in time, or predicts the belief at time step k, by the probabilistic motion model $p\left(\mathbf{x}_k^t | \mathbf{x}_{k-1}^t, \bar{\mathbf{x}}_{k-1}^s\right)$:

$$p\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \bar{\mathbf{x}}_{k-1}^s\right) = \int_{\mathcal{X}^t} p\left(\mathbf{x}_k^t | \mathbf{x}_{k-1}^t, \bar{\mathbf{x}}_{k-1}^s\right) p\left(\mathbf{x}_{k-1}^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \bar{\mathbf{x}}_{k-1}^s\right) d\mathbf{x}_{k-1}^t. \quad (6.1)$$

Note that the motion model is $p\left(\mathbf{x}_k^t | \mathbf{x}_{k-1}^t\right)$ if the target is not reactive to the robot. The observation update, or the correction process, is performed using the Bayes theorem. The target belief is corrected using the new observation ${}^s\tilde{\mathbf{z}}_k^t$ as

$$p\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k}^t, \bar{\mathbf{x}}_k^s\right) = \frac{q\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k}^t, \bar{\mathbf{x}}_{k-1:k}^s\right)}{\int_{\mathcal{X}^t} q\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k}^t, \bar{\mathbf{x}}_{k-1:k}^s\right) d\mathbf{x}_k^t}, \quad (6.2)$$

where $q(\cdot) = l\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t\right) p\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \bar{\mathbf{x}}_{k-1}^s\right)$ and $l\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s\right)$ represents the observation likelihood of $\mathbf{x}_k^t$ given ${}^s\tilde{\mathbf{z}}_k^t$, $\bar{\mathbf{x}}_k^s$.

$$l\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s\right) = \prod_j l_j^a\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s\right) \tag{6.3}$$

where $l_j^a(\cdot)$ are the likelihoods of jth acoustic sensor.

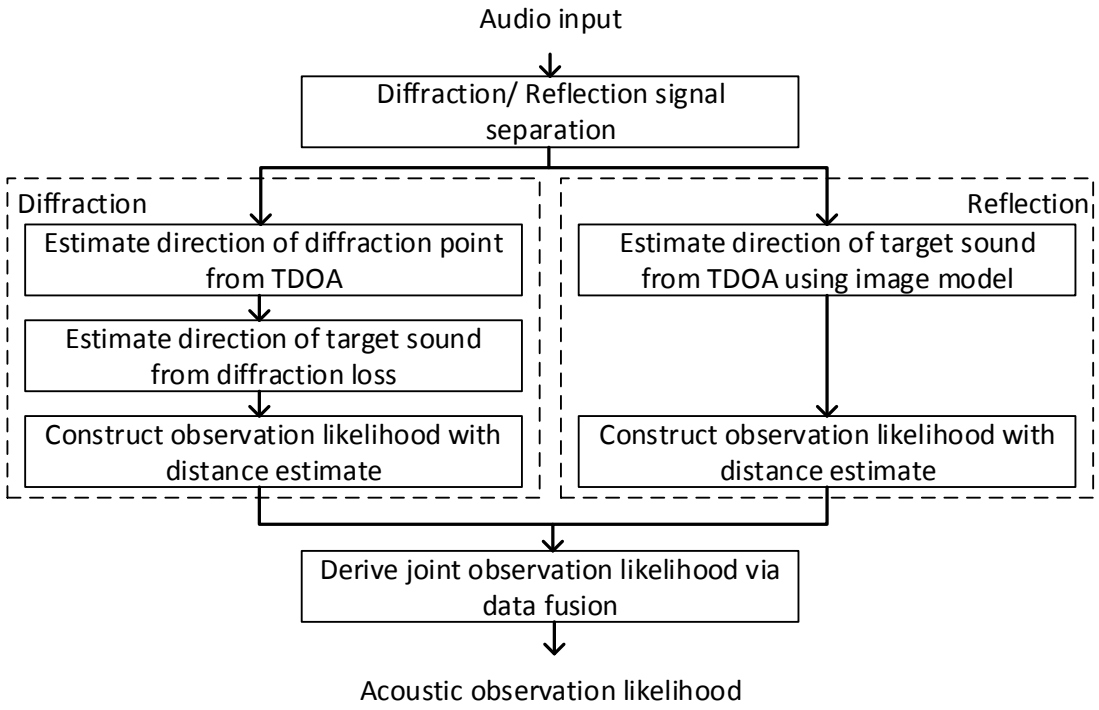


Figure 6.1: Construction of auditory NFOV target likelihood

6.2.2 Construction of Auditory NFOV Target Observation Likelihood

Figure 6.1 shows the overview of the approach proposed for constructing an NFOV target observation likelihood using auditory sensors. The proposed approach extracts the first-arrival diffraction and reflection signals by taking the wave propagation physical properties into account. The approach begins with obtaining a time-domain signal of a relatively

impulsive sound at each microphone. In each curve, notable peaks are then extracted as candidates for first-arrival diffraction and reflection signals. When each candidate signal is described in the frequency domain, the first-arrival diffraction and reflection signals can be identified since they are the first signals correlated in the low-frequency range. The diffraction signal is then used to identify the so-called diffraction point by deriving the TDOA for each pair of microphones. Beyond the diffraction point, the target direction is estimated by using the property of sound energy loss through diffraction, or the diffraction loss. An observation likelihood is eventually constructed by additionally estimating the distance from the sound magnitude and features. The reflection signal estimates the target direction directly from the TDOA by mirroring and creating a virtual target. It also creates an observation likelihood with a distance estimate by considering the sound magnitude and characteristics and environmental properties. A joint observation likelihood is finally created by the fusion of the diffraction and reflection observation likelihoods.

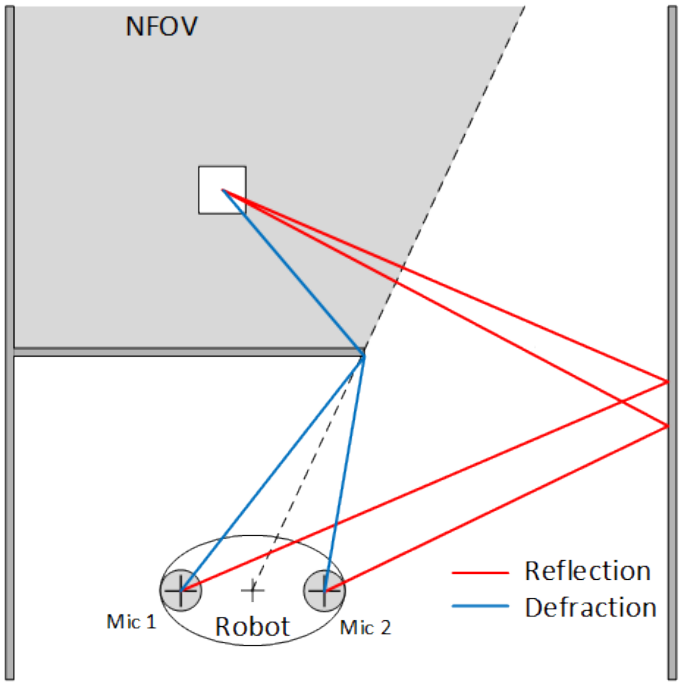


Figure 6.2: Auditory NFOV target observation

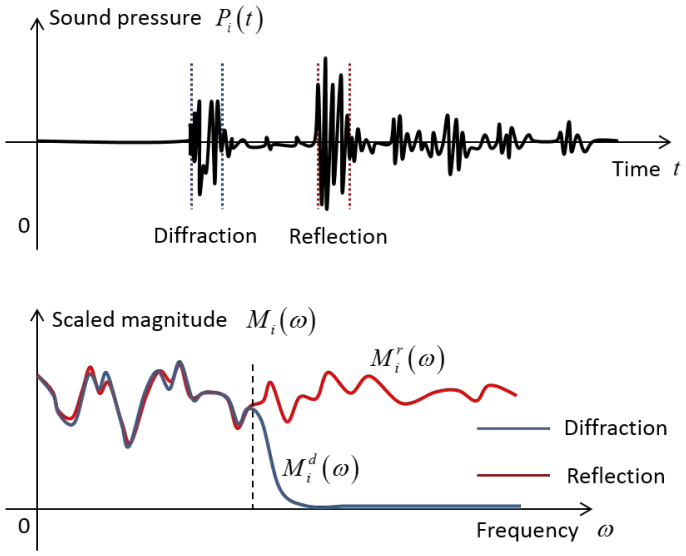


Figure 6.3: Auditory NFOV target observation: Diffraction and Reflection

The proposed approach infers the location of the sound target using both the first-arrival diffracted and reflected sound signals. The next subsection describes the extraction of the first-arrival diffraction and reflection signals, followed by the target estimation, using the diffracted and reflected signals, in the subsequent two subsections. The final goal of this project is to develop a probabilistic RBE-based framework, but the preliminary study has succeeded in the proof-of-concept in deterministic formulations. The two subsections will present the deterministic NFOV target estimation using diffraction and reflection sound waves. The final subsection derives the joint observation likelihood, as a result of data fusion.

6.3 Extraction of First-arrival Diffraction and Reflection Signals

Figure 6.3 illustrates the extraction process of the diffraction and reflection signals in a simple scenario, where a robot carrying two microphones receives sound emitted by an NFOV

target in a two-dimensional indoor environment with three walls (Figure 6.2). As shown in the figure, sound waves emitted from the target reach the robot first through diffraction and second through reflection. If the sound is relatively impulsive, the first-arrival diffraction and reflection signals can be extracted clearly. The assumption of the sound specular reflection holds for wall textures that are smooth compared to the wavelength. Figure 6.3 shows not only the sound pressure in the time domain, $P_i(t)$, but also the magnitude of the resulting first-arrival diffraction and reflection signals in the frequency domain, $M_i^d(\omega)$ and $M_i^r(\omega)$, where $i \in \{1, 2\}$ is the index of microphone. Note that the magnitude is scaled to examine the correlation. Signals are considered to be from the same sound source if they share the same low frequency characteristics, because low-frequency signals reflect and diffract. The proposed approach thus selects the first set of signals that have the same low-frequency characteristics, but are dissimilar in high frequency as the first-arrival diffraction and reflection signals of all candidate signals. Diffraction signals have little high-frequency components, whilst reflection signals see components in all frequencies. Each microphone, 1 and 2, constructs a different data set.

6.3.1 Estimation of Sound Direction from Diffraction Signals

Figure 6.4 shows the notations used for target estimation from diffraction signals in the scenario introduced in the last subsection. The diffraction signal that microphones 1 and 2 receive originated from the LOS location at which the sound diffracts. The proposed approach starts target estimation from diffraction signals with the selection of diffraction points from all candidates, which are corners of all structures. The measured quantity used for the selection is the TDOA, $\Delta t_d = t_{d2} - t_{d1}$, where t_{d1} and t_{d2} are the TOAs at Microphones 1 and 2 respectively. The diffraction point can be easily found from

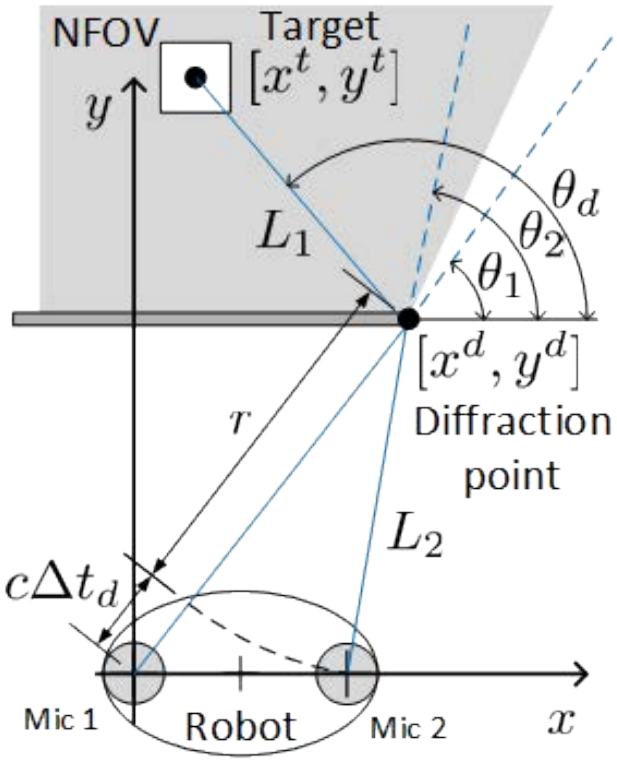


Figure 6.4: Proposed approach

candidates, as it satisfies the following equation:

$$\left(x^d\right)^2 + \left(y^d\right)^2 = (c\Delta t_d + r)^2 . \tag{6.4}$$

where $\left[x^d, y^d\right]$ is the location of a candidate diffraction point, c is the speed of sound and r is a shorter distance between a microphone and the candidate diffraction point. With the diffraction point identified, the proposed approach further identifies the direction of the sound target from the diffraction point by analyzing the magnitudes of diffraction and reflection signals $M_i^d(\omega)$ and $M_i^r(\omega)$. The loss of high-frequency signal components is assumed to be less with a microphone closer to the LOS, microphone 2 in this case, as there is no loss with a microphone in the LOS of the sound target. Medwin [2] proved the validity of this assumption over a quarter of a century ago, as shown in Fig. 6.5.

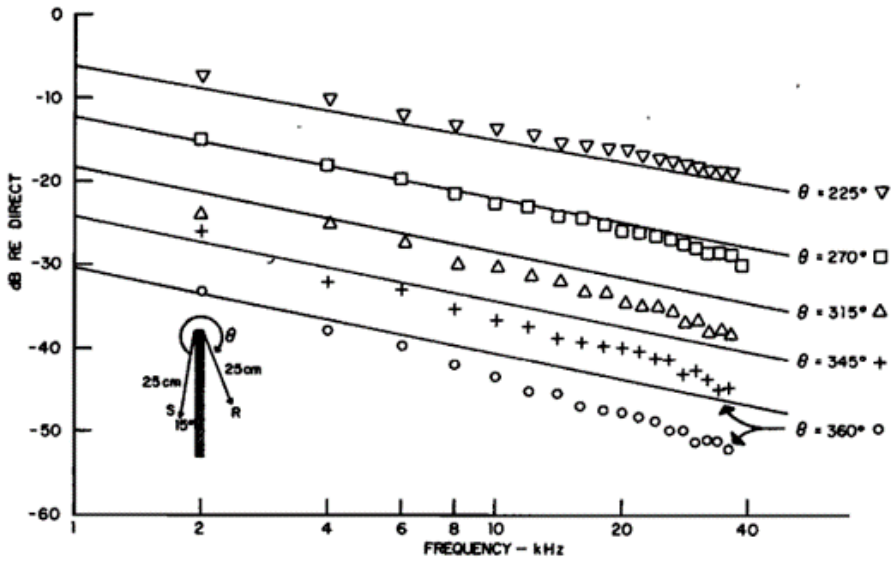


Figure 6.5: Magnitude with different orientation angles [2]

The magnitude of diffraction sound drops when the "degree of NLOS" represented by the orientation angle is increased. This makes the proposed approach define the diffraction loss as

$$DL_i = \int \left[M_i^r(\omega) - M_i^d(\omega) \right] \mathrm{d}\omega \geq 0, \forall i \in \{1, 2\} \tag{6.5}$$

and associate it with the degree of NLOS. The work of Medwin also proved that the diffraction loss is approximately proportional to the degree of NLOS. The sound direction from the diffraction point is given by

$$\theta_d = f(\theta_1, \theta_2 | DL). \tag{6.6}$$

6.3.2 Estimation of Sound Direction from Reflection Signals

Figure 6.6 shows the proposed approach for the estimation of sound direction from reflection signals. Reflection makes the sound propagation and the subsequent target estimation complicated, but if the wall is smooth and yields specular reflection, the sound direction can be estimated easily by introducing a virtual target [149], which is located symmetri-

cally to the real target, relative to the wall of reflection. Let the position of the virtual target be $[\hat{x}^t, \hat{y}^2]$. The measured TDOA can be associated with the position of the virtual target as

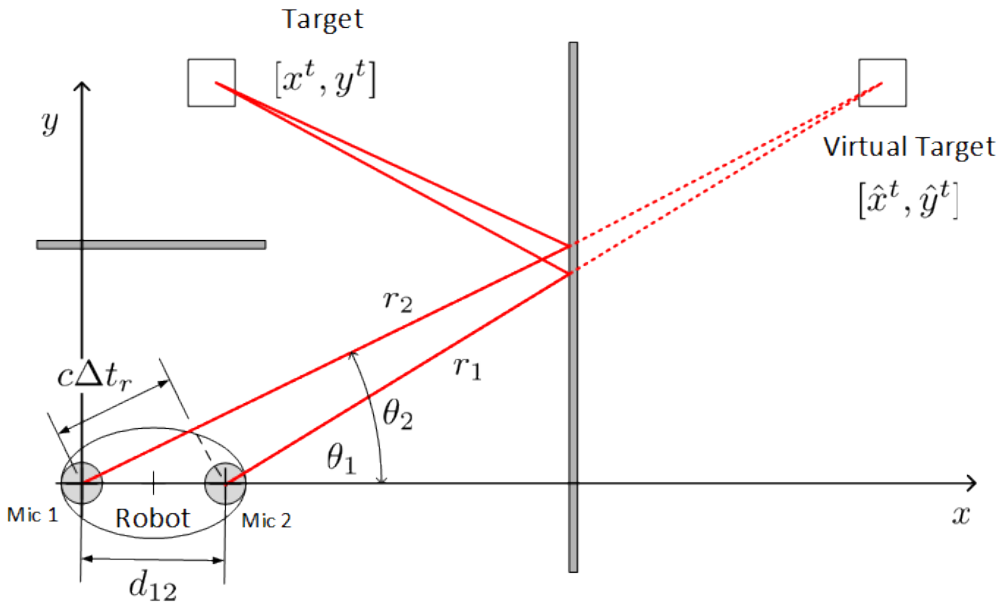


Figure 6.6: Estimation of sound direction from reflection signals by proposed approach

$$\begin{cases} c\Delta t_r = \Delta r_{12} = \|r_1 - r_2\| \\ \|r_1\| \sin\theta_1 = \|r_2\| \sin\theta_2 \\ \|r_2\| \cos\theta_2 - \|r_1\| \cos\theta_1 = d_{12} \end{cases}, \quad (6.7)$$

where d_{12}, $\{r_1, r_2\}$ are the distance between Microphones 1 and 2 and vectors from sensor to the virtual sound source, respectively. Derivation attempted as a preliminary study for this project yields the relationship as a TDOA curve expressing r_1 as a function of θ_1

$$\|r_1(\theta_1)\| = \frac{-d_{12}^2 + \Delta r_{12}^2}{2(\Delta r_{12} + d_{12}\cos(\theta_1))}. \quad (6.8)$$

Further mathematical manipulation shows that this equation asymptotically yields the sound direction as

$$\frac{\theta_2 + \theta_1}{2} = \lim_{\|r_1\| \to \infty} \tan^{-1} \frac{\hat{y}^t}{\hat{x}^t} = \cos^{-1} \frac{c\Delta t_d}{d_{12}}. \tag{6.9}$$

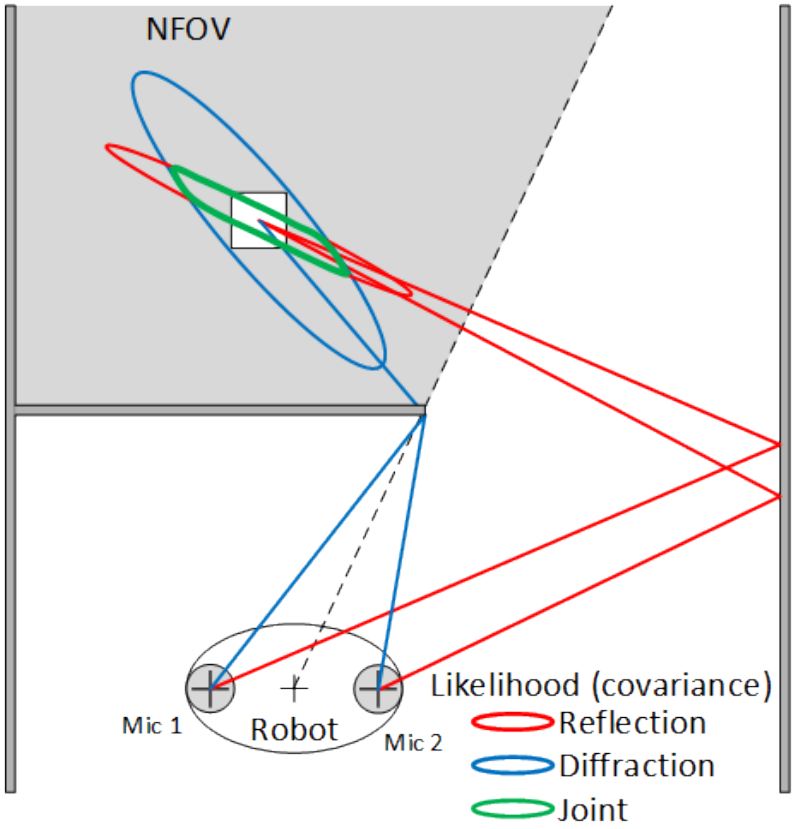


Figure 6.7: Construction of joint observation likelihood through data fusion

6.3.3 Construction of Joint Observation Likelihood through Data Fusion

While the sound can be better identified in direction rather than distance, it is also possible to make an estimate as to how far away the sound target is. The proposed approach estimates the positions by utilizing any available information including the magnitude, sound

patterns, or sound features stored in a knowledge base, and constructs an observation likelihood for each of the diffraction and reflection signals by modeling uncertainties. For the j-th pair of microphones, the diffraction and reflection likelihoods are then combined to create an auditory joint observation likelihood via the canonical data fusion formula:

$$l_j^a\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t,\bar{\mathbf{x}}_k^s,\bar{\mathbf{m}}_k\right)=l_j^d\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t,\bar{\mathbf{x}}_k^s,\bar{\mathbf{m}}_k\right)l_j^r\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t,\bar{\mathbf{x}}_k^s,\bar{\mathbf{m}}_k\right) \tag{6.10}$$

where $l_j^d(\cdot)$ and $l_j^r(\cdot)$ are the diffraction and reflection observation likelihoods. Figure 6.7 illustrates the diffraction and reflection observation likelihoods, as well as the joint observation likelihood, where the observation likelihood is represented by an ellipsoid indicating a probability distribution with a covariance. The diffraction and reflection likelihoods are shown to have high eccentricity due to more accuracy in direction than in distance. Since the difference of the diffraction and reflection likelihoods in orientation may not be significant, the resulting auditory joint likelihood is also given by an ellipsoid with high eccentricity. However, the proposed approach, utilizing the diffraction and reflection physics of sound, could estimate the location of the sound target.

6.4 Numerical Analysis

The preliminary simulation experiments demonstrate the validity of the proposed approach. The experiments were set up with the parameters in Table 6.1. Figure 6.8 shows an example of localization. The dashed line, square, circles and red cross are the reflective wall, target, microphone sensors and target estimation, respectably. As shown in the figure, diffraction and reflection observation likelihoods construct the green lined joint likelihood to estimate the target location under the NFOV condition. Based on the initial simulation, the parametric study was performed to measure the sensitivity of estimation based on the sensor distance d_{12} in a range of 1 cm to 50 cm. The sensor position was also varied at three different locations. Figure 6.9 shows the certainty of the target es-

timation as KL-Divergence at a different sensor location with increasing d_{12}. As shown in the figure, increasing the distance between the microphones leads to a better estimate of the target. This is expected because increasing d_{12} will increase the time of arrival difference, resulting in a better angle estimation. However, increasing the sensor position in the y-direction leads to a reduction in KL-divergence.

Table 6.1: Dimensions and parameters for the simulation

Parameter	Value
x^t	$[0.5m, 1.5m]$
$[x^d, y^d]$	$[1m, 1m]$
x_{wall}	$2m$
x_s	$[0.6m, 0.3m]$

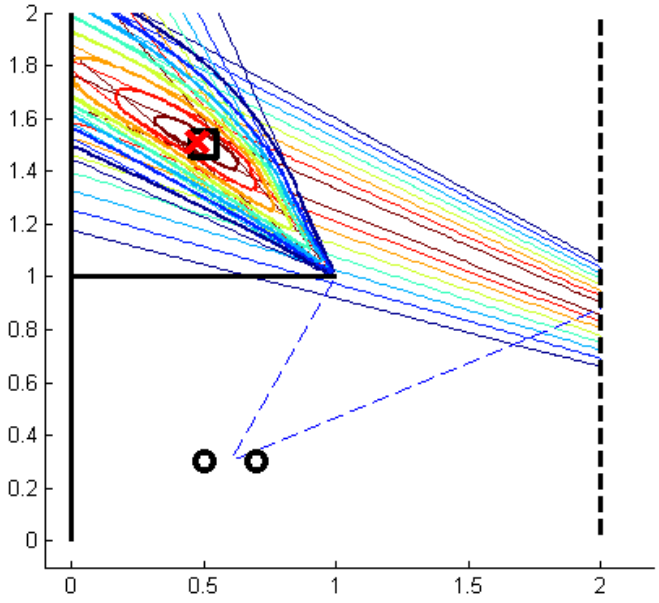


Figure 6.8: Target estimation

Figure 6.10 shows the experimental testing environment. As shown in the figure, the pink cylinder on top is the omni-directional target sound source, and the separator wall

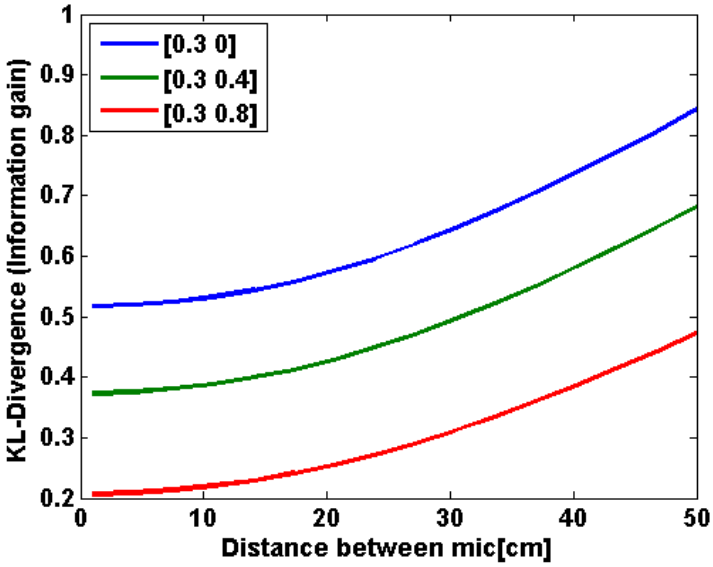


Figure 6.9: d_{12} variation effect in certainty

and microphone array are shown. The sound insulator was installed at the top and the bottom of the environment to mitigate additional reflection and minimize reverberation. A removable smooth reflective wall was used for introducing the reflection. The experiments included microphones in a circular pattern as shown in Fig. 6.11 with the dimensions in Table 6.2.

Table 6.2: Dimensions and parameters for the experiments

Parameter	Value
L_{wd}	62cm
r_t	$[-20\text{cm}, 25\text{cm}]$
$\|r\|$	35cm
θ	15°
L_{wr}	112cm

Figure 6.12 depicts incoming sound to the microphone array without the reflection wall. The figure consists of a reference microphone at the target and the microphone array signal. As shown in the figure, the diffraction loss increases when the angle of

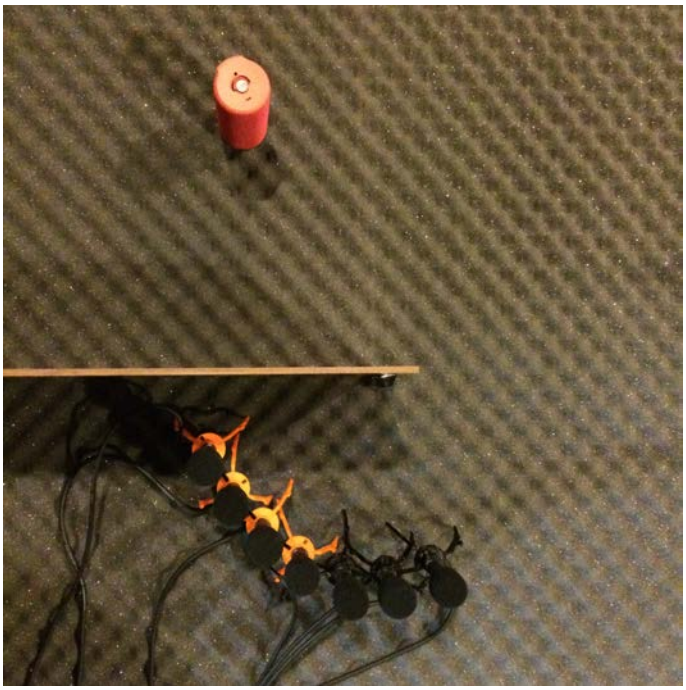


Figure 6.10: Experimental setup

diffraction is enlarged from mic-1 to mic-4. The region A, in particular, displays the first arrival diffraction signal. The diffraction loss is more clearly identified in the frequency domain depicted in Fig. 6.13.

Figure 6.14 is a comparison of mic-2 signals with and without the reflection wall. The figure clearly shows the effect of the wall between the two cases where region A only contains diffraction, and region B contains a mixture of diffraction and reflection. The duration of region A was measured to be less than a 5% error from the actual distance measurements.

6.5 Chapter Summary

This chapter demonstrated the target estimation capability of a new acoustic approach in NFOV, using the sound wave physical properties, reflection and diffraction. This approach used sound source reflection and diffraction signals to construct two distinct observation

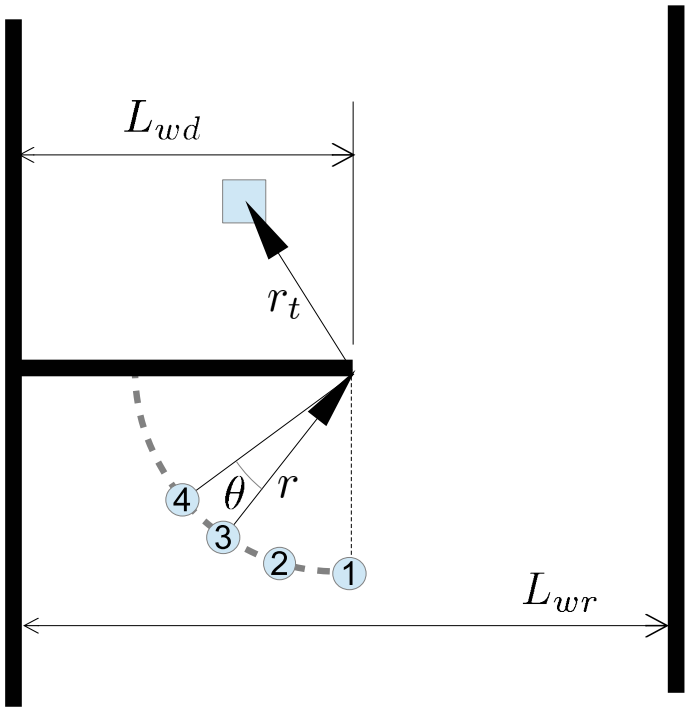


Figure 6.11: Experimental environment

likelihoods from the target sound. The joint acoustic observation likelihood can then be used to estimate the target location. This process was formulated mathematically with two-dimensional assumptions. The proposed approach was tested and parametrically studied under different conditions in the simulation. Finally, an experimental environment was constructed for further validation.

Future work consists of extending the application to an indoor environment. This includes the incorporation of an actual mobile target estimation using a mobile robot equipped with a microphone array.

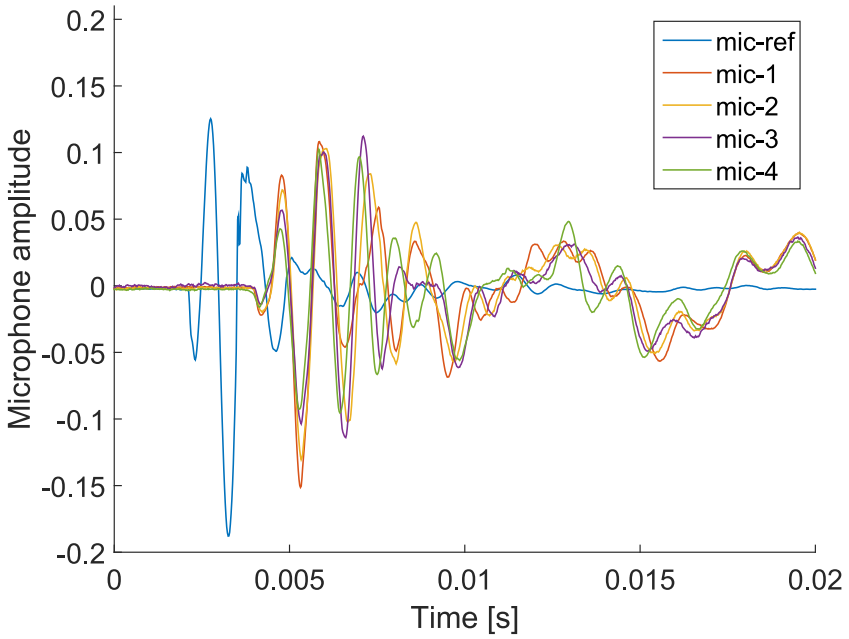


Figure 6.12: Reference mic signal at the target and signal of mic array

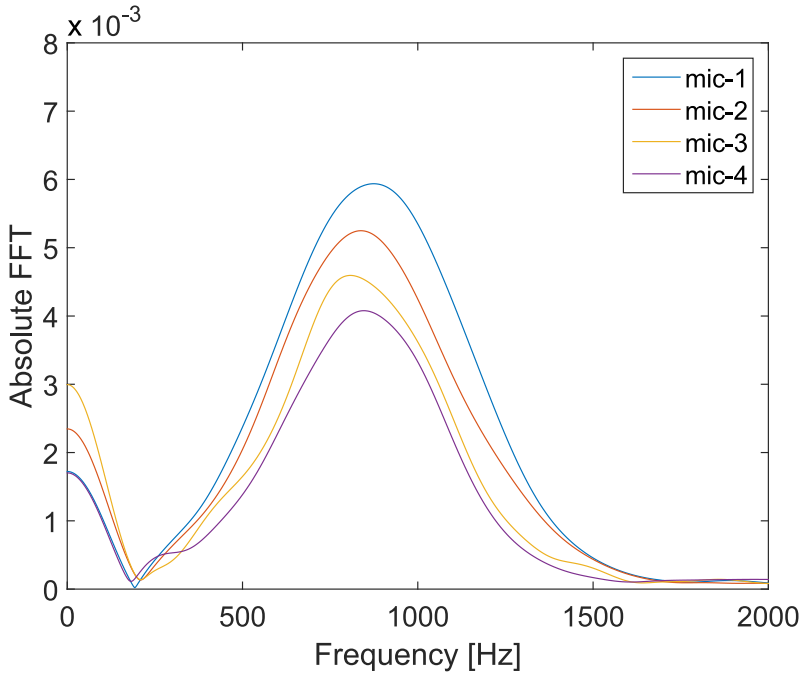


Figure 6.13: Magnitude with different orientation angles

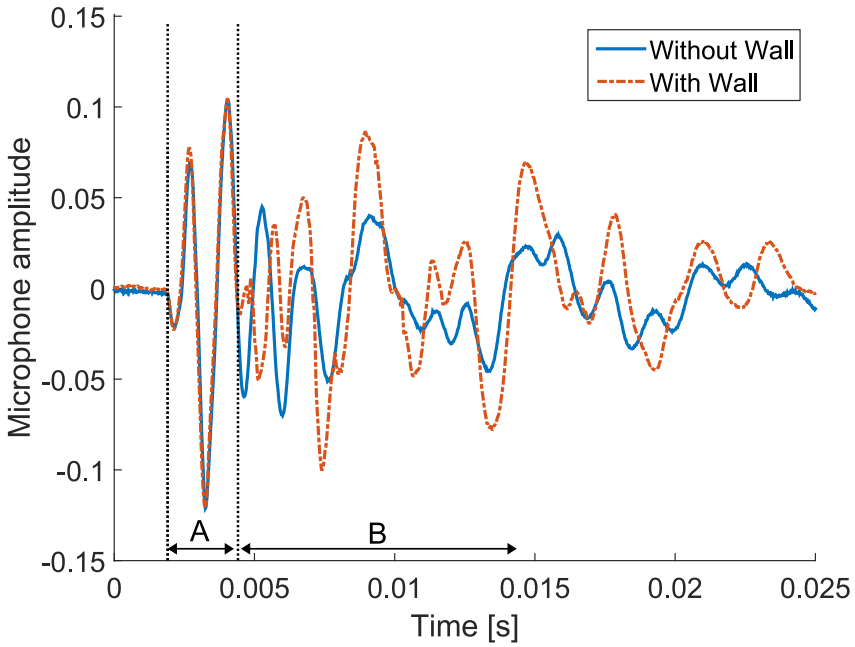


Figure 6.14: Signal with and without the reflective wall

CHAPTER 7

Conclusions and Future Work

In this final chapter, the progress for proposed NFOV acoustic target localization framework and findings are summarizes and discussed. This chapter begins with a summary of original contributions in NFOV target estimation in Section 7.1, followed by Section 7.2 where limitations of the proposed technique and future research suggestions are discussed. Finally, this dissertation ends with closing remarks in Section 7.3.

7.1 Summary of Contributions

The primary objectives of this dissertation are: a localization technique that can estimate the position of the unknown target, together with a development of a testing system capable of parametrically and quantitatively studying the performance and effectiveness of

a proposed technique. The proposed framework was mathematically formulated probabilistically to describe the highly nonlinear non-Gaussian process of NFOV acoustic target localization. The recursive mechanism predicts and updates the estimation of a target location. With this mechanism multimodality of the observation likelihood are suppressed, leading to the better estimate of the target position. The proposed framework is applied to the formulation of the three techniques.

The first proposed technique uses a fingerprint-based technique to construct the location-dependent database from a microphone array located in the NFOV of a target. Following the database collection phase, the NFOV acoustic target location is estimated by calculating the observation likelihood within the RBE framework.

The second approach takes a hybrid of visual/auditory information. Auditory sensors are used to assist visual sensors when the event of "no detection" by an optical sensor is converted into an observation likelihood, and utilized to positively update probabilistic belief on the target that is dynamically maintained by the RBE. Acoustic observation likelihoods are constructed based on *a priori* acoustic cues in the given environment.

The last proposed technique focuses on wave propagation properties, based on sound physics, by extracting the first-arrival diffraction and reflection signals. The fixed microphone array technique requires prior data collection, which could be a time-consuming process. The process time increases exponentially when the microphones are placed on a mobile platform. These shortcomings are addressed on the final approach, with the use of diffraction and reflection techniques to eliminate the prior database construction phase. However, final approach is still in a proof of concept, and assumes geometrical propagation of the sound signal, and only considers well-controlled ideal environments.

A probabilistic formulation within the RBE framework maintains nonlinear, non-Gaussian distributions. Numerical studies performed NFOV localization for the parametrically controlled environment for each proposed localization framework, demonstrated their capabilities and effectiveness.

7.2 Direction for Future Work

There are several possible extensions to the proposed framework, despite the success of NFOV localization. In real environment a single target assumption is not guaranteed, thus the capability of multiple target localization is necessary. The aid of sound source separation, speech recognition, and other signal processing algorithms could be easily combined to provide possible solutions.

Mobile robots require adaptation to not only static, but also dynamic environments. The tested environments for the proposed framework were solely relying on static assumption. Figure 7.1 shows an example of an autonomous mobile robot that can coexist with a human. With the increase in demand for the coexistence of robots with humans and for HRI, the mobile robots will inevitably be exposed to dynamic environments. With an evolving robots, an implementation of a real-time accurate Simultaneous localization and mapping (SLAM) could update the state of the dynamic environment accordingly to match the change to surroundings. This could be one of the solutions to the problem.

Despite numerical studies that have been performed, the actual application for robot audition on mobile robots needs to be implemented and validated. Fortunately, the diffraction/ reflection based approach has been granted by National Science Foundation Early-concept Grant Exploratory Research (EAGER) to pursue an investigation of the approach.

7.3 Closing Remarks

In the world of an increasing introduction of mobile devices, and the rapid development of emerging technologies, the underlying fundamental physics remains the same. When it comes to mobile robots, it is essential to observe as much information as possible from the environment. This dissertation provides a glimpse into NFOV target localization with the use of acoustic signals in conjunction with other sensors. The advancement of proposed

techniques, in collaboration with other sensing capabilities, will potentially be applied to the future applications.

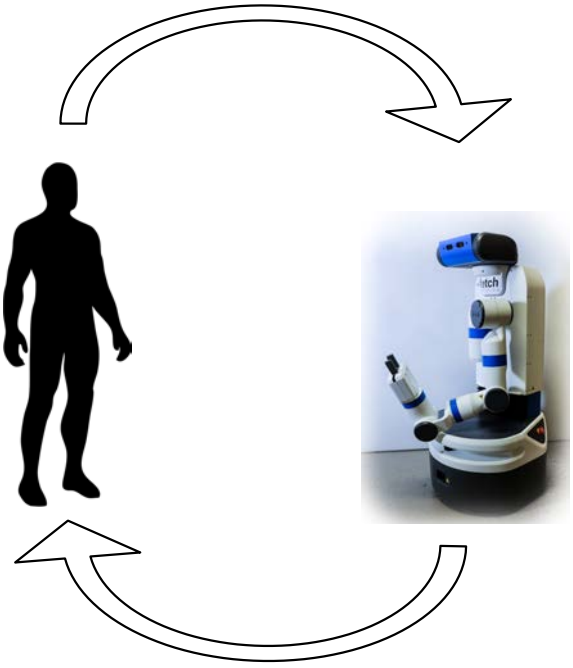


Figure 7.1: Future step towards HRI

References

[1] Hassan A Karimi. Indoor wayfinding and navigation. 2015. (document), 2.1

[2] H Medwin. Shadowing by finite noise barriers. *The Journal of the Acoustical Society of America*, 69(4):1060–1064, 1981. (document), 6.3.1, 6.5

[3] George A. Bekey. *Autonomous Robots: From Biological Inspiration to Implementation and Control (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005. ISBN 0262025787. 1.1

[4] Alonzo Kelly. *Mobile Robotics: Mathematics, Models, and Methods*. Cambridge University Press, 2013. 1.1

[5] Stuart Russell and Peter Norvig. Artificial intelligence: a modern approach. 1995. 1.1

[6] Jean M Whitney, **Takami, Kuya**, Scott T Sanders, and Yasuhiro Okura. Design of system for rugged, low-noise fiber-optic access to high-temperature, high-pressure environments. *Sensors Journal, IEEE*, 11(12):3295–3302, 2011.

[7] Xinliang An, Thilo Kraetschmer, **Takami, Kuya**, Scott T Sanders, Lin Ma, Weiwei Cai, Xuesong Li, Sukesh Roy, and James R Gord. Validation of temperature imaging by h 2 o absorption spectroscopy using hyperspectral tomography in controlled experiments. *Applied optics*, 50(4):A29–A37, 2011.

[8] **Takami, Kuya**, Tomonari Furukawa, Makoto Kumon, Daisuke Kimoto, and Gamini Dissanayake. Estimation of a nonvisible field-of-view mobile target incorporating optical and acoustic sensors. *Autonomous Robots*, pages 1–17, 2015.

[9] **Takami, Kuya**, Tomonari Furukawa, Makoto Kumon, and Gamini Dissanayake. Non-field-of-view acoustic target estimation in complex indoor environment. *Springer Tracts in Advanced Robotics*, 2015.

[10] **Takami, Kuya**, Tomonari Furukawa, and Daniel Watman. High-resolution measurement system for rotating deformable body. *Journal of Visualization*, 2015. Submitted.

[11] **Takami, Kuya**, Tomonari Furukawa, Makoto Kumon, and Lin Chi Mak. Non-field-of-view acoustic target estimation using reflection and diffraction. In preparation.

[12] Makoto Kumon, Daisuke Kimoto, **Takami, Kaya**, and Tomonari Furukawa. Acoustic recursive bayesian estimation for non-field-of-view targets. In *Image Analysis for Multimedia Interactive Services (WIAMIS), 2013 14th International Workshop on*, pages 1–4. IEEE, 2013.

[13] Makoto Kumon, Daisuke Kimoto, **Takami, Kuya**, and Tomonari Furukawa. Bayesian non-field-of-view target estimation incorporating an acoustic sensor. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 3425–3432. IEEE, 2013.

[14] **Takami, Kuya**, Saied Taheri, Mehdi Taheri, and Tomonari Furukawa. Prediction of railroad track foundation defects using wavelets. In *2013 Joint Rail Conference*, pages V001T01A008–V001T01A008. American Society of Mechanical Engineers, 2013.

[15] **Takami, Kuya**, Tomonari Furukawa, Makoto Kumon, and Lin Chi Mak. Non-field-of-view indoor sound source localization based on reflection and diffraction. 2015.

[16] **Takami, Kuya**, Tomonari Furukawa, Makoto Kumon, and Gamini Dissanayake. Non-field-of-view acoustic target estimation in complex indoor environment. 2015.

[17] Tomonari Furukawa, **Takami, Kuya**, Xianqiao Tong, Daniel Watman, Abbi Hamed, Ravindra Ranasinghe, and Gamini Dissanayake. Map-based navigation of a autonomous car using grid-based scan-to-map matching. 2015.

[18] **Takami, Kuya** and Tomonari Furukawa. High-resolution deformation measurement system for fast rotating tires towards noise prediction. 2015.

[19] **Takami, Kuya** and Tomonari Furukawa. Study of tire noise characteristics with high-resolution synchronous images. 2015.

[20] Guoqiang Mao, Barış Fidan, and Brian DO Anderson. Wireless sensor network localization techniques. *Computer networks*, 51(10):2529–2553, 2007. 2.2

[21] Giovanni Bellusci, Junlin Yan, Gerard JM Janssen, and Christian CJM Tiberius. An ultra-wideband positioning demonstrator using audio signals. In *Positioning, Navigation and Communication, 2007. WPNC'07. 4th Workshop on*, pages 71–76. IEEE, 2007. 2.2

[22] Piergiorgio Svaizer, Alessio Brutti, and Maurizio Omologo. Environment aware estimation of the orientation of acoustic sources using a line array. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pages 1024–1028. IEEE, 2012. 2.2, 2.3.5

[23] Paul M Hofman, Jos GA Van Riswick, and A John Van Opstal. Relearning sound localization with new ears. *Nature neuroscience*, 1(5):417–421, 1998. 2.2, 2.3.3

[24] Chahé Nerguizian, Charles Despins, and Sofiène Affès. Geolocation in mines with an impulse response fingerprinting technique and neural networks. *Wireless Communications, IEEE Transactions on*, 5(3):603–611, 2006. 2.2, 2.3.3

[25] Mussa Bshara, Umut Orguner, Fredrik Gustafsson, and Leo Van Biesen. Fingerprinting localization in wireless networks based on received-signal-strength measurements: A case study on wimax networks. *Vehicular Technology, IEEE Transactions on*, 59(1):283–294, 2010. 2.2

[26] Sinan Gezici, Zhi Tian, Georgios B Giannakis, Hisashi Kobayashi, Andreas F Molisch, H Vincent Poor, and Zafer Sahinoglu. Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks. *Signal Processing Magazine, IEEE*, 22(4):70–84, 2005. 2.3.1, 2.3.2

[27] Ali H Sayed, Alireza Tarighat, and Nima Khajehnouri. Network-based wireless location: challenges faced in developing techniques for accurate wireless location information. *Signal Processing Magazine, IEEE*, 22(4):24–40, 2005. 2.3.1, 2.3.2

[28] James A Storer and John H Reif. Shortest paths in the plane with polygonal obstacles. *Journal of the ACM (JACM)*, 41(5):982–1012, 1994. 2.3.1

[29] Charles Knapp and G Clifford Carter. The generalized correlation method for estimation of time delay. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 24(4):320–327, 1976. 2.1, 2.3.2

[30] Yanchao Zhang, Wei Liu, Yuguang Fang, and Dapeng Wu. Secure localization and authentication in ultra-wideband sensor networks. *Selected Areas in Communications, IEEE Journal on*, 24(4):829–835, 2006. 2.1

[31] James Scott and Boris Dragovic. Audio location: Accurate low-cost location sensing. In *Pervasive Computing*, pages 1–18. Springer, 2005. 2.1

[32] Yimin Zhang, Moeness G Amin, and Shashank Kaushik. Localization and tracking of passive rfid tags based on direction estimation. *International Journal of Antennas and Propagation*, 2007, 2007. 2.1

[33] Andy Ward, Alan Jones, and Andy Hopper. A new location technique for the active office. *Personal Communications, IEEE*, 4(5):42–47, 1997. 2.1, 2.3.2

[34] Nissanka B Priyantha, Anit Chakraborty, and Hari Balakrishnan. The cricket location-support system. In *Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 32–43. ACM, 2000. 2.1, 2.3.2

[35] Paramvir Bahl and Venkata N Padmanabhan. Radar: An in-building rf-based user location and tracking system. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 775–784. Ieee, 2000. 2.1, 2.3.3

[36] Stephen P Tarzia, Peter A Dinda, Robert P Dick, and Gokhan Memik. Indoor localization without infrastructure using the acoustic background spectrum. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*, pages 155–168. ACM, 2011. 2.1

[37] Lionel M Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P Patil. Landmarc: indoor location sensing using active rfid. *Wireless networks*, 10(6):701–710, 2004. 2.1, 2.3.2, 2.3.3

[38] Hong Lu, Wei Pan, Nicholas D Lane, Tanzeem Choudhury, and Andrew T Campbell. Soundsense: scalable sound sensing for people-centric applications on mobile phones. In *Proceedings of the 7th international conference on Mobile systems, applications, and services*, pages 165–178. ACM, 2009. 2.1

[39] Veljo Otsason, Alex Varshavsky, Anthony LaMarca, and Eyal De Lara. Accurate gsm indoor localization. In *UbiComp 2005: Ubiquitous Computing*, pages 141–158. Springer, 2005. 2.1

[40] Nadia Aloui, Kosai Raoof, Ammar Bouallegue, Stephane Letourneur, and Sonia Zaibi. Performance evaluation of an acoustic indoor localization system based on a fingerprinting technique. *EURASIP Journal on Advances in Signal Processing*, 2014 (1):1–16, 2014. 2.1

[41] Aylin Aksu and Prashant Krishnamurthy. Sub-area localization: A simple calibration free approach. In *Proceedings of the 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems*, pages 63–72. ACM, 2010. 2.1, 2.3.4

[42] Anja Bekkelien, Michel Deriaz, and Stéphane Marchand-Maillet. Bluetooth indoor positioning. *Master's thesis, University of Geneva*, 2012. 2.1

[43] Jongchul Song, Carl T Haas, and Carlos H Caldas. A proximity-based method for locating rfid tagged objects. *Advanced Engineering Informatics*, 21(4):367–376, 2007. 2.1

[44] Lewis Girod and Deborah Estrin. Robust range estimation using acoustic and multimodal sensing. In *Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on*, volume 3, pages 1312–1320. IEEE, 2001. 2.3.2, 2.3.5

[45] Hoang Do, Harvey F Silverman, and Ying Yu. A real-time srp-phat source location implementation using stochastic region contraction (src) on a large-aperture microphone array. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 1, pages I–121. IEEE, 2007. 2.3.2

[46] Harvey F Silverman, Ying Yu, Joshua M Sachar, William R Patterson, et al. Performance of real-time source-location estimators for a large-aperture microphone array. *Speech and Audio Processing, IEEE Transactions on*, 13(4):593–606, 2005. 2.3.2

[47] Jean-Marc Valin, François Michaud, Jean Rouat, and Dominic Létourneau. Robust sound source localization using a microphone array on a mobile robot. In *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 2, pages 1228–1233. IEEE, 2003. 2.3.2

[48] Darren B Ward, Eric A Lehmann, and Robert C Williamson. Particle filtering algorithms for tracking an acoustic source in a reverberant environment. *Speech and Audio Processing, IEEE Transactions on*, 11(6):826–836, 2003. 2.3.2

[49] J. Riba and A. Urruela. A non-line-of-sight mitigation technique based on ml-detection. In *Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing*, volume 2, pages 153–156, May 2005. 2.3.2

[50] Kenneth D Frampton. Acoustic self-localization in a distributed sensor network. *Sensors Journal, IEEE*, 6(1):166–172, 2006. 2.3.2

[51] Enzo Mumolo, Massimiliano Nolich, and Gianni Vercelli. Algorithms for acoustic localization based on microphone array in service robotics. *Robotics and Autonomous systems*, 42(2):69–88, 2003. 2.3.2

[52] Darren B Ward and Michael S Brandstein. Grid-based beamformer design for room-environment microphone arrays. In *Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on*, pages 23–26. IEEE, 1999. 2.3.2

[53] Jean-Marc Valin, François Michaud, and Jean Rouat. Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering. *Robotics and Autonomous Systems*, 55(3):216–228, 2007. 2.3.2

[54] Marco Bertinato, Giulia Ortolan, Fabio Maran, Riccardo Marcon, Alessandro Marcassa, Filippo Zanella, Matrizio Zambotto, Luca Schenato, and Angelo Cenedese. Rf localization and tracking of mobile nodes in wireless sensors networks: Architectures, algorithms and experiments. 2008. 2.3.2

[55] Huan Dai, Zhao-Min Zhu, and Xiao-Feng Gu. Multi-target indoor localization and tracking on video monitoring system in a wireless sensor network. *Journal of Network and Computer Applications*, 2012. 2.3.2

[56] Dian Zhang, Yanyan Yang, Dachao Cheng, Siyuan Liu, and Lionel M Ni. Cocktail: an rf-based hybrid approach for indoor localization. In *Communications (ICC), 2010 IEEE International Conference on*, pages 1–5. IEEE, 2010. 2.3.2

[57] Hui Liu, H. Darabi, P. Banerjee, and Jing Liu. Survey of wireless indoor positioning techniques and systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(6):1067–1080, 2007. ISSN 1094-6977. doi: 10.1109/TSMCC.2007.905750. 2.3.2

[58] Sinan Gezici. A survey on wireless position estimation. *Wireless Personal Communications*, 44(3):263–282, 2008. 2.3.2

[59] Ismail Guvenc and Chia-Chin Chong. A survey on toa based wireless localization and nlos mitigation techniques. *Communications Surveys & Tutorials, IEEE*, 11(3):107–124, 2009. 2.3.2

[60] Guolin Sun, Jie Chen, Wei Guo, and KJ Ray Liu. Signal processing techniques in network-aided positioning: a survey of state-of-the-art positioning designs. *Signal Processing Magazine, IEEE*, 22(4):12–23, 2005. 2.3.2, 2.3.3

[61] James J Caffery Jr. A new approach to the geometry of toa location. In *Vehicular Technology Conference, 2000. IEEE-VTS Fall VTC 2000. 52nd*, volume 4, pages 1943–1949. IEEE, 2000. 2.3.2

[62] Saipradeep Venkatraman, James Caffery, and Heung-Ryeol You. A novel toa location algorithm using los range estimation for nlos environments. *Vehicular Technology, IEEE Transactions on*, 53(5):1515–1524, 2004. 2.3.2

[63] Wang Wei, Xiong Jin-Yu, and Zhu Zhong-Liang. A new nlos error mitigation algorithm in location estimation. *Vehicular Technology, IEEE Transactions on*, 54(6):2048–2053, 2005. 2.3.2

[64] Neiyer S Correal, Spyros Kyperountas, Qicai Shi, and Matt Welborn. An uwb relative location system. In *IEEE Conference on Ultra Wideband Systems and Technologies*, pages 394–397, 2003. 2.3.2

[65] Robert Fleming and Cherie Kushner. Low-power, miniature, distributed position location and communication devices using ultra-wideband, nonsinusoidal communication technology. *Semi-annual technical report, darpa, fbi, Aetherwire Location Inc*, 1995. 2.3.2

[66] Robert J Fontana and Steven J Gunderson. Ultra-wideband precision asset location system. In *Ultra Wideband Systems and Technologies, 2002. Digest of Papers. 2002 IEEE Conference on*, pages 147–150. IEEE, 2002. 2.3.2

[67] Neal Patwari, Joshua N Ash, Spyros Kyperountas, Alfred O Hero, Randolph L Moses, and Neiyer S Correal. Locating the nodes: cooperative localization in wireless sensor networks. *Signal Processing Magazine, IEEE*, 22(4):54–69, 2005. 2.3.2

[68] B Denis, JK Keignart, and N Daniele. Impact of nlos propagation upon ranging precision in uwb systems. In *Ultra Wideband Systems and Technologies, 2003 IEEE Conference on*, pages 379–383. IEEE, 2003. 2.3.2

[69] Joon-Yong Lee and Sungyul Yoo. Large error performance of uwb ranging in multipath and multiuser environments. *Microwave Theory and Techniques, IEEE Transactions on*, 54(4):1887–1895, 2006. 2.3.2

[70] ZN Low, JH Cheong, CL Law, WT Ng, and YJ Lee. Pulse detection algorithm for line-of-sight (los) uwb ranging applications. *Antennas and Wireless Propagation Letters, IEEE*, 4:63–67, 2005. 2.3.2

[71] Zoubir Irahhauten, Giovanni Bellusci, Gerard JM Janssen, Homayoun Nikookar, and CCJM Tiberius. Investigation of uwb ranging in dense indoor multipath environments. In *Communication systems, 2006. ICCS 2006. 10th IEEE Singapore International Conference on*, pages 1–5. IEEE, 2006. 2.3.2

[72] Chi Xu and Choi L Law. Delay-dependent threshold selection for uwb toa estimation. *IEEE communications letters*, 12(5):380–382, 2008. 2.3.2

[73] Ismail Guvenc, Chia-Chin Chong, and Fujio Watanabe. Joint toa estimation and localization technique for uwb sensor network applications. In *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pages 1574–1578. IEEE, 2007. 2.3.2

[74] Joan Borras, Paul Hatrack, and Narayan B Mandayam. Decision theoretic framework for nlos identification. In *Vehicular Technology Conference, 1998. VTC 98. 48th IEEE*, volume 2, pages 1583–1587. IEEE, 1998. 2.3.2

[75] Sauoradeep Venkatraman and Jr Caffery. Statistical approach to non-line-of-sight bs identification. In *Wireless Personal Multimedia Communications, 2002. The 5th International Symposium on*, volume 1, pages 296–300. IEEE, 2002. 2.3.2

[76] Pi-Chun Chen. A non-line-of-sight error mitigation algorithm in location estimation. In *Wireless Communications and Networking Conference, 1999. WCNC. 1999 IEEE*, pages 316–320. IEEE, 1999. 2.3.2

[77] Li Cong and Weihua Zhuang. Non-line-of-sight error mitigation in tdoa mobile location. In *Global Telecommunications Conference, 2001. GLOBECOM'01. IEEE*, volume 1, pages 680–684. IEEE, 2001. 2.3.2

[78] Li Cong and Weihua Zhuang. Nonline-of-sight error mitigation in mobile location. *Wireless Communications, IEEE Transactions on*, 4(2):560–573, 2005. 2.3.2

[79] Mohammad Heidari and Kaveh Pahlavan. Identification of the absence of direct path in toa-based indoor localization systems. *International Journal of Wireless Information Networks*, 15(3-4):117–127, 2008. 2.3.2

[80] Wade H Foy. Position-location solutions by taylor-series estimation. 1976. 2.3.2

[81] Xin Wang, Zongxin Wang, Bob O Dea, et al. A toa-based location algorithm reducing the errors due to non-line-of-sight (nlos) propagation. *IEEE Transactions on Vehicular Technology*, 52(1):112–116, 2003. 2.3.2

[82] S Al-Jazzar and Jr Caffery. Ml and bayesian toa location estimators for nlos environments. In *Vehicular Technology Conference, 2002. Proceedings. VTC 2002-Fall. 2002 IEEE 56th*, volume 2, pages 1178–1181. IEEE, 2002. 2.3.2

[83] Cha-Hwa Lin, Juin-Yi Cheng, and Chien-Nan Wu. Mobile location estimation using density-based clustering technique for nlos environments. *Cluster Computing*, 10(1): 3–16, 2007. 2.3.2

[84] Ralph O Schmidt. A new approach to geometry of range difference location. 1972. 2.3.2

[85] W Murphy and Willy Hereman. Determination of a position in three dimensions using trilateration and approximate distances. *Department of Mathematical and Computer Sciences, Colorado School of Mines, Golden, Colorado, MCS-95-07*, 19, 1995. 2.3.2

[86] YT Chan and KC Ho. A simple and efficient estimator for hyperbolic location. *Signal Processing, IEEE Transactions on*, 42(8):1905–1915, 1994. 2.3.2

[87] Kegen Yu and Y Jay Guo. Improved positioning algorithms for nonline-of-sight environments. *Vehicular Technology, IEEE Transactions on*, 57(4):2342–2353, 2008. 2.3.2

[88] Philip E Gill, Walter Murray, and Margaret H Wright. Practical optimization. 1981. 2.3.2

[89] Li Cong and Weihua Zhuang. Hybrid tdoa/aoa mobile user location for wideband cdma cellular systems. *Wireless Communications, IEEE Transactions on*, 1(3):439–447, 2002. 2.3.2

[90] Ji Li, Jean Conan, and Samuel Pierre. Mobile station location estimation for mimo communication systems. In *Wireless Communication Systems, 2006. ISWCS'06. 3rd International Symposium on*, pages 561–564. IEEE, 2006. 2.3.2

[91] Honglei Miao, Kegen Yu, and Markku J Juntti. Positioning for nlos propagation: algorithm derivations and cramer–rao bounds. *Vehicular Technology, IEEE Transactions on*, 56(5):2568–2580, 2007. 2.3.2

[92] Chee Kiat Seow and Soon Yim Tan. Non-line-of-sight localization in multipath environments. *Mobile Computing, IEEE Transactions on*, 7(5):647–660, 2008. 2.3.2

[93] Chee Kiat Seow and Soon Yim Tan. Non-line-of-sight localization in multipath environments. *Mobile Computing, IEEE Transactions on*, 7(5):647–660, 2008. 2.3.2, 2.3.3

[94] NJ Thomas, DGM Cruickshank, and DI Laurenson. Calculation of mobile location using scatterer information. *Electronics Letters*, 37(19):1193–1194, 2001. 2.3.2

[95] Saipradeep Venkatraman and James Caffery. Hybrid toa/aoa techniques for mobile location in non-line-of-sight environments. In *Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE*, volume 1, pages 274–278. IEEE, 2004. 2.3.2

[96] Yoko Sasaki, Satoshi Kagami, and Hiroshi Mizoguchi. Multiple sound source mapping for a mobile robot by self-motion triangulation. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 380–385. IEEE, 2006. 2.3.2

[97] Jean-Marc Valin, François Michaud, Brahim Hadjou, and Jean Rouat. Localization of simultaneous moving sound sources for mobile robot using a frequency-domain steered beamformer approach. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 1, pages 1033–1038. IEEE, 2004. 2.3.2

[98] Jean-Marc Valin, François Michaud, and Jean Rouat. Robust 3d localization and tracking of sound sources using beamforming and particle filtering. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 4, pages IV–IV. IEEE, 2006. 2.3.2

[99] Joseph Hector DiBiase. *A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays*. PhD thesis, Brown University, 2000. 2.3.2

[100] Joseph H DiBiase, Harvey F Silverman, and Michael S Brandstein. Robust localization in reverberant rooms. In *Microphone Arrays*, pages 157–180. Springer, 2001. 2.3.2

[101] Paramvir Bahl and Venkata N Padmanabhan. Radar: An in-building rf-based user location and tracking system. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 775–784. Ieee, 2000. 2.3.3

[102] Andrew M Ladd, Kostas E Bekris, Algis P Rudys, Dan S Wallach, and Lydia E Kavraki. On the feasibility of using wireless ethernet for indoor localization. *IEEE Transactions on Robotics and Automation*, 20(3):555–559, 2004. 2.3.3

[103] Pi-Chun Chen. A non-line-of-sight error mitigation algorithm in location estimation. In *Wireless Communications and Networking Conference, 1999. WCNC. 1999 IEEE*, pages 316–320. IEEE, 1999. 2.3.3

[104] Eric A Prigge. *A positioning system with no line-of-sight restrictions for cluttered environments*. PhD thesis, Stanford University, 2004. 2.3.3

[105] Hiam M Khoury and Vineet R Kamat. Evaluation of position tracking technologies for user localization in indoor construction environments. *Automation in Construction*, 18(4):444–457, 2009. 2.3.3

[106] Julie Letchner, Dieter Fox, and Anthony LaMarca. Large-scale localization from wireless signal strength. In *Proceedings of the national conference on artificial intelligence*, volume 20, page 15. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005. 2.3.3

[107] Brian Ferris, Dirk Haehnel, and Dieter Fox. Gaussian processes for signal strength-based location estimation. In *In Proc. of Robotics Science and Systems*. Citeseer, 2006. 2.3.3

[108] Andrew M Ladd, Kostas E Bekris, Algis P Rudys, Dan S Wallach, and Lydia E Kavraki. On the feasibility of using wireless ethernet for indoor localization. *IEEE Transactions on Robotics and Automation*, 20(3):555–559, 2004. 2.3.3

[109] Mikkel Baun Kjærgaard. A taxonomy for radio location fingerprinting. In *Location- and Context-Awareness*, pages 139–156. Springer, 2007. 2.3.3

[110] Paramvir Bahl, Venkata N Padmanabhan, and Anand Balachandran. A software system for locating mobile users: Design, evaluation. Technical report, and lessons. Technical report, Microsoft Research, MSR-TR-2000-12, 2000. 2.3.3

[111] Michael Robinson and Ioannis Psaromiligkos. Received signal strength based location estimation of a wireless lan client. In *Wireless Communications and Networking Conference, 2005 IEEE*, volume 4, pages 2350–2354. IEEE, 2005. 2.3.3

[112] David Madigan, E Einahrawy, Richard P Martin, W-H Ju, Parameshwaran Krishnan, and AS Krishnakumar. Bayesian indoor positioning systems. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, volume 2, pages 1217–1227. IEEE, 2005. 2.3.3

[113] Brian Ferris, Dieter Fox, and Neil D Lawrence. Wifi-slam using gaussian process latent variable models. In *IJCAI*, volume 7, pages 2480–2485, 2007. 2.3.3

[114] Sinno Jialin Pan, James T Kwok, Qiang Yang, and Jeffrey Junfeng Pan. Adaptive localization in a dynamic wifi environment through multi-view learning. In *Proceedings of the national conference on artificial Intelligence*, volume 22, page 1108. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2007. 2.3.3

[115] Qiang Yang, Sinno Jialin Pan, and Vincent Wenchen Zheng. Estimating location using wi-fi. *IEEE Intelligent Systems*, 23(1):8–13, 2008. 2.3.3

[116] Andreas Haeberlen, Eliot Flannery, Andrew M Ladd, Algis Rudys, Dan S Wallach, and Lydia E Kavraki. Practical robust localization over large-scale 802.11 wireless networks. In *Proceedings of the 10th annual international conference on Mobile computing and networking*, pages 70–84. ACM, 2004. 2.3.3

[117] Jeffrey Junfeng Pan, James T Kwok, Qiang Yang, and Yiqiang Chen. Multidimensional vector regression for accurate and low-cost location estimation in pervasive computing. *Knowledge and Data Engineering, IEEE Transactions on*, 18(9):1181–1193, 2006. 2.3.3

[118] Mauro Brunato and Roberto Battiti. Statistical learning theory for location fingerprinting in wireless lans. *Computer Networks*, 47(6):825–845, 2005. 2.3.3

[119] Zheng Yang, Chenshu Wu, and Yunhao Liu. Locating in fingerprint space: wireless indoor localization with little human intervention. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pages 269–280. ACM, 2012. 2.3.3

[120] Hyuk Lim, Lu-Chuan Kung, Jennifer C Hou, and Haiyun Luo. Zero-configuration, robust indoor localization: Theory and experimentation. 2005. 2.3.3

[121] Andreas Savvides, Chih-Chieh Han, and Mani B Strivastava. Dynamic fine-grained localization in ad-hoc networks of sensors. In *Proceedings of the 7th annual international conference on Mobile computing and networking*, pages 166–179. ACM, 2001. 2.3.3

[122] Kamin Whitehouse, Chris Karlof, and David Culler. A practical evaluation of radio signal strength for ranging-based localization. *ACM SIGMOBILE Mobile Computing and Communications Review*, 11(1):41–52, 2007. 2.3.3

[123] Jie Yin, Qiang Yang, and Lionel Ni. Adaptive temporal radio maps for indoor location estimation. In *Pervasive Computing and Communications, 2005. PerCom 2005. Third IEEE International Conference on*, pages 85–94. IEEE, 2005. 2.3.3

[124] Chuang-Wen You, Polly Huang, Hao-hua Chu, Yi-Chao Chen, Ji-Rung Chiang, and Seng-Yong Lau. Impact of sensor-enhanced mobility prediction on the design of energy-efficient localization. *Ad Hoc Networks*, 6(8):1221–1237, 2008. 2.3.3

[125] Shih-Hau Fang, Tsung-Nan Lin, and Kun-Chou Lee. A novel algorithm for multipath fingerprinting in indoor wlan environments. *Wireless Communications, IEEE Transactions on*, 7(9):3579–3588, 2008. 2.3.3

[126] Shih-Hau Fang and Tsung-Nan Lin. Robust wireless lan location fingerprinting by svd-based noise reduction. In *Communications, Control and Signal Processing,*

2008. ISCCSP 2008. 3rd International Symposium on, pages 295–298. IEEE, 2008. 2.3.3

[127] D Jourdan, John J Deyst Jr, Moe Z Win, and Nicholas Roy. Monte carlo localization in dense multipath environments using uwb ranging. In *Ultra-Wideband, 2005. ICU 2005. 2005 IEEE International Conference on*, pages 314–319. IEEE, 2005. 2.3.3

[128] Damien Jourdan. *Wireless sensor network planning with application to UWB localization in GPS-denied environments*. PhD thesis, Massachusetts Institute of Technology, 2006. 2.3.3

[129] Frédéric Evennou and François Marx. Advanced integration of wifi and inertial navigation systems for indoor mobile positioning. *Eurasip journal on applied signal processing*, 2006:164–164, 2006. 2.3.3

[130] Seon-Woo Lee and Kenji Mase. Activity and location recognition using wearable sensors. *IEEE pervasive computing*, 1(3):24–32, 2002. 2.3.3

[131] A. Takanishi, S. Masukawa, Y. Mori, and T. Ogawa. Study on anthropomorphic auditory robot continuous localization of a sound source in horizontal plane,. In *11th Annual Conference of the Robotics Society of Japan*, pages 793–796, 1993. 2.3.3

[132] Makoto Kumon, Tomoko Shimoda, Ryuichi Kohzawa, Ikuro Mizumoto, and Zenta Iwai. Audio servo for robotic systems with pinnae. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 1881–1886. IEEE, 2005. 2.3.3

[133] Tomoko Shimoda, Toru Nakashima, Makoto Kumon, Ryuichi Kohzawa, Ikuro Mizumoto, and Zenta Iwai. Spectral cues for robust sound localization with pinnae. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 386–391. IEEE, 2006. 2.3.3

[134] Ashutosh Saxena and Andrew Y Ng. Learning sound location from a single microphone. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 1737–1742. IEEE, 2009. 2.3.3

[135] Li-Wei Wu, Wei-Han Liu, Chieh-Cheng Cheng, and Jwu-Sheng Hu. Gaussian mixture-sound field landmark model for robot localization applications. *Advanced Robotics*, 21(5-6):619–643, 2007. 2.3.3

[136] Jwu-Sheng Hu, Wei-Han Liu, Chieh-Cheng Cheng, and Chia-Hsing Yang. Location and orientation detection of mobile robots using sound field features under complex environments. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 1151–1156. IEEE, 2006. 2.3.3

[137] Sudarshan S Chawathe. Low-latency indoor localization using bluetooth beacons. In *Intelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference on*, pages 1–7. IEEE, 2009. 2.3.4

[138] J Abel Smith et al. closed-form. least-squares source location estimation from range-difference measurements. 1987. 2.3.5

[139] Lin C Mak and Tomonari Furukawa. Non-line-of-sight localization of a controlled sound source. In *Advanced Intelligent Mechatronics, 2009. AIM 2009. IEEE/ASME International Conference on*, pages 475–480. IEEE, 2009. 2.3.5

[140] JeanMarc Valin, François Michaud, Jean Rouat, and Dominic Létourneau. Robust sound source localization using a microphone array on a mobile robot. In *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 2, pages 1228–1233. IEEE, 2003. 2.3.5

[141] YanChen Lu and Martin Cooke. Binaural estimation of sound source distance via the direct-to-reverberant energy ratio for static and moving sources. *Audio, Speech, and Language Processing, IEEE Transactions on*, 18(7):1793–1805, 2010. 2.3.5

[142] Gautam Narang, Kentaro Nakamura, and Kazuhiro Nakadai. Auditory-aware navigation for mobile robots based on reflection-robust sound source localization and visual slam. In *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on*, pages 4021–4026. IEEE, 2014. 2.3.5

[143] Jani Even, Yaileth Morales, Nagasrikanth Kallakuri, Carlos Ishi, and Norihiro Hagita. Audio ray tracing for position estimation of entities in blind regions. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 1920–1925. IEEE, 2014. 2.3.5

[144] Tomonari Furukawa, Frederic Bourgault, Benjamin Lavis, and Hugh F Durrant-Whyte. Recursive bayesian search-and-tracking using coordinated uavs for lost targets. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 2521–2526. IEEE, 2006. 3.2.1

[145] Makoto Kumon Daisuke Kimoto. Optimization of the ear canal position for sound localization using interaural level difference. In *36th Meeting of Special Interest Group on AI Challenges*, 2012. 4.3, 5.1

[146] Thomas W Sederberg, Jianmin Zheng, Almaz Bakenov, and Ahmad Nasri. T-splines and t-nurccs. In *ACM transactions on graphics (TOG)*, volume 22, pages 477–484. ACM, 2003. 4.3, 5.1

[147] Daisuke Kimoto and Makoto Kumon. On sound direction estimation by binaural auditory robots with pinnae. In *35th Meeting of Special Interest Group on AI Challenges*, 2011. 5.1

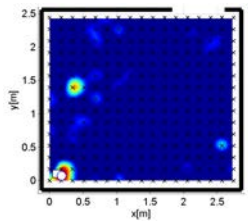
[148] Y. Noda and Makoto Kumon. Sound source direction estimation in the median plane by two active pinnae. In *13th SICE System Integration Division Annual Conference*, 2012. 5.1

[149] Ville Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of Audio Engineering Society*, 45(6):456–466, June 1997. 6.3.2

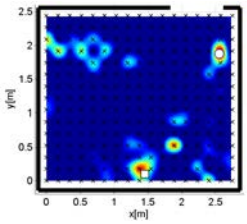
APPENDIX A

A.1 Indoor Acoustic Observation Likelihood

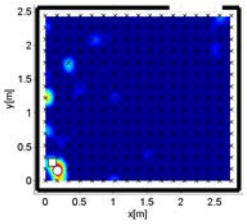
Fused observation likelihoods from Six microphone pair combinations.



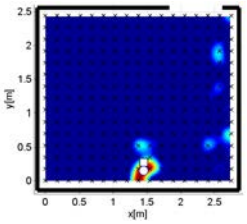
(a) Position = 001



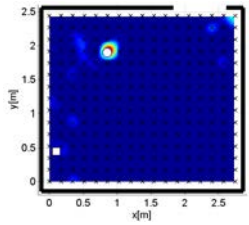
(b) Position = 009



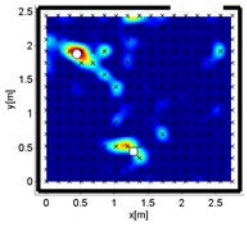
(c) Position = 017



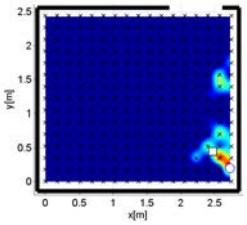
(d) Position = 025



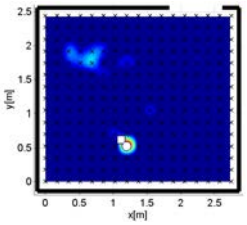
(e) Position = 033



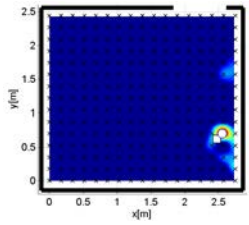
(f) Position = 040



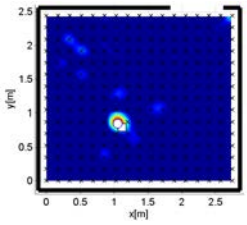
(g) Position = 047



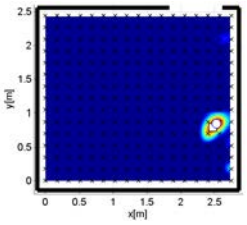
(h) Position = 055



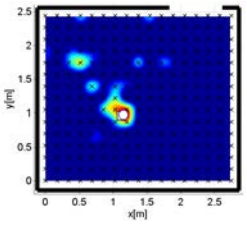
(i) Position = 063



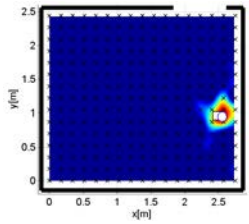
(j) Position = 071



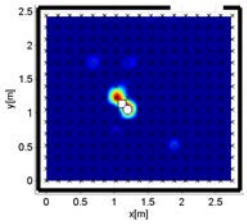
(k) Position = 079



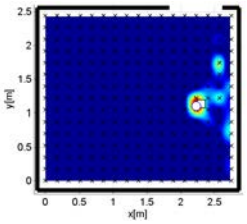
(l) Position = 087



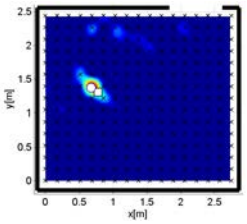
(m) Position = 095



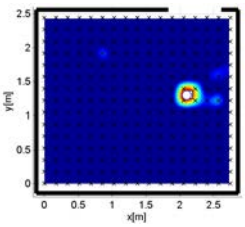
(n) Position = 103



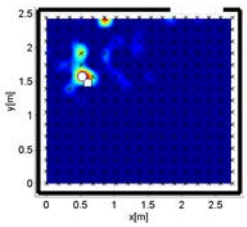
(o) Position = 110



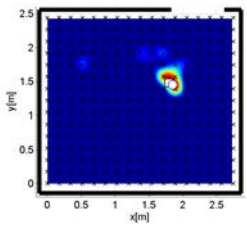
(p) Position = 117



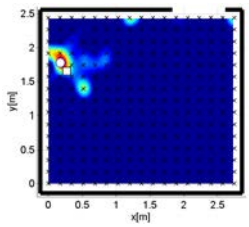
(a) Position = 125



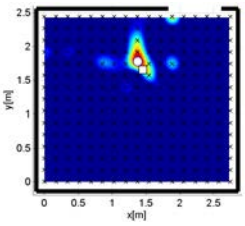
(b) Position = 132



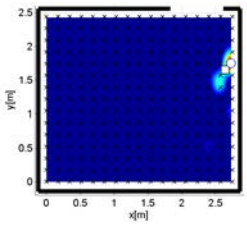
(c) Position = 139



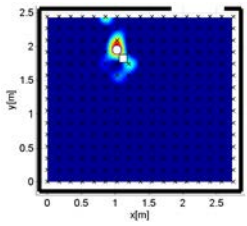
(d) Position = 146



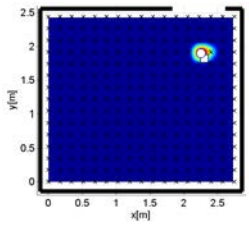
(e) Position = 153



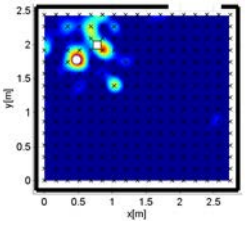
(f) Position = 160



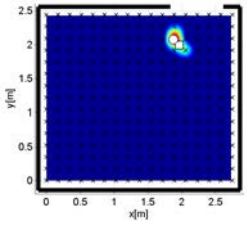
(g) Position = 167



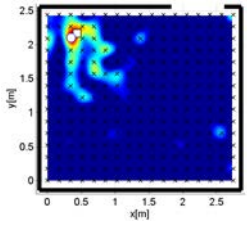
(h) Position = 174



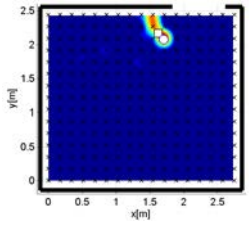
(i) Position = 181



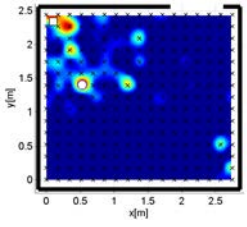
(j) Position = 188



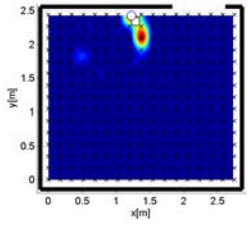
(k) Position = 195



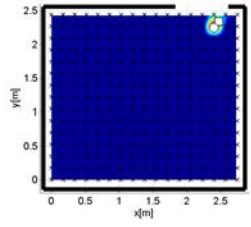
(l) Position = 202



(m) Position = 209



(n) Position = 216



(o) Position = 223